

EXTREME EVENT RISK ASSESSMENT FOR OFFSHORE SYSTEMS
DESIGN AND OPERATION IN HARSH ENVIRONMENTS

By

© Mohammad Hizbul Bahar Arif

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
Faculty of Engineering and Applied Science
Memorial University of Newfoundland

October 2023

St. John's, Newfoundland.

*Dedicated to all the COVID-19 heroes around the world,
who tirelessly strive for the well-being of humanity.*

ABSTRACT

Operations of the offshore systems in harsh environments require better understanding, precise assessment, and effective management of risks. The harsh environmental conditions, such as strong ocean currents, extreme wave conditions, complex subsurface geology, frigid temperatures, and icebergs, exert extreme load on the offshore systems. Environmental factors are interconnected, and when they occur at a higher rate or in extreme conditions, they are likely to cause a catastrophic event. Such scenarios are prone to occur in the current changing conditions of climate. Assessment of extreme loads that may cause a rare event situation is critical to define risk scenarios. This study focuses on the assessment of these extreme event risk scenarios. By integrating extreme load and its likelihood of occurring, this research investigates the current state of knowledge in extreme event risk analysis. The extreme load consideration task considers three dominating aspects: stationary and non-stationary conditions; univariate and multivariate analysis; and dependence of the variables. This study also focuses on the flexible risk-based design methodology that integrates the traditional Extreme Value Theory (EVT) with climate change.

The key environmental parameters considered in this study are iceberg speed, wind speed, and wave height. The developed methodologies use the above parameters from the Atlantic Continental Shelf, specifically the Flemish Pass basin, Grand Bank, and the Jeanne d'Arc basin. Due to limited data for certain environmental phenomena, such as large iceberg data in the Flemish Pass basin, the iceberg load assessment problem is treated as a rare event scenario. Traditional methods, including Peak Over Threshold (POT) based Generalized Pareto Distribution (GPD) and Block Maxima (BM) based Generalized Extreme Value (GEV), were found to be inadequate to capture the present-day extreme characteristics in the rare event cases. As an alternative, this study proposes and validates the use of POT-based Heavy Right Tail Distribution (HRTD) for iceberg

load analysis at the Flemish Pass basin. The research also observes that Maximum Likelihood Estimator (MLE) provides a biased estimate for model parameter estimation in rare event scenarios, whereas the Hill, SmooHill, and Bayesian approaches offer better estimates. The methodology is extended to multivariate settings to capture extreme dependencies using extreme value copula function for investigating rare event risk profiles. The proposed low-resolution risk profile methodology offers a more efficient and cost-effective alternative to computationally expensive numerical models in the offshore domain. Climate change is observed to have an impact on the correlation between various environmental factors, including wind speed and wave height. Because of climate change, 100-year events are becoming more frequent. Consequently, the study adopts a 1000-year time frame to adjust for the increasing frequency of 100-year events under the influence of climate change, enabling predictions beyond standard lifetimes. The conditional return level function is utilized to construct rare events return level predictions under climate change threats. Finally, a non-stationary process is considered to generate a dynamic risk profile. Outcomes of this research provide a clear understanding of how climate change affects the Newfoundland offshore region. By incorporating predicted extreme loads and their likelihood of occurring, the traditional EVT-based methodologies are combined with adaptable risk-based design methodologies. The proposed dynamic, flexible, and small-scale ($0.1^{\circ} \times 0.1^{\circ}$ latitude/longitude grid) risk assessment methodology aids in offshore design decision-making for safer design and operation.

ACKNOWLEDGMENTS

I wish to extend my heartfelt appreciation to Dr. Faisal Khan (supervisor), Dr. Salim Ahmed (co-supervisor), and Dr. Syed Imtiaz (co-supervisor) for their invaluable guidance, unwavering support, and remarkable patience throughout my PhD journey. Their extensive knowledge has significantly enriched both my academic research and daily life. Furthermore, I am deeply grateful to Dr. Veeresh Gadag for technical advice and assistance with my research work. I would also like to express my thanks to every member of the Process Engineering department and the Centre for Risk, Integrity, and Safety Engineering (C-RISE) for their contributions and cooperation.

I am thankful for the financial support provided by the Natural Sciences and Engineering Council of Canada through the Discovery Grant and the Canada Research Chair (CRC) Tier 1 Program in offshore safety and risk engineering.

Above all, I recognize that none of this would have been possible without the unwavering love, patience, and support of my family. Throughout the years, my beloved wife, parents, and immediate family have been a constant source of love, care, strength, and encouragement. I am deeply grateful to each one of you for helping me reach this milestone, as I truly couldn't have ever done it without you all.

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	x
ABBREVIATIONS	xiii
SYMBOLS	xv
1 INTRODUCTION	1
1.1 Research Context	1
1.2 Research problem and opportunity	3
1.3 Research Objective	6
1.4 Contribution and Novelty	7
1.4.1 Task 1: Univariate extreme model (frequentist approach) and Iceberg collision risk analysis	9
1.4.2 Task 2: Climate change and a tail based univariate extreme model (Bayesian approach)	10
1.4.3 Task 3: Bivariate extreme model and extreme load analysis for offshore system design	11
1.4.4 Task 4: Risk-based non-stationary approach (univariate).	12
1.5 Research Outcomes	13
1.6 Thesis overview	14
1.7 Authorship contribution Statement	19
2 Literature review	20
2.1 Introduction	20
2.2 Bibliometric Analysis	21
2.3 Extreme Event modeling	26
2.4 Risk analysis	30
2.5 Research gap/opportunity	31
3 RARE EVENT RISK ANALYSIS – APPLICATION TO ICEBERG COLLISION	35

Preface	35
Abstract	35
3.1 Introduction	36
3.2 Methodology	39
3.2.1 Defining a rare event	39
3.2.2 The proposed framework	40
3.2.2.1 Phase 1: outlier-based approach (traditional)	41
3.2.2.2 Phase 2: tail based (non-traditional) approach	44
3.2.3 Uncertainty estimates for model parameters and return level	45
3.2.4 Risk analysis	47
3.3 Case study (Iceberg extreme speed estimation for the Flemish Pass basin)	49
3.3.1 Descriptive statistics	50
3.3.2 Result and discussion	54
3.3.2.1 Threshold and block maxima	54
3.3.2.2 Distributions parameters estimate	55
3.3.2.3 Return level estimate	56
3.3.2.4 Goodness of fit test	57
3.3.2.5 Risk estimation	57
3.4 Conclusions	59
4 EVOLVING EXTREME EVENTS CAUSED BY CLIMATE CHANGE: A TAIL-BASED BAYESIAN APPROACH FOR EXTREME EVENT RISK ANALYSIS	62
Preface	62
Abstract	62
4.1 Introduction	63
4.2 Methodology	67
4.2.1 The proposed framework	67
4.2.2 Heavy Right Tail Distribution (HRTD)	68
4.2.3 Bayesian inference	69
4.2.4 Risk estimation	72
4.3 Case Study	73

4.3.1	Descriptive statistics	74
4.3.2	Result and Discussion	77
4.3.2.1	Threshold selection	77
4.3.2.2	Parameter estimate (Bayesian inference)	78
4.3.2.3	Return level and Risk estimate	82
4.4	Conclusions	84
5	A GENERALIZED FRAMEWORK FOR RISK-BASED EXTREME LOAD ANALYSIS IN OFFSHORE SYSTEM DESIGN	86
	Preface	86
	Abstract	86
5.1	Introduction	87
5.2	Methodology	90
5.2.1	Dependency measure	92
5.2.2	Marginal distributions	94
5.2.3	The Copula Model	95
5.2.4	Goodness of fit test	96
5.2.5	Extreme Bivariate Distribution	97
5.2.6	Parameter estimation	98
5.2.7	Consideration of Return level	99
5.2.8	Risk estimation	100
5.2.9	Design wind speed and wave height	102
5.3	The Case Study	102
5.3.1	Descriptive statistics	103
5.3.2	Marginal distributions (GEV fit) and Copula function	105
5.3.3	Risk estimation	109
5.4	Conclusions	111
6	EXTREME WIND LOAD ANALYSIS USING NON-STATIONARY RISK BASED APPROACH	114
	Preface	114
	Abstract	114
6.1	Introduction	115

6.2	Methodology	117
6.2.1	Data processing and findings trend	118
6.2.2	Extreme Distributions	119
6.2.3	GEV fit: Optimal block size and parameter estimation	120
6.2.4	Goodness of fit test	121
6.2.5	Risk analysis	122
6.2.6	Uncertainty estimate	123
6.3	Case study	123
6.3.1	Descriptive statistics	124
6.3.2	Fit distributions and model validation	126
6.3.3	Risk analysis	128
6.4	Conclusions	130
7	SUMMARY, CONCLUSIONS AND FUTURE RESEARCH	132
	RECOMMENDATIONS	
7.1	Novelty and contribution	133
7.2	Conclusions	134
7.2.1	Univariate risk assessment methodology (frequentist modeling approach)	134
7.2.2	Univariate risk assessment methodology (Bayesian modeling approach)	135
7.2.3	Joint extreme events load analysis and risk assessment	136
7.2.4	Extreme event load analysis and a non-stationary risk-based approach	137
7.3	Future research opportunity	137
	References	142

LIST OF TABLES

Table 3.1.	Descriptive statistics: Flemish Pass basin iceberg data (2002-2015) categorize in three different size (Small, medium and large).	52
Table 3.2.	Distribution parameters estimate. Corresponding 95% confidence intervals for model parameters are in parenthesis.	55
Table 3.3.	Iceberg exceedance occurrence probability, extreme speed (km/hr), Kinetic Energy (KE) in Mega Jules (MJ) and Risk (MJ/year) corresponding to the three different icebergs' weight (Megatons, MT). Corresponding 95% confidence intervals for return periods are in parenthesis (Lower Limit, Upper Limit).	58
Table 4.1.	Descriptive statistics: Jeanne d'Arc basin iceberg speed data (2002-2017).	76
Table 4.2.	HRTD parameter (α) and its range range estimate through Hill and SmooHill estimator.	80
Table 4.3.	HRTD parameter (α) and Standard error (Bayesian inference).	81
Table 4.4.	Iceberg exceedance occurrence probability, extreme speed (km/hr), Kinetic Energy (KE) in Megajoules (MJ) and Risk (MJ/year) corresponding to the three different thresholds (km/hr).	82
Table 5.1.	Data statistics: Hourly wind speed (m/s) and wave height (m) data for three grid locations (with extremes) for the Flemish Pass basin. The most extreme grid locations are highlighted as bold.	104
Table 5.2.	Distribution parameter values for Case I and Case II, considering time-space maximum data. Corresponding 95% confidence intervals for model parameters are in parenthesis.	107
Table 5.3.	Copula parameter value and goodness of fit test statistics for Case I and Case II.	107
Table 5.4.	Joint exceedance occurrence probability, conditional wind speed (m/s) ($x y>14$), conditional wave height (m) ($y x > 28$), and Impact energy (MJ) and Risk (MJ/year).	110
Table 6.1.	Stationary and non-stationary GEV models.	119

Table 6.2.	Descriptive data statistics: hourly wind speed (m/s) data at the Grand Banks (1988-2015).	
Table 6.3	Different GEV models are provided, along with parameter values and test statistics. In parenthesis, the standard errors are indicated.	126
Table 6.4.	Stationary risk profile: Different expected return levels are used.	129
Table 6.5.	Non-Stationary risk profile: Risk in a certain year based on different return levels.	129

LIST OF FIGURES

Figure 1.1.	Proposed Extreme load assessment methodology divided into four different sub methodologies in offshore Engineering domain.	7
Figure 1.2.	Thesis outline.	18
Figure 2.1.	Cone diagram for rare event modeling for the period 1968-2018.	21
Figure 2.2.	Framework for Bibliometric analysis.	22
Figure 2.3.	Modeling of extreme events' contribution across different geographical locations.	23
Figure 2.4.	Density plot: Extreme events modelling contribution across different geographical locations.	24
Figure 2.5.	The cooperation network of notable authors from peer-reviewed publications in the field of extreme events.	24
Figure 2.6.	Cooperation network of authors of peer reviewed publications in the domain of process accident risk analysis.	25
Figure 2.7.	Cooperation network of affiliated institution in the domain of process accident risk analysis.	26
Figure 3.1.	The proposed framework for risk analysis.	41
Figure 3.2.	Iceberg frequency distribution at Flemish Pass basin (2002-2015).	50
Figure 3.3.	Iceberg temporal (2002 to 2015) distribution (small, medium and large).	51
Figure 3.4.	Iceberg speed distribution (maximum vs average) over the period 2002 to 2015 (merge small, medium and large iceberg).	51
Figure 3.5.	Box plot for the small, medium, and large iceberg. The highlighted cyan circles represent rare events (extreme outliers).	53
Figure 3.6.	Speed distributions for different size of icebergs (2002 -2015). The iceberg sizes are labeled on the top right corner of the plot.	53
Figure 3.7.	Mean residual life plot. The vertical dotted blue line indicates the chosen threshold in all three cases. The units of X-axis and Y-axis are km/hr.	54

Figure 3.8.	The return levels (GPD, GEV, Hill, and SmooHill) plot for three different iceberg cases. The dotted horizontal lines indicate the most extreme outliers. Data tables (return periods, extreme iceberg speed, km/hr) added.	56
Figure 4.1.	The proposed framework for heavy tail event risk analysis	67
Figure 4.2.	Bayesian inference flow chart (MCMC, Metropolis-Hasting algorithm).	70
Figure 4.3.	Iceberg frequency distribution at the Jeanne d'Arc basin (2002-2017).	74
Figure 4.4.	Iceberg (Small, Medium and Large) temporal (2002 to 2017) distribution.	75
Figure 4.5.	Large Iceberg speed (maximum vs. average) over the period 2002 to 2017.	75
Figure 4.6	Box plot for the large iceberg. All data points above the red line are rare events (extreme outliers).	76
Figure 4.7.	Histogram: The large iceberg speed distribution (2002 -2017).	77
Figure 4.8.	Normal Q-Q plot. The red vertical lines are showing the threshold range.	77
Figure 4.9.	Threshold selection procedure (mean excess vs iceberg speed). The number of excesses is labeled at the top of the plots. Threshold values are highlighted with vertical lines.	78
Figure 4.10.	The posterior distribution for the HRTD fit for three different threshold values. The threshold values are leveled on top of each plot.	80
Figure 4.11.	Bayesian inference: The cumulative distribution function fit comparison for three different thresholds. The threshold values are labelled on top of each plot.	81
Figure 5.1.	The proposed methodology for bivariate extreme event risk analysis.	93
Figure 5.2.	A simple topside representation of a random platform.	102
Figure 5.3.	Wind and wave correlated distribution for two grid locations. The grid location is levelled on top of each plot. The horizontal and vertical lack ines indicate the amount of data in the area of $x > 28 \text{ m/s}$, $y > 14 \text{ m}$.	105

Figure 5.4.	Q-Q and density plot of wind and wave extreme spatial-temporal distribution (empirical vs model).	107
Figure 5.5.	Correlation map plot (Empirical vs Model): yearly maximum wind speed vs. wave height. The bottom plot shows the error (Model-Empirical) correlation.	108
Figure: 5.6.	Map plot (50 years vs. 100 years): Predicted extreme joint wind speed, wave height and its corresponding risk.	110
Figure 6.1.	The proposed framework for extreme wind load risk analysis.	117
Figure 6.2.	(Left) The frequency distribution of wind speed data (1988-2015). (Right) the box plot, showing data profile and extreme outliers are separated by the red lines.	124
Figure 6.3.	Annual wind speed distribution. The observational data vs Hindcast model data.	124
Figure 6.4.	Different return levels estimate (Stationary vs non-stationary)	126

ABBREVIATIONS

AIC	Akaike's information criterion
BIC	Bayesian information criterion
EVT	Extreme Value Theory
GEV	Generalized Extreme Value
GPD	Generalized Pareto Distribution
MLE	Maximum Likelihood Estimator
POT	Peak Over Threshold
HRTD	Heavy Right Tail Distribution
CDF	Cumulative distribution function
FEA	Finite element analysis
CFD	computational fluid dynamics
PWM	Probability-Weighted Moments
MCMC	Markov Chain Monte Carlo
BM	Block Maxima
IQR	Interquartile Range
IID	independent and identical distribution
RP	Return period
KE	Kinetic Energy
IIP	International Ice Patrol
MED	Medium
LG	Large
GR	Growlers
BB	berg bits
VLG	very large
RAD	Radian
MJ	Mega Joule
MT	Mega Tone
GH	Gumbel–Hougaard copula

Q-Q plot

Quantile-Quantile plot

P-P plots

Probability-Probability plot

GMLE

Generalized Maximum Likelihood Estimator

RMSE

Root Mean Square Error

SYMBOLS

A	Wind projected area
C	Copula
C	Extreme Event Consequence
CS	Drilling derrick shape coefficient
D	Deviation statistic
F	Wind force
M	Mass
P	Extreme Event Occurrence Probability
$Q1$	1 st Quantile
$Q3$	3 rd Quantile
R	Extreme Event Risk
u	Threshold
U	Wind speed
V	Speed
α	Tail index
ρ	Density
k	Model Shape parameter
β	Mass coefficient
μ	Model Location parameter
σ	Model Scale parameter
τ	Kendall's tau

1. INTRODUCTION

1.1 Research Context

Extreme climatic conditions, such as strong winds, high waves, ocean currents, icebergs, and low temperatures, significantly influence the design, operation, maintenance, and risk management of offshore systems and marine activities. These conditions can lead to challenging offshore environments where unexpected or unpredictable climatic loads may damage the integrity and functionality of offshore systems. Such extreme and infrequent events have been characterized in past research by various names, including "extreme events" (Castillo et al., 2005), "black swan events" (Taleb, 2007), "rare events" (Rubino and Tuffin, 2009), and "heavy tail events" (Resnick, 2007). Despite the different names, they all refer to infrequent events with potentially significant impacts or consequences. These events are characterized by their exceptional intensity, duration, or magnitude, surpassing what is typically observed in the given domain. Their manifestations span various fields, including health science (as seen with the COVID-19 pandemic), weather and climate (hurricanes, heatwaves, extreme winds, low temperatures, etc.), natural disasters (such as earthquakes and tsunamis), financial markets (like stock market crashes), accidental events at offshore engineering domain (Deepwater Horizon disaster, Gulf of Mexico, 2010; Ocean Ranger oil rig disaster, Canadian Atlantic Ocean, 1982 etc.) and other complex systems. The understanding and management of these extreme events are crucial for ensuring the safety and resilience of offshore systems and marine activities, as they pose unique challenges due to their infrequency and severity. Proper risk assessment and preparedness strategies are vital to mitigate the potential impacts of such extreme events across different domains. Handling such extreme

events is extremely challenging, especially in the offshore domain, where operations take place in remote and harsh environments characterized by strong currents, uneven surfaces, freezing temperatures, and icebergs. Different offshore regions also present unique challenges. For instance, the Newfoundland offshore region, including areas like the Grand Bank or Flemish Pass basin, may have less extreme wind and wave conditions but poses challenges with icing and iceberg hazards (MANICE Canadian Ice Service, 2005). Icebergs typically follow predictable/usual paths, but under extreme environmental loads/harsh environment, unexpected changes in their trajectory can result in significant impacts when they collide with engineering systems. Natural incidents like these are beyond control and cannot be prevented. Nevertheless, understanding current extreme events is crucial for risk assessment, planning, and decision-making to ensure safer offshore operations. Offshore operators can optimize design costs and achieve process safety by adopting modeling outputs that offer precise predictions of environmental event loads and their occurrence probabilities through improved modeling approaches that incorporate climate change trends and current extremeness. Most of the existing approaches for extreme event modeling are based on Extreme Value Theory (EVT); using a 100-year return level is a common practice to estimate design environmental parameters (e.g., Castillo et al., 2005; Levine, 2009; Das et al., 2016, etc.). However, it is essential to question whether the existing EVT/100-year return level-based methodologies are sufficient to capture the current extremeness. If not, what alternative approaches or return levels are more suitable for determining the engineering design criteria for safer offshore operations? Furthermore, how can modeling results be effectively utilized in offshore risk assessments plans? This research was conducted in response to these challenges. It placed significant emphasis on extreme event modeling and integrated the modeling

outcomes into a risk-based framework. Towards this objective, the primary challenges lie in three key areas:

- Formulating a robust mathematical representation of extreme event scenarios. This involves developing a comprehensive framework to accurately describe a extreme event for the offshore engineering domain.
- Conducting extreme event modeling, which encompasses precise load estimation and accurate estimation of occurrence probabilities. This step is crucial for understanding the magnitude and likelihood of extreme events to enhance risk assessments.
- Integrating the impact of regional climate change into the modeling process. Recognizing the evolving nature of climate patterns and their influence on extreme events is essential to ensure that the risk assessment methodologies are up-to-date and capable of accounting for changing environmental conditions.

By addressing these challenges, this research endeavors to establish an effective risk-based approach that captures the complexities of extreme events in the offshore engineering domain, ultimately leading to enhanced safety, resilience, and decision-making in offshore operations.

1.2 Research problem and opportunity

Testing in extreme conditions is recommended for developing any engineering system. Designing an extreme state and considering it in the design of a system or its operation is not a straightforward task. Below are few key challenges in this area in the prospect of extreme event modeling and risk assessment methodology:

- Data availability: Obtaining reliable and sufficient data for risk analysis in offshore engineering is challenging. Limited operational history of offshore structure and scarcity of observed extreme events make it difficult to gather comprehensive data for accurate risk assessment.
- Uncertainty and variability: Offshore engineering involves dealing with uncertainties and variability associated with environmental conditions. In addition, in the modeling prospect, there are uncertainties in the input parameters, and data variability. Quantifying and managing those uncertainties is crucial for reliable risk analysis.
- Computational complexity: Due to low probability, extreme event modeling needs an efficient computational methodology. Offshore engineering risk analysis often requires computationally intensive modeling approach, due to low probability of occurrence; traditional simulation technique might be time consuming and may not be feasible for long-scale offshore system.
- Non-stationary: The statistical properties of environmental conditions or other risk factors may change over time due to climate variability; so, accounting for non-stationarity is essential for risk analysis. Identifying and incorporating these changes into the models pose additional challenges.
- Multivariate dependencies: Offshore systems are subject to complex dependencies and interactions among various factors, such as environmental conditions, structural integrity and operational factors. Developing multivariate models by capturing the dependencies in different variables is a significant research challenge.
- Validation and verifications: validating and verifying extreme event model can be challenging due to data scarcity.

- Risk assessment and decision making: Incorporating extreme event modeling into risk assessment and decision-making process is an ongoing challenge. Quantifying the probabilities and consequences of extreme events and integrating them into risk analysis framework is essential for making an adequate decision regarding the design, operation, and maintenance of offshore structure.

All the above challenges are more complicated when modeling extreme events in the offshore engineering domain (compared to other domains, for example onshore engineering, or financial domain). The strong currents, uneven surfaces, extremely low temperatures, staggering depths, and iceberg threats are making offshore engineering systems/equipment operations more complicated. In addition, climate change/global warming issues make the task to be further complicated. We cannot ignore the recent patterns in weather related phenomena. For example: the recent European heat wave (Climate signal beta, European heat wave, 2019; repeating in 2023), the global sea level was 3.2 inches in 2018 above the 1993 average (NOAA, 2019), Greenland's loss of 532 billion tons of ice in a record melt in 2019 (CBC report, 2020); and global temperature prediction between 2030 and 2052, expected to rise by 1.5°C, compared to the pre-industry era, if it continues to rise at the current rate (Ogunbode et al., 2020). This being the case, then what will future Newfoundland offshore domain look like? The above poses significant challenges to researchers interested in relevant fields. However, the existing literature does not adequately address many research questions necessary to develop a framework to model extreme events considering the recent phenomena. This work is an attempt to answer a list of questions focusing on the following:

- Is the existing EVT approach applicable/capable to capture present extreme characteristics in extreme event modeling contexts under climate change for offshore engineering domain?
- Which return level adaptation is more appropriate for monitoring and managing offshore system design/upgrading existing facilities?
- Does climate change have any impact on the correlation in environmental parameters? If this is the case, then how to deal with bivariate or multivariate cases in the extreme load analysis?
- How to generate a computationally efficient “small-scale risk profile, say $(0.1^0 \times 0.1^0$ latitude/longitude grid)” for offshore operations safer?

This research takes the opportunity to address the above concerns and challenges. It seeks to provide valuable insights into improving the modeling of extreme events in offshore engineering, incorporating the effects of climate change. By answering these critical questions, the research aims to contribute to the development of more reliable risk assessment and management strategies, ultimately enhancing the safety and resilience of offshore operations.

1.3 Research Objective

The extreme load assessment problem is considered in several sub problems in the prospect of offshore engineering domain. To meet the research objectives, the following tasks are listed in Figure. 1.1 are in consideration.

The study achieves its objective of enhancing the understanding and prediction of extreme events in the offshore engineering domain, contributing to more effective risk assessment and decision-making for safer offshore operations.

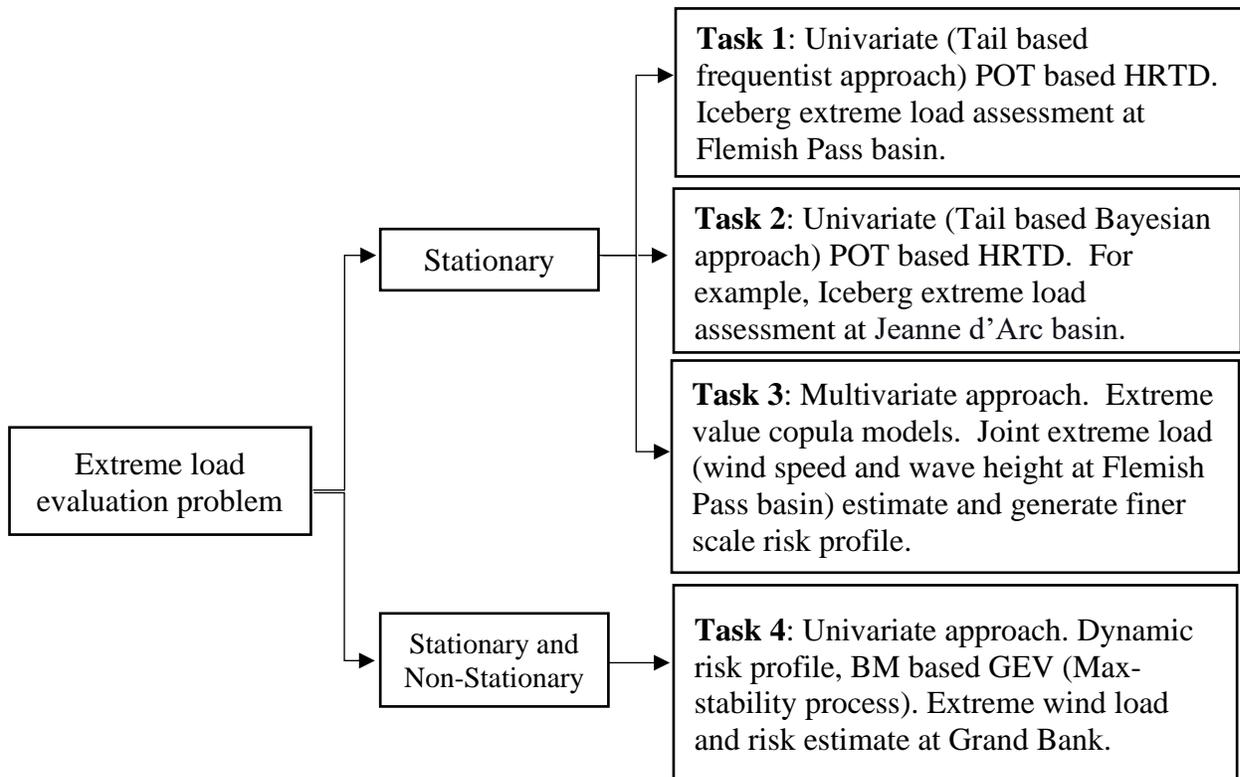


Figure 1.1. Proposed Extreme load assessment methodology divided into four tasks for offshore Engineering domain.

1.4 Contribution and Novelty

This section provides a detailed overview of the findings, challenges, and contributions of this research work. Notably, the distinctiveness of this study lies in its development of a mathematical representation for extreme events in the offshore engineering domain, incorporating heavy tail modeling concepts to address climate change issues, and employing an appropriate return level

function. By adopting a risk-based framework, the research proposes a flexible risk profile tailored for safer offshore operations, setting it apart from existing methodologies. Extensive investigation into extreme modeling literature across various fields, including engineering, finance, economics, and hydrology, has enabled the development of an appropriate methodology/model capable of capturing present-day extremeness. The research diligently explores each identified contribution in the relevant chapters of the thesis, presenting a comprehensive understanding of the novel approaches and their applications. Key contributions and novelties of this study are briefly summarized as follows:

- **Mathematical representation for extreme events:** The research establishes a robust mathematical framework for characterizing extreme events in the offshore engineering domain. This provides a solid foundation for accurate modeling and analysis of extreme events.
- **Heavy tail modeling and climate change:** By incorporating heavy tail modeling concepts, the study addresses the challenges posed by climate change and its impact on extreme events, and uncertainty estimate. This allows for a more nuanced understanding of extreme event occurrences in a changing climate.
- **Appropriate return level function:** The research proposes an appropriate return level function for estimating extreme event probabilities. This selection ensures the reliability and accuracy of risk assessments.
- **Adoption of risk-based framework:** By adopting a risk-based approach, the research offers a comprehensive and practical methodology for assessing offshore risks. The framework facilitates informed decision-making and proactive risk management.

Through a review of literature in various disciplines, this research successfully develops a methodology that effectively captures and quantifies present-day extremeness, providing valuable insights into risk assessment and management in the offshore engineering domain. Each contribution is expanded upon in the relevant chapters, showcasing the exploration and innovative solutions presented in this study.

1.4.1 Task 1: Univariate extreme event model (frequentist approach) and iceberg collision risk analysis

Developed a flexible risk analysis approach, and implemented at the Flemish Pass basin for iceberg collision risk analysis, yielding several notable findings and contributions:

- **Mathematical representation of extreme events:** This research adopts Devore's (Devore, 2011) definition of "outliers" and "extreme outliers" to provide a precise mathematical representation of extreme events; originally proposed by Tukey's (Tukey, 1977). The "Extreme outliers" criterion is considered as a "Extreme event" criterion for the offshore engineering domain.
- **Tail-based approach:** While outlier-based Generalized Extreme Value (GEV) or Generalized Pareto Distribution (GPD) methods are popular in extreme event analysis across various domains, this study reveals their limitations in capturing present extreme characteristics when dealing with small data sets (as is the case with extreme events). As an alternative, a tail-based approach is proposed to overcome these limitations.
- **Distribution parameter estimation:** To address the challenges of small data sets, the Hill and SmooHill estimators are utilized for the estimation of distribution parameters. These estimators serve as alternatives to the commonly used Maximum Likelihood Estimator (MLE) for cases with limited data.

- **Threshold selection and model validation:** Threshold selection is a critical step in extreme event modeling. Mean residual life plot, Normal Q-Q plot, and parameter stability plot are employed to guide threshold determination. Model validation is performed using various numerical and graphical evidence, including log-likelihood values, Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC), Q-Q plot, and return levels plot.
- **Uncertainty estimation:** A method for uncertainty estimation is proposed for model parameters and return levels, akin to the concept of parametric bootstrap. This allows for a more comprehensive understanding of the uncertainties associated with the extreme event model.

Finally, this research offers valuable insights and innovative methodologies to improve the analysis and modeling of extreme events, particularly in the context of iceberg risk collision at the Flemish Pass basin. By addressing the challenges specific to small data sets and providing robust model validation and uncertainty estimation techniques, this study significantly contributes to the field of extreme event analysis in offshore engineering.

1.4.2 Task 2: Climate change and a tail based univariate extreme event model (Bayesian approach)

The study developed a univariate risk-based extreme modeling (Bayesian) approach and applied it to Iceberg risk analysis in the Jeanne d'Arc basin. The key findings and contributions of the research are outlined as follows:

- **Implementation of Peak Over Threshold (POT) based Heavy Right Tail Distribution (HRTD):** The study utilized a POT-based HRTD, like the POT-based Generalized Pareto Distribution (GPD), for modeling extreme events.

- Utilization of Bayesian approach for tail index estimation: Bayesian techniques were employed to estimate the tail index, providing a more robust parameter estimation process to capture extreme events effectively.
- Application of Hill and SmooHill Estimators for parameter range estimation: The Hill and SmooHill estimators were used to determine parameter ranges, facilitating the Bayesian process and ensuring an accurate fit and return level estimate.
- Incorporation of climate change impact into risk analysis: The research integrated the impact of climate change into the risk analysis methodology, enabling a comprehensive assessment of risks under changing environmental conditions.
- Generation of risk profiles with multiple threshold Values: Rather than using a fixed threshold, the study generated risk profiles for three different threshold values, allowing for a more flexible and comprehensive analysis of extreme events.
- Justification of threshold selection: The selection of thresholds was justified using various plots, including Normal Q-Q plot, Mean residual plot, and parameters stability plot. Additionally, the distribution CDF was directly compared with data CDF to validate the appropriateness of the proposed distribution.

By implementing this novel approach and providing justifications for its key components, the research contributes to a deeper understanding of extreme event modeling and risk analysis in the offshore domain.

1.4.3 Task 3: Bivariate extreme load model: its impact analysis on offshore system design

This research proposes a bivariate extreme model for estimating the joint wind and wave height risks at Flemish Pass basin. The key findings and contributions of the study are outlined below:

- Examination of climate change effects: The research investigates the impact of climate change on the extreme behavior of wind and wave height, as well as their correlation. This analysis provides valuable insights into how climate change influences extreme events in the offshore environment.
- Small-scale risk profile generation: The research introduces a unique approach to risk profile generation on a small scale, using a $(0.1^0 \times 0.1^0)$ latitude/longitude grid. This method sets it apart from traditional offshore extreme load analysis methods (references 43-45). The small-scale grid allows for a more localized and detailed assessment of risks.
- Innovative distribution selection: Traditional goodness-of-fit tests, such as AIC, BIC, and correlation map plot, are complemented by the application of the Max stable process theorem and extreme value copula for distribution selection. This novel approach enhances the accuracy and reliability of the chosen distributions.
- Risk assessment at Flemish pass basin: The research findings reveal that the smaller grid areas at Flemish Pass basin exhibit lower risks compared to higher grid areas. Additionally, climate change has led to approximately 30% less correlation between wind speed and wave height. Over the past three decades, wind speed has increased by approximately 19%, and wave height has increased by about 8%.

By presenting these contributions, the research significantly advances the understanding of bivariate extreme modeling for risk estimation and provides valuable insights into the impact of climate change on extreme events in the Flemish Pass basin area. The innovative methods and localized risk profile generation make this study distinct from traditional offshore extreme load analyses.

1.4.4 Task 4: Risk-based non-stationary approach (univariate).

This research introduces and applies a dynamic risk profile for extreme wind risk analysis in the Grand Banks region. The key findings and contributions of the study are outlined below:

- Comparison of stationary and non-stationary extreme predictions: The research conducts a comparative analysis of stationary and non-stationary extreme predictions. By examining both scenarios, the study gains insights into the influence of temporal variations on extreme wind events.
- Incorporation of climate change impact in return levels: The research incorporates meaningful considerations of climate change in return levels. This integration enables a more accurate assessment of extreme wind risks considering changing environmental conditions.
- Use of observational wind speed data: The study utilizes observational wind speed data, which exhibits a similar trend to Hindcast model data. This data comparison enhances the reliability and validity of the risk analysis.
- Similar increasing data trends in non-stationary return levels: The non-stationary return levels demonstrate consistent increasing trends across different levels. This observation further supports the understanding of the impact of temporal changes on extreme wind events.

By proposing and implementing a dynamic risk profile, this research advances the understanding of extreme wind risk analysis in the Grand Banks region. The comparative analysis between stationary and non-stationary predictions, coupled with the consideration of climate change impact, provides valuable insights into the dynamic nature of extreme wind events. The use of

observational wind speed data and the consistent trends observed in non-stationary return levels contribute to the robustness and relevance of the risk assessment methodology.

1.5 Research Outcomes:

The research conducted in this thesis has resulted in the publication of four papers in peer reviewed journals:

- Task 1: Rare event risk analysis – application to iceberg collision. M. Arif, F. Khan, S. Ahmed, S. Imtiaz, *Journal of Loss Prevention in the Process Industries*, 66 (July, 2020) 104199. <https://doi.org/10.1016/j.jlp.2020.104199>
- Task 2: Evolving extreme events caused by climate change: A tail-based Bayesian approach for extreme event risk analysis. M. Arif and F. Khan, S. Ahmed, S. Imtiaz. *Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* (February 2021). <https://doi.org/10.1177/1748006X21991036>
- Task 3: A generalized framework for risk based extreme load analysis in offshore system design. M. Arif and F. Khan, S. Ahmed, S. Imtiaz. *Journal of offshore mechanics and Arctic engineering*. 145(2): 021701, 2022. DOI: <https://doi.org/10.1115/1.4055553>
- Task 4: Extreme wind load analysis using non-stationary risk-based approach. M. Arif and F. Khan, S. Ahmed, S. Imtiaz. *Journal of Safety in Extreme Environments*, 4, 247–255, 2022. DOI:[10.1007/s42797-022-00064-2](https://doi.org/10.1007/s42797-022-00064-2)

1.6 Thesis overview

This thesis is presented in manuscript format and consists of seven chapters. The research findings are reported in four chapters, namely Chapter 3 to Chapter 6, and have been published in peer-reviewed journals. The introductory aspects, literature review, and conclusive remarks are covered in Chapters 1, 2, and 7, respectively. The current chapter provides a comprehensive overview of

the research scope, background, and contributions. The remaining chapters of this thesis encompass the following contents:

Chapter 2

This chapter serves as a comprehensive review and analysis of existing research in the areas of extreme event modeling, offshore engineering, and risk assessment. It carefully evaluates the strengths and weaknesses of the current literature, providing valuable insights into the state of the field. Through this assessment, gaps in past research are identified, highlighting areas that require further exploration and providing opportunities for current research to make significant contributions to the domain. By synthesizing and critically examining the existing body of knowledge, this chapter lays the foundation for advancing the understanding and application of extreme event modeling in the context of offshore engineering and risk assessment.

Chapter 3

This chapter introduces a novel heavy right tail modeling approach designed to effectively capture current extreme characteristics in a univariate context. Unlike traditional extreme value theory-based techniques that struggle to capture present-day extremeness, this approach proves to be more suitable. The Hill and SmooHill-estimators are employed for precise parameter estimation, enhancing the accuracy of the model. The modeling outcomes are subsequently integrated into a risk assessment methodology specifically tailored for offshore engineering. This innovative methodology offers flexibility and adaptability, providing valuable insights for risk assessment.

This chapter's significant contributions were recognized and resulted in its publication in the *Journal of Loss Prevention in the Process Industries* in July 2020. By addressing the limitations of existing techniques and proposing a more effective approach, this research makes a noteworthy advancement in the field of risk assessment for offshore engineering.

Chapter 4

This chapter introduces an innovative univariate flexible risk-based methodology that incorporates a Bayesian approach for model parameter estimation. By utilizing Bayesian techniques, the model parameter estimation process becomes more robust and reliable, leading to improved risk assessments for offshore engineering. Furthermore, the methodology considers the issue of climate change and its impact on risk assessment. By integrating climate change considerations, the risk assessment becomes more comprehensive and adaptive to changing environmental conditions.

The significance of this research was recognized, and the chapter was published in the *Journal of Risk and Reliability* in February 2021. Through its novel approach and inclusion of climate change factors, this chapter makes a valuable contribution to the field of risk assessment for offshore engineering.

Chapter 5

This chapter introduces a novel multivariate methodology that encompasses risk-based approaches presented in Chapter 3 and Chapter 4. The research delves into the examination of correlations and dependencies between extreme events, considering the impact of current climate change. Joint modeling of extreme events is a key feature of this methodology, which enhances the accuracy and comprehensiveness of risk assessments. One distinct aspect of this research lies in the proposition of a low-resolution risk profile concept and the adaptation of bivariate modeling outcomes in the risk assessment methodology. By incorporating these innovative ideas, the study introduces a unique perspective to risk assessment in the offshore domain. The chapter's significance was recognized, leading to its publication in the *Journal of Offshore Mechanics and Arctic Engineering* in October 2022. Through its multivariate approach and incorporation of

climate change impact, this research makes a noteworthy contribution to advancing risk-based methodologies for offshore engineering.

