Probabilistic cyclone damage assessment on large spatially distributed civil infrastructure systems

by

Yu Liang

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

School of Civil Engineering
Faculty of Engineering
The University of Sydney
November 03, 2025

CERTIFICATE OF ORIGINALITY

As the author of this thesis, I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at The University of Sydney or any other educational institution, except where due acknowledgment is made in the thesis.

I also declare that the thesis has been written by me. Any help that I have received in my research work and the presentation of the thesis itself has been acknowledged. I clarify that all information sources and literature used are indicated in the thesis.

Student name: Yu Liang

GENERATIVE AI USAGE STATEMENT

During the preparation of this thesis, ChatGPT was used for the purposes of text enhancement, including translation, paraphrasing, and grammar checking. Where any text was modified by generative AI, the author then reviewed the resulting content for any errors, inaccuracies or biases, and modified it as required. The author takes full responsibility for the submitted thesis and ensures the work is their own and has used generative AI within the parameters of use, see University of Sydney generative AI guide for researchers.

Student name: Yu Liang

This thesis is dedicated to my parents,

Taizhong Liang and Junxia Wu.

ACKNOWLEDGEMENTS

Completing this doctoral thesis marks one of the most significant milestones in my life. From the moment I decided to embark on a PhD journey, I understood that earning this degree was not solely about the title itself but about the profound and transformative experience it represents. Each step forward, every challenge overcome, and each moment of sudden insight has enriched the meaning of this demanding yet rewarding endeavour. Although the journey was often arduous, it was also filled with unexpected moments of joy.

I am deeply grateful to my homeland. The safety, stability, and prosperity of my country have enabled me to pursue my aspirations freely and with confidence. Even while studying far from home, I have always felt its warmth and support, whether it was sharing mooncakes during the Mid-Autumn Festival, eating zongzi at the Dragon Boat Festival, pasting Spring Festival couplets at Chinese New Year, or simply receiving timely fraud-prevention reminders from the consulate. These small yet meaningful gestures kept me closely connected to my roots, regardless of the distance.

My heartfelt thanks also go to the University of Sydney. The unique balance of a relaxed yet dignified atmosphere enabled me to maintain equilibrium between academic pursuits and personal well-being. I will always treasure the memories of lying on the campus lawns, allowing the fatigue of research to fade under the open sky. I have come to deeply appreciate the university's historic buildings, expansive green spaces, spectacular jacaranda trees, and the enchanting and slightly mysterious chime of the Quadrangle clock tower. Above all, I value the friendships

I have built here, which made this experience truly meaningful and unforgettable.

I owe my deepest gratitude to the person who has been the most important throughout my PhD journey, Professor Hao Zhang. Over the past three and a half years, Professor Zhang patiently guided me in learning how to conduct rigorous scientific research. This journey would not have been possible without his generosity in sharing his time, expertise, and unwavering support. Through countless cycles of drafting, revising, and starting anew, I came to realise that every decision in research must be grounded in sound reasoning and logical rigour, otherwise it will inevitably be challenged and require reworking. This iterative process profoundly shaped the way I think critically and express my ideas persuasively. As Professor Zhang often reminded me, the PhD journey resembles a marathon. There were moments when I felt eager to sprint ahead, only to become exhausted and tempted to stop. At such times, his advice helped me to slow down, focus on building a solid foundation, and then accelerate when the moment was right. His guidance enabled me to maintain a steady and sustainable pace and ultimately to achieve my goals.

I am also sincerely grateful to my senior colleagues for their generous support and guidance. When I felt lost in my research, Dr Cao Wang was always willing to help me analyse problems and untangle my thoughts. I am equally thankful to Dr Diqi Zeng for patiently addressing the doubts and uncertainties I encountered along the way.

My deepest appreciation goes to my family, including my father, mother, and sister, for their unwavering encouragement throughout this journey. I will never forget the words my mother shared with me before she passed away, encouraging me to persevere in the face of adversity. Her memory has been a constant source of strength during the most difficult times. This thesis represents my growth at

the age of twenty-seven. Just as I did when I was a young boy, I now dedicate this achievement to my mother in the hope that it would make her proud.

I am profoundly thankful to my close friend Yangyang Chen, who accompanied me through every success, frustration, and moment of joy, regardless of time or distance. I am also deeply grateful to Haolong, Frank, and Han, who have been wonderful companions on this journey. We studied together, faced challenges side by side, and shared countless moments of laughter. With all of you, the burdens felt lighter and the journey far more meaningful.

This research was made possible through the support of the China Scholarship Council Postgraduate Research Scholarship and the China Scholarship Council Supplementary Scholarship provided by the Faculty of Engineering at the University of Sydney, Australia. I am deeply grateful to both my home country and the University of Sydney for their essential support.

Finally, I would like to express my sincerest thanks to everyone who, in one way or another, has been part of this journey. Your guidance, encouragement, and companionship made this achievement possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Tropical cyclones (TCs) cause billions of dollars in economic losses worldwide each year. Accurately assessing their impacts is therefore essential for guiding disaster response and supporting long-term risk mitigation strategies. However, because TCs evolve dynamically over both space and time and generate widespread impacts, current studies often lack efficient and precise methodologies for evaluating TC-induced damage at the regional scale. In addition, climate change has driven substantial global warming in recent decades, which is expected to significantly influence TC behaviour and associated risks.

This thesis addresses these challenges through three studies. The first study focuses on developing an efficient framework for assessing TC-induced losses for regional buildings. This study establishes relationships between wind speeds and total loss across the region. When the study area is large, spatial heterogeneity, e.g. terrain variability, can hinder accurate modelling of the regional hazard-damage relationship. To address this challenge, the present study employs a spatial clustering algorithm to divide the entire area into multiple sub-regions with relatively homogeneous internal characteristics. For each sub-region, a TC loss model is developed, defined as a function of wind speed at the sub-regional centroid and the corresponding building portfolio loss ratio. In application, the loss in all sub-regions is first assessed individually and then aggregated to estimate the total regional loss. The second study introduces a Bayesian network (BN)-based TC loss model for estimating the spatial distribution of regional building losses. The

model takes key TC parameters as inputs and outputs the loss ratio for each subdivision in the study area. Given the significant impact of discretising continuous variables on model accuracy, this study introduces novel discretisation schemes for both TC and loss variables. The advantage of the model is its ability to efficiently assess the spatial distribution of regional losses while accounting for uncertainties and spatial correlations in the estimation process. The third study investigates the potential impacts of climate change on future TC activity. This study proposes a framework for quantitatively assessing the impacts of climate change on TC frequency, genesis location, and intensity using high-resolution global climate models. The projected TC characteristics are then used as inputs to a synthetic TC model to estimate changes in coastal risk under future climate conditions. The proposed framework is applied to the North Atlantic under the high-emission scenario SSP585. The individual and combined effects of the modified TC characteristics are analysed to evaluate their respective roles in shaping future coastal risk profiles.

CHAPTER 1

INTRODUCTION

1.1 Statement of the problem

Tropical cyclones (TCs), commonly referred to as hurricanes or typhoons, are among the most destructive natural hazards, causing substantial economic losses, widespread infrastructure damage, and large-scale population displacement. These events are often accompanied by extreme precipitation and severe flooding, which exacerbate structural failures in residential buildings and disrupt critical lifeline systems such as power supply, water distribution, and transportation networks (Ellingwood et al., 2004; Lin et al., 2012). Recent examples, including Hurricane Katrina (2005), Hurricane Helene (2024), and Hurricane Milton (2024), have each caused tens of billions of dollars in damages across the United States (Florida Department of Financial Services, 2006; Smith, 2025). In Australia, the average annual insured losses attributed to TCs are estimated to be approximately 0.7 billion Australian dollars (Andrews et al., 2016). Table 1.1 and Table 1.2 (Smith, 2025) present the ten most catastrophic North Atlantic TCs since the 20th century, ranked by fatalities and direct economic losses, respectively. For instance, Hurricane Mitch (1998) resulted in over 11,000 deaths, while Hurricanes Katrina (2005) and Harvey (2017) each generated more than $125 billion in losses.

Urban infrastructure systems, including essential facilities such as hospitals, schools, and commercial and residential buildings, as well as transportation, power,

Table 1.1: Top 10 deadliest North Atlantic TCs since 20th century (Wikipedia, 2025a).

Rank	TC	Season	Category	Deaths
1	Mitch	1998	5	11,374
2	Fifi-Orlene	1974	2	8,210
3	Flora	1963	4	7,193
4	Okeechobee	1928	5	4,075
5	Monterrey	1909	3	4,000
6	Cuba	1932	5	3,142
7	Maria	2017	5	3,059
8	Jeanne	2004	3	3,037
9	Jérémie	1935	1	2,150
10	David	1979	5	2,068

Table 1.2: Top 10 costliest North Atlantic TCs (Wikipedia, 2025b).

Rank	TC	Season	Category	Costs (Billion USD) *
1	KATRINA	2005	5	$125.0
2	Harvey	2017	4	$125.0
3	Ian	2022	5	$113.1
4	Maria	2017	5	$91.6
5	Helene	2024	4	$78.7
6	Irma	2017	5	$77.2
7	Ida	2021	4	$75.3
8	Sandy	2012	3	$68.7
9	Ike	2008	4	$38.0
10	Milton	2024	5	$34.3

* Not adjusted for inflation

water, and communication networks, are highly vulnerable to TC impacts. Variations in structural design, construction materials, and maintenance levels create significant disparities in individual buildings' capacities to withstand extreme loads (Lee and Ellingwood, 2017). Moreover, these systems are interdependent, meaning that the failure of a single component can trigger cascading effects. For example, power outages during a TC can interrupt water supply systems, hinder emergency responses, and severely limit healthcare services. Despite these interconnected vulnerabilities, current design codes primarily focus on the safety of individual structures and do not adequately address the resilience of communities as integrated systems. The persistently high economic losses observed in recent decades indicate that concentrating on individual buildings alone is insufficient. A broader, systemic approach to disaster planning is needed — one that encompasses the entire built environment and emphasizes community-level resilience (McAllister, 2016; Koliou et al., 2020).

Community resilience is commonly defined as the capacity of both physical and social systems to recover to their pre-disaster conditions within an acceptable timeframe following a hazard event (Timmerman, 1981). This perspective shifts the focus from isolated structures to spatially distributed networks and emphasizes collective recovery. The evaluation of system-wide damage is typically conducted using a probabilistic risk assessment framework. This framework not only accounts for uncertainties arising from the inherent randomness of natural hazards and limited knowledge but also incorporates spatial correlations, such as those related to wind speed and structural capacity. However, few studies have comprehensively integrated these uncertainties and correlations into TC risk assessments. Moreover, while probabilistic risk analysis provides a well-established framework, it often demands substantial computational resources, especially for

large urban areas. The dynamic nature of TCs, which requires continuous monitoring of their temporal and spatial evolution, further increases computational burdens and limits practical application in engineering practice.

Furthermore, there is growing evidence that climate change is altering atmospheric and oceanic conditions, which in turn affect the frequency, intensity, and genesis locations of TCs (Rosowsky, 2021; Pant and Cha, 2019). These changes may shift the spatial distribution of TC risk, potentially exposing previously low-risk regions to severe hazards. Intensifying storms may also exceed existing design thresholds for buildings and lifeline systems. Consequently, understanding the evolving nature of TC hazards under climate change and accurately assessing their implications for urban infrastructure resilience have become essential for informed, long-term planning and risk mitigation.

1.2 Research objectives and scope

The primary aim of this thesis is to develop probabilistic models for assessing TC-induced losses, enabling accurate and efficient evaluations of building damage following TC events. In addition, the study seeks to investigate how TC-related risks may evolve in coastal cities under future climate scenarios. To achieve these goals, the research is structured around three specific objectives:

1. Formulate a probabilistic loss model for building portfolios. This model serves a role analogous to a "fragility curve" but at the regional level. By utilising wind speed data from one or more determined locations, it enables rapid estimation of expected building loss across the whole region. Such an approach provides a practical and efficient tool for post-event loss assessment.

2. Develop a BN-based framework for regional loss estimation. The framework

is designed to efficiently capture the spatial distribution of losses across a study area while explicitly accounting for uncertainties and spatial correlations. It takes key TC parameters, including translation speed, heading angle, storm centre location, and central pressure, as inputs and outputs the loss ratio for each sub-region. This approach allows for a more detailed understanding of how TC characteristics translate into localized impacts.

3. Assess the potential effects of climate change on TC behaviour and urban risk. Based on a high-resolution general circulation model (GCM), the study explores how climate change may influence TC frequency, genesis locations, and intensity. The individual and combined effects of these factors are evaluated to quantify how future climatic conditions could alter the magnitude and spatial distribution of TC-induced risks. The findings aim to inform long-term urban planning and climate adaptation strategies.

1.3 Organization of the thesis

This thesis is organized into six chapters, each addressing a key component of the research.

Chapter 2: literature review. This chapter provides a comprehensive review of the current state of knowledge in several areas, including TC wind field modelling, synthetic TC generation techniques, damage assessment methodologies for individual structures and infrastructure systems, applications of BNs in risk modelling, and the effects of climate change on TC behaviour and urban vulnerabilities.

Chapter 3: probabilistic loss modelling for building portfolios. This chapter introduces a probabilistic model for estimating TC-induced losses across large building portfolios. To address the heterogeneity of extensive urban areas, a divide-and-aggregate approach is proposed. In this method, the study area is

divided into multiple sub-regions, losses are assessed individually for each sub-region, and the results are aggregated to estimate overall regional losses. This approach improves computational efficiency while maintaining accuracy.

Chapter 4: BN-based TC loss model construction. Building upon the methodology developed in Chapter 3, this chapter presents a BN-based framework for estimating the spatial distribution of regional building losses. It also introduces novel discretisation schemes for both TC parameters and loss variables, balancing the model's accuracy and the network's complexity.

Chapter 5: climate change impacts on TC-induced risks. This chapter investigates how evolving climate conditions may alter TC characteristics and their associated risks. Combing a high-resolution GCM and synthetic TC simulation techniques, it evaluates how these changes affect the spatial distribution of risk in coastal cities.

Chapter 6: conclusions and future work. The final chapter summarises the key findings and contributions of the research, discusses their practical implications, and outlines potential directions for future work in improving TC risk modelling and resilience planning.

1.4 Publications

Some chapters of this thesis are based on the author's manuscripts published/submitted during his PhD candidature, as outlined below:

1. A probabilistic framework to construct tropical cyclone loss models for building portfolios (based on Chapter 3), published in Structural Safety, 119, 102665. Authors: Yu Liang, Hao Zhang, Cao Wang, Diqi Zeng.

2. A Bayesian network framework for probabilistic assessment of regional building losses from hurricanes (based on Chapter 4), submitted to Reliability

Engineering and System Safety (under review). Authors: Yu Liang, Hao Zhang, Cao Wang.

3. A global climate model–based framework for assessing climate change impacts on hurricane wind risk in coastal cities (based on Chapter 5), in preparation for submission. Authors: Yu Liang, Hao Zhang, Cao Wang.

The author was the primary contributor to all manuscripts listed above. Permission to include these materials in the thesis has been granted by the corresponding authors.

CHAPTER 2

ADVANCES IN TROPICAL CYCLONE RISK ASSESSMENT FOR LARGE-SCALE INFRASTRUCTURE SYSTEMS

2.1 Review of previous studies

This chapter presents a comprehensive and systematic review of TC risk assessment for urban infrastructure systems. It includes key aspects of TC wind field modelling, synthetic TC simulation, wind fragility and risk assessment, rapid post-TC damage assessment techniques, and the projected impacts of climate change on TC activity. The objective of this chapter is to offer a state-of-the-art overview of current methodologies and advances in TC risk assessment for large-scale infrastructure systems.

2.1.1 Concept of community resilience

In recent decades, the economic losses caused by extreme natural disasters have reached unprecedented levels. Between 1980 and 2024, weather- and climate-related disasters in the United States resulted in an average annual economic loss of $64.8 billion (adjusted for the Consumer Price Index). Notably, the cumulative losses over the three-year period from 2022 to 2024 alone amounted to $461.6 billion (Smith, 2025). This escalating financial burden has heightened awareness among scholars and policymakers of the urgent need for cities to endure, adapt to, and rapidly recover from disasters. This concept, widely known as community

resilience, has therefore gained increasing attention worldwide.

Governments and international organizations have launched several initiatives to enhance urban resilience. For instance, Australia's National Strategy for Disaster Resilience promotes a whole-of-community approach to disaster management, emphasising shared responsibilities among individuals, businesses, and government entities (Wilkins and McCarthy, 2011). In the United States, Presidential Policy Directives outline strategies to strengthen the security and resilience of critical infrastructure against both physical and cyber threats, calling for coordinated actions across government agencies, private-sector stakeholders, and international partners (U.S. Department of Energy, 2013). At the global level, the United Nations (United Nations Office for Disaster Risk Reduction, 2025) and the Rockefeller Foundation's 100 Resilient Cities initiative (Foundation, 2025) have recognized resilience as a multidimensional strategy for achieving sustainable urban development.

The concept of resilience originated in ecology, where Holling (1973) defined it as the ability of an ecosystem to absorb disturbances and maintain its essential functions despite external shocks. Timmerman (1981) later extended this notion to the context of natural hazards, emphasizing how systems recover from extreme disruptions. Over time, the scope of resilience has expanded to encompass social, economic, and infrastructure systems. Today, resilience is understood not merely as a return to pre-disaster conditions but also as the capacity to learn, adapt, and transform in response to future risks (Folke et al., 2002; Bruneau et al., 2003).

To systematically evaluate community resilience, Bruneau et al. (2003) identified four key attributes that characterise a resilient system facing extreme events:

1. Robustness – the capacity of a system to withstand the impacts of extreme hazards without significant degradation of core functions.

2. Rapidity – the speed with which a community can restore essential services and return to normal operations.

3. Redundancy – the availability of alternative or backup components that allow critical functions to continue when primary systems fail.

4. Resourcefulness – the ability to identify problems, mobilise personnel, and allocate financial and material resources effectively during and after a disaster.

Building on these attributes, Bruneau et al. (2003) further classified community resilience into four interrelated dimensions: technical, organizational, social, and economic. Technical resilience concerns the performance and recovery of physical systems, including buildings, transportation networks, and utilities. Organizational resilience refers to the ability of governmental and institutional entities to plan, coordinate, and implement effective disaster response strategies. Social resilience addresses the mobilization of community resources, support networks, and actions to reduce vulnerability and promote collective well-being. Economic resilience focuses on mitigating both direct and indirect financial losses to ensure sustainable recovery and long-term economic stability.

Considerable progress has been made in assessing these four dimensions of resilience. In the technical domain, performance-based engineering models are used to estimate functionality loss and recovery trajectories for buildings, roads, and utilities (Ellingwood et al., 2018; Rosowsky and Ellingwood, 2002; Li, 2005; Wang and Rosowsky, 2018). Resilience-based design frameworks integrate resilience objectives into building codes and infrastructure policies (Cimellaro et al., 2010), while optimization models help prioritize retrofit strategies for critical infrastructure (Zhang et al., 2022; Kong et al., 2021). Social resilience is often evaluated

through social vulnerability indices, which incorporate demographic and socioeconomic variables such as age, income, education, ethnicity, health status, and access to housing and healthcare services (Cox and Perry, 2011; Zhang and Nicholson, 2016; Contreras and M, 2019). Recent research has also highlighted the role of social capital, including trust, community cohesion, and informal support networks, in fostering collective coping and resource sharing during recovery (Aldrich, 2017; Kyne and Aldrich, 2020; Western et al., 2005). Organizational resilience is assessed using indicators of governance effectiveness, emergency preparedness, and policy coordination before, during, and after disasters (Col, 2007; Kapucu and Sadiq, 2016; Kalogiannidis et al., 2023). Metrics such as diversity and redundancy in governance structures, polycentric decision-making frameworks, and inclusive stakeholder participation are often employed to evaluate institutional readiness and responsiveness (Grafton et al., 2019; Freitag et al., 2014). For economic resilience, regional input-output models (He et al., 2019), computable general equilibrium models (Xie et al., 2014), and agent-based simulations (Jiang et al., 2024) are used to estimate direct and indirect economic impacts, including business closures, employment disruptions, and reductions in gross domestic product (GDP) (Zhang and Nicholson, 2016; Gao et al., 2020; Chang-Richards et al., 2019).

Despite these advances, several challenges remain in the community resilience assessment. Many current models are hazard-specific and do not address multi-hazard interactions (Koliou et al., 2020). Existing building codes primarily prioritize life safety without explicitly incorporating resilience-based performance goals (Ellingwood et al., 2018). Infrastructure assessments often overlook cascading failures, such as power outages disrupting water supply and emergency services (Ouyang and Wang, 2015). Furthermore, inconsistent definitions of resilience across disciplines contribute to discrepancies in assessment methods and policy

implementation. Addressing these gaps requires the development of standardized, multi-hazard frameworks, enhanced infrastructure design criteria, and equitable policies to promote comprehensive and sustainable resilience planning.

2.1.2 TC wind field modelling

Wind field models play a critical role in TC risk assessment by translating TC intensity into a spatially distributed wind speed field. Understanding the wind field structure is essential for estimating building loads and regional hazard exposure. In the horizontal dimension, the wind field is typically divided into three zones: the eye, the eyewall, and the outer diffusion region. Winds are relatively calm within the eye, reach their maximum intensity in the eyewall, and gradually decrease in strength toward the periphery (Zhao et al., 2024). In the vertical dimension, near-surface winds are governed by the planetary boundary layer (PBL), where frictional effects and turbulent mixing significantly influence wind speed profiles. The height of the PBL generally ranges from several hundred meters to over one kilometre, depending on TC intensity and environmental conditions. Above the PBL, the flow transitions to gradient wind, which is primarily driven by the balance of pressure gradient, Coriolis, and centrifugal forces (Vickery et al., 2009a).

Two primary types of wind field models are commonly used in engineering applications: parametric models and numerical models. Parametric models provide empirical or semi-theoretical representations of TC wind fields (Yan and Zhang, 2022; Zeng et al., 2021). These models typically describe radial and azimuthal wind speed distributions using mathematical formulations derived from observed pressure fields and gradient wind equations. Early parametric approaches assumed axisymmetric wind structures. For instance, Schloemer (1954) proposed one of the earliest models with this assumption, which was later refined by Holland (1980) through pressure-based formulations to estimate wind speeds near the TC centre

and across its radial extent. However, symmetric models proved inadequate for real-world TCs, which often exhibit asymmetries, such as stronger winds on the right side of a moving storm in the Northern Hemisphere. To address this limitation, Georgiou (1986) extended Holland's model by introducing a moving wind field component, resulting in a more realistic representation of asymmetric wind patterns.

Numerical models offer a more complex approach by directly solving the governing equations of atmospheric motion, including the Navier–Stokes, energy, and continuity equations (Meng et al., 1995; Kepert, 2001). These models explicitly simulate the interactions between the TC and its surrounding environment, capturing highly complex wind field structures. Early work by Chow (1971) introduced a two-dimensional model that incorporated surface drag effects to reproduce observed asymmetries. This model utilized a nested finite-difference grid to represent boundary layer processes. With advances in computational resources, Kepert (2001) developed a polar-coordinate formulation of the Navier–Stokes equations, enabling improved simulation of three-dimensional momentum conservation and vertical wind variations. More recently, Hong et al. (2019) proposed a nonlinear three-dimensional wind field model that employed an implicit finite difference scheme to enhance computational efficiency while maintaining physical accuracy.

In engineering and hazard assessments, parametric models are widely used because they require fewer data inputs and are computationally efficient, making them suitable for large-scale and time-sensitive applications (Yan and Zhang, 2022; Zeng et al., 2021; Lee and Ellingwood, 2017). Numerical models, while capable of simulating complex wind field behaviours and their interactions with land, ocean, and atmospheric processes, are often limited by data availability and computational demands. As a result, they are primarily used in research studies

and high-resolution forecasting.

2.1.3 TC track simulation

Historical records of TCs extend back to around 1850. However, reliable and consistent documentation of TC activity only became available after the advent of satellite observations in the 1980s (Harper, 2002; Bloemendaal et al., 2020). As a result, many coastal locations have few, if any, recorded events within the observational dataset. This scarcity of data limits the ability to accurately estimate long-term TC risks. Consequently, relying solely on historical records is insufficient for comprehensive hazard assessment. To overcome these limitations, synthetic TC modelling techniques have been widely adopted in engineering and risk analysis applications.

Early TC simulation models were designed for site-specific risk assessments (Russell, 1969; Jarvinen et al., 1984). These models employed the Monte Carlo method to generate probabilistic estimates of key storm parameters. The typical modelling process involved four steps: 1. Establish site-specific statistical distributions for essential parameters such as closest approach distance, translation speed, heading direction, central pressure, and the pressure profile parameter. 2. Use Monte Carlo simulation to randomly sample values for each parameter. 3. Simulate the storm moving along a straight-line track, assuming constant intensity until landfall. 4. Apply decay models to represent storm weakening after landfall. While effective for individual locations, this approach had two major limitations. First, in data-sparse regions, the parameter distributions could not be reliably estimated. Second, because these models only focused on individual sites, they failed to account for regional variability in TC characteristics and risks.

To address these limitations, researchers expanded TC simulations to encompass broader geographic regions. Vickery et al. (2000) introduced an empirical

track model that divided the study area, such as the North Atlantic basin, into grid cells and used regression techniques to estimate model parameters for each cell. This approach simulated the full TC lifecycle, including genesis, development, and dissipation. Emanuel et al. (2006) proposed a hybrid statistical–deterministic approach that incorporated beta-and-advection models, allowing environmental factors to influence storm track evolution. Vickery et al. (2009b) enhanced storm intensity modelling by integrating the effects of sea surface temperature (SST) and vertical wind shear, improving the realism of TC simulations. Bloemendaal et al. (2020) developed a fully statistical model that uses a resampling algorithm to simulate storm genesis, track movement, central pressure, and maximum wind radius. Cui et al. (2021) applied quadtree decomposition to partition ocean basins into variable-sized grids, enabling localized parameter estimation and improving spatial representation of TC dynamics. Zeng et al. (2024) built upon this concept by introducing a full-track model for the western North Pacific basin. The model employed autoregressive techniques to simulate both track evolution and storm intensity while explicitly accounting for spatial nonstationarity in regression parameters.

One of the most significant practical applications of synthetic TC models is their integration into structural design standards. For example, ASCE 7-16 (ASCE, 2017) classifies structures into four risk categories based on potential consequences of failure. Each category corresponds to a mean recurrence interval (MRI) ranging from 300 years for Category I structures to 3000 years for Category IV. Synthetic TC simulations are used to generate design wind speed maps for these MRIs, providing a scientific basis for performance-based design. Beyond design codes, synthetic TC modelling is also widely applied to regional risk assessment. For example, it is used for evaluating TC risk distribution across coastal

regions (Cui and Caracoglia, 2019; Lee and Rosowsky, 2007) in the United States. Huang et al. (2021) integrated a Markov environment-dependent intensity model with the Vickery approach (Vickery et al., 2000) to estimate typhoon risks for ten major Chinese coastal cities. Arthur (2021) developed a statistical-parametric loss model to perform hazard assessments at both regional and continental scales in Australia.

The evolution of TC track modelling reflects a shift from simplified, site-specific simulations to sophisticated, region-wide frameworks that capture the full lifecycle of TCs. These advancements have significantly improved the accuracy of risk assessments and have become integral to modern engineering design and disaster planning.

2.1.4 Structural fragility analysis

Fragility represents a probabilistic measure that quantifies the likelihood of a structural system reaching or exceeding a specified performance threshold under a given level of hazard intensity. It is commonly expressed as $P(LS \mid D = y)$, where D denotes the hazard intensity parameter, such as wind speed, peak ground acceleration, or spectral acceleration, depending on the objectives of the risk analysis (Ellingwood et al., 2004; Li, 2005); LS represents the limit state, which defines the boundary between acceptable and unacceptable structural performance. Limit states can be categorized into two main types. 1. Deformation-related limit states, such as excessive drift, detachment of cladding from framing, or failure at roof-to-wall connections; and 2. strength-related limit states, including roof or wall collapses, sheathing removal, and foundation failures (Li, 2005).

The HAZUS-MH program provides a standardized framework for classifying building damage into five levels: none, minor, moderate, severe, and destruction. As summarized in Table 2.1, these classifications are based on observable structural

damage, such as the percentage of roof cover failure, the number of damaged windows or doors, and the extent of roof deck and roof structural failures (FEMA, 2014). For instance: minor damage typically involves small areas of roof cover failure (2–15%) or limited window and door damage. Moderate damage reflects more extensive roof cover loss (15–50%) and partial roof deck failures. Severe and destruction levels correspond to widespread structural failures, including roof system collapse and significant wall damage. This standardized categorization allows for consistent assessments of structural vulnerability across different regions and building types.

Table 2.1: Damage state definition for residential construction classes in HAZUS-MH program (FEMA, 2014).

Damage state	Qualitative measure	Roof cover failure	window/door failure	Roof deck	Roof structure failure
0	No	$\leq 2\%$	No	No	No
1	Minor	$> 2\%$ and $\leq 15\%$	1	No	No
2	Moderate	$> 15\%$ and $\leq 50\%$	2 or 3	1-3	No
3	Severe	$> 50\%$	> 3 and $\leq 25\%$	> 3 and $\leq 25\%$	No
4	Destruction	Typically $> 50\%$	$> 25\%$	$> 25\%$	Yes

The development of fragility functions typically relies on structural reliability methods, such as the First-Order Reliability Method (FORM) or Monte Carlo Simulations (MCS) (Lee and Rosowsky, 2005; Ellingwood et al., 2004). Both aleatory and epistemic uncertainties are incorporated to represent the complex factors influencing structural performance. Aleatory uncertainty arises from the inherent variability in external loads and structural capacities. Examples include external and internal pressure coefficients, velocity pressure exposure factors, and material properties (Ellingwood and Tekie, 1999). Epistemic uncertainty stems from limited data availability for statistical parameters or simplifying assumptions made during modelling.

Fragility functions are commonly expressed using a lognormal cumulative distribution function (CDF) (Li, 2005; Ellingwood et al., 2007; Rosowsky and Ellingwood, 2002):

$$P(DS \geqslant v \mid U = u) = \Phi\left(\frac{\ln(u) - \lambda_v}{\xi_v}\right), \tag{2.1}$$

where $\Phi(\cdot)$ is the standard normal CDF, λ_v, ξ_v are the logarithmic mean and standard deviation of the fragility function for the damage state v. In Eq. (2.1), wind speed U is treated as a deterministic variable, while all other sources of uncertainty are captured by the parameters λ_v and ξ_v. These uncertainties primarily arise from factors such as the wind exposure factor, external and internal pressure coefficients, and wind directionality effects (Ellingwood and Tekie, 1999). Fig. 2.1 illustrates a fragility curve showing how the probability of exceeding each damage state increases with wind speed. The example presented corresponds to a representative two-story wood-frame house located in suburban terrain, featuring strapped roof trusses, a hip roof, no garage, and wood-frame walls (FEMA, 2014).

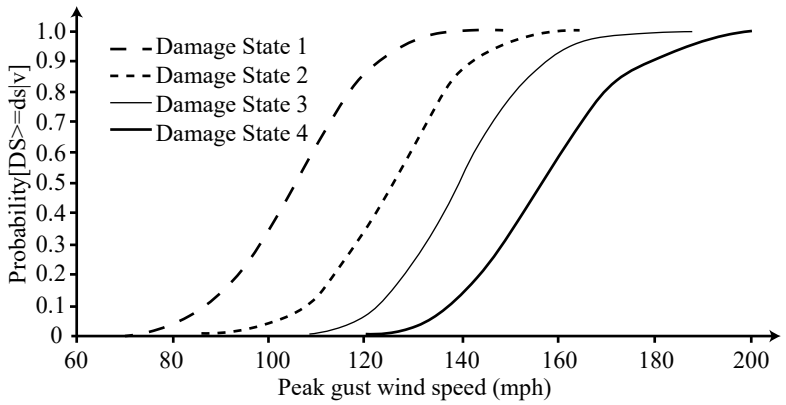


Figure 2.1: Fragility curve for a two-story wood-frame house with strapped trusses and hip roof in suburban terrain.

2.1.5 Risk assessment for distributed infrastructures

Distributed infrastructures in urban environments typically comprise two major categories: building portfolios and lifeline systems. Building portfolios encompass

collections of residential, commercial, and institutional structures dispersed across a region. These buildings are heterogeneous in construction type, occupancy, and spatial arrangement, and they play essential roles in providing shelter, supporting economic activity, and maintaining critical services within cities (Anwar et al., 2023). Lifeline systems include crucial utility and service networks, such as electricity, water supply, telecommunications, and transportation. These systems are spatially extensive, highly interconnected, and indispensable for maintaining urban functionality during and after extreme hazard events.

The risk assessment of building portfolios is typically evaluated along three primary dimensions. The first is technical dimension. This dimension focuses on physical damage metrics, such as the damage ratio and immediate occupancy ratio (Shapiro et al., 2000; Lin and Wang, 2016). For instance, the proportion of buildings exceeding a severe damage state (DS $\geq$ 3) indicates the extent of significant structural impairment (Shapiro et al., 2000). The second is economic dimension. Economic losses are commonly expressed as monetary values or the proportion of total damage. These losses are generally classified into structural and non-structural losses. Structural losses are converted into monetary terms using empirical damage-to-loss mappings. For example, slight, moderate, extensive, and complete damage are typically associated with 2–20%, 10–40%, 50–80%, and 100% of the building replacement cost, respectively (FEMA, 2014; Wang et al., 2017; Do et al., 2020). Non-structural losses involve interior partitions, ceilings, mechanical equipment, and façade elements. These losses are inherently more variable due to differences in materials and systems, resulting in higher uncertainty in estimation (Ahmad et al., 2024). The third is social dimension. Social impacts are commonly represented by indicators such as household dislocation, which reflects the percentage of households displaced due to loss of habitability

or the need for temporary shelter (FEMA, 2014). This indicator is derived by integrating building damage data, demographic information, and post-event surveys (Peacock et al., 2008; Lin, 2009). The number of displaced households can be estimated by identifying buildings that exceed habitability thresholds, multiplying by average household occupancy, and adjusting for shelter capacity and socio-economic vulnerability factors (Lin and Wang, 2016).

Substantial progress has been made in modelling building portfolio risks for various natural hazards, including earthquakes, TCs, floods, among others. Regarding earthquakes, Wang et al. (2020) extended life-cycle analysis to community-scale buildings, incorporating metrics such as life-cycle cost and cumulative prospect value to reflect both risk-neutral and risk-averse decision-making. Ding et al. (2025) utilized probabilistic machine learning models combined with Poisson binomial distributions to assess the probability of functionality failure for entire regional buildings. For TCs, Wang et al. (2023) simulated post-disaster recovery and optimized repair resource allocation under constraints such as deadlines and capacity limits. Nofal et al. (2021) proposed a multi-hazard regional risk framework addressing wind, storm surge, and wave damage to structures and contents. Similarly, for floods, Xie et al. (2025) developed a spatiotemporal forecasting approach that introduced an evacuation-oriented functionality metric integrating damage, utility disruptions, and transportation accessibility to support pre-disaster planning. Additional studies have also examined the impacts of other hazards, such as tornadoes (Memari et al., 2018) and tsunamis (Alhamid et al., 2023).

Lifeline systems are critical to urban resilience because their failure can disrupt multiple essential services. For power networks, a variety of modelling approaches are used to assess the outage risks, including connectivity-based analysis (Zeng et al., 2021); complex network theory (Dueñas-Osorio and Vemuru, 2009;

Ouyang, 2013); power flow modelling (Ouyang and Duenas-Osorio, 2014; Hu and Ma, 2023; Fan et al., 2023); and linear programming and network flow optimization techniques (Wang et al., 2016; Dwivedi and Yu, 2013; Zhang et al., 2018), among others. Specifically, Ouyang and Duenas-Osorio (2014) proposed a multi-model framework that integrates hazard characterization, component fragility, DC power flow analysis, and staged restoration modelling to assess TC-induced power system damage. Zhang et al. (2018) combined network flow theory with building functionality modelling to predict downstream impacts of electricity disruption on building operations. Wang et al. (2016) identified critical transmission lines by analysing the direction and magnitude of power flows to determine which failures would most severely reduce system capacity. Significant advances have also been made in other networks, including water systems (Vertommen et al., 2022; Yu et al., 2024), telecommunications (Liu et al., 2020b; Hagenmeyer et al., 2016), and transportation networks (Mohebbi et al., 2020; Wang et al., 2019).

Lifeline systems are inherently interdependent. For example, water systems depend on electricity for pumping and treatment; power grids require water for cooling operations; and both rely on communication networks for real-time monitoring and control. These interconnections mean that failure in one system can cascade into others, amplifying disruption and complicating recovery efforts. To better understand this connection, Ji et al. (2024) categorized interdependency modelling approaches into three main types: 1. topology-based models that apply graph theory to identify key connections and vulnerabilities; 2. probabilistic models, such as BNs, to represent uncertainty and conditional dependencies; 3. hybrid models that combine simulations, optimization, and data-driven techniques. In addition, Ouyang et al. (2009) simulated the interdependence between power

and oil supply networks using geographically co-located nodes to track cascading failures. Zhang et al. (2018) investigated how power outages disrupt water pumping and ultimately affect building operations. However, current models still rely on simplifying assumptions that limit their realism and applicability. Many models are unable to fully capture the complexity of real-world interactions, particularly under extreme multi-hazard conditions. Future research should focus on improving model fidelity, and advancing the understanding of interdependent behaviours and cascading failure mechanisms to support resilience-based planning and decision-making (Yu et al., 2024; Liu et al., 2022; De Felice et al., 2022).

2.1.6 Challenges in TC risk analysis

Risk assessment for regional infrastructure systems under TCs faces three key challenges: 1. the quantification of uncertainties in both hazard characterization and structural response, 2. the accurate representation of spatial correlations in hazard fields and building damages, and 3. the need for efficient yet reliable methods that scale to community-level evaluations.

Uncertainties are commonly classified into aleatory and epistemic components. Aleatory uncertainty reflects the intrinsic randomness of natural hazards and structural capacities and must be treated probabilistically. Epistemic uncertainty arises from incomplete knowledge in hazard modelling, parameter estimation, and structural response prediction, and can be reduced as data and models improve (Ellingwood et al., 2004, 2018). In TC risk analysis, key sources of uncertainty include track simulation and intensity evolution, wind-field computation, parameter estimation, and the fragility-based mapping from wind demand to damage. In fully probabilistic assessment, the modelling of TC movement and intensity contributes substantially to overall uncertainty (Vickery et al., 2000; Bloemendaal et al., 2020; Ullrich et al., 2021). Under scenario-based analyses, hurricane track

and intensity parameters are prescribed, and the dominant source of uncertainty typically arises from the wind-field model. Zeng et al. (2021) quantified the uncertainty of the Georgiou wind field model using 34 recorded wind fields from six historical TC events in the North Atlantic. The bias of the wind field model was defined as

$$\varepsilon = \frac{U}{U_0}, \tag{2.2}$$

where U is the observed wind speed and U_0 is the modelled wind speed. Their analysis showed that ε follows a lognormal distribution, with its parameters dependent on the selected wind field model, site location, and distance from the storm centre. Similar findings have been reported in other studies (Valamanesh et al., 2016; Arunachalam, 2024; Pang et al., 2012; Li and Ellingwood, 2006). For the Georgiou wind field model, Zeng et al. (2021) expressed the mean of $\ln \varepsilon$ as

$$\overline{\ln \varepsilon} = -0.0515 \frac{r}{R_{\max}} - 0.08321, \tag{2.3}$$

where r is the radial distance from the storm centre, and $R_{\max}$ is the radius of maximum wind. The logarithmic standard deviation of ε was found to be approximately constant at 0.12, as illustrated in Fig. 2.2.

Evaluating spatial correlation is a second challenge in community risk assessments. Because TCs have broad spatial footprints, nearby sites experience wind fields that are statistically dependent. This site-to-site correlation is commonly modelled as a function of the distance between two locations (Goda and Hong, 2008; Baker, 2008). For example, Sheng et al. (2024) examined both inter-event and intra-event variability in TC wind fields and assessed wind-induced damage for a portfolio of structures. Zeng et al. (2020) employed semivariogram functions to characterise spatial correlations in the logarithmic bias of wind speeds across the North Atlantic. Typical semivariogram models include Gaussian, exponential, and spherical functions (Chilès and Delfiner, 2012). Under these models, the

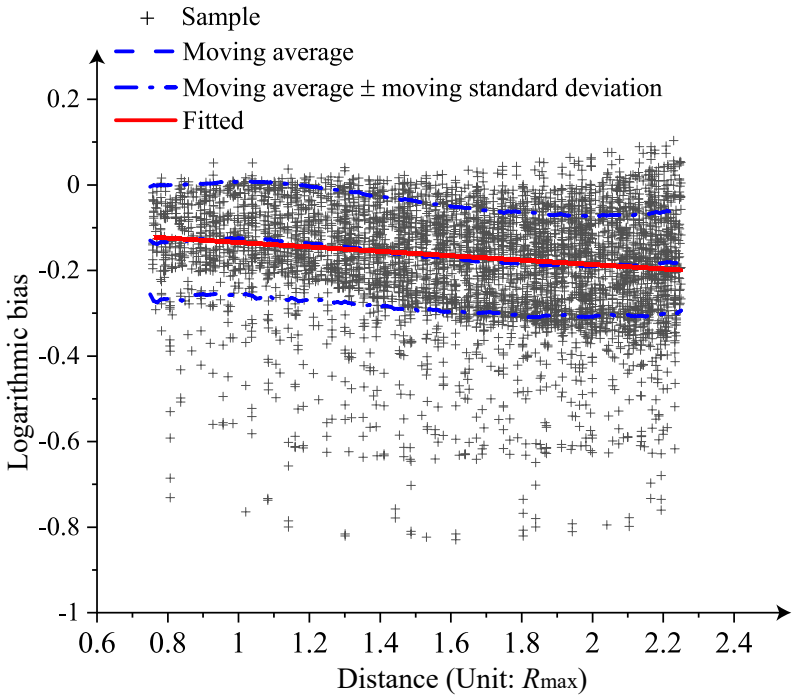


Figure 2.2: Logarithmic model bias as a function of distance from the storm centre (replotted from Zeng et al. (2021)).

site-to-site wind speed correlation is expressed as:

$$\rho_{U_i,U_j} = \begin{cases} 1 - a\left(1 - \exp\left(-\dfrac{h_{ij}^2}{b^2}\right)\right), \\ 1 - a\left(1 - \exp\left(-\dfrac{h_{ij}}{b}\right)\right), \\ 1 - a\left(\dfrac{3h_{ij}}{2b} - \dfrac{h_{ij}^3}{2b^3}\right) \text{ if } h_{ij} \leq b, a \text{ otherwise,} \end{cases} \tag{2.4}$$

where ρ_{U_i,U_j} is the wind speed correlation between site i and site j, h_{ij} is the distance between the two sites, a is the partial sill of the semivariogram and b is the correlation length, which describes the decay rate of the correlation with respect to h_{ij}. The first, second, and third expressions correspond to the Gaussian, exponential, and spherical semivariogram models, respectively. The choice of semivariogram model depends on the wind field representation and the geographic characteristics of the study region. Zeng et al. (2020) found that among the three

models, the Gaussian formulation provides the best fit to empirical spatial correlations of TC wind fields in the North Atlantic (see Fig. 2.3). The optimal Gaussian parameters were estimated as $a = 1$ and $b = 1.322R_{\max}$, where $R_{\max}$ denotes the radius of maximum winds.

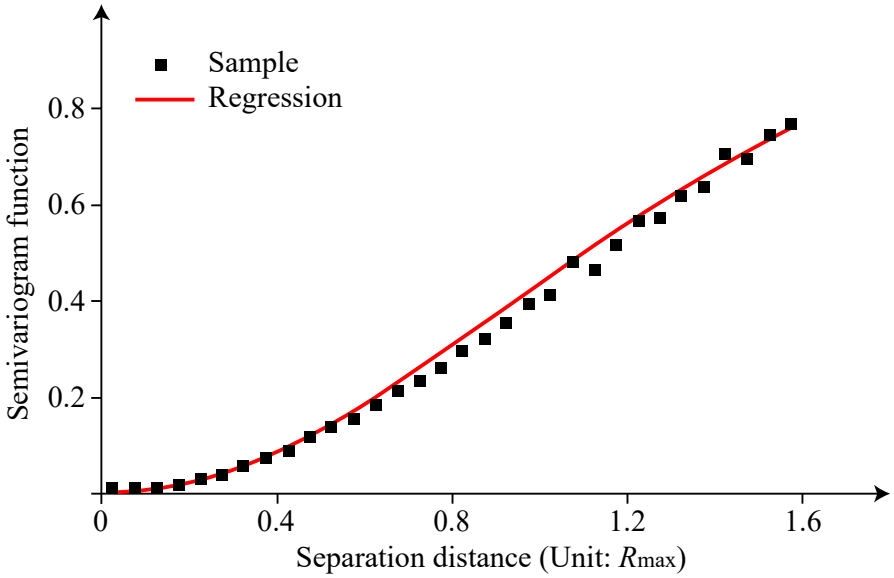


Figure 2.3: Fitting of semivariograms for estimating wind speed correlation between sites (replotted from Zeng et al. (2021)).

Another source of correlation arises from similarities in construction materials, structural designs, and building codes within a region, referred to as structure-to-structure correlation. These similar construction characteristics cause buildings to exhibit positively correlated damage states during TC events. To capture joint damage across regional structures, copula-based models, such as Gaussian copulas, have been applied in TC risk studies and in seismic assessments of bridge networks (Zeng et al., 2020; Vitoontus, 2012; Lee and Kiremidjian, 2007). However, relatively few TC studies simultaneously integrate multiple uncertainty sources with both hazard and structural correlations in community risk evaluations.

A third, practical challenge in regional TC risk analysis concerns computational demands. Scenario-based probabilistic risk analysis is a widely used approach for assessing TC-induced damage and losses. It involves modelling wind speed fields, assessing structural damage, and estimating community losses. This framework explicitly incorporates uncertainties and correlations in wind field modelling and structural response, providing a more realistic estimation of potential impacts than deterministic methods (Ellingwood et al., 2004, 2018; Vitoontus, 2012). However, although scenario-based probabilistic risk analysis provides a well-established framework, it often demands significant computational resources, particularly for large-scale communities. Moreover, the temporal and spatial evolution of TCs requires continuous monitoring to maintain up-to-date information on disaster impacts. This high computational burden limits its practical applicability in engineering contexts. To improve the efficiency of post-disaster assessments, one effective strategy is to establish a relationship between hazard intensity and regional damage, analogous to fragility functions used for individual buildings. Existing methods generally fall into two categories: (1) empirical models based on statistical data, and (2) simulation-based approaches. For instance, Huang et al. Huang et al. (2001) developed a TC loss model that correlates wind speed with insurance claim ratios and damage ratios, using historical insurance data from Hurricanes Hugo (1989) and Andrew (1992). The findings indicated that the wind speed exhibits a near-exponential relationship with the damage ratio (Fig. 2.4). Similarly, Kim et al. Kim et al. (2016) constructed a statistical model based on claim payouts for commercial buildings affected by Hurricane Ike in Texas. This model employed multiple regression analysis, incorporating variables such as wind speed, building age, floor area, and appraised value. However, empirical models based on insurance data are highly dependent on the availability and quality of

the underlying datasets. As a result, constructing such models is challenging in regions with limited or unreliable post-cyclone records. Simulation-based models typically employ physics-based or probabilistic frameworks. For example, Lin and Wang (2016) proposed a probabilistic approach for assessing the vulnerability of building portfolios under extreme wind events. Fig. 2.5 illustrates the relationship between earthquake magnitude and the different damage thresholds defined by the immediate occupancy ratio. Masoomi et al. Masoomi et al. (2019) constructed the relationship between TC-induced hazards (including wind, storm surge, and waves) and building damage by simulating structural responses under various hazard intensities. Similar methods have been explored in other studies (Pita et al., 2012). Nevertheless, most simulation-based approaches neglect the effects of spatial heterogeneity (e.g. terrain variability) on the functional relationship between hazard intensity and regional damage, which may reduce model accuracy in large-scale areas.