Chapter 6

This chapter introduces a dynamic risk profile methodology achieved through the adoption of the Max-stable process for extreme modeling. The methodology provides a comprehensive and detailed risk profile that adapts to changing environmental conditions and offers valuable insights for offshore safety. By utilizing the Max-stable process the method enhances the accuracy and reliability of extreme modeling. The focus of this method is to investigate potential inter-period trends in extreme wind loads, spanning from historical periods to future projections. By analyzing and comparing data from various time frames, the study focuses on identifying significant changes in the occurrence and intensity of extreme wind events in the offshore environment. This chapter's significance and contributions were acknowledged, leading to its publication in the Journal of Safety in Extreme Environments in September 2022. Through its dynamic approach and utilization of the Max-stable process, this research advances the understanding of extreme wind events and their implications in offshore engineering.

Chapter 7

This chapter provides a comprehensive overview of the contributions and conclusions derived from the thesis. The key findings and insights from the research are summarized, highlighting the significance of the study's outcomes in the field of offshore engineering and risk assessment. Moreover, the chapter offers valuable recommendations for future studies, identifying potential areas for further exploration and advancement. By proposing future research directions, the study aims to inspire and guide researchers in their efforts to build upon the current findings and address emerging challenges in offshore engineering. Figure 1.2 illustrates the structure of the thesis,

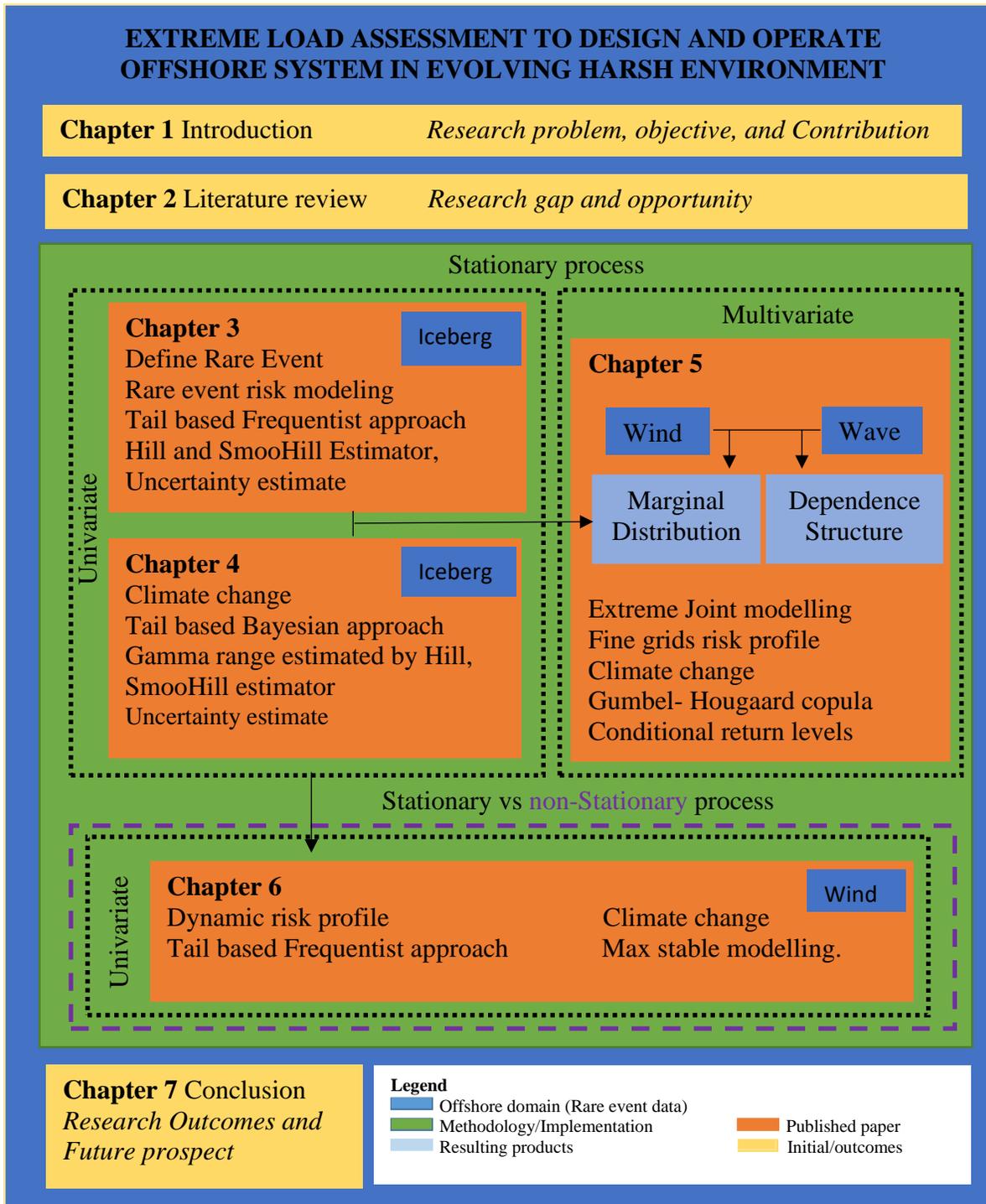


Figure 1.2. Thesis outline.

depicting the organization and flow of the various chapters. This visual representation aids in understanding the sequential development of the research and how each chapter contributes to the overall narrative.

1.7. Authorship contribution Statement

The doctoral candidate, Mohammad Arif, is the original author of this thesis, and played a key role in the conceptualization, methodology development, formal analysis, investigation, and the writing of the original draft as well as the subsequent revisions and editing. The supervisors also made significant contributions; the details of which are as follows:

Faisal Khan (Supervisor): Dr. Khan was involved in the conceptualization of the research, contributed to the methodology development, analysis of results, and provided valuable feedback during the writing process. Also supervised the research project and was involved in project administration and funding acquisition.

Salim Ahmed (Co-supervisor): Dr. Ahmed participated in refining the methodology, provided feedback during the writing phase, and supervised the research.

Syed Imtiaz (Co-supervisor): Dr. Imtiaz contributed to the methodology development, provided valuable feedback during the writing, and editing process, and supervised the research.

2. LITERATURE REVIEW

2.1 Introduction

Extreme event risk assessment is a crucial area of offshore engineering, as it involves identifying and evaluating high-impact, low-probability events (with occurrence probability lower than 10^{-5}) that can pose significant threats to offshore structures, personnel, and the marine environment. This assessment is made more difficult by the problems caused by climate change. The primary goal of this study is to identify a suitable modeling approach for capturing current extremes in the offshore engineering domain, along with an effective risk assessment methodology and a proper framework for incorporating modeling outcomes into risk evaluation. By examining methodologies utilized in diverse fields including business, science, and engineering, the review seeks to identify best practices and adapt them to the offshore context. Furthermore, it will explore the challenges associated with extreme event risk assessment in offshore domain, such as data scarcity, uncertainties in climate change projections, and non-stationarity of extreme events. Key focus will be given to extreme modeling approaches, risk assessment methodologies, and frameworks that can effectively integrate modeling outcomes into a comprehensive risk evaluation process. The objective of this study is to offer valuable insights into the effectiveness of past and present methodologies in the context of current climate change and global warming issues. Specifically, the focus is on assessing how past methodologies can address the challenges posed by extreme events and climate change to ensure the safety and resilience of offshore structures and operations. Additionally, the research aims to identify prospects for further advancements in this

area, with the goal of enhancing the capacity of offshore engineering to cope with evolving environmental conditions and potential risks.

2.2 Bibliometrics Analysis

This research conducts a bibliometric analysis within the Engineering domain to evaluate the current state and potential for modeling extreme events. Initially, the study encompasses various domains such as Business, Science, and Engineering. However, it eventually narrows its focus to Asset Integrity Management, as indicated in Figure 2.1.

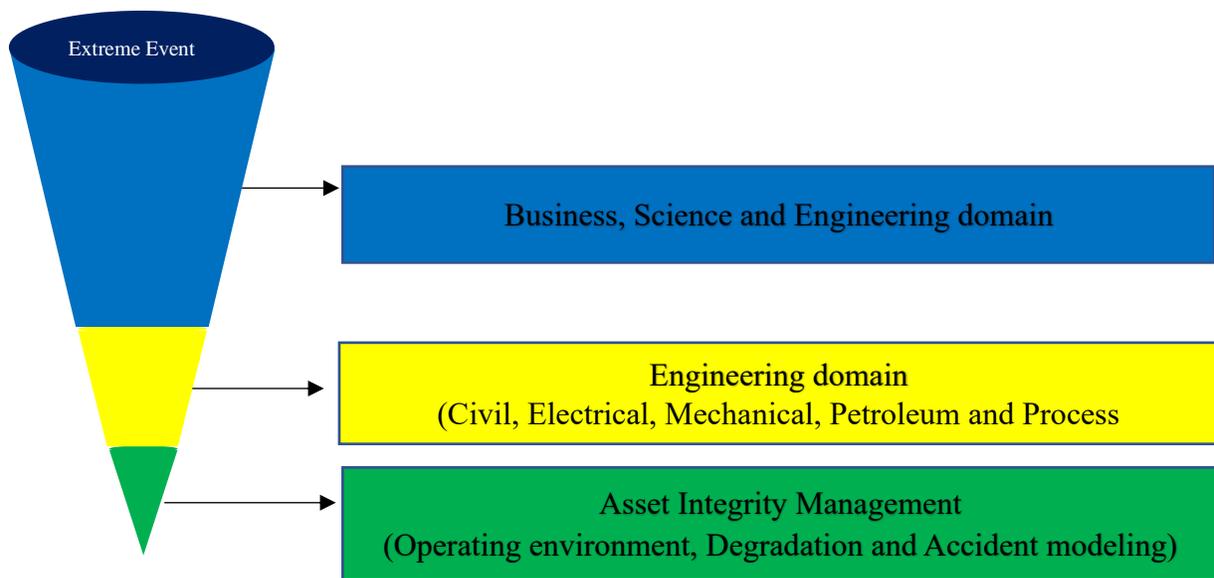


Figure 2.1. Cone diagram for rare event modeling literature for the period 1968-2018.

The analysis centers on three distinct categories: Operating Environment, Degradation Modeling, and Process Accident Events. By thoroughly examining academic research work, the study aims to identify prominent authors, institutions, and nations actively engaged in modeling extreme events. Moreover, the study highlights leading journals in this field and explores their co-citation history, thereby offering valuable insights into the trends and progress within extreme event

research, particularly within the offshore engineering domain. Data for the analysis is gathered from three different databases, namely Web of Science Core Collection, Scopus, and Google Scholar, with access available through Memorial University library database facilities. To obtain relevant data, the researchers conducted searches on all three databases using specific, restricted keywords, as depicted in Figure 2.2.

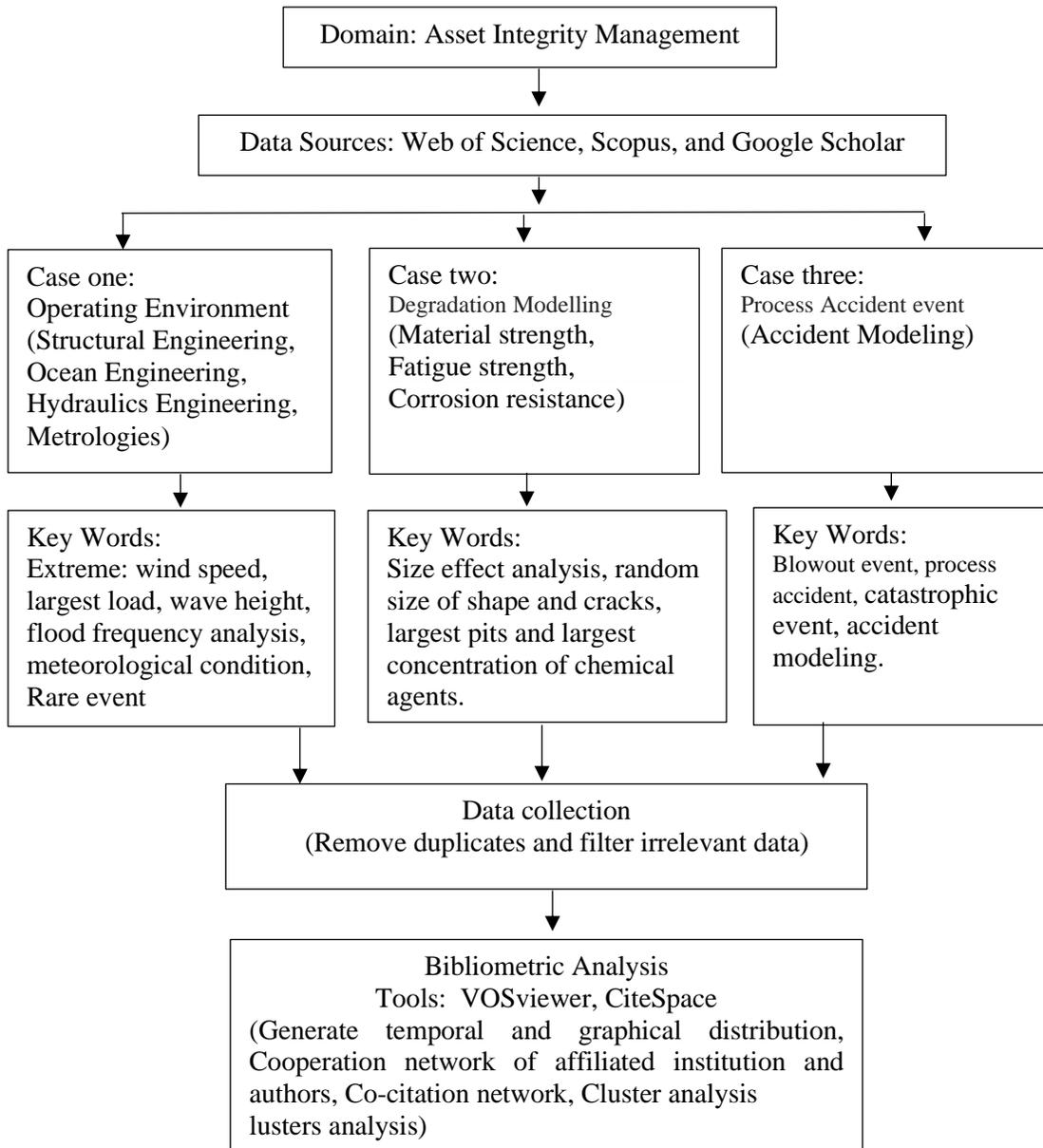


Figure 2.2. Framework for Bibliometric analysis.

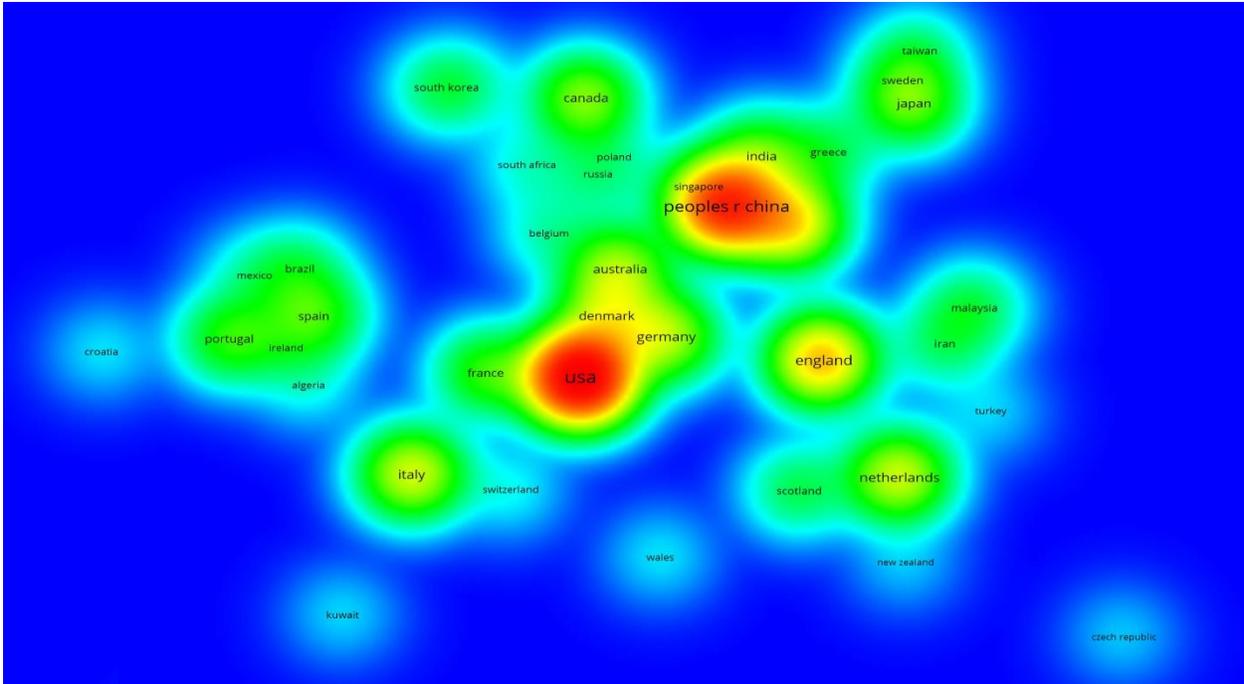


Figure 2.4. Density plot: Extreme events modelling contribution across different geographical locations.

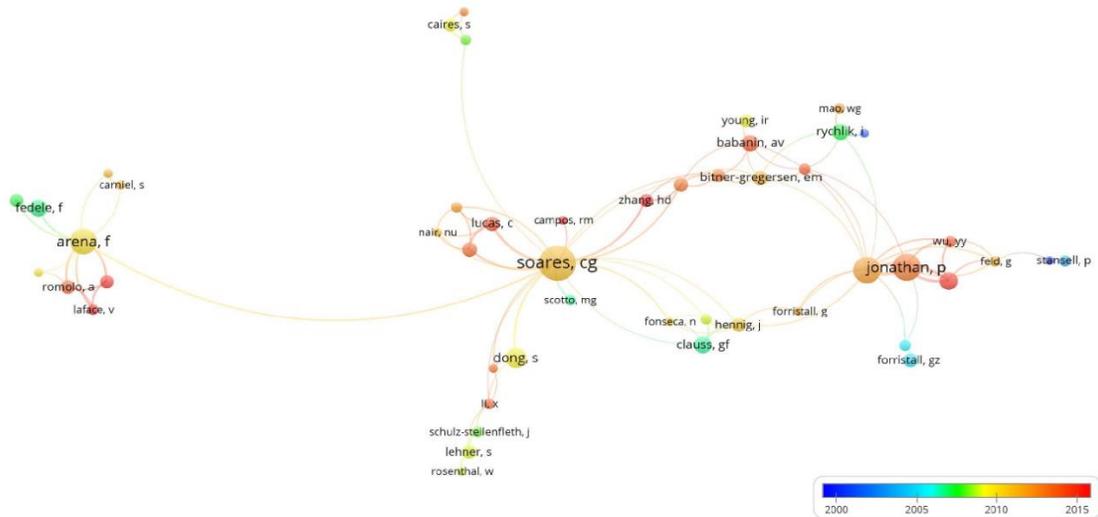


Figure 2.5. The cooperation network of notable authors from peer-reviewed publications in the field of extreme events.

Figure 2.5 shows the prominent authors in extreme event modeling, emphasizing their significant theoretical contributions in different domains as shown in the highlighted sections of the plots.

Within the engineering domain, Castillo et al. (2005), Jan Beirlant et al. (2006), and Sidney Resnick (2007), and Devore (2011) are identified as key contributors to extreme event modeling and its application. In the context of risk analysis, Figure 2.6 emphasizes the authors' contributions in process accident event risk analysis. Faisal Khan's substantial contribution in this area is highlighted, indicating his significant impact on the research.

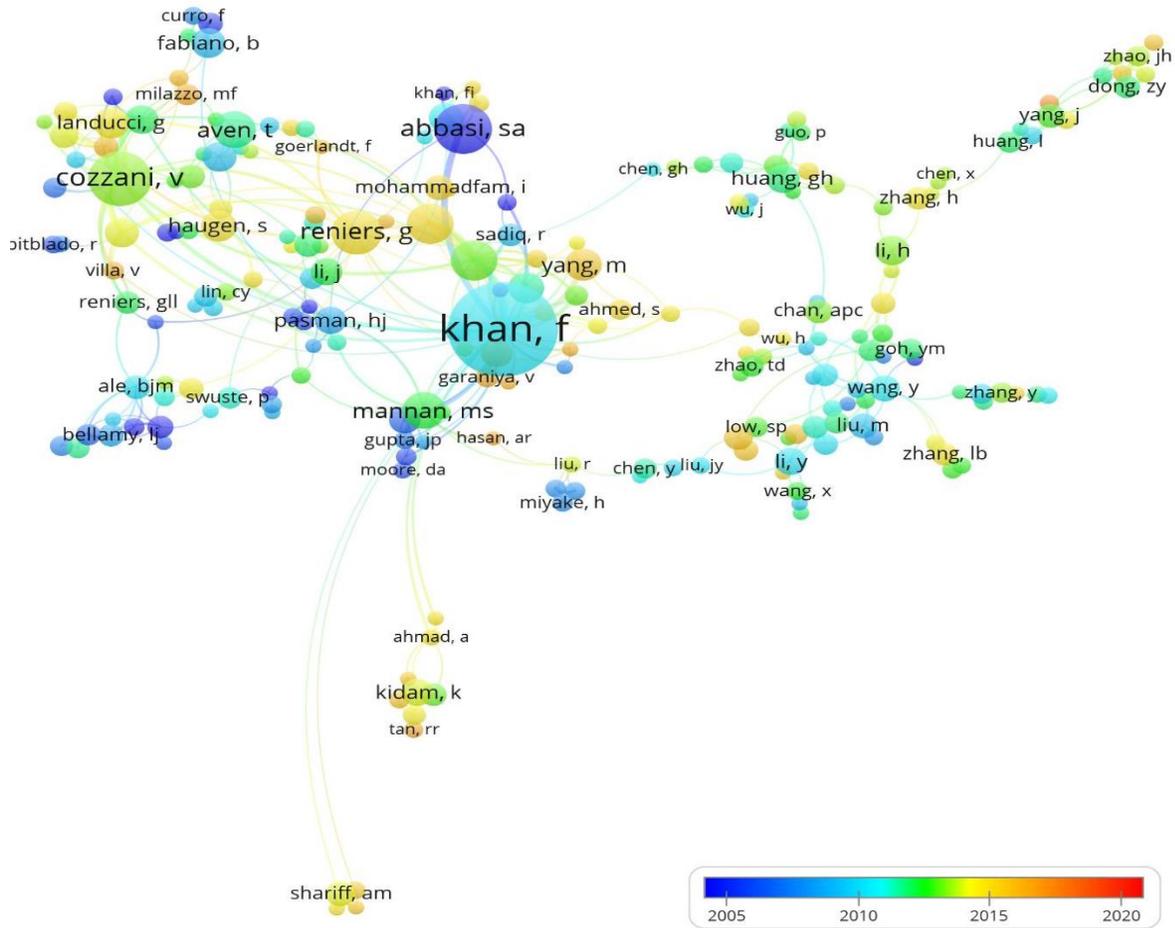


Figure 2.6. Cooperation network of authors of peer reviewed publications in the domain of process accident risk analysis.

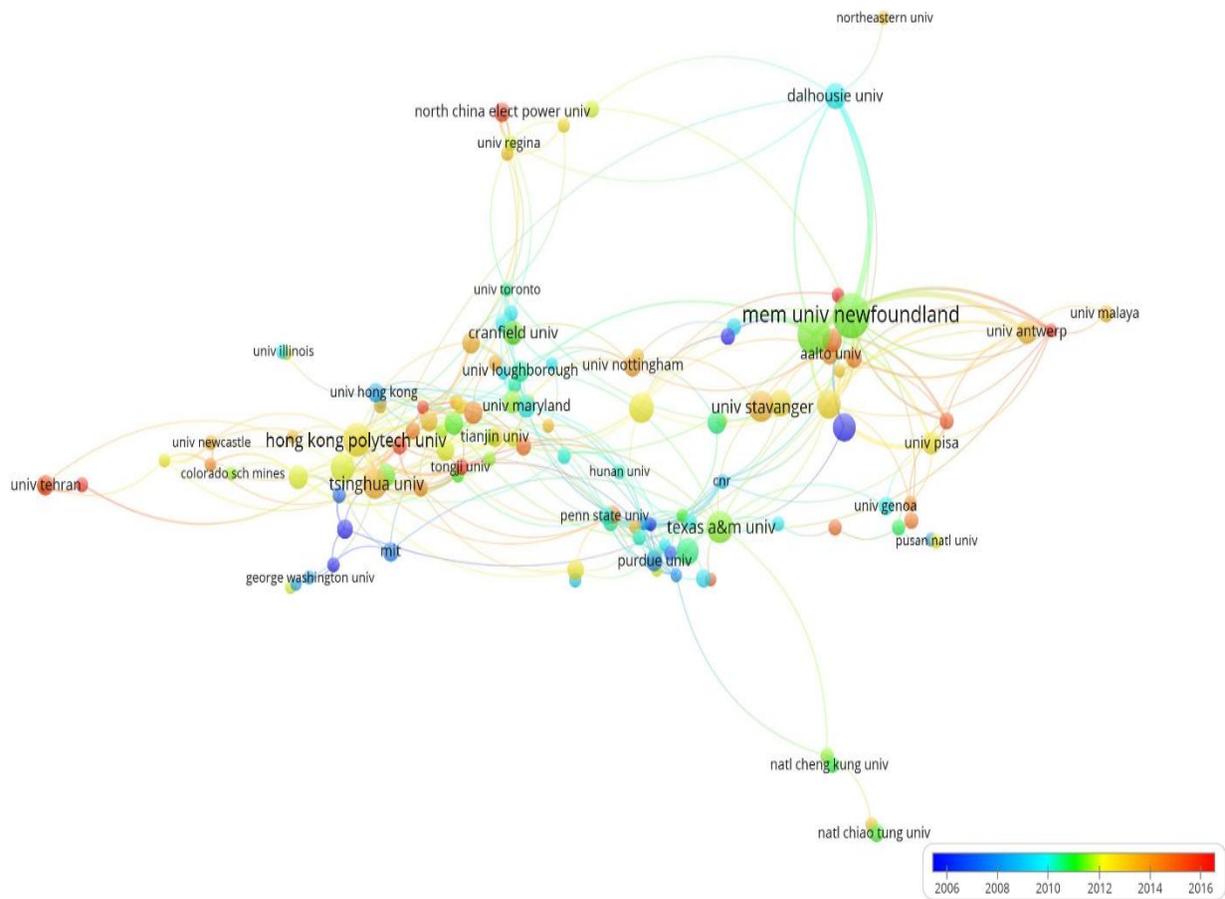


Figure 2.7. Cooperation network of affiliated institution in the domain of process accident risk analysis.

Figure 2.7 highlights the prominent institutions with noteworthy contributions to risk analysis in process accident event modeling. Memorial University is recognized for its substantial research output in this field. These figures and density plots offer a comprehensive view of the influential researchers and institutions in extreme event modeling and risk analysis, providing valuable insights into the distribution of research contributions and the geographical landscape of this critical domain. The following sections 2.3 and 2.4, highlight comprehensive findings from the literature review, including analyses of past and current research on methodologies, challenges, and future directions related to "Extreme Event Modeling" and "Risk Analysis."

2.3 Extreme event modeling

- Definition of extreme event: In the literature, occurrences characterized by "low probability" and "high impact" are commonly referred to as extreme events or rare events. Scholars have used various terms to describe such phenomena, including "extreme events" (Castillo et al., 2005), "black swan events" (Taleb, 2007), "heavy tail events" (Resnick, 2007; Nair et al., 2013 etc.), and "rare events" (Rubino and Tuffin, 2009). These terminologies emphasize the infrequent nature of these events and the significant consequences they entail when they do occur. Devore (2011) introduced precise mathematical definitions for "outliers" and "extreme outliers," making them particularly relevant in the context of offshore engineering, especially when dealing with extreme events. Researchers for example, (McPhillips et al., 2018) have emphasized the necessity of reestablishing a standardized threshold for rarity to ensure consistency in risk analysis. The significance of rare events lies in their potential to cause catastrophic consequences, making their comprehensive investigation imperative in design and decision-making processes.
- Probabilistic modeling techniques: Extreme Value Theory (EVT) is a fundamental approach in rare event modeling, grounded in the statistical theory of extreme values. It provides a means for researchers to model the tail behavior of extreme events, making it well-suited for estimating probabilities of extreme occurrences. The use of methodologies based on Extreme Value Theory (EVT) is prevalent, as evidenced by the works of scholars like Coles, 2001; Agbeyegbe and Leon, 2002; Gilli and Keliezi, 2006; Resnick, 2007; Levine, 2009; Scarrott and Macdonald, 2010; Nair et al., 2013; Das et al., 2016; Asadi and Melchers, Wang and Li, 2016; Hu and Ayyub, 2017; Pryor and Barthelmie, 2021; Debnath

et al., 2021, and others. For a comprehensive theoretical background and its practical application in engineering, Castillo et al., 2005 and Beirlant et al., 2006 are highly recommended sources. Among the above various methodologies, the Peak Over Threshold (POT) based Generalized Pareto Distribution (GPD) and Block maxima (BM) based Generalized Extreme Value (GEV) stand out as the most popular and widely used in diverse domains. Resnick (2007) presents the concept of Heavy Tail Phenomenon, which is more significant in tail probabilistic and statistical modeling approaches under current extremeness. The above methodology offers an advanced statistical technique to model tail behavior and estimate probabilities of extreme events. In addition, the Monte Carlo simulations are commonly employed to assess structural response under extreme conditions (Rubino and Tuffin, 2009; Estecahandy, 2015, etc.). Probabilistic methods are favored for their computational efficiency and ease of implementation.

- Multivariate dependency: Extensive research has been conducted on "Extreme Value Theory" in both bivariate and multivariate scenarios (Tawn 1988; Shiau 2003; Goodarzi et al., 2012, and others). However, certain methods are stringent when it comes to selecting a marginal distribution, as exemplified by studies such as Vanem, 2015; Gaidai et al., 2020; Toshkova et al., 2020, among others. To offer greater flexibility in choosing a marginal distribution, copula-based methods have emerged as a popular alternative (Candela and Aronica, 2017; Manuel et al., 2018; Zhang et al., 2018; Kang, 2019; Hu and Ayyub, 2019; Fang et al., 2020; Liu et al., 2020; Ross et al., 2020, Haselsteiner et al., 2021, and others). These approaches consider the interdependencies among various environmental variables in offshore engineering, a critical factor for accurately modeling the dependency of extreme events.

- Non-stationary analysis: The attention towards the impact of climate change on extreme events in the offshore domain has been considerable. Researchers have extensively utilized methods such as non-stationary approaches to investigate changing environmental conditions over time (Katz et al., 2002, Panagoulia et al., 2014; Vanem, E. 2015; Wang et al., 2017; Paola et al., 2018; Silva and Simonovic, 2020, and others). Moreover, several studies, e.g. Zhang et al., 2018, have demonstrated the effectiveness of time-domain simulations in capturing the dynamic behavior of offshore structures during extreme events. These approaches consider the interplay between climate change and extreme events in offshore settings.
- Numerical simulations: Numerical simulations play a crucial role in understanding the behavior of offshore structures under extreme loading conditions. For example, (Moens and Vandepitte, 2006) utilized Finite Element Analysis (FEA) and (Rij et al., 2019) implemented Computational Fluid Dynamics (CFD) simulations to assess structural response and dynamic behavior during extreme events. However, this methodology is computationally expensive, and was found to be more complicated for the offshore domain.
- Machine learning and data-driven approaches: Machine learning techniques have gained popularity in extreme event modeling due to their ability to identify patterns and relationships in large datasets. For example: (Guth and Sapsis, 2019) have explored the use of machine learning algorithms, such as support vector machines, neural networks, and random forests, to predict extreme events and improve forecasting accuracy. Liu et al., 2020 employed artificial intelligence and machine learning algorithms to optimize extreme event risk analysis. AI-based methods offer opportunities for pattern recognition, anomaly detection, and real-time decision-making, enabling proactive risk management.

- Model parameters and uncertainty estimation: Various methods are widely used in extreme event modeling, including Maximum Likelihood Estimation (MLE), Method of Moments, Probability-Weighted Moments (PWM, Hosking and Wallis, 1987), and L-Moment methods. Each method has its own strengths and limitations. For instance, the efficiency of MLE relies on the sample size, while Bayesian methods offer an alternative approach with wide applicability (Lee, 2007). Additionally, Hill (Hill, 1975) and SmooHill (Resnick, 1997) methods are commonly employed for tail modeling. Choice of an appropriate estimation method depends on data characteristics, modeling assumptions, and the desired level of accuracy for the specific application. Incorporating uncertainty quantification techniques, such as bootstrapping (Efron, 1979), Bayesian inference, model selection (using AIC or BIC), Full sample or subsample bootstraps (Politis et al., 1999), model structural uncertainty methods (like Ensemble modeling), Markov Chain Monte Carlo (MCMC) simulation, and sensitivity analysis, can enhance the reliability of extreme modeling results and support informed decision-making in the face of capturing extremes.

2.4 Risk analysis

Quantifying risk is a fundamental aspect of decision-making and planning in various domains, including finance, insurance, healthcare, and engineering. By assigning numerical values to risks, organizations and individuals can better understand potential hazards, make informed choices, and develop strategies to mitigate adverse outcomes. In a risk context, two critical elements are considered:

- Probability: This refers to the likelihood of a particular event or scenario occurring. It is typically expressed as a numerical value between 0 (indicating no chance of occurrence) and 1 (indicating certainty).

- **Consequence:** This represents the potential impact or severity of the event if it were to happen. Effects can vary in scope, ranging from minor inconveniences to catastrophic outcomes.

Therefore, assessing and managing risks involves understanding the potential hazards, evaluating their likelihood and potential consequences, and implementing measures to reduce their negative impacts. (Bedford and Cooke, 2001) provided a comprehensive and in-depth examination of Probabilistic Risk Analysis (PRA). Over the years, numerous methodologies have been proposed for conducting risk analyses. (Tixier et al. 2002) provided a comprehensive review, encompassing 62 risk analysis methodologies, which were categorized into four distinct groups: Deterministic, probabilistic, qualitative, and quantitative. These approaches are designed to enhance structural robustness, optimize emergency response planning, and facilitate the implementation of risk mitigation measures. Among all of them Faisal Kahn's research group frequently utilizes the quantifying risk-based methodology, which has been applied in various offshore studies, such as Sulistiyono et al. (2015), particularly in the context of low-temperature environments. This research adopts a similar risk-based methodologies to evaluate and address risks, offering valuable insights into quantifying risk and facilitating decision-making processes within the offshore domain.

2.5 Research gap/opportunity

Offshore engineering encounters several challenges when it comes to evaluating and handling risks associated with extreme events. These low-probability, high-impact occurrences can lead to severe consequences for offshore structures, personnel, and the marine environment. Despite considerable advancements in extreme event risk analysis, as described in the previous sections, the literature

review has revealed specific research gaps in this field within the offshore domain. The identified research gaps are as follows:

- **Mathematical representation of an extreme event:** Before conducting extreme modeling in the offshore domain, it is essential to establish a well-defined criterion for identifying which events are considered extreme. This criterion will serve as the basis for selecting and analyzing the specific events that fall within the scope of extreme modeling.
- **Data scarcity and quality:** One of the primary challenges in extreme event risk analysis is the scarcity of data. Offshore structures are designed to withstand low probability, high-impact events, but historical data may not be sufficient/available to accurately assess the risks associated with such rare occurrences. Data collection in offshore environments is often expensive and logistically challenging. Research efforts are required to develop innovative ways to obtain and validate reliable data for accurate risk analysis.
- **Climate change adaptation:** One major research gap is the integration of climate change projections into extreme event risk modeling. As climate patterns change, traditional historical data may become less relevant in predicting future extreme events. There is a need for research that effectively integrates climate change projections to capture potential shifts in the frequency, intensity, and occurrence of extreme events in offshore environments.
- **Advanced modeling techniques:** Extreme Value Theory (EVT) based modeling approaches have proven to be effective in extreme modeling across various domains. However, their performance under the influence of climate change becomes less clear and needs further investigation. If EVT based methodologies fail to adequately capture the current extremeness, alternative methodologies/advanced modeling techniques are required. In

addition, investigating and incorporating non-linear effects in extreme event modeling would provide a more accurate understanding of structural responses and potential failure modes under severe conditions.

- **Non-stationarity of extreme events:** Traditional statistical models often assume stationarity, meaning that the statistical properties of extreme events remain constant over time. However, in the context of climate change and changing environmental conditions, extreme events may exhibit non-stationarity. Bridging this research gap involves developing new approaches that can capture the evolving nature of extreme events and their impact on offshore engineering structures.
- **Multi-hazard risk assessment:** Offshore structures are vulnerable to multiple hazards, such as Icebergs, hurricanes, earthquakes, tsunamis, and more. Currently, most risk assessments consider these hazards individually. There is a research gap in developing comprehensive multi-hazard risk assessment methodologies to better understand and manage the combined effects of different extreme events on offshore structures.
- **Uncertainty quantification:** Uncertainty exists in various aspects of extreme events modeling, including data, model parameters, and climate change projections. Quantifying uncertainties in the risk assessment process and exploring their impacts on the results are crucial to provide decision-makers with a more comprehensive understanding of potential risks.
- **Validation of model predictions:** Validating extreme event risk models using limited real-world data poses a challenge. Establishing reliable validation processes and assessing the robustness of model predictions are crucial to ensure the accuracy and reliability of risk assessments.

It needs significant research efforts to address the above research gap in offshore engineering risk analysis for extreme events modeling for improving the resilience and safety of offshore structures. By integrating climate change projections, accounting for non-stationarity, enhancing data quality, exploring advanced modeling techniques, and adopting multi-hazard risk assessment, this research aims to bridge these gaps and contribute to more accurate and comprehensive extreme event risk analysis in the offshore engineering domain.

3. RARE EVENT RISK ANALYSIS – APPLICATION TO ICEBERG COLLISION

Preface

This manuscript is based on a previously published version in the Journal of Loss Prevention in the Process Industries, and I am the primary author of this research paper. Working under the guidance of co-authors Faisal Khan, Salim Ahmed, and Syed Imtiaz, I implemented a statistical model to calculate iceberg collision risk assessments at the Flemish Pass basin. Throughout the research process, I conducted the literature review, collected the necessary data, developed the methodology, and performed the analysis and modeling. The co-authors played crucial roles in the project as well, contributing to concept development, writing the methodology, reviewing, and editing the manuscript, and validating the modeling outcomes. Initially, I drafted the manuscript, and later I incorporated feedback from the other authors and the peer-review process to make necessary revisions and improvements. Co-author Faisal Khan was responsible for project management and funding acquisition for this research. This collaborative effort allowed us to bring together different expertise and insights to produce a comprehensive and robust study on iceberg collision risk assessment in the offshore domain.

Abstract

To design an engineering system, testing in extreme conditions is at least recommended if not required. There are ambiguities about how to define an extreme state and how to consider it in the design of a system or its operation. The probability estimation of such an event is challenging due to data scarcity, especially in many engineering domains, e.g., offshore development. In this study,

available techniques for analyzing the probability of extreme events are examined for their suitability in engineering applications, and a framework is proposed for rare event risk analysis. The framework is comprised of three phases. In the first phase, the outlier based extreme value theory is implemented to estimate the rare event probability. The maximum likelihood criterion is used to estimate the extreme distribution parameters. In the second phase, the rare event is considered as a heavy tail event, and the tail index is estimated through the Hill and the SmooHill estimator. In the third phase, the uncertainty analysis is conducted, and the risk is computed. The proposed methodology is tested for extreme iceberg risk assessment on large offshore structures in the Flemish Pass basin. For this specific case, the estimated design extreme iceberg speed was 4.31 km/hr, with an occurrence probability of 3.61E-06.

Keywords: Rare event; Risk; Heavy tail event; SmooHill; Block Maxima; Peak Over Threshold.

3.1. Introduction

Rare or infrequent events with high impacts are of great concern in industrial and business operations. There is no precise value for the term “rare”, and it depends on the domain of study. For example, in the aviation industry, a catastrophic failure may be considered as a rare event, and acceptable probability is less than 10^{-9} for an average flight time (about 8 hours journey) (Rubino and Tuffin, 2009). Similarly, in the finance industry, in the case of a stock market crash, the failure probability of 10^{-5} will have a huge impact and may be considered as a rare event (Levine, 2009). Modeling of such an event is a key issue and has attracted the attention of engineers and scientists. An engineer may have to design a tall structure considering the highest magnitude of earthquake in that region over about 100 years. In the offshore domain, the worst hurricane or extreme icebergs can be subjects of concern.