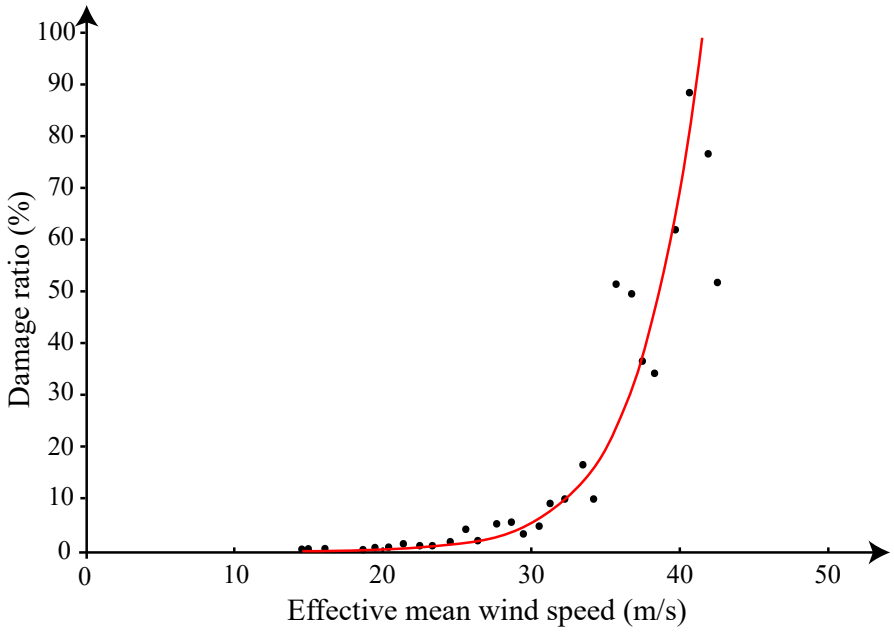


Figure 2.4: The relationship between damage ratio and effective mean surface wind speed (Adapted from Huang et al. (2001)).

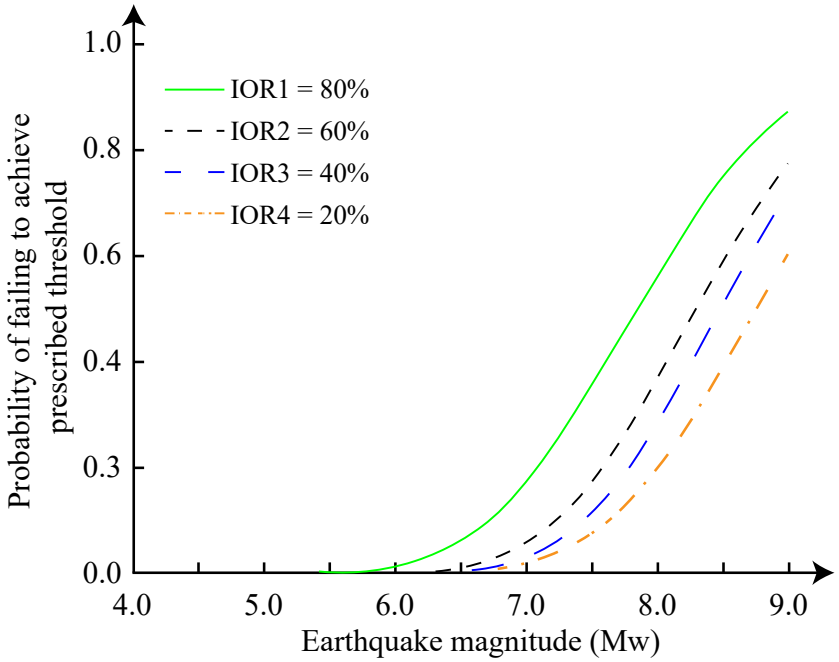


Figure 2.5: The probability of not reaching the four specified damage thresholds under different earthquake magnitudes (Adapted from Lin and Wang (2016)).

2.1.7 Applications of Bayesian networks in risk analysis

BNs are probabilistic graphical models that represent conditional dependencies among random variables by means of a directed acyclic graph (DAG). Each node denotes a random variable, and each directed edge represents a conditional dependence that may, under appropriate assumptions, be interpreted causally. In Fig. 2.6, variables A–F form a small BN: nodes with no parents (e.g., A) are root nodes, while nodes with no children (e.g., E and F) are leaf nodes. The parents of a node are those with arrows pointing into it; the children are those receiving arrows from it.

A key advantage of BNs is their ability to infer unknown variables from partial information, making them well-suited for modelling complex input-output relationships. BNs can also integrate data from diverse sources, including quantitative

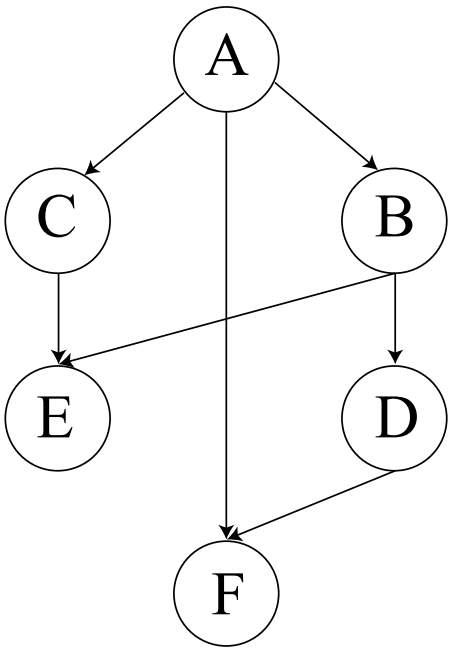


Figure 2.6: Schematic diagram of BNs.

measurements, semi-quantitative data, and expert judgment, while accommodating a wide range of variables such as hazards, climate conditions, losses, and decisions. This flexibility allows BNs to support both deterministic and probabilistic predictions (Zheng et al., 2021; Kaikkonen et al., 2021). Consequently, BNs have been widely adopted across many domains.

In the medical field, BNs have supported diagnosis and treatment planning for breast, lung, and cervical cancers (Roberts et al., 1995; Sesen et al., 2013; Onisko et al., 2019), the assessment of dementia and Alzheimer's disease (Seixas et al., 2014), and the modelling of epidemic transmission dynamics (Beresniak et al., 2012). In engineering diagnostics, BN-based methods have achieved competitive or superior performance relative to traditional approaches (Galagedarage Don and Khan, 2019). For example, Liu et al. (2020a) developed a BN for fault diagnosis in commercial aircraft environmental control systems, attaining a fault isolation accuracy of 89.3%. Galagedarage Don and Khan (2019) leveraged causal structures

between faults and symptoms to identify 8 out of 10 fault types from historical process data. BNs have also been widely applied in disaster-related risk assessments, including natural disasters such as earthquakes (Goerlandt and Islam, 2021), landslides (Moriguchi et al., 2023), and floods (Feng, 2024), as well as human-induced disasters, such as fires (Matellini et al., 2013), explosions (Aydin et al., 2024), and operational errors (Pan et al., 2021). For instance, Matellini et al. (2013) formulated a BN framework with two coupled parts: one for early fire development (fire type, time of day, human detection cues) and another for occupant actions, fire growth, and potential rescue outcomes. Zhou et al. (2020) proposed a BN for risk assessment of sewer pipelines in underground utility tunnels, identifying critical initiating events (e.g., fire sources, earthquakes, human errors, and external interferences) and quantifying their impacts on casualties, economic losses, and pollution (Fig. 2.7).

In TC risk assessment, BNs have been applied in several contexts. Wang (2023) developed a BN model to investigate the factors influencing evacuation decisions during TCs, considering factors such as risk perception, media information, and social influence. Van Verseveld et al. (2015) integrated TC-induced multi-hazard indicators (e.g., inundation depth, flow velocity, wave attack, scour depth) with building characteristics (e.g., type and property value) to estimate the probability of structural damage. However, this approach relies on post-cyclone observation data, which limits applicability in regions lacking reliable TC loss records. Furthermore, the model was developed at the building scale for relatively small study areas. Broader-scale applications may therefore require alternative modelling strategies. Li et al. (2020) proposed a BN for predicting power distribution system outages caused by TCs, incorporating variables related to TC dynamics and power grid topology. The assessment of power outages in this study

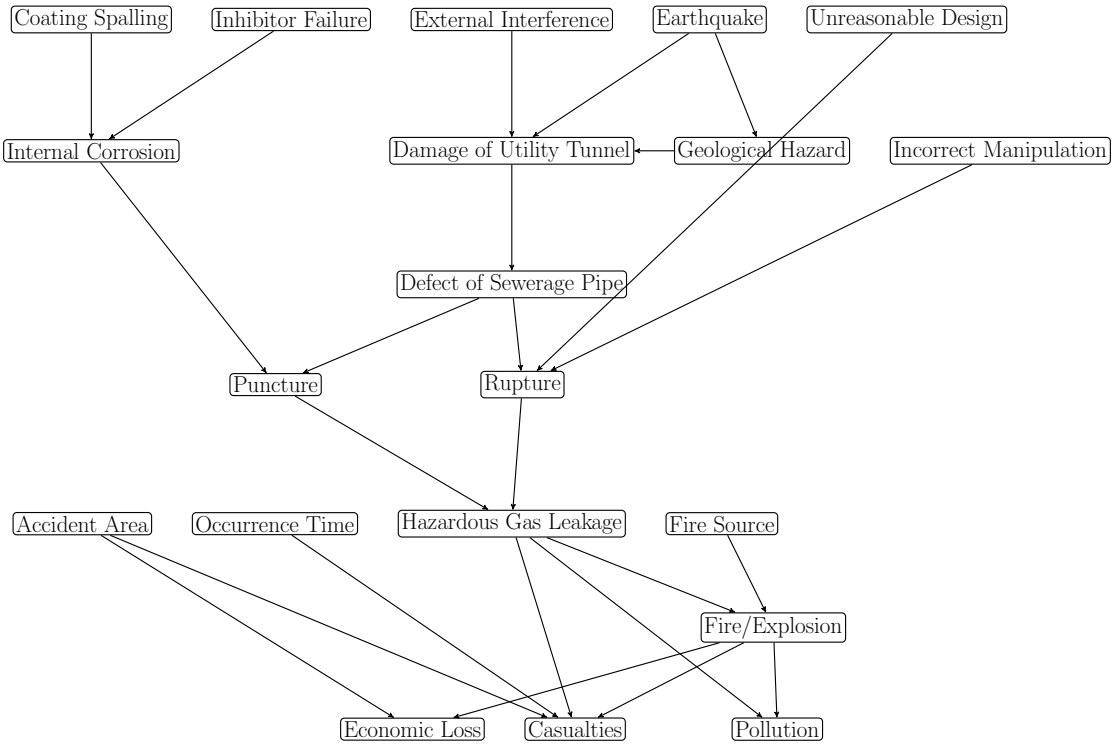


Figure 2.7: A risk assessment model for sewer pipelines in underground utility tunnels based on BNs (Zhou et al., 2020).

was based on network connectivity, whereas actual power systems are more complex and may also depend on power flow conditions. Similar research efforts are reported in (Cai et al., 2021; Lu and Zhang, 2022; Wang et al., 2024). Overall, existing applications of BNs in TC risk assessment have focused mainly on post-disaster decision-making, building-level damage assessment, and early studies in modelling lifeline network disruptions. Applications of BNs to assess TC loss at a regional scale remain limited.

A key challenge in BN modelling is the discretisation of continuous variables, which must balance model complexity against predictive accuracy. Three discretisation schemes are commonly used in BNs: manual, unsupervised, and supervised approaches. Manual discretisation relies on domain expertise and prior knowledge. This approach enhances interpretability and allows the integration of expert insights, but it introduces subjectivity and often lacks applicability across different contexts (Zheng et al., 2021). Unsupervised methods, such as equal-width and equal-frequency discretisation, are computationally efficient but often overlook important data patterns, reducing their predictive performance (Beuzen et al., 2018; Zheng et al., 2021). Supervised discretisation considers the relationship between inputs and outputs, retaining the information most relevant for prediction. Common supervised methods include entropy-based approaches (e.g., minimum description length (Fayyad, 1993)), decision trees (Costa and Pedreira, 2023), and Chi-squared-based methods (e.g., ChiMerge (Kerber, 1992)). A systematic review by Kaikkonen et al. (2021), which analysed 72 studies using BNs in environmental and engineering risk assessments, found that manual and unsupervised discretisation methods remain dominant due to simplicity, interpretability, and low computational cost, while supervised discretisation methods were applied in fewer than 10% of the cases. Nevertheless, supervised discretisation has consistently shown

superior performance. For example, Beuzen et al. (2018) demonstrated that supervised discretisation methods achieved 5%–10% higher accuracy than manual and unsupervised methods.

2.1.8 Effects of climate change on TC risk assessment

In recent decades, climate change has led to a significant increase in SSTs (Salarieh et al., 2023; Kieu et al., 2023). Since the formation and intensification of TCs are highly sensitive to oceanic and atmospheric conditions, such warming is expected to alter TC activity, thereby affecting TC risks in coastal urban regions. As a result, understanding how climate change influences TC behaviour and what this implies for urban infrastructure risk has become increasingly important.

The Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2021) introduced a set of scenario frameworks to support assessments of long-term climate risks under varying socioeconomic and emissions conditions. The Shared Socioeconomic Pathways (SSPs) describe five alternative trajectories (SSP1–SSP5) of global societal, demographic, and economic development over the 21st century, forming a foundation for analysing how different development patterns may influence climate outcomes. For example, SSP1 represents a sustainable development pathway characterised by strong mitigation and low emissions, whereas SSP5 depicts a fossil-fuel–intensive trajectory with high energy demand and limited climate action, leading to substantial emissions growth. In parallel, the Representative Concentration Pathways (RCPs) define trajectories of greenhouse gas concentrations and the corresponding radiative forcing levels by 2100 (2.6, 4.5, 6.0, and 8.5 W/m^2). Historically, climate projections were primarily based on RCPs alone, which specified greenhouse gas concentration outcomes without explicit socioeconomic context. Since the publication of the IPCC's Sixth Assessment Report (AR6), the combined "SSP-RCP" approach has become standard in

climate-change research (IPCC, 2021).

Climate change is generally expected to influence TC intensity, genesis location, and frequency. Regarding TC intensity, most studies indicate that future climate scenarios will likely lead to stronger storms (Knutson et al., 2020; Roberts et al., 2020; Bloemendaal et al., 2022). The projected intensification is consistent with enhanced enthalpy fluxes driven by higher SSTs and increased atmospheric moisture that together promote convective activity. According to the IPCC AR6 (IPCC, 2021), global mean TC wind speeds may increase by approximately 1–10% by 2100 under a 2°C warming scenario, while TC-induced rainfall rates are projected to rise by 6–22% under this condition. Studies on TC genesis suggest potential shifts in the spatial distribution of storm formation under a warming climate. In the North Pacific, some studies report a poleward migration in the mean latitude of TC genesis, consistent with changes in vertical wind shear, mid-level humidity, and potential intensity (Chen et al., 2021). In the North Atlantic, Weaver and Garner (2023) identified a northwestward shift, characterised by reduced TC activity in the traditional main development region and increased formation closer to the southeastern United States. However, Hall et al. (2021) had different conclusions and suggested an eastward displacement of genesis locations associated with elevated SSTs in the main development region. Regarding TC frequency, earlier research reported an apparent increase in recent decades (Mann and Emanuel, 2006; Holland, 2007). These studies collected historical TC records, as shown in Fig. 2.8, which shows annual counts of named storms ($\geq$34 knots), hurricanes ($\geq$64 knots), and major hurricanes ($\geq$96 knots) in the North Atlantic from 1851 to 2024, indicating an apparent increasing trend. However, subsequent analyses attribute this trend largely to improvements in satellite observations and data collection rather than an actual rise in storm occurrences (Vecchi et al., 2021;

Landsea et al., 2010). More recent assessments, including the IPCC AR6 (IPCC, 2021), indicate that the total global number of TCs may decrease, but there is high confidence that the proportion of intense storms (Category 4–5) will increase. This conclusion is consistent with recent findings. For instance, Knutson et al. (2020) projected an approximately 16% reduction in TC frequency in the North Atlantic, accompanied by a 24% increase in the occurrence of Category 4–5 storms.

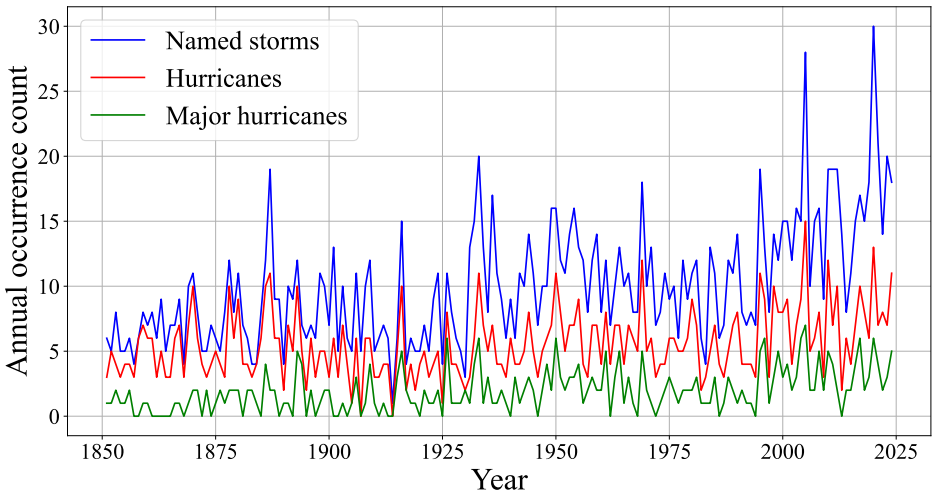


Figure 2.8: Annual frequency of named storms, hurricanes, and major hurricanes in the North Atlantic (1851–2024).

Global climate models (GCMs) are one of the most important tools for assessing the impacts of climate change. GCMs numerically represent the Earth's climate system by solving a set of physical equations that describe atmospheric, oceanic, and land-surface processes. Early generations of GCMs, constrained by limited computational power, typically employed coarse spatial resolutions (often exceeding 100 km). At such resolutions, GCMs could capture large-scale climate patterns such as global temperature distribution and general circulation, but were unable to resolve mesoscale phenomena like the inner-core structure of TCs. Consequently, they tended to underestimate TC frequency and intensity (Han et al., 2022). To improve spatial detail, scientists often employ downscaling techniques,

including dynamical and statistical approaches. Dynamical downscaling embeds a high-resolution regional climate model within a coarser-resolution GCM. The outputs of the coarse GCM provide boundary and initial conditions for the nested model, enabling simulation of fine-scale meteorological processes (Emanuel, 2021). Statistical downscaling, in contrast, establishes empirical or semi-empirical relationships between large-scale atmospheric variables (e.g., sea-level pressure, sea surface temperature) and local-scale climate variables (e.g., wind speed, precipitation), thereby translating coarse-resolution model outputs into local projections (Lee et al., 2020). However, dynamical downscaling is computationally expensive and requires substantial data inputs, making it impractical for producing large synthetic TC event sets. Statistical downscaling, on the other hand, depends heavily on the quality, representativeness, and temporal coverage of historical datasets. Recently, the Coupled Model Intercomparison Project Phase 6 (CMIP6) introduced a new generation of GCMs. These models feature substantially improved spatial resolutions, with some achieving grid sizes of 25 km or finer. Such advancements have significantly improved the capability of GCMs to simulate and identify TCs. Roberts et al. (2020) utilised the HighResMIP ensemble and demonstrated that higher-resolution simulations significantly reduce the biases observed in coarser models. Their results showed that high-resolution GCMs can reproduce more realistic TC counts and genesis locations. Nevertheless, these models still exhibit systematic biases and their simulation periods, typically ranging from 30 to 100 years, remain insufficient for comprehensive regional risk assessments.

2.2 Critical appraisal and pending research issues

Despite substantial advances in TC risk assessment, several critical challenges remain unresolved. These challenges are summarized below:

1. Computational challenges in large-scale probabilistic risk assessment. Traditional approaches rely on scenario-based probabilistic risk assessment frameworks, which are difficult to apply for dynamically evaluating post-disaster conditions due to their intensive computational demands. There is an urgent demand for efficient, accurate, and scalable methods capable of producing rapid regional damage estimates. Such methods must balance computational efficiency with predictive accuracy to enable dynamic support for emergency management.

2. Limited capability for spatially detailed loss assessment. Most existing models struggle to adequately capture the spatial variability of regional losses, particularly in large-scale communities. As a result, decision-makers often lack reliable information on where damage will be most severe. This limitation hinders their ability to prioritize emergency response, allocate limited resources, and plan long-term recovery and mitigation strategies. Therefore, there is a need for probabilistic approaches that incorporate uncertainties and correlations related to both hazard intensity and structural vulnerability, generate spatially detailed loss estimates at fine geographic scales, and remain computationally efficient for use in emergency contexts.

3. Incomplete consideration of climate change impacts. Existing research on the influence of climate change on TC behaviour has primarily focused on changes in TC intensity. However, relatively little attention has been given

to other potential changes, such as shifts in genesis locations and variations in TC frequency, and no effective methods currently exist to assess these factors. A comprehensive understanding of how altered TC activity affects hazard exposure is essential for anticipating emerging vulnerabilities in coastal communities. Therefore, it is necessary to adopt integrated approaches that evaluate the combined effects of climate change on TC intensity, frequency, and genesis locations.

CHAPTER 3

A PROBABILISTIC FRAMEWORK FOR TC LOSS MODELS OF BUILDING PORTFOLIOS

3.1 Scenario-based TC risk assessment of infrastructure systems

Scenario-based TC risk assessment is a widely used framework for estimating the impact of TCs on system performance under a given scenario. Specifically, for a TC scenario S, the performance of the infrastructure can be evaluated as follows:

$$P(Z \leq z \mid S) = \int_{\mathbf{v}} \int_{\mathbf{y}} F_{Z|\mathbf{DS}}(z \mid \mathbf{v}) f_{\mathbf{DS}|\mathbf{W}}(\mathbf{v} \mid \mathbf{y}) f_{\mathbf{W}|S}(\mathbf{y} \mid S) d\mathbf{v} d\mathbf{y}, \tag{3.1}$$

where $f_{\mathbf{W}|S}(\mathbf{y}|S)$ is the probability density function (PDF) of wind speeds, $\mathbf{W}$, at building locations given the TC scenario S. $f_{\mathbf{DS}|\mathbf{W}}(\mathbf{v}|\mathbf{y})$ is the PDF of physical damage states, $\mathbf{DS}$, under the condition of $\mathbf{W}$. $F_{Z|\mathbf{DS}}(z|\mathbf{v})$ is the cumulative distribution function (CDF) of the performance metric for regional buildings, Z, conditional on $\mathbf{DS}$. Finally, $P(Z \leq z|S)$ reflects the convolution of the three conditional probability distributions: $f_{\mathbf{W}|S}(\mathbf{y}|S)$, $f_{\mathbf{DS}|\mathbf{W}}(\mathbf{v}|\mathbf{y})$, and $F_{Z|\mathbf{DS}}(z|\mathbf{v})$.

3.1.1 TC wind field model

The Georgiou gradient wind field model (Georgiou, 1986) is widely used in engineering applications because it balances computational efficiency with the ability to capture the key characteristics of TC wind fields. This model extends the earlier formulation of Holland (1980) by incorporating storm translation effects, providing a more realistic representation of asymmetrical wind structures. The

gradient wind speed (V_g) is expressed as the vector sum of a rotational component and a translational component, as shown in Fig. 3.1. The calculation is given by:

$$V_g = \frac{1}{2}\left(V_t \sin\alpha - fr\right) + \sqrt{\frac{1}{4}\left(V_t \sin\alpha - fr\right)^2 + B\frac{\Delta P}{\rho}\left(\frac{R_{\max}}{r}\right)^B \exp\left(-\left(\frac{R_{\max}}{r}\right)^B\right)}, \tag{3.2}$$

in which V_t is the TC's translational speed, r is the distance from the TC centre, α is the clockwise angle between the translation direction and the radial position of the sites, f is the Coriolis parameter, B is the pressure profile exponent, ΔP represents the pressure difference between the periphery and the TC centre, $R_{\max}$ is the radius to maximum winds, and air density, ρ, is assumed to be 1.15 kg/m^3.

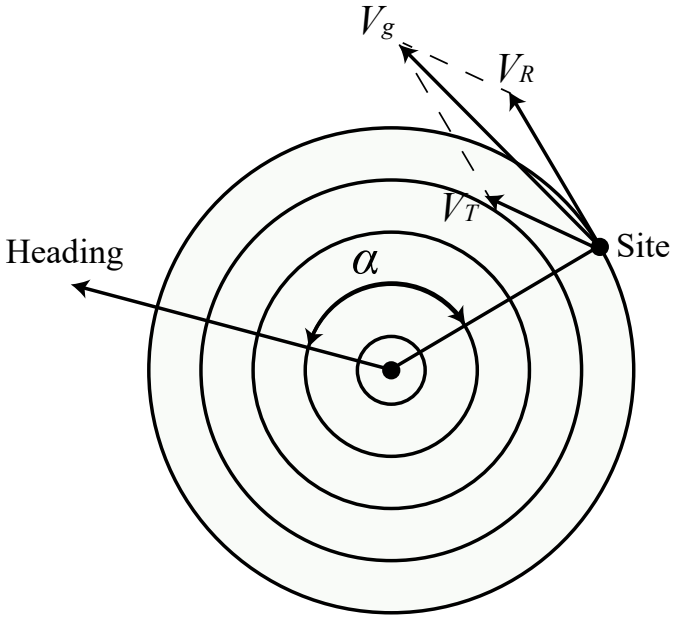


Figure 3.1: Schematic of wind field structure and gradient wind speed decomposition in a TC.

Structural damage assessment commonly relies on surface wind speeds; for example, fragility functions typically require 3-second gust wind speeds at 10 m height in open terrain as input. Therefore, converting V_g to surface wind speeds

is a necessary step, which can be expressed as:

$$U_0 = V_g \cdot F \cdot G, \tag{3.3}$$

where F is a gradient-to-surface conversion factor to transform the gradient wind speed into a surface wind speed. G represents the gust factor, which converts sustained wind speeds over long durations (e.g., 10 minutes to 1 hour) to gust wind speeds (e.g., 3 seconds). The conversion factor, F, in Eq. (3.3), comprises two components: the ratio of 10-meter overwater wind speed to gradient wind speed, $R(x)$, and the ratio of 10-meter overland wind speed to that over water, ϕ_t. The ratio $R(x)$ is defined as follows (Georgiou, 1986):

$$R(x) = \begin{cases} 1 - 0.35x, & 0 \leq x \leq 0.5 \\ 0.825, & 0.5 < x \leq 2 \\ 0.825 - 0.025(x-2), & 2 < x \leq 5 \\ 0.75, & x > 5 \end{cases} \tag{3.4}$$

where $x = r/R_{\max}$. The value of ϕ_t is estimated as (Georgiou, 1986):

$$\phi_t = \begin{cases} 1.0 - 0.015 \cdot DC, & 0 \leq DC \leq 10.0 \\ 0.875 - 0.0025 \cdot DC, & 10.0 < DC \leq 50.0 \\ 0.75, & DC > 50.0 \end{cases} \tag{3.5}$$

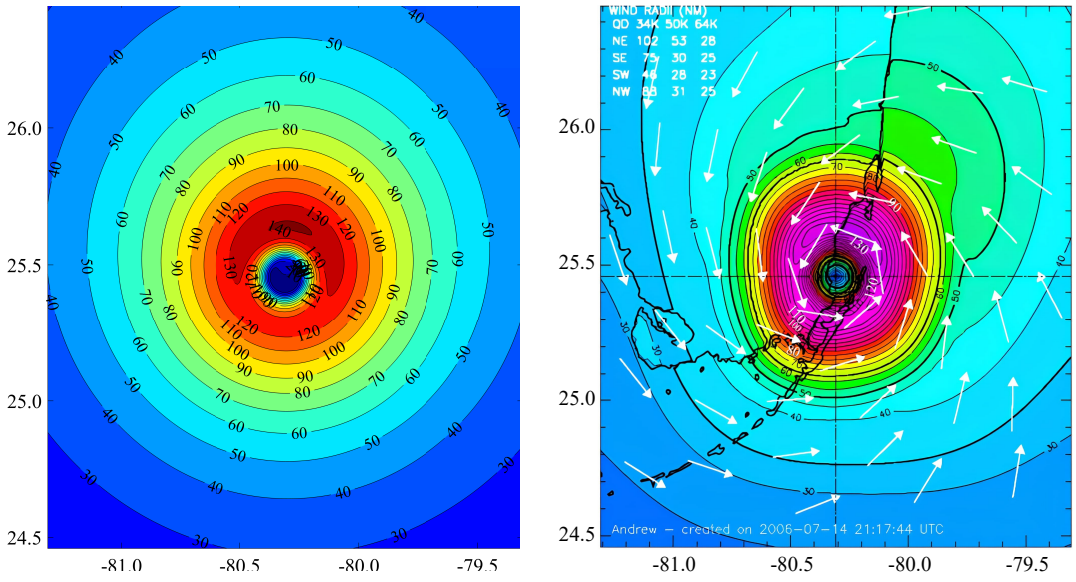
where DC is the inland distance from the coastline, measured in kilometres. The gust factor G is computed as (Vickery and Skerlj, 2005):

$$G = f(U_{10}, u_*, z_0) \cdot \left[1 + I_u \cdot g(y, \tau, 10)\right], \tag{3.6}$$

where U_{10} is the 10 m mean wind speed derived from Eqs. (3.4)–(3.5), u_* is the friction velocity, z_0 is the surface roughness length, I_u is the turbulence intensity at

10 m, and $g(y, \tau, 10)$ is the peak factor depending on turbulence spectral properties y, gust duration τ, and reference height (10 m). Further details can be found in Vickery and Skerlj (2005).

The Georgiou model was demonstrated using Hurricane Andrew (1992). As shown in Fig. 3.2a and Fig. 3.2b, these figures compare the wind field simulated by the Georgiou model with the observed wind field from H∗Wind. The comparison is based on the 1-minute wind speed at a height of 10 meters above the water surface at 0900 UTC on August 24, 1992. The simulated wind field closely matches the observed field from the H∗Wind archive, demonstrating the model's reliability for engineering applications.



(a) (b)

Figure 3.2: Wind field for Hurricane Andrew at 0900 UTC on August 24, 1992, (a) From Georgiou wind field model, (b) From H∗Wind legacy archive (Powell et al., 1998).

Historical TC databases such as HURDAT2 (HRD/NOAA, 2023), the international best track archive for climate stewardship (IBTrACS) (Knapp et al., 2010), the China Meteorological Administration dataset (Hong et al., 2016), the Hong

Kong Observatory (Doberck, 1884), and the Regional Specialized Meteorological Centre Tokyo (Kunitsugu, 2012), provide extensive global records of TCs. These databases typically include TC information such as longitude, latitude, maximum wind speed ($V_{\max}$), and central pressure (P_{c}), among other variables. Due to limitations in previous recording techniques, specific parameters are often missing or incomplete for early TCs. For example, the radius to maximum wind speed ($R_{\max}$) and P_{c} are frequently absent from earlier records. Additionally, the pressure profile exponent B, which characterises the pressure gradient around the storm eye and determines the rate of pressure increase outward from the centre, is not explicitly provided in these databases.

To address this issue, researchers have derived missing parameters through statistical analyses of historical records. For instance, for TCs in the Gulf of Mexico, the $R_{\max}$ is estimated as a function of pressure difference (ΔP):

$$\ln(R_{\max}) = 3.858 - 7.700 \times 10^{-5}\Delta P^2 + \varepsilon_{\ln R_{\max}}, \tag{3.7}$$

where $\varepsilon_{\ln R_{\max}}$ represents the model error for the logarithmic $R_{\max}$, which follows a normal distribution with a mean of zero and a standard deviation of 0.390.

For TCs in the Atlantic Ocean, an optimal model for $R_{\max}$ incorporates both ΔP and latitude (Ψ). In this case, $\varepsilon_{\ln R_{\max}}$ has a standard deviation of 0.466:

$$\ln(R_{\max}) = 3.421 - 4.600 \times 10^{-5}\Delta P^2 + 0.00062\Psi^2 + \varepsilon_{\ln R_{\max}}. \tag{3.8}$$

For TCs making landfall along the Gulf and Atlantic coasts, $R_{\max}$ is estimated as follows, with $\varepsilon_{\ln R_{\max}}$ having a standard deviation of 0.457:

$$\ln(R_{\max}) = 2.377 - 4.825 \times 10^{-5}\Delta P^2 + 0.0483\Psi + \varepsilon_{\ln R_{\max}}. \tag{3.9}$$

Finally, a generalised model applicable to all TCs is expressed as:

$$\ln(R_{\max}) = 3.015 - 6.291 \times 10^{-5}\Delta P^2 + 0.0337\Psi + \varepsilon_{\ln R_{\max}}, \tag{3.10}$$

where the standard deviation of $\varepsilon_{\ln R_{\max}}$ is 0.441.

The estimation of P_c is typically based on a power-law relationship with $V_{\max}$:

$$V_{\max} = m\,(P_{\text{ref}} - P_c)^n\,, \tag{3.11}$$

where m and n are regression coefficients, and P_{ref} represents the ambient atmospheric pressure, which is taken as 1016 hPa for the North Atlantic region and 1010 hPa for the Western North Pacific (Harper, 2002). The values of the regression coefficients vary by study region. For instance, Dvorak (1984) determined the coefficients for the North Atlantic region as $m = 3.45$ and $n = 0.644$.

The parameter B is commonly derived by minimizing the discrepancy between computed and observed $V_{\max}$ (Vickery and Wadhera, 2008; Zeng et al., 2021). This process often involves iterative calculations, which can be computationally demanding, particularly for large number of TC simulations. An alternative approach relies on regression models linking B to other TC parameters. For instance, Vickery et al. (2000) provided an empirical formulation for estimating B in cases where $\Delta P > 25$ mb and flight levels were below 3,000 m:

$$B = 1.34 + 0.000328\Delta P - 0.000522 R_{\max}. \tag{3.12}$$

For storms with $\Delta P > 25$ mb and flight levels below 1,500 m, a slightly adjusted model is used:

$$B = 1.38 + 0.00184\Delta P - 0.00309 R_{\max}. \tag{3.13}$$

These approaches allow for the reconstruction of complete storm datasets required for hazard modelling. A more comprehensive review of TC parameter estimation methods can be found in Vickery et al. (2000, 2009a); FEMA (2014); Harper (2002).

3.1.2 Damage assessment of individual structures under TC events

Fragility functions are commonly employed to estimate the probability that an individual structure will reach or exceed a specific damage state given a particular wind speed. Fragility functions are typically modelled using the CDF of a lognormal distribution, as shown in Eq. (2.1). The probability of a specific DS is expressed as

$$P(DS = v \mid U = u) = \Phi\left(\frac{\ln(u) - \lambda_v}{\xi_v}\right) - \Phi\left(\frac{\ln(u) - \lambda_{v+1}}{\xi_{v+1}}\right), \tag{3.14}$$

where $\Phi(\cdot)$ is the standard normal CDF; λ_v and ξ_v are the logarithmic mean and standard deviation of the fragility function for $DS = v$; and λ_{v+1} and ξ_{v+1} are the corresponding parameters for $DS = v + 1$.

HAZUS-MH defines five damage states, as shown in Table 2.1. Based on these classifications, the expected loss ratio for building i, $\mathrm{E}[LR_i]$, is calculated by:

$$\mathrm{E}[LR_i] = \sum_k \int_u \mathrm{E}[LR_i | DS_i = ds_i^{(k)}] \cdot P(DS_i = ds_i^{(k)} \mid U_i = u) \cdot f_u du, \tag{3.15}$$

where f_u is the PDF of wind speed at site i; $P(DS_i = ds_i^{(k)} \mid U_i = u)$ is the probability that the building reaches damage state k under wind speed u, which is calculated using fragility functions; and $\mathrm{E}[LR_i | DS_i = ds_i^{(k)}]$ is the expected structural loss ratio of building i given damage state k, which is typically estimated as follows: 2%–20% for slight damage, 10%–40% for moderate damage, 50%–80% for extensive damage, and 100% for complete failure (FEMA, 2014; Wang et al., 2017). Based on Eq. (3.15), the variance of the loss ratio for building i, $\mathrm{Var}[LR_i]$, is then calculated as follows:

$$\begin{aligned}\mathrm{Var}[LR_i] = &\sum_k \int_u \mathrm{E}[LR_i^2 | DS_i = ds_i^{(k)}] \cdot P(DS_i = ds_i^{(k)} \mid U_i = u) \cdot f_u du - \\ &\left(\sum_k \int_u \mathrm{E}[LR_i | DS_i = ds_i^{(k)}] \cdot P(DS_i = ds_i^{(k)} \mid U_i = u) \cdot f_u du\right)^2.\end{aligned} \tag{3.16}$$

3.1.3 Evaluation of system losses during TC events

As discussed in Section 2.1.6, assessing system losses during TC events requires consideration of two main types of spatial correlation. The first type is site-to-site correlation, which reflects the similarity of wind speeds between nearby locations. This correlation is commonly represented using an exponential function of the separation distance between two sites, as shown in Eq. (2.4). The specific correlation model and its parameters may vary depending on the wind field formulation and the geographic characteristics of the study area.

The second type is structure-to-structure correlation, which describes the spatial correlation between structural capacities. This correlation reflects similarities in design practices, construction materials, and building codes. It can be modelled as (Vitoontus, 2012):

$$q_{ij} = c \cdot \exp\left(-\frac{h_{ij}}{d}\right) + r, \tag{3.17}$$

where d is the scale of correlation, which determines the rate at which the correlation decays with distance. The parameters c and r reflect the degree of similarity between buildings.

Common performance metrics for regional buildings include the structural damage ratio, direct economic loss, economic loss ratio, and household displacement, among others (Lin and Wang, 2016; Ellingwood et al., 2018). Of these metrics, the loss ratio is frequently used to evaluate regional building performance, as it offers a standardized measure of value that facilitates comparisons and evaluations across various regions and building types. The overall loss ratio in a region, called building portfolio loss ratio (BPLR), is defined as the ratio of the total loss to the total replacement cost of all buildings. It is expressed as follows (Vitoontus,

2012):

$$\text{BPLR} = \frac{\sum_i LR_i \cdot RPC_i}{\sum_i RPC_i}, \tag{3.18}$$

where LR_i is the structural loss ratio of building i, RPC_i is the full structural replacement cost of building i.

Based on Eq. (3.18), the expected BPLR and its variance are given by:

$$\text{E}[\text{BPLR}] = \frac{\sum_{i=1}^{N} \text{E}[LR_i] \cdot RPC_i}{\sum_{i=1}^{N} RPC_i}, \tag{3.19}$$

$$\text{Var}[\text{BPLR}] = \frac{\left(\sum_{i=1}^{N} \left(\text{Var}[LR_i] \cdot RPC_i^2 \right) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \rho_{LR_i, LR_j} \cdot \sigma_{LR_i} \cdot \sigma_{LR_j} \cdot RPC_i \cdot RPC_j \right)}{\left(\sum_{i=1}^{N} RPC_i \right)^2}, \tag{3.20}$$

where ρ_{LR_i, LR_j} is the correlation coefficient between the loss ratios of buildings i and j, influenced by spatial proximity and structural similarities.

In practice, Eq. (3.1) cannot be solved in closed form due to the large number of buildings in a community, multiple sources of uncertainty, and the presence of spatial correlations. Spatial correlations arise both across sites, owing to similarities in wind speeds at nearby locations, and across structures, due to similarities in building performance resulting from common construction materials, structural types, and design codes (Vitoontus, 2012). Consequently, Monte Carlo simulation (MCS) is typically employed to approximate Eq. (3.1) (Ellingwood et al., 2018). In each simulation, a sample wind field is generated using a TC wind field model. A random field model is then applied to the TC wind field to account for the modelling uncertainties of the wind field model, which is a main contributor to TC risk assessment (Li and Ellingwood, 2006). The post-hazard damage states of buildings are then evaluated using structural fragility functions, such as those

provided in HAZUS (FEMA, 2014) or other empirical/analytical studies (Ellingwood et al., 2004; Li and Ellingwood, 2006). In the structural damage assessment, spatial correlations, including both site-to-site and structure-to-structure correlations, should be incorporated. The system performance indicator Z is computed based on the relationship between the individual structure's damage and overall system performance. Repeating this procedure over a sufficiently large number of simulations yields the probability distribution and statistics of the system-level performance metric Z.

3.2 Monte Carlo simulation for probabilistic risk analysis

3.2.1 Introduction to Monte Carlo simulation

The name "Monte Carlo" originates from the famous casino in Monaco, reflecting the method's reliance on randomness, much like games of chance. Monte Carlo simulation (MCS) is a computational technique that relies on random sampling and statistical methods to solve complex problems involving uncertainty. It is particularly useful in situations where analytical solutions are impractical or infeasible due to the complexity of the system, the presence of multiple interacting variables, or the need to account for randomness.

MCS is built upon probability theory, particularly the Law of Large Numbers and the Central Limit Theorem. The fundamental principle of this approach is that with a sufficiently large number of simulations, the statistical properties of the sampled data converge to the true values of the underlying probability distributions. By repeatedly drawing random samples from a specified probability distribution, it is possible to estimate quantities such as expected values, variances, and probabilities of extreme events.

Suppose a random variable X follows a probability density function $p(x)$, and we are interested in computing various statistical properties such as the expectation, variance, and CDF of a function $f(X)$. Theoretically, the expectation of $f(X)$ is expressed as:

$$\mathrm{E}[f(X)] = \int f(x)p(x)dx. \tag{3.21}$$

Instead of solving integrals analytically, MCS approximates these quantities through random sampling. By generating N independent and identically distributed (i.i.d.) samples $X_1, X_2, ..., X_N$, the Monte Carlo estimate of the expectation is:

$$\mathrm{E}[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i). \tag{3.22}$$

Similarly, the variance of $f(X)$ can theoretically be calculated using:

$$\mathrm{Var}[f(X)] = E[f(X)^2] - (E[f(X)])^2. \tag{3.23}$$

Using Monte Carlo sampling, the variance can be approximated by:

$$\mathrm{Var}[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i)^2 - \left(\frac{1}{N}\sum_{i=1}^{N} f(X_i)\right)^2. \tag{3.24}$$

Similarly, the CDF of $f(X)$ is defined as:

$$F_f(a) = P(f(X) \leq a) = \int_{-\infty}^{a} p_f(y)dy, \tag{3.25}$$

where $p_f(y)$ is the probability density function of $f(X)$. MCS estimates this probability by counting the proportion of sampled values satisfying the condition $f(X_i) \leq a$:

$$F_f(a) \approx \frac{1}{N}\sum_{i=1}^{N} I(f(X_i) \leq a), \tag{3.26}$$

where $I(\cdot)$ is an indicator function that equals one if the condition inside holds and zero otherwise.

3.2.2 Generation of a single random variable

In an MCS, it is necessary to generate random numbers that follow a given probability distribution. Based on the complexity of the probability distribution, various methods can be utilized to generate random samples. The most common approaches include the inverse transform method, rejection sampling, and, in more complex cases, Markov Chain Monte Carlo (MCMC) methods.

The inverse transform method is one of the most straightforward and widely used techniques for generating random samples from a given probability distribution. The core idea of the inverse transform method is that if a random variable X follows a given continuous distribution with cumulative distribution function $F_X(x)$, then the variable $U = F_X(X)$ follows a uniform distribution on $[0, 1]$. Specifically, a uniformly distributed random number $U \sim U(0, 1)$ can be transformed into a sample from the target distribution by computing:

$$X = F_X^{-1}(U). \tag{3.27}$$

The inverse transform method is particularly effective for distributions where $F_X(x)$ has a closed-form inverse. For example, in the case of an exponential distribution with parameter λ, the CDF is given by:

$$F_X(x) = 1 - e^{-\lambda x}, \tag{3.28}$$

and solving for x in terms of U yields:

$$X = -\frac{1}{\lambda} \ln(1 - U). \tag{3.29}$$

The method is also applicable to discrete distributions, where the inverse function can be computed iteratively by summing over the probability mass function. However, one limitation of the inverse transform method is that, for many distributions, it is difficult or impossible to compute the inverse CDF analytically. In such cases, this method is not suitable.

When the inverse transform method is not feasible, rejection sampling provides an alternative. The fundamental idea behind rejection sampling is to use an easily sampled proposal distribution $g(x)$ that resembles the target distribution $f(x)$ while ensuring that:

$$f(x) \leq Mg(x), \quad \forall x, \tag{3.30}$$

where M is a constant chosen to ensure that $g(x)$ properly envelops $f(x)$. A candidate sample is drawn from $g(x)$, and an additional acceptance step is performed to determine whether the sample should be retained. Specifically, a uniform random variable $U \sim U(0,1)$ is drawn, and the sample is accepted if:

$$U \leq \frac{f(X^*)}{Mg(X^*)}, \tag{3.31}$$

where X^* is the proposed sample. This acceptance criterion ensures that the final set of accepted samples follows the desired distribution.

To illustrate this method, consider a scenario where the target distribution is:

$$f_X(x) = 3x^2, \quad 0 \leq x \leq 1. \tag{3.32}$$

It is straightforward to verify that this is a valid probability density function, as its integral over $[0,1]$ equals 1. Since directly sampling from $f_X(x)$ is not trivial, we use a simpler proposal distribution $g(x)$, which we take as the uniform distribution:

$$g(x) = 1, \quad 0 \leq x \leq 1. \tag{3.33}$$

To ensure that $f_X(x) \leq Mg(x)$ for all x, we observe that the maximum value of $f_X(x)$ occurs at $x = 1$, where $f_X(1) = 3$. Therefore, we choose $M = 3$, ensuring:

$$f_X(x) \leq 3g(x), \quad \forall x \in [0,1]. \tag{3.34}$$

The rejection sampling procedure then follows these steps:

1. Draw a candidate Y from the uniform distribution $g(x) = U(0, 1)$.

2. Generate an independent, uniformly distributed random number $U \sim U(0, 1)$, which is used to determine whether to accept or reject the process.

3. Accept Y as a sample from $f_X(x)$ if:

$$U \leq \frac{f_X(Y)}{Mg(Y)} = \frac{3Y^2}{3 \times 1} = Y^2. \tag{3.35}$$

   Otherwise, reject Y. Then, repeat the process a certain number of times, for example, 10,000 times.

It is worth noting that the choice of the proposal distribution $g(x)$ has a significant impact on the method's efficiency. If the proposal function closely approximates the shape of $f(x)$, the acceptance rate will be high, leading to a more efficient sampling process. Conversely, a poor choice of $g(x)$ can lead to excessive rejections, making the method computationally expensive.

For even more complex distributions, particularly those in high-dimensional spaces, MCMC methods are often employed. MCMC techniques construct a Markov chain whose stationary distribution is the target distribution, allowing for iterative sampling from otherwise intractable probability distributions. The Metropolis-Hastings algorithm is a commonly used MCMC method where a proposed state is accepted or rejected based on a probabilistic criterion, ensuring that the Markov chain converges to the desired distribution over time. Another widely used approach, Gibbs sampling, is particularly effective when the conditional distributions of the target distribution are known and can be sampled efficiently.

In most cases in this thesis, the inverse transform method and rejection sampling are sufficient to generate random numbers from the required probability distributions. For more algorithms related to MCMC, please refer to Robert and Casella (2004); Suess and Trumbo (2010).

3.2.3 Generation of correlated random variables

Modelling correlated random variables is essential for reliability analysis and risk assessment. This section discusses two key approaches for generating correlated random samples: (1) the case where variables follow a multivariate normal distribution and (2) the case where variables follow non-normal distributions.

When all variables follow a multivariate normal distribution, generating correlated samples is straightforward. Suppose the random vector $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ follows:

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3.36}$$

where $\boldsymbol{\mu}$ is the mean vector, and $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{X}$. A standard approach to generate $\boldsymbol{X}$ involves the Cholesky decomposition of $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^{\mathsf{T}}, \tag{3.37}$$

where $\boldsymbol{L}$ is a lower triangular matrix. By generating an independent standard normal vector $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_n)$ (where $Z_i \sim \mathcal{N}(0, 1)$), we obtain the correlated $\boldsymbol{X}$ through:

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{Z}. \tag{3.38}$$

In many real-world scenarios, variables do not follow a normal distribution. In such a scenario, the Nataf transformation can be applied for correlation modelling of non-normal distributions while maintaining the known marginal distributions. Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a vector of random variables with known marginal distributions $F_{X_i}(x)$ and a known correlation matrix $\boldsymbol{P}_X = [\rho_{X_{ij}}]$. The goal is to generate samples of $\boldsymbol{X}$ while maintaining both the individual distributions and the prescribed correlation structure.

The Nataf transformation introduces an auxiliary vector $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)$, where each Y_i is a standard normal variable correlated via matrix $\boldsymbol{P}_Y = [\rho_{Y_{ij}}]$.

The relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by:

$$X_i = F_{X_i}^{-1}\big(\Phi(Y_i)\big), \tag{3.39}$$

where $F_{X_i}^{-1}(\cdot)$ is the inverse CDF of X_i, $\Phi(\cdot)$ is the standard normal CDF. Thus, the key step in the Nataf approach is determining $\boldsymbol{P}_Y$, the correlation matrix in normal space, such that the transformed variables $\boldsymbol{X}$ maintain their target correlation $\boldsymbol{P}_X$. The correlation transformation from $\boldsymbol{P}_X$ to $\boldsymbol{P}_Y$ follows an integral equation:

$$\rho_{X_{ij}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big[\frac{F_{X_i}^{-1}(\Phi(y_i)) - \mu_i}{\sigma_i}\Big]\Big[\frac{F_{X_j}^{-1}(\Phi(y_j)) - \mu_j}{\sigma_j}\Big]\phi_2(y_i, y_j; \rho_{Y_{ij}})\,\mathrm{d}y_i\,\mathrm{d}y_j, \tag{3.40}$$

where $\phi_2(y_i, y_j; \rho_{Y_{ij}})$ is the bivariate normal PDF with correlation $\rho_{Y_{ij}}$. μ_i and σ_i are the mean and standard deviation of X_i. This equation is solved iteratively to determine $\rho_{Y_{ij}}$ for each pair (i, j). Once $\boldsymbol{P}_Y$ is obtained, the standard normal samples $\boldsymbol{Y}$ are generated via Cholesky decomposition and then mapped back to $\boldsymbol{X}$ using the Eq. (3.37), Eq. (3.38) and Eq. (3.27).

In practical applications, when the coefficient of variation (CoV) of X_i is relatively small (typically CoV < 0.5), a polynomial approximation can be used to directly estimate $\rho'_{X_{ij}}$ from $\rho_{X_{ij}}$, avoiding expensive numerical integration. A commonly used approximation is:

$$\rho'_{X_{ij}} \approx R_{ij}(\mathrm{CoV}, \rho_{X_{ij}}) \cdot \rho_{X_{ij}}, \tag{3.41}$$

where R_{ij} is an empirical function determined by regression or lookup tables (Der Kiureghian and Liu, 1986; Melchers and Beck, 2018).

3.2.4 Latin Hypercube sampling

Latin Hypercube sampling (LHS) is a stratified sampling technique designed to improve the efficiency of Monte Carlo simulations, particularly in high-dimensional

spaces. For instance, in Chapter 4, we employ LHS to generate multiple TC-related variables for building the training dataset. The advantage of LHS is that it preserves the probabilistic characteristic of traditional MCS and, at the same time, ensures a more uniform and thorough exploration of the input space.

The fundamental idea of LHS is to divide the cumulative probability distribution of each input variable into N equally probable, non-overlapping intervals. For a problem with d random input variables and a desired sample size of N, the method generates one sample from each interval in each dimension, ensuring that all regions of the input space are represented. The resulting N samples are constructed such that each one draws values from unique intervals across all dimensions, reducing clustering and oversampling in localized regions. Mathematically, let $\boldsymbol{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ be a vector of d input random variables, each with a known cumulative distribution function $F_i(x)$. The unit interval $[0, 1]$ is partitioned into N equally spaced subintervals:

$$\left[0, \frac{1}{N}\right), \left[\frac{1}{N}, \frac{2}{N}\right), \ldots, \left[\frac{N-1}{N}, 1\right]. \tag{3.42}$$

Within each subinterval, a value is randomly sampled. For the j-th sample, the value in the i-th dimension is generated as:

$$x_j^{(i)} = F_i^{-1}\left(\frac{\pi_i(j) - 1 + u_{ij}}{N}\right), \quad j = 1, 2, \ldots, N, \quad i = 1, 2, \ldots, d, \tag{3.43}$$

where π_i is a random permutation of the integers $\{1, 2, \ldots, N\}$ for dimension i, u_{ij} is a uniform random variable sampled in the interval $[0, 1)$, and F_i^{-1} is the inverse CDF of the i-th variable. This process guarantees that each marginal distribution is well-sampled across its entire range. By shuffling the sample order in each dimension using the permutation π_i, the method avoids correlations induced by aligned sampling across dimensions. It promotes a space-filling design in the multidimensional input space.

When input random variables exhibit statistical dependence, the direct application of LHS may not adequately capture the underlying correlation structure. In such cases, the LHS method can be combined with the Nataf transformation to generate samples that not only preserve the marginal distributions of each variable but also approximate the desired correlation matrix. The process typically proceeds as follows:

First, the marginal distribution and the correlation structure of the input variables need to be specified. Assume that each $X^{(i)}$ in $\boldsymbol{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ has a known cumulative distribution function $F_i(x)$, and let $\boldsymbol{\rho}$ denote the desired correlation matrix for $\boldsymbol{X}$.

Second, the original variables are mapped into a standard normal space using the Nataf transformation. Each variable is independently transformed via its cumulative distribution and the inverse of the standard normal distribution function Φ^{-1}:

$$Z^{(i)} = \Phi^{-1}(F_i(X^{(i)})), \quad i = 1, 2, \ldots, d, \tag{3.44}$$

where the vector $\boldsymbol{Z} = (Z^{(1)}, Z^{(2)}, \ldots, Z^{(d)})$ is a standard Gaussian vector. However, to ensure that the transformation maintains the correct correlation in the original space, the Gaussian vector must have a correlation matrix $\boldsymbol{\Sigma}_Z$ that is consistent with the target correlation matrix $\boldsymbol{\rho}$ under the Nataf framework (see Section 3.2.3).

Third, LHS is performed in the standard normal space. For a total of N samples, the unit interval $[0, 1]$ is divided into N equally spaced subintervals in each dimension. A value is randomly selected from each subinterval, and a random permutation is applied to determine the order of values in each dimension. This process refers to Eq. (3.43), and the standard normal sample is generated as:

$$\tilde{Z}_j^{(i)} = \Phi^{-1}\left(\frac{\pi_i(j) - 1 + u_{ij}}{N}\right), \tag{3.45}$$

where $\tilde{\boldsymbol{Z}}_j = (\tilde{Z}_j^{(1)}, \ldots, \tilde{Z}_j^{(d)})$ are approximately independent and standard normally distributed. To introduce the desired correlation $\boldsymbol{\Sigma}_Z$, the Cholesky decomposition of the matrix is applied using Eq. (3.37) and each independent sample $\tilde{\boldsymbol{Z}}_j$ is then transformed to a correlated standard normal sample via:

$$\hat{\boldsymbol{Z}}_j = \boldsymbol{L}\tilde{\boldsymbol{Z}}_j. \tag{3.46}$$

Finally, the correlated Gaussian samples $\hat{\boldsymbol{Z}}_j$ are mapped back to the original variable space using the inverse Nataf transformation. For each dimension i, this involves computing the cumulative standard normal probability and applying the inverse of the marginal distribution function:

$$X_j^{(i)} = F_i^{-1}\left(\Phi(\hat{Z}_j^{(i)})\right), \quad i = 1, 2, \ldots, d, \tag{3.47}$$

where the samples $\boldsymbol{X}_j = (X_j^{(1)}, \ldots, X_j^{(d)})$ jointly follow the prescribed marginal distributions and approximate the target correlation matrix $\boldsymbol{\rho}$.

3.3 Development of TC loss models for building portfolios

3.3.1 Motivation and challenges

Inspired by fragility functions, which estimate structural damage based on wind speed without requiring structural analysis, this study aims to develop a TC loss model capable of estimating damage to building portfolios using only basic TC information. Once the model is established, TC damage assessments can be performed without conducting the scenario-based TC risk analysis described in Section 3.1, thereby significantly improving assessment efficiency.

However, the accuracy of TC loss models is significantly affected by regional characteristics. For example, when developing a wind-loss function for a large-scale community with uneven building distribution, the same regional wind speed

(typically measured at the regional centroid) may correspond to substantially different losses due to spatial heterogeneity: a TC impacting densely built areas will cause significantly more damage than one affecting sparsely populated regions. As a result, directly constructing a TC loss model may lead to a weak correlation between regional wind speed and the associated losses.

To address this issue, this study proposes a "divide-and-aggregate" approach. Instead of developing a TC loss model for the entire study area, the region is first divided into smaller sub-regions, and loss functions are then established for each sub-region individually. The total damage for the whole area is calculated by aggregating the loss estimates from all sub-regions. This approach has several advantages. First, because each sub-region is much smaller than the whole study area, the variation in wind speeds within a sub-region is significantly reduced. Second, the spatial division groups areas with similar characteristics together. Therefore, although the overall region may be heterogeneous, each sub-region tends to exhibit relatively uniform characteristics. Compared to constructing a wind–loss function for the entire study area, developing functions at the sub-regional level can lead to substantially improved accuracy.

3.3.2 Criteria for regional division

This section introduces three primary criteria to guide the regional division.

The first criterion is the spatial distribution of buildings. In most communities, buildings are not uniformly distributed but tend to cluster in specific areas. By grouping regions with similar building densities, we can distinguish between densely and sparsely populated zones, ensuring a relatively uniform distribution of structures within each sub-region. Government statistical data typically includes the number of buildings and the area of administrative regions. Building density ρ_b can be calculated using the formula $\rho_b = N_b/A$, where N_b represents the number of

buildings within a given area A. Alternatively, Kernel Density Estimation (KDE) can be employed to provide a more spatially continuous assessment of building density.

The second criterion is topography. Structures located in areas with dense vegetation or natural barriers experience significantly lower wind loads than those in open, unobstructed terrain. Grouping regions with similar topographical features into the same sub-region, therefore, helps ensure more consistent wind load effects. In wind engineering, topography is commonly quantified using the surface roughness length z_0, which represents the height at which the extrapolated logarithmic wind profile reaches zero velocity (FEMA, 2014). A higher z_0 value indicates rougher terrain, leading to greater wind attenuation due to surface friction. Roughness length can be obtained from government databases that document land use and vegetation characteristics. It may also be estimated using remote sensing or GIS methods by analysing land cover data, including vegetation patterns, built environments, and terrain features (FEMA, 2014).

The third criterion is the regional TC risk. While the precise location of a future TC cannot be predicted, the statistical distribution of TC risk across the study area can be assessed. Higher risk areas are more likely to experience TCs with greater frequency and severity, and vice versa. Therefore, regions with similar levels of risk should be grouped together. In this study, TC risk is measured using return period (RP) wind speed, which is defined as the wind speed expected to be exceeded, on average, once within a specified RP (Vickery et al., 2009b). Generally, regions closer to the coastline are exposed to higher return period wind speeds. These wind speeds are estimated by simulating thousands of years of TC activity and statistically evaluating the associated regional risks. For instance, Vickery et al. (2009b) developed a synthetic TC model to generate contour maps

of return period wind speeds for coastal cities in the United States. While different RPs correspond to varying wind speed levels, the spatial distribution patterns are similar across RPs.

Other criteria, such as structural types, building heights, and year of construction, also have an impact on improving sub-regional uniformity. These structural characteristics often exhibit spatial clustering. When similar structures are grouped, they tend to exhibit more consistent structural responses within sub-regions during TC events. However, such detailed information is often difficult to obtain in practice, as it requires not only data on these attributes but also the precise spatial location of each building. For this reason, these factors are not treated as primary considerations in this study.

3.3.3 Spatial clustering using the SKATER algorithm

The SKATER (Spatial 'K'luster Analysis by Tree Edge Removal) algorithm (Teixeira et al., 2015) is a graph-based spatial clustering method that groups spatial units with similar characteristics while preserving spatial contiguity within each cluster. In this study, we apply the SKATER algorithm, guided by the three regional characteristics introduced in Section 3.3.2, to partition the study area.

The study area is first discretised into spatial grids with a predefined resolution. These grids serve as the basic spatial units of analysis, each containing corresponding values for building density, roughness length, and return period wind speed. The choice of grid resolution should balance computational efficiency with modelling accuracy. These three attributes have different scales: building density ranges from tens to thousands per square kilometre, roughness length spans from 0 to 1, and return period wind speed ranges from tens to hundreds of miles per hour. Therefore, before clustering, these attributes are standardized to prevent any single feature from disproportionately influencing the clustering outcome.

The objective of the SKATER algorithm is to minimize within-cluster variance while ensuring that each cluster remains spatially contiguous. This objective can be expressed as:

$$\text{Minimize} \sum_{i=1}^{k} \sum_{x \in C_i} (x - \mu_{C_i})^2, \tag{3.48}$$

where μ_{C_i} is the mean value of cluster C_i, x is an individual data point within the cluster, and k is the number of clusters. This objective function is similar to that of the K-means algorithm; the key difference is that the SKATER algorithm includes an additional constraint — each cluster C_i must form a connected subgraph. That is, for any two spatial units within C_i, there must exist a path connecting them that includes only units belonging to C_i. For further details, refer to Teixeira et al. (2015). In this study, the SKATER algorithm is implemented using the PySAL library in Python.

The SKATER algorithm typically requires the number of clusters to be specified as an input parameter. However, manual selection of this number can be subjective. To address this, an adaptive approach is adopted to determine the optimal number of clusters: the sum of within-cluster variance is calculated for a range of cluster counts and plotted against the corresponding number of clusters. As within-cluster variance decreases monotonically with an increasing number of clusters, we identify the "elbow point" of the curve, where the rate of variance reduction significantly diminishes, as the optimal number of clusters. This approach ensures that features within each sub-region are sufficiently uniform while avoiding an excessive number of clusters.

To illustrate the SKATER algorithm, we present an example using a 3×3 spatial grid. The grid consists of 9 spatial units arranged in row-major order, each

uniquely indexed from 0 to 8:

$$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}. \tag{3.49}$$

Each spatial unit is associated with a single attribute X, with values randomly assigned in the range of 1 to 10. In this example, we assume $X = [2,\ 7,\ 10,\ 3,\ 7,\ 2,\ 5,\ 8,\ 4]$. Spatial adjacency is defined using the Rook criterion, where each unit is adjacent to its immediate north, south, east, or west neighbour. The goal is to partition these nine spatial units into two spatially contiguous clusters.

The objective function of the SKATER algorithm is equivalent to that of the classical K-means clustering, as shown in Eq. (3.48). What distinguishes SKATER from standard K-means is that it incorporates a spatial constraint: each cluster C_j must form a spatially connected subgraph. That is, if two spatial units i and j are assigned to the same cluster, there must exist a path between them consisting entirely of units within C_j. SKATER achieves this objective through the following steps. First, it constructs a spatial adjacency graph, where each node represents a spatial unit, and each edge connects a pair of neighbouring units. The edge weights represent the dissimilarity between connected units and are defined as:

$$w_{ij} = |X_i - X_j|. \tag{3.50}$$

Since this example involves only one attribute, using the absolute difference effectively captures the dissimilarity between units in terms of this attribute. For multivariate data, it is common practice to standardize each attribute and compute a multivariate distance (e.g., Euclidean distance) to account for all dimensions. Applying the above definition to each adjacent pair in the 3×3 grid yields the edge weights shown in Table 3.1.

Table 3.1: Edge weights derived from attribute differences for MST construction in the SKATER algorithm.

Edge	Nodes	$\lvert X_i - X_j \rvert$
e_1	(0–1)	5
e_2	(0–3)	1
e_3	(1–2)	3
e_4	(1–4)	0
e_5	(2–5)	8
e_6	(3–4)	4
e_7	(3–6)	2
e_8	(4–5)	5
e_9	(4–7)	1
e_{10}	(5–8)	2
e_{11}	(6–7)	3
e_{12}	(7–8)	4

Next, SKATER constructs a minimum spanning tree (MST) from this graph, using the edge weights to determine the optimal structure with minimal total dissimilarity. The MST connects all nine nodes with eight edges and avoids forming any cycles. The MST is formed by selecting the edges in ascending order of weight, provided they do not introduce loops. The resulting MST includes the following edges:

$$\{(1,4),\ (0,3),\ (4,7),\ (3,6),\ (5,8),\ (1,2),\ (6,7),\ (7,8)\}. \tag{3.51}$$

To partition the data into k clusters, the SKATER algorithm removes the $k-1$ most "inconsistent" edges from the MST. These are typically the edges with the highest weights as they represent the most significant attribute differences. In this example, we target $k = 2$ clusters. The edge (7–8) has the highest weight of 4 and is therefore removed. This splits the MST into two connected components: Cluster 1 (C_1): {0, 1, 2, 3, 4, 6, 7}; Cluster 2 (C_2): {5, 8}. To evaluate the clustering

result, we compute the within-cluster sum of squared deviations (SSW):

$$\text{SSW} = \sum_{j=1}^{2} \sum_{i \in C_j} (X_i - \bar{X}_j)^2. \tag{3.52}$$

For Cluster 1, the attribute values are $[2, 7, 10, 3, 7, 5, 8]$, the within-cluster sum of squares is $\sum_{i \in C_1} (X_i - 6.0)^2 = 48$. For Cluster 2, the values are $[2, 4]$, the within-cluster sum of squares is $\sum_{i \in C_2} (X_i - 3.0)^2 = 2$. Thus, the total within-cluster sum of squares is $48 + 2 = 50$. In comparison, if no clustering is performed (i.e., all nodes are in a single group), the total sum of squared deviations is about 64. This demonstrates a reduction in within-cluster variance from 64 to 50 after partitioning.

It is worth noting that the MST-based cut used in this example is a heuristic approach. In many cases, removing the edge with the most significant attribute difference yields a reasonable clustering result but not necessarily the optimal one. In practice, SKATER implementations employ a more flexible optimisation strategy. Rather than simply removing the edge with the highest weight, the algorithm may instead remove an edge with a smaller weight if doing so results in a lower overall SSW. For instance, in this example, SKATER choose to remove the edge between nodes 1 and 2 (with weight 3) rather than the edge between nodes 7 and 8 (with weight 4). This results in the following two clusters: Cluster 1 includes nodes {0, 1, 3, 4, 5, 6, 7, 8}, and Cluster 2 consists solely of node {2}. Although the removed edge does not have the highest weight, the resulting SSW is reduced to 39.5, which is lower than the 50 obtained from the naive heuristic. For further details on the optimisation constraints and improvements in SKATER, refer to AssunÇão et al. (2006); Teixeira et al. (2015); Aydin et al. (2018).

3.4 Example: TC loss models for building portfolios in Miami-Dade County

3.4.1 Overview of Miami-Dade County

Miami-Dade County, Florida, is highly vulnerable to cyclones and has experienced a sufficient number of historical events to support model development. Post-cyclone reports indicate that single-family residential buildings are among the most frequently observed types of damaged structures (Ellingwood et al., 2004). Accordingly, this section uses single-family residential buildings (hereafter referred to as residential buildings) as an example to illustrate the procedure for constructing TC loss models in Miami-Dade County.

According to building statistics reported in the Hazus Inventory Technical Manual (FEMA, 2014), single-family residential buildings are classified into four structural types: WSF1 (wood-framed, single-story), WSF2 (wood-framed, multi-story), MSF1 (masonry, single-story), and MSF2 (masonry, multi-story). Each category is further divided into subcategories based on variations in structural components, including roof shape, sheathing, and shutters. Specifically, WSF1 and WSF2 are each divided into eight subcategories, while MSF1 and MSF2 are each divided into twelve. The number of buildings in each subcategory is determined from the proportional combinations of these component types relative to the total number of buildings in the category, as summarised in Table 3.2.

Fragility functions for these structures are obtained from HAZUS-MH (FEMA, 2014). To account for variations in terrain, represented by surface roughness length, HAZUS-MH provides four fragility functions for each structure type: $z_0 = 0.03$ m (open terrain), $z_0 = 0.35$ m (typical suburban terrain), $z_0 = 0.7$ m (suburban terrain with some trees or densely spaced homes), $z_0 = 1.0$ m (treed suburban terrain). For locations with surface roughness values that fall between the defined

Table 3.2: Classification and quantity of residential buildings in Miami-Dade County (FEMA, 2014).

Building types	Component types	Number
WSF1	Roof shape: Gable: 62%, Hip: 38%; Sheathing: 8d@6"/6": 43%, 6d@6"/12": 57%; Shutters: Yes: 30%, No: 70%.	72379
WSF2	Roof shape: Gable: 62%, Hip: 38%; Sheathing: 8d@6"/6": 43%, 6d@6"/12": 57%; Shutters: Yes: 30%, No: 70%.	31022
MSF1	Roof shape: Gable: 62%, Hip: 38%; Sheathing: 8d@6"/6": 43%, 6d@6"/12": 25%, 8d@6"/12": 32% ; Shutters: Yes: 30%, No: 70%.	240325
MSF2	Roof shape: Gable: 62%, Hip: 38%; Sheathing: 8d@6"/6": 43%, 6d@6"/12": 25%, 8d@6"/12": 32% ; Shutters: Yes: 30%, No: 70%.	49219

categories, building damage is estimated by interpolating the fragility functions of the adjacent categories. To calculate loss ratios, the empirical relationship between DS and loss ratio is applied as follows (Wang et al., 2017): no damage = 0%, minor damage = 20%, moderate damage = 40%, severe damage = 80%, and destruction = 100%. Because absolute structural replacement costs can vary over time, while the relative cost ratios among building types tend to remain stable (FEMA, 2014), this study uses replacement cost ratios and assumes they remain constant across time and census tracts. According to construction cost estimates from (Homeguide, 2025), the average cost is approximately $30 per square foot for wood-framed structures and $47 per square foot for masonry structures. Based on these values, the structural replacement cost ratio of MSF1 to WSF1 is set to 1.56. In addition, the average height of one-story buildings is about 9 feet, while two-story buildings average about 17 feet (FEMA, 2014). Accordingly, the structural replacement cost ratios for WSF2/WSF1 and MSF2/MSF1 are both set to 1.88.

3.4.2 Spatial division according to three criteria

The spatial distribution of residential buildings is shown in Fig. 3.3. Building density is estimated using building count data from HAZUS at the census tract level (FEMA, 2014). Based on this data, buildings are randomly generated, and KDE is applied using the Spatial Analyst tool in ArcGIS to calculate density values. The results indicate that the highest concentrations of residential buildings are in the eastern and northeastern regions, while the western and southern areas contain significantly fewer buildings.

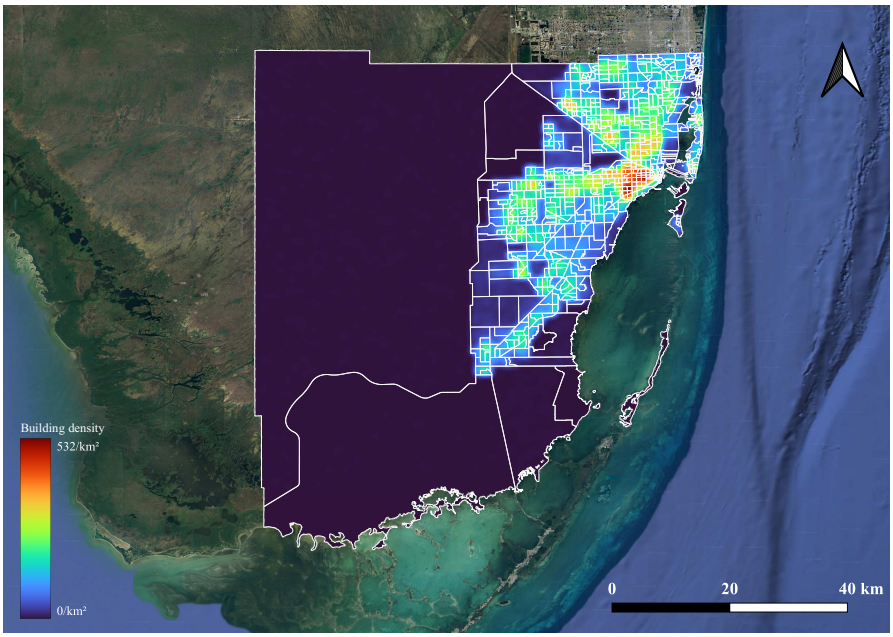


Figure 3.3: Residential building density in Miami-Dade County.

The HAZUS-MH manual (FEMA, 2014) provides surface roughness length z_0 for each census tract in Miami-Dade County. The distribution of z_0 is presented in Fig. 3.4. A correlation can be observed between regional terrain characteristics and building density: areas with relatively high building density generally have z_0 values between 0.35 and 0.70, whereas regions with lower building density typically exhibit z_0 values ranging from 0.03 to 0.35.

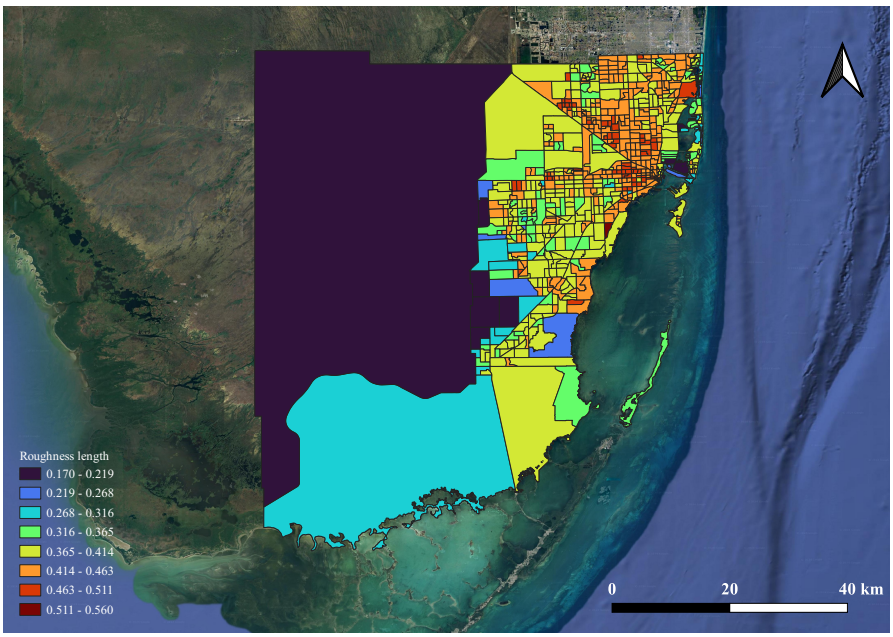


Figure 3.4: Surface roughness length at the census tract level in Miami-Dade County.

Vickery et al. (2009b) estimated TC risk contours for U.S. cities based on return period wind speeds. In this study, the 100-year return period wind speed is used to represent TC risk in the study area. Values between contour lines are estimated through interpolation, and the spatial distribution of TC risk is shown in Fig. 3.5.

The study area is partitioned using the three criteria and the SKATER algorithm described in Section 3.3. To determine the optimal number of clusters, the relationship between the sum of within-cluster variance and the number of clusters is plotted in Fig. 3.6a. An imaginary line is then drawn between the first and last observations. The perpendicular distance from each point on the curve to this line is calculated, and the results are shown in Fig. 3.6b. The "elbow point", which represents the maximum perpendicular distance from the line, indicates the optimal number of clusters (Satopaa et al., 2011), which is 13 in this example. To clarify sub-regional boundaries, a smoothing process is applied using the GRASS

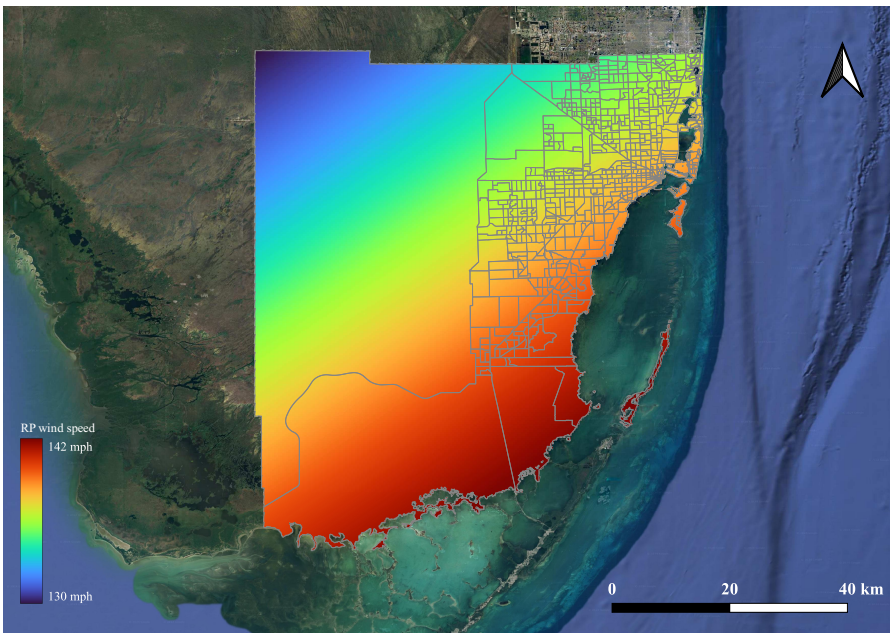
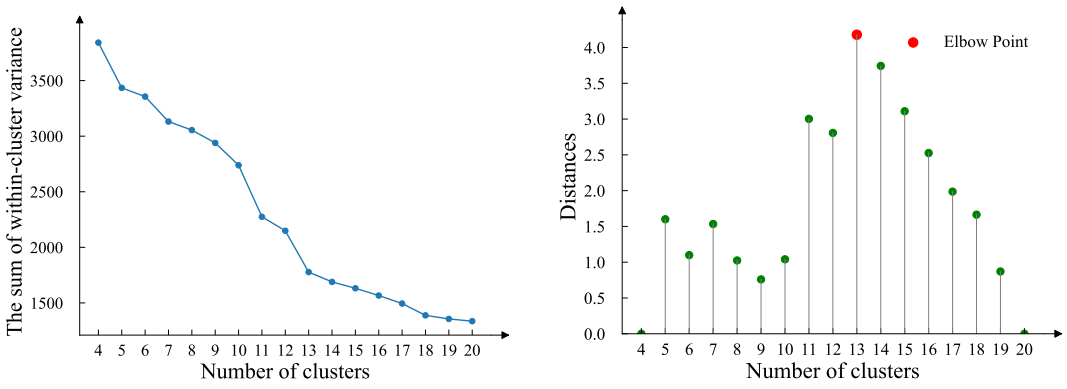


Figure 3.5: 100-year return period wind speeds in Miami-Dade County.

toolbox in QGIS. The spatial relationships among the 13 sub-regions are shown in Fig. 3.7. The centroid coordinates of each segment are shown in Table 3.3.

3.4.3 Construction of TC loss models

According to HURDAT2 (HRD/NOAA, 2023), a total of 122 TCs have passed within a 100-kilometre radius of Miami-Dade County since 1900. Of these, 27 reached Category 2 or higher on the Saffir-Simpson Hurricane Wind Scale, as illustrated in Fig. 3.8. In the figure, red traces represent all TCs, while yellow traces indicate those that reached Category 2 or above. TC loss models require a sequence of events encompassing both moderate and severe impacts. Historical statistics indicate that TC damage typically begins at Category 2 (Wikipedia, 2025c). Therefore, these 27 Category 2 or higher TCs are selected for developing the TC loss model. The basic information of these 27 TCs is presented in Table 3.4. In regions with limited historical TC records, synthetic TCs can be used. Previous studies (Vickery et al., 2000; Bloemendaal et al., 2020) have developed



(a) (b)

Figure 3.6: Elbow method for determining the optimal number of clusters: (a) sum of within-cluster variance vs. number of clusters; (b) elbow point identification.

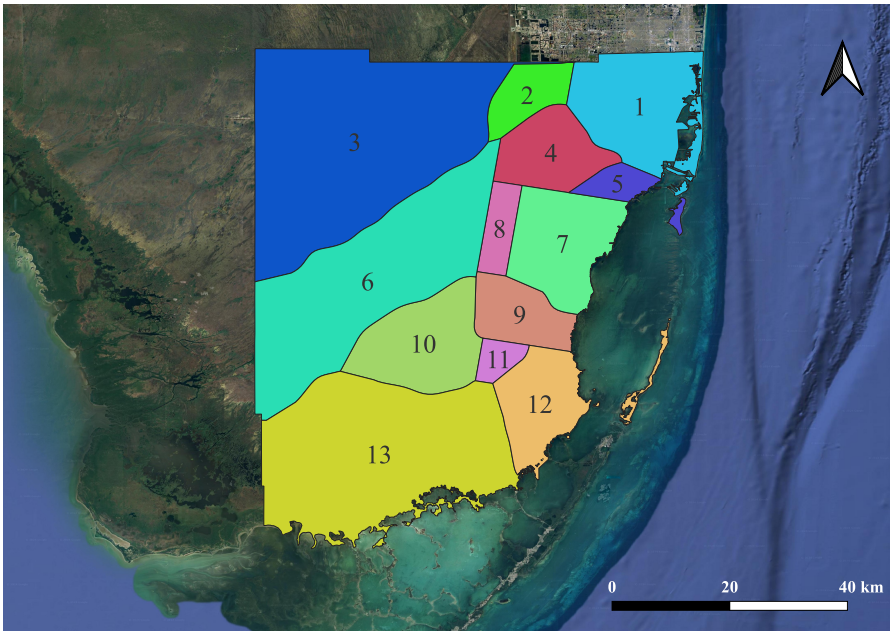


Figure 3.7: Spatial distribution of 13 sub-regions.

Table 3.3: Centroid coordinates of the 13 sub-regions.

Segment id	Longitude	Latitude
1	-80.224	25.880
2	-80.415	25.901
3	-80.705	25.822
4	-80.375	25.804
5	-80.250	25.746
6	-80.683	25.588
7	-80.355	25.652
8	-80.461	25.677
9	-80.427	25.532
10	-80.590	25.483
11	-80.463	25.457
12	-80.381	25.384
13	-80.664	25.297

synthetic cyclone models to simulate TC tracks and intensities. Further details on these methodologies are available in Snaiki and Wu (2020).

Due to the limitations of early recording techniques, certain parameters of historical cyclones, such as central pressure (P_c) and radius to maximum wind ($R_{\max}$), were not fully documented. The relationship between maximum wind speed ($V_{\max}$) and P_c is commonly modelled using Eq. (3.11). To estimate the regression coefficients, 570 historical TC snapshots with complete P_c and $V_{\max}$ data within a 500 km radius of the study area were collected for regression analysis, as shown in Fig. 3.9. These coefficients differ slightly from those reported by Dvorak (1984). The best-fit results from both studies are presented in Table 3.5. The discrepancy likely stems from the broader geographic scope of the Dvorak method, which covers the entire North Atlantic, whereas the present study focuses specifically on TCs near Miami-Dade County. Additionally, $R_{\max}$ and the B value are estimated using Eq. (3.9) and Eq. (3.13), respectively.

Each of the 27 TCs is used to calculate the BPLR for each sub-region, following

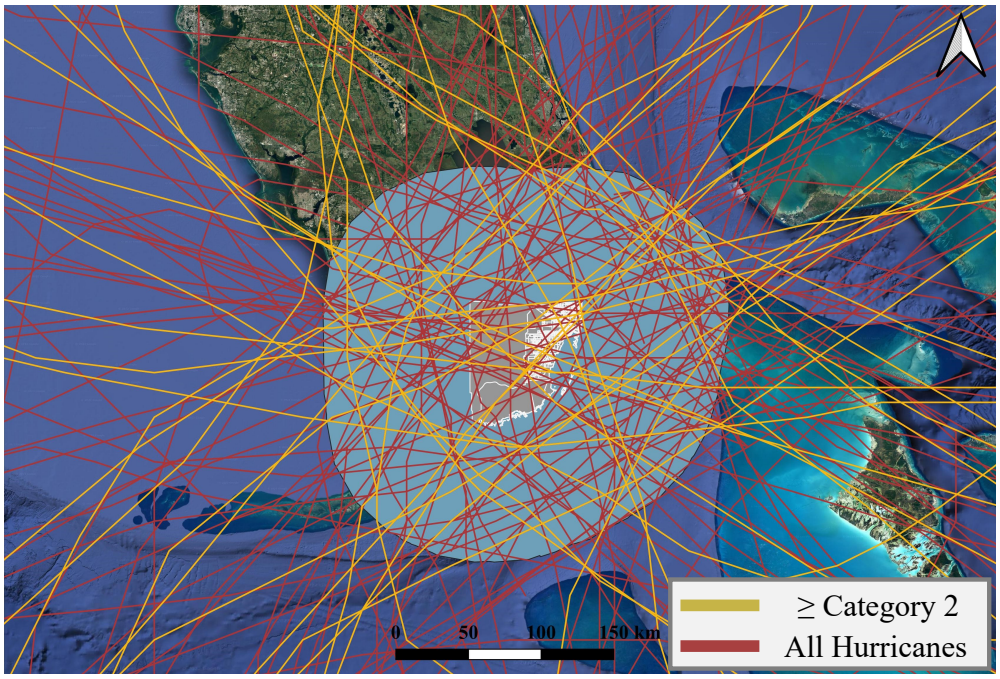


Figure 3.8: Historical cyclones within 100 km of Miami-Dade County based on HURDAT2 (HRD/NOAA, 2023).

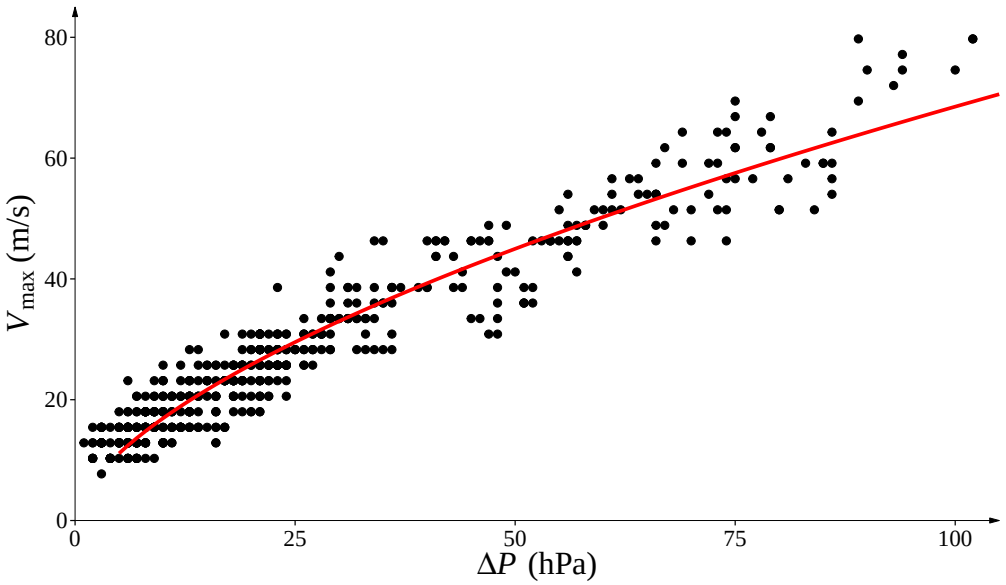


Figure 3.9: The relationship between the $V_{\max}$ and ΔP in the vicinity of Miami-Dade county.

Table 3.4: Cyclone data for developing TC loss models (Category 2+).

No.	Storm name	Duration	Category
1	Hurricane Eight	October 8 – October 23, 1906	3
2	Hurricane Five	October 6 – October 13, 1909	3
3	Great Miami	11 September - 22 September, 1926	4
4	Havana–Bermuda	October 14 - October 28, 1926	4
5	Fort Pierce	August 3 - August 14, 1928	2
6	Okeechobee	September 6 - September 21, 1928	5
7	Great Bahamas	September 19 - October 1, 1929	4
8	Treasure Coast	August 31 - September 7, 1933	4
9	Cuba–Bahamas	October 1 - October 9, 1933	3
10	Labor Day	September 6 - September 10, 1935	5
11	Cuba	September 23 - October 2, 1935	4
12	Yankee	October 30 - November 8, 1935	2
13	Florida	October 3 - October 13, 1941	3
14	Homestead	September 12 - September 20, 1945	4
15	Fort Lauderdale	September 4 - September 21, 1947	4
16	Florida	September 18 - September 26, 1948	4
17	Miami	October 3 - October 16, 1948	3
18	Florida	August 23 - August 31, 1949	4
19	KING	October 13 - October 20, 1950	4
20	DONNA	August 29 - September 14, 1960	4
21	CLEO	August 21 - September 5, 1964	4
22	ISBELL	October 8 - October 19, 1964	3
23	BETSY	September 12 - September 13, 1965	4
24	DAVID	August 25 - September 8, 1979	5
25	ANDREW	August 16 - August 29, 1992	5
26	WILMA	October 15 - October 27, 2005	5
27	IRMA	August 30 - September 13, 2017	5

Table 3.5: Comparison of regression coefficients in the $V_{\max}$–P_c relationship.

Source	m	n
This study	4.185	0.607
Dvorak (1984)	3.45	0.644

the scenario-based probabilistic risk assessment framework introduced in Section 3.1. Gradient wind speed is computed using Eq. (3.2) at 15-minute intervals while the TC centre remains within 100 km of the study area. The conversion factors F and G are determined using Eq. (3.5) and Eq. (3.6), respectively. Wind field uncertainty, ε, in Eq. (2.2), is quantified using the model proposed by Zeng et al. (2021). The variable ε follows a normal distribution, with its mean and standard deviation defined in Eq. (2.3). The spatial correlation of the wind field, i.e., site-to-site correlation described in Eq. (2.4), is estimated using a Gaussian model. The best-fit parameters applied in Eq. (2.4) are $a = 1$ and $b = 1.322R_{\max}$ (Zeng et al., 2021). It is important to note that the uncertainty and correlation models developed by Zeng et al. (2021) are based on overwater wind speeds. Therefore, these models should be applied before the water-to-land conversion in Eq. (3.5).

The parameters of spatial correlation between structures, i.e. the structure-to-structure correlation in Eq. (3.17), depend on construction material, structural type, and building code. Vitoontus (2012) provided relevant parameters for assessing the correlation between different structural types in seismic risk analysis. For TC risk assessment, Zeng et al. (2021) revised these parameters as follows: for buildings of the same type, $c = 0.5$ and $r = 0.2$; for buildings of different types, $c = 0.35$ and $r = 0.14$, and these values are adopted in this example. The scale of correlation, d, adopted 15 km for estimating TC-induced damage to the power network supports (Zeng et al., 2021). Given that the structural capacity of residential buildings tends to exhibit greater variability than that of power network supports, their correlation is expected to decline more rapidly with distance. Therefore, in this example, d is adjusted to 10 km.

To improve computational efficiency, a random sampling method is employed.

Fig. 3.10 shows the relationship between the estimated BPLR (y-axis) for Miami-Dade County and the number of random samples (x-axis) during Hurricane Andrew (1992). The results show that both the mean and standard deviation of the BPLR stabilise when the sample size reaches approximately 4,000, and this sample size is therefore adopted for the subsequent analysis.

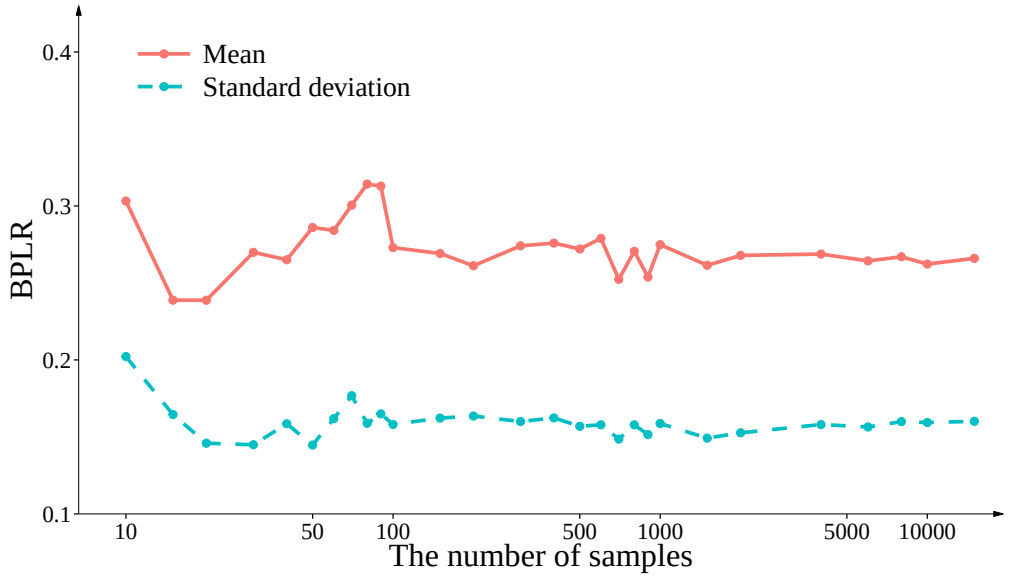


Figure 3.10: Convergence of BPLR with increasing sample size.