Towards rare event modeling, Taleb (Taleb, 2007) defined the term “rare event” as a “black swan”. There is a growing concern regarding the study of black swan events with insufficient data (Aven, 2015). For hydrology, the extreme conditions such as extreme wind, extreme wave height, extreme floods, and others are generally estimated using the Extreme Value Theory (EVT). The ultimate objective of such a study is to prevent and/or to reduce the impact of such potential events. To identify the engineering problems related to rare events, the ocean and offshore engineering domain is categorized into three subdomains, namely, operating environment, structural degradation and process events (accidents). Within the category of operating environment, extreme wind, extreme ice load, extreme wave height, extreme meteorological condition and extreme flood are the risk sources. Extreme fatigue, the largest pits and the largest concentration of chemical agents may contribute to structural degradation. As process events, extreme consequences such as a blowout (Yang et al., 2015) or major process accidents (El-Gheriani et al., 2017a) may also be considered as rare events. Several theoretical approaches have been proposed to model such an event, for example, the Monte Carlo simulation (Rubino and Tuffin, 2009; Estecahandy et al., 2015; Rocchetta et al., 2015), the Mixer model (Hanum et al., 2015), and the alternative approach (Arima et al., 2010). Two practical approaches widely used in the case of extreme analysis are the Peak Over Threshold (POT) based Generalised Pareto Distribution (GPD) (for example (Das et al., 2016), and the Block Maxima (BM) based Generalized Extreme Value (GEV) (for example, see (Levine, 2009; Asadi and Melchers, 2017; Agbeyegbe and Leon, 2002; Gilli and Keliezi, 2006)). However, questions remain regarding whether all the preceding extreme cases should be considered as rare events or not. In the engineering domain, an event occurrence probability less than one in hundred thousand is often considered as rare. The study presented here highlights the rare event scenario which is unique for ocean engineering and offshore system design (domain

specific). It is recognized that classical statistical approaches are not the most effective way of modeling rare event. It is therefore in the present study a logic-based approach is considered to model a rare event. The logic-based approach provide reasonable source of the data on rare event and thus enable use of advanced statistical approach to model rare event frequency and probability. For modeling such events, one needs to precisely define the term “rare event”. Devore proposed a scale to measure the outliers and extreme outliers in a data set (Devore, 2011). In the present work, Devore's concept is adopted to classify an event as rare or not. A data set having extreme outliers is considered as rare event data. This paper examines the applicability and efficiency of some well-established models from other fields of study such as finance, and hydrology, to the offshore engineering domain for the purpose of rare event modeling; more specifically, the GPD and the GEV are tested. If GPD or GEV are not able to capture extreme characteristics of a rare event, then the rare event may be considered as heavy tail events. The tail index is estimated through Hill (Hill, 1975) and SmooHill (Resnick and Starica, 1997) estimators.

In general, rare event modeling suffers from uncertainty to a high degree. Therefore, it leads to how engineering design can deal with this vast uncertainty. However, modern engineering designs have been used the concept of flexibility to effectively improve the expected performance of a system in an uncertain environment (Neufville and Scholtes, 2011). The flexible design risk estimation is a common problem in risk analysis. The flexibility concept was applied to design a water supply system (Zhang and Babovic, 2011) and a water management system (Deng et al., 2013; Manocha and Babovic, 2018). However, to develop such an idea for the rare event is challenging, and climate changes make this task more complicated in the ocean engineering domain. This study implements the flexibility concept in the ocean engineering domain under rare event condition for the design purpose. The proposed extreme event risk-based design approach is

a flexible engineering design approach. The flexibility is achieved due to adoptable risk acceptance criteria in design purpose. For example, in the current case study of the iceberg risk-based design approach is explained using minimum acceptable risk criteria. The risk is estimated considering impact energy which is a function of iceberg size and speed. This criterion is different compared to traditional iceberg analysis (C-CORE, 2015; Husky oil operations limited as operator, 2000). This paper is dedicated to the iceberg risk estimate from the prospect of engineering design in the Flemish Pass basin. The basin is situated in the North Atlantic Ocean (around 400 km east of St. John's, Newfoundland and Labrador) and might be the site of a new offshore oil bonanza. Offshore drilling operators might find the results useful in avoiding any future unwanted situation by considering the estimated iceberg risk (through extreme iceberg speed prediction) in the iceberg risk management plan for the Flemish Pass basin. The manuscript is structured as follows: following the introduction, Section 3.2 defines a rare event and describes the proposed methodology to model rare events. Section 3.3 provides an overview of the case studies and presents results. Finally, Section 3.4 provides conclusions and recommendations for future works.

3.2 Methodology

3.2.1 Defining a rare event

A rare event is traditionally defined as an event with a low probability which may have a high impact. This research attempts to precisely define the term “rare event” and justify it with a specific criterion. To accomplish this, first, the terms “outliers” and “extreme outliers” are defined. According to Devore (Devore, 2011), in a data set, any data point greater than $Q3 + 1.5 \times (Q3 - Q1)$ is an outlier, and a data point greater than $Q3 + 3 \times (Q3 - Q1)$ is an extreme outlier, where $Q3 - Q1 =$ Interquartile Range (IQR), $Q3 =$ third quartile and $Q1 =$ first quartile. Therefore, the data sets have been classified into three different categories: the sample data set X may be a case of having “no

outliers”, “outliers” or “extreme outliers (rare events)” and mathematically expressed (for right tail) as

$$\left\{ \begin{array}{l} \text{no outliers if } x \leq (Q3 + 1.5 \times IQR) ; \text{ for all } x \in X \\ \text{has outliers if } (Q3 + 1.5 \times IQR < x \leq (Q3 + 3 \times IQR)); \text{ for any } x \in X \\ \text{has extreme outliers (Rare event) if } x > (Q3 + 3 \times IQR); \text{ for any } x \in X \end{array} \right. \quad (3.1)$$

This research considers an extreme outlier event as a rare event. Data sets with outliers have tails (regular or fat), and these tails become heavier for data sets having “extreme outliers”. Tails may be right or left and this classification is equally applicable for data with left tails. If the data sets have no outliers, then any classical statistics (the case for central limit theory) may be applied for modeling which is not of interest for the present study. The focus here is on rare event cases (extreme outliers), where the data sets have a heavy right tail.

3.2.2 The Proposed framework

A heavy right tail describes a system behaviour that is driven by the large values or, in extreme cases, by a single large value. The tail index is characterized using two well-known methods, namely, the Hill estimator and the SmooHill estimator. Using this tail index, the Heavy Right Tail Distribution (HRTD) is implemented and the return levels are estimated. The 100-year return periods are most common in extreme analysis e.g. (Das et al., 2016; Sulistiyono et al., 2015) and in rare event modeling context, 100 years are not the most extreme. Therefore, a 1000-year aim is considered to adjust the time window of a rarest event and predictability of the event beyond a typical lifetime. In addition to this principal aim, the framework attempts to implement “EVT” as a part of outlier-based approach. This gives an opportunity to compare the results and check the efficiency of the conventional approaches in the case of rare event modeling with limited sample size. The proposed framework to analyze rare events is presented in Figure 3.1.

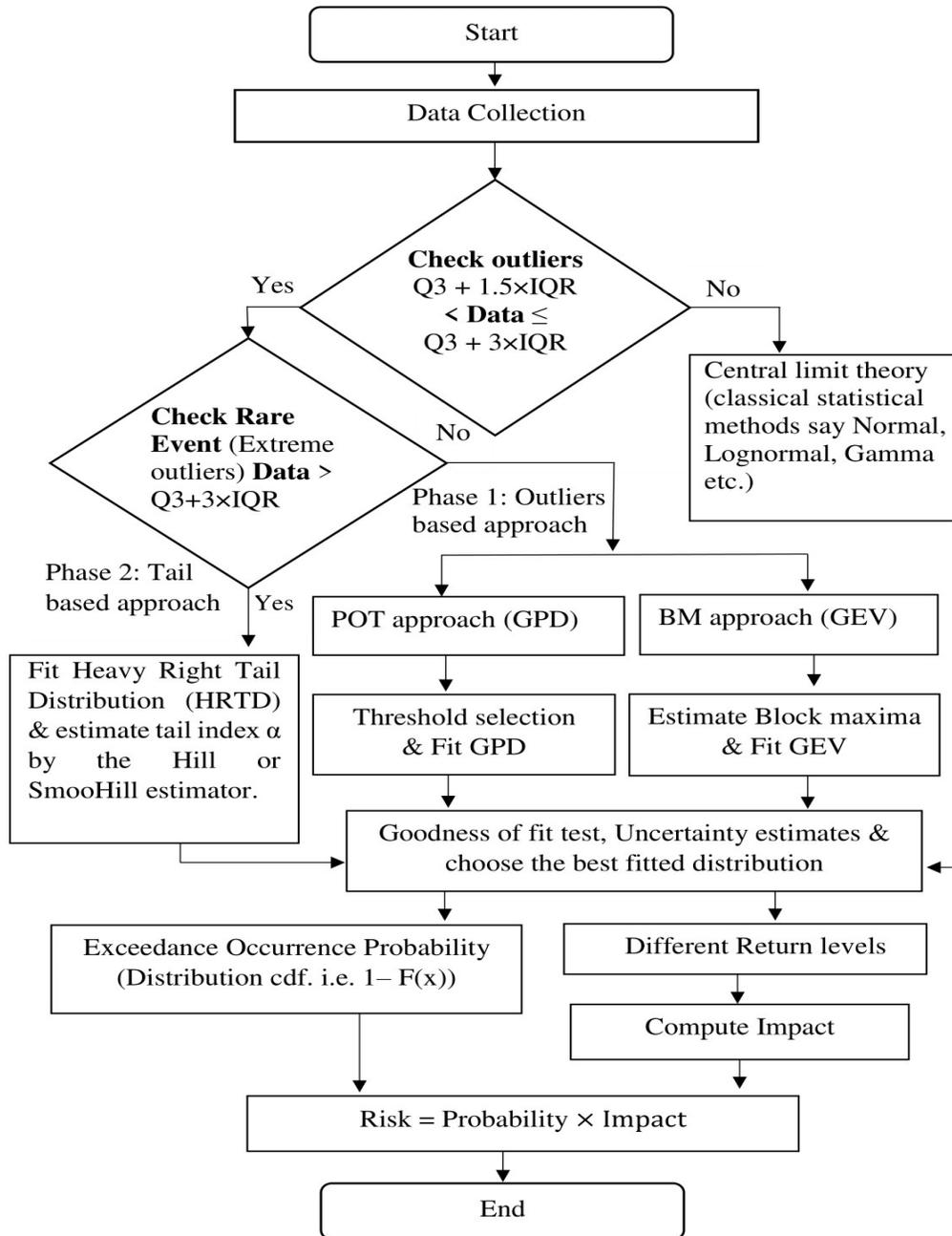


Figure 3.1. The proposed framework for risk analysis.

3.2.2.1 Phase 1: Outlier based approach (traditional)

The EVT is implemented under two fundamental concepts: the BM and POT. The POT approach uses all significant observations, while BM misses some significant observations which may fall beyond the block. For that reason, in extreme value statistics, most researchers prefer the POT

method over the BM method. However, in the case of independent and identical distribution (IID) assumption, the convergence rates of the two ways are comparable (Bucher and Zhou, 2021). The Generalized Extreme Value (GEV) is the standard form of three distributions; namely Gumbel, Frechet, and Weibull distributions (Castillo et al., 2005). The Frechet distribution has a long tail compared to the Weibull and Gumbel. On the other hand, according to the Pickand theorem (Pickands, 1975), for any extreme analysis, the distribution of extreme events above a high threshold is a Generalised Pareto Distribution (GPD). The Cumulative Distribution Function (CDF) of GPD is defined as (Castillo et al., 2005):

$$F(x) = \begin{cases} 1 - \left[1 - k \left(\frac{x-\mu}{\sigma}\right)\right]^{\frac{1}{k}} & \text{if } k \neq 0, \\ 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)} & \text{if } k = 0, x \geq 0, \end{cases} \quad (3.2)$$

where $\sigma > 0$, and stands for the scale parameter; k is the shape parameter and μ is the location parameter. For a random variable X , (in the case of POT approach) the excess distribution over a threshold u is defined as:

$$F_u(x) = P(X - u \leq x | x > u). \quad (3.3)$$

The expected return level in the annual scale is denoted by x_m (the likelihood of extreme events and m -year return level stands for the level expected once in every m -year), and is determined as (Das et al., 2016; Castillo et al., 2005):

$$x_m = \begin{cases} u + \left(\frac{\sigma}{k}\right) [(m \times n_e \times p_u)^k - 1] & \text{for } k \neq 0, \\ u + \sigma \log(m \times n_e \times p_u) & \text{for } k = 0, \end{cases} \quad (3.4)$$

where u is the threshold, n_e is the number of mean events per year, p_u is the probability (exceed rate) and computed as $P(X > u)$. The BM approach deals with all maximum observations from several blocks and the GEV family is appropriate to model the data set consisting of all maximum

observations. The CDF of Generalized Extreme Value Distribution (GEV) is given as (Castillo et al., 2005):

$$F(x) = \begin{cases} \exp \left\{ - \left[1 - k \left(\frac{x-\mu}{\sigma} \right) \right]^{\frac{1}{k}} \right\} & \text{if } k \neq 0, \\ \exp \left\{ - \frac{(x-\mu)}{\sigma} \right\} & \text{if } k = 0. \end{cases} \quad (3.5)$$

The case of interest here is only when $k > 0$, which is bounded and approaches to Pareto distribution in the right tail. The return level of the GEV is defined as (Castillo et al., 2005):

$$x_m = \begin{cases} \mu + \frac{\sigma}{k} [1 - (-\log(1-p))^k] & \text{for } k \neq 0, \\ \mu - \sigma \log\{-\log(1-p)\} & \text{for } k = 0. \end{cases} \quad (3.6)$$

The occurrence probability p is related to m -return level as $\frac{1}{m}$. The POT and the BM approaches have merits and demerits based on implementation, sample size and the case of study. This research implements both methods as part of an outlier-based approach, to check their capabilities of identifying a rare event. This study uses Maximum Likelihood Estimation (MLE) method to estimate the distribution parameters. In MLE, the likelihood function of independent observations x_1, x_2, \dots, x_n is defined as:

$$L(x_i; \theta) = \prod_{i=1}^n f(x_i; \theta), \quad (3.7)$$

where $f = \frac{dF}{dx}$, and θ represents the distribution parameters. For the case of GPD, $\theta = \sigma, k$ and in the case of GEV, $\theta = \mu, \sigma$ and k . The efficiency of the MLE depends on the sample size and its shape parameter. If the sample size is not sufficient to fulfill the convergence criteria of the MLE method, then the Probability Weighted Moment (PWM) (Hosking and Wallis, 1987); is an alternative. However, if $k < 0$, PWM performances become a little worse than with MLE, otherwise, these two methods are comparable at $k \approx 0.2$ (Deidda and Puliga, 2009).

3.2.2.2 Phase 2: Tail based (Non-traditional) approach

In the case of “extreme outliers” or a rare event, the tail must be heavier than for the case of “outliers”. From the rare event prospect, this work focuses on the extreme of extremes. More precisely, the fundamental interest is in heavy tail, which is heavier than the exponential distribution and the regular extreme distribution (GPD or GEV). However, there are limited data for rare event scenarios in engineering. The rare event characteristic is examined as a heavy tail event (Resnick, 2007), and (Nair et al., 2013). The objective here is to focus on the tail instead of looking at full distribution. More precisely, the focus is on the shape parameter instead of the scale parameter as a part of the semiparametric modeling approach. In such situations, finding the tail index (exact position where the tail starts) and how this tail index is used for rare event prediction will be vital. To estimate tail index, several methods are proposed in the literature e.g. (Resnick, 2007). Among them, the Hill estimator (Hill, 1975) is widely used to estimate the tail index; however, it has two difficulties. If the distribution is far from the Pareto, then the estimates can be biased (Resnick, 2007); also, the Hill estimator is highly sensitive to the choice of the higher order statistics (Resnick and Starica, 1997). To minimize this sensitivity issue, the SmooHill estimator was proposed by Resnick and Starica, 1997. The bias issue has less relevance in the present study, as the purpose is to model rare events, where the sample distribution is likely to be of Pareto type. Besides those methods, the weighted Hill estimator, the qq estimator, and the moment estimator described in (Resnick, 2007) are also applicable, however, are not considered in the present study. Let X be a heavy tail random variable. The Heavy Right Tail Distribution (HRTD) is mathematically defined as (Resnick, 2007):

$$P[X > x] \sim f(x) = \alpha x^{-\alpha-1}; \text{ for } \alpha > 0, x > 0, \quad (3.8)$$

the probability, $X \leq x$, is defined as (Resnick, 2007):

$$P[X \leq x] \sim F(x) = 1 - x^{-\alpha}; \text{ for } \alpha > 0, x > 1, \quad (3.9)$$

where α is the tail index of the heavy right tail distributions. Equations 3.8 and 3.9 are also known as pdf and cdf of the special case of Pareto distribution with a scale parameter of one and tail index of α considered as its shape parameter. The expected return level of HRTD is defined as:

$$x_m = u + \alpha[(m \times n_e \times p_u)^{\frac{1}{\alpha}} - 1] \text{ for } \alpha \neq 0, \quad (3.10)$$

where u is the threshold (considered as the location parameter), m is the return period, n_e is the number of mean events per year, p_u is the probability (exceed rate) and computed as $P(X > u)$ and α is the tail index. The Equation 3.10 is equivalent to Equation 3.4 with assumptions, GPD scale parameter, $\sigma = 1$ and shape parameter, $k = \frac{1}{\alpha}$, where α is the tail index. Considering that x_1, x_2, \dots, x_n are independent and identically distributed with order $x_1 > x_2 > \dots > x_n$, then the Hill estimator of α is defined as (Hill, 1975):

$$H_{r,n} = \frac{1}{r} \sum_{i=1}^r \ln \left(\frac{x_{(i)}}{x_{(r+1)}} \right), \quad (3.11)$$

where r stands for the order of higher order statistics. The computation process is the same as that of MLE, except r top order statistics must be considered. In practice, the Hill plot is a plot of $H_{r,n}$ against the values r ($1 \leq r \leq n$). The SmooHill estimator is defined as (Resnick, 1997):

$$SmooH_{r,n} = \frac{1}{(v-1)r} \sum_{j=r+1}^{vr} H_{j,n}, \quad (3.12)$$

where v is an integer (greater than one) and $vr < n$, where n is the sample size.

3.2.3 Uncertainty estimate for model parameters and return level

Efron introduced the term Bootstrap in 1979 (Efron, 1979). The proposed method is applicable for parametric and non-parametric estimation of the uncertainty. Full sample or subsample bootstraps (Politis et al., 1999) may give biased estimates for a small data set and more so in the case of extreme events. A parametric bootstrap is used and integrated with the MLE to avoid the bias issue.

Bootstrap data sets are generated randomly from the fitted distribution instead of being sampled from the original data. More precisely, in the case of GPD, from the original data, GPD distribution parameters $\{\sigma, k\}$ are estimated. Using the estimated parameters, a random sample (bootstrap data set) generated from the GPD distribution. The length of the random sample as the same length as the original data set. Then the GPD is fitted on the bootstrap dataset and update the values of the parameters and return level calculated for any specific year (e.g., 100 years). This process is repeated, e.g., 1000 times, which gives 1000 different parameter sets $[\{\sigma_1, k_1\}, \{\sigma_2, k_2\} \dots, \{\sigma_{1000}, k_{1000}\}]$, and “100 year return level” $[x_1, x_1, \dots, x_{1000}]$. According to the central limit theory, the distributions of these parameters and return levels are normally distributed. Finally, the mean value and a 95% confidence interval for the model parameters and return level for different time periods are computed. In the full repetition process, the thresholds are fixed as estimated from the original data.

The proposed approach has the following steps:

1. Generate a bootstrap data set: for the outliers-based method, use a GPD (GEV) to generate a random sample with estimated model parameters. In the case of a tail-based approach, the random samples are generated through the HRTD (more precisely, a random Pareto distribution is generated with a scale parameter assumed one and shape parameter as the tail index). The sample size and threshold are considered to be the same as for the original data sets.
2. Fit GPD (GEV) on the bootstrap dataset and update model parameters with a fixed threshold (block size). For the case of tail-based approach, fit a Pareto distribution on bootstrap data sets and update the shape parameter.
3. Determine the return level using the new estimated parameters.

4. Repeat these steps 1000 times and obtain 1000 values for model parameters and 1000 values for each return period, m.
5. Determine the mean value and a 95% confidence interval for the model parameters and return level for different periods.

3.2.4 Risk analysis

The risk estimation process follows a similar approach as in (Sulistiyono, et al., 2015) with some modifications. The iceberg collision risk is denoted as $R(\text{iceberg})$ and defined as:

$$R(\text{iceberg}) = P(O) \times C, \quad (3.13)$$

where $P(O)$ is occurrence (extreme iceberg speed) probability and C is the iceberg consequence (impact). The occurrence probability is defined as:

$$P(O) = (RP)^{-1}, \quad (3.14)$$

where RP is the return period. In the present study, predictions are conditioned on certain threshold values. Therefore, to make the probability unconditional, i.e. using all past history data the Equation 3.14 is modified as:

$$P(O) = p_e \times p_u \times (RP)^{-1}, \quad (3.15)$$

where p_e is the probability of the event that is more extreme being more likely to occur than the estimated extreme event for a given return period. p_u is the exceed rate ($\Pr [X > u]$, where u is the threshold). As a part of consequence (C) or impact, this research considered the kinetic energy of the iceberg. The Kinetic Energy is denoted by KE , and defined as (Husky oil operations limited as operator, 2000):

$$KE = \frac{1}{2} M \beta v^2, \quad 3.16$$

where M is the iceberg mass, v is the iceberg speed (calculated by model prediction) and β is a constant (added mass coefficient). The mass coefficient (β) is added to find the entrained mass of

water surrounding the iceberg, and its value is chosen as 1.2 (Husky oil operations limited as operator, 2000). To estimate an acceptable extreme design load of iceberg impact (considering iceberg size and extreme speed), operators need to define an acceptable risk level. The risk level might vary considering the region of operation, regulatory requirement, technology maturity as well as environmental factors. As Low As Reasonably Practicable (ALARP) framework is often used to define the acceptable risk (Pike et al., 2020). In the present case study, an acceptable risk level is defined as 0.001 MJ/year. This acceptable risk is derived from past studies related to offshore structure design (Sulistiyono et al., 2015). The risk is estimated considering the size of the iceberg, iceberg speed and the exceedance probability. The maximum Iceberg speed (corresponds to all identified acceptable risk values) is selected as the design iceberg's extreme speed. This approach avoids hardening against every possible scenario and thus simplifies the choice of design criterion. For example, to assess the risk of wind load in a specific region, a set of four extreme wind speeds (25 m/s, 29 m/s, 32 m/s, 35m/s and 39 m/s) are considered. The wind speeds exceedance probabilities (0.045, 0.0046, 0.001, 0.00032, and 0.000041), wind loads (as impact energy; 379 N, 522 N, 614 N, 682N, 901N) are the return periods 5, 20, 50, 100 and 1000 years are used to estimate risk. Here the risk (due to extreme wind speed) is defined as a combination of the likelihood of extreme wind and its impact energy (may be referred to as consequences, and for design, purposes computed as load in N). The likelihood is the measure of the probability of occurrence for a return period (/yr). Therefore, multiplying the corresponding likelihood with impact energy, the risk values are estimated as 17.1, 2.3, 0.61, 0.22, 0.04 N/year, respectively. Based on an acceptable risk level (say 0.61 N/year), the design extreme wind speed is then obtained as 32 m/s.

3.3 Case Study (Iceberg extreme speed estimation for the Flemish Pass basin)

The proposed framework is implemented to estimate the iceberg risk to an offshore structure. The iceberg risk is evaluated considering the probability of iceberg collision with the structure and its impact. The impact energy of the iceberg collision is dependent on iceberg speed and size. The present study focuses on estimating extreme iceberg speed in three different categories (based on iceberg size), namely small, medium, and large icebergs. There is no direct data sources for iceberg speed. Iceberg speed mainly depends on size, shape, currents, wind speed and wave height; here, iceberg speed was computed through the generic method (distance/time). Iceberg sight locations and corresponding times were taken from the International Ice Patrol (IIP) iceberg sighting database (International Ice Patrol, 1998). Most iceberg speed analysis (for example (Husky oil operations limited as operator, 2000; King et al., 2015) focuses on the average iceberg speed over space. However, here single iceberg speed was computed in a specific region. The primary study region of interest is the Flemish Pass basin (47° N to 48° N, 46° W to 47.3° W) for the period 2002 to 2015. Two random areas are considered to compute the average iceberg speed. Finally, our computed average speed is compared with C-CORE. (2015). to check the accuracy of the speed computations. For both areas, computed speeds are comparable. More precisely, the iceberg speed listed in (C-CORE, 2015) versus estimated average speed in this study is 0.88 km/hr versus 1.02 km/hr in the areas 46° N to 50° N and 45° W to 50° W and 0.9 km/hr versus 1.15 km/hr in the regions 47° N to 48° N and 46° W to 48° W. Although this study is focused on single iceberg speed, this average speed comparison gives confidence that the computed iceberg speed in this study is acceptable. Moreover, the validation of these two different areas gives confidence for iceberg speed computation in the primary study area, i.e., the Flemish Pass basin (47° N to 48° N and 46° W to 47.3° W). The present study uses open-source platform R (freely available at

<https://www.rstudio.com>) with extreme R packages such as POT, extRemes, evmix, ismev, EnvStats and fExtremes. Minitab is used for the descriptive statistical analysis.

3.3.1 Descriptive statistics

Icebergs pose serious concerns for offshore drilling and marine facilities. This work focusses on specific iceberg speed instead of average speed. Every year around 20,000 to 40,000 icebergs move across the Baffin Bay from west Greenland, and the Labrador current transports some of them to the Flemish Pass basin. Icebergs are of different sizes and shapes. According to the descriptive statistical analysis on iceberg data from the Flemish Pass basin, medium (MED) size icebergs were found to be more frequent (548) during the period from 2002 to 2015, which is about 51% of total

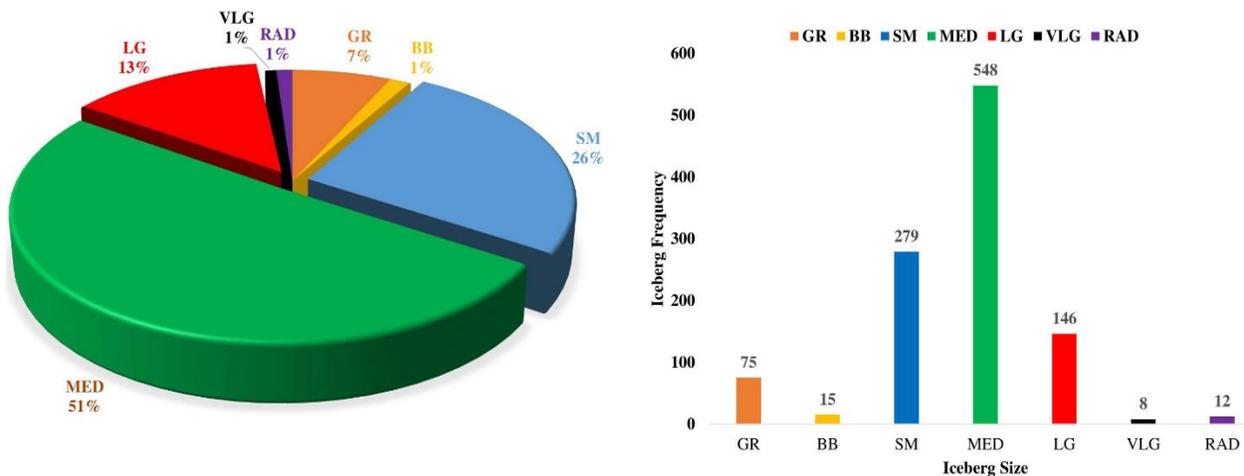


Figure 3.2. Iceberg frequency distribution at Flemish Pass basin (2002-2015).

icebergs sighted in that area (Figure 3.2). As shown in Figure 3.2, 279 small (26%) and 146 (13%) large icebergs were also sighted. This study only focusses on small (SM), medium (MED) and large (LG) icebergs. There are other icebergs such as (10%) growlers (GR), bergy bits (BB), very large (VLG), and radian (RAD). The size of radian (RAD) icebergs are not determined. The very large icebergs are not considered in this study because of their very low speed (maximum speed

computed 1.1 km/hr) compared to other sizes; also, their sizes are not precise. Moreover, VLG, GR and BB icebergs are easily manageable using an existing offshore iceberg risk protection plan.

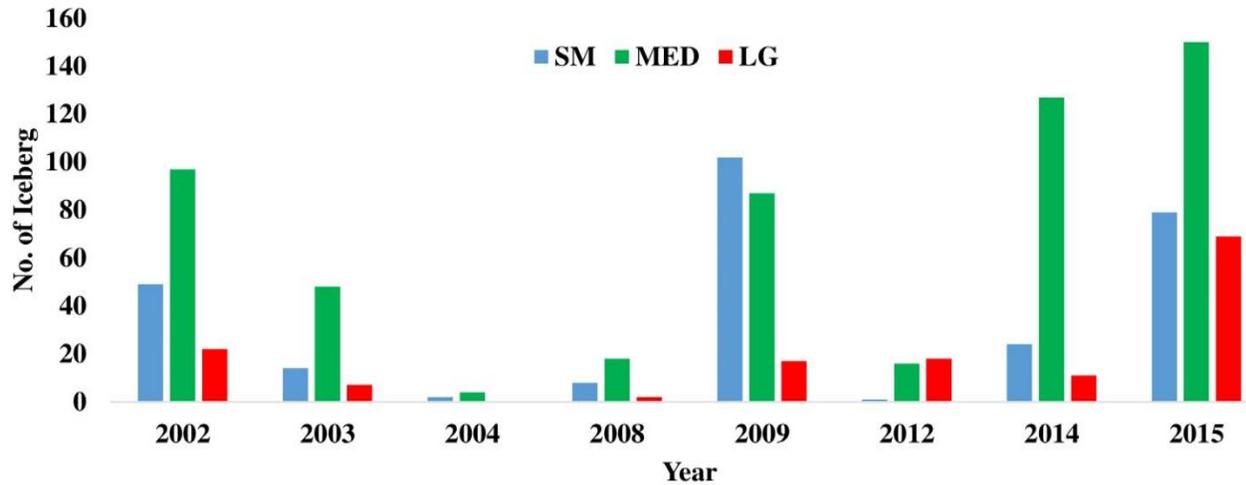


Figure 3.3. Iceberg temporal (2002 to 2015) distribution (small, medium and large).

The temporal iceberg distribution (2002-2015) is illustrated in Figure 3.3. In all three cases, the number of icebergs has an increasing trend during the period 2012 to 2015. Iceberg sight location and time are recorded to compute iceberg speed. In Figure 3.4, the calculated iceberg speeds display maximum and average iceberg speeds. The highest maximum iceberg speed is 5.78 km/hr, and the lowest maximum is

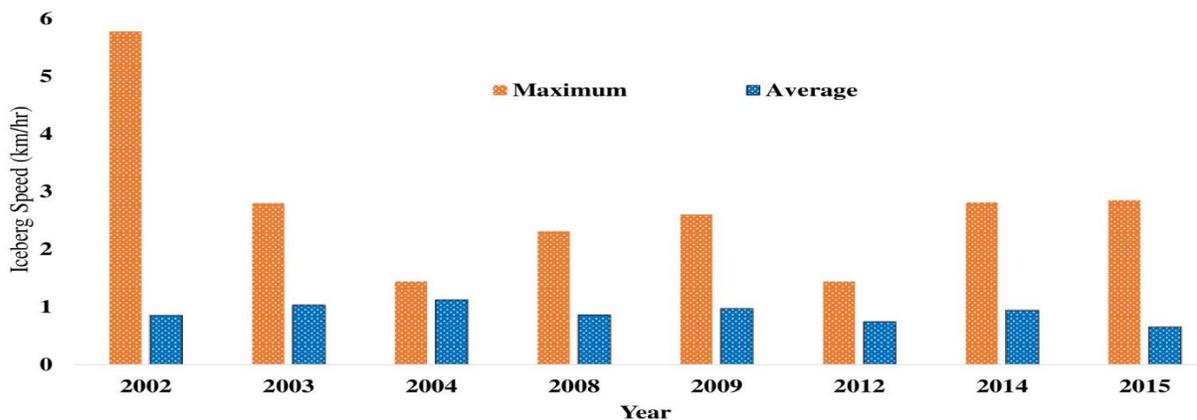


Figure 3.4. Iceberg speed distribution (maximum vs average) over the period 2002 to 2015 (merge small, medium and large iceberg).

1.44 km/hr. However, average maximum speed is only 1.12 km/hr during the study period. The average speed is listed as lower than the lowest single iceberg speed and compared to extreme individual speed, it is about one fifth. Therefore, instead of the average iceberg speed, the individual extreme iceberg speed needs to be considered as a part of an ice management plan. Table 3.1 presents the statistics of the iceberg data in the study region. In the case of the small iceberg, seven observations were identified as outliers, and two of them are extreme outliers (data > 3.08). For the medium size, five views are listed as outliers, and among them, only one is an extreme outlier (data > 2.97). Finally, for the large iceberg, there are three outliers, and one is an extreme outlier (data > 2.33), also clearly visible in the boxplot Figure 3.5 (extreme outliers are highlighted with cyan circles). For the small iceberg data in Table 3.1, skewness (3.37) and kurtosis (19.7) indicate that this data set does not follow the normal distribution. The positive value of

Table 3.1. Descriptive statistics: Flemish Pass basin iceberg data (2002-2015) categorize in three different size (Small, medium and large).

Data Statistics (Iceberg speed, km/hr)											
Iceberg Size	Minimum	Q1	Median	Q3	Maximum	IQR	Q3+1.5IQR	Q1-1.5IQR	Q3+3IQR	Skewness	Kurtosis
Small	0.134	0.497	0.764	1.14	5.78	0.645	2.11	-0.470	3.08	3.37	19.7
Mediu	0.0140	0.460	0.704	1.09	3.45	0.628	2.03	-0.483	2.97	1.38	3.10
Large	0.0566	0.444	0.671	0.915	2.58	0.471	1.62	-0.261	2.33	1.33	3.66

skewness indicates that the size of the right-handed tail is larger than the left-handed tail. The positive Kurtosis also implies that the data set has a heavy tail. The distributions of medium and large icebergs also have heavy tails (medium: skewness 1.39, kurtosis 3.10 and large: skewness

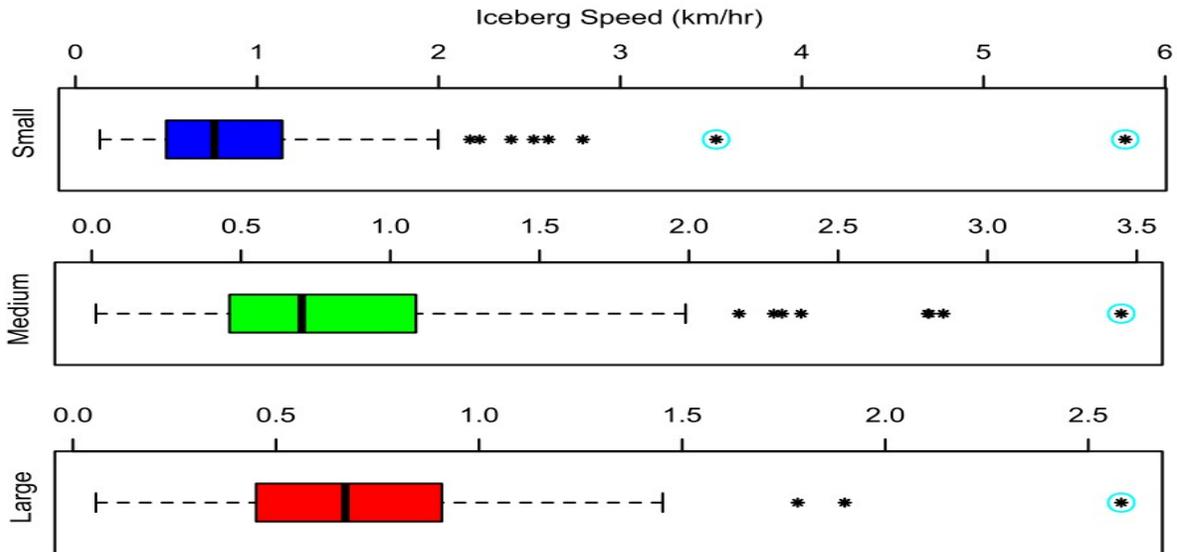


Figure 3.5. Box plot for the small, medium, and large iceberg. The highlighted cyan circles represent rare events (extreme outliers).

1.33 and kurtosis 3.66.). Moreover, for all the three cases in Table 3.1, $Q1-1.5IQR$ gives negative values, and all observed values are positive; this indicates that none of the data sets have a left-tail (Myriam and Pascal, 2013), and visible in Figure 3.6.

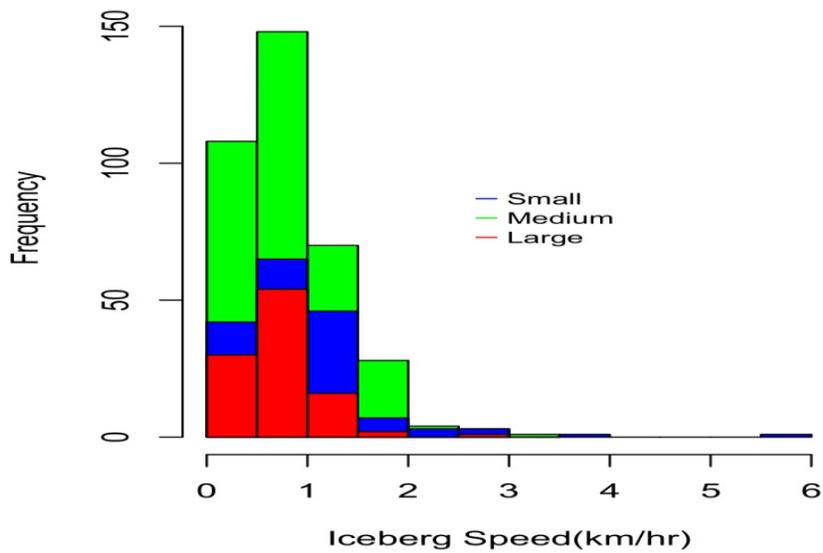


Figure 3.6. Speed distributions for different size of icebergs (2002 -2015). The iceberg sizes are labeled on the top right corner of the plot.

So, in all the three cases means and variances become irrelevant because they fail to explain the tail of the distribution. Data sets for all three iceberg cases include extreme outliers and represent rare events. Therefore, according to the proposed methodology, the classical statistic is not sufficient to model them and, either an outlier or a tail-based approach is required.

3.3.2 Result and Discussion

3.3.2.1 Threshold and Block maxima

The thresholds are chosen using the mean residual life plot, the threshold versus parameter plot, and the Hill plot (Scarrott and MacDonald, 2012). In the mean residual life plot, mean excesses are plotted against the thresholds, and a value is chosen from the threshold range (where the plot shows linearity). A low threshold leads to bias and a high threshold causes high variance, so the threshold is chosen as a trade-off between bias and variance. In the case of the small iceberg, as seen in Figure 3.7, linear behaviour is observed above 1.4 km/hr. The right side of the plot shows high variability due to limited data above such a high threshold, and a confidence interval is not visible for the case

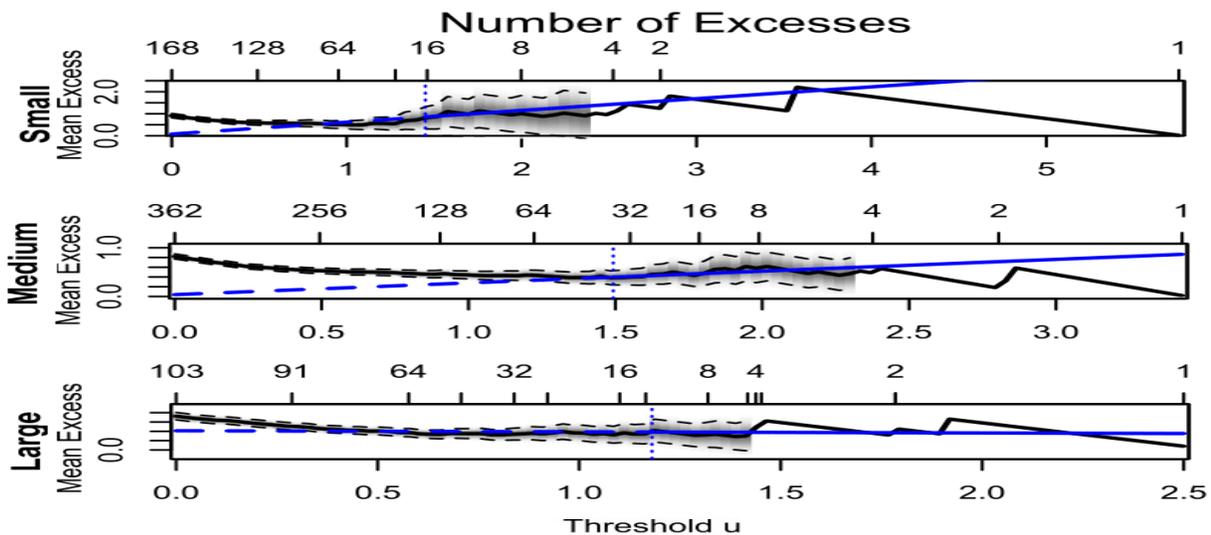


Figure 3.7. Mean residual life plot. The vertical dotted blue line indicates the chosen threshold in all three cases. The units of X-axis and Y-axis are km/hr.

of fewer than five exceedances. The selected threshold is validated through the threshold versus parameter plot. Using similar analyses, the thresholds chosen for the case of medium and large icebergs are 1.5 and 1.2, respectively. In this research, in all three cases, the block size is chosen as equivalent to the number of exceed observations considered in the GPD case. This assumption provides a better comparison framework for comparing the GEV outputs with GPD, instead of splitting the full data sets into annual scales. The maximum value from each block was extracted, and the GEV was fitted on the new data set, which contained only all block maxima.