The TC loss model is then developed for each sub-region. The input variable is the maximum wind speed, defined as the 3-second gust at 10 meters height over open terrain, experienced at the centroid of the sub-region during the TC event. The output variables are the mean and standard deviation of the BPLR for each sub-region. The TC loss models for the 13 sub-regions are presented in Fig. 3.11 and Fig. 3.12. The observed relationships indicate that both the mean and standard deviation of BPLR follow the CDF of a lognormal distribution with

respect to wind speed:

$$\mu_{\mathrm{BR}_i} = \Phi\left(\frac{\ln(U_i) - a_1}{b_1}\right), \tag{3.53a}$$

$$\sigma_{\mathrm{BR}_i} = \Phi\left(\frac{\ln(U_i) - a_2}{b_2}\right), \tag{3.53b}$$

where μ_{BR_i} and σ_{BR_i} are the mean and standard deviation of the BPLR for sub-region i, U_i is the wind speed at the centroid of sub-region i, a_1, a_2, b_1 and b_2 are fitting coefficients, and $\Phi(\cdot)$ is the standard normal CDF. The regression coefficients of each segment (including $\mu_{\mathrm{BR}_i}(U_i)$ and $\sigma_{\mathrm{BR}_i}(U_i)$)) are presented in Table 3.6. The strong regression relationships observed in Figs. 3.11 and 3.12 demonstrate the effectiveness of the proposed spatial division method.

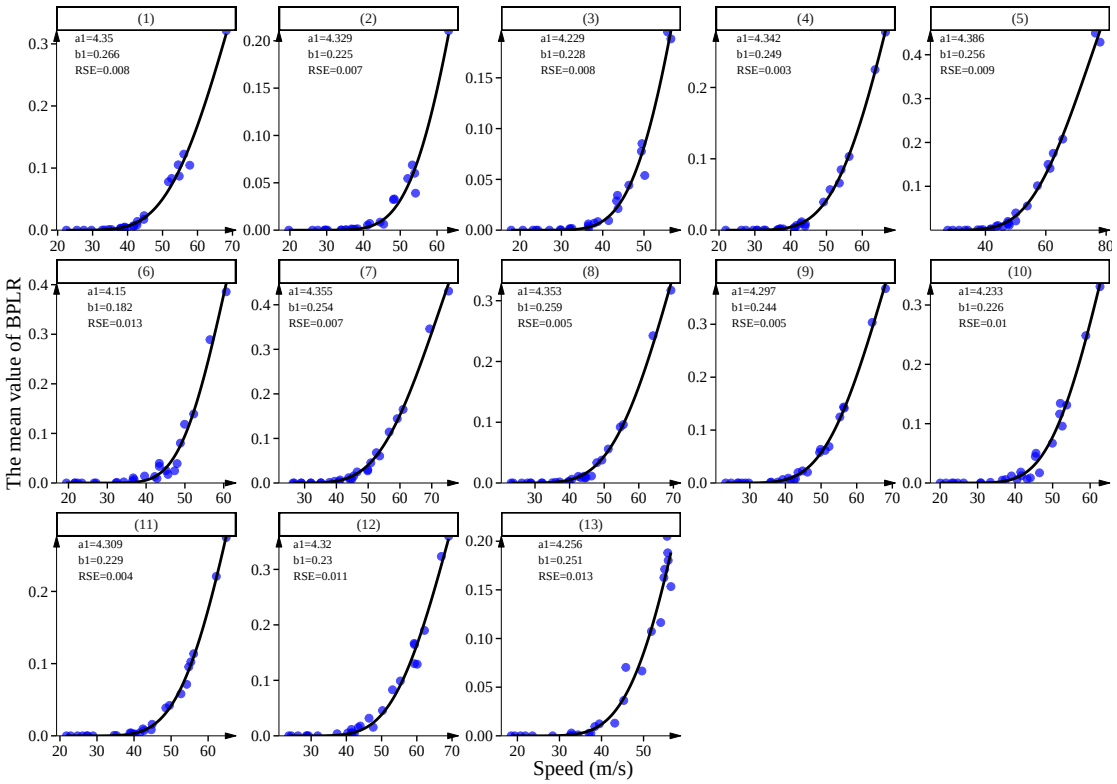


Figure 3.11: Relationship between centroid wind speed and mean BPLR across 13 sub-regions.

Based on the 13 values of μ_{BR_i} and σ_{BR_i}, the mean and standard deviation of the BPLR for Miami-Dade County, denoted as μ_{BR} and σ_{BR}, are calculated as

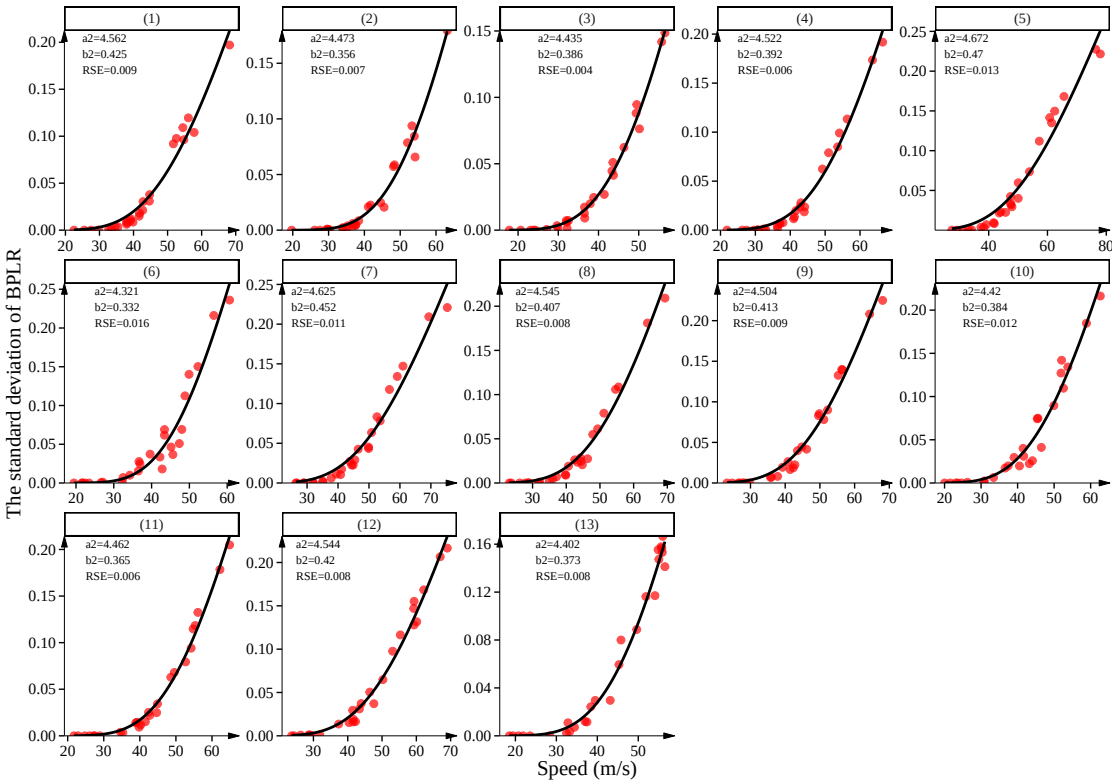


Figure 3.12: Relationship between centroid wind speed and standard deviation of BPLR across 13 sub-regions.

Table 3.6: Regression coefficients for the TC loss models for both $\mu_{\mathrm{BR}_i}(U_i)$ and $\sigma_{\mathrm{BR}_i}(U_i)$.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
a1	4.35	4.33	4.23	4.34	4.39	4.15	4.36	4.35	4.3	4.23	4.31	4.32	4.26
b1	0.27	0.23	0.23	0.25	0.26	0.18	0.25	0.26	0.24	0.23	0.23	0.23	0.25
a2	4.56	4.47	4.44	4.52	4.67	4.32	4.63	4.55	4.5	4.42	4.62	4.54	4.40
b2	0.43	0.36	0.39	0.39	0.47	0.33	0.45	0.41	0.41	0.38	0.37	0.42	0.37

follows:

$$\mu_{\mathrm{BR}} = \sum_{i=1}^{N} w_i \mu_{\mathrm{BR}_i}, \tag{3.54a}$$

$$\sigma^2_{\mathrm{BR}} = \sum_{i=1}^{N} w_i^2 \sigma^2_{\mathrm{BR}_i} + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} w_i w_j \rho_{ij} \sigma_{\mathrm{BR}_i} \sigma_{\mathrm{BR}_j}, \tag{3.54b}$$

where N is the total number of sub-regions (13 in this example), w_i is the proportion of sub-region i relative to the total residential building value, and ρ_{ij} is the correlation coefficient between the BPLR in sub-regions i and j. This coefficient reflects sub-regional characteristics (e.g., terrain features, structural types) and may also be influenced by the selection of TC events. In this example, the correlation matrix is estimated based on the 27 TC events in Fig. 3.8, including both moderate and severe cases, reflecting an average effect across all levels of severity. The correlation matrix shows that spatially adjacent sub-regions tend to exhibit stronger correlations:

$$\rho_{ij} = \begin{bmatrix}
1.00 & 0.94 & 0.77 & 0.87 & 0.85 & 0.67 & 0.76 & 0.72 & 0.75 & 0.68 & 0.69 & 0.74 & 0.29 \\
0.94 & 1.00 & 0.72 & 0.85 & 0.79 & 0.64 & 0.70 & 0.70 & 0.69 & 0.60 & 0.61 & 0.66 & 0.23 \\
0.77 & 0.72 & 1.00 & 0.88 & 0.89 & 0.91 & 0.89 & 0.92 & 0.86 & 0.81 & 0.78 & 0.78 & 0.37 \\
0.87 & 0.85 & 0.88 & 1.00 & 0.96 & 0.85 & 0.93 & 0.93 & 0.88 & 0.80 & 0.78 & 0.79 & 0.31 \\
0.85 & 0.79 & 0.89 & 0.96 & 1.00 & 0.88 & 0.96 & 0.94 & 0.91 & 0.85 & 0.83 & 0.84 & 0.35 \\
0.67 & 0.64 & 0.91 & 0.85 & 0.88 & 1.00 & 0.92 & 0.94 & 0.91 & 0.87 & 0.85 & 0.84 & 0.44 \\
0.76 & 0.70 & 0.89 & 0.93 & 0.96 & 0.92 & 1.00 & 0.97 & 0.96 & 0.90 & 0.88 & 0.87 & 0.42 \\
0.72 & 0.70 & 0.92 & 0.93 & 0.94 & 0.94 & 0.97 & 1.00 & 0.91 & 0.85 & 0.82 & 0.80 & 0.35 \\
0.75 & 0.69 & 0.86 & 0.88 & 0.91 & 0.91 & 0.96 & 0.91 & 1.00 & 0.94 & 0.95 & 0.94 & 0.49 \\
0.68 & 0.60 & 0.81 & 0.80 & 0.85 & 0.87 & 0.90 & 0.85 & 0.94 & 1.00 & 0.94 & 0.90 & 0.52 \\
0.69 & 0.61 & 0.78 & 0.78 & 0.83 & 0.85 & 0.88 & 0.82 & 0.95 & 0.94 & 1.00 & 0.95 & 0.53 \\
0.74 & 0.66 & 0.78 & 0.79 & 0.84 & 0.84 & 0.87 & 0.80 & 0.94 & 0.90 & 0.95 & 1.00 & 0.57 \\
0.29 & 0.23 & 0.37 & 0.31 & 0.35 & 0.44 & 0.42 & 0.35 & 0.49 & 0.52 & 0.53 & 0.57 & 1.00
\end{bmatrix}. \tag{3.55}$$

For comparison, an additional loss model is developed using the regional centroid wind speed and the total BPLR of the entire study area, as shown in Fig. 3.13.

Compared with the regression performance illustrated in Figs. 3.11 and 3.12, the relationship between the regional centroid wind speed and the total BPLR is significantly weaker.This indicates that directly constructing loss models at a regional scale may result in poor correlation due to spatial heterogeneity, potentially leading to suboptimal predictive performance.

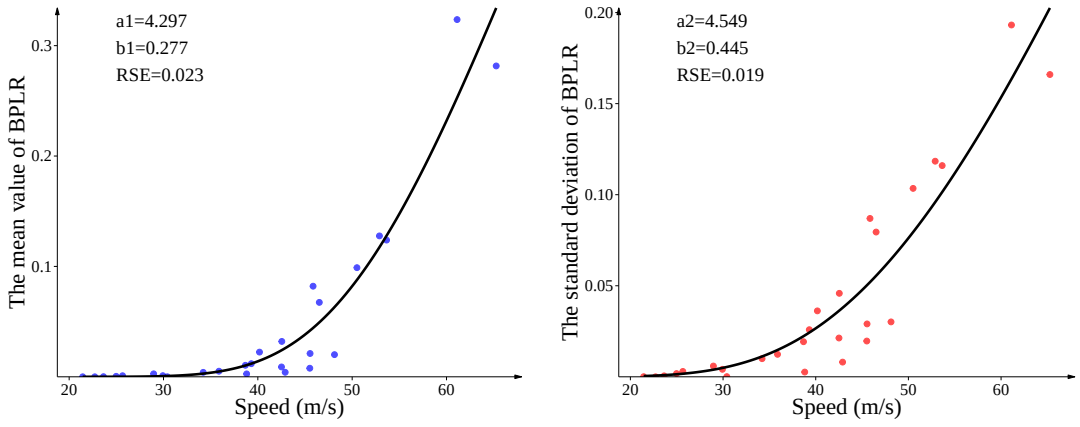


Figure 3.13: Fitted relationship between the regional centroid wind speed and the total BPLR (a) the mean value of BPLR, (b) the standard deviation of BPLR.

3.4.4 Model validation

To evaluate the validity of the TC loss models established in Section 3.4.3, a validation is performed using 100 synthetic cyclones from Bloemendaal et al. (2020). These TCs are selected based on their trajectories, ensuring that each passes within 100 km of Miami-Dade County. The spatial distribution of these cyclones is shown in Fig. 3.14.

Fig. 3.15a and Fig. 3.15b compare the true and predicted values of the mean and standard deviation of BPLR across 100 TC events, respectively. The true values are calculated using the probabilistic risk assessment framework described in Section 3.1, while the predicted values are obtained from the TC loss models

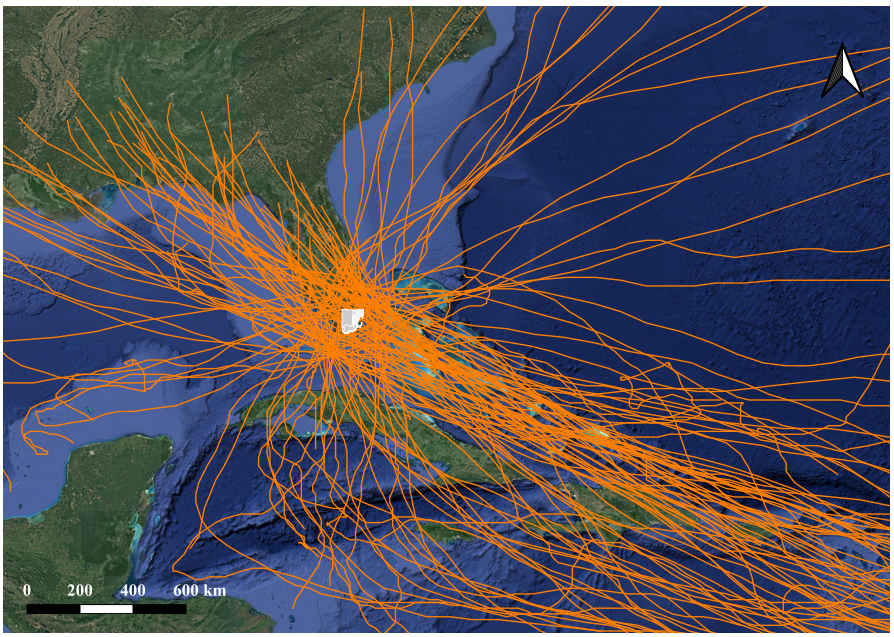
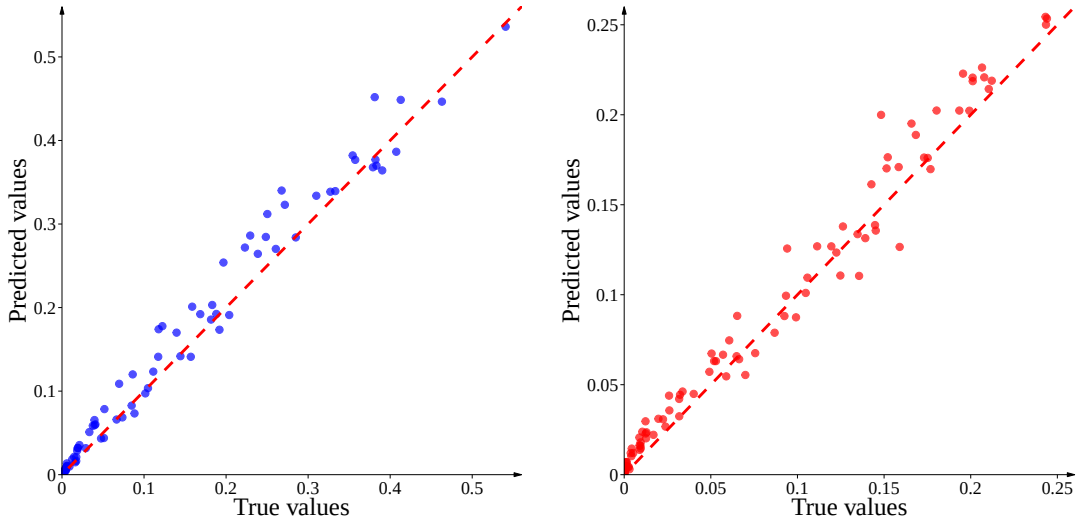


Figure 3.14: Spatial distribution of 100 test TCs used for model validation (Bloemendaal et al., 2020).

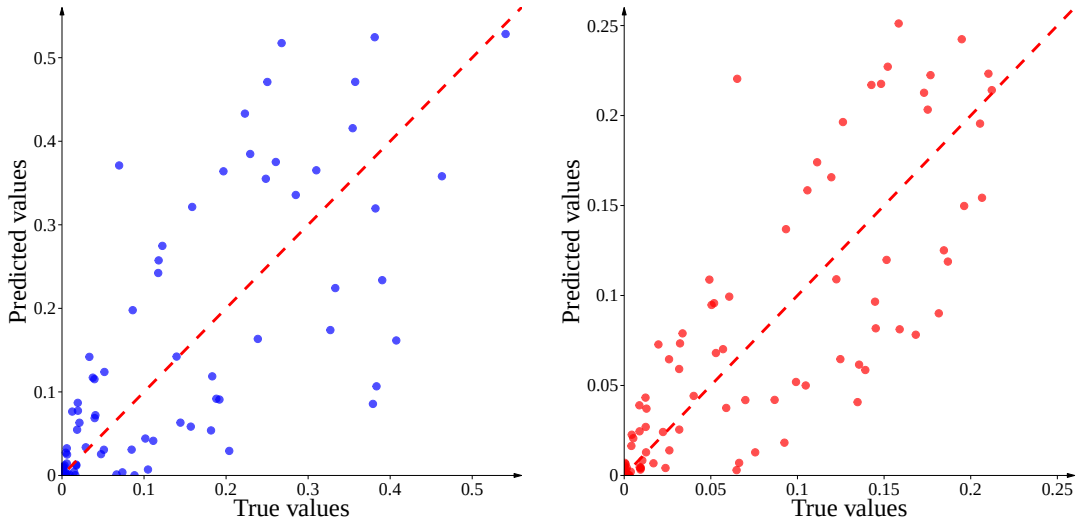
defined in Eq. (3.54). Most data points are tightly clustered around the 1:1 diagonal line. Even at higher BPLR levels, the model maintains satisfactory accuracy with only minor deviations. The close agreement between the two sets of values supports the validity of the proposed TC loss models and highlights their potential for efficient regional risk estimation.

For comparison, the prediction results based on the TC loss model constructed using the regional centroid wind speed and the total BPLR of the entire study area are presented in Fig. 3.16. The figure shows that both the predicted mean and standard deviation have considerably lower accuracy.



(a) (b)

Figure 3.15: Comparison of true and predicted BPLR for Miami-Dade County across 100 test TCs: (a) μ_{BR}; (b) σ_{BR}.



(a) (b)

Figure 3.16: Comparison of true and predicted BPLR for Miami-Dade County across 100 test TCs: (a) μ_{BR}; (b) σ_{BR}, based on the TC loss model developed without spatial division.

3.5 Conclusion

The heterogeneity of regional characteristics significantly influences the relationship between wind speed and regional losses in a large-scale community. To enhance the accuracy of regional loss estimation, this study proposes a "divide-and-aggregate" approach. This method involves developing TC loss models at the sub-regional level and then aggregating the results to estimate the regional BPLR. The key findings of this study are summarised as follows:

1. A spatial clustering approach is employed to divide a large region into sub-regions based on three criteria: the spatial distribution of buildings, the terrain, and regional TC risk characteristics.

2. The optimal number of clusters is determined by plotting the relationship between the sum of within-cluster variance and the number of clusters. The point at which the reduction in variance begins to level off, known as the elbow point, is identified as the optimal number of clusters.

3. The mean and standard deviation of the total BPLR are calculated as weighted sums of the sub-regional BPLR. When estimating the total variance, the correlation between sub-regions should be considered; neglecting this correlation can lead to an underestimation of the total variance.

CHAPTER 4

BAYESIAN NETWORK-BASED ESTIMATION OF REGIONAL BUILDING LOSSES FROM TROPICAL CYCLONES

4.1 Fundamentals of BNs

A BN represents probabilistic dependencies among variables using a directed graphical structure. It consists of a set of nodes (random variables) and directed arcs (dependencies). The direct predecessors or successors of a node are referred to as parents and children, respectively. Nodes without parents are called root nodes and represent input variables, while nodes without children are called leaf nodes and represent output variables.

Each node is associated with a conditional probability table (CPT), which specifies the probability distribution of the node given the states of its parents. For root nodes, the CPT is defined by a marginal distribution, which may be obtained from statistical data or expert judgment (Fenton and Neil, 2018).

A key property of BNs is conditional independence: a node is conditionally independent of its non-descendants given its parents. This property allows a complex joint probability distribution to be decomposed into local conditional probabilities, significantly simplifying the computation of the joint probability. According to the chain rule, the joint probability of variables X_1 to X_n can be expressed as:

$$P(X_1, X_2, ..., X_n) = P(X_1)P(X_2 \mid X_1) \dots P(X_n \mid X_1, X_2, ..., X_{n-1}). \tag{4.1}$$

Based on the property of conditional independence, the joint probability can be simplified to:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Pa}(X_i)), \tag{4.2}$$

where $\text{Pa}(X_i)$ denotes the set of parent nodes of X_i, and $P(X_i \mid \text{Pa}(X_i))$ is the corresponding CPT of X_i.

Inference in BNs can be classified into two types depending on the position of evidence variables: predictive inference and diagnostic inference. Predictive inference involves reasoning forward from cause to effect, estimating the probability distribution of an outcome variable based on a known cause. In contrast, diagnostic inference reasons backward from observed effects to potential causes. For predictive inference, the distribution of a target variable X_i can be updated given evidence $\mathbf{E}$, using the following expression:

$$P(X_i \mid \mathbf{E}) = \frac{P(X_i, \mathbf{E})}{P(\mathbf{E})} = \frac{P(X_i, \mathbf{E})}{\sum_{X_i} P(X_i, \mathbf{E})}, \tag{4.3}$$

where $\sum_{X_i}$ denotes the summation over all possible states of the variable X_i.

For diagnostic inference, the posterior distribution of a potential cause variable X_j can be updated based on observed evidence $\mathbf{E}$, using Bayes' Theorem:

$$P(X_j \mid \mathbf{E}) = \frac{P(\mathbf{E} \mid X_j) \cdot P(X_j)}{P(\mathbf{E})} = \frac{P(\mathbf{E} \mid X_j) \cdot P(X_j)}{\sum_{X_j} P(\mathbf{E} \mid X_j) \cdot P(X_j)}. \tag{4.4}$$

4.1.1 An illustrative example of a BN

Fig. 4.1 illustrates a simple example of a BN. This example includes three general connections in BN: serial (A-C-D), diverging (C-D and C-E), and converging connections (A, B - C). The prior distribution and conditional probabilistic table for this example are shown in Table 4.1.

The first calculation estimates the marginal probability distribution of node C in the absence of any observed evidence. This calculation uses the chain rule to

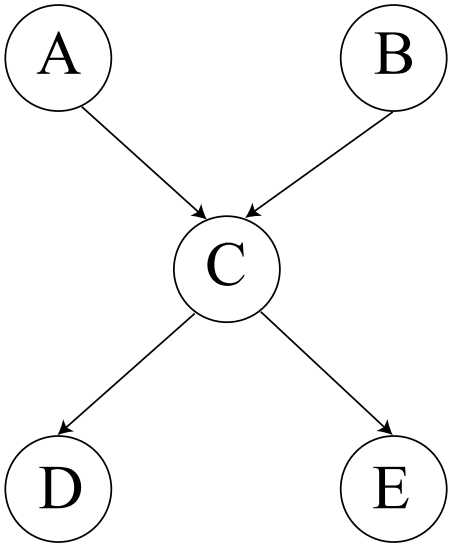


Figure 4.1: A Bayesian network example.

Table 4.1: The prior probability and CPT of the BN example.

Node	Parents	Values of parents	$P(Node\|Parents)$
A	None	N/A	$P(a_1) = 0.1, P(a_2) = 0.9$
B	None	N/A	$P(b_1) = 0.2, P(b_2) = 0.8$
C	A, B	$A = a_1, A = a_2$; $B = b_1, B = b_2$	$P(c_1\|a_1, b_1) = 0.3, P(c_2\|a_1, b_1) = 0.7$ $P(c_1\|a_1, b_2) = 0.4, P(c_2\|a_1, b_2) = 0.6$ $P(c_1\|a_2, b_1) = 0.5, P(c_2\|a_2, b_1) = 0.5$ $P(c_1\|a_2, b_2) = 0.6, P(c_2\|a_2, b_2) = 0.4$
D	C	$C = c_1, C = c_2$	$P(d_1\|c_1) = 0.7, P(d_2\|c_1) = 0.3$ $P(d_1\|c_2) = 0.8, P(d_2\|c_2) = 0.2$
E	C	$C = c_1, C = c_2$	$P(e_1\|c_1) = 0.8, P(e_2\|c_1) = 0.2$ $P(e_1\|c_2) = 0.9, P(e_2\|c_2) = 0.1$

expand the joint probability, simplifies it by the conditional independence among variables, and computes based on the known prior distributions and conditional probability relationships:

$$\begin{aligned} P(c_1) &= \sum_{a_i} \sum_{b_j} P(a_i, b_j, c_1) \\ &= \sum_{a_i} \sum_{b_j} P(a_i) \cdot P(b_j \mid a_i) \cdot P(c_1 \mid a_i, b_j) \\ &= \sum_{a_i} \sum_{b_j} P(a_i) \cdot P(b_j) \cdot P(c_1 \mid a_i, b_j) \\ &= 0.1 \cdot 0.2 \cdot 0.3 + 0.1 \cdot 0.8 \cdot 0.4 + 0.9 \cdot 0.2 \cdot 0.5 + 0.9 \cdot 0.8 \cdot 0.6 \\ &= 0.56. \end{aligned} \tag{4.5}$$

Given that $P(c_1) = 0.56$, the complementary probability is calculated as $P(c_2) = 1 - P(c_1) = 0.44$. Using the same procedure, the marginal probabilities for nodes D and E can also be computed as follows: $P(d_1) = 74.4\%$, $P(d_2) = 25.6\%$; and $P(e_1) = 84.4\%$, $P(e_2) = 15.6\%$.

The second calculation determines the conditional probability distribution of the child node C, given that node A is observed to be in state a_1. This inference involves a combination of the Bayes theorem, the law of total probability, and the chain rule:

$$\begin{aligned} P(c_1 \mid a_1) &= \frac{P(a_1, c_1)}{P(a_1)} \\ &= \frac{P(a_1, b_1, c_1) + P(a_1, b_2, c_1)}{P(a_1)} \\ &= \frac{P(a_1)P(b_1 \mid a_1)P(c_1 \mid a_1, b_1) + P(a_1)P(b_2 \mid a_1)P(c_1 \mid a_1, b_2)}{P(a_1)} \\ &= \frac{0.1 \cdot 0.2 \cdot 0.3 + 0.1 \cdot 0.8 \cdot 0.4}{0.1} \\ &= 0.38. \end{aligned} \tag{4.6}$$

Based on this result, the probability of c_2 given a_1 is obtained as $P(c_2 \mid a_1) = 1 - P(c_1 \mid a_1) = 0.62$. Following the same procedure, the conditional distributions

of other child nodes given $A = a_1$ are computed as $P(d_1|a_1) = 76.2\%$, $P(d_2|a_1) = 23.8\%$ and $P(e_1|a_1) = 86.2\%$, $P(e_2|a_1) = 13.8\%$.

The third calculation involves inference in the reverse direction, estimating the probability distribution of parent nodes given evidence from a child node. Specifically, we compute the posterior probability of node A being in state a_1, given that node C is observed in state c_1. This is a typical inference task in BNs:

$$P(a_1 \mid c_1) = \frac{P(a_1, c_1)}{P(c_1)} = \frac{0.038}{0.56} = 0.0679. \tag{4.7}$$

Once $P(a_1 \mid c_1)$ is known, the posterior probability of the complementary state is given by $P(a_2 \mid c_1) = 1 - P(a_1 \mid c_1) = 0.932$. Applying the same procedure, we obtain the conditional distributions of the other parent nodes given c_1 as $P(b_1 \mid c_1) = 17.1\%$ and $P(b_2 \mid c_1) = 82.9\%$.

4.1.2 Discretisation methods for BNs

In BNs, the presence of continuous variables often necessitates discretisation. Discretisation approaches can be broadly categorized into three types: manual, unsupervised, and supervised methods.

Manual discretisation is one of the most straightforward and interpretable approaches, relying on domain expertise or predefined standards to divide a continuous variable into discrete intervals. This process does not consider the statistical properties of the dataset or any target variable but instead reflects human judgment, regulatory guidelines, or physical classifications commonly used in practice. For instance, TC wind speed is discretised based on the Saffir-Simpson Hurricane Wind Scale. A tropical depression is defined as having sustained wind speeds less than 17.5 m/s (or 39 mph), while a tropical storm corresponds to wind speeds between 17.5 and 32.6 m/s (39–73 mph). Once the wind speed exceeds 32.7 m/s (74 mph), the event is classified as a TC, which is further subdivided into

Categories 1 (74–95 mph) through 5 (exceeding 157 mph). Other examples of manual discretisation include the use of temperature bands, such as "cold" (e.g., $\leq$10°C), "moderate" (10–25°C), and "hot" ($\geq$25°C), or air quality indices derived from pollutant concentration thresholds. This type of discretisation method is proper when interpretability and consistency with external standards are prioritized. However, they may not always align optimally with the underlying data structure or predictive performance.

Unsupervised discretisation methods rely solely on the statistical properties of the variable to be discretised. One of the most commonly employed unsupervised techniques is equal-width discretisation, where the range of a continuous variable is divided into k intervals of identical width. Formally, given a variable X with minimum and maximum values $\min(X)$ and $\max(X)$, the width of each bin is defined as:

$$w = \frac{\max(X) - \min(X)}{k}. \tag{4.8}$$

The bin boundaries are then placed at $\min(X) + iw$, where $i = 0, 1, ..., k$. While simple and computationally efficient, this method can be insensitive to skewed distributions, resulting in bins with highly imbalanced frequencies.

Another common unsupervised method is equal-frequency discretisation, also known as quantile-based discretisation. Instead of equal bin widths, this approach ensures that each interval contains approximately the same number of data points. For a dataset with n instances and k bins, each bin ideally contains n/k data points. This strategy effectively prevents sparsely populated bins, though it may create intervals with widely varying widths, particularly in highly skewed distributions.

A more adaptive unsupervised approach involves clustering-based discretisation. Here, data points are grouped into k clusters using unsupervised algorithms

such as K-means, where the goal is to minimize within-cluster variance. Let μ_j denote the centroid of cluster j, and let x_i represent the data point. The algorithm minimizes the objective function:

$$\sum_{j=1}^{k} \sum_{x_i \in C_j} (x_i - \mu_j)^2. \tag{4.9}$$

Each cluster is subsequently treated as a discrete bin. This method captures non-linear data distributions more effectively but introduces additional complexity and may require predefining the number of clusters. To demonstrate the procedure of K-means clustering, a two-dimensional example is shown in the following. Consider a simple dataset consisting of six two-dimensional points:

$$A(1,1),\ B(1.5,2),\ C(3,4),\ D(5,7),\ E(3.5,5),\ F(4.5,5). \tag{4.10}$$

We aim to partition this dataset into $K = 2$ clusters using the K-means algorithm. We begin by selecting two initial centroids: $\mu_1^{(0)} = A(1,1)$ and $\mu_2^{(0)} = D(5,7)$. We compute the squared Euclidean distance from each point to both centroids, as shown in Table 4.2.

Table 4.2: Squared Euclidean distance of each point to the two centroids.

Point	Distance to μ_1	Distance to μ_2	Assigned Cluster
A	0.00	52.00	C_1
B	1.25	37.25	C_1
C	13.00	13.00	C_1
E	22.25	6.25	C_2
F	28.25	4.25	C_2
D	52.00	0.00	C_2

After this step, the cluster assignments are: $C_1^{(1)} = \{A, B, C\}$ and $C_2^{(1)} = \{D, E, F\}$. Then, the new centroids are computed as the mean of the points in each cluster: For cluster C_1:

$$\mu_1^{(1)} = \frac{A + B + C}{3} = \left(\frac{1 + 1.5 + 3}{3}, \frac{1 + 2 + 4}{3}\right) = (1.833,\ 2.333). \tag{4.11}$$

For cluster C_2:

$$\mu_2^{(1)} = \frac{D+E+F}{3} = \left(\frac{5+3.5+4.5}{3}, \frac{7+5+5}{3}\right) = (4.333,\ 5.667). \tag{4.12}$$

Using the updated centroids, we again calculate distances and reassign points. The results remain unchanged: $C_1 = \{A, B, C\}$ and $C_2 = \{D, E, F\}$. Since there are no changes in cluster assignments, the algorithm converges.

Supervised discretisation uses information from a target variable to guide the binning process. This approach allows the discretisation to focus on regions of the input variable space that are most informative for predicting the output. In this section, the decision tree algorithm is used to illustrate the discretisation process. Decision trees are non-parametric supervised methods that determine optimal splits by maximizing information gain or, equivalently, by minimizing impurity (Jun, 2021).

When the target variable is categorical, discretisation is commonly performed using decision tree-based algorithms. These algorithms identify cut points that maximize the separability of the target classes. The quality of a candidate split is typically evaluated using information gain, which measures the reduction in entropy after the split. Entropy, denoted as $H(S)$, quantifies the impurity or uncertainty of a dataset S with respect to the class distribution and is defined as:

$$H(S) = -\sum_{c} p(c) \log_2 p(c), \tag{4.13}$$

where $p(c)$ is the proportion of instances in S that belong to class c. A perfectly pure dataset, where all instances belong to a single class, has an entropy of zero, while maximum entropy occurs when all classes are equally likely. Given a candidate split that divides S into subsets $S_1, S_2, \ldots, S_k$, the information gain from the split is:

$$\text{Gain}(S, A) = H(S) - \sum_{i=1}^{k} \frac{|S_i|}{|S|} H(S_i), \tag{4.14}$$

where A denotes the splitting attribute (i.e. the continuous variable being discretised), $|S_i|$ is the number of instances in subset S_i, and $|S|$ is the total number of instances in the parent dataset S. The term $\frac{|S_i|}{|S|}$ represents the relative size of subset S_i. The goal is to select the cut point that results in the greatest reduction in entropy, thereby producing subintervals that are more homogeneous in terms of the class label.

When the target variable is continuous, supervised discretisation follows similar principles. During discretisation, the trees recursively partition the TC variables by minimizing the weighted variance of the resulting subsets. The objective function can be expressed as:

$$\text{Minimize } \frac{1}{n} \sum_{m=1}^{M} \sum_{y \in S_m} (y - \bar{y})^2, \tag{4.15}$$

where S_m is the m-th subset, M is the total number of subsets, y is a sample of the PCs in S_m, $\bar{y}$ is the mean of the PCs in S_m, and n is the total number of samples.

To illustrate how a regression tree performs supervised discretisation on continuous variables, we demonstrate a simple example with 10 data points. Each data point contains two input features, x_1 and x_2, and one continuous target variable y. The values of these 10 samples are shown in Table 4.3. In this example, a regression tree with a maximum depth of two is built.

The initial sum of squared errors (SSE) at the root node (i.e. before any splits) is calculated as:

$$\text{SSE}_{\text{root}} = \sum_{i=1}^{10} (y_i - \bar{y})^2 = 42.976. \tag{4.16}$$

To grow the first layer of the tree, all possible feature-threshold pairs are evaluated to determine the split that minimizes the total SSE. Feature x_1 takes values

Table 4.3: Dataset with two input features and one target variable.

Sample	x_1	x_2	y
1	1	1	1.2
2	1	2	1.0
3	2	1	2.5
4	2	2	2.3
5	3	1	4.0
6	3	2	4.1
7	4	1	4.9
8	4	2	5.0
9	5	1	7.2
10	5	2	7.0

in the set $\{1, 2, 3, 4, 5\}$, and x_2 in $\{1, 2\}$. The candidate split thresholds are typically chosen between adjacent values. For x_1: thresholds at $\{1.5, 2.5, 3.5, 4.5\}$. For x_2: one threshold at 1.5. Each threshold divides the dataset into two groups, and the sum of the SSEs of the two resulting groups is calculated. The total SSEs corresponding to these five splits are: $x_1 < 1.5$: 16.87; $x_1 < 2.5$: 7.57; $x_1 < 3.5$: 13.60; $x_1 < 4.5$: 15.24; $x_2 < 1.5$: 28.53. Among all candidates, the split at $x_1 < 2.5$ yields the lowest combined SSE of 7.57. This threshold is, therefore, selected for the first-level split. After the first-level split, the resulting two groups are as follows:

1. Left node (Region A): $x_1 < 2.5$, containing 4 samples:
   $(1, 1, 1.2), (1, 2, 1.0), (2, 1, 2.5), (2, 2, 2.3)$, with an SSE of 1.035;

2. Right node (Region B): $x_1 \geq 2.5$, containing 6 samples:
   $(3, 1, 4.0), (3, 2, 4.1), (4, 1, 4.9), (4, 2, 5.0), (5, 1, 7.2), (5, 2, 7.0)$, with an SSE of 6.53.

Next, the second-level splits are determined for both Region A and Region B. For Region A, candidate thresholds include: $x_1 < 1.5$, yielding an SSE of 0.04,

and $x_2 < 1.5$, yielding an SSE of approximately 0.85. The former results in a significantly reduced SSE compared to the original 1.035 and is thus chosen as the second-level split for Region A. For Region B, the candidate thresholds are: $x_1 < 3.5$: SSE = 3.98, $x_1 < 4.5$: SSE = 0.84, and $x_2 < 1.5$: SSE = 9.08. The best split for Region B occurs at $x_1 < 4.5$, which reduces the SSE from 6.53 to 0.84 and is therefore selected. After two layers of splitting, the regression tree partitions the input space by discretising the continuous feature x_1 into the following non-overlapping intervals:

1. Leaf 1: $x_1 < 1.5$;
2. Leaf 2: $1.5 \leq x_1 < 2.5$;
3. Leaf 3: $2.5 \leq x_1 < 4.5$;
4. Leaf 4: $x_1 \geq 4.5$.

It this example, the feature x_2 is not used in any split. This indicates that the feature x_2 does not provide a significant reduction in error under the current tree depth constraint.

4.1.3 Parameter learning in BNs

CPTs in BNs are typically constructed using two main approaches: knowledge-based expert assessment and data-driven statistical estimation. When sufficient data are available, conditional probabilities are estimated using methods such as maximum likelihood estimation (MLE) or Bayesian estimation with prior distributions. Conversely, when data are limited or incomplete, domain experts provide initial probability values based on their knowledge and experience. These expert-defined CPTs can then be refined using limited observational data through methods such as the Expectation-Maximisation (EM) algorithm or other probabilistic

inference approaches (Fenton and Neil, 2018).

MLE is the most fundamental approach to parameter learning in BNs, relying solely on observed data without incorporating prior distributions. It assumes that the dataset is complete, meaning that the values of all variables in the network are fully observed for each instance. The goal of MLE is to find the set of parameters that maximizes the likelihood of the observed dataset under the assumed network structure. For a variable X, the conditional probability $\hat{P}(X \mid \mathrm{Pa}(X))$ estimated using MLE is given by:

$$\hat{P}(X = x \mid \mathrm{Pa}(X) = \mathbf{p}) = \frac{\sum_i \mathbb{I}\big(X^{(i)} = x \ \wedge \ \mathrm{Pa}(X)^{(i)} = \mathbf{p}\big)}{\sum_i \mathbb{I}(\mathrm{Pa}(X)^{(i)} = \mathbf{p})}, \tag{4.17}$$

where $\mathbf{p}$ is specific states or values of $\mathrm{Pa}(X)$, x is a specific state or value of X, i denotes the i-th sample in the dataset, $\mathbb{I}(\cdot)$ is the indicator function that equals one if the condition is satisfied and zero otherwise, and $\wedge$ represents the logical AND operation.

While MLE provides a straightforward way to estimate parameters from fully observed data, it often yields unreliable estimates when data is sparse or when specific configurations of variables are not present in the dataset. Bayesian Estimation offers a more robust alternative by incorporating prior knowledge into the learning process through the use of probability distributions over the parameters themselves. In the Bayesian framework, each conditional probability table associated with a variable X_i and its parent configuration $\mathrm{Pa}(X_i) = \boldsymbol{p}$ is treated as a multinomial distribution whose parameters are modelled as random variables. A conjugate prior for the multinomial distribution is the Dirichlet distribution, denoted $\mathrm{Dir}(\cdot)$. Let $x_1, x_2, \ldots, x_r$ denote the possible discrete states of X_i. Then the prior distribution over the conditional probability vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_r]$,

where $\theta_k = P(X_i = x_k \mid \text{Pa}(X_i) = \boldsymbol{p})$, is given by:

$$\boldsymbol{\theta} \sim \text{Dir}(\alpha_1, \alpha_2, \ldots, \alpha_r), \tag{4.18}$$

where $\alpha_k > 0$ are hyperparameters encoding prior beliefs about the distribution of X_i given its parent configuration. After observing the data, the posterior distribution remains Dirichlet, with updated parameters $\alpha_k + N_k$, where N_k is the number of times $X_i = x_k$ and $\text{Pa}(X_i) = \boldsymbol{p}$ occur together in the dataset. The posterior mean, which is typically used as the Bayesian estimate, is given by:

$$\hat{P}(X_i = x_k \mid \text{Pa}(X_i) = \boldsymbol{p}) = \frac{N_k + \alpha_k}{N + \sum_{j=1}^{r} \alpha_j}, \tag{4.19}$$

where $N = \sum_{k=1}^{r} N_k$ is the total number of times $\text{Pa}(X_i) = \boldsymbol{p}$ occurs in the dataset. This estimate is effectively a smoothed version of the MLE, where the prior counts α_k are added to the observed counts N_k. In practice, symmetric priors with $\alpha_k = \alpha$ for all k are often used when no strong prior preference exists. The Bayesian approach prevents zero probabilities and produces more stable estimates, particularly when the data is limited.

In many real-world applications, data is incomplete or contains hidden (latent) variables that are not directly observed. In such cases, neither MLE nor standard Bayesian Estimation can be directly applied. The EM algorithm provides a principled framework for parameter learning under these circumstances. EM is an iterative optimisation technique that alternates between inferring missing information based on current parameter estimates and updating the parameters based on these inferences. Let $\boldsymbol{X}$ denote the observed variables and $\boldsymbol{Z}$ the hidden or missing variables. The EM algorithm seeks to maximize the marginal likelihood $P(\boldsymbol{X} \mid \boldsymbol{\theta})$ by iteratively optimising a lower bound on this likelihood. At iteration t, the algorithm performs the following two steps.

E-step (Expectation step): Compute the expected value of the complete-data log-likelihood, with respect to the conditional distribution of the missing data given the observed data and the current parameter estimates $\boldsymbol{\theta}^{(t)}$. This expectation is given by:

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = \mathbb{E}_{\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{(t)}}[\log P(\boldsymbol{X}, \boldsymbol{Z} \mid \boldsymbol{\theta})], \tag{4.20}$$

where $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ is the expected complete-data log-likelihood. It quantifies how well the parameters $\boldsymbol{\theta}$ explain the complete data (both observed and unobserved), under the assumption that the current model $\boldsymbol{\theta}^{(t)}$ describes the hidden part $\boldsymbol{Z}$ of the data.

M-step (Maximization step): Maximize the expected complete-data log-likelihood computed in the E-step to obtain new parameter estimates:

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}). \tag{4.21}$$

This process is repeated until convergence, typically when the change in the likelihood or parameter values falls below a predefined threshold.

4.2 BN-based TC loss model for regional buildings

4.2.1 Construction of the BN-based TC loss model

The primary purpose of a BN-based TC loss model is to estimate the spatial distribution of loss based on fundamental TC parameters, without performing the scenario-based probabilistic risk assessment described in Section 3.1. The model consists of two main components: TC-related input data and damage-related output data. To capture the spatial distribution of damage, the study area is divided into multiple geographic subdivisions. Common subdivision units include ZIP codes, census tracts, and block groups (FEMA, 2014), which generally

decrease in size in that order. Alternatively, subdivisions may be custom-defined using algorithms that consider regional characteristics, such as SKATER (Teixeira et al., 2015).

TC parameters include longitude, latitude, translation speed, heading angle, central pressure (P_c), maximum wind speed ($V_{\max}$), radius to maximum winds ($R_{\max}$), and the Holland parameter (B), among others. These variables can be grouped into three categories: geographical location, motion vector, and intensity. Selecting appropriate TC variables is essential to capture key characteristics while avoiding redundancy and controlling model complexity. In this study, five parameters are selected: the distance and azimuth of the TC relative to the study area, translation speed, heading angle, and central pressure. The first two describe the TC's geographical location. Polar coordinates are used instead of longitude and latitude because they more effectively represent the TC's position relative to the study area. Translation speed and heading angle together define the TC's motion vector. The final parameter, P_c, represents TC intensity and the potential extent of its impact. P_c is selected because it can be used to estimate other intensity-related parameters. For instance, the relationship between $V_{\max}$ and P_c is commonly expressed as $V_{\max} = m\,(P_{\text{ref}} - P_c)^n$ (Harper, 2002), where m and n are regression coefficients and P_{ref} denotes ambient atmospheric pressure. Previous studies have also established relationships between P_c and $R_{\max}$, as well as between P_c and B (Vickery and Wadhera, 2008; FEMA, 2014). Thus, including P_c in the loss model effectively reflects the intensity characteristics of TCs.

In this study, the CPT of the TC loss model represents the probabilistic relationship between TC characteristics and building losses. It is constructed through three steps: (1) generating a set of TC scenarios defined by combinations of the five TC parameters introduced earlier; (2) calculating the regional loss ratio for

each scenario using the probabilistic risk assessment framework described in Section 3.1; and (3) employing these TC scenarios and corresponding regional losses as training data for parameter learning via MLE (shown in Eq. (4.17)), thereby constructing the CPT. This training procedure allows the BN to directly learn the statistical dependence between TC characteristics and regional losses, effectively transferring the accuracy of the probabilistic risk assessment framework into a computationally efficient surrogate model. Once trained, the BN can rapidly update loss predictions under new TC conditions while retaining the uncertainty and correlation structure captured in the training data.

4.2.2 Bayesian inference under partial loss information

One of the key advantages of BNs is their ability to perform probabilistic inference, allowing unknown variables to be estimated based on available evidence. In the context of TC-induced losses, this capability enables the estimation of loss statistics in regions where loss information is unavailable by making use of loss observations from other regions. This reverse inference is particularly useful in rapid post-event assessments, where detailed hurricane hazard information may be uncertain or still under evaluation, while preliminary loss reports from some regions become available earlier. In such situations, the BN can utilise the observed loss data to update the underlying storm characteristics and subsequently infer losses in unobserved regions, thereby supporting timely decision-making when information is incomplete.

When loss data are available for only a subset of regions, the loss ratios in unobserved regions are updated through Bayesian inference, as expressed by Eq. (4.4). To illustrate this process, consider a simplified example in which the TC parameter vector $\boldsymbol{H}$ can take two discrete configurations, $\boldsymbol{H}^1$ and $\boldsymbol{H}^2$. Each configuration represents a different combination of TC characteristics (e.g., $\boldsymbol{H}^1$ may correspond

to a strong, slow-moving storm, while $\boldsymbol{H}^2$ represents a moderate, fast-moving storm). A uniform prior distribution is assumed: $P(\boldsymbol{H}^1) = 0.5, \quad P(\boldsymbol{H}^2) = 0.5$. Suppose there are two affected regions, A and B. The loss ratio in region A is observed ($LR_A = 0.3$), whereas the loss ratio in region B is unobserved. Our objective is to compute the posterior probability that $LR_B = 0.6$ given the observed $LR_A = 0.3$. The conditional likelihoods under each TC configuration are assumed as:

$$\begin{aligned} P(LR_A = 0.3 \mid \boldsymbol{H}^1) &= 0.2, \qquad P(LR_B = 0.6 \mid \boldsymbol{H}^1) = 0.3, \\ P(LR_A = 0.3 \mid \boldsymbol{H}^2) &= 0.4, \qquad P(LR_B = 0.6 \mid \boldsymbol{H}^2) = 0.5. \end{aligned} \tag{4.22}$$

First, the marginal probability of the observed loss ratio in region A is calculated as:

$$\begin{aligned} P(LR_A = 0.3) &= P(\boldsymbol{H}^1) \cdot P(LR_A = 0.3 \mid \boldsymbol{H}^1) + P(\boldsymbol{H}^2) \cdot P(LR_A = 0.3 \mid \boldsymbol{H}^2) \\ &= 0.1 + 0.2 = 0.3. \end{aligned} \tag{4.23}$$

Next, Eq. (4.4) is applied to compute the posterior probability of interest:

$$\begin{aligned} P(LR_B = 0.6 \mid LR_A = 0.3) = &\sum_{\boldsymbol{H} \in \{\boldsymbol{H}^1, \boldsymbol{H}^2\}} \Bigg(P(LR_B = 0.6 \mid \boldsymbol{H}) \\ &\cdot \frac{P(\boldsymbol{H}) \cdot P(LR_A = 0.3 \mid \boldsymbol{H})}{P(LR_A = 0.3)} \Bigg) \\ = &\frac{0.3 \cdot 0.1 + 0.5 \cdot 0.2}{0.3} = \frac{0.13}{0.3} \approx 0.433. \end{aligned} \tag{4.24}$$

This result indicates that, based on the observed loss ratio in region A, the inferred probability that the loss ratio in region B equals 0.6 is approximately 43.3%. In practice, this approach can be extended to more realistic TC parameter distributions, a larger number of regions, and continuous-valued variables by discretisation techniques.

4.3 Discretisation of continuous variables in TC loss models

4.3.1 Supervised discretisation for TC variables

The discretisation of TC variables directly affects the predictive accuracy of regional losses. Supervised discretisation methods are generally more effective than manual or unsupervised approaches because they preserve the predictive relationship between inputs and outputs (Fenton and Neil, 2018; Chen et al., 2024). Standard supervised discretisation is typically guided by a single output variable. In this study, multiple output variables are considered, specifically the losses associated with individual subdivisions. Discretising input variables separately for each output can result in inconsistent schemes across outputs. Conversely, combining all outputs into a single aggregated output during discretisation treats each subdivision equally, without considering differences in their relative importance, which may not reflect actual priorities. Some subdivisions, such as those with higher building density or greater economic value, should be given greater weight in the discretisation process.

PCA is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional representation while retaining as much of the original variance (i.e., information) as possible. The standard procedure involves three steps. First, the original variables are standardised to have a zero mean and a variance of one to ensure comparability across different scales. Second, the loadings of the original variables are calculated, representing the contribution of each variable to the principal components. Finally, the number of principal components is determined using criteria such as the cumulative explained variance (Greenacre et al., 2022). In this study, PCA is used not only for dimensionality reduction but also to incorporate the relative importance of the original variables. To achieve

this, a weighting factor, $\sqrt{\omega}$, is applied to the standardised loss ratio. This adjustment scales the variance of each loss variable so that subdivisions with higher importance exhibit greater variance. More information from these variables is therefore retained during the PCA process. This procedure is expressed as:

$$L' = \sqrt{\omega} \cdot \frac{L - \mu}{\sigma}, \tag{4.25}$$

where L' is the weighted and standardised loss variable for a subdivision, and μ and σ denote the mean and standard deviation of L, respectively.

The principal components, denoted by $\mathbf{Y}$, are then computed from the adjusted loss ratios of all subdivisions, $\mathbf{L}'$, using PCA:

$$\mathbf{Y} = \mathbf{L}'\mathbf{Q}, \tag{4.26}$$

where $\mathbf{Q}$ is the loading matrix, which serves as a projection matrix that maps the original data into the PCA space. The number of principal components is selected based on a cumulative explained variance threshold, such as 95%.

Next, these principal components are used to guide the discretisation of TC variables using a decision tree algorithm, which is shown in Section 4.1.2. The number of discretisation intervals for each TC variable depends on both the depth of the decision tree and the variable's importance. Deeper trees generate more intervals, while variables with greater influence on the output variables are allocated more intervals. To determine the appropriate number of intervals, the elbow method can be applied. This graphical approach plots the relationship between the number of intervals and the weighted variance, identifying the point at which additional intervals result in only marginal reductions in variance. The overall process of supervised discretisation for TC variables is illustrated in Fig. 4.2.

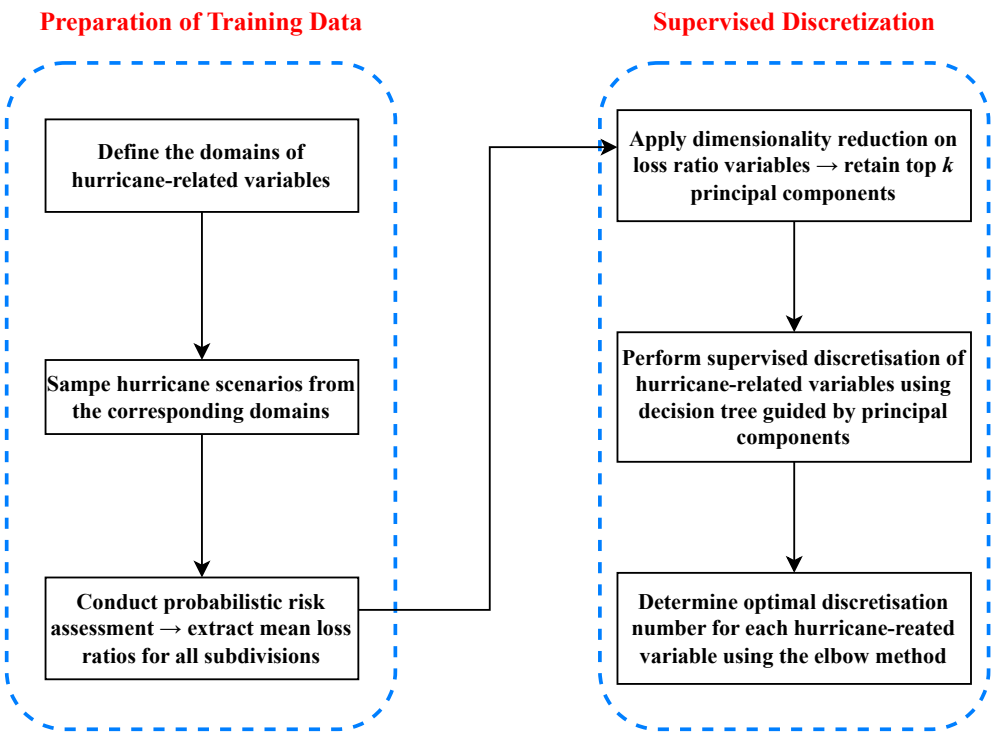


Figure 4.2: Workflow for supervised discretisation of TC-related variables.

4.3.2 Clustering-based discretisation for loss variables

Regional losses are influenced by factors such as structural types, terrain, and TC characteristics. Losses across different regions may be correlated, with adjacent areas typically exhibiting stronger spatial correlations, as shown in Section 3.4. Therefore, the discretisation of loss variables should capture these regional loss characteristics. Traditional methods, such as expert judgment or simple schemes like equal-width and equal-frequency discretisation, are inadequate due to the wide range of influencing factors and may lead to suboptimal outcomes.

To address these challenges, this study adopts a more flexible approach: K-means clustering. This method partitions the principal components into K clusters by iteratively updating cluster centroids to minimise the within-cluster variance, thereby capturing the underlying data structure. Conducting K-means clustering

in the PCA space, rather than on individual loss variables, eliminates the need to cluster each variable separately, which reduces computational complexity and enables the global loss patterns represented by the principal components to be used to capture the overall characteristics of community damage.

The optimal number of clusters, K, is determined using the elbow method, as introduced in Section 4.3.1. The resulting cluster centroids, $\mathbf{C}$, are then mapped back to the original loss variable space, denoted as $\mathbf{C}'$, using the inverse PCA transformation:

$$\mathbf{C}' = (\mathbf{C}\mathbf{Q}^{\mathrm{T}})\boldsymbol{\Omega}^{-\frac{1}{2}}\mathbf{D} + \boldsymbol{\mu}, \tag{4.27}$$

where $\boldsymbol{\Omega}$, $\mathbf{D}$, and $\boldsymbol{\mu}$ represent the weight, standard deviation, and mean of the original loss variables, respectively. Discretisation thresholds are then determined as the midpoints between the centroids of adjacent clusters in $\mathbf{C}'$. Fig. 4.3 illustrates the discretisation process of the loss-related variables. The black text represents the same procedure used for supervised discretisation of TC variables, as described in Section 4.3.1. The blue text indicates the main procedures for discretising loss variables using a clustering-based approach.

4.4 Example: BN-based TC loss model for Miami-Dade County

This study continues to focus on Miami-Dade County as the study region. Miami-Dade County contains 706 census tracts, which are used as the spatial units for loss estimation in this study. Census tracts are selected because they generally have relatively uniform population sizes and building counts, represent an intermediate spatial scale (larger than block groups but smaller than ZIP code areas) and have clearly defined boundaries. These characteristics make census tracts widely used in regional risk assessments (Salim et al., 2024; Rouhana et al., 2025). To reduce

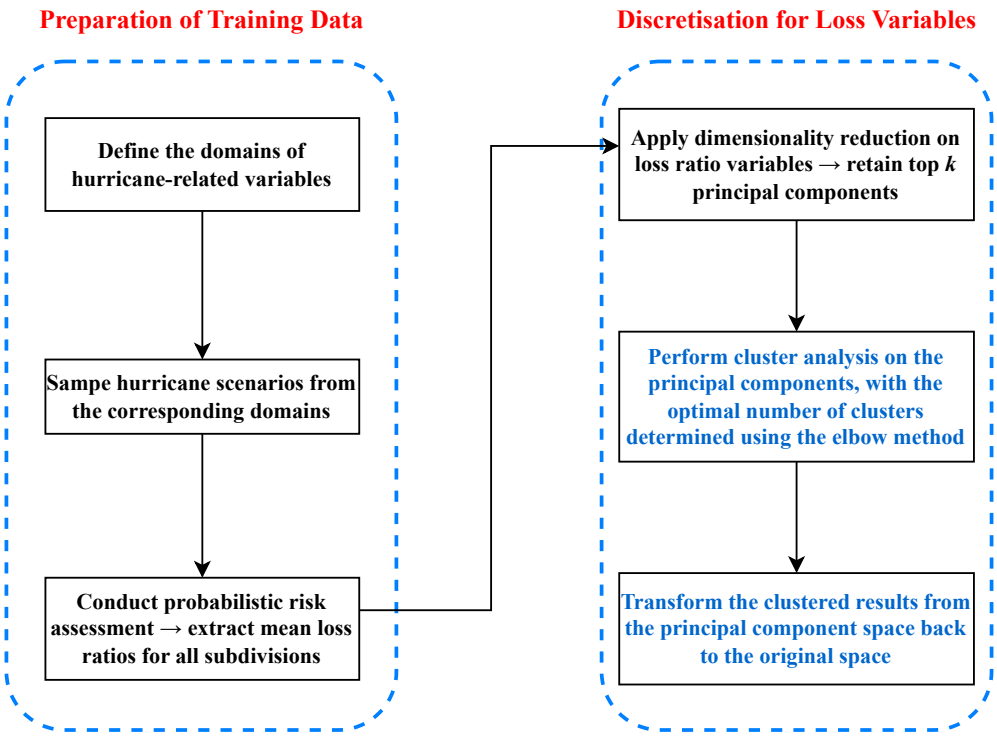


Figure 4.3: Flowchart of the clustering-based discretisation process for loss variables.

computational demands, this study focuses on the 608 tracts that contain more than 10 residential buildings for constructing the BN-based TC loss model.

The gradient wind field, wind speed conversion, parameter selection, and wind speed uncertainty and correlation follow the same procedures described in Section 3.4. Because census tracts are defined with consideration of homogeneity in living conditions (Krieger et al., 2002), structural capacity correlation is assumed to exist among buildings within the same census tract. The correlation coefficient is set to 0.7 for buildings of the same type and 0.4 for buildings of different types (Zeng et al., 2023).

Using the parameters defined in Eq. (3.2) through Eq. (3.6) and the procedure described in Section 3.1.1, the mean, standard deviation, and probability distribution of the loss ratio for each census tract are estimated. Given the large

number of buildings, random sampling is employed to improve computational efficiency. Lin and Wang (2016) evaluated different sample sizes for a 6 km^2 study area containing 4,246 buildings and found that the sampling estimates converged to the exact solution when the sample size reached 100. In the present study, the average census tract area is 7.3 km^2, with an average of 557 buildings. Although the average area is slightly larger than that considered by Lin and Wang (2016), the average number of buildings is substantially lower. Accordingly, a sample size of 100 is used for tracts with more than 100 buildings, while all buildings are included when the total count is below 100.

4.4.1 Model construction

The BN-based TC loss model includes five TC variables: distance (D_{st}) and azimuth (A_z) relative to the centroid of the study area, translation speed (V_t), heading angle (θ), and central pressure (P_c), along with 608 loss variables (L_1 to L_{608}). The BN structure is shown in Fig. 4.4.

These five TC variables serve as root nodes in the BN, implying an underlying assumption of mutual independence. To evaluate this assumption, historical TC data within a 100 km radius of the study area (hereafter referred to as nearby TCs) are obtained from HURDAT2 for events occurring after 1900 (HRD/NOAA, 2023). Values of the five variables are extracted from these records and used to compute the correlation matrix:

$$\boldsymbol{\rho} = \begin{bmatrix} 1.000 & -0.189 & -0.181 & -0.191 & -0.073 \\ -0.189 & 1.000 & 0.030 & 0.061 & 0.027 \\ -0.181 & 0.030 & 1.000 & -0.129 & -0.029 \\ -0.191 & 0.061 & -0.129 & 1.000 & -0.050 \\ -0.073 & 0.027 & -0.029 & -0.050 & 1.000 \end{bmatrix}, \tag{4.28}$$

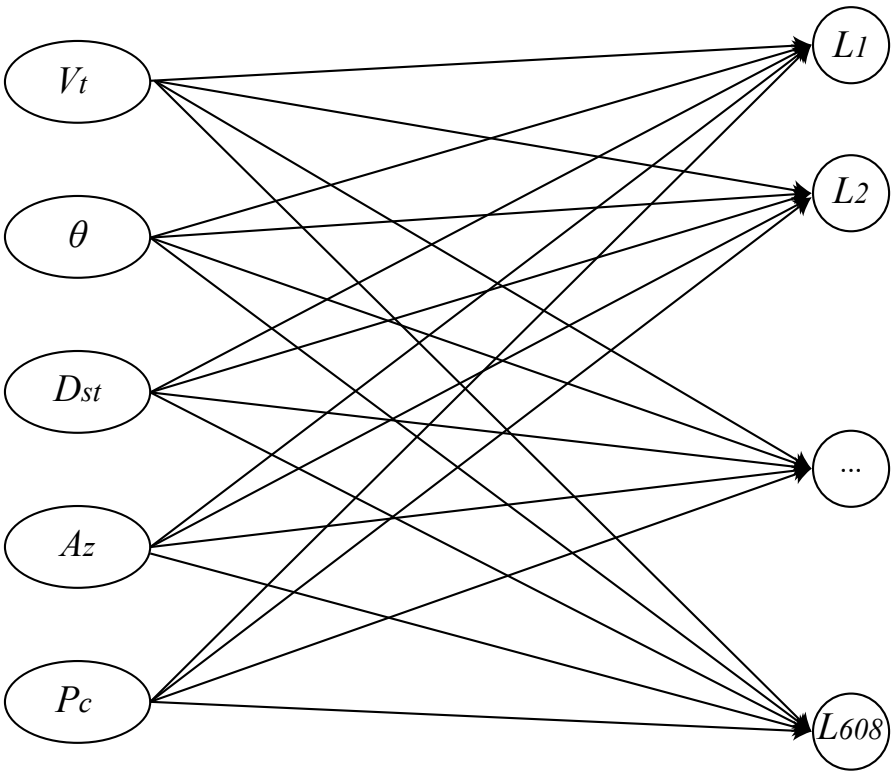


Figure 4.4: Structure of the TC loss model for Miami-Dade County.

where the rows and columns correspond to A_z, D_{st}, V_t, θ, and P_c, respectively. The low correlation coefficients indicate that the assumption of mutual independence is reasonably satisfied.

To prepare for the discretisation of TC and loss variables, a training dataset is generated consisting of TC scenarios and the corresponding mean loss ratios for all census tracts, calculated using the scenario-based probabilistic loss assessment framework. The TC scenarios are randomly sampled from the domains of the five TC variables, which are determined based on historical nearby TCs: 0–15 m/s for V_t, 0–360° for θ, 0–120 km for D_{st}, 0–360° for A_z, and 900–1010 hPa for P_c. To determine an appropriate dataset size, the relationship between sample size and the stability of the discretisation results is evaluated. Sample sizes range from 1,000 to 12,000, in increments of 1,000. The decision tree depth is set to 4, and P_c is

chosen as the observation variable to examine how sample size affects the stability of discretisation. The selection of P_c and a tree depth of four provides a reasonable discretisation level (five splits in this case) and allows for consistent comparison across sample sizes. For each sample size, 10 experiments are conducted, and the normalised root mean square error (RMSE) of the discretisation results for P_c is calculated. Previous studies suggest that a normalised RMSE of 5–10% indicates relatively small variation (Law, 2013). Given the significant impact of discretisation on the BN, a stricter threshold of 3% is applied. The results show that this threshold is reached when the sample size reaches 10,000. Accordingly, a training sample size of 10,000 is adopted for the subsequent discretisation process.

The first step in discretising TC variables is to perform a weighted PCA on the loss variables to reduce dimensionality while preserving the relative importance of each tract in the principal components. In this example, the importance factor is defined by building density, as shown in Fig. 3.3. The threshold for the cumulative explained variance ratio is set to 95%. The cumulative variances of the first ten principal components are shown in Table 4.4. The first five principal components, which collectively explain 96.8% of the variance, are selected for subsequent analysis. The PCA process significantly reduces dimensionality because of substantial redundancy in the loss data. Fig. 4.5 presents a heatmap of the correlation coefficients among the 608 loss variables, where darker colours indicate stronger correlations. Overall, correlations are relatively high, particularly between geographically adjacent census tracts.

In the second step, these five principal components are used to guide the discretisation of the TC variables through a decision tree, implemented using the DecisionTreeRegressor class in the scikit-learn library. By specifying the tree depth, the algorithm identifies optimal split points for each TC variable, generates

Table 4.4: Cumulative explained variance of PCA components (PCA1–PCA10).

Component	Cumulative variance	Component	Cumulative variance
PCA1	0.7350	PCA6	0.9775
PCA2	0.8686	PCA7	0.9840
PCA3	0.9117	PCA8	0.9873
PCA4	0.9436	PCA9	0.9903
PCA5	0.9681	PCA10	0.9924

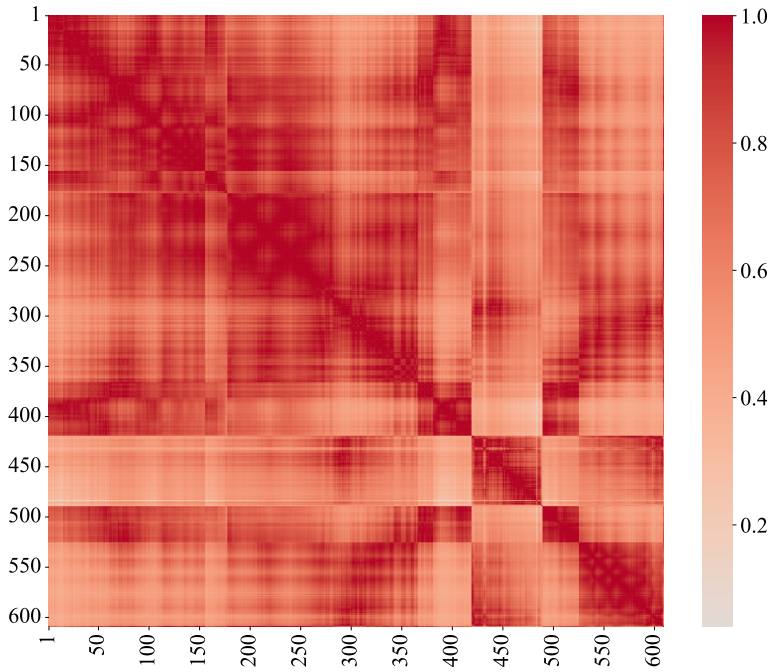


Figure 4.5: Correlation heatmap of mean loss ratios across the 608 census tracts.

the corresponding discretisation thresholds, and determines the total number of bins. Because the decision tree uses binary splitting, the total number of bins approximately follows a 2^n relationship, where n is the tree depth. To select an appropriate discretisation level, Fig. 4.6 plots the weighted variance, defined in Eq. (4.15), against the number of bins, with each point corresponding to a specific tree depth. Both the weighted variance and the number of bins are normalised to the range [0,1] to facilitate identification of the elbow point. Following the method proposed by Satopaa et al. (2011), a straight line is drawn between the first and last points, and the perpendicular distance from each point on the curve to this line is computed. The maximum distance is taken as the optimal discretisation level. At this level, the discretisation numbers and thresholds for each TC variable are shown in Fig. 4.7. The number of bins for A_z, D_{st}, and P_c is substantially higher than for V_t and θ, suggesting that the former three variables have a greater influence on losses.

The discretisation of loss variables is performed by applying K-means clustering to the five principal components, implemented using the KMeans class in the scikit-learn library. Fig. 4.8 shows the number of clusters ranging from 5 to 30 and the corresponding reduction in inertia, defined as the sum of squared distances to the cluster centres. Following the same procedure used to determine the optimal discretisation level for TC variables, the optimal number of clusters for loss variables in this example is 15. The cluster centroids in the PCA space are then mapped back to the original variable space using Eq. (4.27). Discretisation points are determined as the midpoints between adjacent cluster centroids, with 0 and 1 added as the lower and upper bounds of the discretisation intervals, respectively. The discretisation results vary across tracts, and Fig. 4.9 presents the outcomes for L_1 to L_5 as an illustration.

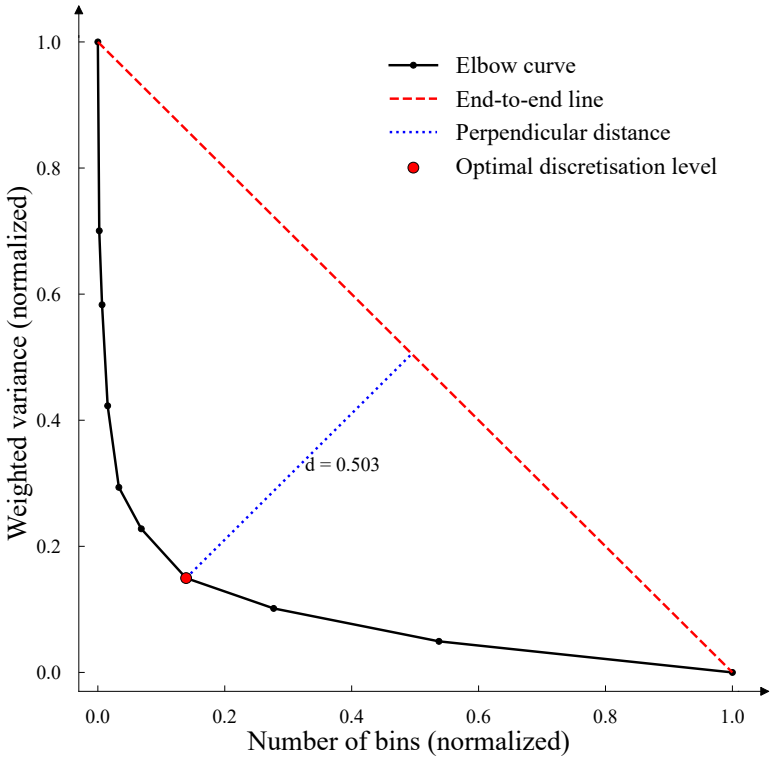


Figure 4.6: Weighted variance versus number of discretisation bins for TC variables.

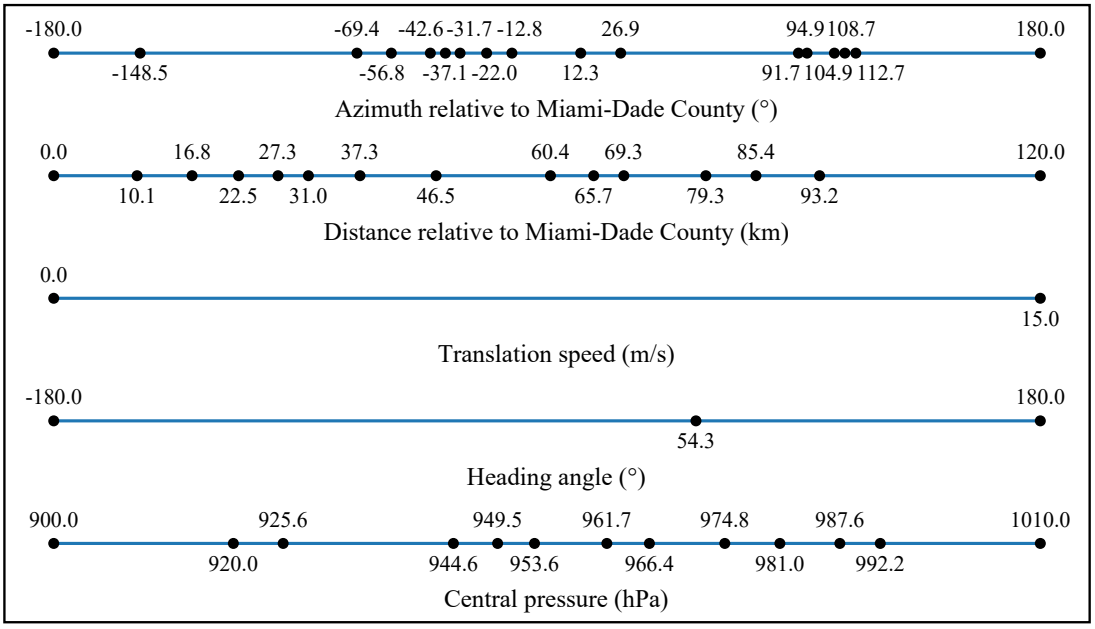


Figure 4.7: Discretisation results of the five TC variables.

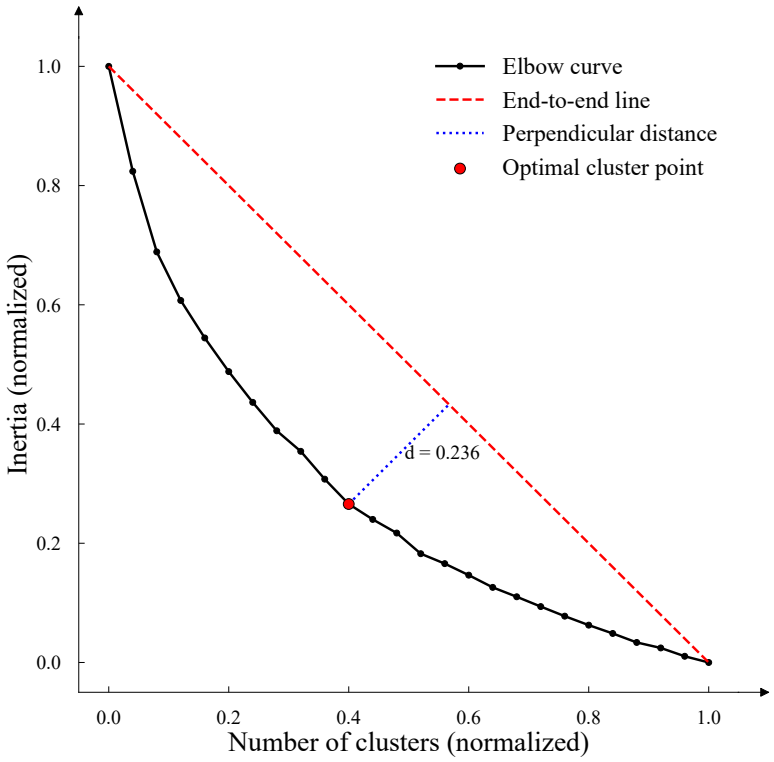


Figure 4.8: Reduction of inertia with increasing number of clusters.

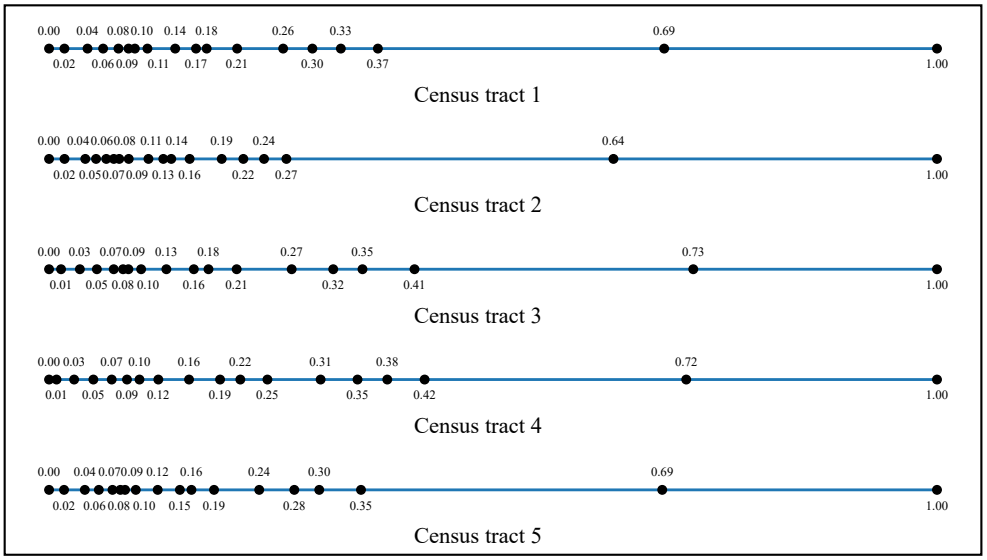


Figure 4.9: Discretisation of loss ratios for census tracts 1–5.

Constructing the CPT for the TC loss model involves two steps. The first step is to determine the marginal distributions of the variables A_z, D_{st}, V_t, θ, and P_c. The second step is to establish the probabilistic relationships between the TC variables and the loss variables. The marginal distributions are derived from historical TC data, as illustrated in Fig. 4.10, where the area of each bin represents the marginal probability of the corresponding discretisation interval. The relationships between TC and loss variables are estimated using maximum likelihood estimation based on the training samples.

4.4.2 Model validation

To evaluate the effectiveness of the TC loss model, this section assesses its predictive performance using the landfall moments of 100 TC events. The selection criteria require TCs to reach at least Category 2 intensity and to pass within 100 km of Miami-Dade County. These thresholds are chosen because post-cyclone reports indicate that noticeable community damage typically begins with Category 2 storms (Wikipedia, 2025c). The 100 km distance ensures the inclusion of TCs with substantial local impact, approximately twice the typical $R_{\max}$ of 40–50 km (Hsu and Yan, 1998). Since 1900, 27 TCs have met these criteria and are included as historical cases, while the remaining 73 events are obtained from a synthetic TC database (Bloemendaal et al., 2020). Fig. 4.11 illustrates the tracks of these 100 TCs, with red lines representing historical events and yellow lines representing synthetic events.

Three representative historical TCs were selected for model verification. The first is Hurricane Great Miami (1926), which made landfall at 12:00 on September 18, 1926. The second is Hurricane Homestead (1945), which made landfall at 20:00 on September 15, 1945. The third is Hurricane Andrew (1992), which made landfall at 08:40 on August 24, 1992. These three TCs are among the most severe

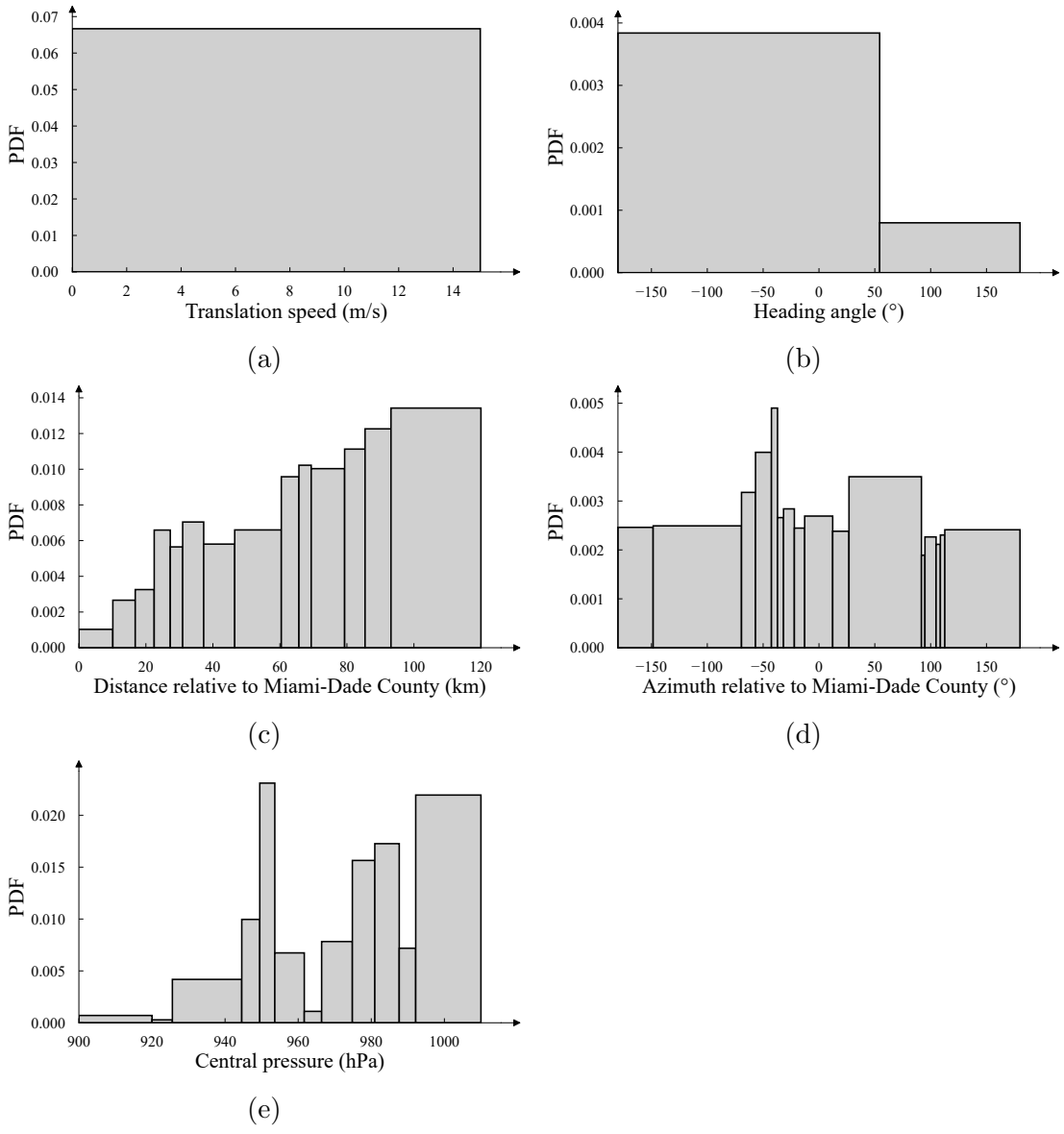


Figure 4.10: Marginal distributions of TC variables: (a) V_t, (b) θ, (c) D_{st}, (d) A_z, and (e) P_c.

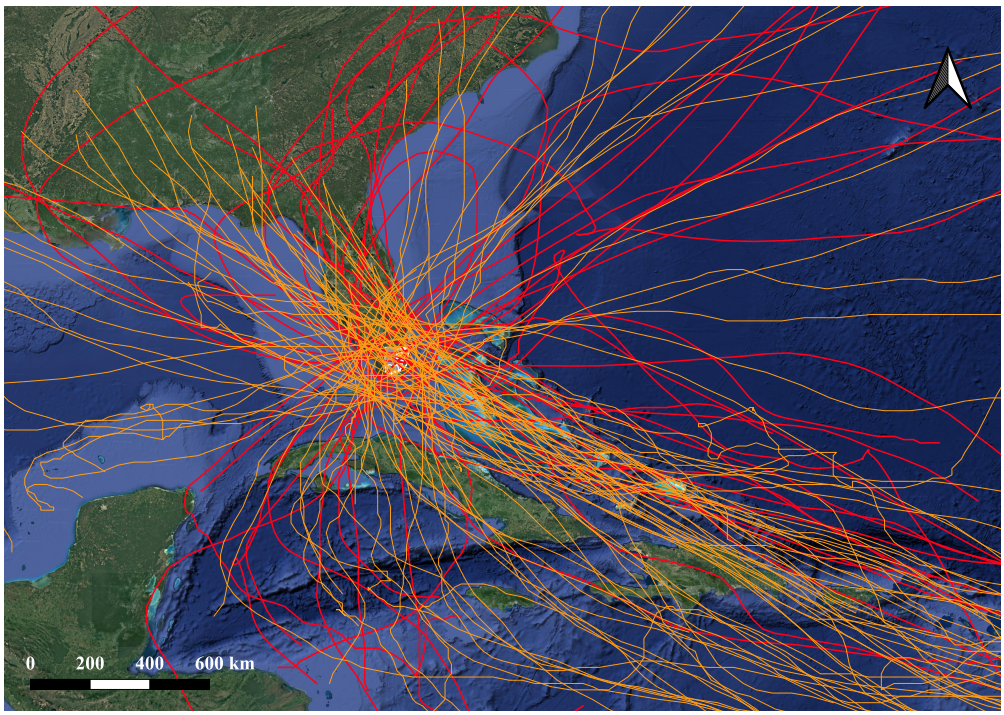


Figure 4.11: Spatial distribution of 100 test TCs used for model validation.

in Florida's history. Fig. 4.12 to Fig. 4.14 illustrate the spatial distribution of the mean and standard deviation of loss ratios in Miami-Dade County during these events. Taking Hurricane Andrew as an example, Figs. 4.14a and 4.14c show the reference values obtained from the scenario-based TC risk assessment framework, while Figs. 4.14b and 4.14d present the corresponding estimates from the BN-based TC loss model. In these figures, darker colours indicate higher loss ratios, and grey areas denote regions with fewer than 10 residential buildings, which are excluded from the analysis. For Hurricane Great Miami (1926), the RMSE across all census tracts is 0.028 for the estimated mean loss ratio and 0.019 for the standard deviation. For Hurricane Homestead (1945), the RMSE values are 0.015 and 0.012, respectively. For Hurricane Andrew (1992), the RMSE values are 0.030 and 0.021, respectively. Overall, the proposed model produces results that closely match the reference values for both the mean and standard deviation of

loss ratios.

Fig. 4.15a and Fig. 4.15b present the RMSE distributions for the mean and standard deviation of loss ratios across 608 census tracts, based on 100 test TCs. The results demonstrate strong predictive performance of the loss model, with the 90th percentile RMSE values for the mean and standard deviation being 0.025 and 0.023, respectively. However, some census tracts exhibit noticeably higher RMSE values, likely due to the weighting factor (in this example, building density) used in the discretisation process.

To examine this hypothesis, a comparative analysis is performed between census tracts in the top 5% and bottom 5% of building density, as shown in Table 4.5. The results indicate that in high-density areas, the average RMSE for both the mean and standard deviation is about 30% lower than in low-density areas. These findings demonstrate that incorporating importance factors into the discretisation process effectively reduces prediction errors in high-priority areas.

Table 4.5: RMSE comparison of loss ratios between high- and low-density census tracts.

Region type	RMSE (mean)	RMSE (standard deviation)
High-density regions (top 5%)	0.0145	0.0117
Low-density regions (bottom 5%)	0.0200	0.0170

4.4.3 Application of Bayesian inference with partial loss information

Having introduced the Bayesian inference formulation in Section 4.2.2, we now demonstrate its practical utility. In this analysis, we consider a scenario in which loss information from only a subset of regions is available, and the TC characteristics are not specified. The BN-based model is then used to infer the loss ratios in the remaining unobserved regions based solely on the available partial loss data.

In practice, post-disaster reports typically provide two summary statistics:

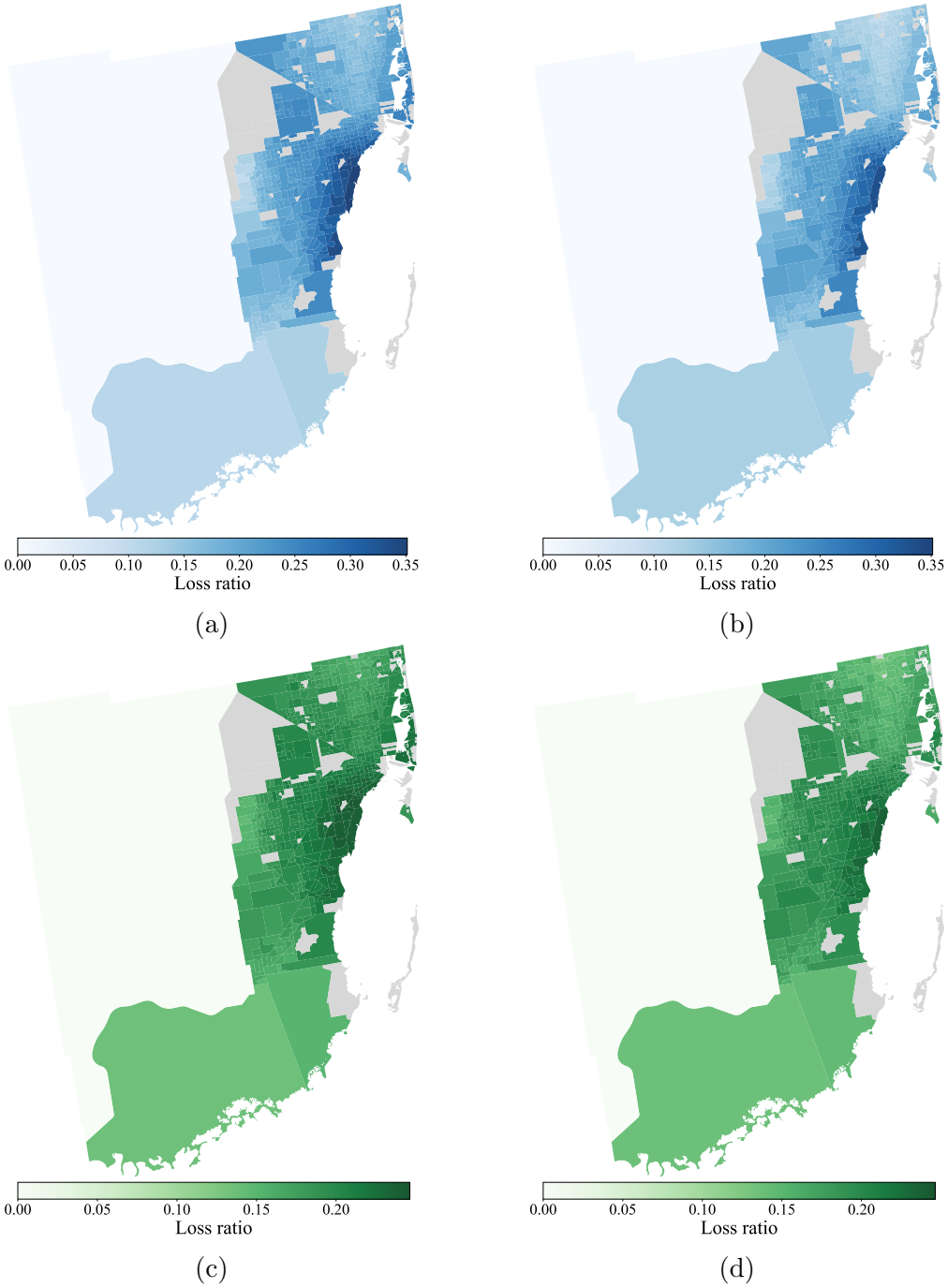


Figure 4.12: Comparison of reference and BN-based model estimated loss ratios for Hurricane Great Miami (1926): (a) mean (reference), (b) mean (BN-based), (c) standard deviation (reference), and (d) standard deviation (BN-based).