3.3.2.2 Distributions parameters estimate

The GPD parameters are computed using the MLE. In all three cases, for higher threshold values, the standard error has an increasing trend, and for the lower threshold values, the error has a decreasing trend. Beside MLE, the GPD parameters are estimated with PWM and found the MLE

Table 3.2. Distribution parameters estimate. Corresponding 95% confidence intervals for model parameters are in parenthesis.

Distrib ution Name	Method	Distribution Parameters	Iceberg size		
			Small	Medium	Large
GEV	MLE	Location, μ	1.32 (1.31, 1.33)	1.43 (1.42, 1.44)	1.14 (1.13, 1.15)
		Scale, σ	0.548 (0.546,0.550)	0.367 (0.365,0.368)	0.365 (0.363, 0.367)
		Shape, k	0.230 (0.227,0.233)	0.216 (0.213,0.218)	0.01 (0.005, 0.015)
GPD	MLE	Scale, σ	0.354 (0.351,0.357)	0.318 (0.317,0.319)	0.454 (0.450,0.458)
		Shape, k	0.613 (0.606,0.620)	0.199 (0.198,0.200)	-0.119 (-0.126,-0.112)
HRTD	Hill	Tail index, α	2.60 (2.59, 2.61)	4.70 (4.68, 4.72)	3.80 (3.78, 3.82)
HRTD	SmooHill	Tail index, α	3.10 (3.09, 3.11)	4.00 (3.99, 4.01)	3.60 (3.58, 3.62)

provides a smaller error compared to the PWM estimates in the case of the small and large icebergs. For the medium iceberg, PWM has a marginally better estimate (minimum uncertainty), but this does not have any significant impact on the return level estimates. Therefore, for this study in all three cases, MLE estimates are chosen to estimate return levels. Since MLE was found to give a better fit for GPD, the GEV parameters were estimated using the MLE approach. The parameters values are listed in Table 3.2. In the tail based approach, the tail indices are estimated using the Hill estimator and the SmooHill estimator and listed in Table 3.2. In all cases, the parameter uncertainty is estimated through the proposed method as explained in section 3.4 and placed in Table 3.2.

3.3.2.3 Return level estimate

Compared to GPD, the GEV return level is more robust with data statistics. However, the GPD took a longer time to capture the most extreme outliers in all three cases. The Hill and SmooHill

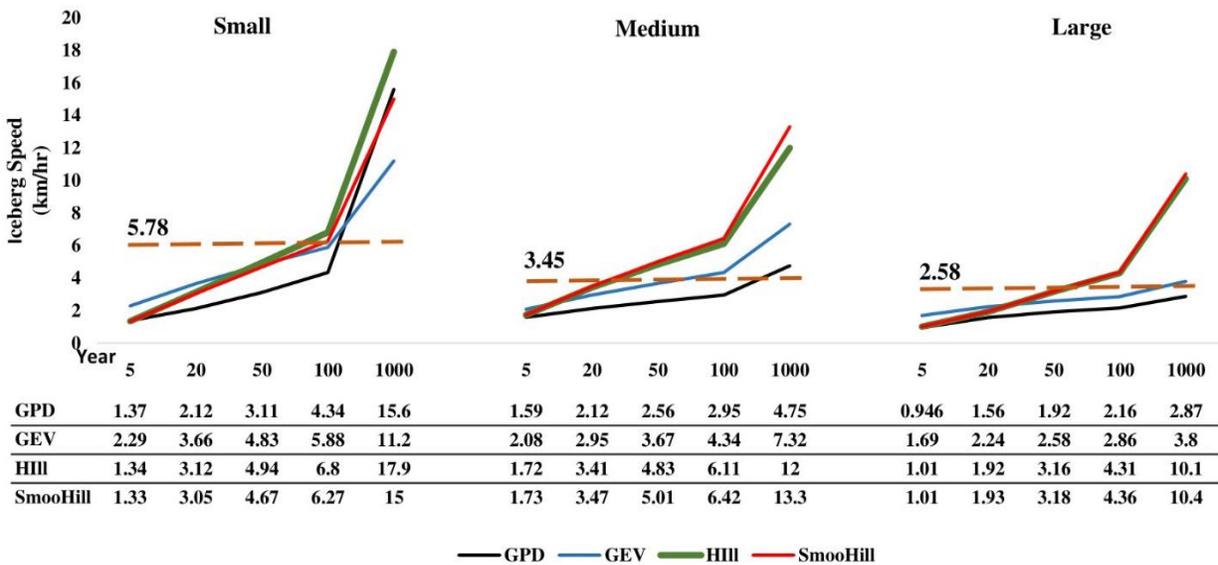


Figure 3.8. The return levels (GPD, GEV, Hill, and SmooHill) plot for three different iceberg cases. The dotted horizontal lines indicate the most extreme outliers. Data tables (return periods, extreme iceberg speed, km/hr) added.

return levels are comparable and have significant differences with GPD and GEV estimate. The return levels are computed (attached in Figure 3.8) by using Equation 3.4 for GPD, Equation 3.6 for GEV and Equation 3.10 for Hill and SmooHill.

3.3.2.4 Goodness of fit test

To choose a best-fitted distribution among Hill and SmooHill, this study considered the tail index. The higher tail index implies a better fit with tail events. In all three iceberg cases, the Hill and Smoohill return level estimates are comparable; however, the highest tail index captures the most extreme event in the early time (Figure 3.8). This study found, in all three cases, based on numerical evidence, the GPD has higher log-likelihood values, and lower Akaike information criterion (AIC) and Bayesian Information Criteria (BIC) which imply a better fit by GPD compared to GEV. However, according to Q-Q (Quantile-Quantile) plot, both GPD and GEV failed to capture the most tail event. The GPD and GEV return levels take a longer time to capture the most extreme events (Figure 3.8), compared to Hill and SmooHill. The SmooHill is more robust, based on data statistics (the most extreme outliers) for the case of the small iceberg; for all other cases, the Hill is more robust. Therefore, in the present study, the iceberg speed estimated by the SmooHill has been chosen to compute the iceberg risk for the case of the small iceberg, and the Hill used for the two other cases.

3.3.2.5 Risk estimation

The iceberg risk is estimated in all three cases by Equation 3.13 and presented in Table 3.3. The occurrence probabilities for different icebergs are computed by Equation 3.15 and summarized in Table 3.3. For example, if the iceberg speed (small) is 1.33 km/hr once in every five years, the

Table 3.3. Iceberg exceedance occurrence probability, extreme speed (km/hr), Kinetic Energy (KE) in Mega Joules (MJ) and Risk (MJ/year) corresponding to the three different icebergs' weight (Megatons, MT). Corresponding 95% confidence intervals for return periods are in parenthesis (Lower Limit, Upper Limit).

Size	Return periods					
	5 years	20 years	50 years	100 years	1000 years	
Small (0.1 MT)	Exceedance Probability	1.03E-02	1.97E-04	2.10E-05	4.22E-06	2.83E-08
	Speed, km/hr (LL, UL)	1.33 (1.32, 1.34)	3.05 (3.04, 3.06)	4.67 (4.66, 4.68)	6.27 (6.25, 6.29)	15.0 (14.9, 15.1)
	KE, MJ (LL, UL)	8.33 (8.20, 8.45)	43.8 (43.5, 44.1)	103 (102, 104)	185 (184, 186)	1060 (1050, 1070)
	Risk, MJ/year (LL, UL)	8.60E-02 (8.47E-02, 8.73E-02)	8.63E-03 (8.57E-03, 8.69E-03)	2.17E-03 (2.15E-03, 2.19E-03)	7.81E-04 (7.77E-04, 7.85E-04)	2.99E-05 (2.97E-05, 3.01E-05)
Medium (2 MT)	Exceedance Probability	1.55E-03	1.56E-05	1.21E-06	2.01E-07	8.42E-10
	Speed, km/hr (LL, UL)	1.72 (1.71, 1.72)	3.41 (3.40, 3.42)	4.83 (4.82, 4.84)	6.11 (6.10, 6.12)	12.0 (11.9, 12.1)
	KE, MJ (LL, UL)	279 (275, 283)	1090 (1088, 1092)	2200 (2190, 2210)	3510 (3500, 3520)	13600 (13300, 13900)
	Risk, MJ/year (LL, UL)	4.34E-01 (4.28E-01, 4.40E-01)	1.70E-02 (1.69E-02, 1.71E-02)	2.67E-03 (2.66E-03, 2.68E-03)	7.06E-04 (7.04E-04, 7.08E-04)	1.15E-05 (1.13E-05, 1.17E-05)
Large (10 MT)	Exceedance Probability	1.87E-02	4.07E-04	2.39E-05	3.61E-06	1.33E-08
	Speed, km/hr (LL, UL)	1.01 (1.00, 1.02)	1.92 (1.91, 1.93)	3.16 (3.15, 3.17)	4.31 (4.30, 4.32)	10.1 (10.0, 10.2)
	KE, MJ (LL, UL)	480 (471, 489)	1740 (1720, 1760)	4700 (4670, 4730)	8740 (8700, 8780)	48000 (47100, 48900)
	Risk, MJ/year (LL, UL)	8.97E+00 (8.81E+00, 9.13E+0)	7.08E-01 (7.00E-01, 7.16E-01)	1.12E-01 (1.11E-01, 1.13E-01)	3.15E-02 (3.13E-02, 3.17E-02)	6.36E-04 (6.24E-04, 6.48E-04)

probability of an iceberg speed less than or equal to 1.33 km/hr is 0.587 (for the small iceberg, $\alpha = 3.1$). Therefore, the probability of iceberg speed being greater than 1.33 km/hr is 0.413. This is a conditional probability; for unconditional probability, 0.413 should be multiplied with $p_u(0.125)$. Finally, the occurrence probability of the iceberg speed of more than 1.33 km/hr is 0.0103. The threshold's exceedance rate, p_u , is computed by finding the data percentile from raw data. This study uses R package "ismev" to find the exceedance rate. Estimated iceberg speed and kinetic energy computed by Equation 3.16 are presented in Table 3.3. Iceberg weights are considered as 0.1, 2 and 10 megatons for the small, medium and large iceberg, respectively (MANICE, 2005). As shown in Table 3.3, in the case of the small iceberg, for a 100 years return level is 6.27 km/hr, which means that in the Flemish Pass basin, once in every 100 years, the iceberg speed may be higher than 6.27 km/hr (with KE 185 MJ). For the case of the medium and large icebergs, it is 6.11 km/hr (with KE 3510 MJ) and 4.31 km/hr (with KE 8740 MJ), respectively. If the average speed is considered over the years (for the period 2002 to 2015) for small, medium and large icebergs speeds are 0.914 km/hr, 0.822 km/hr, and 0.726 km/hr and the corresponding KE are 3.94 MJ, 3.18 MJ, and 2.48 MJ. If a large iceberg collides with a structure in the Flemish Pass basin, the computed KE 480 MJ (as a 5-year return level) is more likely to create a more severe unwanted situation compared to KE 2.48 MJ (computed based on average large iceberg speed). The risk is estimated by assuming iceberg speed exceeds the estimated speeds and that the iceberg collides in the study area. In Table 3.3, the extreme risk is highlighted in bold (**3.15E-02 MJ/year**). The corresponding speed is selected as the extreme iceberg speed for the design purpose, which is estimated as 4.31 km/hr.

3.4. Conclusions

Rare event risk analysis is a critical exercise to design and operate engineering systems in harsh and remote environments. The traditional risk analysis approaches (based on classical statistics or

extreme value theory) fail to capture the extreme characteristics of rare events and, therefore, are unable to provide accurate outcomes. This study proposes a simple yet rigorous framework to categorize the existing extreme theories for rare event risk analysis for engineering design perspective. The methodology is new, while models used in the methodology are well known and practiced models. The proposed methodology, based on the flexibility concept, provides a transparent, auditable and flexible means to quantify risk and its use in design consideration. The framework comprises three phases: outlier-based analysis, tail-based analysis and risk estimation. The framework is applied to a natural hazard (iceberg collision) scenario associated with an offshore facility. The extreme iceberg speed is estimated using the rare event modeling framework, where the rare event (extreme outlier event) is considered as a heavy tail event. The tail index is estimated through the Hill estimator and the SmooHill estimator. An uncertainty quantification approach is introduced for model parameters, return levels and iceberg risk. The proposed methodology has been applied in three cases for three different sizes of icebergs (small, medium and large icebergs). Although the POT based GPD and the BM based GEV are frequently used in various fields for extreme analysis, both were found to be not adequate for the case study. The tail-based approaches predict the most extreme outliers in shorter time periods compared to the outlier-based approach. The SmooHill is more robust compared to the Hill in the case of the small iceberg; for the other two cases, the Hill is more robust. Finally, if an iceberg collides with a structure in the Flemish Pass basin area, the estimated risk provides a more accurate indication of damage compared to that obtained using average iceberg speed. Uncertainty analysis is also considered that provide certain confidence on the analysis and robustness. The result from the case study shows that the estimated iceberg risk associated with individual extreme iceberg speed is more significant compare to the traditional iceberg risk analysis (predicated on average iceberg speed).

Therefore, to avoid any unwanted situation (for drilling or engineering facilities) or designing a system flexible under rare event scenario, the estimated risk must be considered in the iceberg risk management plan at the Flemish Pass basin. For future studies, a non-stationary approach will be implemented for different natural hazards (for example, wind speed, wave height etc.) relevant to the Flemish Pass Basin.

4.

EVOLVING EXTREME EVENTS CAUSED BY CLIMATE CHANGE: A TAIL-BASED BAYESIAN APPROACH FOR EXTREME EVENT RISK ANALYSIS

Preface

A version of this manuscript has already been published by the Journal of Risk and Reliability. This research paper was primarily written by me. Under the supervision of co-authors Faisal Khan, Salim Ahmed, and Syed Imtiaz, a heavy right tail model (Bayesian inference) is implemented to compute iceberg collision risk assessments at the Jeanne d’Arc basin by incorporating climate change issue. I conducted the research review, collected the data, developed methodology, and carried out the analysis and modelling. The co-authors contributed to concept development, methodology writing, review & editing, and verified modeling outcomes. The initial draught of the manuscript was written by myself, and I later made changes considering feedback from the other authors and the peer-review process. Project management and funding acquisition came within the responsibility of co-author Faisal Khan. This collaborative effort allowed us to bring together different expertise and insights to produce a comprehensive and robust study on iceberg collision risk assessment at the Jeanne d’Arc basin, considering the complexities of climate change.

Abstract

Natural hazards are of significant concern for engineering development in the offshore environment. Climate change phenomena are making these concerns even greater. The frequency and extent of natural hazards are undesirably evolving over time; so risk estimation for such events

require special consideration. In most cases the existing extreme models (based on the extreme value theory) are unable to capture the changing frequency and extreme characteristics of natural hazards. To capture the evolving frequency and extreme characteristics of natural hazards and their effects on offshore process operations, an advanced probabilistic approach is proposed in this paper. The approach considers a heavy right tail probability model. The model parameter is estimated through the Bayesian inference. Hill and the SmooHill estimators are used to evaluate the lowest and highest exponent of the probability model. The application of the approach is demonstrated through extreme iceberg risk analysis for the Jeanne d'Arc basin. This study shows climate change or global warming is causing to appear a significant number of icebergs every year in the study area. Offshore structures are often designed to withstand the impact of 1 MT icebergs weight; however, the study observes large icebergs (10 MT weight) are sighted in recent years (14% of the total number of cited icebergs for the period of 2002-2017). As a result, the design philosophy needs to be revised. The proposed risk-based approach provides a robust design criterion for offshore structures.

Keywords: Climate change, rare event, iceberg risk, Heavy tail, Bayesian inference, Hill, SmooHill estimator.

4.1. Introduction

Offshore development and existing offshore facilities face several risk factors; the “natural hazard” is one of them. Depending upon the region, this may include hurricanes, earthquakes, icebergs, sea ice, extreme waves or combinations of different hazards. The catastrophic behaviour of these hazards is hard to predict and extremely challenging to control (e.g. Hurricane Katrina; August 2005, Gulf of Mexico, extreme wind speed was recorded as 280 km/hr). Currently, the climate change is treated as the prime factor for all the extreme effects of such natural hazards. For this

reason, "climate change" or "global warming" issues are the priority for most research. The Intergovernmental Panel on Climate Change IPCC report (IPCC, 2014) states the global temperature increased 1.2 degrees Celsius, compared to the pre-industry era. As a direct consequence, the global mean sea level rose by 17 to 21 cm during the period from 1901 to 2010 (Castillo et al., 2005). In particular, global warming or extreme heat waves lead to ice sheets melting, increase the frequency of icebergs, and cause the sea level to rise. Scientists around the globe have a concern regarding the rising sea level. To monitor or mitigate the effects of such natural threats, a list of recent works offer alternative ideas for the production/development flexibility of offshore fields to ensure higher degree of protection in a harsh environment. For example, in the case of offshore facilities, please refer to the Deepwater Artificial Seabed system (Zhen et al., 2020a), the Offshore Resource Centre (Rahman et al., 2020) and Quantitative Risk Analysis approach (Zhen et al., 2020b). The focus of this study is, however, risks from floating icebergs. Every year a significant number of icebergs come across Baffin Bay from west Greenland and include some from east Greenland. The Labrador current transports some of the icebergs to Newfoundland oil drilling zone (e.g. Jeanne d'Arc basin). Such a natural hazard cannot be ignored, due to the iceberg alley. Climate change causes, the present day extreme characteristics of natural hazards to have more impact, compared to the pre-industrial time. For example, ocean waves and wind have been getting higher and stronger in the past 30 years (Young and Ribal, 2019a). This has a direct impact on iceberg speed and direction. Moreover, the recent European heat wave (Climate signal beta, 2019) cannot be ignored and might dangerously increase iceberg frequency and size in the near future. In an extremely harsh environment, some icebergs might have a chance to move with enormous kinetic energy and cause damage or create an unwanted situation for

offshore engineering facilities. Therefore, integrating climate change into the iceberg risk estimation by incorporating iceberg extreme speed is necessary.

In an average (say wind speed 6 to 10 m/s and current speed 0.1 to 0.5 m/s) weather conditions, icebergs move at a certain average speed, following a regular path. For this reason, average iceberg speed is used in traditional iceberg risk analysis (King et al., 2015; Husky energy, 2000). However, how do icebergs react in extreme weather conditions? There is no precise answer to this question because the weather has a chaotic nature, and climate change is causing the increasing likelihood and intensity of extreme events. In the present climate change scenario, it is likely that an iceberg moving with unexpected speed in an irregular direction or path could collide with any offshore facilities. In this research, such an extreme scenario (iceberg collision) is treated as a rare event. The iceberg risk estimation problem is considered as a heavy tail event problem: a most extreme iceberg speed and its occurrence probability prediction are the key focus. There is no database for iceberg speed which makes extreme iceberg speed estimation and its risk analysis more complicated.

In most fields, the well-established Extreme Value Theory (EVT) based models are applied to model such an extreme event (Castillo, 2005; Das et al., 2016; Damon, 2009; Scarrott and Macdonald, 2010; Asadi, and Melchers, 2017). The Peak Over Threshold (POT) based Generalized Pareto Distribution (GPD) and Block maxima (BM) based Generalized Extreme Value (GEV) are frequently used in all domains. Some present real-life rare phenomena such as the 9/11 terrorist attack, blowout events like the Macondo blowout accident in the Gulf of Mexico, hurricane Katrina make the implementation of the EVT questionable. Insufficient data is a major issue when modelling a rare event. In the engineering domain it is more complicated (Rocchetta et al., 2015; Yang et al., 2015; Rathnayaka et al., 2013; Sulistiyono et al., 2015; El-Gheriani et al.,

2017b). In such cases, the EVT based methodology fails to capture most extreme events, and as an alternative, a rare event is considered as a heavy tail event. A tail-based approach has been proposed by (Arif et al., 2020), and considered the 1000 year return period to capture most extreme scenarios. A similar approach was also proposed by (Clauset et al., 2013), The Monte Carlo method (Rubino and Tuffin, 2009; Estecahandy, 2015) and the Mixer model (Hanum et al., 2015) have been considered in past research for rare event modelling. Model parameter estimation in the case of an unusual event may not be straightforward. For example, the maximum likelihood estimator may result in a biased estimate if the sample space is small (Deidda and Puliga, 2009). The Probability Weighted Moment (PWM) also gives a poor estimate compared to MLE if shape parameter values are less than zero (Pham et al., 2019). To address this issue the Hill (Hill, 1975) and SmooHill (Resnick et al., 1997) estimator are implemented to estimate the tail index (Arif³ et al., 2020). However, all point-based methods does not provide a probability distribution over the possible range of the model's parameter and does not incorporate prior information (Deidda and Puliga, 2009). To overcome this issue, Bayesian inference has shown some potential in recent research (Deidda and Puliga, 2009; Pham et al., 2019). The main objective of the present work is to model heavy tail events (e.g. extreme iceberg speed, process accident, terrorist act, etc.) that have taken place in recent times. In this paper, a most extreme event is considered as a heavy tail event (Resnick, 2007), and a POT based Heavy Right Tail distribution (HRTD) is implemented similar to (Arif et al., 2020a; Clauset et al., 2013). Bayesian inference is used to estimate the distribution parameter, and finally, the extreme risk is estimated by using the model's prediction of extreme iceberg speed.

The remainder of the paper is structured as follows: Section 4.2 describes the proposed methodology to model most extreme events. Section 4.3 provides an overview of the case studies and presents results. Section 4.4 provides conclusions and recommendations for future works.

4.2. Methodology

4.2.1 The proposed framework

Finding a best-fitted probabilistic distribution is a straightforward task in some cases. However, a heavy tail event has two characteristics: data are scarce and the system's behaviour is highly influenced by one or more than one large data value.

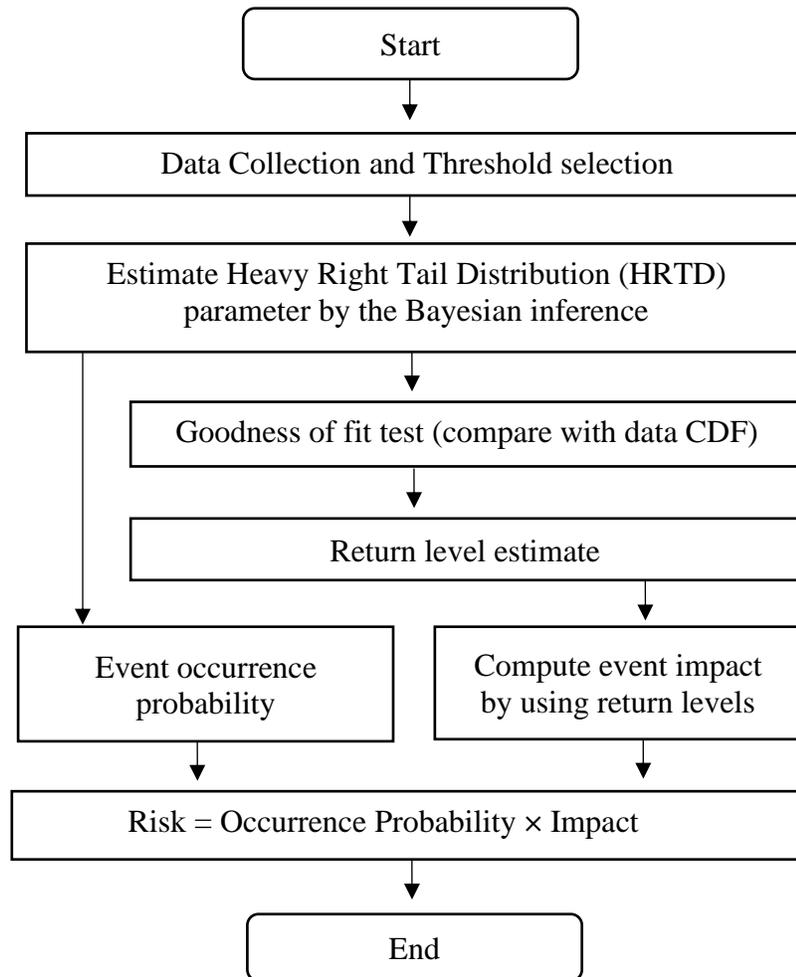


Figure 4.1. The proposed framework for heavy tail event risk analysis

Those two characteristics making the modeling approach complicated. In this research, a Bayesian framework is proposed for the HRTD parameter estimation. The parameter is examined for three different threshold values instead of for a fixed threshold. These three different estimates allow comparing the model prediction capability. The Hill (Hill, 1975) and the SmooHill (Resnick and Starica, 1997) estimators are used to evaluate the possible parameter range of HRTD to execute Bayesian inference. This also gives a frame of reference to justify the Bayesian estimate. Finally, the tail event occurrence probability and its impact are used to compute risk. The occurrence probability is computed directly from the probability distribution function of HRTD. The impact is calculated as far interest through estimated return levels. In the present study, iceberg kinetic energy is computed as iceberg impact. The proposed framework to study the rare event is presented in Figure 4.1 and assumes data are Independent and Identically Distributed (IID).

4.2.2 Heavy Right Tail Distribution (HRTD)

Let X be a heavy tail random variable. The HRTD is mathematically defined as (Clauzet et al., 2009):

$$P[X > x] \sim f(x) = \frac{\alpha-1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}, \quad (4.1)$$

for the probability, $X \leq x$, HRTD is defined as (Clauzet et al., 2009):

$$P[X \leq x] \sim F(x) = 1 - \left(\frac{x}{x_{min}} \right)^{-\alpha+1}, \quad (4.2)$$

where $x_{min} > 0$ is the scaling factor, $\alpha > 1$ is the shape parameter (tail index) and $x > 0$. In this research we assume the scaling factor (x_{min}) to be the same as the threshold value (u), as the proposed rare event modeling approach keeps the large values separate from the bulk of distributions, similar to the aspect of the Peak Over Threshold (POT) approach (Das et al., 2016). Equations 4.1 and 4.2 give a clear indication that, as the tail index goes higher, the distribution has a more massive tail. In this research, return value function is derived with a similar approach as in

the work of (Bommier, 2014). Consider that p_u is the probability of occurrence of an exceedance of a high threshold u . Mathematically, it is expressed as $p_u = P(X > u)$. Therefore, by considering $x_{min} = \text{threshold}, u$ from Equation 4.1,

$$P[X > x] = p_u \left[\frac{\alpha-1}{u} \left(\frac{x}{u} \right)^{-\alpha} \right].$$

Then the return level x_m that is expected on an average, once every m observation is obtained by solving:

$$p_u \left[\frac{\alpha-1}{u} \left(\frac{x}{u} \right)^{-\alpha} \right] = \frac{1}{m},$$

which implies (replace x by x_m),

$$x_m = u^{(1-\frac{1}{\alpha})} [p_u \times m \times (\alpha - 1)]^{\frac{1}{\alpha}}. \quad (4.3)$$

The N-year return level is defined as $m = N \times n_y$, where n_y is the number of events per year.

Therefore Equation 4.3 can be rewritten as

$$x_m = u^{(1-\frac{1}{\alpha})} \times [p_u \times N \times n_y \times (\alpha - 1)]^{\frac{1}{\alpha}} \quad (4.4)$$

Equation 4.4 gives the N-year return level and is called the return level function of HRTD. In order to determine N-year return level, four parameters, namely, threshold u , the exceedance rate p_u , the number of events per year, n_y and the HRTD parameter (α) need to be determined. This research computes the exceedance rate (p_u) as a ratio of the number of events exceeding the threshold to the total number of observations. Finally, the HRTD parameter α is estimated through Bayesian inference.

4.2.3 Bayesian inference

To sufficiently estimate uncertainty, a large sample size is a prerequisite for the case of a Maximum Likelihood Estimator (MLE). In the Bayesian approach, a prediction consists of an expected value together with an associated uncertainty (Lee, 2007) and there is no need for a separate step for

uncertainty analysis because the Bayesian framework derives a full posterior distribution of the parameters. The posterior distribution $f(\alpha|X)$ of the exponent variable α given the observed data, $X = x_1, x_2, \dots, x_n$, can be written as:

$$f(\alpha|X) = \frac{f(X|\alpha)f(\alpha)}{f(X)}, \quad (4.5)$$

because the data are conditionally independent, and the likelihood function can be written as

$$f(X|\alpha) = \prod_{i=1}^n f(x_i|\alpha). \quad (4.6)$$

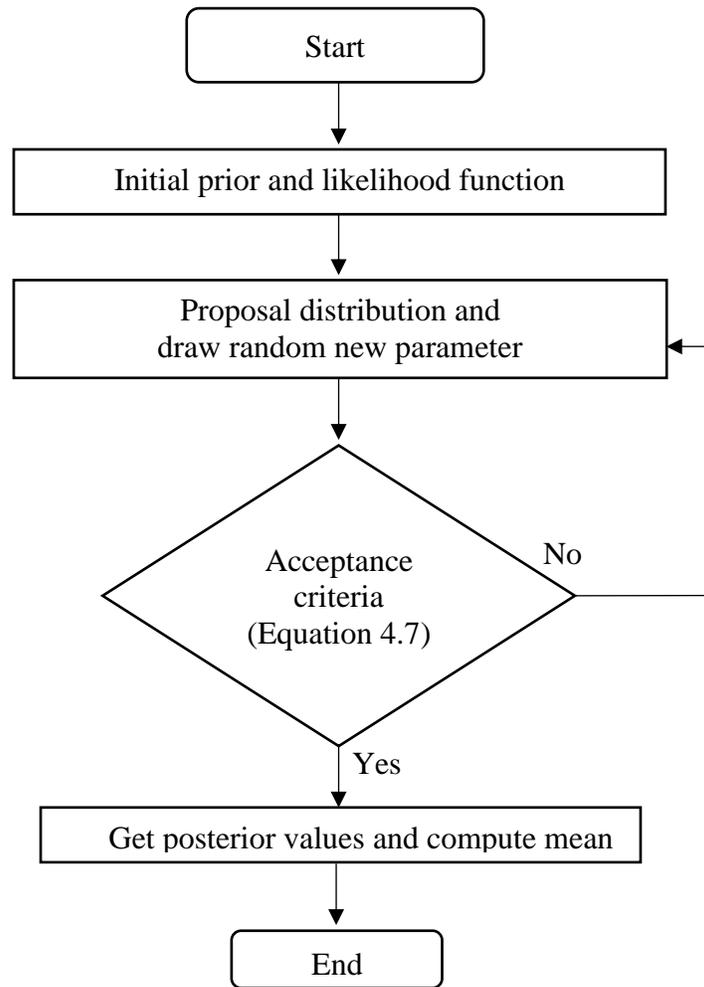


Figure 4.2. Bayesian inference flow chart (MCMC, Metropolis-Hasting algorithm).

The prior $f(\alpha)$ is considered as Jeffrey’s prior (Foreman-Mackey et al., 2013). In this research, Markov chain Monte Carlo (MCMC) sampling (Figure 4.2) with the Metropolis-Hasting algorithm

(Estecahandy et al., 2015) is implemented to approximate the desired distribution. The Metropolis-Hastings algorithm needs a proposal distribution such as $Q(\alpha'|\alpha)$, from which samples can be drawn. This proposal distribution is used to do a random walk (Markov chain) in the distribution space, by accepting or rejecting samples on the basis of how well the parameter fits the distribution $f(X|\alpha)$. At each iteration the new parameter α' drawn from $Q(\alpha'|\alpha)$ is conditioned on the current sample α . To decide whether to accept or reject the new value α' , the following ratio needs to be computed for each new proposed α' : $\frac{f(\alpha'|X)}{f(\alpha|X)}$ Using Equation 4.6, this expression can be expressed

as: $\frac{f(X|\alpha')f(\alpha')}{f(X|\alpha)f(\alpha)}$, and is equivalent to $\frac{\prod_{i=1}^n f(x_i|\alpha')f(\alpha')}{\prod_{i=1}^n f(x_i|\alpha)f(\alpha)}$, where f is the PDF of the distribution to the sample as expressed in (4.1). The rules for acceptance can be formulated as (Foreman-Mackey et al., 2013):

$$f(\text{accept}) = \min \left(1, \frac{\prod_{i=1}^n f(x_i|\alpha')f(\alpha')}{\prod_{i=1}^n f(x_i|\alpha)f(\alpha)} \right). \quad (4.7)$$

Therefore, if α' is more likely than the current α , it is accepted. A plausible starting point is always demanding for convergence to the target distribution. However, in the case of rare event modelling, one or few data points lead the distributions and cause a heavy tail. For focus on the tail event, instead of using any random values, the Hill and the SmooHill estimators have been used to estimate the possible parameter range. The 95% confidence interval is estimated for both the Hill and the SmooHill estimate, as explained in (Arif et al., 2020a); and it has following steps:

- 1 Generate a bootstrap data set: Use HRID to generate a random sample with estimated model parameter.
- 2 Fit a HRTD on bootstrap data set, update model parameter and estimate return levels.

- 3 Repeat steps 1 and 2 say 1000 times and obtain 1000 values for model parameter and 1000 values for each return period, m.

Determine the mean value and a 95% confidence interval for the model parameter and return level for different time periods. The 95% confidence interval is computed as (mean $\pm 1.96 \times$ standard error), where 1.96 = significance level and standard error = $s/\sqrt{1000}$, and s is the standard deviation.

4.2.4 Risk estimation

The iceberg risk estimation process follows the same as those in (Sulistiyono et al., 2015) and (Arif^a, 2020). The risk is defined as:

$$R(\text{iceberg}) = P(O) \times C, \quad (4.8)$$

where $P(O)$ is occurrence probability and C is the event consequences (impact). The event occurrence probability is defined as:

$$P(O) = p_e \times p_u \times (RP)^{-1}, \quad (4.9)$$

where p_e is the probability of the estimated event, p_u is the exceedance rate ($\Pr [X > u]$, where u is the threshold) and RP is the return period. The terms p_e and p_u are used to get the unconditional probability estimate. As a part of the consequences (C) or impact, this research considered the kinetic energy of the iceberg. The Kinetic energy is denoted by KE, and defined as (Husky energy, 2000):

$$KE = \frac{1}{2} M \beta v^2, \quad (4.10)$$

where M is the iceberg mass, v is the iceberg speed, and β is a constant (added mass coefficient).

The extreme impact and corresponding occurrence probability might be used to generate extreme risk scenarios for different return periods. The risk-based design is proposed considering appropriate risk guidelines. The offshore operator needs to select acceptable risk criteria. The acceptance criteria are influenced by the regulatory regime, corporate culture of the operator as

well as operating conditions. As Low As Reasonably Practicable (ALARP) risk acceptance framework is often used to determine the acceptable risk level (Pike et al., 2020). This study considered an acceptable risk value of 0.5 MJ/year as a guiding example on how iceberg extreme design speed might be chosen. For a specific operation, the acceptable value can be significantly different. For example, under rare event conditions, a risk engineer has a plan to estimate extreme wave height risk in any specific region, and estimates several risk levels $3.1E-01$, $4.2E-02$, $5.2E-03$, and $5.9E-05$ kw/m for different return periods: 5, 20, 50 and 100 years respectively. Therefore, by incorporating an acceptable risk (say $3E-03$ kw/m) for a specific engineering system, a proposed design risk might be $4.2E-02$ kw/m. This design risk is associated with the corresponding extreme wave height, its occurrence probability and wave power (impact).

4.3. Case Study

The proposed methodology is implemented for a large iceberg collision risk estimate, incorporating estimated extreme iceberg speed and its occurrence probability for the Jeanne d'Arc basin. This area is an offshore sedimentary basin located about 340 kilometers from the basin centre, east-southeast of St. John's, Newfoundland and Labrador. The Hibernia, Terra Nova and the White Rose oil drilling operations are making Jeanne d'Arc basin an important oil drilling zone. However, through the iceberg alley, every year a significant number of icebergs arrive at the Jeanne d'Arc basin from west Greenland. Between iceberg formations and the melting process, in extreme weather conditions, the icebergs might have a chance of moving with a heavy kinetic energy, the key focus of this study. The iceberg risk is evaluated considering the probability of iceberg collision with the structure and its impact on the study area. In this research, the iceberg speed was estimated through the generic method (distance/time). Iceberg sight locations and corresponding times were taken from the International Ice Patrol (IIP) iceberg sighting database

(International Ice Patrol, 1995). The primary study region of interest is the Jeanne d'Arc basin (46° N to 47.5° N, 47° W to 50° W) for the period 2002 to 2017. In this study period, the average iceberg speed was computed as 0.91 km/hr through the study area, which is equivalent to the average iceberg speed listed for the same area, of 1.07 km/hr (King et al., 2015). The average value computation is not of interest. Instead of average speed, this research focuses on extreme individual iceberg speed, which is listed as 8.48 km/hr for the period 2002 to 2017 at Jeanne d'Arc basin. The open-source platform Python (freely available at www.python.org) and R (freely available at www.r-project.org) are used to do the data modeling.

4.3.1 Descriptive statistics

Every year a significant number of icebergs are sighted in the Jeanne d'Arc basin area. The only source of icebergs is the Greenland ice sheet, melting as a direct consequence of climate change. According to data statistics, during the period 2002 to 2017, in Jeanne d'Arc basin, medium (MED) size icebergs were found to be more frequent (3167), which is about 45% of total icebergs sighted in that area (Figure 4.3). As shown in Figure 4.3, 1992 small (29%) and 988 (14%) large icebergs were sighted. Other icebergs, Growlers (GR), Bergy Bits (BB), very large (VLG) and Radian (Rad) were also sighted in this area, and together, their number is 834 (12%). Here, the large iceberg is taken into consideration to implement the proposed methodology. Large iceberg weight is listed as 10 MT (MANICE, 2005) (megatons) and this moves with huge kinetic energy. In general, the iceberg frequency distribution has followed a sine wave in the full study period, and the highest peak was recorded in 2015, and after that, there is a clear decreasing trend (except for small icebergs). In the case of large icebergs, the highest frequency is 225 in 2015, and 13 large icebergs

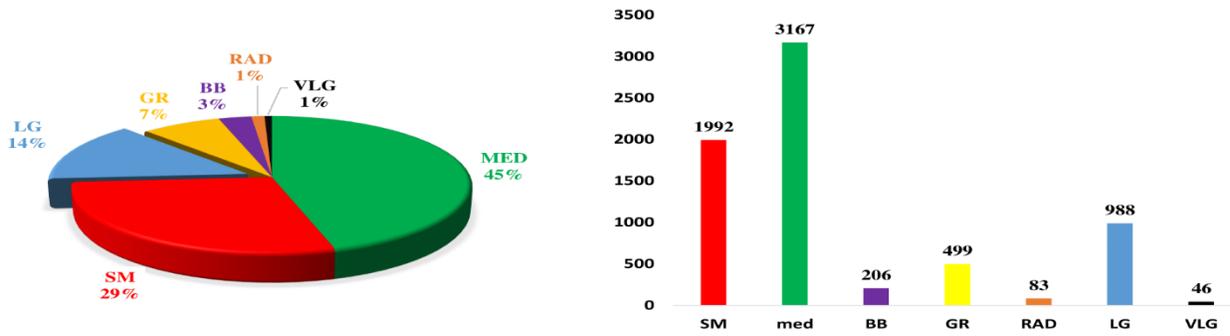


Figure 4.3. Iceberg frequency distribution at the Jeanne d'Arc basin (2002-2017).

are listed in 2017. (Figure 4.4). In Figure 4.5, yearly maximum iceberg speed is displayed along with yearly average iceberg speed. The highest maximum iceberg speed is 8.48 km/hr (in 2002), and the lowest is 1.35 km/hr (in 2007). However, the maximum average speed is only 1.28 km/hr (in 2002), which is around one- sixth of the maximum highest speed (1.28 vs. 8.48 km/hr).

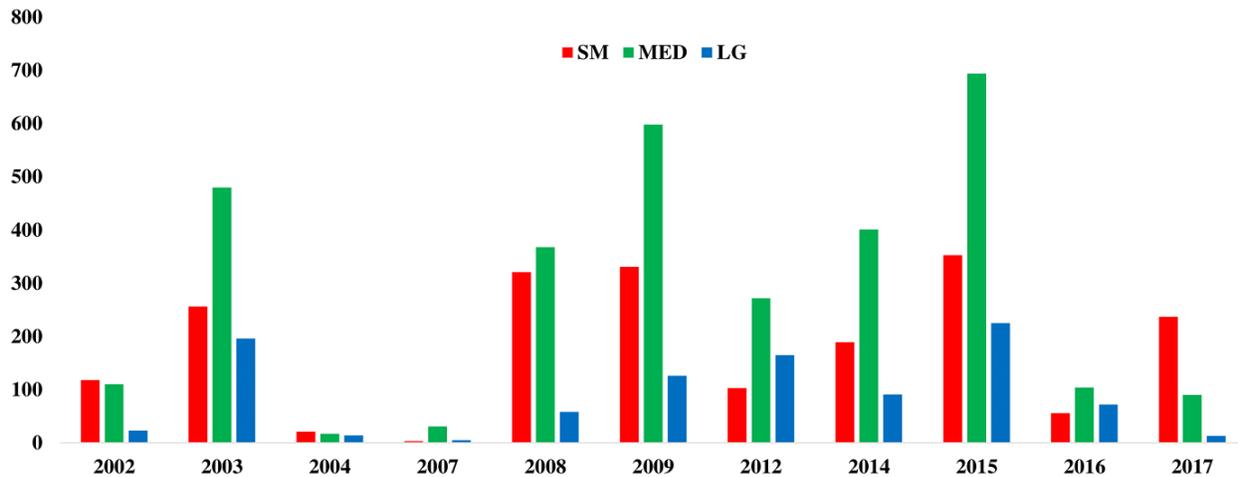


Figure 4.4. Iceberg (Small, Medium and Large) temporal (2002 to 2017) distribution.

Table 4.1 presents the large iceberg speed data statistics. According to the definition of a rare event in Arif et al., 2020a, five observations are considered as rare events (data > 3 km/hr), which are also

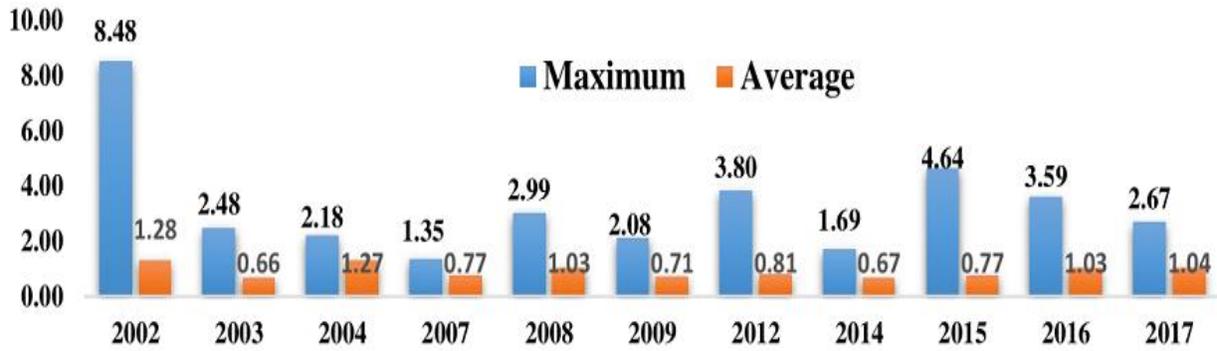


Figure 4.5. Large Iceberg speed (maximum vs. average) over the period 2002 to 2017.

displayed in the boxplot Figure 4.6. The rare events are separated from the regular outliers by the horizontal red line. However, skewness (4.080) and kurtosis (39.10) indicate that this data set does

Table 4.1. Descriptive statistics: Jeanne d'Arc basin iceberg speed data (2002-2017).

Large Iceberg data Statistics (Iceberg speed, km/hr)								
Minimum	Q1	Median	Q3	Maximum	IQR	Q3+3IQR	Skewness	Kurtosis
0.0270	0.4105	0.6692	1.058	8.483	0.6475	3.000	4.080	39.10

not follow the normal distribution, as the skewness and kurtosis of the data are far from 0. The positive skewness indicates that the size of the right tail is larger than the left tail. The kurtosis > 3 ; implies the data set has a heavy tail, which is also displayed in Figure 4.6 and Figure 4.7. These statistics provide clear evidence that the Greenland ice sheet is melting, and forming icebergs, which may create an unwanted situation in the study area for marine facilities. Moreover, the climate is changing, and the recent European heat wave (Climate signal beta, 2019) also might cause increasing iceberg frequency and intensity in the near future.

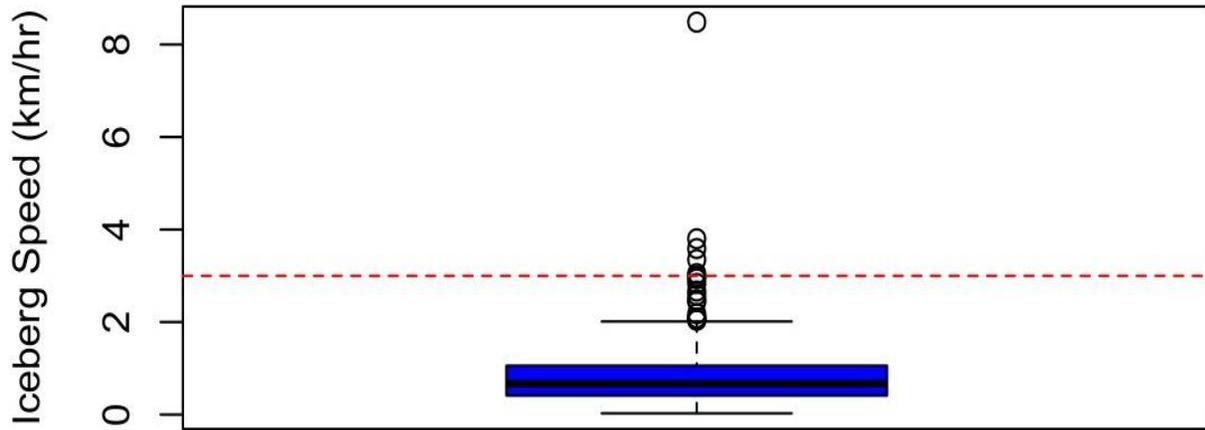


Figure 4.6. Box plot for the large iceberg. All data points above the red line are rare events (extreme outliers).

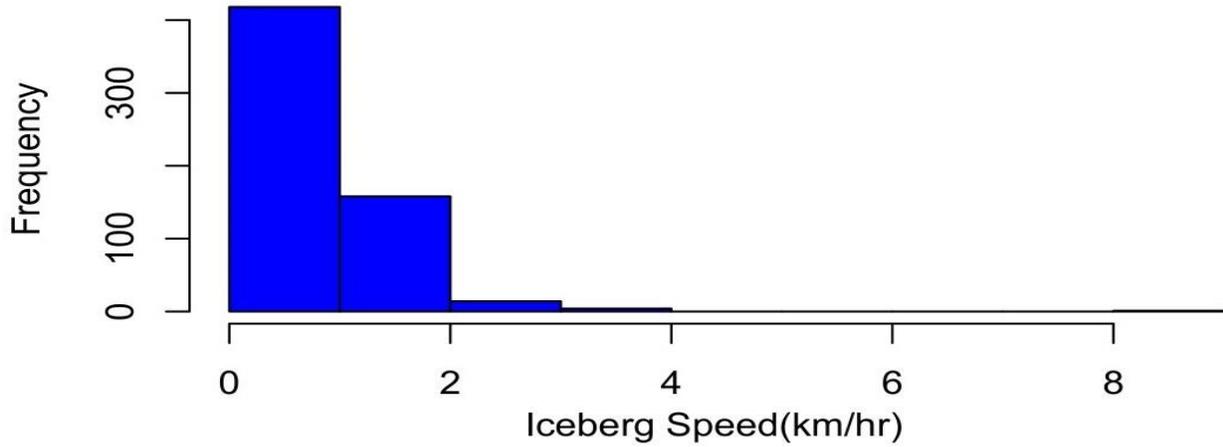


Figure 4.7. Histogram: The large iceberg speed distribution (2002 -2017).

4.3.2. Result and Discussion

4.3.2.1 Threshold selection

The thresholds are chosen using the Normal Q-Q plot (Figure 4.8) and mean residual life plot (Figure 4.9). The thresholds values are also justified with a parameter stability plot. A low threshold leads

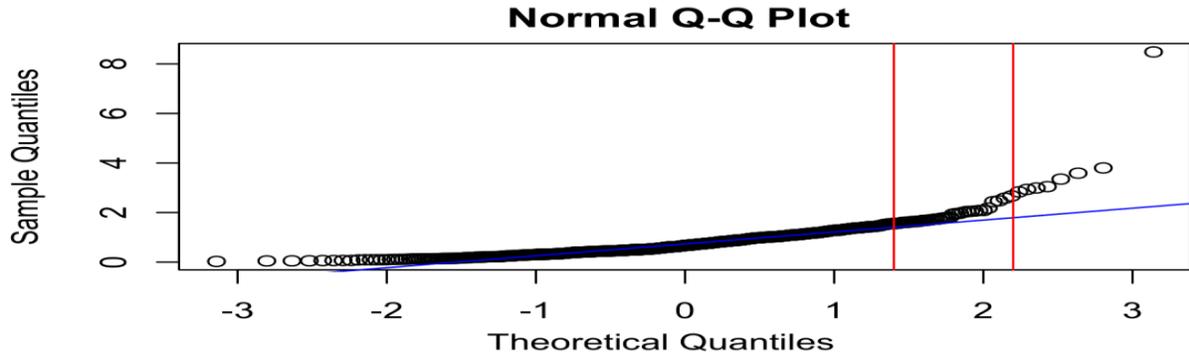


Figure 4.8. Normal Q-Q plot. The red vertical lines are showing the threshold range.

to bias and a high threshold causes high variance (Scarrott and MacDonald, 2012). Therefore, the threshold is chosen as a tradeoff between bias and variance. In the Normal Q-Q plot (Figure 4.8), the extreme values start to deviate in the range of 1.4 to 2.2, and above the threshold points, the mean residual life plot (Figure 4.9) has reasonable linearity; however, below this level, the curve shows biased curvature.

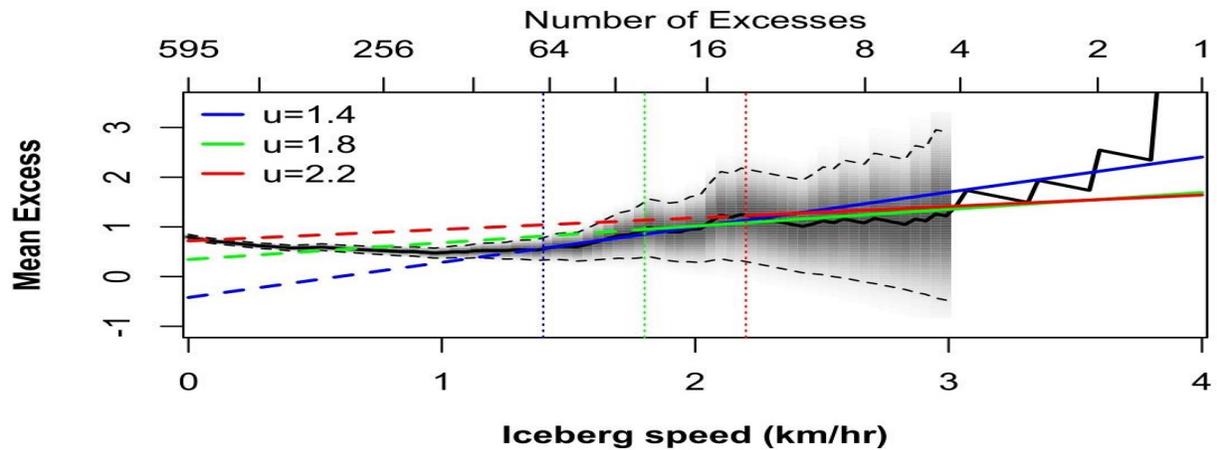


Figure 4.9. Threshold selection procedure (mean excess vs iceberg speed). The number of excesses is labeled at the top of the plots. Threshold values are highlighted with vertical lines.

The model parameter stability plots have a similar trend. Instead of a fixed threshold, finally, three thresholds (1.4, 1.8, and 2.2) are chosen, and the HRTD parameter is estimated for their corresponding threshold values.

4.3.2.2 Parameter estimate (Bayesian inference)

To estimate the HRTD parameter, this research used `Bayespowerlaw.bayes` function of the Python `BayesPowerlaw` package. This function fits the data to HRTD and determines its parameter by using Bayesian inference (MCMC, Metropolis-Hastings algorithm). To execute this function requires several initial setups.

- Parameter range estimate: The HRTD parameter range is estimated through the Hill (Hill, 1975) and SmooHill (Resnick and Starica, 1997) estimators, and estimated values are listed in Table 4.2. The details of the implementation of the Hill and SmooHill estimators are described in (Arif et al., 2020a). Table 4.2 gives an estimate of the minimum and maximum values of the parameters of the HRTD.
- The thresholds values are used as the lowest value from the data to fit HRTD.
- Number of MCMC iterations as set as 10000.
- Standard deviation of the sampling step size during MCMC for both gamma (first value) and weight (second value) are estimated from the Standard Deviation (SD) of the data set. More precisely, for the threshold 1.4, SD is estimated as 0.96 (1.37 for the threshold 1.8 and 1.96 for the threshold 2.2)
- A Jeffrey prior is used for the prior estimate.

In practice, this study uses Python package `BayesPowerlaw` with a Metropolis Hastings algorithm, as explained earlier, to estimate parameters. The Python package uses HRTD parameter range values between 1.1 to 6. The parameter range for the corresponding threshold values are listed in

Table 4.2. This research found that if default values are used, then the estimated parameter value fails to capture the most extreme event.

Table 4.2. HRTD parameter (α) and its range range estimate through Hill and SmooHill estimator.

Threshold, km/hr	Methods		Parameter range
	Hill estimator	SmooHill estimator	
1.4	3.524 (3.519, 3.529)	3.323 (3.318, 3.327)	[3.318, 3.529]
1.8	2.814 (2.811, 2.818)	3.221 (3.217, 3.225)	[2.811, 3.225]
2.2	2.628 (2.624, 2.632)	2.932 (2.928, 2.936)	[2.624, 2.936]

The HRTD parameter is estimated for three different threshold values using Bayesian inference.

Among all 10000 posterior values, the parameter values which are convergent in the MCMC (Metropolis-Hasting) algorithm are considered in the final posterior distribution. The mean and standard errors are computed and presented in Table 4.3. The mean of all accepted posterior values is used for the HRTD parameter estimate. The posterior distribution for the HRTD fit is placed in Figure 4.10. In the case of threshold 1.8, the standard error is a little high (as the parameter range is high), but all three are reasonable.

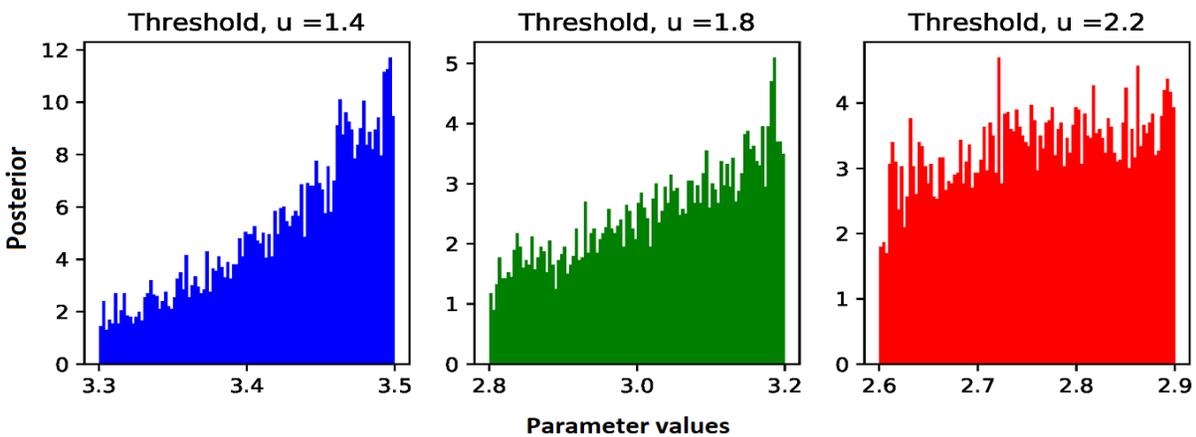


Figure 4.10. The posterior distribution for the HRTD fit for three different threshold values. The threshold values are leveled on top of each plot.

The fitted model (for different threshold values) directly compare with data CDF in Figure 4.11 as a part of model validation.

Table 4.3. HRTD parameter (α) and Standard error (Bayesian inference).

Threshold, u	HRTD parameter	Standard error
1.4	3.429	0.0005
1.8	3.028	0.0011
2.2	2.843	0.0005

The study findings demonstrate that the Bayesian inference estimate provides a superior fit compared to the Hill and SmooHill estimators across all three different threshold values. This conclusion is based on the comparison of cumulative distribution functions (CDFs) and parameter error estimates. Notably, Figure 4.11 showcases the remarkable fit achieved by the Bayesian inference method, supporting its effectiveness in accurately capturing the extreme characteristics of the data. A higher threshold value leads the more extreme values to fit the distribution line.

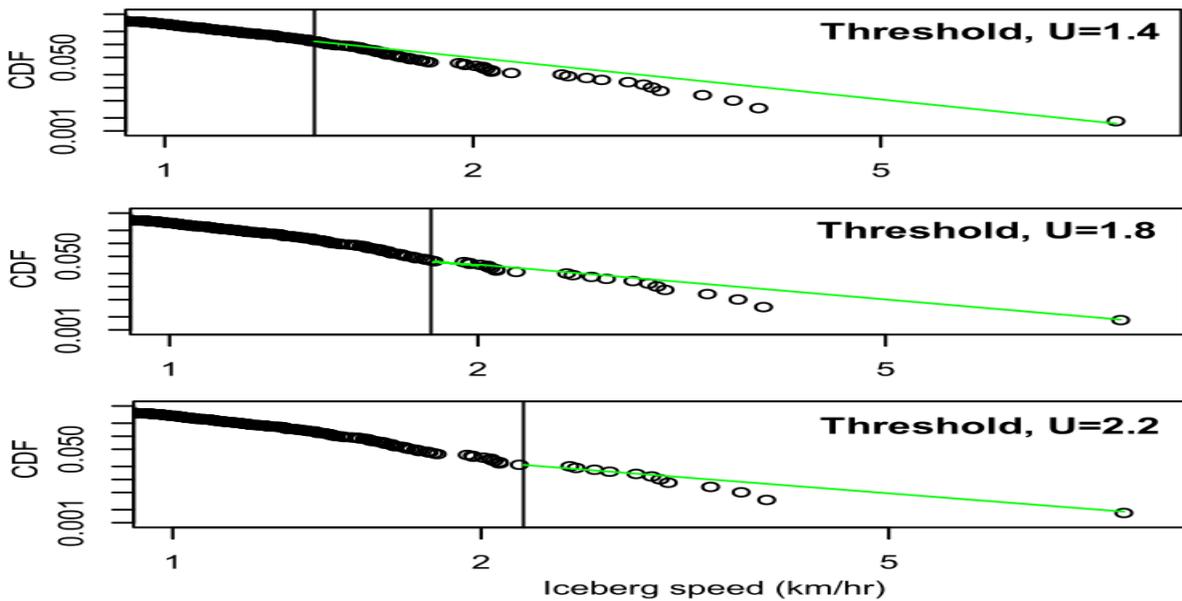


Figure 4.11. Bayesian inference: The cumulative distribution function fit comparison for three different thresholds. The threshold values are labelled on top of each plot.

4.3.2.3 Return level and Risk estimate

In this section, model return levels, i.e. the extreme iceberg speed (by the Equation 4.4) and corresponding KE (by Equation 4.10) are estimated for three different threshold values, and are

Table 4.4 Iceberg exceedance occurrence probability, extreme speed (km/hr), Kinetic Energy (KE) in Megajoules (MJ) and Risk (MJ/year) corresponding to the three different thresholds (km/hr).

Threshold, (u, km/hr)	Risk factor	Return periods				
		5 years	20 years	50	100 years	1000 years
1.4	Exceedance Probability	1.603E-03	1.511E-04	3.165E-05	9.680E-06	1.887E-07
	Speed, km/hr	4.217E+00	6.318E+00	8.254E+00	1.010E+01	1.977E+01
	KE, MJ	4.910E+03	1.080E+04	1.830E+04	2.720E+04	1.020E+05
	Risk MJ/year	7.869E+00	1.638E+00	5.797E-01	2.637E-01	1.925E-02
1.8	Exceedance Probability	1.728E-03	1.707E-04	3.702E-05	1.148E-05	2.483E-07
	Speed, km/hr	3.769E+00	5.957E+00	8.063E+00	1.013E+01	2.168E+01
	KE, MJ	3.780E+03	8.450E+03	1.440E+04	2.150E+04	8.210E+04
	Risk, MJ/year	6.539E+00	1.442E+00	5.323E-01	2.469E-01	2.040E-02
2.2	Exceedance Probability	1.745E-03	1.779E-04	3.920E-05	1.252E-05	2.814E-07
	Speed, km/hr	3.459E+00	5.632E+00	7.775E+00	9.924E+00	2.230E+01
	KE, MJ	2.260E+03	5.390E+03	9.600E+03	1.490E+04	6.370E+04
	Risk, MJ/year	3.947E+00	9.593E-01	3.761E-01	1.860E-01	1.792E-02

presented in Table 4.4. According to Table 4.4, in the case of threshold 1.4, for 100 years, the return level is 10.10 km/hr (KE 27200 MJ), which means that once every 100 years, the extreme iceberg speed will be 10.10 km/hr. For the case of thresholds 1.8 and 2.2, it is 10.13 km/hr (21500 MJ) and 9.92 km/hr (14900 MJ) respectively. The occurrence probabilities are computed by Equation 4.9 for different return periods with corresponding threshold values and presented in Table 4.4. In the case of threshold 1.4, iceberg speed is 4.217 km/hr once every five years (Table 4.4), so the probability of iceberg speed being less than or equal to 4.217 km/hr is 0.9318 (Equation 4.2). Therefore, the probabilities for iceberg speed to be more than 4.217 km/hr is 0.0682. This is a conditional probability; for the unconditional probability, 0.2094 is multiplied with $p_u = 0.1176$. The occurrence probability (once every 5 years) of an iceberg speed more than 4.217 km/hr is 0.0016 (Table 4.4). A similar procedure has been used for the case of thresholds 1.8 and 2.2. Finally, the risk is estimated using Equation 4.8, and is presented in Table 4.4, and based on acceptable risk criteria, a reasonable design risk is proposed. Data scarcity for extreme conditions is one of the reasons for uncertainty in risk calculation. At threshold 1.4, there are 70 data points used; however, in the case of 2.2, it is only the case of 12 data points. This causes significant variation in the extreme iceberg speed prediction for the three cases, and thus the risk estimate.

Most engineering offshore facilities have been designed to withstand the impact between the small and medium iceberg weights. For example, the Terra Nova FPSO has been designed to withstand the impact of a 0.1 MT iceberg (Husky energy, 2000) and the Hibernia oil drilling platform can withstand a direct collision with a 1 MT iceberg (Kennedy, 2014). The FPSO can move off the station if an iceberg of greater size threatens the platform; however, the other offshore facility like Hibernia faces challenges. According to the descriptive statistics (Figure 4.3), every year a significant number of large icebergs are sighted in the study area (14% of the total number of

icebergs listed for the period 2002 to 2017) and moving with 10 MT impacts (only its weight (Toshkova et al., 2020)) in normal weather conditions. Moreover, according to the observational data, during the study period the maximum individual iceberg speed is listed as 8.48 km/hr (Figure 4.5), which is also nine times higher than the average iceberg speed (0.91 km/hr) for the period 2002-2017. The computed risk based on average iceberg speed (0.91 km/hr) has ~70 times less compared to extreme individual iceberg risk for the period of 100 years (389.8 MJ vs 27200 MJ). Therefore, extreme iceberg speed is more significant compared to the average iceberg speed. In a harsh environment, if a large iceberg collides with a structure in Jeanne d'Arc basin, the estimated iceberg impact energy is good enough to lead any catastrophic scenario. This gives a clear justification for the primary goal of this research. The design risk considered as $3.761E-01$ MJ/year, and the corresponding iceberg speed is proposed for the design purpose, which is estimated as 7.78 km/hr (exceed occurrence probability $3.920E-05$).

4.4. Conclusions

Capturing the extreme characteristics of natural hazards which are the direct consequences of climate change, and their risk estimations is a challenging and time demanding agenda. To address this, a tail-based Bayesian approach is proposed and implemented in an extreme event scenario (iceberg collision). The methodology is new; however, model is well known. The extreme computed risk (based on iceberg size and speed) is an outcome of climate change. The Hill and SmooHill estimators are used to estimate parameter range to start the Bayesian process, which allows a robust parameter to estimate to capture the most extremeness, a reasonable fit, and a return level estimate. In all three threshold cases, especially in the case of the smallest sample, the proposed methodology shows a consistent fit to capture the extremeness. The 1000-year return period helped to achieve more extreme projection. The estimated design risk (based on extreme

individual iceberg speed) shows a significant impact compared to the average iceberg speed (the impact due to average iceberg speed is only about 4% of the impact of extreme iceberg speed). Therefore, the calculated risk needs to be considered in the iceberg risk management plan for the Jeanne d'Arc basin to avoid any unwanted situation. As a next step, a non-stationary approach for iceberg risk analysis for the different natural hazards (for example, wind speed, wave height, etc.) will be considered.

5. A GENERALIZED FRAMEWORK FOR RISK-BASED EXTREME LOAD ANALYSIS IN OFFSHORE SYSTEM DESIGN

Preface

This manuscript is based on a previously published version in the journal of Offshore Mechanics and Arctic Engineering, and I am the primary author of this research paper. Working under the supervision of co-authors, the study focuses on jointly investigating two crucial environmental parameters: wind speed and wave height. As part of the multivariate methodology development, a bivariate extreme joint model was developed and implemented to compute risk assessments for extreme wind and wave height at the Flemish Pass basin. Throughout the research process, I conducted an extensive literature review, collected, and analyzed the necessary data, and developed the methodology for the joint modeling. The analysis and modeling tasks were also carried out by me. The co-authors made significant contributions to the project, including concept development, methodology writing, manuscript review, and editing.

The manuscript underwent modifications based on the valuable feedback received from the co-authors and the peer-review process, resulting in further improvements and refinements. Co-author Faisal Khan was responsible for project management and funding acquisition for this research.

This collaborative effort has enabled us to present a comprehensive study on joint modeling of extreme wind and wave height at the Flemish Pass basin, providing valuable insights for offshore engineering in consideration of varying environmental conditions.

Abstract

The primary aim of this research is to consider the correlation among environmental factors in calculating 100 and 1000 years of extreme load design criteria. This is done by considering load

as energy transferred from external environment to the offshore system. Also, incorporating spatial and temporal dependence of environmental variables in the context of offshore design. A bivariate extreme value distribution and a conditional joint return level function are developed using the Gumbel- Hougaard copula. The offshore design risk criteria are developed for the finer grid locations ($0.1^{\circ} \times 0.1^{\circ}$ latitude/longitude grid) considering joint extreme wind and wave energy. The developed approach is tested using data for the Flemish Pass basin off the east coast of Canada. Along with the primary aim, the impact of climate change is investigated (time and space variability) by implementing the proposed methodology in two cases: the periods from 1959 to 1988 and 1989 to 2018. This study observed that climate change has caused 30% less correlation between wind speed and wave height in recent years [1989-2018] compared to the period of 1959 to 1988. The proposed extreme design wind speed is 39.7 m/s, significant wave height is 16.4 m; their joint exceeding probability is $5.80E-05$ over an annual basis for a scenario of 100- year.

Keywords Extreme energy; Copula function; Tail dependence; Max-stable process; Risk-based design.

5.1. Introduction

Engineering design needs to focus on extreme environmental conditions. Extreme loads due to earthquakes, tide, wind, waves etc., often lead to the failure of engineering systems. In addition, offshore engineering systems/equipment operates in an environment with strong currents, uneven surfaces, extremely low temperatures, and staggering depths. All this cause environmental extreme load analysis to be more complicated in the offshore domain. Furthermore, the financial impact of downtime and repair work is high. Therefore, proposing adequate or optimal structural design parameters value is challenging, and climate change issues make this task more complicated. The optimal design parameter estimation process is a case of compromising the safety and cost. It

generates the "flexibility" concept adopted in modern engineering designs in an uncertain environment (Neufville and Scholtes, 2011). The offshore operator must rely on a flexible risk-based design approach to optimize cost and achieve desired process safety by adopting the events' extreme probability and impact. Towards this, a risk-based flexible design approach is implemented for a univariate case (Arif et al., 2020a; Arif et al., 2020b). This could be compared to the offshore Accidental Limit State (ALS) design requirement (Paik and Thayamballi, 2007). However, environmental variables are correlated; for example, floating offshore structures are exposed to wind, wave height, spectral peak periods etc. Therefore, in offshore structural design, to make the impact analysis more realistic, the extreme joint conditions and their association needs to adopt (RP2A-WSD. 2000).

Wind and waves are prime factors for offshore engineering design. Being correlated, their joint association should be considered for relevant offshore engineering design and operations. However, the effect of changing climatic conditions on the correlation between extreme wind speed and significant wave height is not well understood. In 2018, global sea level was 3.2 inches (81 mm) above the 1993 average, with the highest annual average in the satellite record from 1993-present (NOAA, 2019). Additionally, the recent European heat wave (Climate signal beta, European heat wave, 2019) and the past 30 years of changing patterns of ocean waves and wind (Young and Ribal, 2019b) cannot be ignored. There are clear indications of global warming (NOAA, 2019; Climate signal beta, European heat wave, 2019, and Ogunbode et al., 2020 etc.) and its direct consequences include Greenland's loss of 532 billion tonnes of ice in a record melt in 2019 (CBC report, 2020). Between 2030 and 2052, global temperature is expected to rise by 1.5°C, compared to the pre-industry era, if it continues to rise at the current rate (Ogunbode et al., 2020). This being the case, then what will the future climate looks like? There is no exact answer;

however, the future climate will likely be dominated by more extreme natural events (Wang et al., 2017). Therefore, changing climate conditions need to be considered in risk assessment approaches to manage unexpected extreme impacts.

Flexibility based risk assessment approaches have been applied in many areas, including iceberg collision risk analysis (Arif et al., 2020a; Arif et al., 2020b), extremely low temperatures for vessels' operations (Sulistiyono et al., 2015), design of a water supply system (Zhang et al., 2011) and water management system (Deng et al., 2013; Manocha and Babovic, 2015); the met ocean study of an offshore Newfoundland and the Labrador basin (C-CORE, 2015), iceberg impact analysis (HUSKY oil operations limited as operator, 2000.) and the working stress design guidelines (RP2A-WSD, 2000). However, many of these studies were done for a large spatial scale, and unable to provide useful information for the design of a structure in a small scale. For example, the risk profile generated in (Arif et al., 2020a) failed to provide information on any specific grid location in the Flemish Pass basin area. In addition, most of these analyses were done for a single variable (Arif et al., 2020a; Arif et al., 2020b; Sulistiyono et al., 2015; Zhang et al., 2011; Deng et al., 2013; Manocha and Babovic, 2018), or in case of multivariable, without considering the correlation (C-CORE, 2015; Lombardo and Ayyub; 2014). Specifically, for the Flemish Pass basin, the 100-year annual extreme wind speed and significant wave height are predicted as 33.4 m/s and 16.3 m respectively (C-CORE, 2015); however wind and wave dependency are ignored. In the context of offshore oil drilling operations/offshore structure design, a small scale, for example, $(0.1^0 \times 0.1^0)$ latitude/longitude grid and multivariable dependency are more appropriate to capture the site specific information. Several approaches have been proposed to estimate the extreme joint distribution by considering short term and long-term dependency, in the context of offshore design and marine operations (Johannessen et al., 2002;

Bay et al., 2013; Elzbieta, 2015; Horn et al., 2018; Lin and Dong, 2019). Moreover, the seasonal effect and short-term variability are considered in building the joint met ocean model (Vanem,, 2016; Lin et al., 2020) and environmental contours are used in the structural analysis of marine and coastal structures (Haver and Winterstein, 2009; Ross et al., 2020). The proposed approach is different and captures a unique condition of extreme joint maxima of variables. The joint occurrence of maxima has a lower probability of occurrence; however, it is the extreme impact condition. The authors have used the joint occurrence of maxima as the extreme condition and then modelled different load occurrences along with their likelihood of occurrence. The load value and its occurrence likelihood are used to estimate risk. Then the risk is used as a guiding parameter. The highest risk (based on pre-defined acceptable risk level) is recommended as the design criterion. In this study, load is calculated as energy, and an extreme joint distribution of wind speed and wave height is created using the extreme value copula technique. The model presents an extremely high-resolution flexible risk profile for the Flemish Pass basin. The developed model has been used to study the impact of climate change on the dependency between wind speed and wave height over the past 60 years. The remainder of the article is organized as follows: section 5.2 describes the proposed methodology in detail; section 5.3 presents the case study, methodology validation and key findings. The concluding remarks of this study are placed in section 5.4.

5.2. The Methodology

The probabilistic approach, with conditional dependence and evidence-based updating, is considered to capture the extreme joint scenario and associated uncertainty. The risk depends on the severity of the consequences and their likelihood. Therefore, to model an extreme dependency structure, the authors focused on developing an extreme joint distribution (to calculate the joint occurrence probability) and an efficient way to calculate joint extreme return levels for impact

(energy) estimation. Towards construct a joint distribution (bivariate/multivariate cases), the models have two components: the marginal distributions and the dependence structure. Univariate extreme value-based approaches are appropriate for choosing a marginal distribution (Arif et al, 2020a; Arif et al., 2020b; Vanem 2015; Staid, 2015; Das et al., 2016; Hu and Ayyub, 2017). However, capturing the extreme dependency structure is complicated. To do this, the classical approaches, for example, implementing the "Extreme Value Theory" in bivariate/multivariate cases are common (Tawn 1988; Shiau 2003; Goodarzi et al., 2012). However, this does not allow flexibility in choosing a marginal distribution. For example, two characteristic variables for a flood event, flood volume and flood peak, have been modelled, and the marginal distribution is chosen as a Gumbel distribution (Shiau, 2003; Goodarzi, 2012). In reality, all considered variables do not always follow the same distribution. For instance, in a rare event (for the case of few data), the Gumbel distribution fails to capture the most extreme circumstances (Vanem, 2015; Gaidai, et al., 2020; Toshkova et al., 2020; Vanem 2020). As an alternative, copula-based approaches are taking the lead. The flexibility in marginal distribution choice has made the copula approach a widely used tool for modelling extreme random variables' dependency. The extreme dependence may be a case of asymptotic dependence (case of max stable models) or asymptotic independence (the case for inverted max-stable models) (Coles, et al., 1999; Rodriguez et al., 2007). When variables are asymptotically dependent, they tend to have big values at the same time. Large values never appear together if they are asymptotically independent. Different copulas have been proposed for tail dependencies and implemented in bivariate/multivariate cases (Ross et al., 2020; Candela and Aronica, 2017; Zhang et al., 2018; Kang, 2019; fang et al., 2020; Liu et al., 2020; Hu and Ayyub, 2019). In a similar aspect, the present study takes advantage of the copula-based method to model the joint distribution of wind speed and wave height.

Return level-based modelling is the most popular method for effect estimation. The return levels for extreme events such as floods and storms are standard criteria employed in offshore engineering design. They describe the severity and likelihood of an extreme event and use it to quantify the risk. For example, in a specific area, a 100-year extreme wind speed of 35 m/s means that the 100-year extreme wind speed is expected to exceed 35 m/s average once every 100 years. The return period is defined clearly in a univariate context. Its definition is more challenging when the problem at hand requires consideration of the dependence between two or more variables in a bivariate/multivariate framework. Several ways of defining a multivariate return period have been proposed in the literature, all of which rely on different probability concepts. These definitions may use conditional probability, joint probability, or can be based on Kendall's distribution or survival function (Toshkova et al., 2020). Here, the conditional return period is considered to model extreme dependency. The time scale considered in this current study is years, and extreme yearly paired (wind and wave) data is generated from the available hourly data for a specific year. The proposed methodology is shown in Figure 5.1. Each component of the methodology is described in the following sections.

5.2.1 Dependency measure

To evaluate the aggregate extreme event risk for offshore environmental risk analysis, a risk analyst needs to consider whether the individual risk factors are dependent or independent. Most variables associated with natural hazards, for example, icebergs, floods, wind, and waves are correlated to some extent. Therefore, instead of calculating the sum of random variables' distribution function, their dependence structure needs to be considered to build the joint distributions. Dependency structure measured by Pearson's correlations coefficients is widely used because of its simplicity; however, it can only represent linear dependency. Therefore, the rank

based correlation coefficient, Kendall's tau (Kendall, 1938) and Spearman's rho (Zar, 2005) is widely used in calibrating copula. The copula function depends on the rank correlations, not on marginal distributions, and is invariant in this case under strictly increasing transformations. However, Pearson's correlations do not satisfy this property and require the existence of second-order moments (Joe 2014). The Kendall's tau dependency measure concept is more robust than

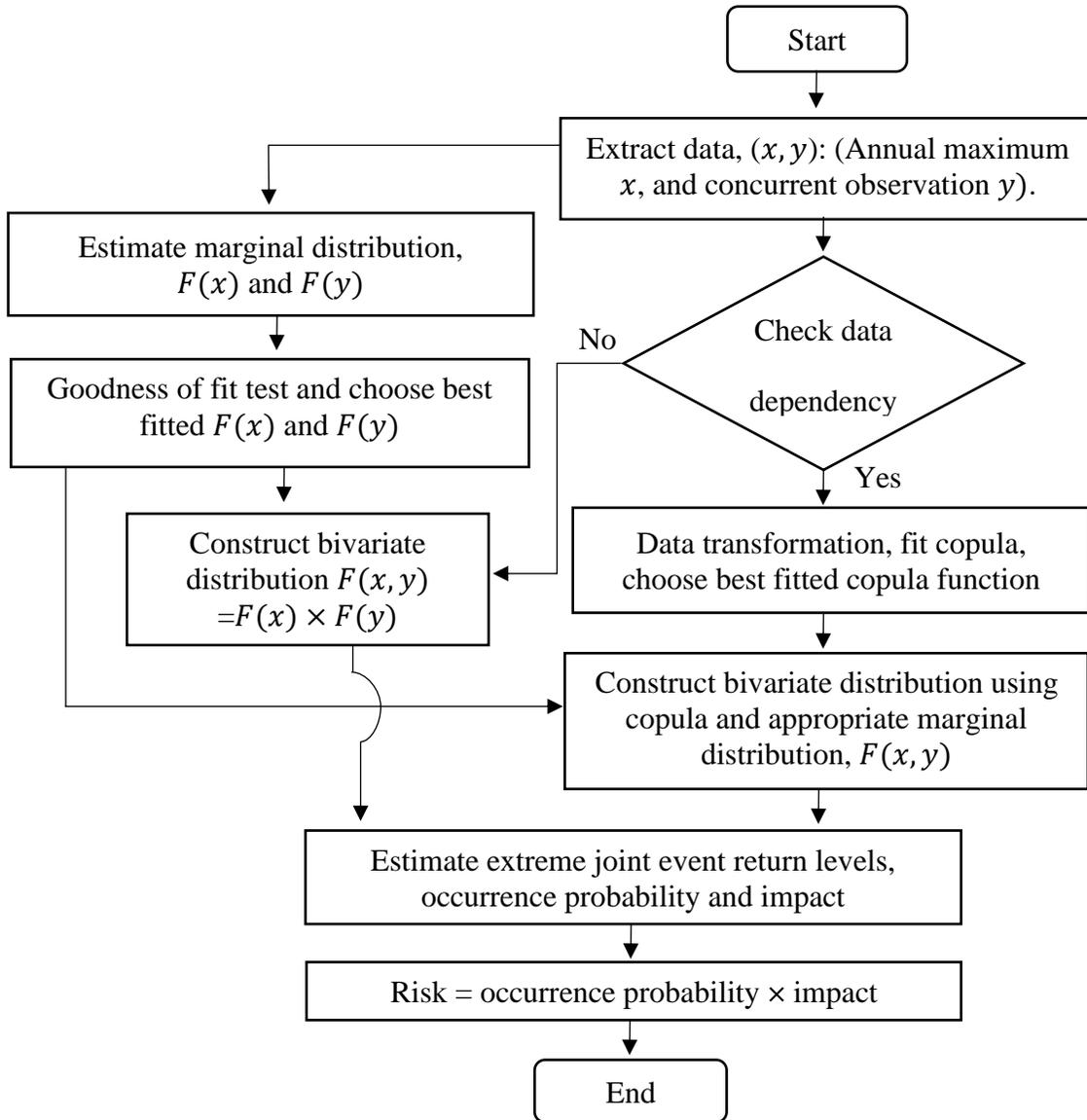


Figure 5.1. The proposed methodology for bivariate extreme event risk analysis

Pearson's correlations coefficients; it is also useful for calibrating copula (Haugh, 2016). Kendall's tau is often a good way to characterize the extreme dependence structure of a particular copula.