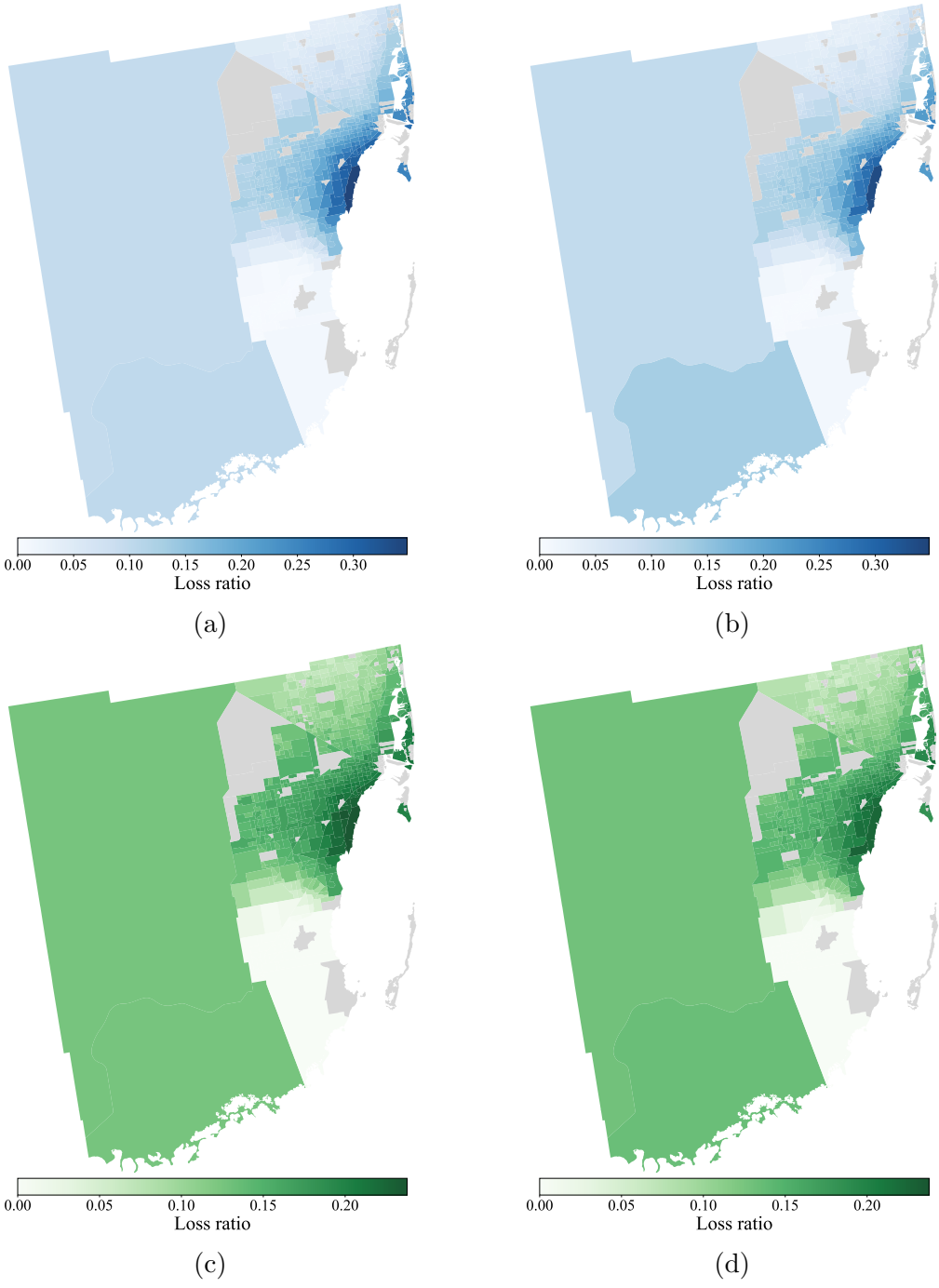
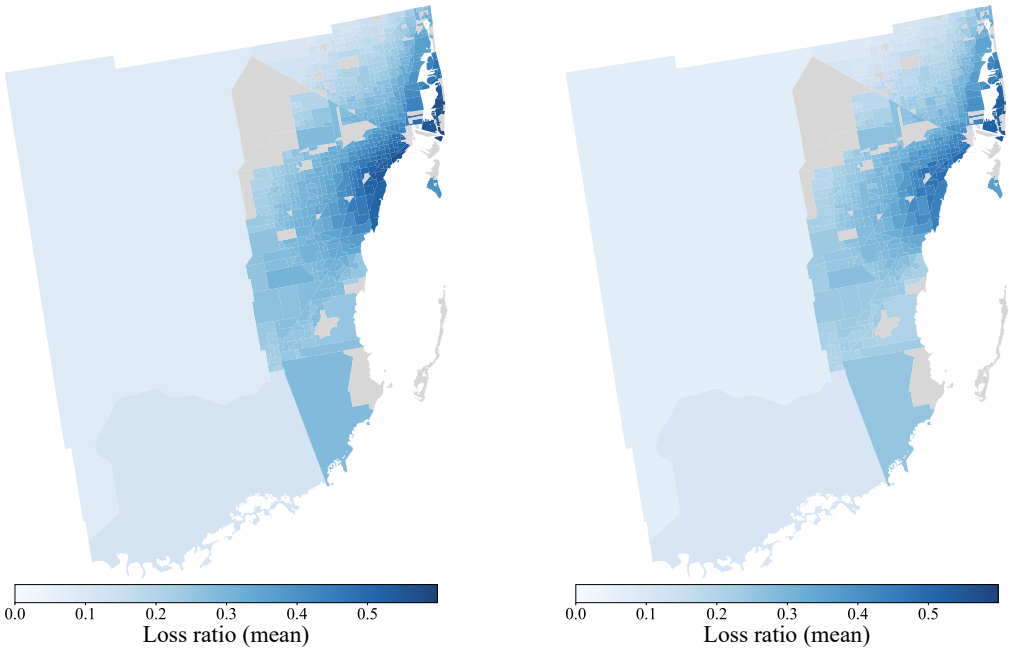
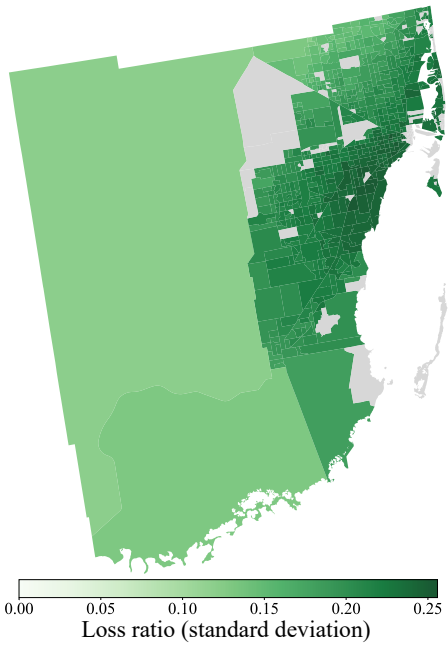


Figure 4.13: Comparison of reference and BN-based model estimated loss ratios for Hurricane Homestead (1945): (a) mean (reference), (b) mean (BN-based), (c) standard deviation (reference), and (d) standard deviation (BN-based).

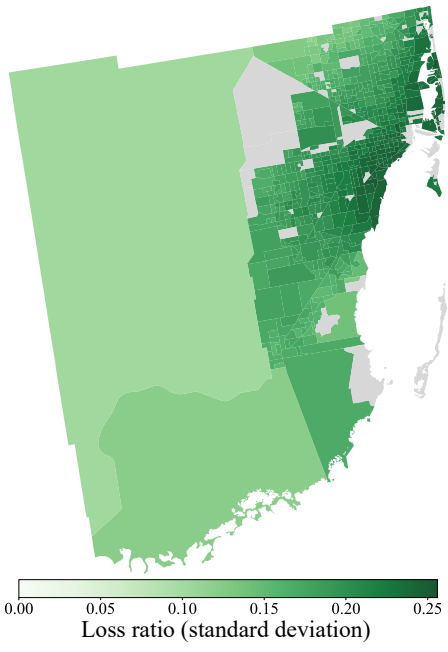


(a)

(b)

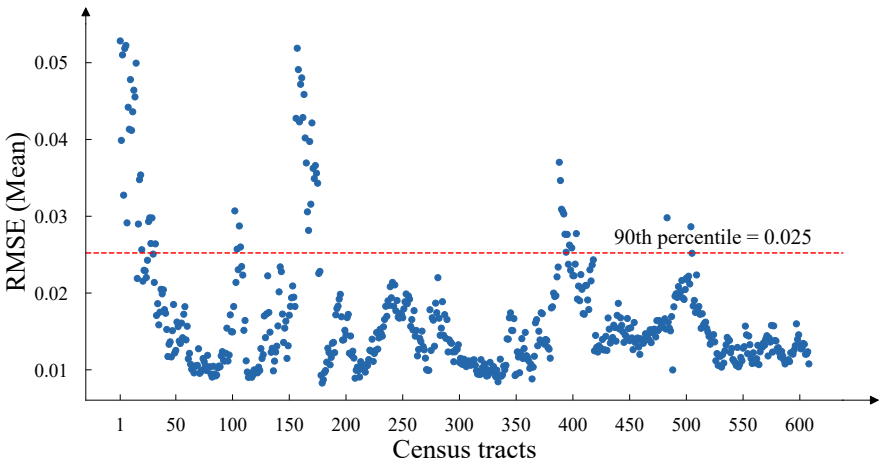


(c)

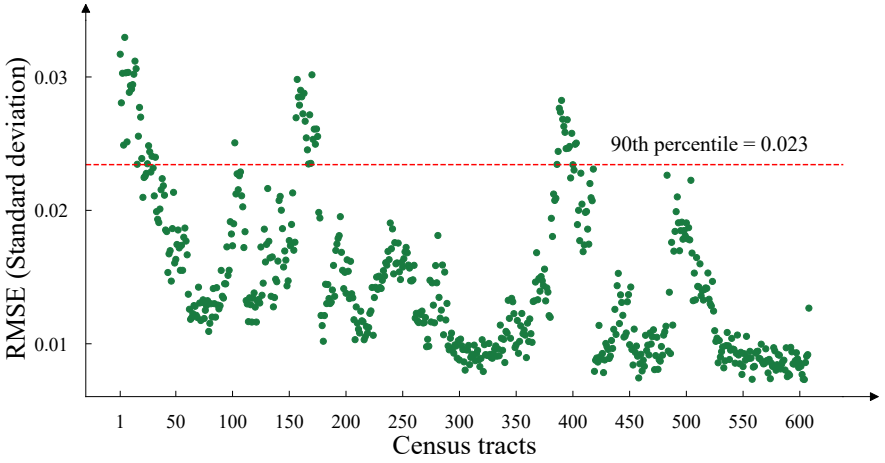


(d)

Figure 4.14: Comparison of reference and BN-based model estimated loss ratios for Hurricane Andrew (1992): (a) mean (reference), (b) mean (BN-based), (c) standard deviation (reference), and (d) standard deviation (BN-based).



(a)



(b)

Figure 4.15: RMSE of predicted loss ratios across 608 census tracts using 100 test TCs: (a) mean, (b) standard deviation.

the mean and the standard deviation of loss ratios. However, since each node in a BN represents a probabilistic variable, the inference process requires a complete description of the underlying probability distribution. To characterise this distribution, the outputs of scenario-based probabilistic risk assessment for three major historical TCs mentioned in Section 4.4.2 were analysed across 608 regions, producing a total of 1,824 regional loss distributions. The results showed that the loss ratios closely followed a Beta distribution. To verify this assumption, the Kolmogorov–Smirnov (K–S) test was applied. The distribution of K–S D-values is shown in Fig. 4.16. Most D-values were below 0.1, indicating a strong agreement between the empirical data and the Beta distribution. Therefore, using the observed mean and standard deviation, along with the Beta distribution, the probability distribution for these regions can be estimated.

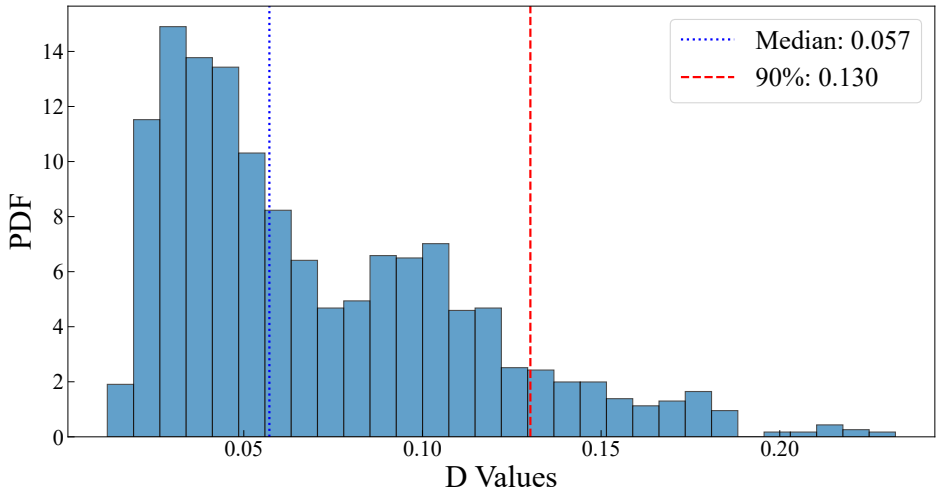
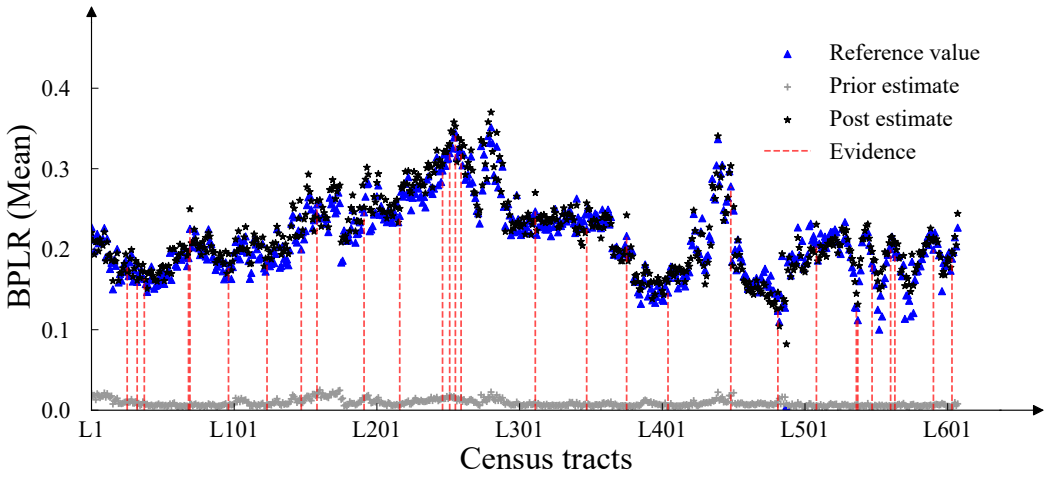


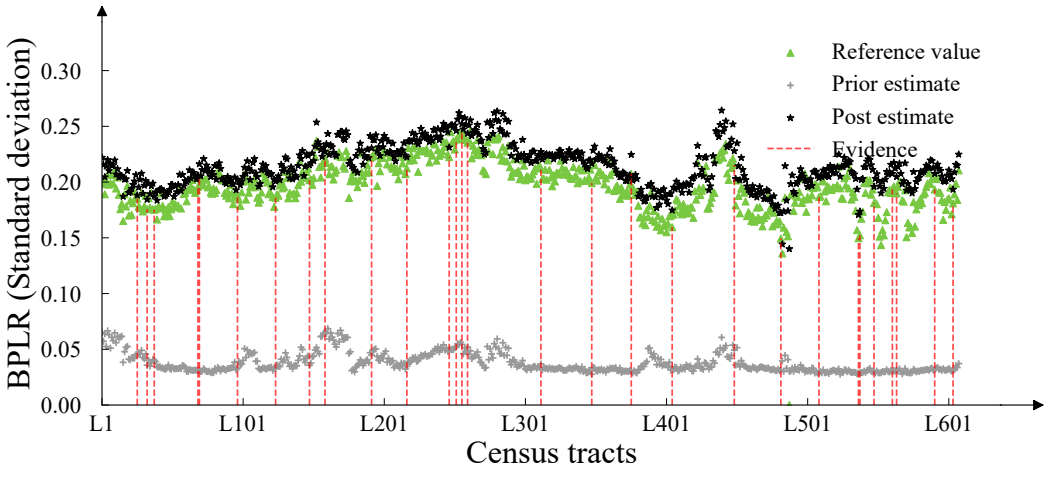
Figure 4.16: Distribution of K-S test D-values for fitting the Beta distribution.

To assess the performance of Bayesian inference, an experiment was conducted in which 5% of census tracts were assumed to have known loss ratios, including both the mean and standard deviation. These data were used to infer the losses in the remaining 95% of the regions. Figs. 4.17–4.19 illustrate the inferred loss

distributions for three historical TCs: the Great Miami Hurricane of 1926, Hurricane Homestead of 1945, and Hurricane Andrew of 1992. In each figure, the light grey points represents the uninformed prior estimate based solely on historical TC patterns, while the dark grey points represents the posterior distribution incorporating the 5% known information. The vertical dashed lines denote the known observations. It is evident that the posterior estimates substantially improve the accuracy of the loss evaluation.

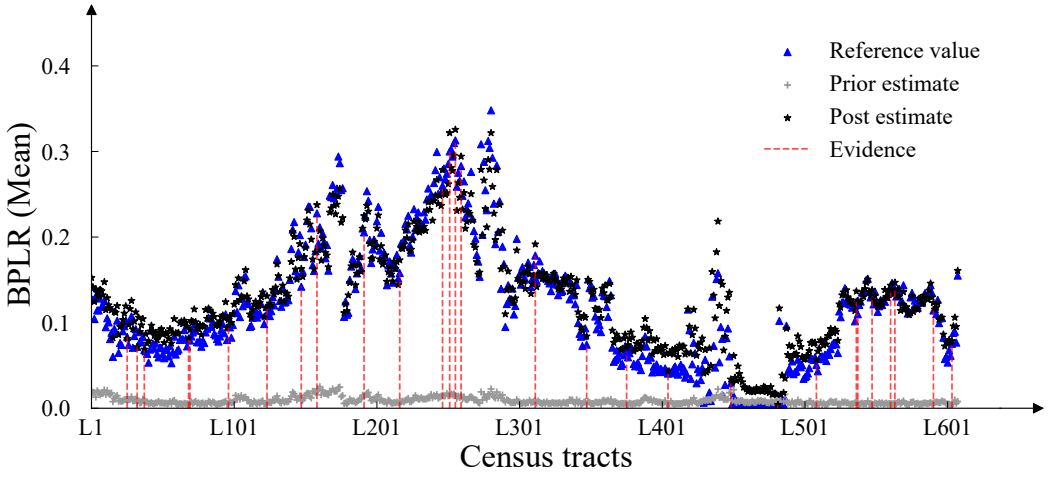


(a)

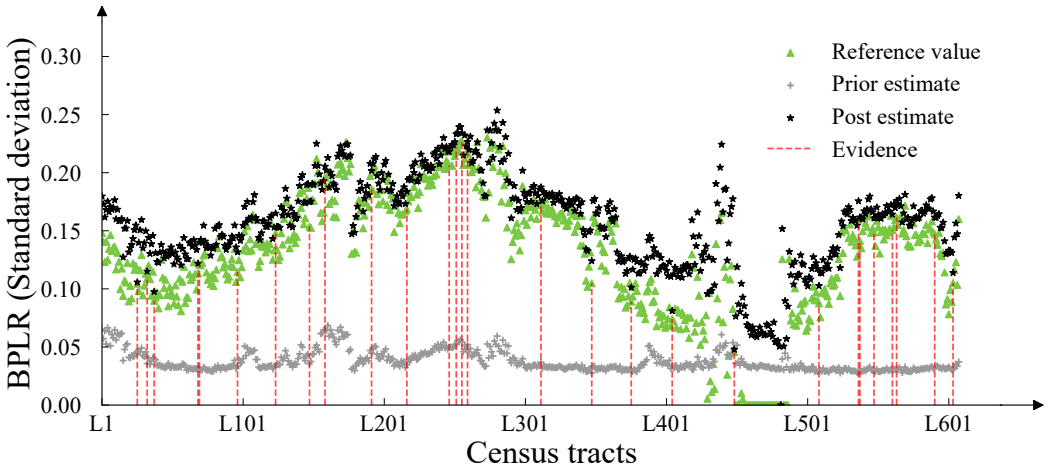


(b)

Figure 4.17: Inference results for Hurricane Great Miami (1926): (a) inferred mean loss ratios; (b) inferred standard deviation of loss ratios.

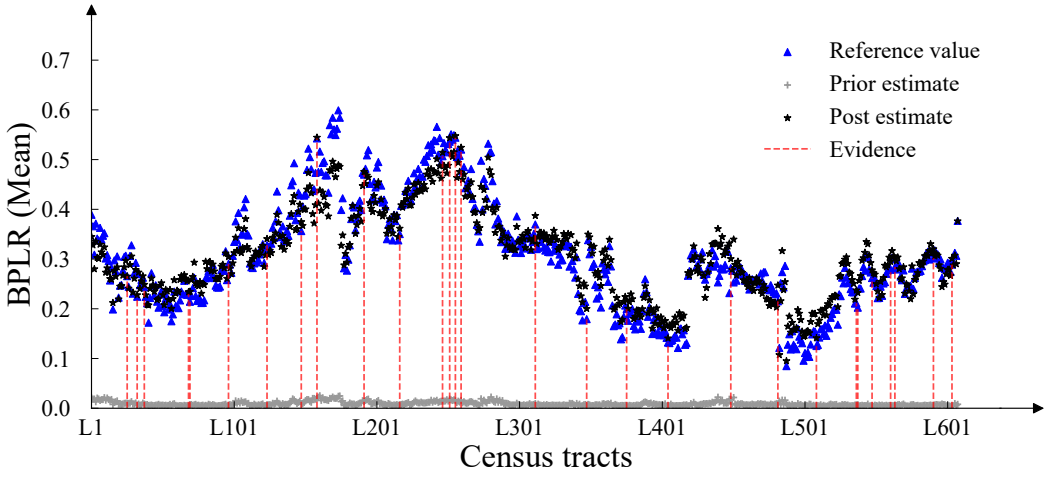


(a)

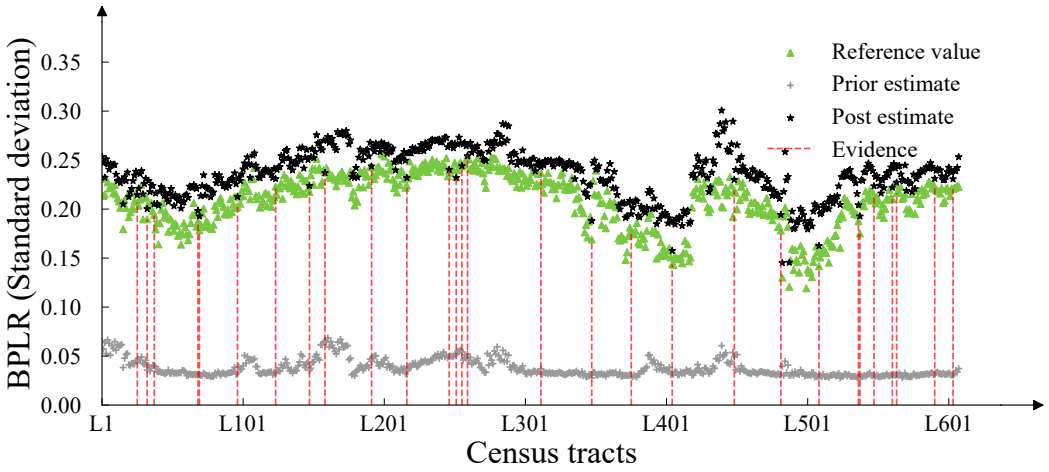


(b)

Figure 4.18: Inference results for Hurricane Homestead (1945): (a) inferred mean loss ratios; (b) inferred standard deviation of loss ratios.



(a)



(b)

Figure 4.19: Inference results for Hurricane Andrew (1992): (a) inferred mean loss ratios; (b) inferred standard deviation of loss ratios.

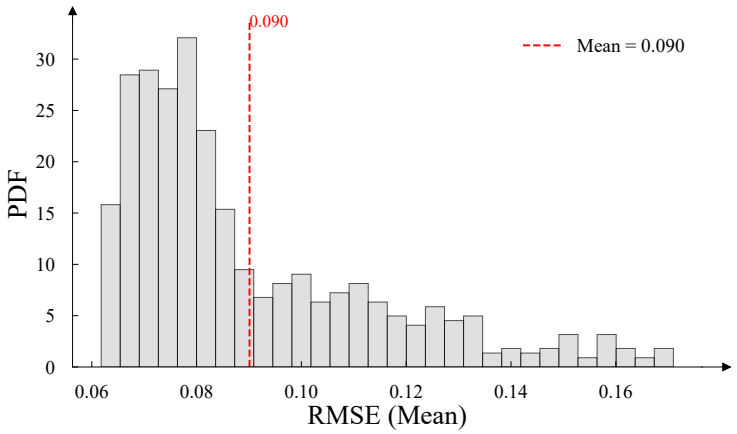
Figs. 4.20 and 4.21 further examine the Bayesian inference results by showing the distributions of the RMSE for the inferred mean and standard deviation of loss ratios based on 100 test TCs. Three scenarios were compared: (1) no known post-disaster information, (2) 5% of census tracts with known mean and standard deviation, and (3) 10% of census tracts with known statistics. When no regional data were available, the model relied solely on historical TC patterns, resulting in the largest estimation errors. Introducing 5% known data significantly reduced the RMSE for both the mean and standard deviation, while increasing the proportion of known data to 10% provided only marginal additional improvements. This suggests that 5% of known data is sufficient to achieve reliable inference.

The above analysis assumes that both the mean and standard deviation of damage can be accurately obtained from post-disaster surveys. In practice, while mean damage estimates are usually easier to derive, estimating the standard deviation is more challenging due to data limitations or small sample sizes. For example, when only a limited number of buildings are inspected, an average damage estimate may be available, but the sample may be too small to provide reliable measures of variability. To overcome this limitation, the relationship between the mean damage and the coefficient of variation is examined across all 608 regions affected by the 100 historical TCs. As shown in Fig. 4.22, the coefficient of variation, y, exhibits a power-law relationship with the mean, which can be expressed as:

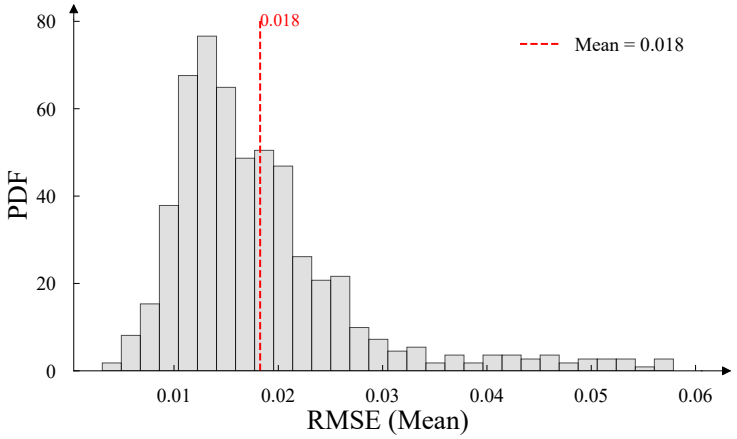
$$y = 0.426 \cdot \bar{x}^{-0.419}, \tag{4.29}$$

where $\bar{x}$ is the mean loss ratio. This empirical relationship enables the estimation of the standard deviation when only mean values are available, thereby allowing the inference process to continue under partial data conditions.

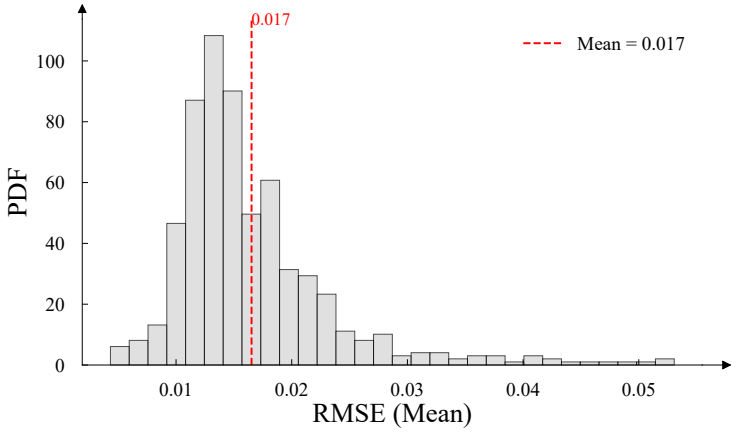
To evaluate the performance of this approach, an additional experiment was



(a)

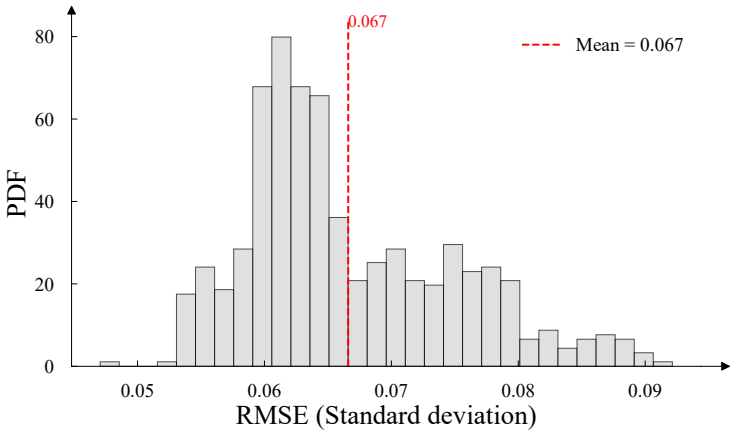


(b)

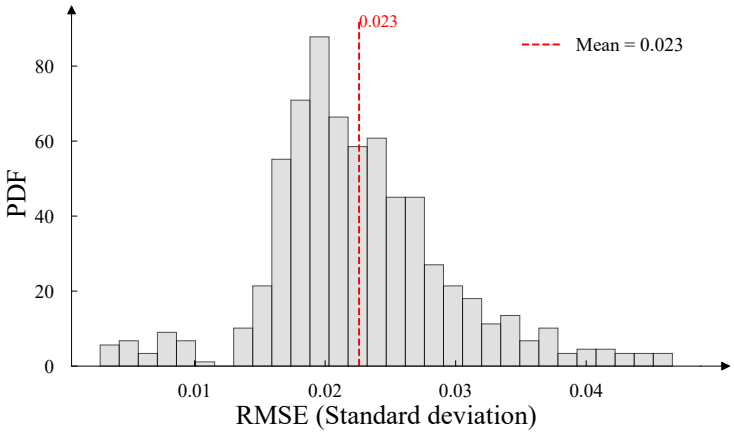


(c)

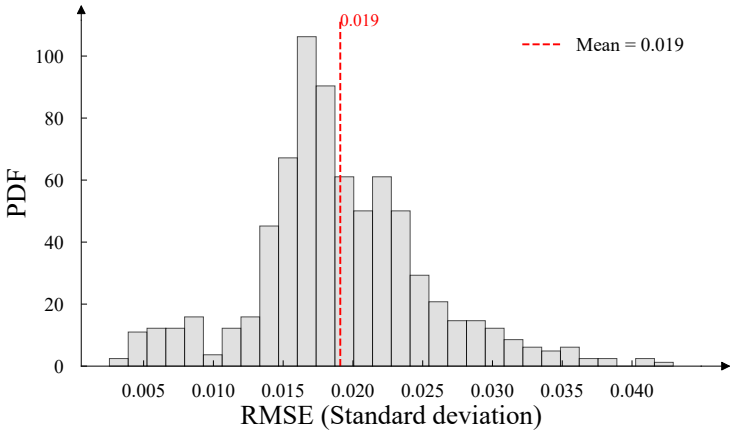
Figure 4.20: RMSE of inferred mean loss ratios under different levels of known observations: (a) no observation; (b) 5% observation; (c) 10% observation.



(a)



(b)



(c)

Figure 4.21: RMSE of inferred standard deviations of loss ratio under different levels of known observations: (a) no observation; (b) 5% observation; (c) 10% observation.

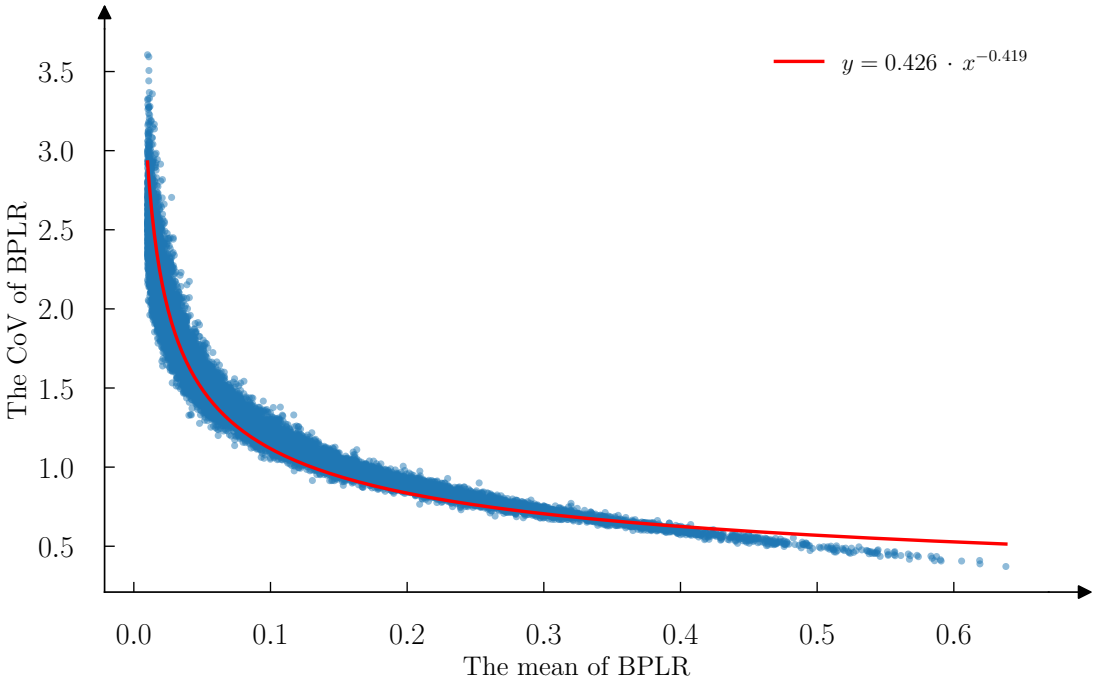


Figure 4.22: Empirical relationship between mean damage and coefficient of variation across 608 regions.

conducted assuming that 5% of regions had known mean damage ratios only. The fitted power-law relationship was used to estimate the corresponding standard deviations, after which Bayesian inference was performed on the remaining 95% of regions. Table 4.6 and Table 4.7 compare the RMSE of inferred mean and standard deviation between the full-information case (both mean and standard deviation known) and the partial-information case (only mean known) for the three historical TCs. The results demonstrate that even under these more limited conditions, the inference remains reasonably accurate, although performance is slightly degraded compared to when both statistics are available.

Table 4.6: The RMSE of inferred mean loss ratios for Miami-Dade County under partial statistical information.

TC name	Both mean and SD	Only mean
The Great Miami (1926)	0.013	0.015
Hurricane Homestead (1945)	0.018	0.020
Hurricane Andrew (1992)	0.029	0.032

Table 4.7: The RMSE of inferred standard deviations of loss ratios for Miami-Dade County under partial statistical information.

TC name	Both mean and SD	Only mean
The Great Miami (1926)	0.019	0.019
Hurricane Homestead (1945)	0.033	0.034
Hurricane Andrew (1992)	0.030	0.031

4.5 Comparative analysis of two TC loss models

The approaches presented in Chapters 3 and 4 both aim to evaluate regional building losses induced by TCs. For ease of reference, the framework in Chapter 3 is denoted as Method 1, and that in Chapter 4 as Method 2. While the two methods share certain conceptual similarities, they differ in terms of data requirements, modelling techniques, and model outputs. This chapter provides a comparative analysis of these approaches across three dimensions: data sources, modelling methodologies, and their respective advantages and limitations.

Data sources. Method 1 primarily relies on statistical datasets, including: (1) spatial data describing building distribution and density, typically obtained from government statistics; (2) topographic and land-surface roughness data used to characterise surface friction and wind attenuation; (3) TC risk metrics, such as return-period wind speeds assessed from long-term synthetic TC simulations; and (4) records of historical TC events. Method 2, in contrast, requires: (1) historical records of TC parameters, such as distance and bearing relative to the study area,

central pressure, translation speed, and heading angle; (2) a set of generated TC scenarios; and (3) computed losses for each spatial unit under these scenarios.

Modelling methodologies. Method 1 is developed using spatial clustering and empirical regression modelling. The process involves three key steps: (1) subdividing the study area into contiguous and relatively homogeneous subregions using the SKATER algorithm, guided by factors such as building density, surface roughness, and TC risk; (2) fitting wind–loss regression functions for each subregion; and (3) aggregating subregional losses to estimate the total regional loss. By employing the maximum wind speed as a predictor, Method 1 enables the estimation of total losses for the entire lifecycle of a TC in a single computation. Method 2, by contrast, is grounded in probability theory and Bayesian inference. The modelling workflow includes: (1) selecting appropriate TC and loss variables; (2) discretising continuous variables using data-driven approaches, including weighted PCA and decision-tree-based supervised discretisation for input variables, and clustering for output variables; (3) generating a set of TC scenarios to represent a full range of possible storm behaviours and computing loss ratios for each scenario using a probabilistic risk assessment framework; and (4) estimating the CPTs through maximum likelihood estimation. The model output is a spatial probability distribution of losses across all subregions. In addition, the inferential capabilities of the BN enable the estimation of losses in unobserved areas based on observed data.

Advantages and limitations. Method 1 can evaluate losses over the full lifecycle of a TC, from genesis to dissipation, or at a specific point in time. Its data requirements are relatively modest, and its model construction process is straightforward, making it suitable for rapid assessments where the primary objective is to estimate total regional losses for a TC event. However, Method 1 cannot capture

spatial variations in damage, as it provides only an aggregated loss value for the entire region. Its accuracy is also sensitive to the quality of spatial partitioning. Moreover, the model assumes that maximum wind speed is the dominant damage indicator, while other influencing factors (e.g., wind duration) are not considered. Method 2, on the other hand, captures both the spatial distribution and probabilistic characteristics of regional losses. Its flexible BN framework allows the representation of complex causal relationships between TC characteristics and resulting losses. Another advantage of Method 2 lies in its scalability; for instance, it could be extended into a dynamic BN by incorporating time-series data, allowing the model to track and predict evolving damage throughout a TC's progression. This dynamic capability would provide valuable insights for emergency planning and real-time decision-making. Furthermore, additional hazard attributes (e.g., wind duration) or TC-related hazards (e.g., precipitation intensity, flood risk) can be integrated to enable multi-hazard risk assessments. Nevertheless, Method 2 entails substantial computational demands and requires extensive, high-quality datasets. In addition, as the model produces a snapshot of the loss distribution at a given moment, assessing losses over an entire TC lifecycle requires multiple computations and a post-processing step to determine either the maximum or cumulative impact.

4.6 Conclusion

This section presents a BN framework for estimating regional building losses caused by TCs. The framework efficiently and accurately quantifies the spatial distribution of losses across large geographic areas under specified TC scenarios. Because the discretisation of continuous variables significantly affects model accuracy, a new discretisation scheme is introduced. The main findings are summarised

as follows:

1. The model incorporates two types of variables: TC variables (including translation speed, heading angle, distance and azimuth of the TC centre relative to the study area, and central pressure) and loss variables (loss statistics for each geographic subdivision). The relationship between TCs and losses is established through a scenario-based probabilistic risk assessment framework.

2. TC variables are discretised using a supervised approach that combines weighted PCA with decision tree algorithms, allowing the relative importance of each subdivision to be incorporated into the discretisation process. Loss variables are discretised using a clustering-based method to capture the characteristics of regional losses.

3. The framework is demonstrated through a numerical example, showing high predictive accuracy. The 90th percentile RMSE values for the mean and standard deviation of loss estimates across the study area are 0.025 and 0.023, respectively. In high-importance areas, prediction errors are approximately 30% lower than in low-importance areas due to the integration of importance factors into the discretisation process.

CHAPTER 5

A GCM–BASED FRAMEWORK FOR ASSESSING CLIMATE CHANGE IMPACTS ON TROPICAL CYCLONE WIND RISK IN COASTAL CITIES

5.1 TC risk assessment under the historical climate in the North Atlantic

TC risk is commonly assessed using return period wind speeds. In ASCE 7-16 (ASCE, 2017), return period wind speeds provide the basis for structural design; for example, mean return intervals (MRIs) of 300, 700, 1,700 and 3,000 years correspond to Risk Categories I–IV, respectively. These wind speeds are typically derived from statistical analyses of tens of thousands of years of simulated TCs generated by synthetic track models. Among the available approaches, the method developed by Vickery et al. (2000) (hereafter the Vickery method) is widely adopted in engineering applications.

In this study, the Vickery method is employed to simulate 30,000 years of synthetic TC activity. On the basis of this dataset, return period wind speeds are evaluated for five selected coastal cities in the United States. The resulting wind hazard estimates serve as the baseline against which the potential influence of future climate change on wind risk is assessed in subsequent sections.

5.1.1 Synthetic TC generation model

The synthetic TC generation model developed by Vickery et al. (2000) is adopted in this study. It is an empirical full-track model that simulates the entire TC lifecycle, including genesis, development, and dissipation. The model operates in two stages: (1) estimation of regression parameters specific to spatial grids across the North Atlantic, and (2) iterative simulation of TC tracks based on these fitted parameters.

In the first stage, the basin is divided into 5° latitude × 5° longitude grids, as shown in Fig. 5.1. Within each grid, the changes of translation speed and heading direction are modelled as:

$$\Delta \ln V_t = a_1 + a_2\psi + a_3\lambda + a_4 \ln V_{ti} + a_5\theta_i + \varepsilon_1, \tag{5.1a}$$

$$\Delta\theta = b_1 + b_2\psi + b_3\lambda + b_4 V_{ti} + b_5\theta_i + b_6\theta_{i-1} + \varepsilon_2, \tag{5.1b}$$

where i denotes the time step (6-hour interval). V_t is the translational velocity, θ is the heading direction (clockwise from north), and ψ and λ are the latitude and longitude of the TC centre, respectively. The error terms ε_1 and ε_2 follow zero-mean Gaussian distributions with grid-specific standard deviations. $\Delta \ln V_t$ and $\Delta\theta$ represent changes between the two successive time steps. Separate sets of coefficients, a and b, are estimated for eastward- and westward-moving storms to account for directional differences.

TC intensity is expressed in terms of the relative intensity I, defined as the ratio of the actual central pressure drop to the maximum possible central pressure drop (Darling, 1991). This dimensionless variable prevents the simulated central pressure from reaching non-physical values. The changes of I in each grid is given

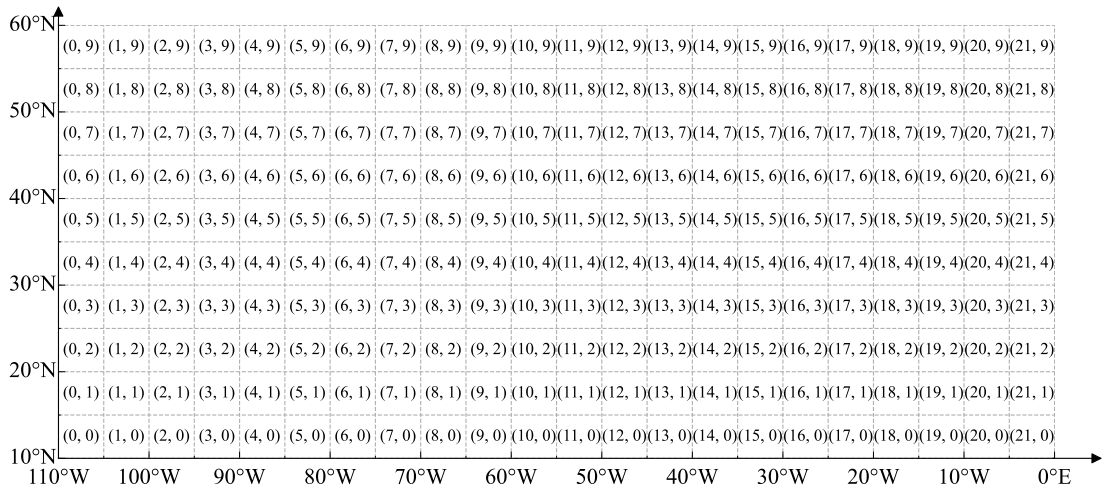


Figure 5.1: The index for 5° latitude by 5° longitude grids in the North Atlantic.

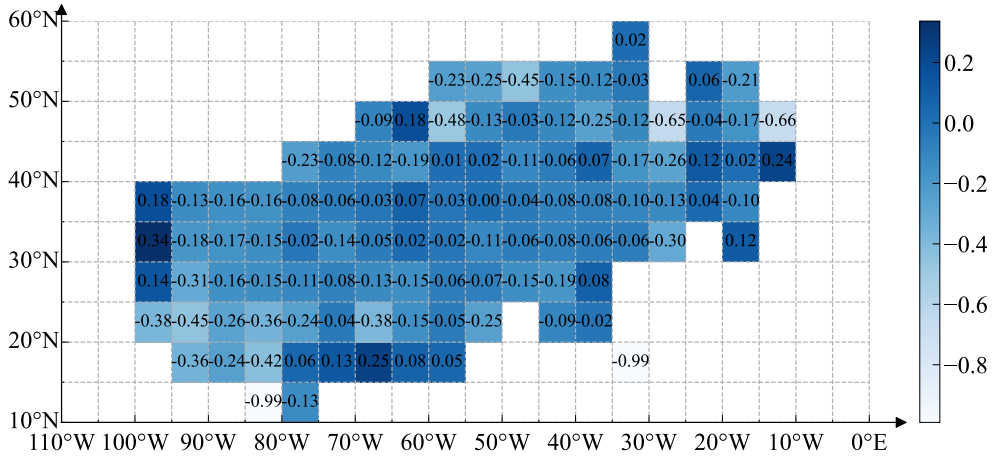
by:

$$\begin{aligned} \ln\left(I_{i+1}\right) = d_1 + d_2 \ln\left(I_i\right) + d_3 \ln\left(I_{i-1}\right) + d_4 \ln\left(I_{i-2}\right) \\ + d_5 T_{s_i} + d_6 \left(T_{s_{i+1}} - T_{s_i}\right) + \varepsilon_3, \end{aligned} \tag{5.2}$$

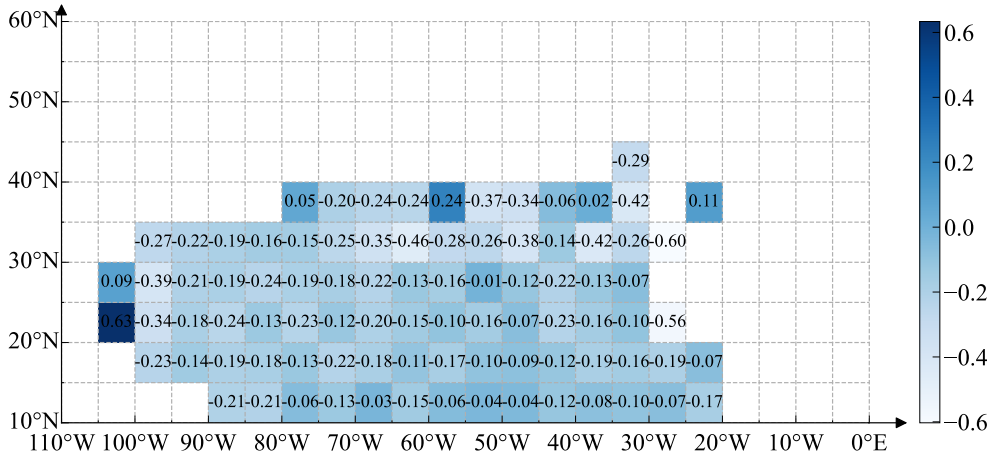
where T_{s_i} is the SST at the TC centre at step i. SST data are obtained from the HadISST database, which provides monthly mean values with a 1° × 1° spatial resolution (Rayner et al., 2003). ε_3 is a zero-mean Gaussian error with grid-specific variance. Coefficients d_1 to d_6 are also grid-specific and differ for eastward- and westward-moving storms.