For example, the coefficient of tail dependence for some copulas can be expressed as a function of Kendall's tau for the copula. This study considered Kendall's tau to measure the data/model dependency. Let (X_1, Y_1) and (X_2, Y_2) be an independent random pair with continuous marginal distributions, and copula C. The two pairs are concordant if $(X_1 - X_2)(Y_1 - Y_2) > 0$ and discordant if $(X_1 - X_2)(Y_1 - Y_2) < 0$. The empirical Kendall's tau is defined as (Kendall, 1938):

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}, \quad (5.1)$$

where n_c and n_d are the numbers of concordant and discordant pairs. The total amount of pairing is $\frac{1}{2}n(n-1)$, where n is the sample size.

5.2.2 Marginal distributions

In univariate cases, the "Extreme Value Theory (EVT)" based approach is frequently used in modeling extreme events. Among all the methods, the Block Maxima (BM) based Generalized Extreme Value (GEV) and Peak Over Threshold (POT) based Generalized Pareto Distribution (GPD) are the most common. The authors implemented these methods in extreme offshore risk analysis in a univariate case (Arif et al., 2020a; Arif et al, 2020b). In this research, BM based GEV is considered because GEV is a particular case of max stable distribution. It is also validated through the chosen goodness of fit criteria (for example, probability, Q-Q, density, and return level plot). The GEV distribution is defined as (Castillo et al., 2005):

$$F(z, ; \mu, \sigma, k) = \exp \left\{ - \left[1 + k \left(\frac{z - \mu}{\sigma} \right) \right]^{\frac{1}{k}} \right\}, \quad (5.2)$$

where μ is the location parameter, σ is the scale and k is the shape parameter ($-\infty < \mu < \infty, \sigma > 0$, and $-\infty < k < \infty$). The GEV is known as combinations of three extreme value distributions

based on its shape parameter values; i.e. if $k = 0$, GEV has Gumbel, if $k > 0$, it has Frechet and if $k < 0$, it has Weibull distribution.

5.2.3 The Copula Model

A copula is a function that couples different marginal distributions. Mathematically, a copula is defined as a distribution function, $C: [0,1]^n \rightarrow [0,1]$, for integer $n \geq 2$ with marginal distributions. Sklar's theorem (Sklar, 1959) clarifies the concept of the copula in terms of its importance and prospects for implementation. According to Sklar's theorem, a distribution function $F_{1:n}: R^n \rightarrow [0,1]$ having univariate marginal distribution F_1, F_2, \dots, F_n , $n \geq 2$, can be written as $F_{1:n}(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$, where C is a copula, and it is unique when all F_1, F_2, \dots, F_n are continuous. The converse is also true, i.e. if C is a copula and F_1, F_2, \dots, F_n are distribution functions, then the function $F_{1:n}$ is a multivariate distribution function with marginal distributions F_1, F_2, \dots, F_n . Therefore, multivariate distributions can be modelled using copulas and the well known univariate methods for modeling marginal distributions. However, multivariate extreme-value analysis is concerned with the extremes in a multivariate random sample; that is, at least some components have exceptionally large values, and mathematical theory suggests the use of max-stable models for univariate and multivariate extremes (Segers, 2012). Therefore, the univariate extreme modelling approach (Arif et al, 2020a; Arif et al., 2020b) can be merged with a copula if it is an extreme value copula, and max-stable. In addition, a copula is an extreme value copula if and only if it is max-stable (Theorem 2.1, Gudendorf et al., 2010). Therefore, a max-stable copula is also an extreme-value copula, being in its own domain of attraction (GEV families are explained in section 5.2). Conversely, each extreme-value copula can be shown to be max-stable. For example, the Gumbel–Hougaard (GH) copula (Hutchinson and Lai, 1990) is an extreme

value copula (Balakrishnan, 2009) and therefore is max stable (Theorem 2.1, Gudendorf et al., 2010). The Gumbel Hougaard (GH) copula is defined as (Hutchinson, 1990):

$$C(u, v) = \exp \{ - [(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}} \}; \theta \geq 1 \quad (5.3)$$

This study focuses on the GH copula to build extreme bivariate models. However, Clayton and Joe copulas are considered to determine the best model. The Clayton copula is defined as (Clayton, 1978):

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}; 0 < \theta < \infty. \quad (5.4)$$

The Joe copula is defined as (Hosking et al., 1985):

$$C(u, v) = 1 - [(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta]^{\frac{1}{\theta}}; 1 \leq \theta < 1 \quad (5.5)$$

5.2.4 Goodness of fit test

In this research, GEV fit is justified with a model versus empirical probability, Q-Q, density, and return level plot. With the Q-Q and the P-P plots, the feasibility of capturing tail data is examined. On the Q-Q and the P-P plots, most of the points should be located close to the diagonal. Significant variations from linearity would mean poor fit of the chosen model or inaccuracy in the parameter estimation technique. For more details refer to (Arif et al., 2020a). For the copula, the goodness of fit is investigated with the likelihood values (log-lik), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are most common. A larger Log-lik and smaller AIC and BIC give a better fit. However, in this case, all considered copulas' have one parameter, so AIC or BIC statistics do not provide any additional information. In this research, best fitted copulas are justified with log-lik values, P-values and testing null hypothesis ($H_0: C = C_\theta$ against $H_1: C \neq C_\theta$ where C_θ is a specific copula (Genest, 2009). Smaller P-values and higher test statistics provide a better fit. Along with the numerical evidence, the fitness of a copula is justified with the correlation map plot (empirical vs. model). More precisely, correlation map plot differences (Model -

Empirical) less than or equal to ~ 0.1 are considered benchmarks to justify the captured dependency. Finally, the predicted 100-year return levels (wind speed and wave height) are compared with the result in (C-CORE, 2015). According to the (C-CORE, 2015) climate study analysis, 100-year extreme wind speed listed for the Flemish Pass basin is 33.4 m/s and wave height is 16.3 m (Hindcast data; the 60 years from 1954 to 2013). The present study considered Hindcast data from 1959 to 2018 in two phases. This direct comparison justifies use of the proposed methodology and verifies its prediction.

5.2.5 Extreme Bivariate Distribution

Let x (wind) and y (wave) be two random variables. Assume data are Independent and Identically Distributed (IID) and that each variable contains the yearly data maximum value. Then the joint cumulative distribution $F(x, y)$ of any pair of continuous random variables (x, y) with marginal distributions $F(x)$ and $F(y)$ is given by

$$F(x, y) = C(F(x), F(y); \theta) = C(u, v; \theta), \quad (5.6)$$

where C represents the copula function, and the dependence structure is defined by the parameter(s), θ . The variables u and v are the normalised ranks between 0 and 1 (transformed from the variables x and y). The $F(x)$ and $F(y)$ are chosen as GEV distribution (as explained in section 5.2). Therefore:

$$F(x) = \exp \left[- \left\{ 1 + k_1 \left(\frac{x - \mu_1}{\sigma_1} \right) \right\}^{-\frac{1}{k_1}} \right] \quad (5.7)$$

$$F(y) = \exp \left[- \left\{ 1 + k_2 \left(\frac{y - \mu_2}{\sigma_2} \right) \right\}^{-\frac{1}{k_2}} \right] \quad (5.8)$$

The Extreme Bivariate Distribution is based on the Gumbel–Hougaard copula (Equation 5.3):

$$F(x, y) = C(u, v) = \exp - \left[\left\{ 1 + k_1 \left(\frac{x - \mu_1}{\sigma_1} \right) \right\}^{-\frac{\theta}{k_1}} + \left\{ 1 + k_2 \left(\frac{y - \mu_2}{\sigma_2} \right) \right\}^{-\frac{\theta}{k_2}} \right]^{1/\theta} \quad (5.9)$$

Similarly, the extreme bivariate distribution can be derived based on the Clayton copula (Equation 5.4) and the Joe copula (Equation 5.5).

5.2.6 Parameter estimation

The likelihood function of independent observations x_1, x_2, \dots, x_n is defined as:

$$L(x_i; \theta) = \prod_{i=1}^n f(x_i; \theta), \quad (5.10)$$

where $f = \frac{dF}{dx}$, and θ represent the distribution parameters. According to (Hosking, 1985), for a small size sample, the MLE parameter estimator is unstable (due to the large bias and RMSE in extreme upper quantile estimators) and the probability weighted moments estimator is recommended (Greenwood, 1979), which is equivalent to the L-moment estimator (Hosking, 1990). This study found that the MLE and the Generalized Maximum Likelihood Estimator (GMLE) are unstable to estimate GEV parameters for the case of few grid location data. In some cases, the GMLE estimator (Martins and Stedinger, 2000) gives negative values for the scale parameter and causes a biased estimate in the shape parameter. The marginal GEV parameters are estimated independently, assuming data are identical and independently distributed. In this research, the GEV parameters are estimated using the L-moment and the copula parameter θ is estimated through MLE. The L-moment estimator for GEV distribution is defined as (Hosking,

$$1990): \mu = \lambda_1 - \frac{\sigma}{k} \{1 - \Gamma(1 + k)\}, \quad \sigma = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)}, \quad k = 7.8590c + 2.955c^2, \quad \text{and } c = \frac{2}{3 + \lambda_3/\lambda_2} - \frac{\log 2}{\log 3},$$

where, $\lambda_1 = \beta_0, \lambda_2 = 2\beta_1 - \beta_0, \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$ and the estimator of β_r is defined as: $\beta_r = n^{-1} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-r)}{n(n-1)(n-2)\dots(n-r)} x_i$, where $r = 0, 1, 2$ and x_i are the ordered observations from a sample size n .

5.2.7 Consideration of Return level

The return level gives the event magnitude for a specific return period. In the univariate context, it is straightforward (Arif et al, 2020a; Arif et al, 2020b; Das et al, 2016) etc. However, the bivariate/multivariate framework needs to solve a non-linear system of equations. If only one extreme random variable is significant in the design criteria or the two random variables are independent or less dependent (dependency less than 2), the return period is defined as (Gaidai, et al, 2020): $T_X = \frac{1}{1-F(x)}$, and $T_Y = \frac{1}{1-F(y)}$, where the numerator, 1, indicates that most extreme events are considered for every year. In the case of a joint return period, the random variables x and y may be considered either jointly or conditionally. The joint return period is defined as:

$$T_{X \text{ or } Y} = \frac{1}{P(X \geq x \text{ or } Y \geq y)} = \frac{1}{1-F(x,y)}, \text{ and } T_{X \text{ and } Y} = \frac{1}{P(X \geq x \text{ and } Y \geq y)} = \frac{1}{1-F(x)-F(y)+F(x,y)}.$$

The conditional distribution Y given $X \geq x$ and the conditional distribution of X given $Y \geq y$ can be defined as (Shiau et al., 2003):

$$F(y|X \geq x) = \frac{F(y)-F(x,y)}{1-F(x)} \quad (5.11)$$

$$F(x|Y \geq y) = \frac{F(x)-F(x,y)}{1-F(y)} \quad (5.12)$$

As a result, the conditional joint return period is as by using Equation 5.11 and Equation 5.12 as follows:

$$\begin{cases} T_{X|Y \geq y} = \frac{1-F(y)}{1-F(x)-F(y)+F(x,y)} \\ T_{Y|X \geq x} = \frac{1-F(x)}{1-F(x)-F(y)+F(x,y)} \end{cases} \quad (5.13)$$

In this study, the authors are interested in the conditional return period defined in Equation 5.13. Therefore, to estimate the bivariate conditional return levels, one needs to solve the system of a nonlinear equations for x and y . The present study used iterative Newton's method to solve systems of nonlinear Equations 5.13. To do this, here the R package "nleqslv" is used. Details

about the algorithm, method and the implementation process for “nleqslv” are available at <https://www.rdocumentation.org/packages/nleqslv/versions/3.3.2/topics/nleqslv>.

5.2.8 Risk estimation

The risk is denoted as R and defined as (Arif et al, 2020a; Sulistiyono et al, 2015):

$$R = P(O) \times C, \quad (5.14)$$

where $P(O)$ is occurrence probability (extreme wind speed and significant wave height) and C is the extreme wind-wave consequence (impact). Impact and the possibility of an incident are the two components of risk (Equation 5.14). As a result, the likelihood represents the possibility of an event occurring over a return time (/yr). Therefore, the risk values are estimated by multiplying the corresponding likelihood by impact energy. In this study, the impact is considered in terms of energy. The occurrence probability is defined as (Arif et al, 2020a):

$$P(O) = (RP)^{-1} \times P_e, \quad (5.15)$$

where RP is the return period and P_e is the joint exceedance probability. The joint exceedance probability computed using the joint cumulative distribution function is defined in Equation 5.9. The probability of occurrence is calculated considering two elements, one return period (which provide scenario in the window of operation), and the second event’s occurrence probability, which maybe treated as short-term variability). For example, for a 100-year return period, the probability of the scenario is 1/100. Now, in this scenario, the event’s joint exceedance probability $P(x > 32.2 \text{ and } y > 15.2)$ is 0.0016 (Computed through Equation 5.9). Therefore, the occurrence probability per year is $0.01 * 0.0016 \sim 1.6E-05$. The extreme joint wind-wave consequence (impact), C is computed as total energy (MJ) imposed through extreme wind and wave profile predicted on section 5.7. This research, instead of measuring the outcomes of this impact, focused on how much total energy (to access load) is available to make any unexpected

situation. The total energy, C is computed by incorporating energy created by the wind (E_{wind}) and energy created by the extreme wave height (E_{wave}). In the current study, wind energy is defined as wind moment, which is defined as force times distance in the direction of force (work done, which is source of failure and thus considered as load). The design wind force, F , is defined as; $F = \frac{1}{2} \times \rho \times u^2 \times CS \times A$ (Greenwood et al., 1979); Therefore, the wind energy, E_{wind} is computed according to the following formula:

$$E_{wind} = F \times d = \frac{1}{2} \times \rho \times U^2 \times CS \times A \times d, \quad (5.16)$$

where U is proposed extreme wind speed (m/s, predicted return levels), ρ is the density of air (1.225 kg/m³), CS is the drilling derrick shape coefficient, is considered as 1.25 (Oil and GmbH, 2013), A = wind projected area and d is the distance of wind speed measure (10 m height). The energy created by the extreme wave height (E_{wave}), is calculated as (Blackledge et al., 2013):

$$E_{wave} = \frac{1}{8} \times \rho \times g \times H^2 \times A; \quad (5.17)$$

where ρ the density of the water and is g the acceleration of gravity, H is the predicted wave height (Concurrent with extreme wind), and A is the wave projected area. Present study, the area, A is considered as 3332 m² (Hibernia oil drilling platform top side length 98 m and width 34 m (HIBERNIA, 1997). Equation 5.17, provides wave energy, which is the work done on the project area by vertical wave force. The present study considered the summation of these two energy terms from a failure mode perspective. Different parts of the object will be affected by the wind and wave forces, and the impact mechanism will differ. However, in this study, the same area (A) is considered for both wind and wave energy computation. We assume that when these two types of energy are combined and given to an existing facility (or a floating object), they will have a synergistic effect and lead to failure. So, from the perspective of failure load calculation, we have

taken the sum of these two energies into joint impact consideration. Additionally, the impact ratio of these two energies is presented and employed in the analysis of the consequences.

5.2.9 Design wind speed and wave height

What is the highest risk needed to adopt for proposing a design parameter? There is no precise answer for this. For example, a 50-year scenario might give us less risk compare to adopting a 1000 years scenario. In addition, anyone can adopt more years' scenario and comes off with another highest risk value. Towards this, an acceptable risk level needs to be defined by the operator (for example, say 10^{-2} MJ/year). Defining this value is to be a subject of optimizing two key factors: safety and cost. This needs to adjust based on an operation's needs. Based on accepted risk value, the corresponding risk components (wind speed, wave height, and occurrence probability) are proposed for design purposes.

5.3. The Case Study

Wind speed and wave height are significant environmental variables with direct impacts on offshore structures and marine facilities. In general, wind speed and wave height show high dependency (time and space). The wind speed and wave height hourly data have been collected from Hindcast model-simulated data (Swail et al., 2006) for 154 grid locations from 1959 to 2018. In this research, the wind is considered the primary variable, then takes the annual maximum of this together with concurrent observations of the waves to extract all extreme data sets. The proposed methodology is used to assess the extreme risk by incorporating the joint probability of extreme wind speed, wave height and their impact energy on an offshore structure in the Flemish Pass basin (47° N to 48° N, 46° W to 47.3° W). Topside engineering facilities must take wind load effects into account while designing offshore structures. Random topside area data from an oil drilling rig are used to estimate the impact of energy. Although the projected area perpendicular

to the direction of force is preferred for wind load effects, the authors primarily focus on the amount of extreme load that would be put on any platform in an extreme situation rather than getting into specifics about platforms. More specifically, topside area data from the Hibernia oil drilling platforms (using top side length 98 m and breadth 34 m; additional information about the platform can be obtained at (HIBERNIA, 1997).

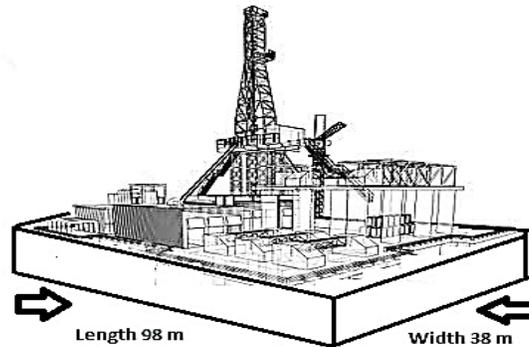


Figure 5.2. A simple topside representation of a random platform.

Figure 5.2 presents a basic topside layout such a random platform. The data set is segmented to formulate two cases. In “Case I”, data from the period 1959 to 1988 are used, and in “Case II”, data from 1989 to 2018 are considered. Splitting the time scale allows the examination of climate change’s effects on wind and wave extreme behaviour and their correlation. The R (freely available at <https://www.rstudio.com>) packages `extRemes`, `evmix`, `fExtremes`, `Copula`, `Vinecopula`, and `nleqslv` are used to implement the methodology and the descriptive statistical data analysis.

5.3.1 Descriptive statistics

According to data statistics, for Case I vs. Case II, the most extreme wind speed listed is 30.85 vs. 36.44 m/s. For wave height, it is 15.54 vs. 15.38 m. Both cases show that the wind speeds have an

Table 5.1. Data statistics: Hourly wind speed (m/s) and wave height (m) data for three grid locations (with extremes) for the Flemish Pass basin. The most extreme grid locations are highlighted as bold.

		Hourly data statistics							
	Case	Grid location	Median	Maximum	Q1	Q3	Q3+3*IQR	skewness	Kurtosis
Wind speed (m/s)	Case I (1959-1988)	(47 ⁰ N, 47.3 ⁰ W)	8.06	29.33	5.31	11.25	29.07	0.57	3.14
		(47.4 ⁰ N, 47.3 ⁰ W)	8.17	29.53	5.45	11.37	29.13	0.59	3.06
		(48⁰N, 46⁰W)	8.50	30.85	5.55	11.85	30.75	0.52	3.03
	Case II (1989-2018)	(47 ⁰ N, 47.3 ⁰ W)	8.36	31.66	5.56	11.51	29.36	0.56	3.22
		(47.4 ⁰ N, 47.3 ⁰ W)	8.45	32.53	5.7	11.61	29.34	0.58	3.24
		(48⁰N, 46⁰W)	8.84	36.44	5.92	12.18	30.96	0.52	3.14
Wave Height (m)	Case I (1959-1988)	(47 ⁰ N, 47.3 ⁰ W)	2.53	15.28	1.74	3.62	9.25	1.47	6.74
		(47.4⁰N, 47.3⁰W)	2.53	15.54	1.74	3.63	9.30	1.43	6.61
		(48 ⁰ N, 46 ⁰ W)	2.67	14.89	1.84	3.83	9.80	1.55	6.75
	Case II (1989-2018)	(47 ⁰ N, 47.3 ⁰ W)	2.64	14.51	1.85	3.72	9.33	1.38	6.59
		(47.4 ⁰ N, 47.3 ⁰ W)	2.65	14.64	1.85	3.74	9.40	1.36	6.50
		(48⁰N, 46⁰W)	2.81	15.38	1.97	3.96	9.91	1.54	6.80

increasing trend in all considered 154 grid locations; however, wave heights have the opposite direction in a few grid locations, for example at (48⁰N, 46⁰W). The median and maxima show significant differences for both variables (Table 5.1), so the grid-points data look similar to the heavy tail. In Case II, the first quantile, and third quantile values suggest that, for the wind, 25% of the values are less than 5.92 m/s and 25% of the data are higher than 12.18 m/s at (48⁰N, 46⁰W). Consequently, the Inter Quartile Range, $IQR = Q_3 - Q_1 = 12.18 - 5.92 = 6.26$, and data values greater

than 30.96 m/s (for the case of wind) 9.91 m (for the case of the waves) are considered as extreme outliers (Devore, 2011). Similar statistics are found in all grid location data. Therefore, the wind and wave data sets have extreme outliers, and their distributions deviate from the normal distribution. Besides, all skewness and kurtosis values are positive, so the data are not normally distributed, and the right tail is heavier than the left tail. Compared to wind data, the kurtosis values of the wave height data are more significant, as the wave height distributions' tail is larger than the wind speed data distribution. Table 5.1 presents information for three grid locations; however, all 154 grid locations have similar descriptive statistics. According to descriptive statistics, this seems to capture extremity of all 154 grid locations' data, so the GEV might be an option.

5.3.2 Marginal distributions (GEV fit) and Copula function

In this research, the optimal block size “one year” is chosen as a trade off between the bias and the variance. The small block sizes lead to bias, and the large block sizes lead to significant variances. The study has found that two years, five years and ten years block data leads to increased error in GEV parameters estimation. Therefore, yearly extreme wind data are extracted along with concurrent wave data for 154 grid locations. For both cases, for a specific location, 30 most extreme data are used to fit GEV distribution. As we are interested in the conditional return level, the conditional extreme criteria are chosen as 28 m/s for wind speed and 14 m for wave height. For all data sets, the normal Q-Q plot, mean residual life plot, and parameters stability plot are used to justify these threshold values. Those threshold values are also verified with historical storm behaviour. In Figure 5.3, the correlation plots between wind speed and wave height are presented for two grid locations. Figure 5.3 represents pairs of annual extreme value of wind together with concurrent value of waves. This representation is also treated as temporal extreme dependency for a specific location. When comparing Case II to Case I (Figure 5.3), it is clear that climate change

causes an unusual pattern of extreme dependency. To do a spatial causality, a data set is generated by considering annual maximum wind and concurrent wave data from all grid locations combined.

This data set is renamed as the extreme spatial-temporal data set and used for the

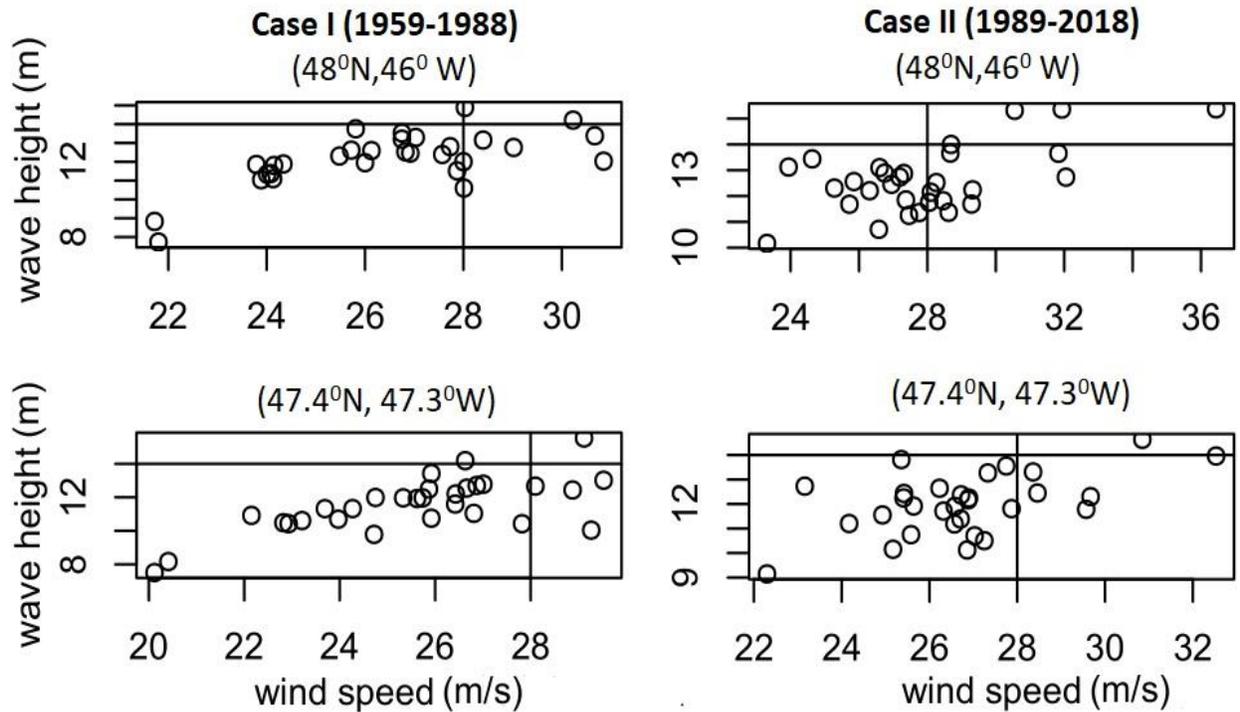


Figure 5.3. Wind and wave correlated distribution for two grid locations. The grid location is levelled on top of each plot. The horizontal and vertical black lines indicate the amount of data in the area of $x > 28 \text{ m/s}$, $y > 14 \text{ m}$.

remaining analyses (GEV fit, choice of copula, validation and risk estimation). Then all steps are repeated for all data sets to generate a small scale joint extreme event risk profile. The GEV model parameters are estimated using the L-moment method and are listed in Table 5.2. The MLE and GMLE are also used to estimate the distribution parameters. However, unstable estimates result in few grid locations data. The GMLE estimator gives negative values for scale parameters and causes bias in shape parameters for some grid locations. This issue may be resolved by imposing a restriction; however, this study considers the L-moment method for estimating GEV parameters.

Table 5.2. Distribution parameter values for Case I and Case II, considering time-space maximum data. Corresponding 95% confidence intervals for model parameters are in parenthesis.

		GEV Distribution parameters		
		Location, μ	Scale, σ	Shape, k
Case I	Wind speed	26.38 (25.34, 27.49)	2.71 (1.91, 3.40)	-0.44 (-0.76, -0.15)
	Wave height	12.0 (11.52, 12.54)	1.36 (0.94, 1.74)	-0.37 (-0.66, -0.08)
Case II	Wind speed	27.28 (26.52, 28.13)	2.02 (1.42, 2.62)	-0.007 (-0.34, 0.23)
	Wave height	12.33 (11.88, 12.82)	1.22 (0.90, 1.56)	-0.19 (-0.48, -0.06)

Table 5.3. Copula parameter value and goodness of fit test statistics for Case I and Case II.

	Copula	Parameter	Kendall's tau	Log-lik	Test-statistic	P-value
Case I (1959-1988)	Clayton	1.94	0.49	10.14	0.29	0.45
	GH	1.99	0.50	9.17	0.78	0.16
	Joe	2.72	0.48	10.44	0.57	0.19
Case II (1989-2018)	Clayton	1.15	0.37	5.38	0.02	0.75
	GH	1.59	0.37	5.23	0.33	0.31
	Joe	1.95	0.34	5.55	0.20	0.44

Based on the GEV fit criterion mentioned in Section 5.4, GEV fits are reasonable for spatial-temporal data sets. Figure 5.4 displays the Q-Q and the density plots for Case II (1989–2018). In addition to these two plots, the P-P plot and the return level graph are also used. Most of the points on the (Q-Q plot, Figure 5.4) are relatively close to the unit's diagonal, which supports how tail events were captured in this study. A similar method is used to confirm the GEV fit in all 154 extreme grid location data. This study considered three copulas for modeling dependency. The copula parameters are estimated through the MLE estimator and listed in Table 5.3. The dependence between wind speed and wave height is measured using Kendall's tau. According to the test statistics listed in Table 5.3, the GH copula shows a smaller log-lik value, smaller P-value,

and higher test statistics. The GH copula is chosen for this research because for most grid location data, GH shows better fit (for some grid locations, the Joe and GH fits were comparable).

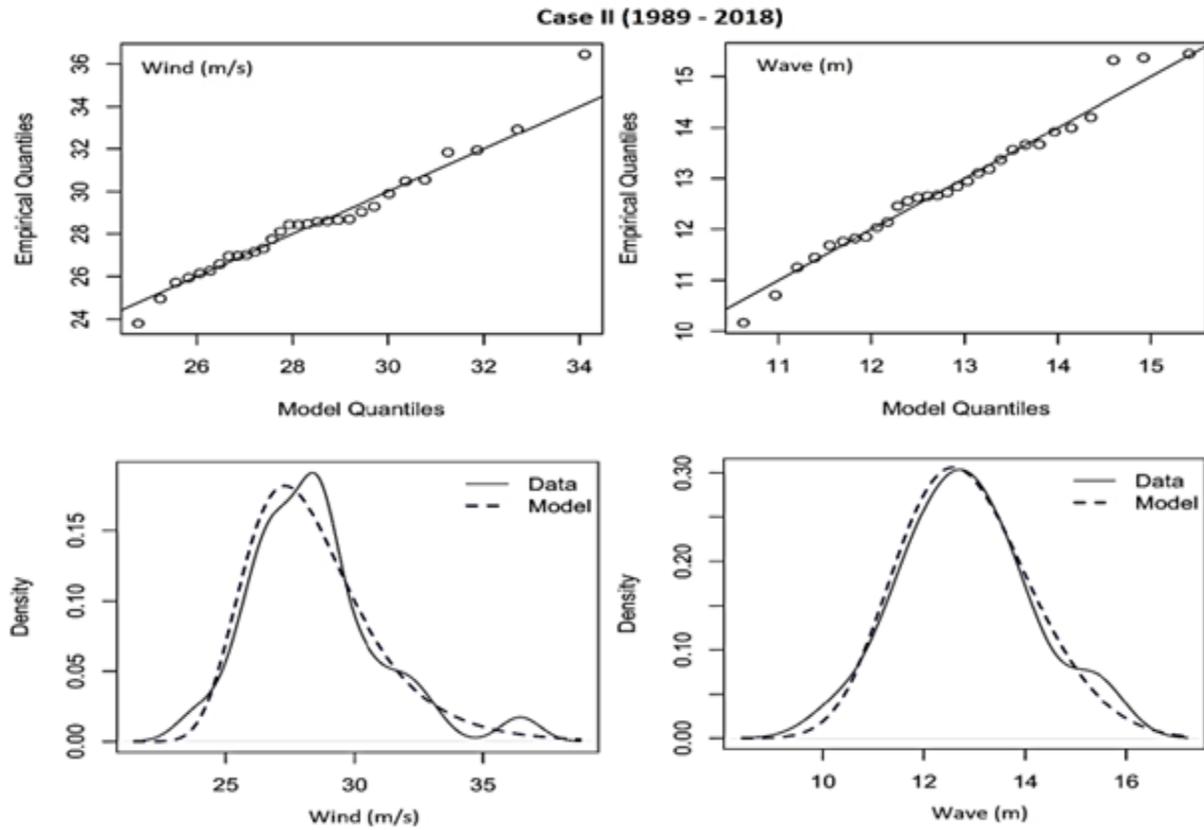


Figure 5.4. Q-Q and density plot of wind and wave extreme spatial-temporal distribution (empirical vs model).

Extreme Wind speed and significant wave height have a stronger correlation in Case I compared to Case II (Figure 5.4). It seems that currently, extreme events show less correlation (Case I vs Case II ~ 0.42 to 0.6 vs 0.18 to 0.42). The correlation is better modelled using the GH copula for Case I compared to Case II (Figure 5.4 and Figure 5.5). However, in both cases, the GH copula shows a reasonable fit (correlation difference plot: Model – Empirical, most grid location areas show ~ 0.1). Therefore, the GH copula is used to build joint extreme distribution with GEV marginal distribution.

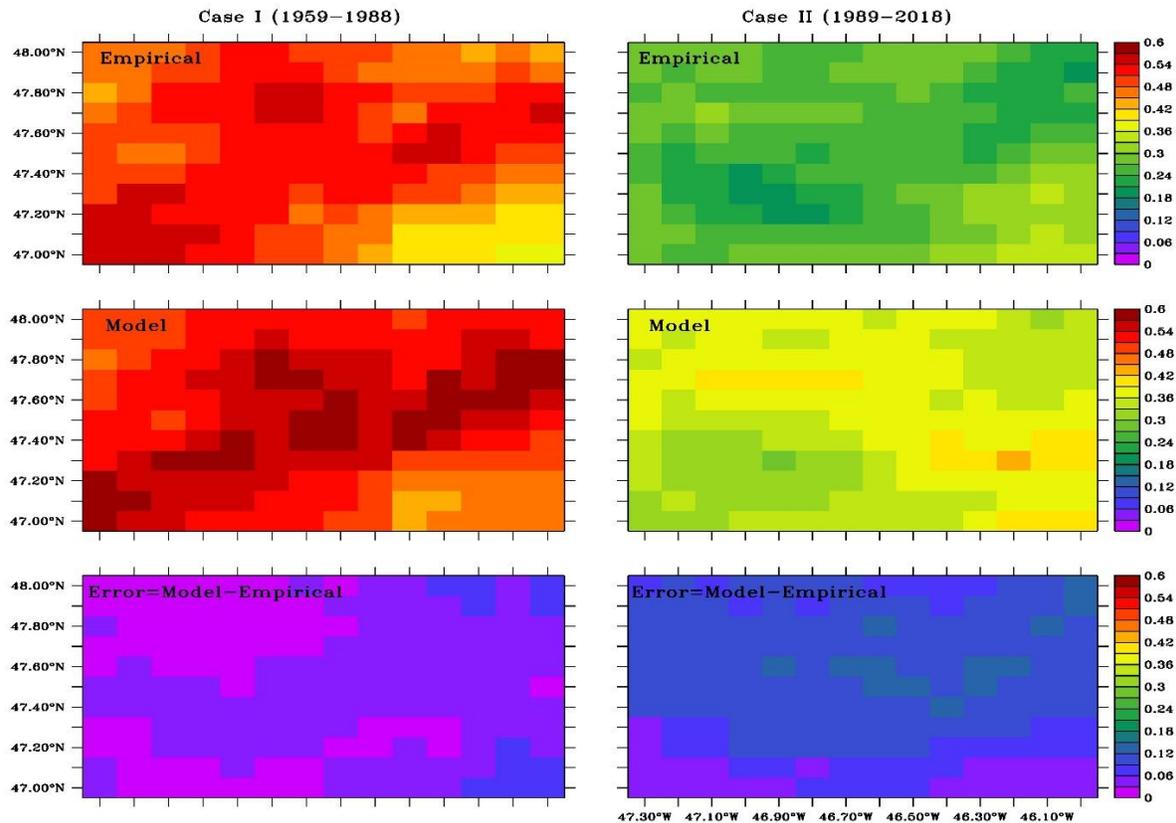


Figure 5.5. Correlation map plot (Empirical vs Model): yearly maximum wind speed vs. wave height. The bottom plot shows the error (Model-Empirical) correlation.

5.3.3 Risk estimation

Extreme joint impact energy and its occurrence probability are used to generate risk profile. All computed extreme events occurrence probabilities, extreme wind speed, significant wave height, and impact energy, for the extreme spatial-temporal data set, are placed in Table 5.4. The energy ratios in Table 5.4 show how both trams (wind and wave) dominate one another as well as the energy efficiency for various return levels. Higher return levels appear to be dominated by wave energy. The joint impact energy is computed by incorporating the wind energy calculated as defined in Equation 5.16 and wave energy as described in Equation 5.17. The conditional extreme wind speed, u , and the wave height, H , are estimated by solving the system of Equation 5.13. The

corresponding joint occurrence probability is computed from Equation 5.9. In both cases, the predicted results are comparable with (C-CORE, 2015). For example, for C-CORE (C-core, 2015) vs. the author’s predicted output (100-year, listed in Table 5.4) for Case I, wind speed is 33.4 m/s vs. 32.2 m/s, and wave height is 16.3 m vs. 15.2 m. However, in Case II, wind speed is 33.4 m/s vs. 39.7 m/s and wave height 16.3 m vs. 16.4 m. This comparison indicates that the proposed

Table 5.4. Joint exceedance occurrence probability, conditional wind speed (m/s) ($x|y>14$), conditional wave height (m) ($y | x > 28$), and Impact energy (MJ) and Risk (MJ/year).