Fig. 5.2–5.4 display the fitted regression coefficients for $\ln V_{ti}$, θ_i and $\ln I_i$ in Eqs. (5.1a)–(5.2). Because observational records are limited, some grids lack sufficient samples for stable estimation. Shaded cells indicate grids where parameter fitting is feasible; unshaded cells denote data-sparse grids. During simulation, if a TC enters a grid without fitted coefficients, the nearest available grid's parameters are adopted to preserve continuity of the track simulation.

Figs 5.5 and 5.6 illustrate, for grids (6,3) and (8,4), the regression performance for Eqs. (5.1) and (5.2). In each figure, the left panels show residuals between



(a)



(b)

Figure 5.2: The regression coefficient for $\ln V_{ti}$ in Eq. (5.1): (a) east-heading direction, (b) west-heading direction.

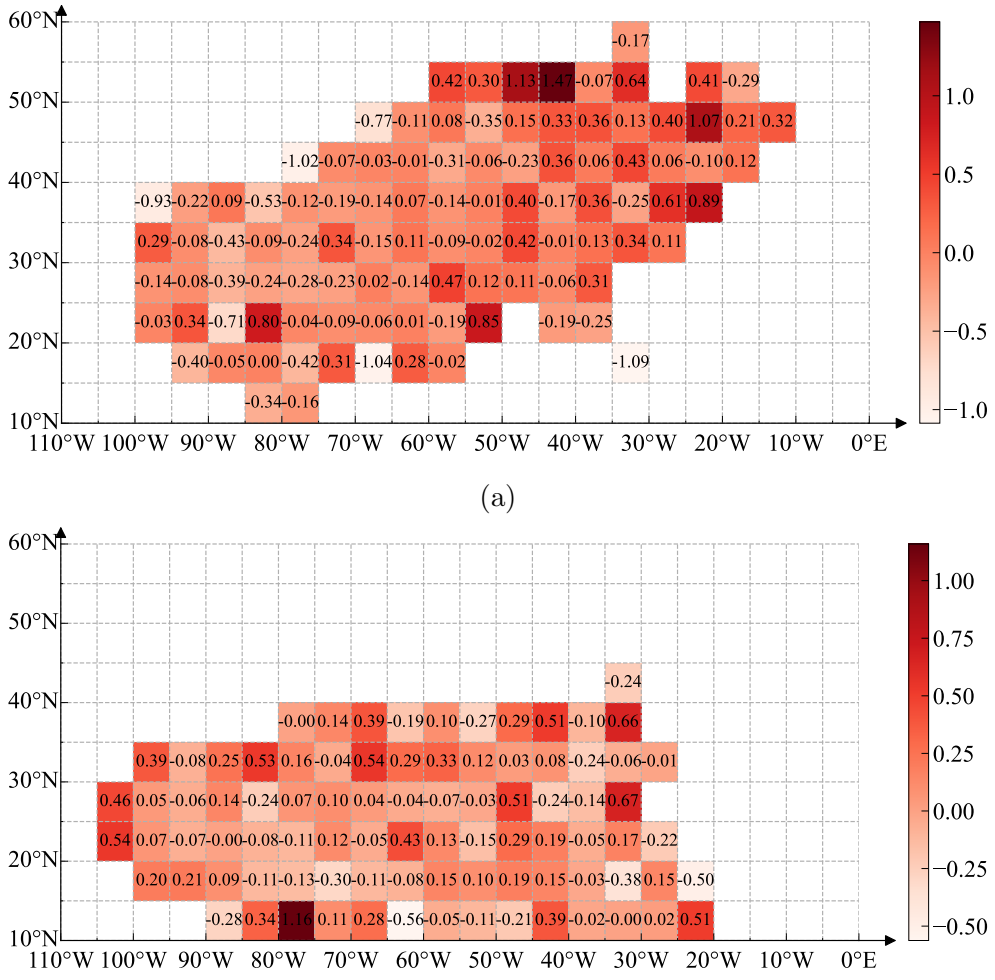
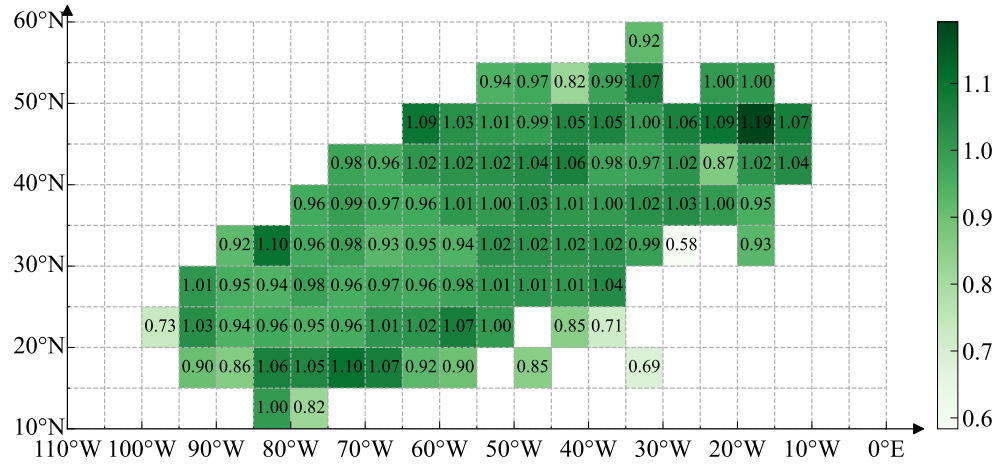
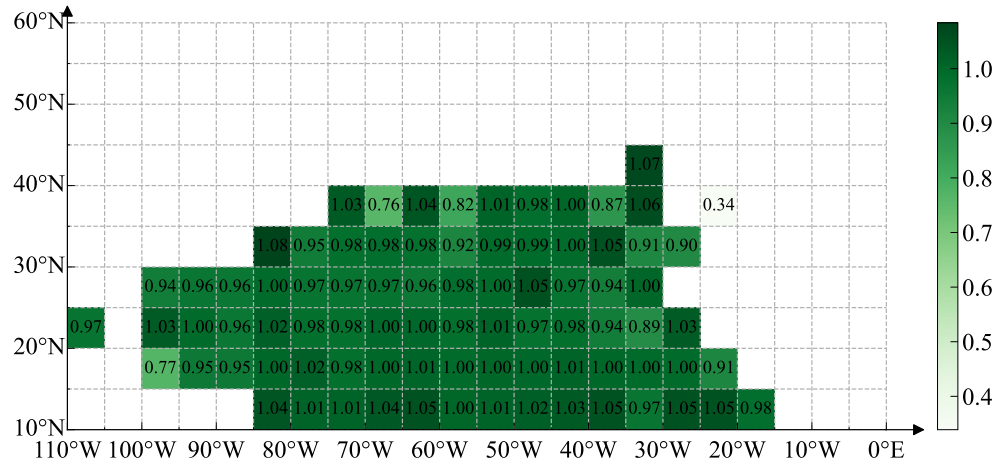


Figure 5.3: The regression coefficient for θ_i in Eq. (5.1): (a) east-heading direction, (b) west-heading direction.



(a)



(b)

Figure 5.4: The regression coefficient for $\ln I_i$ in Eq. (5.2): (a) east-heading direction, (b) west-heading direction.

computed and observed values; circular and triangular markers correspond to westward- and eastward-moving cases, respectively. Residuals are centred around zero with no evident systematic bias. The right panels compare distributions of modelled and observed values. Red markers represent translation speed, heading and intensity derived from HURDAT2, whereas blue markers show the corresponding quantities from the regression equations. The close alignment of these distributions demonstrates satisfactory goodness-of-fit.

The second stage utilises the coefficients in Eq. (5.1a), Eq. (5.1b), and Eq. (5.2) to iteratively generate synthetic TC events. TC genesis points are first sampled from the starting locations in HURDAT2 (HRD/NOAA, 2023), including the initial date, position, heading, translation speed, and central pressure. The TC tracks are then advanced at 6-hour intervals using these three equations. Upon landfall, the central pressure deficit ΔP decays following the filling-rate model (Vickery et al., 2000):

$$\Delta P(t_h) = \Delta P_0 \cdot \exp(-a \cdot t_h), \tag{5.3}$$

where t_h is the time since landfall in hours, ΔP_0 is the pressure deficit at landfall, and a is the decay rate. The decay parameter a is region-specific and can be modelled as (Vickery, 2005):

$$a = a_0 + a_1 \Delta P_0 + \varepsilon_a, \tag{5.4}$$

where a_0 and a_1 regression coefficients and ε_a a zero-mean normal error with standard deviation σ_{ε_a}; regional values for these parameters are listed in Table 5.1. If the storm re-enters the ocean, the intensity simulation reverts to Eq. (5.2). The simulation proceeds until the intensity drops below a threshold or the track exits the basin.

Fig. 5.7 illustrates the coastal mileposts along the U.S. and Mexican coastlines. The model's performance is evaluated by comparing the statistics of translation

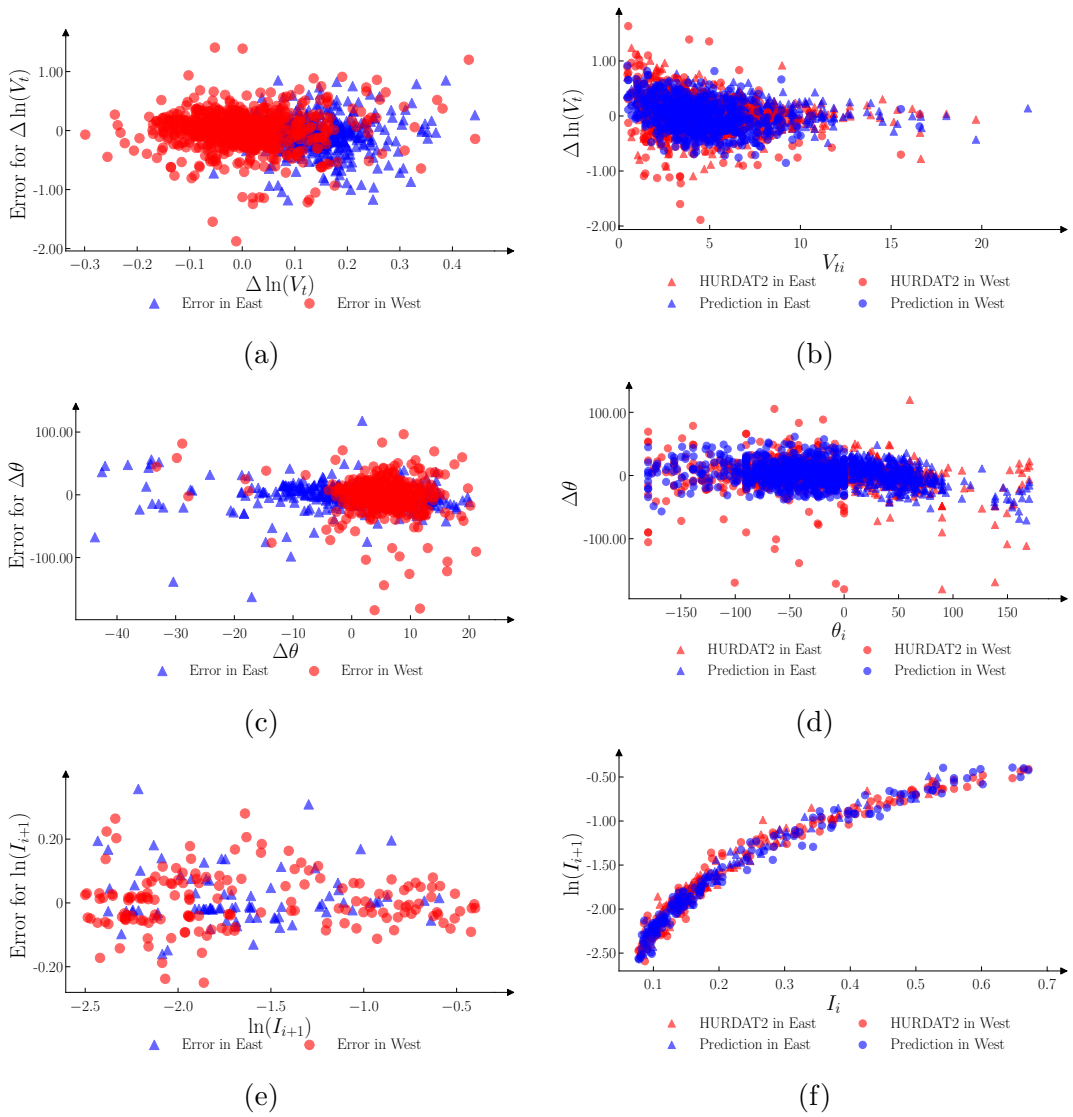


Figure 5.5: The regression results for the grid (6,3): (a) Prediction errors for $\Delta \ln V_t$, (b) Predicted values for $\Delta \ln V_t$, (c) Prediction errors for $\Delta\theta$, (d) Predicted values for $\Delta\theta$, (e) Prediction errors for $\ln I_{i+1}$, (f) Predicted values for $\ln I_{i+1}$.

Table 5.1: Regression coefficients for TC decay models across regions.

Region	a_0	a_1	σ_{ε_a}
Florida Peninsula	0.006	0.00046	0.0025
Gulf-of-Mexico	0.035	0.00050	0.0355
Atlantic Coast	0.038	0.00029	0.0093

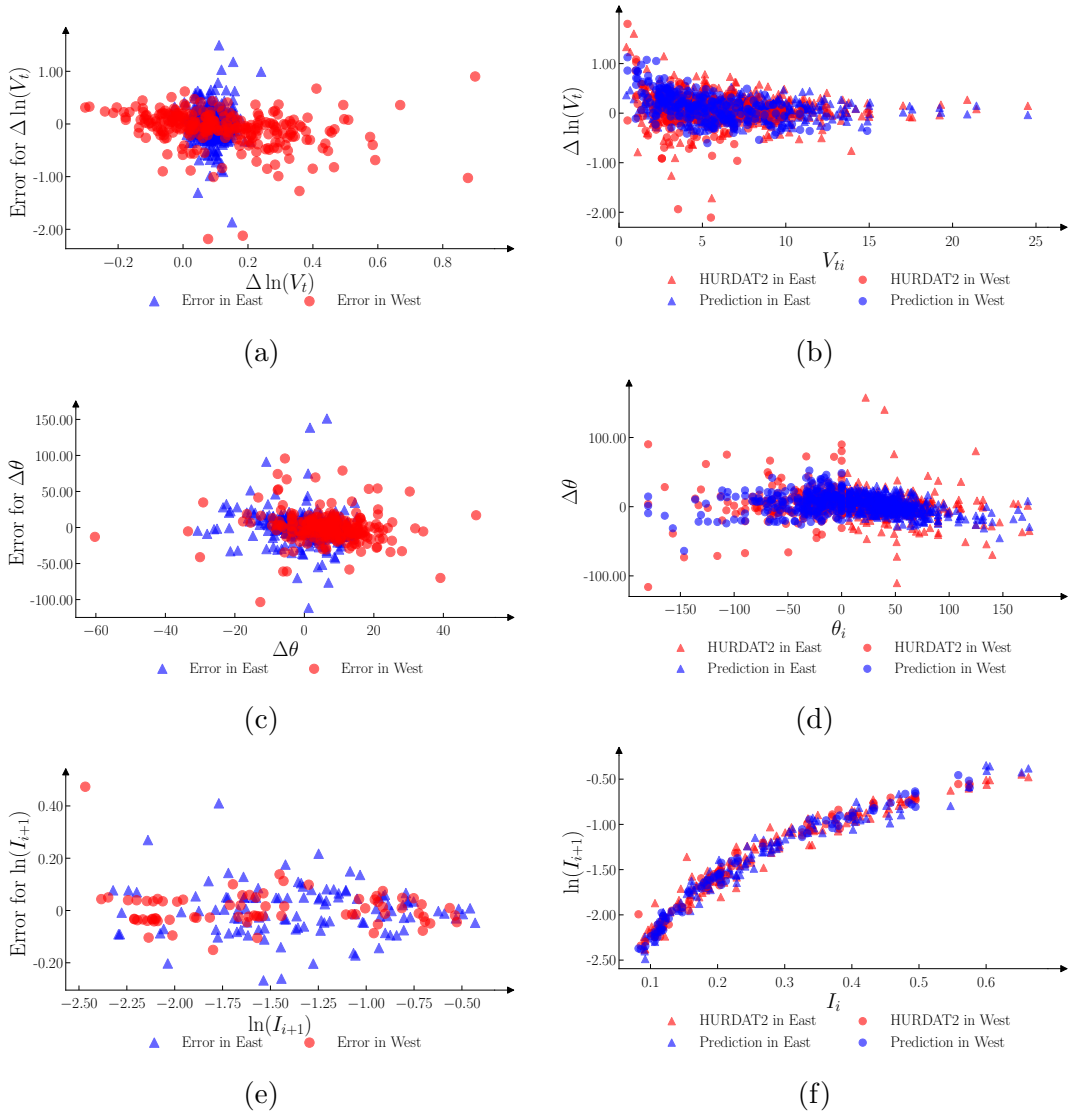


Figure 5.6: The regression results for the grid (8,4): (a) Prediction errors for $\Delta \ln V_t$, (b) Predicted values for $\Delta \ln V_t$, (c) Prediction errors for $\Delta\theta$, (d) Predicted values for $\Delta\theta$, (e) Prediction errors for $\ln I_{i+1}$, (f) Predicted values for $\ln I_{i+1}$.

speed, heading direction, annual occurrence rate, and central pressure deficit between historical records and simulated TCs. Fig. 5.8 presents the mean and standard deviation of these four variables across all mileposts, computed using the HURDAT2 database and 30,000 years of simulated TCs. The 30,000 simulated years are grouped into multiple 175-year subsets, approximately matching the duration of the observational record, to illustrate the stochastic variability of these synthetic TCs. For most mileposts, the observational curves fall within the envelope of simulated results, indicating that the synthetic TC model effectively reproduces the statistical characteristics of TC behaviour.

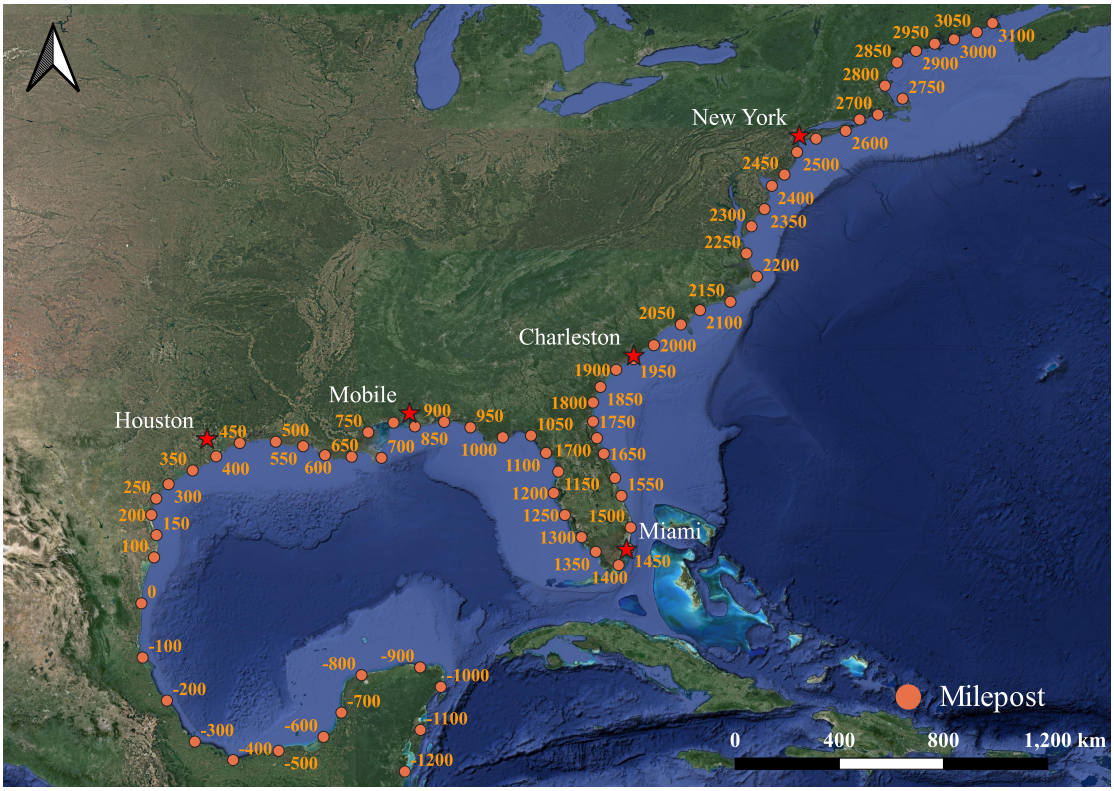


Figure 5.7: Milepost locations along U.S. and Mexican Coastline.

Finally, Fig. 5.9 shows a comparison of spatial densities between simulated and HURDAT2 storm tracks over the North Atlantic, with the close correspondence further demonstrating the effectiveness of the framework.

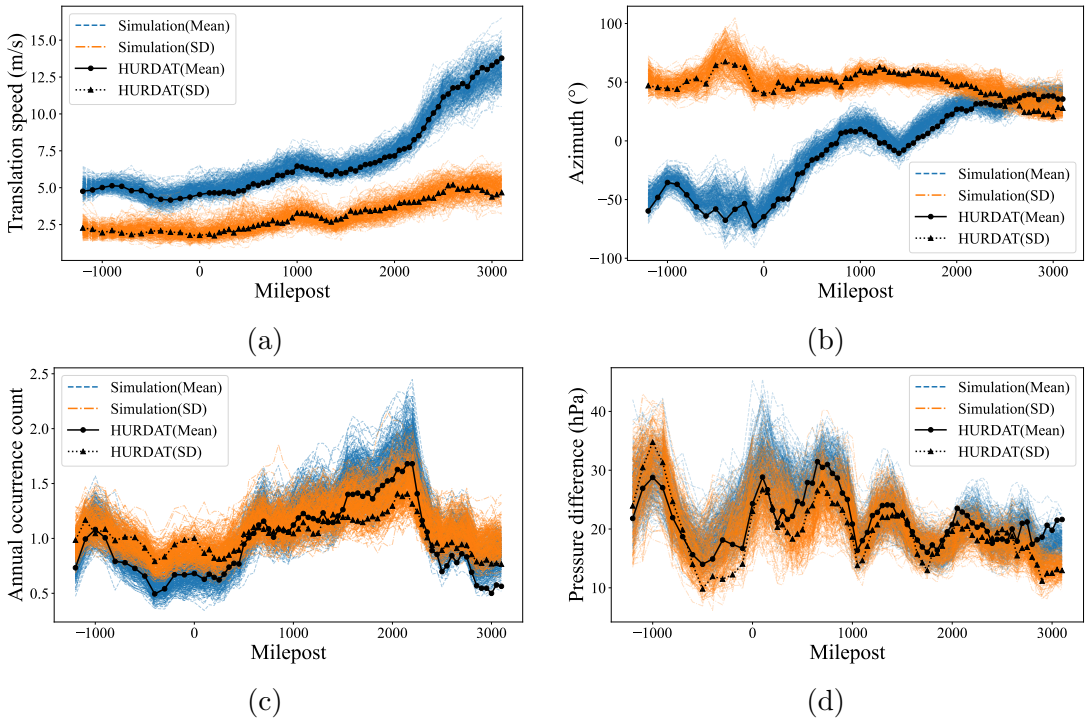
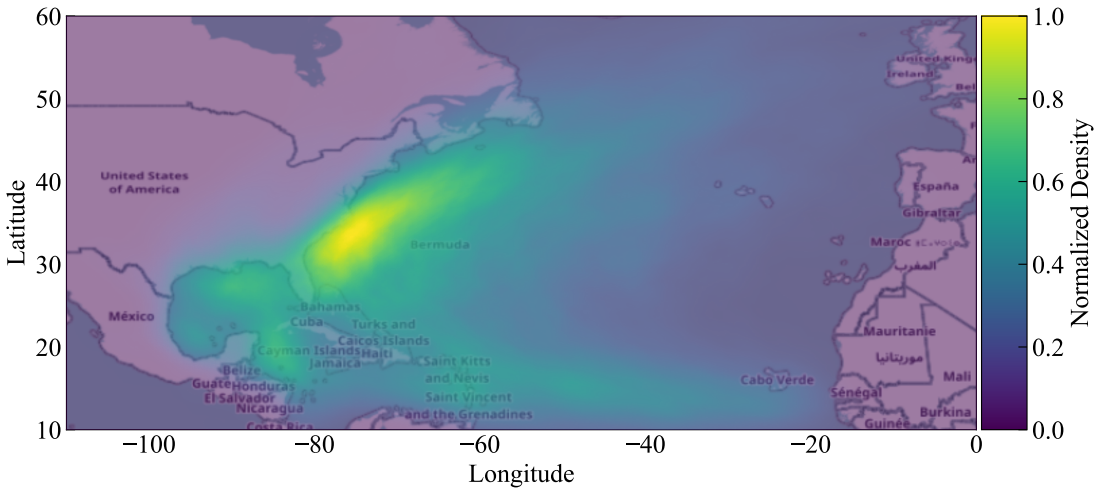
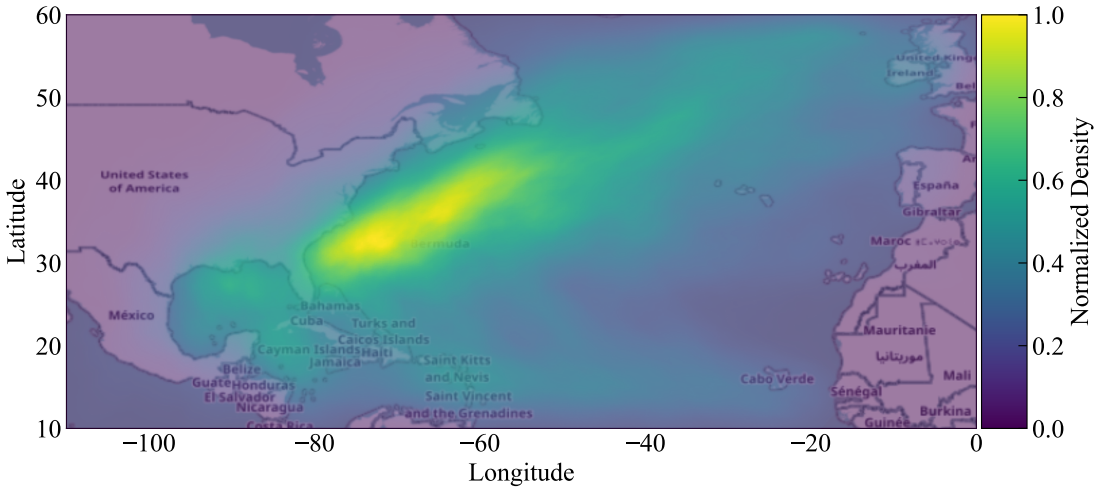


Figure 5.8: Comparison of TC parameters between simulated TCs and HURDAT2 database: (a) translation speed, (b) heading direction, (c) annual occurrence count and (d) central pressure deficit.



(a)



(b)

Figure 5.9: The spatial density distribution of TCs in the North Atlantic based on (a) HURDAT2, (b) Simulated.

5.1.2 Estimation of return period wind speeds

In this study, five coastal cities, Charleston, Houston, Miami, Mobile, and New York, are selected to illustrate the spatial distribution of TC risk along the U.S. coastline, as shown in Fig. 5.7. Synthetic TCs generated in Section 5.1.1 are used to compute the return period wind speeds for these cities.

Following the approach proposed in previous studies (Vickery et al., 2000; Bloemendaal et al., 2022), the wind speeds of 30,000 synthetic TCs are first evaluated using the wind field model described in Section 3.1.1. For each city, the maximum 3-second gust wind speed at 10 m height over open terrain produced by each TC is recorded. Wind speeds exceeding 9 m/s are retained to estimate the return period wind speed based on

$$P_{\text{exc}}(v) = \frac{i}{n+1} \cdot \frac{n}{m}, \tag{5.5}$$

$$T(v) = \frac{1}{P_{\text{exc}}(v)}, \tag{5.6}$$

where $P_{\text{exc}}(v)$ denotes the exceedance probability of wind speed v, i is the rank of v in descending order, n is the number of TC events considered, m is the simulation period in years, and $T(v)$ is the return period corresponding to wind speed v.

Fig. 5.10 compares the return period wind speeds calculated in this study (solid lines) with those specified in ASCE 7-16 (ASCE, 2017) (markers). Miami exhibits the highest return period wind speeds, followed by Mobile, Charleston, and Houston, while New York shows the lowest values. Table 5.2 summarises the RMSE for the five MRIs considered (50, 100, 300, 700 and 1,700 years) at each city. The simulated return period wind speeds align well with the ASCE provisions, indicating that the TC simulation method effectively captures regional variations in TC risk. These return period wind speeds form the basis for assessing future climate change impacts on coastal cities in subsequent sections.

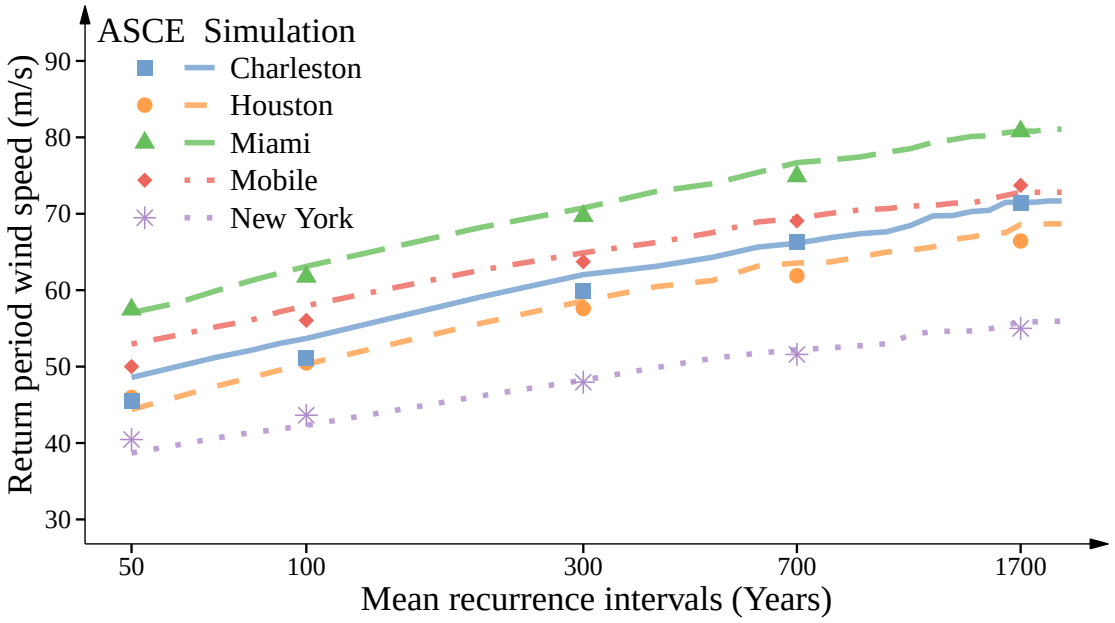


Figure 5.10: The return period wind speed for the five cities in the historical climate.

Table 5.2: RMSE of estimated return period wind speeds for five cities.

City	RMSE (m/s)
Charleston	2.02
Houston	1.48
Miami	1.12
Mobile	1.72
New York	1.10

5.2 Projected changes in TC characteristics under future climate scenarios

5.2.1 Overview of the framework

This section presents a framework for estimating the impacts of climate change on three key TC characteristics, including intensity, frequency, and genesis location, using high-resolution GCMs.

Historical TCs are obtained from established databases such as HURDAT2 (HRD/NOAA, 2023) and the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010), which serve as reference datasets for bias correction. Two additional TC datasets are then generated from GCM simulations corresponding to the present and future climates, respectively. The GCM-based TCs are identified using a TC tracking algorithm that applies multiple selection criteria to distinguish TCs from other vortical systems. These three datasets are separately analysed to evaluate TC characteristics in terms of intensity, frequency, and genesis location. Comparison between historical observations and GCM-simulated historical-climate TCs allows for the estimation of systematic model biases, which are subsequently applied to the future-climate simulations to produce bias-adjusted projections of future TC characteristics.

This study adopts the SSP585 emission scenario, representing a high-end radiative forcing pathway of approximately 8.5 W/m^2 by 2100. This scenario provides an upper bound for impact and risk assessments and is widely used in climate change studies (Kropf et al., 2025; Balaguru et al., 2023; Meiler et al., 2023). The CNRM-CM6-1-HR model is employed due to its relatively high spatial resolution (approximately 50 km in the atmosphere and 25 km in the ocean). As a member of CMIP6/HighResMIP, this model is designed to evaluate how enhanced resolution

influences climate simulations. Compared with its predecessor, CNRM-CM6-1-HR exhibits notable improvements in climate reproducibility (Saint-Martin et al., 2021; Voldoire et al., 2019) and has been widely applied in climate impact assessments (Kropf et al., 2025; Meiler et al., 2023). This study employs a single GCM to demonstrate the framework's applicability rather than to conduct ensemble-based uncertainty analysis. Nevertheless, the proposed workflow is readily extendable to multiple GCMs in future research. The periods 1979–2014 and 2015–2050 are defined as the historical and future climates, respectively. The historical period corresponds to the satellite era, during which reliable global TC observations became available, while 2015 marks the beginning of the CMIP6 projection period and the implementation phase of the Paris Agreement. The year 2050 represents a mid-century projection horizon that is widely adopted in climate impact assessments (Bloemendaal et al., 2022).

5.2.2 GCM-based TC detection and tracking method

TC trackers are automated algorithms that identify cyclonic vorticity maxima in the lower and mid-troposphere and track their temporal evolution. These algorithms apply a series of filtering criteria (e.g. a warm-core structure, sustained cyclonic rotation) to distinguish TCs from other vortical systems, such as midlatitude disturbances or transient eddies.

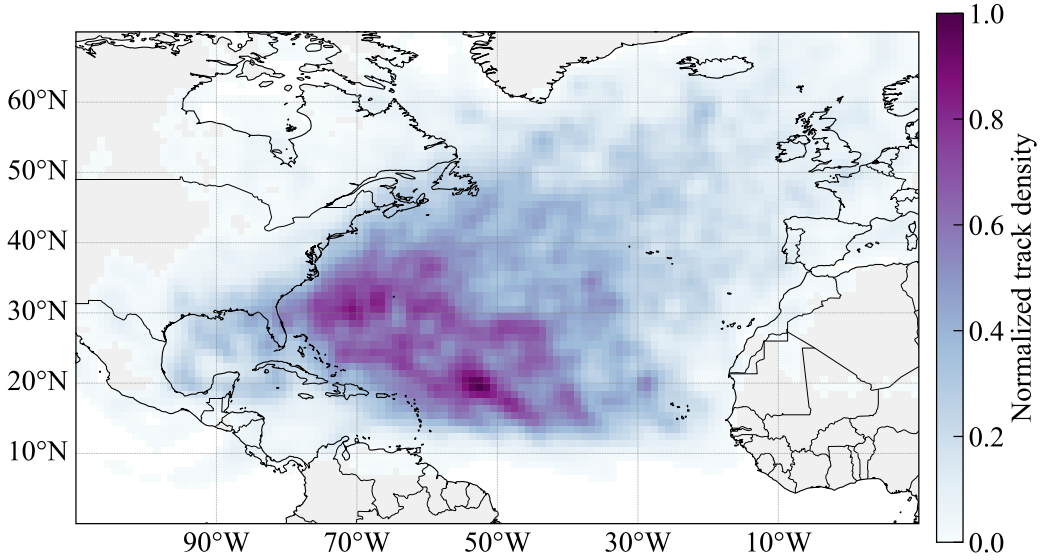
In this study, the TRACK algorithm (Hodges et al., 2017) is adopted for TC detection. The algorithm begins with the isolation of regions exhibiting enhanced cyclonic relative vorticity, generally focused at the 850 hPa level, and incorporates spectral filtering to reduce background noise and highlight coherent storm structures. Unlike some identification schemes that apply thresholds during the tracking stage, TRACK initially captures a broader range of disturbances, including pre-TC and post-TC stages, by using vertically averaged vorticity across

multiple levels (850, 700, and 600 hPa). This enhances temporal coherence and ensures a more continuous representation of the storm lifecycle, particularly when the vorticity maximum shifts between levels. The candidate vortices are initially linked using a nearest-neighbour approach and subsequently optimised by minimizing a cost function to produce smooth and dynamically consistent trajectories. Tracks must persist for a minimum of two days to ensure robust detection while reducing the likelihood of spurious identifications. Once the tracking stage is completed, additional filtering is applied to identify TCs using objective criteria based on storm structure and intensity. These criteria include a specified threshold for low-level vorticity, a demonstrable warm-core vertical profile (inferred from vorticity differences between the lower and upper troposphere), a coherent vertical structure between 850 and 200 hPa, and the requirement that these conditions persist for at least four consecutive time steps. To focus on tropical systems, only tracks originating between 30°S and 30°N and located over oceanic regions are considered. Previous studies have demonstrated that the algorithm successfully identifies nearly all observed TCs; for further details, see Hodges et al. (2017).

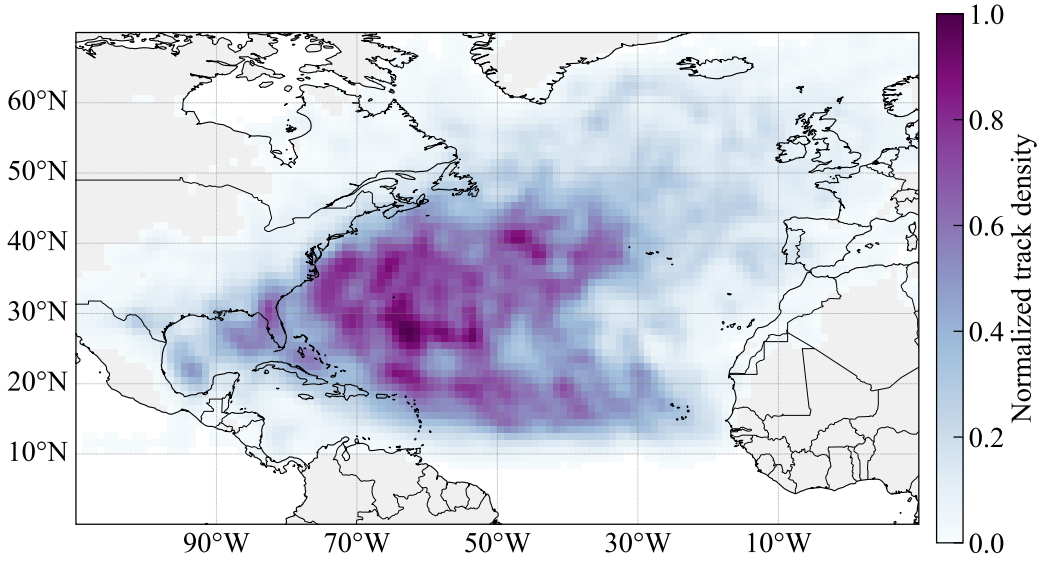
The TRACK algorithm is applied to identify TCs using atmospheric and oceanic outputs from the CNRM-CM6-1-HR model. The resulting datasets include both historical (1979–2014) and future (2015–2050) climates. The spatial distribution of these TCs is presented in Fig. 5.11. It should be noted that these GCM-derived TCs are affected by biases associated with both the underlying climate model and the tracking algorithm. Therefore, bias correction is a necessary step before using these identified tracks in risk analyses.

5.2.3 SST adjustment for TC intensity estimation

SST directly affects TC intensity through its control on ocean heat content and energy fluxes that fuel TC development. The relationship between SST and TC



(a)



(b)

Figure 5.11: Spatial distribution of identified TCs from the CNRM-CM6-1-HR model: (a) historical period (1979–2014) and (b) future period (2015–2050).

intensity is represented in Eq. (5.2). This section evaluates the influence of climate change on SST as a proxy for assessing its impact on TC intensity.

The systematic biases in SSTs simulated by the GCMs are first evaluated. Observed SSTs for 1979–2014 are obtained from the HadISST dataset (Rayner et al., 2003), which provides monthly mean values on a $1^\circ \times 1^\circ$ grid. Model-derived SSTs for both the historical and future periods are obtained from the CNRM-CM6-1-HR model (Voldoire, 2019). To ensure consistency with the observations, the native $0.25^\circ \times 0.25^\circ$ outputs are resampled to a $1^\circ \times 1^\circ$ grid, consistent with HadISST. A bias correction is applied to mitigate the systematic discrepancies in modelled SSTs. Common bias-correction techniques include quantile mapping, empirical distribution fitting, and mean adjustment (Mehrotra and Sharma, 2015), which generally aim to align the statistical properties of model outputs with observed climatology. Among these, the mean bias correction, one of the simplest yet most widely adopted approaches, is used in this study. The mean SST bias ΔT for each grid cell is computed as the average difference between observed and simulated SSTs during the historical period (1979–2014):

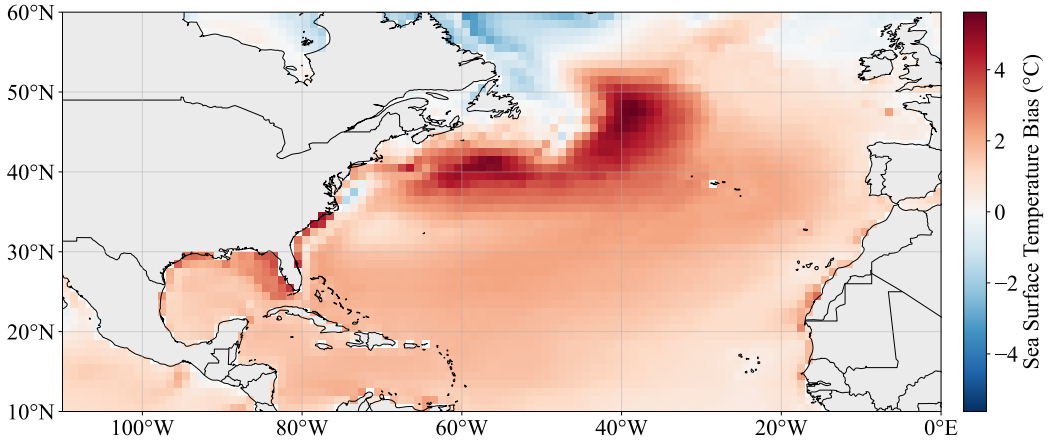
$$\Delta T(x, y) = \frac{1}{N} \sum_{t=1}^{N} \Big(T_{\text{obs}}(x, y, t) - T_{\text{GCM}}(x, y, t) \Big), \tag{5.7}$$

where $T_{\text{obs}}(x, y, t)$ and $T_{\text{GCM}}(x, y, t)$ denote the observed and simulated SSTs at longitude (x), latitude (y), and month (t), respectively. A total of $N = 432$ monthly values (36 years $\times$ 12 months) are averaged per grid cell to obtain the mean bias field. The bias-adjusted SST projections for 2015–2050 are then obtained by adding this bias field to the raw SST:

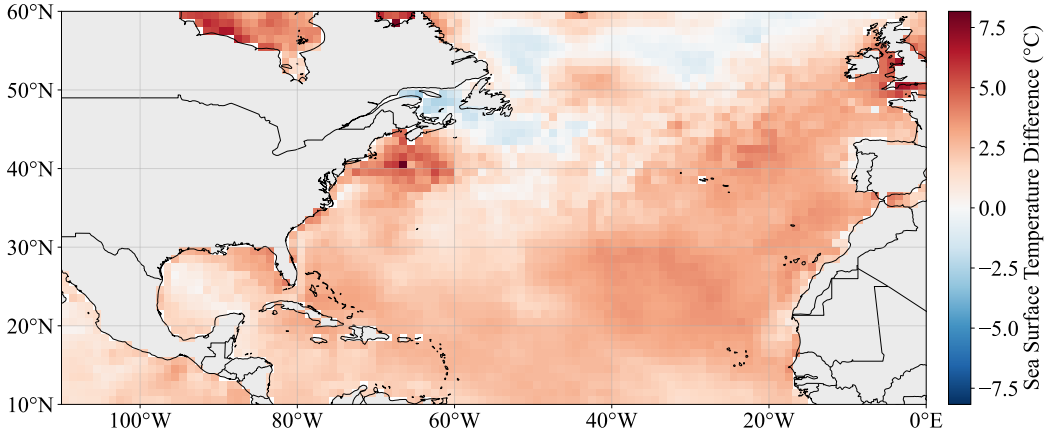
$$T(x, y, t') = T_{\text{GCM}}(x, y, t') + \Delta T(x, y), \quad t' \in [2015, 2050]. \tag{5.8}$$

The spatial distribution of the mean SST bias is shown in Fig. 5.12a. The CNRM-CM6-1-HR model systematically underestimates SSTs across much of the

North Atlantic, with the largest cold biases (up to 4 °C) observed south of Greenland, around the Labrador Sea, and across the western subtropical Atlantic. These anomalies are likely associated with limitations in representing the Gulf Stream and related mesoscale circulation. If uncorrected, such biases could lead to underestimated TC intensities and, consequently, to an underestimation of climate-induced risks. The bias-corrected SST changes between August 2014 and August 2050 are shown in Fig. 5.12b for illustration, revealing widespread warming across most of the North Atlantic, with localised cooling in subtropical regions.



(a)



(b)

Figure 5.12: SST correction for the CNRM-CM6-1-HR model: (a) mean SST bias, (b) projected SST difference between August 2014 and August 2050 under SSP585.

5.2.4 Assessment of future TC frequency

In probabilistic TC risk assessment, the annual TC frequency is commonly modelled using count distributions such as the Poisson distribution (Jing and Lin, 2020; Zeng et al., 2024) or the negative binomial (NB) distribution (Huang et al., 2022; Vickery et al., 2000). To determine the appropriate distribution for the North Atlantic basin, this study calculates the mean and standard deviation of annual TC frequency from 1979 to 2014, which are 12.38 and 4.95, respectively. Because the variance ($4.95^2 = 24.50$) is substantially greater than the mean, the Poisson distribution, which assumes equality between the mean and variance, is unsuitable in this situation. Therefore, the NB distribution is adopted in this study. The empirical cumulative distribution function (CDF) and the fitted NB distribution are compared in Fig. 5.13, showing good agreement.