	5 years	20 years	50 years	100 years	1000 years	
Case I	Occurrence Probability	7.36E-03	4.30E-04	6.80E-05	1.60E-05	2.00E-07
	Wind speed (m/s)	3.13E+01	3.19E+01	3.21E+01	3.22E+01	3.24E+01
	Wave height (m)	1.42E+01	1.48E+01	1.51E+01	1.52E+01	1.55E+01
	Wind impact energy (MJ)	2.50E+01	2.60E+01	2.63E+00	2.65E+00	2.68E+00
	Wave impact energy (MJ)	8.21E+02	8.94E+02	9.25E+02	9.42E+02	9.76E+02
	Impact energy ratio	3.04E-02	2.91E-02	2.84E-03	2.81E-03	2.75E-03
	Joint Impact energy (MJ)	8.46E+02	9.20E+02	9.27E+02	9.45E+02	9.79E+02
	Risk (MJ/year)	6.23E+00	3.96E-01	6.31E-02	1.51E-02	1.96E-04
	Case II	Occurrence Probability	2.59E-02	1.50E-03	2.34E-04	5.80E-05
Wind speed (m/s)		3.33E+01	3.64E+01	3.83E+01	3.97E+01	4.41E+01
Wave height (m)		1.45E+01	1.55E+01	1.61E+01	1.64E+01	1.72E+01
Wind impact energy (MJ)		2.83E+01	3.38E+01	3.74E+01	4.01E+01	4.96E+01
Wave impact energy (MJ)		8.51E+02	9.82E+02	1.05E+03	1.10E+03	1.21E+03
Impact energy ratio		3.33E-02	3.44E-02	3.56E-02	3.66E-02	4.10E-02
Joint Impact energy (MJ)		8.80E+02	1.02E+03	1.09E+03	1.14E+03	1.26E+03
Risk (MJ/year)		2.28E+01	1.52E+00	2.55E-01	6.59E-02	7.56E-04

methodology and prediction listed in Table 5.4 are acceptable. Based on acceptable engineering design criterion (annual occurrence failure probability of 10^{-5} and pre-defined risk level criteria); the 100 years' events are considered in extreme spatial-temporal data for the design purpose. The proposed extreme design wind speed and wave height are 39.7 m/s, 16.4 m respectively, and the joint event likelihood is $5.80E-05$ per year. Case II having a clear indication of climate change impact compare to Case I. More precisely, in Case II, 100 years events wind speed increased 7.5

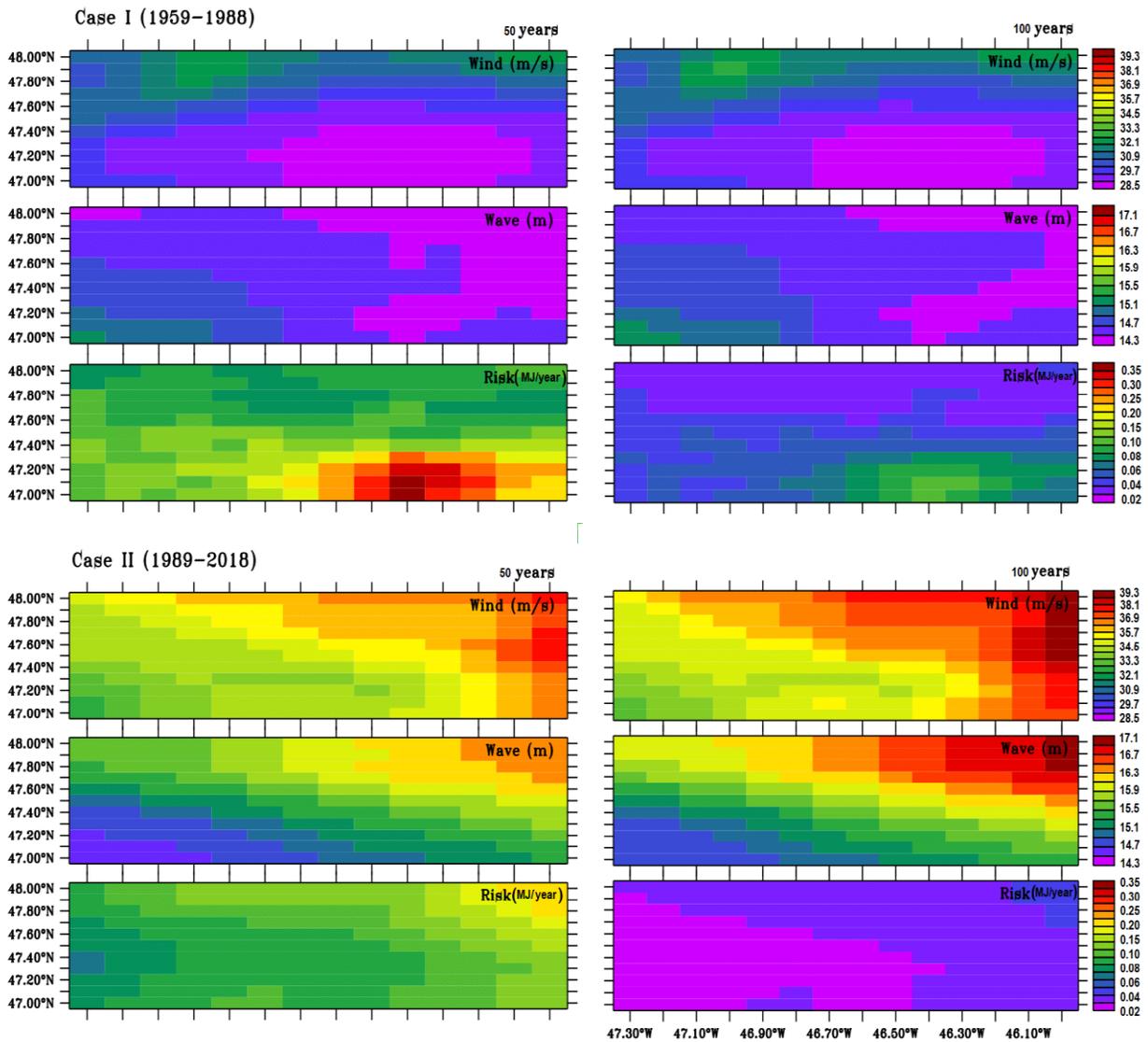


Figure: 5.6. Map plot (50 years vs. 100 years): Predicted extreme joint wind speed, wave height and its corresponding risk.

m/s and wave height 1.2 m compare to case I. Finally, a 100-years risk profile is generated for all 154 grid locations; presented using the map plot in Figure 5.6. In addition, a 50-year risk profile is added for justification of a 100-year risk profile in Figure 5.6. In Case II, the wind speed and wave height are higher in all grid locations compared to Case I (Figure 5.6). More precisely, the location (47°N, 47.3°W) is a better choice compared to (48°N, 46°W). The climate-changing pattern is visible in 5 years extreme events in Case II (Table 5.4). The present study focuses on the Case II scenario from a design perspective; the higher grid locations show more extreme events and risks. So for safer offshore operations, small grid areas might be a better choice. This significant concern needs to adapt to an offshore risk management plan for future designs upgrading policies of existing offshore operations as a part of climate change adaptation.

5.4. Conclusions

This study proposed a joint extreme event risk assessment methodology for offshore system design, which provides a bivariate extreme joint distribution and the conditional joint return level function. The Gumbel-Hougaard (GH) copula and the Generalized Extreme Value (GEV) marginal distribution are used to build the joint distribution. The GH copula and GEV are chosen as “extreme value copula” and “max stability” concepts. The copula parameters are estimated through the Maximum Likelihood Estimator (MLE), and the GEV distribution is determined using the L-moment methods. The MLE and GMLE estimators provide biased estimates in GEV parameters’ computation (for a few grid locations data). The proposed methodology is validated using the Hindcast model simulated data, and an extreme wind speed and a significant wave height are proposed for the Flemish Pass basin. The “small scale risk profile” and “climate change impact adaptation” make this research distinct. The results are validated against test statistics and direct comparison with a published report. The smaller grid locations area was found to yield less risk

compared to the higher grid locations area. In addition, climate change was found to cause ~30% less correlation between wind speed and wave height. Over the last 30 years, wind speed increased ~ 19%, and wave height increased ~8%. Therefore, to avoid any unwanted situation for design and operation, the corresponding estimated risk must be considered in the offshore risk management plan for the Flemish Pass basin. The present study has been done based on the assumption that data are independent and identically distributed. Therefore, for future studies, a non-stationary multivariate approach, or stationarity with respect to covariates such as wave and wind direction, seasonality etc. will be considered for different natural hazards such as icebergs, wind speed, wave height, and wave periods relevant to the Flemish Pass Basin.

6. **Extreme Wind Load Analysis using Non-stationary Risk-based Approach.**

Preface

This work is a part of my graduate study, and a version of this research has been published in the Journal of Safety in Extreme Environments. I am the primary author of this research paper, with the guidance and support of co-authors Faisal Khan, Salim Ahmed, and Syed Imtiaz. The main objective of this research is to propose a probabilistic approach to effectively capture the uncertain nature of the extreme wind trend at the Grand Bank area. The modeling outputs are then incorporated to construct a dynamic risk profile. Apart from the main goal, the research also involves comparing the results obtained using the proposed non-stationary approach with those obtained using a stationary methodology. The comparison highlights that the non-stationary approach is more accurate in capturing the high trend variability of extreme winds. This work has been modified based on valuable feedback received from the co-authors and peer reviewers during the publication process.

Abstract

Probabilistic approaches under stationary conditions are used to design offshore systems. However, the frequency and impact of extremes in offshore environments have been changing. The present study investigates whether there are significant inter-period trends in extreme wind loads from the historical period to the future. This study comprises two elements. The first element

considers the traditional Block Maxima based Generalized Extreme Value (GEV) approach and generates different return levels-based risk profiles (under stationary condition). In the second element, the GEV location parameters are considered as time-dependent and a dynamic risk profile of different returns for wind events is developed. Distribution parameters are computed using the well-known Maximum Likelihood Estimate (MLE) method. The proposed methodology is tested using data from the Grand Banks region of the east coast of Canada. The study observed a 100-year extreme wind speed of 33.7 m/s (occurrence probability, $9.99\text{E-}05$) assuming a stationary condition, whereas for the non-stationary model, the predicted extreme wind speed for 2030 is 41.5 m/s (occurrence probability $2.82\text{E-}05$) in the Grand Banks region, Canada. The study highlights that a 100- year return period is not an adequate design criterion under the current situation; one must consider the non-stationary behaviour or a higher return period such as 500 or 1000 years.

Keywords: Risk-based design, Return level, Offshore System Extreme value, Wind load

6.1. Introduction

The risk associated with an extreme environmental condition is of great interest for the design and operation of offshore and coastal installations. Among the climate factors, wind can cause a significant threat to such structures and thus has a major influence on their design. Underestimation of risk can lead to infrastructure failures, whereas overestimation can lead to expensive construction and inefficient resource allocation. As a result, developing a realistic methodology for determining an appropriate design wind speed requires striking a balance between safety and cost. Existing Extreme Value Theory (EVT) or tail-based probabilistic modelling methodologies (Levine, 2009; Das et al., 2016, Wang and Li, 2016; Hu and Ayyub, 2017, Pryor and Barthelmie, 2021; Debnath et al., 2021. etc.) permit 100 or 500 years of extreme wind speed projection, which

the author is interested in. However, choosing an adequate return level for suggesting a design wind speed is not easy. The problem of climate change complicates the estimation process further. Towards this, the risk-based methodologies have been used to evaluate the design variables in offshore engineering facilities (e.g., Sulistiyono et al., 2015; Kang et al., 2017; Hallowell et al., 2018 etc.). In a similar aspect, the stationary risk-based methodology (Arif et al., 2020a and 2020b) was proposed for adopting extreme modelling outputs in order to overcome this problem. However, the above works did not consider non-stationary processes. Furthermore, events that span a century or a millennium might happen at any time. The possibility of 100/500-year events occurring this year or in the future, as well as the risk scenario related to them, when non-stationary factors are taken into account, have not been satisfactorily addressed in previous studies. This study aims to develop a non-stationary risk-based approach for anticipating extreme design wind speed by combining risk factors, event occurrence probability and wind load. If so, finding an extreme distribution and formulating adequate return levels in non-stationary conditions will be difficult and the main emphasis of this research.

Towards identify an extreme distribution, the Extreme Value Theory (EVT) based modelling approach focuses on the tail of the sample distribution and has been applied successfully in a variety of fields, including hydrology (e.g. Paola et al., 2018), coastal engineering (e.g. Leo et al., 2021), ocean engineering (e.g. Mackay and Jonathan, 2020), offshore load estimate (e.g. Ramadhani, 2021) and climate change issues (e.g. Lombardo and Ayyub 2014; Erik Vanem, 2015; Hu and Ayyub, 2017 etc.). The Block Maxima-based Generalized Extreme Value (GEV) distribution and the Peak Over Threshold (POT)-based Generalized Pareto Distribution (GPD) are prevalent among all of them. The POT method employs all significant observations, whereas the BM method overlooks some significant observations that may occur outside of the block. In both

systems, determining thresholds and the optimal block size is difficult. Aside from that, there are several unknowns in the implementation process, such as the size of the data set, its quality, the methodology employed, a lack of knowledge, and so on. Higher return levels frequently increase the level of risk. Furthermore, it is unknown how to incorporate the climatic trend into an extreme value analysis from the standpoint of climate change, which adds to the overall uncertainty. In this context, adopting extreme climate trends and capturing uncertainty is challenging. The block maxima based GEV distribution was implemented in several fields with parameters (location or scale) that changed linearly with time (for example, Vanem, E. 2015; Paola et al., 2018 etc.), or an appropriate polynomial (for example, Silva and Simonovic, 2020, Panagoulia et al., 2014 etc.) or sinusoidal/log-sinusoidal functions (Katz et al., 2002) to capture the non-stationarity of a process. In a similar vein, to adapt the non-linearity extreme behaviours of nature, the authors investigated a variety of GEV parameter linear/non-linear trends for the location/scale parameters and their varied combinations. In order to predict return levels across several years, the return level function is used in the estimation of wind load. The risk factor wind load is expressed in terms of wind energy in this study. Finally, a dynamic risk profile is generated by merging extreme modelling predictions on the Grand Banks area (Newfoundland offshore area). The current methodology differs from Arif et al., 2020a, Arif et al., 2020b, and Arif et al., 2022a in that it constructs a non-stationary risk profile. The organization of this work is as follows: the suggested methodology is described in depth in Section 6.2, and the case study, methodology validation, and major conclusions are presented in Section 6.3. Section 6.4 contains the study's concluding remarks.

6.2. Methodology

Risk estimation involves estimation of the probability of an event and its impact. In this work, the occurrence probability of an extreme event is determined by the best fitted “extreme distribution”

and the impact (wind energy) is calculated using the projected "extreme return levels." To accomplish this, the current study aims to develop an acceptable EVT-based extreme model to reflect the current extreme characteristics of wind speed, the methodology is shown in Figure 6.1.

6.2.1 Data processing and findings trend

The random behavior or statistical parameters of a non-stationary stochastic process change with time. Wind activity could be considered a non-stationary process with temporal volatility in the context of climate change. As a result, distinct data blocks are analyzed with statistical behavior (mean and standard deviation) changes during the data processing, and the trend is determined.

The process is classified as non-stationary if any data trends are found. Finally, various block

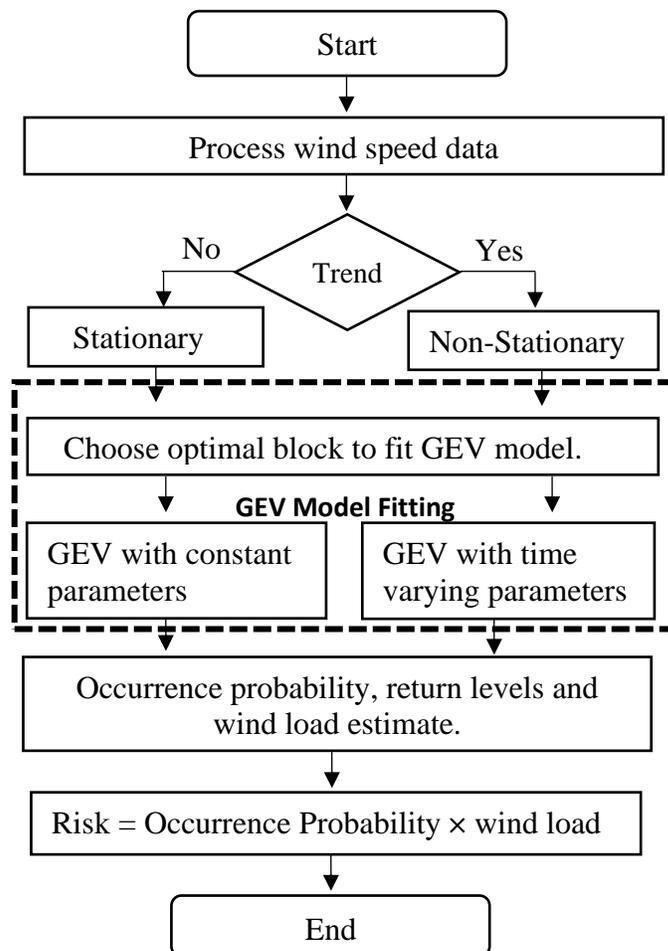


Figure 6.1. The proposed framework for extreme wind load risk analysis.

maxima data are extracted to suit the extreme model. The goal of this study is to look at extreme distributions in both stationary and non-stationary situations.

6.2.2 Extreme Distributions

Extreme analysis is commonly performed using the Generalized Extreme Value (GEV) distributions or the generalized Pareto Distributions (GPD). The GEV is the standard generalized form of three different distributions: Gumbel, Frechet, and Weibull. For a brief description and their different applications, please refer to Castillo et al., 2005. The GEV and GPD have been used to predict extreme wind speed (Liu et al. 2018; Campos and Soares, 2018), wave height (RM Campos et al.,2019, V Petrov et al.,2019), and extreme natural hazard estimation (Mazas, 2019) and so on. Arif et al. (2020a) proposed a heavy right tail distribution for the condition of data scarcity for iceberg extreme risk collision analysis, in addition to GPD and GEV. Appropriate distributions can be chosen depending on the study's interest, data statistics, and factors' suitability.

Table 6.1. Stationary and non-stationary GEV models.

Type	Symbol	Model parameter Combinations
M_I	$GEV(\mu, \sigma, k)$	$\mu(t) = \mu, \sigma(t) = \sigma, k(t) = k$
M_{II}	$GEV(\mu(t), \sigma, k)$	$\mu(t) = \mu_0 + \mu_1 t, \sigma(t) = \sigma, k(t) = k$
M_{III}	$GEV(\mu, \sigma(t), k)$	$\mu(t) = \mu, \sigma(t) = \sigma_0 + \sigma_1 t, k(t) = k$
M_{IV}	$GEV(\mu(t), \sigma(t), k)$	$\mu(t) = \mu_0 + \mu_1 t, \sigma(t) = \sigma_0 + \sigma_1 t, k(t) = k$
M_V	$GEV(\mu(t), \sigma, k)$	$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2, \sigma(t) = \sigma, k(t) = k$
M_{VI}	$GEV(\mu(t), \sigma(t), k)$	$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2, \sigma(t) = \sigma_0 + \sigma_1 t, k(t) = k$

In this research, five distinct GEV models (mentioned in Table 6.1) are considered. The model M_I is for stationary circumstances, whereas the remaining models (M_{II} to M_{VI}) are for non-stationary cases (by making the assumptions that there are linear and quadratic trends to the location parameter, linear trends to the scale parameter, and other combinations of these). The shape parameter is considered constant because the current study used limited data; however, the stability of the shape parameter with temporal change necessitates a substantial amount of data. The cumulative distribution function of Generalized Extreme Value (GEV) distribution is given (Castillo et al., 2005):

$$F(x) = \begin{cases} \exp \left\{ - \left[1 + k \left(\frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{k}} \right\} & \text{if } k \neq 0, \\ \exp \left\{ - \frac{(x-\mu)}{\sigma} \right\} & \text{if } k = 0. \end{cases} \quad (6.1)$$

where μ is the location parameter, σ is the scale and k is the shape parameter. The return level of the GEV model is defined as for the stationary case (Castillo et al., 2005):

$$x_{m_s} = \begin{cases} \mu - \frac{\sigma}{k} [1 - (-\log(1-p))^{-k}] & \text{for } k \neq 0, \\ \mu - \sigma \log\{-\log(1-p)\} & \text{for } k = 0. \end{cases} \quad (6.2)$$

The occurrence probability p is proportional to the level of m -return as $1/m$. For example, the value of p is 0.01 for a 100-year return level.

6.2.3 GEV fit: Optimal block size and parameter estimation

The best block size is determined by balancing bias and variance. Small block sizes lead to bias, whilst large block sizes lead to high variations. The current study looked at different blocks based on different time periods, such as yearly, two years, five years, and ten years, and determined the best block size to apply GEV. The appropriate block size is determined by the parameter estimates with minimal error. Smaller error in parameter estimate took the priority to choose the optimal

block. The most prevalent methods for estimating GEV parameters include Maximum Likelihood Estimation (MLE), Probability Weighted Moments (PWM), and Bayesian approaches, with the optimum method chosen based on a small error in parameter values. The current investigation determined that all of the listed approaches have similar accuracy; lastly, the GEV model parameters are calculated using the common MLE in both stationary and non-stationary scenarios.

6.2.4 Goodness of fit test

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used as performance criteria for GEV fit in this study.

The AIC is defined as (Katz et al., 2013):

$$AIC = 2 * nllh + 2n, \quad (6.3)$$

where $nllh = -\log$ likelihood, and n is the number of parameter of GEV models.

The BIC is defined as (Katz et al., 2013):

$$BIC = 2 * nllh + n * \ln(L), \quad (6.4)$$

where L is the total number of the data points.

In this study, smaller AIC and smaller BIC provide a better fit since negative log-likelihood is taken into account. To see if the non-stationary model is valuable, or if the observed trend is significant (in other words, if the non-stationary model is better than the stationary model), a simple test was carried out. A deviation statistic is defined as follows with models $M_0 \subset M_1$ (Katz et al., 2013):

$$D = 2\{nllh(M_1) - nllh(M_0)\} \quad (6.5)$$

The asymptotic distribution of D is given by the χ^2 distribution. The calculated values of D can be compared to critical values of χ^2 ; large values of D indicate that model M_1 explains significantly more variation in the data than model M_0 .

6.2.5 Risk analysis

The risk evaluation approach follows the procedure defined in Arif et al. 2020a. The wind risk is denoted by R_{wind} , and defined as follows (Sulistiyono et al., 2015):

$$R_{wind} = P(O) \times C, \quad (6.6)$$

where $P(O)$ is the wind occurrence probability and C is the wind consequences (impact). The wind occurrence probability is defined as:

$$P(O) = p_e \times (RP)^{-1} \quad (6.7)$$

where p_e is the event exceedance probability and RP is the return period. The current study evaluated wind energy (E_{wind}) as a consequence or impact, which is defined as the product of wind force on a certain area and the distance at which the wind is measured. As a result, the wind energy, E_{wind} , is calculated using the formula (Germanischer, 2013):

$$E_{wind} = \frac{1}{2} \times \rho \times U^2 \times CS \times A \times d, \quad (6.8)$$

where U is proposed extreme wind speed (m/s, predicted return levels), ρ is the density of air (1.225 kg/m³), CS is the drilling derrick shape coefficient, considered as 1.25 (Germanischer, 2013), and d is the distance of wind measured from surface (10 m height). The Area A is considered as 3332 m² (the Hibernia oil drilling platform top side length of 98 m and width of 34 m (HIBERNIA, 1997)).

The proposed risk-based methodology offers flexibility in determining the risk of wind load in a particular area. More precisely, based on various return levels anticipated by the model, a set of extreme wind loads (in terms of energy) is estimated together with associated occurrence probability. Finally, the risk values are evaluated by multiplying the associated likelihood and impact energy; an appropriate design extreme wind speed is chosen based on offshore operation standards for an acceptable risk level. Arif et al. 2020a provides additional details.

6.2.6 Uncertainty estimate

To quantify uncertainty in parameter distributions and return level estimates, a parametric bootstrap is used. The steps of the parametric bootstrap are as follows (Arif et al 2020a):

1. From the fitted GEV model, generate a simulated sample with the same length as the actual data.
2. Fit the GEV model to the simulated sample and record the parameter estimates that arise (return levels).
3. Repeat steps 1 and 2 a thousand times to get 1,000 model parameter values and 1,000 return period values.
4. For each time interval, calculate the mean value and a 95% confidence interval for the model parameters and return level. The 95% confidence interval is computed as (mean $\pm 1.96 \times$ standard error), where 1.96 = significance level, standard error = $s/\sqrt{1000}$, and s is the standard deviation.

By using the estimated confidence interval of the return level, the wind energy and finally, risk uncertainty, are estimated.

6.3 Case study

The study area is located about 340 kilometres east-southeast of St. John's, Newfoundland and Labrador. The Grand Banks region (Terra Nova, Hibernia, Hebron, and White Rose fields) produces thousands of barrels of oil per day. From an economic standpoint, this is an area of significant interest. However, the risk associated with drilling operations resulting from natural hazards, e.g. icebergs, wind speed, and wave height, is also significant. The wind is the most powerful natural hazard since it influences the effects of other hazards. Hourly wind speed data (observation and Hindcast model) are taken into account in this research at (44.250N, 53.620W).

The Hindcast model data (Swail et al., 2006) for the period 1988-2015 is compared to the observational data (DFO, 2021) for the period 1988-2013. Missing values from one year (2001) are analysed using linear interpolation in the observational data. For validation of the approach, two current years' data (2014, 2015) from the Hindcast model are added to the observational data. Finally, the proposed methodology is tested on data from 1988 to 2015. Minitab is used for descriptive data analysis, and the open-source platform R (available at <https://www.rstudio.com>) packages (extRemes, and ismev) are utilised for data modelling.

6.3.1 Descriptive statistics

The numerical and graphical representations of the data allow the reader to learn from the past and establish a foundation for selecting a better modelling methodology. Table 6.1 shows the statistics of the wind speed data in the study region. During the study period, the average wind speed was 7.3 m/s (computed through data median), with a maximum of 35.2 m/s. The data set does not match the normal distribution, since the skewness and kurtosis values are not zero (Table 6.2). Furthermore, the positive skewness of 0.4129 shows that the data set has a right tail, while the positive kurtosis of 3.0986 suggests that the data has a long tail. Data values between 18.13 and 26.15 are "outliers," whereas data points larger than 26.15 are "severe outliers," according

Table 6.2. Descriptive data statistics: hourly wind speed (m/s) data at the Grand Banks (1988-2015).

Data Points	Max	Median	Skewness	Kurtosis	Q3+1.5* IQR	Q3+3*IQR
124,473	35.2	7.3	0.4129	3.0986	18.13	26.15

to the Devore definition (Devore, 2011). According to the aforementioned data statistics, the data set has a lengthy right tail, as illustrated in Figure 6.2. The tail's weight was visualised using a histogram plot (Figure 6.2 (left)). In Figure 6.3, the hourly wind speed time series plot shows that

the majority of the data falls below 16.7 m/s, with a vertical red line separating all the severe events. Except for the year 2011, the Hindcast results are shown to be high, when compared to observational data. For this location, observation data is available for the years 1988 through 2013, with one year missing (2001). The missing data for 2001 is computed using linear interpolation. The Hindcast data are listed as high compared to observational data, except for the year of 2011. Finally, observational data are combined with simulated data for the years 2014 and 2015. As a trade-off between bias and variance, the best block size is set at one year. The present study discovered that using blocks of two, five, and ten years to predict GEV parameters caused an increase in the error term for the parameters. Finally, the entire data set is divided into 28 blocks

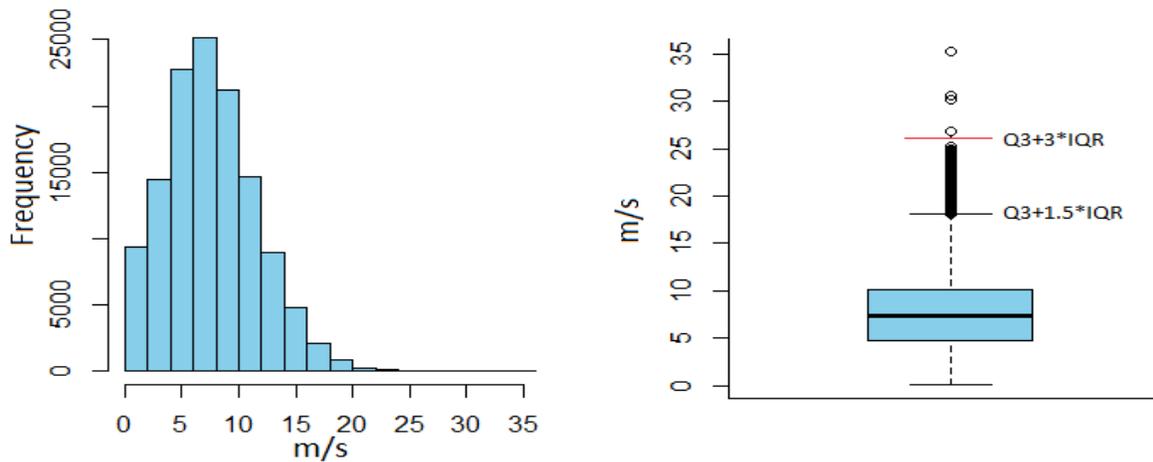


Figure 6.2. (Left) The frequency distribution of wind speed data (1988-2015). (Right) the box plot, showing data profile and extreme outliers are separated by the red lines.

(1 year block); each block maximum value is computed, and the GEV is fitted to the data set of yearly block maxima.

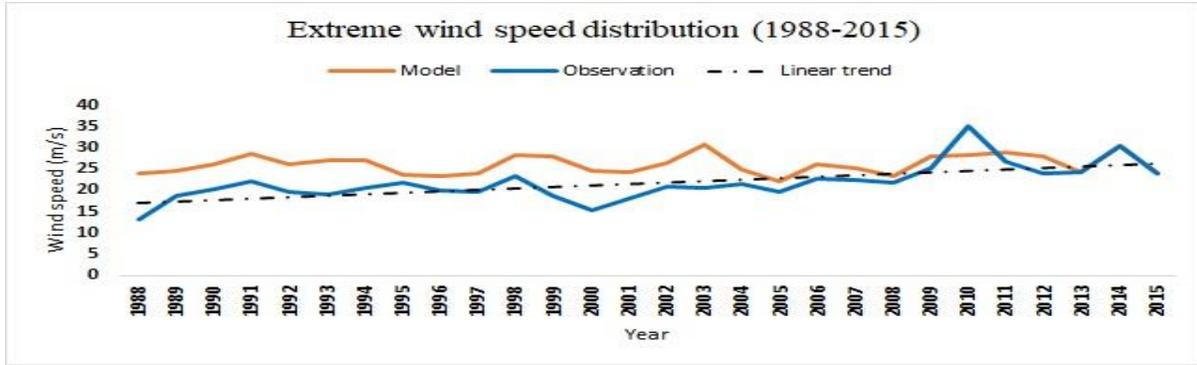


Figure 6.3. Annual wind speed distribution. The observational data vs Hindcast model data.

6.3.2 Fit distributions and model validation

On the annual block maxima data, various GEV models are fitted. Table 6.3 lists the GEV parameter values as well as the accompanying AIC and BIC values. Compared to all other listed

Table 6.3. Different GEV models are provided, along with parameter values and test statistics. In parenthesis, the standard errors are indicated.

GEV Models	Estimated GEV Model Parameters values (error)				
	Location	Scale	Shape	AIC	BIC
M_I	$\mu = 20.06 (0.739)$	$\sigma = 3.60(0.495)$	$k = -0.087 (0.10)$	162.56	166.56
M_{II}	$\mu_0 = 15.60(1.14)$ $\mu_1 = 0.33(0.07)$	$\sigma = 2.59(0.37)$	$k = -0.06 (0.10)$	147.38	152.71
M_{III}	$\mu = 19.47 (1.03)$	$\sigma_0 = 2.77 (0.96)$ $\sigma_1 = 0.08 (0.1)$	$k = -0.19 (0.15)$	163.86	169.19
M_{IV}	$\mu_0 = 15.58 (1.19)$ $\mu_1 = 0.34 (0.07)$	$\sigma_0 = 2.63 (0.87)$ $\sigma_1 = 0.01 (0.06)$	$k = -0.05 (0.13)$	149.38	156.04
M_V	$\mu_0 = 16.65 (1.49)$ $\mu_1 = 0.12 (0.21)$ $\mu_2 = 0.008 (0.01)$	$\sigma = 2.55 (0.37)$	$k = -0.06 (0.11)$	148.26	154.92
M_{VI}	$\mu_0 = 16.89 (1.56)$ $\mu_1 = 0.08 (0.22)$ $\mu_2 = 0.009 (0.007)$	$\sigma_0 = 2.74 (0.74)$ $\sigma_1 = -0.02 (0.05)$	$k = -0.03 (0.14)$	150.24	158.23

models, GEV model type II (M_{II}), has the better fit (based on smaller error in GEV parameters value, and smaller AIC, BIC values). In addition, the type of GEV model is justified by the deviation statistics discussed in section 6.3. For instance, $M_{II} \subset M_V$, the deviation statistics are $D = 2 \times (-69.13 + 69.69) = 1.12$, and $\chi^2(0.05) = 4.57$; hence, the M_V model should be rejected. Except for M_{III} , all models, M_{IV} to M_{VI} , have similar statistics as the M_{II} model, and all are rejected. The M_{III} model has the same statistics as the model M_{II} . Furthermore, in the case of $M_I \subset M_{II}$, M_I is rejected, indicating that the non-stationary model accounts for far more volatility in the data. In addition, when compared to all other models, the M_{II} model has a small error. Finally, as explained in section 6.5, the 95 percent confidence interval is calculated. In the case of stationary model (M_I) it is computed as $\mu = 20.06$ [18.61, 21.50], $\sigma = 3.60$ [2.63, 4.57] and $k = -0.087$ [-0.267, 0.093]. For the case of the non stationary model, $\mu_0 = 15.60$ [13.37, 17.83], $\mu_1 = 10.33$ [0.19, 0.47], $\sigma = 2.59$ [1.86, 3.32] and $k = -0.06$ [-0.26, 0.14]. Therefore Equation 6.2 is modified for the best fitted GEV models (M_{II}) for the case of non-stationary (*for* $k \neq 0$):

$$x_{m_{n,s}} = (15.60 + 0.33t) - \frac{2.59}{0.06} [1 - (-\log(1-p))^{0.06}] \quad (6.9)$$

For $t = 29, 30, 31$, and so on, the wind speed estimates for the years 2016, 2017, 2018 and so on are provided with specific p values. If an offshore operator is interested in 100-years return levels, then p value is 0.01; and Equation 6.3 might be treated as the forecast of 100-year return level. More precisely, for the year 2022 ($t = 35$), the (a 100-year return level) predicted wind speed is 40.87 m/s, and for a 50-year return level, 2022's wind speed is estimated as 38.54 m/s. In the Grand Banks area, the 100-year extreme wind speed was predicted by C-CORE (2015) to be 33 m/s under stationary conditions, which is comparable to our predicted wind speed (33.7 m/s). Table 6.5 clearly shows the effects of climate change in comparison to Table 6.4. In more detail, the

wind speed was calculated for the stationary scenario at 38.8 m/s (with an occurrence probability of $9.95E-07$), and the non-stationary case at 39.8 m/s is predicted to occur in 2030

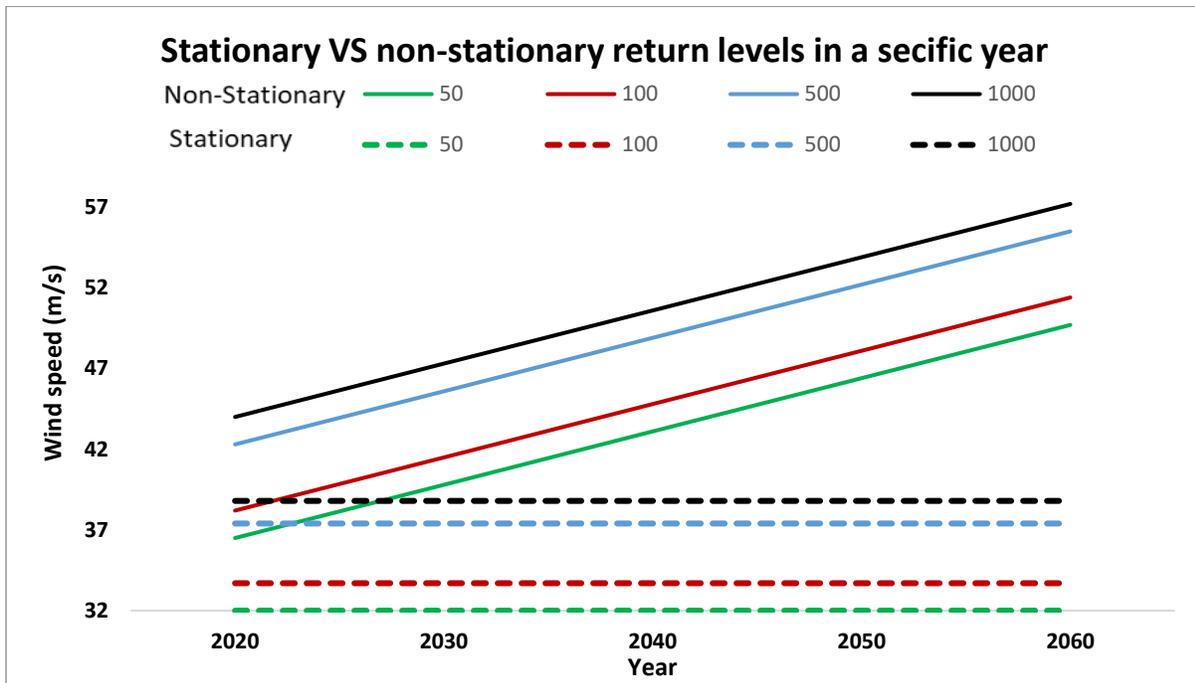


Figure 6.4. Different return levels estimate (Stationary vs non-stationary)

(with occurrence probability $9.56E-05$). A 1000-year event occurring more frequently in recent decades is, therefore, a visible sign of climate change. As a result, the 100-year return period mentioned in C-CORE (2015) is no longer extreme according to the criterion of climate change. In Figure 6.4, the various return levels for both stationary and non-stationary situations for the next five decades are indicated.

6.3.3 Risk analysis

Equations 6.7 and 6.8 are used to calculate the event occurrence probability and wind energy, respectively. Finally, using equation 6.6, the risk is calculated. Wind energy (based on various return levels) and the associated event occurrence probability are both included in Table 6.4 to create a flexible risk profile. The offshore operator determines a design wind speed depending on

Table 6.4. Stationary risk profile: Different expected return levels are used.

Risk factors	Return levels (stationary)					
	5 years	20 years	50 years	100 years	500 years	1000 years
Probability	4.00E-02	2.49E-03	3.99E-04	9.99E-05	3.98E-06	9.95E-07
Wind speed (m/s)	2.51E+01	2.95E+01	3.20E+01	3.37E+01	3.74E+01	3.88E+01
Wind Energy (MJ)	1.61E+01	2.22E+01	2.61E+01	2.90E+01	3.56E+01	3.84E+01
Risk (MJ/year)	6.43E-01	5.54E-02	1.04E-02	2.89E-03	1.42E-04	3.82E-05

Table 6.5. Non-stationary risk profile: Risk in a certain year based on different return levels.

Return Levels	Risk factors	In a specific Year				
		2020	2030	2040	2050	2060
50 years	Probability	2.96E-04	9.56E-05	2.85E-05	7.59E-06	1.82E-06
	Wind speed (m/s)	3.65E+01	3.98E+01	4.31E+01	4.64E+01	4.97E+01
	Wind energy (MJ)	3.39E+01	4.04E+01	4.73E+01	5.49E+01	6.30E+01
	Risk (MJ/year)	1.01E-02	3.86E-03	1.35E-03	4.17E-04	1.15E-04
100 years	Probability	9.08E-05	2.82E-05	7.99E-06	2.05E-06	4.64E-07
	Wind speed (m/s)	3.82E+01	4.15E+01	4.48E+01	4.81E+01	5.14E+01
	Wind energy (MJ)	3.73E+01	4.40E+01	5.13E+01	5.91E+01	6.75E+01
	Risk (MJ/year)	3.39E-03	1.24E-03	4.10E-04	1.21E-04	3.13E-05
500 years	Probability	8.01E-06	2.34E-06	6.17E-07	1.45E-07	2.97E-08
	Wind speed (m/s)	4.23E+01	4.56E+01	4.89E+01	5.22E+01	5.55E+01
	Wind energy (MJ)	4.56E+01	5.30E+01	6.10E+01	6.95E+01	7.86E+01
	Risk (MJ/year)	3.66E-04	1.24E-04	3.76E-05	1.01E-05	2.34E-06
1000 years	Probability	5.05E-07	1.24E-07	2.68E-08	4.96E-09	7.58E-10
	Wind speed (m/s)	4.40E+01	4.73E+01	5.06E+01	5.39E+01	5.72E+01
	Wind energy (MJ)	4.94E+01	5.71E+01	6.53E+01	7.41E+01	8.35E+01
	Risk (MJ/year)	2.50E-05	7.09E-06	1.75E-06	3.68E-07	6.33E-08

the proper risk threshold. In details, Table 6.4. shows that in the event of stationary, the risk level of 10^{-4} arises for the first time in 500 years. As a result, if offshore operators choose the risk level of 10^{-4} , they may choose a design wind speed of 37.4 m/s in the Grand Banks region in the stationary case; the event occurrence probability is $3.98E-06$. In a non-stationary example, the likelihood of 500-year occurrences occurring in 2030 is 42.3 m/s, with an occurrence probability of $2.34E-06$ (Table 6.5). Additionally, a prediction of a 1000-year risk (stationary case) is more common in recent decades when compared to non-stationary. As a result, operators can select a standard risk level (depending on their interests), construct a non-stationary risk profile (using Equation 6.9), and determine a design wind speed for developing or designing engineering facilities for the Grand Banks.

6.4 Conclusions

In both stationary and non-stationary scenarios, the BM-based GEV is utilised to forecast catastrophic wind speeds in the Grand Banks area. The dynamic behaviour of extreme occurrences is examined using a non-stationary GEV distribution. Finally, model projections are used to calculate the impact (wind energy) and risk. Non-stationary techniques is found to be useful in determining return levels that accurately reflect the current state of nature's extremes. Prior studies had utilised the 100-year return time as the most common measure for offshore facilities (CORE, 2015); however, the study found that, under the non-stationary scenario, recent decades have exceeded the 1000-year extreme (stationary case) due to climate change. As a result, the risk is evaluated using a 1000-year return time to approximate the current climatic trend. The data for this study come from only one locaion on the Grand Banks. Several areas may need to be added in the future to account for geographic differences. The proposed approaches could be used to estimate other environmental variables like extreme wave height or iceberg speed. Inaddition,

event duration is a significant risk factor because extreme events that last longer have catastrophic effects. Author is interested in a multivariate risk-based approach by combining wave height, wind speed, duration, and iceberg speed.

7. Contributions, Findings, and Future Research Recommendations

Extreme event risk analysis in the offshore engineering domain plays a pivotal role in safeguarding critical infrastructure and ensuring the resilience of offshore structures under extreme conditions/harsh environment. Through the utilization of sophisticated probabilistic methods and advanced modeling techniques, the analyses provided in this thesis are expected to enable engineers and offshore operators to identify and address risk of catastrophic events with precision and accuracy. By incorporating historical data, simulations, and climate change projections, this research gained valuable insights into capturing the nature of extreme events and assessed their potential consequences on offshore assets. By implementing both univariate and multivariate approaches under stationary and non-stationary conditions, the proposed methods provide deeper understanding of the complex interactions among various environmental variables, such as iceberg speed, wind speed, wave height, and others; and modeling outcomes incorporated into Risk assessment. Despite challenges related to data availability and uncertainties in climate projections, this research demonstrates a successful application of the developed frameworks in real-world case studies. The proposed methodologies are crucial in guiding risk management strategies, emergency response planning, and the implementation of appropriate safety measures in the offshore engineering domain. By understanding and mitigating rare event risks, offshore operators can ensure the sustainable development of offshore infrastructure and safer offshore operations on marine environment.

7.1 Novelty and contribution:

This research makes a significant contribution through the development of an innovative methodology that better captures current extreme events, estimates extreme occurrence probabilities, and evaluates environmental loads. Moreover, this methodology integrates modelling outcomes to enhance offshore operation safety through a risk-based approach. By combining these elements, the study provides a more robust and reliable framework for mitigating risks and ensuring the security of offshore structures in extreme conditions. Furthermore, what sets this methodology apart is its utilization of a joint modeling approach, exploring extreme correlations and dynamic risk profile methodology, along with uncertainty estimates that specifically address climate change concerns. These unique elements collectively strengthen the approach and its applicability to address the challenges of today's offshore environments. The highlighted contributions are outlined below:

- A novel mathematical representation for rare events within the offshore engineering domain was introduced and proposed (by adopting the definition of outliers and extreme outliers as presented by Devore (2011), which was originally introduced by Tukey (Tukey, 1977)).
- Evaluated climate change's impact on environmental variables at offshore domain and incorporated in offshore risk assessment methodology.
- Proposed univariate models (frequentist and Bayesian approaches) to capture present day extreme characteristics and came up with an uncertainty estimate for flexibility to adopt climate change impact and incorporated them in risk methodology.
- Developed models (bivariate/multivariate) to capture the extreme dependency in environmental load and its uncertainty.

- This research stands out by incorporating rare event modelling outcomes to estimate extreme loads/energy and integrating them into the risk assessment methodology. This unique approach enhances the understanding of extreme events and contributes to a more comprehensive and accurate evaluation of potential risks.
- The proposed probabilistic low-resolution risk assessment technique offers a computationally efficient and viable alternative to computationally expensive numerical models in the offshore domain. Its streamlined approach sets it apart as a practical and valuable method for risk assessment.
- The dynamic risk assessment methodology developed in this study enables the continuous monitoring of current risk levels over time. This capability facilitates the timely updating of risk policies as required, ensuring a proactive and adaptive approach to risk management.

7.2 Conclusions

Extreme event risk assessment in offshore engineering is a critical process to ensure safety, reliability, and resilience of offshore structures and operations. By carefully studying and quantifying low-probability, high-impact events, this analysis allows us to identify potential vulnerabilities and develop robust strategies to mitigate their consequences.

7.2.1 Univariate risk assessment methodology (frequentist modeling approach)

A univariate methodology is presented for modeling rare events and developed a risk assessment framework. The methodology is applied to analyze iceberg collision risk in the Flemish Pass basin. During the study, it was observed that popular extreme models such as the Generalized Pareto Distribution (GPD) or the Generalized Extreme Value (GEV) failed to accurately capture present-day extremeness. The occurrence of 100-year events (return levels)

became more frequent due to climate change. To address those issue, the rare event problem was considered as a heavy-tailed event and a Heavy Right Tail Model (HRTM) was implemented to better capture present-day extremeness. Due to limited data availability, the model parameters were estimated using Hill and SmooHill estimators, as the commonly used Maximum Likelihood Estimator (MLE) was unsuitable for this case study. Furthermore, adopting 1000-year return level window instead of the commonly used 100-year return level, resulted in a more suitable estimation of extreme load. Finally, incorporating extreme load and its corresponding occurrence probability, a risk profile was proposed, specifying the design iceberg speed for future structures, and ensuring safer engineering operations in the existing facilities at the Flemish Pass basin. The offshore operators may adopt this risk profile to upgrade current iceberg management policies and enhance the safety of current existing facilities. Details of these outcomes are listed on Chapter 3.

7.2.2 Univariate risk assessment methodology (Bayesian modeling approach)

According to this study, between 2002 and 2017, a limited number of large icebergs (weighing 10 MT) were observed in the Jeanne d'Arc basin, constituting approximately 14% of all icebergs recorded during the study period. Consequently, the estimation of large iceberg speeds was deemed a rare event problem in this research. To address the challenge of data scarcity, Bayesian inferences were employed for HTRD (Hill Tail Range Distribution) parameter estimation in conjunction with the Hill and SmooHill estimator method presented in Chapter 3. The study found that Bayesian inference yielded a superior estimate for HRTD parameter estimation compared to the Hill and SmooHill estimators. This conclusion was supported by the Cumulative Distribution Function (CDF) comparison plot (listed as 4.11) and the error analysis presented in Table 4.3. The modeling outcomes were then utilized for iceberg risk

assessment, culminating in the proposal of a flexible risk profile. This risk-based approach offers a dependable design standard for offshore constructions, considering the variability and extreme characteristics associated with iceberg events. By incorporating Bayesian inference, this methodology contributes to a more accurate and reliable understanding of iceberg risks, thereby aiding in effective decision-making for offshore engineering projects.

7.2.3 Joint extreme events load analysis and risk assessment

The offshore domain's complexity and the interdependence of environmental variables necessitate a multivariate approach to comprehend their interactions. By employing multivariate dependence modeling, a more inclusive risk assessment can be achieved, considering factors such as wave height, wind speed, ocean currents, and structural response. This approach enables a more precise representation of potential hazards, enhancing risk evaluation accuracy. In addition, environmental variables naturally exhibit correlations, however, under extreme conditions, and climate change impact, these correlations may undergo changes, demanding a thorough understanding of their manifestations. In this study, extreme wind speed and wave height data in the Flemish Pass basin were jointly analyzed and modeled. The outcomes of this joint modeling were then integrated into a risk assessment methodology, yielding a flexible risk profile tailored for the Flemish Pass basin. Apart from its primary objective, the proposed finer scale risk profile methodology ($0.1^{\circ} \times 0.1^{\circ}$ latitude/longitude grid) offers operators a valuable tool for managing offshore operations in specific areas, serving as a cost-effective alternative to computationally expensive numerical models. The research found due to impact of climate change, revealing a noteworthy 30% decrease in the correlation between wind speed and wave height during the recent years [1989-2018] compared to the earlier period of 1959 to 1988. As a result, the study proposes design

wind speed and wave height recommendations to enhance offshore operations' safety in the Flemish Pass basin. These research methodologies help to bolster risk management strategies, ensuring a safer operating environment in light of changing environmental conditions.

7.2.4 Extreme event load analysis and a non-stationary risk-based approach

Environmental variables possess an inherent chaotic nature, necessitating the capturing of their variability over time to enable realistic load prediction. In this study, a methodology is proposed to establish a dynamic risk profile at the Grand Banks area. The focus of this methodology lies in extreme wind load prediction, which is then integrated into the overall risk assessment. The Block Maxima-based Generalized Extreme Value (GEV) approach is employed to handle both stationary and non-stationary scenarios in the Grand Banks region. Six different GEV models (as listed in Table 6.1) are considered to capture linear and nonlinear trends. By utilizing metrics like AIC, BIC, and deviation statistic, the most suitable model is selected. Wind energy estimation is then conducted, leading to the development of a dynamic risk profile for the upcoming five decades (as presented in Table 6.5). Notably, the study reveals that in the non-stationary scenario, the extreme wind speeds observed in recent decades have surpassed the 1000-year extreme recorded in the stationary case. This non-stationary approach proves to be highly valuable for offshore operators as it allows them to update current policies based on the latest risk estimates.

7.3 Future research opportunity

The current key questions to address: Is the proposed methodology sufficient to fully capture and prevent any uncertain events in the future? The answer is NO. While we may not be able to entirely prevent unforeseen situations, this research empowers us to be proactive and better prepared in safeguarding offshore operations against extreme events. For more realistic

modeling outcomes/capturing accurate extreme characteristics over the time, the proposed methodology might have opened a door for future research opportunities for rare event risk assessment in offshore engineering. It presents exciting avenues for enhancing safety, sustainability, and efficiency in the industry. Here are some potential areas for future research:

- **Incorporating climate change projections:** As climate change continues to impact weather patterns and environmental conditions, it becomes imperative to integrate climate change projections into rare event risk assessment models. Proposed methodology might improve by focusing on understanding how climate change alters the frequency and intensity of extreme events, enabling better anticipation and preparation for potential hazards.
- **Data collection and monitoring:** This research utilizes data collected from two sources: the International Ice Patrol (IIP) iceberg sighting database and Hindcast model-simulated data as presented by Swail et al. in 2006. However, improving data collection methods, including the use of remote sensing technologies, IoT devices, and real-time monitoring systems, can provide better data for rare event risk assessment. In addition, the proposed methodology can explore innovative ways to collect and integrate data from multiple sources to enhance the accuracy and reliability of risk assessment methodology. In the present thesis, the influence of environmental controls on extreme outliers has not been considered. For instance, during certain storm periods, there could be a noteworthy number of extreme outliers, potentially affecting the occurrence of typical outliers. Therefore, rather than encompassing all extreme events during a storm, considering only the most extreme event may yield more accurate estimates.

- **Uncertainty quantification:** The assessment of rare event risks requires addressing uncertainties present in both data and model parameters. The current study emphasizes the use of bootstrapping, where bootstrap data sets are randomly generated from the fitted distribution instead of directly sampling from the original data. However, there remains an opportunity for future research (say for example: machine learning, Monte Carlo simulation, Fuzzy logic etc.) to concentrate on developing robust uncertainty quantification techniques. Such advanced methods would provide a more comprehensive consideration of uncertainties in the analysis, ultimately leading to more informed decision-making processes. “Regarding load uncertainty, the current study addresses the uncertainty related to wind, waves, and iceberg speed. However, in addition to the variables we have considered load uncertainty which may be influenced by other factors, such as ocean depth, ocean current, wave periods etc. Therefore, in the future, a more comprehensive assessment of load uncertainty may be necessary to account for these additional phenomena. By incorporating these techniques, a deeper understanding of rare event risks can be achieved, enabling more effective and reliable risk management strategies.
- **Multivariate approaches:** Enhancing the understanding of the interdependence among various offshore engineering parameters and environmental factors can lead to more comprehensive multivariate risk assessment methods. Investigating the correlations between different variables and their influence on rare event occurrences can improve the accuracy of risk models. The current multivariate methodology has been successfully implemented in a bivariate case, specifically in Chapter 5, where wind speed and wave height estimation at Flemish Pass basin were examined. Chapter 3 and

Chapter 4 presented univariate analyses on Iceberg collision risk. However, as a potential future study, expanding the scope to jointly analyze iceberg collision risk with other variables such as wind speed, wave height, ocean currents, and wave periods could present a promising avenue for multivariate approaches, akin to the bivariate methodology demonstrated in Chapter 5. Such an extension would provide a more holistic understanding of the interplay between multiple factors and their collective influence on rare event scenarios, ultimately leading to a more comprehensive risk assessment framework.

- **Non-stationary modeling:** Environmental parameters are subject to changes over time, making it essential for research to explore non-stationary modeling approaches. These approaches aim to capture the dynamic nature of rare events while considering the evolving environmental conditions and their influence on risk. The univariate non-stationary methodology introduced in Chapter 6 and the proposed dynamic risk profile offer valuable insights that can be easily adapted to bivariate or multivariate scenarios. By incorporating multiple environmental variables, a more realistic risk assessment can be achieved, enhancing the safety and resilience of offshore operations.
- **New approaches:** The current methodology for predicting rare events relies on return level functions, which can be further enhanced by the implementation of Artificial Intelligence (AI) and Machine Learning (ML) techniques. By utilizing AI and ML algorithms, rare event predictions can be optimized and incorporated into the risk assessment methodology, leading to a more realistic and accurate risk profile. The integration of advanced data-driven approaches enables a deeper understanding of rare events, improving the overall effectiveness of risk analysis and ensuring better

preparedness for extreme scenarios. AI and machine learning algorithms can provide opportunities for automating aspects of rare event risk assessment and optimizing decision-making processes. These technologies can help identify patterns and trends in large datasets and improve the efficiency of risk analysis.

In summary, the future of rare event risk assessment in offshore engineering is characterized by advancements in data collection, modeling techniques, and interdisciplinary collaboration. By embracing these research opportunities, the industry can enhance its ability to anticipate, mitigate, and respond to rare events, ultimately ensuring a safer and more sustainable offshore environment.

References

Agbeyegbe, T. D. and Leon, G. (2002). The tail behavior of stock index returns on the Jamaican Stock Exchange. 23rd Annual Review Seminar of the Central Bank of Barbados. September 2002. New York.

Arif, M., Khan, F., Ahmed, S., and Imtiaz, S. (2020a). Rare event risk analysis—application to iceberg collision. *Journal of Loss Prevention in the Process Industries*, 66. 104199. <https://doi.org/10.1016/j.jlp.2020.104199>

Arif, M., Khan, F., Ahmed, S., and Imtiaz, S. (2020b). Evolving extreme events caused by climate change: A tail-based Bayesian approach for extreme event risk analysis. *Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*. 235. 6. <https://doi.org/10.1177/1748006X21991036>.

Arif, M., Khan, F., Ahmed, S., and Imtiaz, S. (2022a). A generalized framework for risk based extreme load analysis in offshore system design. *Journal of offshore mechanics and Arctic engineering*. 145(2): 021701 (11 pages). <https://doi.org/10.1115/1.4055553>

Arif, M., Khan, F., Ahmed, S., and Imtiaz, S. (2022b). Extreme wind load analysis using non-stationary risk-based approach. *Journal of Safety in Extreme Environments*. 4: 247–255. DOI:10.1007/s42797-022-00064-2

Arima, S., Petris, G. and Tardella, L. (2010). Alternative rare events probability estimation. 45th Scientific Meeting of the Italian Statistical Society. May 9, 2010.

Asadi, Z. S., and Melchers, R. E. (2017). Extreme value statistics for pitting corrosion of old underground cast iron pipes. *Reliability Engineering and System Safety*. 162: 64-71. <https://doi.org/10.1016/j.ress.2017.01.019>

- Aven, T. (2015). Implications of black swans to the foundations and practice of risk assessment and management. *Reliability Engineering & System Safety*. 134: 83-91.
<https://doi.org/10.1016/j.ress.2014.10.004>
- Balakrishnan, N., and Lai, C. D. (2009). *Continuous bivariate distributions*. Springer, New York.
<https://doi.org/10.1007/b101765>
- Bay, B., Yang, X. C., and Zhang, Q. H. (2013). Joint probability distribution of winds and waves from wave simulation of 20 years (1989-2008). *Water Science and Engineering*, 63: 296-307.
<https://doi.org/10.3882/j.issn.1674-2370.2013.03.006>
- Bedford, T. and Cooke, R. (2001). *Probabilistic risk analysis: foundations and methods*. United Kingdom: Cambridge University press.
- Beirlant, J., Y., Goegebeur, J., Segers, L., Teugels, J. L. (2006). *Statistics of Extremes: Theory and Applications*. England: John Wiley and Sons, ltd.
- Bitner-Gregersen, E. M. (2015). Joint met-ocean description for design and operations of marine structures. *Applied Ocean Research*. 51: 279-292. <https://doi.org/10.1016/j.apor.2015.01.007>
- Blackledge, J., Coyle, E., Kearney, D., McGuirk, R., and Norton, B. (2013). Estimation of wave energy from wind velocity. *IAENG Engineering Letters*. 4: 158-170.
- Bommier, E. (2014). *Peaks-over-threshold modeling of environmental data*. U.U.D.M. Project Report. Uppsala: Uppsala University.
- Bucher, A. and Zhou, C. (2021). A horse racing between the block maxima method and the peak-over-threshold approach. *Statist. Sci.* 36(3): 360-378. DOI: 10.1214/20-ST5795
- Campos, M., and Soares, C. (2018). Spatial distribution of offshore wind statistics on the coast of Portugal using Regional Frequency Analysis. *Renewable energy*. 123: 806-816.
<https://doi.org/10.1016/j.renene.2018.02.051>

Campos, M., Soares, C, and Alves, J., Parente, C. E., and Guimaraes, L.G. (2019). Regional long-term extreme wave analysis using hindcast data from the South Atlantic Ocean, *Ocean Engineering*, 179: 202-212. <https://doi.org/10.1016/j.oceaneng.2019.03.023>

Candela, A. and Aronica, G. T. (2017). Probabilistic flood hazard mapping using bivariate analysis based on copulas. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*. American Society of Civil Engineers. 3(1): A4016002. <https://doi.org/10.1061/AJRUA6.0000883>

Castillo, E., Hadi, A.S., Balakrishnan, N., and Balakrishnan, J. N. (2005). *Extreme value and related models with applications in engineering and science*. New Jersey: John Willey & Sons, Inc.

CBC report. (2020). <https://www.cbc.ca/news/science/greenland-ice-loss-1.5693549>. Toronto: CBC (Science). The Associated Press. Posted: Aug 20, 2020, 2:25 PM NDT.

C-CORE. (2015). *Metocean Climate Study Offshore Newfoundland & Labrador. Study main report. Volume 2: Regional trends and comparisons with other regions.*
[NALCOR-MetOcean-Vol.-2-Final-2017.pdf \(oilconl.com\)](https://www.oilconl.com/NALCOR-MetOcean-Vol.-2-Final-2017.pdf)

Clauset, A., Woodard, R. (2013). Estimating the historical and future probabilities of large terrorist events. *The Annals of Applied Statistics, Institute of Mathematical Statistics* 7(4): 1838-1865.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*. 65 (1): 141–151. <https://doi.org/10.1093/biomet/65.1.141>

Climate signals. (2019). European heat wave June 2019. Updated: October 15, 2021. <https://www.climatesignals.org/events/european-heat-wave-june-2019>.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*. 2: 339-365.

Foreman-Mackey, D., Hogg, D., Lang, D., and Goodman, J. (2013). emcee: the MCMC hammer. *Publications of the Astronomical Society of the Pacific*. 125. 925. 306.

Levine, D. (2009). Modeling tail behavior with extreme value theory. *Risk Management*. 17: 14-18.

Das, K. P., Sams, C. A. and Psingh, V. (2016). Characterization of the tail of river flow data by generalized Pareto distribution, *Journal of Statistical research*. 48-50 (2): 55-70.

Debnath, M., Doubrawa, P., Optis, M., Hawbecker, P., and Bodini, N. (2021). Extreme wind shear events in US offshore wind energy areas and the role of induced stratification. *Wind Energy Science*. 6 (4): 1043-1059.

Deidda, R., and Puliga, M. (2009). Performances of some parameter estimators of the generalized Pareto distribution over rounded-off samples, *Physics and Chemistry of the Earth, Parts A/B/C*. 34 (10-12): 626-634. <https://doi.org/10.1016/j.pce.2008.12.002>

Deng, Y., Cardin, M. A. V., and Babovic, D. Santhanakrishnan, P. Schmitter, and A. Meshgi. (2013). Valuing flexibilities in the design of urban water management systems. *Water research*. 47. 20: 7162-7174. <https://doi.org/10.1016/j.watres.2013.09.064>

Devore, J. L. (2011). *Probability and Statistics: For Engineering and the Sciences* (8th Edition). Boston: Cengage Learning.

DFO. (2021). Marine Environmental Data Section Archive. Ecosystem and Oceans Science, Department of Fisheries and Oceans Canada. <https://meds-sdmm.dfo-mpo.gc.ca>

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*. 7, 1–26. Institute of mathematical Statistics.

El-Gheriani, M., Khan, F., Chen, D., and Abbassi, R. (2017a). Major accident modelling using spare data. *Process Safety and Environmental Protection*. 106: 52-59.

El-Gheriani, M., Khan, F., and Zuo, M. J. (2017b). Rare event analysis considering data and model uncertainty. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, American Society of Mechanical Engineers. 3 (2): 021008.

Erik, V. (2015). Non-stationary extreme value models to account for trends and shifts in the extreme wave climate due to climate change. *Applied Ocean Research*.52: 201-211.

<https://doi.org/10.1016/j.apor.2015.06.010>

Estecahandy, M., Bordes, L., Collas, S. and Paroissin, C. (2015). Some acceleration methods for Monte Carlo simulation of rare events. *Reliability Engineering and System Safety*. 144: 296-310.

Fang, G., Pan, R., and Hong, Y. 2020. Copula-based reliability analysis of degrading systems with dependent failures. *Reliability Engineering and System Safety*. 193. 106618.

<https://doi.org/10.1016/j.ress.2019.106618>

Mazas, F. (2019). Extreme events: a framework for assessing natural hazards. *Nat Hazards* 98: 823–848. <https://doi.org/10.1007/s11069-019-03581-9>

Gaidai, O., Xu, X., Wang, J., Ye, R., Cheng, Y., and Karpa, O. (2020). SEM-REV offshore energy site wind-wave bivariate statistics by hindcast. *Renewable Energy*. 156: 689-695.

<https://doi.org/10.1016/j.renene.2020.04.113>

Genest, C., Remillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*. 44. 2: 199-213.

<https://doi.org/10.1016/j.insmatheco.2007.10.005>

Germanischer, L. S. (2013). Rules for the Certification and Construction Industrial Services-Offshore Substations. Hamburg: Germanischer Lloyd SE.

<https://doi.org/10.1115/1.4036155>

Gilli, M., and Keliezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*. 27: 207–228.

Goodarzi, E., Mirzaei, M., and Ziaei, M. (2012). M. Evaluation of dam overtopping risk based on univariate and bivariate flood frequency analyses. *Canadian Journal of Civil Engineering*. 39. 4: 374-387. <https://doi.org/10.1139/l2012-012>

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions express able in inverse form. *Water resources research*. 15 (5): 1049-1054. <https://doi.org/10.1029/WR015i005p01049>

Gudendorf, G., and Segers, J. (2010). Extreme-value copulas. *Copula theory and its applications*. Springer. 127-145. Part of the Lecture Notes in Statistics book series (LNSP, volume 198).

Guth, S. and Sapsis, T. P. (2019). Machine learning predictors of extreme events occurring in complex dynamical systems. *Entropy*. 21 (10): 925. MDPI. <https://doi.org/10.3390/e21100925>

Hallowell, S., Myers, A., Arwade, S., Pang, W. Rawal, P., Hines, E., Hajjar, J., Qiao, C., Valamanesh, V., and Wei, K. (2018). Hurricane risk assessment of offshore wind turbines. *Renewable Energy*. 125. 234-249. <https://doi.org/10.1016/j.renene.2018.02.09>

Hanum, H., Aji Hamim Wigena, Anik Djuraidah and Wayan Mangku. (2015). Modeling extreme rainfall with Gamma-Pareto distribution. *Appl. Math. Sci*. 9 .121: 6029-6039.

Haselsteiner, A. F., Coe, R. G., Manuel, L., Chai, W., Leira, B., Clarindo, G., Soares, C. G., Hannesdottir, A., Dimitrov, N., Sander, A., Ohlendorf, J., Thoben, k., Hauteclocque, G., Mackay, E., Jonathan, P., Qiao, C., Myersl, A., Rode, A., Hildebrandt, A., Schmidt, B., Vanem, E., and Huseby, A. (2021). A benchmarking exercise for environmental contours. *Ocean Engineering*. 236. 109504.

Haugh, M. (2016). An introduction to copulas. IEOR E4602: quantitative risk management. Lecture Notes. New York: Columbia University.

Haver, S., and Winterstein, S. R. (2009). Environmental contour lines: A method for estimating long term extremes by a short-term analysis. Transactions of the Society of Naval Architects and Marine Engineers. 116: 116-127.

Hibernia oil platform [Internet]. (1997). <https://structurae.net/en/structures/hibernia-platform>.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution, The annals of statistics. 3 (5): 1163-1174.

Horn, J. T., Gregersen, E. B., Krokstad, J. R., Leira, B. J., and Amdahl, J. (2018). A new combination of conditional environmental distributions. Applied Ocean Research. 73: 17-26. <https://doi.org/10.1016/j.apor.2018.01.010>

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society. 52 (1): 05-124. <https://doi.org/10.1111/j.2517-6161.1990.tb01775.x>

Hosking, J. R. M. and J. R. Wallis (1987), Parameter and quantile estimation for the generalized Pareto distribution. Technometrics. 29 (3): 339-349.

Hosking, J. R., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics. 27 (3): 251-261. <https://www.climatesignals.org/events/european-heat-wave-june-2019>

Hu, H., and Ayyub, B. M. (2017). Extreme precipitation analysis and prediction for a changing climate. ASCE-ASME J. Risk Uncertainty Eng. Syst. Part A: Civ. Eng. 4 (3): 04018029. <https://doi.org/10.1061/AJRUA6.0000980>

Hu, H. and Ayyub, B. M. (2019). Validating and enhancing extreme precipitation projections by downscaled global climate model results and copula methods. *Journal of Hydrologic Engineering*. 24 (7): 04019019.<https://orcid.org/0000-0003-2692-241X>

Husky oil operations limited as operator. (2000). “Appendix E: Assessment of Iceberg Impact Frequencies, White Rose Development Application.” (http://www.huskyenergy.com/downloads/areasofoperations/eastcoast/DevelopmentApplication/Vol5_AppendixE.pdf).

Hutchinson, T. P., and Lai, C. D. (1990). *Continuous bivariate distributions emphasising Applications*. Adelaide: Rumsby Scientific.

International Ice Patrol. (1995, updated 2018). “International Ice Patrol (IIP), Iceberg Sighting Database, Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center”. doi: <https://doi.org/10.7265/N56Q1V5R>.

IPCC. (2014). *Climate change 2014: synthesis report, contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change*. IPCC. Hdl:10013/epic.45156.d001

Johannessen, K., Meling, T. S., and Haver, S. (2001). Joint distribution for wind and waves in the northern North Sea. ISOPE International Ocean and Polar Engineering Conference. ISOPE-I-01-239 Norway: ISOPE.

Kang, J. S., Hu, X., and Li, C. (2019). Evaluation of return period and risk in bivariate non-stationary flood frequency analysis. *Water*. 11.1.79. <https://doi.org/10.3390/w11010079>

Kang, J., Sun, L., Sun, H., and Wu, C. (2017). Risk assessment of floating offshore wind turbine based on correlation-FMEA. *Ocean Engineering*. 129. 382-388. <https://doi.org/10.1016/j.oceaneng.2016.11.048>

Katz, R.W. (2013). Statistical methods for nonstationary extremes. *Extremes in a changing climate: Detection, analysis and uncertainty*. P15-37. Dordrecht: Springer.

Katz, W and Parlange, B and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*. 25. 8-12. 1287-1304.

[https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8)

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, JSTOR*. 30 (1/2): 81-93.

<https://doi.org/10.2307/2332226>

Kennedy, T. (2014). Protecting offshore operations from ice hazards, new ice management system helps operators keep Arctic offshore drilling and production facilities safe. *Heart Energy, E&P*.

<https://www.hartenergy.com/exclusives/protecting-offshore-operations-ice-hazards-20073>

King, T., R. Wright, K. Drover, G. Fleming, E. Gillis. (2015). Offshore Newfoundland and Labrador Metocean Study. OTC Arctic Technology Conference. OTC-25495. OTC.

<https://doi.org/10.4043/25495-MS>

Lee, S. (2007). *Structural equation modeling: A Bayesian approach*, volume 711. England: John Wiley & Sons, Ltd.

Leo, F., Besio, G., Briganti, R., and Vanem, E. (2021). Non-stationary extreme value analysis of sea states based on linear trends. Analysis of annual maxima series of significant wave height and peak period in the Mediterranean Sea. *Coastal Engineering*. 167. 103896.

<https://doi.org/10.1016/j.coastaleng.2021.103896>

Levine, D. (2009). Modeling tail behavior with extreme value theory, *Risk Management*: 17, 14-18.

- Lin, Y., and Dong, S. (2019). Wave energy assessment based on trivariate distribution of significant wave height, mean period and direction. *Applied Ocean Research*. 87: 47-63. <https://doi.org/10.1016/j.apor.2019.03.017>
- Lin, Y., Dong, S., and Tao, S. (2020). Modelling long-term joint distribution of significant wave height and mean zero-crossing wave period using a copula mixture. *Ocean Engineering*. 197. 106856. <https://doi.org/10.1016/j.oceaneng.2019.106856>
- Liu, X. D., Pan, F., Cai, W., and Peng, R. (2020). Correlation and risk measurement modeling: A Markov-switching mixed Clayton copula approach. *Reliability Engineering and System Safety*. 197. 106808. <https://doi.org/10.1016/j.res.2020.106808>
- Liu, Y., Chen, D.Y., Li, S.W., and Chan, P. W. (2018). Revised power-law model to estimate the vertical variations of extreme wind speeds in China coastal regions. *Journal of Wind Engineering and Industrial Aerodynamics*, 173, 227-240. <https://doi.org/10.1016/j.jweia.2017.12.002>
- Lombardo, F. T., and Ayyub, B. M. (2014). Analysis of Washington, DC, wind and temperature extremes with examination of climate change for engineering applications. *ASCE-ASME J. Risk Uncertainty Eng. Syst. Part A: Civ. Eng.* 1 (1): 04014005. <https://doi.org/10.1061/AJRU6.0000812>.
- Mackay, E. Jonathan, P. (2020). Assessment of return value estimates from stationary and non-stationary extreme value models. *Ocean Engineering* 207. 107406. <https://doi.org/10.1016/j.oceaneng.2020.107406>
- Mackey, D., Hogg, D., Lang, D., and Goodman, J. (2013). *emcee: the MCMC hammer*. *Publications of the Astronomical Society of the Pacific*. 125. 925. 306. DOI 10.1086/670067

[MANICE] Manual of Ice. (2005). Canadian Ice Service – Environment Canada. Manual of Standard Procedures for Observing and Reporting Ice Conditions. Environment Canada. (http://globalcryospherewatch.org/bestpractices/docs/MANICE_2005.pdf).

Manocha, N., and Babovic, V. (2018). Sequencing Infrastructure Investments under Deep Uncertainty Using Real Options Analysis. *Water*. 10.2.229. <https://doi.org/10.3390/w10020229>

Martins, E. S., and Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*. 36(3): 737-744. <https://doi.org/10.1029/1999WR900330>

Manuel, L., Nguyen, P., Canning, J., Coe, R., Eckert-Gallup, A., and Martin, N. (2018). Alternative approaches to develop environmental contours from metocean data. *Journal of Ocean Engineering and Marine Energy*. 4: 293-310.

McPhillips, L. E., Chang, H., Chester, M. V., Depietri, Y., Friedman, E., Grimm, N. B., Kominoski, J. S., McPhearson, T., Mendez-Lazaro, P., Rosi, E. J and others. (2018). Defining extreme events: A cross-disciplinary review. *Earth's Future*. 6.3: 441-455.

Moens, D., and Vandepitte, D. 2006. Recent advances in non-probabilistic approaches for non-deterministic dynamic finite element analysis. *Archives of Computational Methods in Engineering*. 13: 389-464.

Myriam, C-G., and Pascal, L. (2013). Extreme value analysis: an introduction. *Journal de la Société Française de Statistique*, 154(2): 66-97. <http://www.sfds.asso.fr/journal>

Nair, J., Wierman, A. and Zwart, B. (2013). The fundamentals of heavy-tails: properties, emergence and identification. *Proceedings of the ACM SIGMETRICS/international conference on Measurement and modeling of computer systems*. 17 June 2013. 387-388.

Neufville, R. D. and Scholtes, S. (2011). *Flexibility in engineering design*. England: MIT Press.

[NOAA] National Oceanic and Atmospheric Administration. (2019). <https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level>

Ogunbode, C. A., Doran, R. and Bohm, G. (2020). Exposure to the IPCC special report on 1.5 C global warming is linked to perceived threat and increased concern about climate change. *Climatic Change*.158 (3): 361-375.

Paik, J. K., and Thayamballi, A. K. (2007). *Ship-shaped offshore installations: design, building*, Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511546082.010>.

Panagoulia, D., Economou, P., and Caroni, C. (2014). Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. *25. 1. 29-43*.

Paola, D. F., Giugni, M., Pugliese, F., Annis, A., and Nardi, F. (2018). GEV parameter estimation and stationary vs. non-stationary analysis of extreme rainfall in African test cities. *Hydrology. 5. (2): 28*.

Petrov, V., Lucas, C., and Soares, C. G. (2019). Maximum entropy estimates of extreme significant wave heights from satellite altimeter data. *Ocean Engineering*,187. 106205. <https://doi.org/10.1016/j.oceaneng.2019.106205>

Pham, H., Nur, D., Pham, H. T.T., and Branford, A. (2019). A Bayesian approach for parameter estimation in multi-stage models. *Communications in Statistics-Theory and Methods. 48. (10): 2459-2482*.

Pickands, III. J. (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics. 3 (1): 119-131*.

Pike, H., Khan, F., and Amyotte, P. (2000). Precautionary Principle (PP) versus As Low As Reasonably Practicable (ALARP): Which one to use and when. *Process Safety and Environmental Protection*. 548 (137): 158-168. <https://doi.org/10.1016/j.psep.2020.02.026>

Politis, D. N., Romano, P.R., and Wolf, M. (1999). *Subsampling*. USA: Springer Science and Business Media.

Clauset, A., Rohilla, S. C., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM review*. (51) 4, 661-703.

Pryor, S., and Barthelmie, R. (2021). A global assessment of extreme wind speeds for wind energy applications. *Nature Energy*. 6(3): 268-276.

Rahman, M. S., Colbourne, B., and Khan, F. (2020). Conceptual development of an offshore resource centre in support of remote harsh environment operations. *Ocean Engineering*. 203. 107236. <https://doi.org/10.1016/j.oceaneng.2020.107236>

Ramadhani, A. and Khan, F., Colbourne, B., Ahmed, S., and Taleb-Berrouane, M. (2021). Environmental load estimation for offshore structures considering parametric dependencies. *Safety in Extreme Environments*. 3: 75-101.

Rathnayaka, S., Khan, F., and Amayotte, P. (2013). Accident modeling and risk assessment framework for safety critical decision-making: application to deepwater drilling operation. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of risk and reliability*. 227 (1): 86-105.

Resnick, S. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. USA: Springer Science and Business Media.

Resnick, S., and Starica, C. (1997). Smoothing the Hill estimator, *Advances in Applied Probability*. 29 (1): 271-293.

Rij, V. J., Yu, Y., McCall, A., and Coe, R. G. (2019). Extreme load computational fluid dynamics analysis and verification for a multibody wave energy converter. International Conference on Offshore Mechanics and Arctic Engineering. June 9, 2019. American Society of Mechanical Engineers. (58899): V010T09A042.

Rocchetta, R., Li, Y., and Zio, E. (2015). Risk assessment and risk-cost optimization of distributed power generation systems considering extreme weather conditions. Reliability Engineering and System Safety. 136: 47-61. <https://doi.org/10.1016/j.ress.2014.11.013>

Rodriguez, J. C. (2007). Measuring financial contagion: A copula approach. Journal of empirical finance. 14 (3): 401-423. <https://doi.org/10.1016/j.jempfin.2006.07.002>

Ross, E., Astrup, O. C., Gregersen, E. B., Bunn, N., Feld, G., Gouldby, B., Huseby, A., Liu, Y., Randell, D., Vanem, E., and Jonathan, P. (2020). On environmental contours for marine and coastal design. Ocean Engineering. 195. 106194. <https://doi.org/10.1016/j.oceaneng.2019.106194>

RP2A-WSD. (2000). Recommended practice for planning, designing and constructing fixed offshore platforms-working stress design. Washington DC: American Petroleum Institute.

Rubino, G. and Tuffin, B. (2009). Rare event simulation using Monte Carlo methods. UK: John Wiley and Sons Ltd. ISBN:9780470745403

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. REVSTAT-Statistical Journal. 10 (1): 33-60.

Scarrott, C., and Macdonald, A. (2010). Extreme-value-model-based risk assessment for nuclear reactors. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability. 224 (4): 239-252.

Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*. Bernoulli Society for Mathematical Statistics and Probability. 18 (3): 764-782. DOI: 10.3150/11-BEJ387

Shiau, J. T. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic environmental research and risk assessment*. 17 (1-2): 42-57.

Silva, D. and Simonovic, S. (2020). Development of Non-Stationary Rainfall Intensity Duration Frequency Curves for Future Climate Conditions. *Water Resources Research Report*. 106.

Sklar, A. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8: 229-231.

Staid, A., Pinson, P., and Guikema, S. D. (2015). Probabilistic maximum-value wind prediction for offshore environments. *Wind Energy*. 18(10): 1725-1738. DOI: 10.1002/we.1787

Sulistiyono, H., Khan, F., Lye L., and Yang, M. (2015). A risk based approach to developing design temperatures for vessels operating in low temperature environment. *Ocean. Eng.* 108: 813–819. <https://doi.org/10.1016/j.oceaneng.2015.08.040>

Swail, V. R, Cardone, V. J., Ferguson, M., Gummer, D. J., Harris, E. L., Orelup, E. A. and Cox, A. T. (2006). The MSC50 wind and wave reanalysis. *Proceedings of the 9th International Workshop on Wave Hindcasting and Forecasting*. Victoria. BC. Canada. 25. 29.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*. 2. USA: Random House.

Tawn, J. A. (1988). *Bivariate extreme value theory: models and estimation*. *Biometrika*. Oxford University Press. 75 (3): 397-415. <https://doi.org/10.1093/biomet/75.3.397>

Tixier, J., Dusserre, G., Salvi, O., and Gaston, D. (2002). Review of 62 risk analysis methodologies of industrial plants. *Journal of Loss Prevention in the process industries*. 15 (4): 291-303.

Toshkova, D., Asher, M., Hutchinson, P., and Lieven, N. (2020). Automatic alarm setup using extreme value theory. *Mechanical Systems and Signal Processing*. 139. 106417. <https://doi.org/10.1016/j.ymssp.2019.106417>

Tukey, J. W. (1977). *Exploratory data analysis*. Reading. 2: 131-160

Vanem, E. (2015). Uncertainties in extreme value modelling of wave data in a climate change perspective, *Journal of Ocean Engineering and Marine Energy*. 1 (4): 339-359.

Vanem, E. (2016). Joint statistical models for significant wave height and wave period in a changing climate. *Marine Structures*. 49: 180-205. <https://doi.org/10.1016/j.marstruc.2016.06.001>

Vanem, E. (2020). Bivariate regional extreme value analysis for significant wave height and wave period. *Applied Ocean Research*. 101. 102266. <https://doi.org/10.1016/j.apor.2020.102266>

Wang, L., and Li, J. (2016). Estimation of extreme wind speed in SCS and NWP by a non-stationary model. *Theoretical and Applied Mechanics Letters*. 6 (3): 131-138. <https://doi.org/10.1016/j.taml.2016.04.001>

Wang, X., Jiang, D., and Lang, X. (2017). Future extreme climate changes linked to global warming intensity. *Science Bulletin*. 62 (24): 1673—1680. <https://doi.org/10.1016/j.scib.2017.11.004>

Yang, M., Khan, F., Lye, L., and Amyotte, P. (2015). Risk assessment of rare events. *Process Safety and Environmental Protection*. 98, 102-108. <https://doi.org/10.1016/j.psep.2015.07.004>

Young, I., and Ribal, A. (2019a), Ocean waves and winds are getting higher and stronger. University Melbourne. (<https://pursuit.unimelb.edu.au/articles/ocean-waves-and-winds-are-getting-higher-and-stronger>).

Young I. R., and Ribal, A. (2019b). Multiplatform evaluation of global trends in wind speed and wave height. *Science*. American Association for the Advancement of Science. 364 (6440): 548-552. DOI: 10.1126/science.aav9527

Zar, J. H. (2005). Spearman rank correlation. *Encyclopedia of Biostatistics*. <https://doi.org/10.1002/0470011815.b2a15150>

Zhang, X., S., and Babovic, V. (2011). A real options approach to the design and architecture of water supply systems using innovative water technologies under uncertainty. *J. Hydro inf.* 14 (1): 13–29. <https://doi.org/10.2166/hydro.2011.078>

Zhang, Y., Kim, C.W., Beer, M., Dai, H., and Soares, C. G. (2018). Modeling multivariate ocean data using asymmetric copulas. *Coastal Engineering*. 135: 91-111. <https://doi.org/10.1016/j.coastaleng.2018.01.008>

Zhen, X., Vinnem, J. E., Han, Y., Peng, C. Yang, X., and Huang, Y. (2020a). New risk control mechanism for innovative deep water artificial seabed system through online risk monitoring system, *Applied Ocean Research*. 95:1-9. doi.org/10.1016/j.apor.2020.102054.

Zhen, X., Vinnem, J. E., Yang, X. and Huang, Y. (2020b). Quantitative risk modelling in the offshore petroleum industry: integration of human and organizational factors. *Ships and Offshore Structure*. 15 (1):1-18. doi.org/10.1080/17445302.2019.1589772.