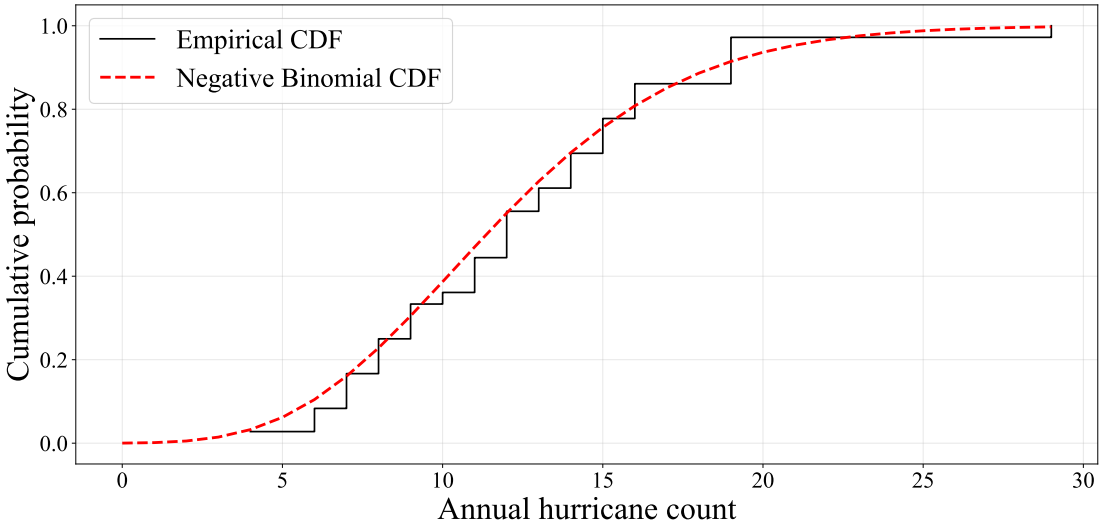


Figure 5.13: Empirical and fitted negative binomial CDFs for annual TC frequency in the North Atlantic.

For estimating the two parameters of the NB distribution under future climate conditions, the same bias-correction principle described in Section 5.2.3 is applied, that is, using observed values to correct model biases before assessing future TC

characteristics. The mean and standard deviation of annual TC frequency under the future climate are evaluated as:

$$\mu = \mu_{\text{obs}} \left(1 + \frac{\mu'_{\text{GCM}} - \mu_{\text{GCM}}}{\mu_{\text{GCM}}}\right), \tag{5.9a}$$

$$\sigma = \sigma_{\text{obs}} \left(1 + \frac{\sigma'_{\text{GCM}} - \sigma_{\text{GCM}}}{\sigma_{\text{GCM}}}\right), \tag{5.9b}$$

where μ_{GCM} and σ_{GCM} denote the mean and standard deviation of modelled TC frequencies for the historical period, and μ'_{GCM} and σ'_{GCM} represent the corresponding statistics for the future period. μ_{obs} and σ_{obs} are the observed mean and standard deviation derived from HURDAT2. Substituting these values yields mean and standard deviation estimates of 11.43 and 4.73 for the future climate, respectively, indicating a reduction compared with the historical climate. This decreasing trend in frequency is consistent with recent studies (Fisher et al., 2020; Knutson et al., 2020; Bloemendaal et al., 2022).

5.2.5 Projected shifts in genesis locations under climate change

Unlike frequency, which is expressed as a scalar value, TC genesis locations are represented as a set of discrete points within a two-dimensional domain. Fig. 5.14 shows the spatial distribution of TC genesis locations in the North Atlantic during 1979–2014. To evaluate projected spatial shifts in genesis distribution, these discrete points are first converted into a continuous spatial field.

Kernel density estimation (KDE) is applied to perform this transformation. KDE is a non-parametric statistical technique that converts discrete events into a continuous probability surface by applying a kernel function to each observation. In this study, a Gaussian kernel is used to estimate genesis density across the North Atlantic, and the Spatial Analyst tool in ArcGIS is employed to compute the corresponding density values. It should be noted that the original genesis data contains both spatial distribution information and event counts. The impact

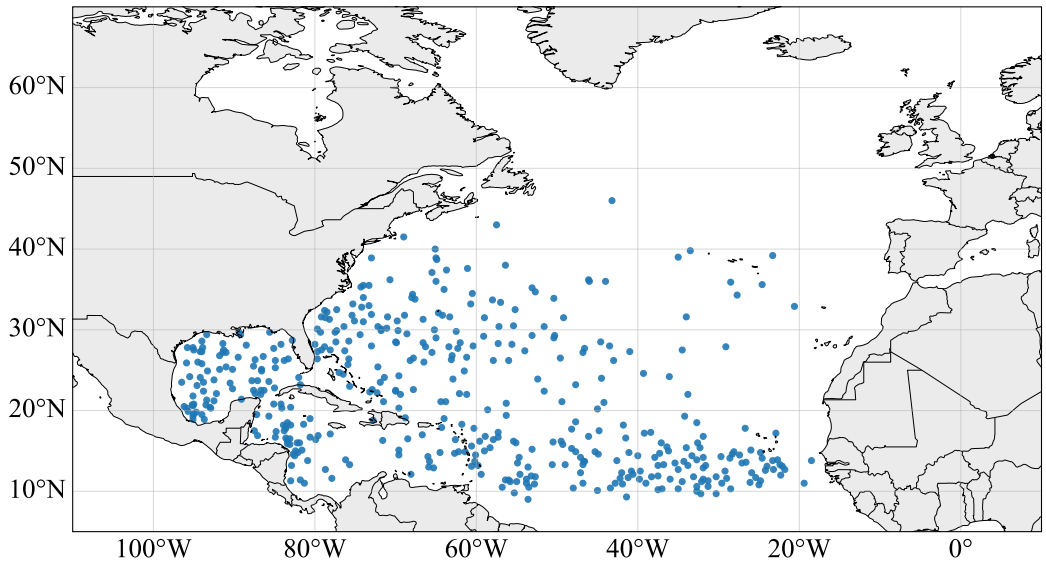


Figure 5.14: The distribution of TC genesis locations in the North Atlantic from 1979 to 2014.

of climate change on TC frequency has been analysed in Section 5.2.4, and this section focuses solely on spatial distribution changes. To eliminate the influence of event counts, the spatial density is normalised so that the resulting values represent the probability density function (PDF) of genesis location. Following the same calibration procedure used for storm frequency, the future genesis density is adjusted by comparing model-simulated and observationally derived density fields. The corrected density field is then obtained as:

$$D(x,y) = D_{\text{obs}}(x,y)\left(1 + \frac{D'_{\text{GCM}}(x,y) - D_{\text{GCM}}(x,y)}{D_{\text{GCM}}(x,y)}\right), \tag{5.10}$$

where $D_{\text{obs}}(x,y)$ denotes the normalised historical genesis density derived from HURDAT2, and $D_{\text{GCM}}(x,y)$ and $D'_{\text{GCM}}(x,y)$ represent the historical and future normalised genesis densities, respectively, obtained from the CNRM-CM6-1-HR model. The resulting $D(x,y)$ represents the calibrated future genesis density.

The spatial differences between historical and future genesis densities are shown

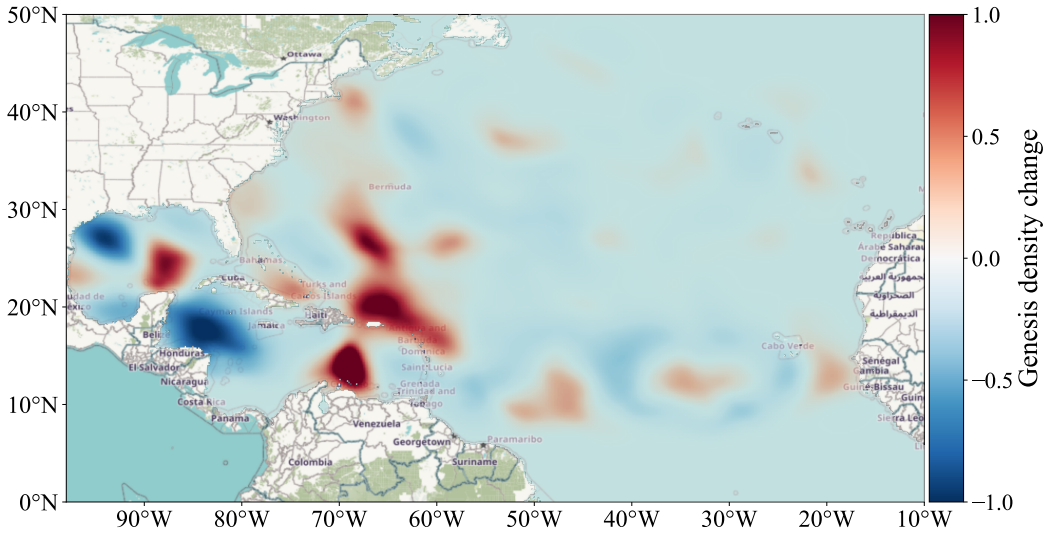
in Fig. 5.15a, with values normalised to the range of –1 to 1 for ease of interpretation. Blue regions indicate reduced genesis activity, whereas red regions suggest increases. A notable reduction in genesis activity is projected across the western Caribbean and adjacent portions of the western tropical Atlantic. Reduced activity is also observed in parts of the Gulf of Mexico and the subtropical western Atlantic. Conversely, increased genesis activity is projected over the central and eastern Caribbean, with a broad region of enhancement extending northeastward across the western–central North Atlantic. Moderate increases are also evident over the eastern tropical Atlantic near the main development region. Overall, the projected pattern indicates a basin-scale reorganisation of genesis conditions, characterised by a relative shift from the western Caribbean toward the eastern Caribbean and the western–central North Atlantic. Such spatial redistribution may influence the initiation regions of TCs, with implications for regional TC hazards along the coasts of North America. The PDF of the spatial distribution of genesis density is presented in Fig. 5.15b.

5.3 Climate change impacts on coastal TC risks

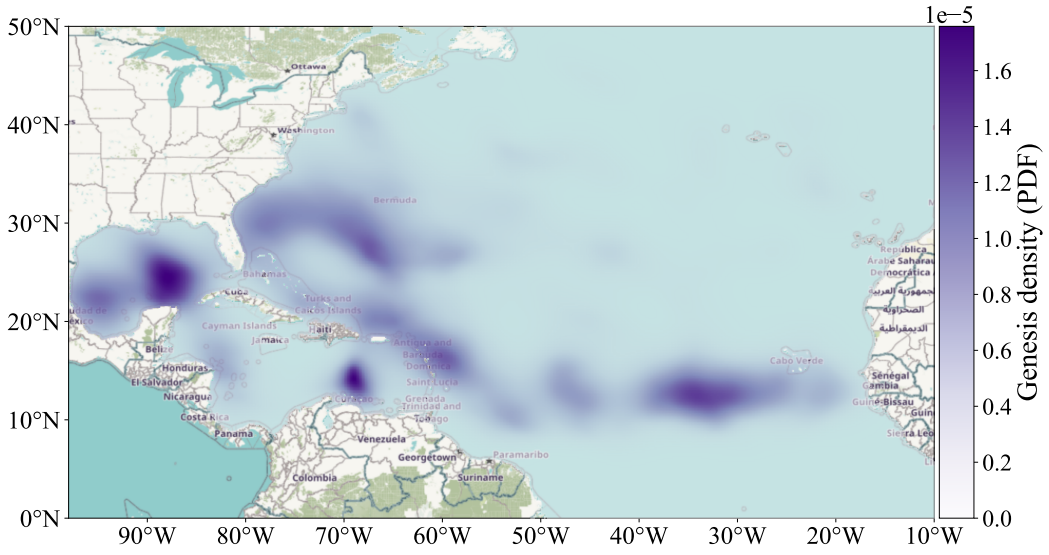
5.3.1 Comparison of return period wind speeds between historical and future climates

Section 5.2 presents the framework for evaluating the impacts of climate change on TC intensity, genesis location, and frequency. Building on this foundation, the present section quantifies the individual and combined contributions of these factors to the return period wind speeds across five coastal cities.

Fig. 5.16 illustrates the variations in return period wind speeds under different climate scenarios. The impacts of changes in TC frequency and genesis location are relatively modest. A reduction in storm frequency consistently decreases return period wind speeds across all cities. The influence of genesis location varies



(a)



(b)

Figure 5.15: Projected changes in TC genesis: (a) spatial differences in TC genesis density between historical and future climates, (b) spatial distribution of future TC genesis density.

spatially: slight decreases are observed in Houston and Mobile, while small increases appear in Miami, Charleston, and New York. This pattern aligns with the projected redistribution of TC genesis, which shows reduced activity in the western Caribbean and Gulf of Mexico but enhanced activity over the eastern Caribbean and the western–central North Atlantic. Overall, both the frequency- and genesis-driven effects remain relatively small.

In contrast, TC intensification produces the most significant increases in return period wind speeds, dominating the climate-change signal across all cities. When all three factors are considered together, Houston exhibits an apparent reduction in return period wind speeds relative to the intensity-only case, followed by Mobile. This reduction arises from both decreased storm frequency and a westward reduction in genesis activity. Conversely, the combined-scenario return-period curves for Miami, Charleston, and New York remain close to the intensity-only projections, as the decrease in frequency is partially compensated by redistributed genesis activity, maintaining similar overall levels of wind hazard.

The return period wind speeds under the combined impacts of changes in TC frequency, genesis location, and intensity are shown in Fig. 5.17. The relative ordering of wind hazard among the five cities remains consistent with the historical climate, with Miami exhibiting the highest wind speeds, followed by Mobile and Charleston, then Houston, and New York showing the lowest levels. However, the magnitude of projected changes varies across cities. Table 5.3 summarises the projected increases in the 100-, 500-, 1000-, and 2000-year return period wind speeds under the combined-impact scenario. New York, Miami, and Houston exhibit the largest increases, with projected changes exceeding 10 m/s at the 2000-year return period. Miami lies within a high TC-risk region, and the projected increase suggests a further amplification of existing hazards. In contrast, New York

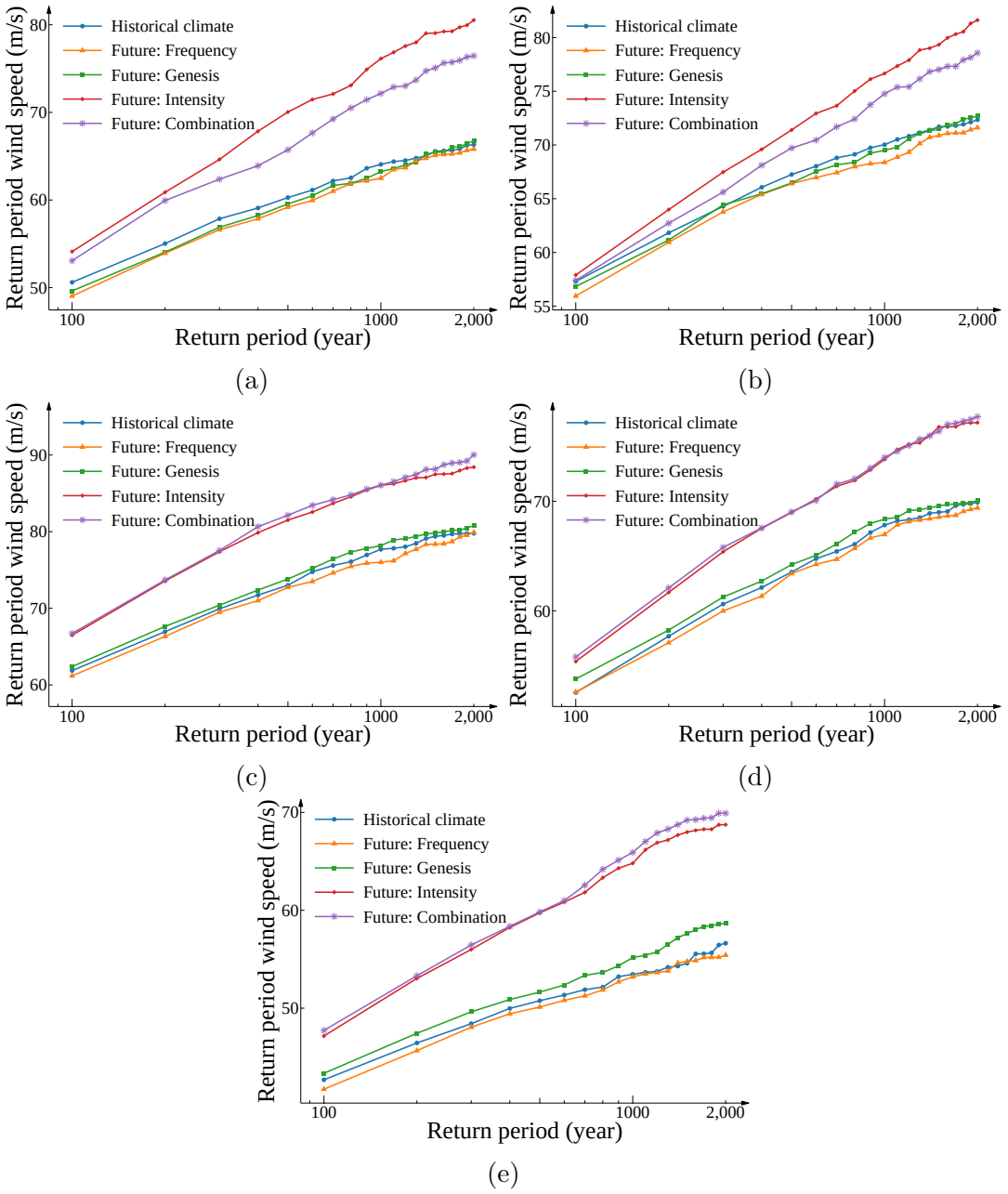


Figure 5.16: Return period wind speed comparison under historical and future climates for five U.S. coastal cities: (a) Houston, (b) Mobile, (c) Miami, (d) Charleston, and (e) New York.

and Houston, historically among the relatively lower-risk cities, show comparatively larger relative increases, implying a notable escalation in hazard exposure under future climate conditions. These results highlight the regional variability of climate-induced TC risks along the U.S. coastline. Furthermore, the increases become more pronounced at longer return periods, indicating that intensifying future TCs may heighten the likelihood of rare, extreme wind events.

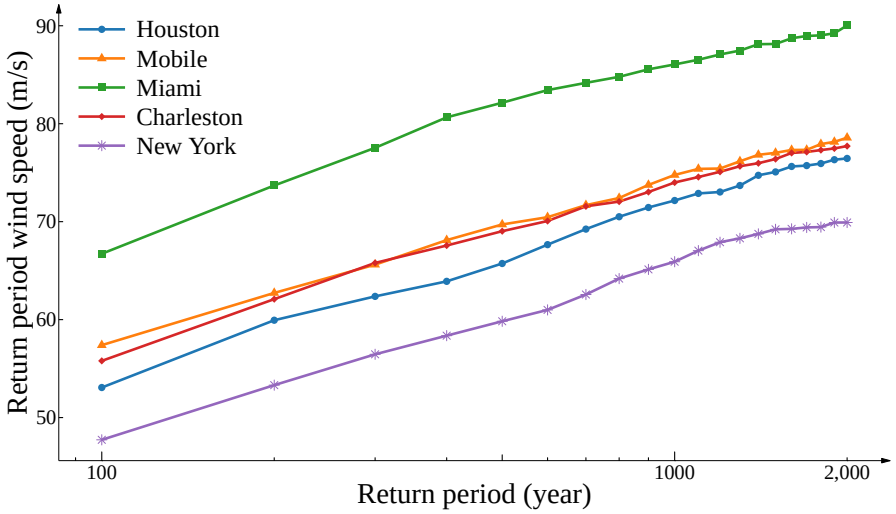


Figure 5.17: Comparison of return period wind speeds across five cities under climate-induced changes in TC frequency, genesis location, and intensity.

Table 5.3: Increase in return period wind speed (m/s) across five cities under climate-induced changes in TC frequency, genesis location, and intensity.

City	ΔV_{100}*	ΔV_{500}	ΔV_{1000}	ΔV_{2000}
Houston	2.46	5.45	8.09	10.14
Mobile	0.10	2.46	4.74	6.22
Miami	4.85	9.13	9.66	10.25
Charleston	3.25	5.49	6.18	7.82
New York	5.05	9.07	12.45	13.28

* V_T denotes the wind speed corresponding to a return period of T years.

5.3.2 Assessment of residential building losses in Miami-Dade County under climate change

This section uses Miami-Dade County as an example to evaluate the impacts of climate change on residential building losses. The loss estimation is conducted using the TC loss model developed in Section 3.4. For each of the 30,000 simulated years, all TCs are processed sequentially. The maximum wind speed at the centroid of each of the 13 sub-regions (Table 3.3) is used as the input to the loss model. Based on the model coefficients for each sub-region (Table 3.6), the corresponding loss ratio is estimated. Finally, the county-level loss ratio is obtained using Eq. (3.54). The annual loss ratio represents the total damage caused by all TCs occurring in a given year. If multiple storms occur in the same year, complete recovery between events is assumed; thus, the annual loss ratio may theoretically exceed one.

Because the number of TCs affecting the region can vary substantially from year to year, the stability of the estimated annual loss distribution depends on the number of simulation years. Fig. 5.18 illustrates the convergence of the estimated average annual loss under the historical climate as the number of simulated years increases. The results show that the average annual loss stabilises once the simulation length reaches approximately 18,000 years. The total simulation length of 30,000 years used in this study is therefore sufficient to obtain stable estimates for Miami-Dade County.

Fig. 5.19 compares the projected distributions of annual loss ratios under historical and future climates. The intensification of TCs produces a clear upward shift across the entire distribution, whereas the effects of changes in TC frequency and genesis location are comparatively modest. The impacts of TC frequency

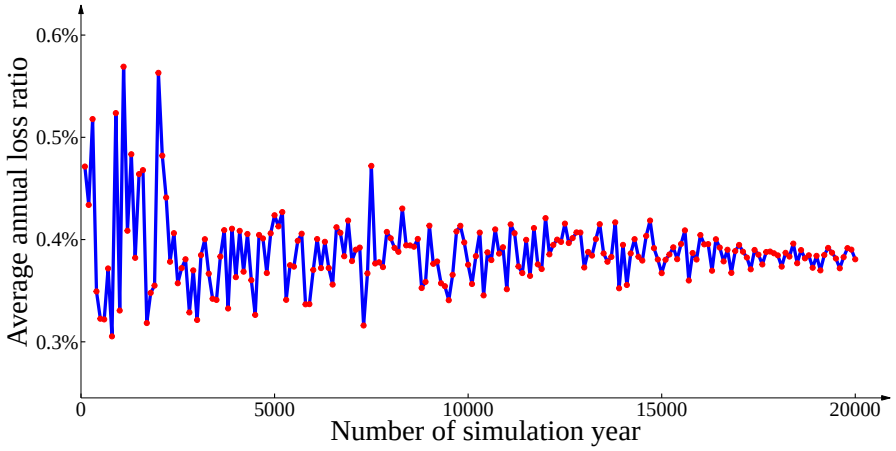


Figure 5.18: Convergence of average annual loss with increasing simulation length for Miami-Dade County.

and genesis location primarily arise through changes in the number of storms affecting the region, while TC intensity influences the severity of individual storm impacts. Based on statistics of storms passing within 200 km of Miami-Dade County, the annual TC arrival rate is 1.25 in the historical climate and 1.16 and 1.28 under the frequency-change and genesis-change scenarios, respectively, indicating only minor differences in the occurrence rate. In contrast, the number of strong storms (Category 4 and above) increases by approximately 28% under the intensity-change scenario, leading to a substantial increase in annual loss. The combined-impact scenario yields an annual loss distribution that closely matches the intensity-only case, further indicating the dominant role of TC intensity in driving future loss increases. Pant and Cha (2019) conducted a similar assessment of climate change impacts on losses in Miami-Dade County. However, their analysis primarily focused on a specific building class or total regional economic loss, whereas the present study evaluates structural damage ratios for the residential building stock. Furthermore, Pant and Cha (2019) considered climate change effects mainly through changes in TC intensity, while the present work incorporates the combined influences of TC intensity, frequency, and genesis location. The two

studies also differ in their SST data sources: Pant and Cha (2019) used SST fields from the GFDL model, whereas the present study employs the CNRM-CM6-1-HR model with a bias-correction procedure to better represent SST characteristics.

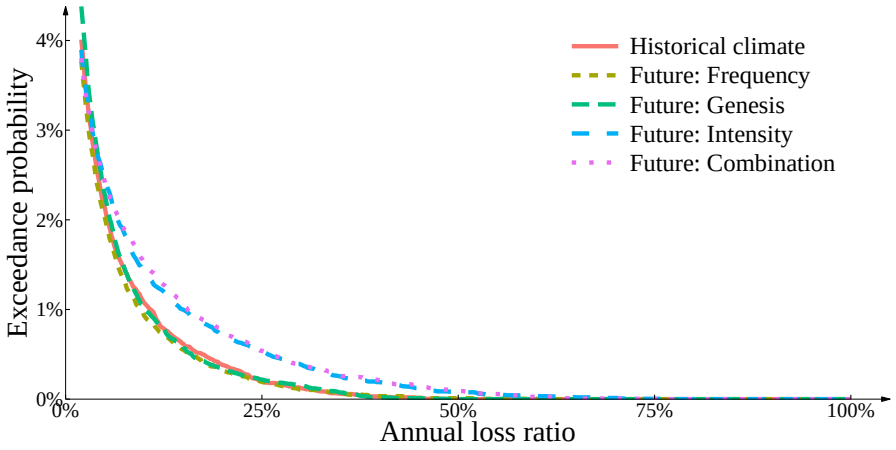


Figure 5.19: Projected distributions of annual residential building loss ratios in Miami-Dade County under historical and future climates.

Fig. 5.20 presents the changes in average annual loss ratios under future climate conditions. Although reduced storm frequency leads to a decrease in average annual loss, the effect of genesis redistribution counteracts this reduction. As a result, the combined-impact scenario produces average annual losses comparable to those under the intensity-only scenario. Overall, incorporating changes in TC frequency, genesis location, and intensity results in a 34.3% increase in average annual loss relative to the historical climate.

5.4 Conclusion

This study proposes a framework for assessing the impacts of climate change on TC behaviour, including intensity, frequency, and genesis locations, using high-resolution GCMs. Projected TC characteristics under future climate conditions are used as inputs to a synthetic TC model to evaluate changes in TC risks for

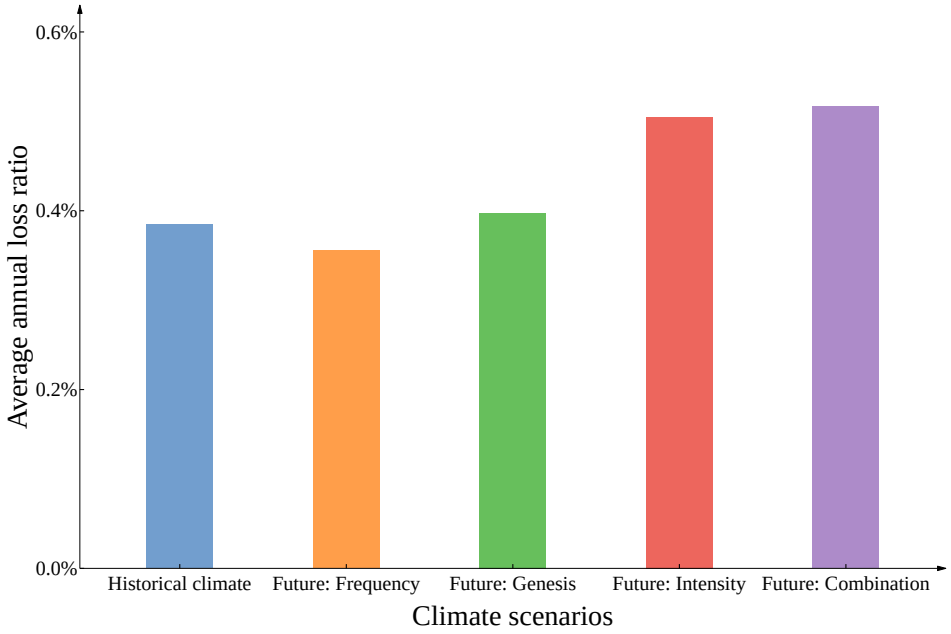


Figure 5.20: Average annual residential building loss ratios in Miami-Dade County under historical and future climates.

coastal cities. The framework is applied to the North Atlantic, using the CNRM-CM6-1-HR model under the SSP585 high-emission scenario to estimate return period wind speeds and associated losses. The main findings are summarised as follows:

1. Climate change is projected to result in a decrease in TC frequency across the North Atlantic, accompanied by a spatial redistribution of genesis activity from the western Caribbean toward the eastern Caribbean and the western–central North Atlantic. SSTs exhibit widespread warming across the basin. The CNRM-CM6-1-HR model shows a systematic cold bias in SST; without bias correction, this bias would lead to an underestimation of TC intensification and consequently an underestimation of coastal hazard.

2. TC intensification is the primary driver of increased risk in the examined

coastal cities. Neglecting concurrent changes in storm frequency and genesis locations leads to an overestimation of risk for cities adjacent to the Gulf of Mexico (e.g., Houston and Mobile), where decreases in frequency and westward reductions in genesis partially offset intensification effects. In contrast, for cities along the southeastern and northeastern U.S. coastline (e.g., Miami, Charleston, and New York), the reduction in storm frequency is partially compensated by the eastward redistribution of genesis activity. As a result, the difference between the intensity-only and combined-impact scenarios is relatively small for these cities.

3. The projected increases in return period wind speeds vary across cities. Among the five locations, Miami and New York exhibit the most pronounced increases. Additionally, increases are larger for long return periods than for short return periods, indicating a heightened likelihood of rare, high-intensity storm events under future climate conditions.

4. The average annual loss ratio for residential buildings in Miami–Dade County is projected to increase by 34.3% relative to historical climate conditions.

This research provides a comprehensive assessment of how climate change may alter TC activity and its impacts on coastal regions. The findings can support the development of more robust engineering design standards and inform climate adaptation strategies for urban planning and disaster management.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Summary and conclusions

This thesis presents a comprehensive and systematic framework for assessing TC-induced risks in urban regions, with an emphasis on improving the efficiency and accuracy of regional loss estimation and evaluating the influence of climate change on TC activity and associated urban risks. The research comprises three studies that focus on quantifying, modelling, and projecting TC risks under both current and future climate conditions.

Chapter 3 introduces a novel framework for constructing a TC loss model for regional building portfolios. Traditional large-area risk assessments face challenges in establishing functional relationships between TC characteristics and losses due to spatial heterogeneity in hazard intensity, terrain conditions, and building distributions. To address this issue, a spatial clustering algorithm was applied to partition the study region into sub-regions with relatively uniform internal characteristics. The clustering considered three key attributes: (1) building spatial density, (2) terrain type represented by surface roughness, and (3) wind hazard risk characterised by return period wind speeds. By grouping areas with similar features, the framework reduces intra-region variability, thus improving the robustness and accuracy of sub-regional wind–loss regressions. The total regional loss was estimated as a weighted sum of sub-regional losses, while the total variance was calculated by incorporating both intra-subregional variances and positive

inter-subregional correlations. Neglecting these correlations would lead to an underestimation of total variance. Testing with 100 TC events showed that the proposed approach achieved RMSEs of 0.023 and 0.016 for the mean and standard deviation of losses, respectively. Compared to direct modelling at the regional scale, this divide-and-aggregate method significantly enhances the precision and reliability of large-scale risk assessments.

Chapter 4 develops a BN-based TC loss model to capture the spatial distribution of building losses across a large region. The model integrates key TC parameters, including translation speed, movement direction, distance and azimuth between the storm centre and the study area, and central pressure, along with loss indicators expressed as loss ratios. This enables the estimation of loss distributions while explicitly accounting for uncertainties and spatial correlations based on only basic TC information. To address the challenge of discretising continuous variables, a novel data-driven discretisation strategy was introduced. For TC variables, weighted PCA combined with decision tree-based methods ensured dimensionality reduction while preserving key information, thereby maintaining higher predictive accuracy in critical areas. For loss variables, K-means clustering was employed to characterise patterns across regions. The model demonstrated high predictive accuracy, with RMSEs for both the mean and standard deviation of loss ratios below 0.02 for most areas. The 90th percentile RMSE values for the mean and standard deviation of the loss estimates across the study area are 0.025 and 0.023, respectively. Furthermore, the BN exhibited strong inferential capabilities under partial observation scenarios: when data for only 5% of locations were available, the model efficiently estimated losses for the remaining 95% of the region.

Chapter 5 explores the potential impacts of climate change on TC behaviour

and associated urban risks. Future climate conditions were projected using the high-resolution CNRM-CM6-1-HR GCM under the SSP585 high-emission scenario. A bias correction was applied to address the model's systematic cold bias in SST prior to storm simulation. Using a synthetic TC generation technique, 30,000 years of stochastic storm activity were simulated for both historical and future climate scenarios. Five coastal cities, Houston, Mobile, Miami, Charleston, and New York, were selected for detailed risk analysis. Results indicate that climate change will lead to a decrease in TC frequency and a spatial redistribution of genesis activity from the western Caribbean toward the eastern Caribbean and the western–central North Atlantic, while TC intensity increases across the basin. TC intensification is the primary driver of heightened future wind hazards. Neglecting concurrent changes in frequency and genesis locations leads to an overestimation of risk for cities adjacent to the Gulf of Mexico, where decreases in storm occurrence and westward reductions in genesis partially offset intensification effects. In contrast, for southeastern and northeastern U.S. cities, the reduction in frequency is partially compensated by the eastward shift in genesis activity, resulting in only small differences between intensity-only and fully combined projections. Projected increases in return period wind speeds are larger for longer return periods, with Miami and New York exhibiting the greatest increases. The average annual loss ratio for residential buildings in Miami–Dade County is projected to rise by approximately 34.3% relative to historical conditions.

Collectively, these three studies form an integrated framework that links hazard simulation, regional loss estimation, and climate modelling. The framework provides an essential foundation for developing long-term strategies to enhance climate resilience and urban adaptation to TC-related hazards.

6.2 Recommendations for future research

This thesis highlights several directions for future work to further improve TC risk assessment and resilience planning:

1. Integration of interdependent network systems. While this study primarily focuses on building damage, other critical infrastructure networks, such as water supply, power grids, and communication systems, are highly interdependent. For example, power systems rely on water for cooling, while water supply depends on electricity for pumping and treatment. Future research should integrate these networked systems into risk models to better capture cascading failures and the full scope of potential disruptions during TC events.

2. Incorporating recovery processes into resilience modelling. Community resilience encompasses four dimensions: robustness, rapidity, resourcefulness, and redundancy. This study addresses robustness by assessing structural damage but does not explicitly consider recovery dynamics. Future work should model the speed and effectiveness of recovery, including resource allocation, and redundancy in system design to provide a more comprehensive assessment of resilience and long-term community functionality.

3. Multi-hazard risk assessment. TCs often bring intense rainfall and storm surge in addition to high winds, and these hazards may occur simultaneously or in close succession. Their combined effects can be significantly greater than those of a single hazard alone. Expanding the current framework to incorporate multi-hazard scenarios, such as wind-driven flooding, would improve risk quantification and support the design of more effective mitigation strategies.

4. Expanding the use of BNs. Beyond regional damage assessment, BNs can serve as powerful tools for causal inference and decision support. Future extensions could integrate emergency response processes, evacuation strategies, and recovery planning. Moreover, adopting dynamic BN frameworks would enable the inclusion of temporal dimensions, facilitating the analysis of evolving risks under changing climate conditions and long-term adaptation planning.

REFERENCES

Ahmad, Z., Ahmed, H. A., Shahzada, K., and Li, Y. (2024). "Vulnerability of non-structural elements (NSEs) in buildings and their life cycle assessment: A review." *Buildings*, 14(1), 170.

Aldrich, D. P. (2017). "The importance of social capital in building community resilience." *Rethinking Resilience, Adaptation and Transformation in a Time of Change*, Springer International Publishing, Cham, Switzerland, 357–364.

Alhamid, A. K., Akiyama, M., Aoki, K., Koshimura, S., and Frangopol, D. M. (2023). "Life-cycle risk assessment of building portfolios subjected to tsunamis under non-stationary sea-level rise based on a compound renewal process." *Earthq. Eng. Struct. Dyn.*, 52(7), 1961–1982.

Andrews, T., Chau, S., Fu, Y., Lubarsky, K., Smith, D., and Waite, J. (2016). "The cost and funding of natural disasters in Australia." *Actuaries Institute General Insurance Seminar*, Melbourne, Australia.

Anwar, G. A., Dong, Y., and Khan, M. A. (2023). "Long-term sustainability and resilience enhancement of building portfolios." *Resilient Cities Struct.*, 2(2), 13–23.

Arthur, W. C. (2021). "A statistical–parametric model of tropical cyclones for hazard assessment." *Nat. Hazards Earth Syst. Sci.*, 21(3), 893–916.

Arunachalam, S. (2024). "Performance-based wind engineering through stochastic simulation and high-fidelity computational modeling." Ph.D. thesis, University of Michigan.

ASCE (2017). *Minimum design loads and associated criteria for buildings and other structures*. ASCE 7-16, Virginia, the United States.

AssunÇão, R. M., Neves, M. C., Câmara, G., and Da Costa Freitas, C. (2006). "Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees." *Int. J. Geogr. Inf. Sci.*, 20(7), 797–811.

Aydin, M., Akyuz, E., and Boustras, G. (2024). "A holistic safety assessment for cargo holds and decks fire & explosion risks under fuzzy Bayesian network approach." *Saf. Sci.*, 176, 106555.

Aydin, O., Janikas, M. V., Assunção, R., and Lee, T.-H. (2018). "SKATER-CON: Unsupervised regionalization via stochastic tree partitioning within a consensus framework using random spanning trees." *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, New York, USA, 33–42.

Baker, J. W. (2008). "An introduction to probabilistic seismic hazard analysis (PSHA)." *Report No. 1.3*, US Nuclear Regulatory Commission, Maryland, the United States.

Balaguru, K., Xu, W., Chang, C.-C., Leung, L. R., Judi, D. R., Hagos, S. M., Wehner, M. F., Kossin, J. P., and Ting, M. (2023). "Increased U.S. coastal hurricane risk under climate change." *Sci. Adv.*, 9(14), 1–11.

Beresniak, A., Bertherat, E., Perea, W., Soga, G., Souley, R., Dupont, D., and Hugonnet, S. (2012). "A Bayesian network approach to the study of historical epidemiological databases: Modelling meningitis outbreaks in the Niger." *Bull. World Health Organ.*, 90, 412–417a.

Beuzen, T., Marshall, L., and Splinter, K. D. (2018). "A comparison of methods for discretizing continuous variables in Bayesian networks." *Environ. Model. Softw.*, 108, 61–66.

Bloemendaal, N., Haigh, I. D., Moel, H., Muis, S., Haarsma, R. J., and Aerts, J. C. J. H. (2020). "Generation of a global synthetic tropical cyclone hazard dataset using STORM." *Sci. Data*, 7(1), 40.

Bloemendaal, N., Moel, H., Martinez, A. B., Muis, S., Haigh, I. D., Wiel, K., Haarsma, R. J., Ward, P. J., Roberts, M. J., Dullaart, J. C. M., and Aerts, J. C. J. H. (2022). "A globally consistent local-scale assessment of future tropical cyclone risk." *Sci. Adv.*, 8(17), 1–13.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., and Von Winterfeldt, D. (2003). "A framework to quantitatively assess and enhance the seismic resilience of communities." *Earthq. Spectra*, 19(4), 733–752.

Cai, B., Zhang, Y., Wang, H., Liu, Y., Ji, R., Gao, C., Kong, X., and Liu, J. (2021). "Resilience evaluation methodology of engineering systems with dynamic-Bayesian-network-based degradation and maintenance." *Reliab. Eng. Syst. Saf.*, 209, 107464.

Chang-Richards, A., Seville, E., Wilkinson, S., and Walker, B. (2019). "Effects of disasters on displaced workers." *Resettlement Challenges for Displaced Populations and Refugees*, Springer International Publishing, Cham, Switzerland, 185–195.

Chen, X., Ma, X., Jia, L., Zhang, Z., Chen, F., and Wang, R. (2024). "Causative analysis of freight railway accident in specific scenes using a data-driven Bayesian network." *Reliab. Eng. Syst. Saf.*, 243, 109781.

Chen, Y., Duan, Z., Yang, J., Deng, Y., Wu, T., and Ou, J. (2021). "Typhoons of western North Pacific basin under warming climate and implications for future wind hazard of east Asia." *J. Wind Eng. Ind. Aerodyn.*, 208, 104415.

Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty.* John Wiley & Sons, Chichester, England.

Chow, S.-h. (1971). "A study of the wind field in the planetary boundary layer of a moving tropical cyclone." M.S. thesis, New York University.

Cimellaro, G. P., Christovasilis, I. P., Reinhorn, A. M., De Stefano, A., and Kirova, T. (2010). "L'aquila earthquake of April 6, 2009 in Italy: Rebuilding a resilient city to withstand multiple hazards.

Col, J.-M. (2007). "Managing disasters: The role of local government." *Public Adm. Rev.*, 67(s1), 114–124.

Contreras, C. and M, P. (2019). "A comparative analysis of population dislocation models for multi-objective community resilience optimization." M.S. thesis, The University of Oklahoma.

Costa, V. G. and Pedreira, C. E. (2023). "Recent advances in decision trees: An updated survey." *Artif. Intell. Rev.*, 56(5), 4765–4800.

Cox, R. S. and Perry, K.-M. E. (2011). "Like a fish out of water: Reconsidering disaster recovery and the role of place and social capital in community disaster resilience." *Am. J. Community Psychol.*, 48(3), 395–411.

Cui, W. and Caracoglia, L. (2019). "A new stochastic formulation for synthetic hurricane simulation over the North Atlantic Ocean." *Eng. Struct.*, 199, 109597.

Cui, W., Zhao, L., Cao, S., and Ge, Y. (2021). "Bayesian optimization of typhoon full-track simulation on the Northwestern Pacific segmented by QuadTree decomposition." *J. Wind Eng. Ind. Aerodyn.*, 208, 104428.

Darling, R. W. R. (1991). "Estimating probabilities of hurricane wind speeds using a large-scale empirical model." *J. Clim.*, 4(10), 1035–1046.

De Felice, F., Baffo, I., and Petrillo, A. (2022). "Critical infrastructures overview: Past, present and future." *Sustainability*, 14(4), 2233.

Der Kiureghian, A. and Liu, P.-L. (1986). "Structural reliability under incomplete probability information." *J. Eng. Mech.*, 112(1), 85–104.

Ding, J.-Y., Feng, D.-C., and Galasso, C. (2025). "Seismic fragility assessment of regional building portfolios using machine learning and Poisson binomial distribution." *Int. J. Disaster Risk Reduct.*, 116, 105044.

Do, T. Q., Lindt, J. W., and Cox, D. T. (2020). "Hurricane surge-wave building fragility methodology for use in damage, loss, and resilience analysis." *J. Struct. Eng.*, 146(1), 4019177.

Doberck, W. (1884). "The Hong Kong Observatory." *Nature*, 29(756), 596–597.

Dueñas-Osorio, L. and Vemuru, S. M. (2009). "Cascading failures in complex infrastructure systems." *Struct Saf*, 31(2), 157–167.

Dvorak, V. F. (1984). "Tropical cyclone intensity analysis using satellite data." *Report No. NESDIS 11*, National Oceanic and Atmospheric Administration, Washington, D.C.

Dwivedi, A. and Yu, X. (2013). "A maximum-flow-based complex network approach for power system vulnerability analysis." *IEEE Trans. Ind. Informat.*, 9(1), 81–88.

Eberenz, S., Lüthi, S., and Bresch, D. N. (2021). "Regional tropical cyclone impact functions for globally consistent risk assessments." *Nat. Hazards Earth Syst. Sci.*, 21(1), 393–415.

Ellingwood, B. R., Celik, O. C., and Kinali, K. (2007). "Fragility assessment of building structural systems in Mid-America." *Earthq. Eng. Struct. Dyn.*, 36(13), 1935–1952.

Ellingwood, B. R., Rosowsky, D. V., Li, Y., and Kim, J. H. (2004). "Fragility assessment of light-frame wood construction subjected to wind and earthquake hazards." *J. Struct. Eng.*, 130(12), 1921–1930.

Ellingwood, B. R. and Tekie, P. B. (1999). "Wind load statistics for probability-based structural design." *J. Struct. Eng.*, 125(4), 453–463.

Ellingwood, B. R., Wang, N., Harris, J. R., and McAllister, T. P. (2018). "Performance-based engineering to achieve community resilience." *Sustainable and Resilient Infrastructure*, Routledge, Oxfordshire, England, 94–111.

Emanuel, K. (2011). "Global warming effects on U.S. hurricane damage." *Weather Clim. Soc.*, 3(4), 261–268.

Emanuel, K. (2021). "Response of global tropical cyclone activity to increasing CO2: Results from downscaling CMIP6 models." *J. Clim.*, 34(1), 57–70.

Emanuel, K., Ravela, S., Vivant, E., and Risi, C. (2006). "A statistical deterministic approach to hurricane risk assessment." *Bull. Amer. Meteor. Soc.*, 87(3), 299–314.

Fan, B., Shu, N., Li, Z., and Li, F. (2023). "Critical nodes identification for power grid based on electrical topology and power flow distribution." *IEEE Syst. J.*, 17(3), 4874–4884.

Fayyad, U. M. (1993). "Multi-interval discretization of continuous-valued attributes for classification learning." *International Joint Conference on Artificial Intelligence*, Vol. 93, Savoie, France, 1022–1029.

FEMA (2014). "Multi-hazard loss estimation methodology: Hurricane model." *Report No. MH 2.1*, Department of Homeland Security, Washington, D.C.

Feng, D. (2024). "Improving hurricane-related flood risk assessments in urban areas: Integrating a Bayesian network modelling approach with an indicator-based approach." Ph.D. thesis, University of Glasgow.

Fenton, N. and Neil, M. (2018). *Risk assessment and decision analysis with Bayesian networks*. CRC Press, Florida, the United States.

Fisher, T. J., Lund, R., and Robbins, M. W. (2020). "A statistical analysis of North Atlantic tropical cyclone changes." *Evaluating Climate Change Impacts*, Chapman and Hall, Abingdon, England, 25–43.

Florida Department of Financial Services (2006). "Task force on long-term solutions for Florida's hurricane insurance market." *Report No. 1*, Florida.

Folke, C., Carpenter, S., Elmqvist, T., Gunderson, L., Holling, C. S., and Walker, B. (2002). "Resilience and sustainable development: Building adaptive capacity in a world of transformations." *AMBIO*, 31(5), 437–440.

Foundation, T. R. "100 resilient cities, <https://www.rockefellerfoundation.org/100-resilient-cities>. Accessed: 2025-03-17.

Freitag, R. C., Abramson, D. B., Chalana, M., and Dixon, M. (2014). "Whole community resilience: An asset-based approach to enhancing adaptive capacity before a disruption." *J. Am. Plann. Assoc.*, 80(4), 324–335.

Galagedarage Don, M. and Khan, F. (2019). "Dynamic process fault detection and diagnosis based on a combined approach of hidden Markov and Bayesian network model." *Chem. Eng. Sci.*, 201, 82–96.

Gao, Z., Geddes, R. R., and Ma, T. (2020). "Direct and indirect economic losses using typhoon-flood disaster analysis: An application to Guangdong province, China." *Sustainability*, 12(21), 8980.

Georgiou, P. (1986). "Design wind speeds in tropical cyclone-prone regions." Ph.D. thesis, The University of Western Ontario.

Goda, K. and Hong, H. P. (2008). "Estimation of seismic loss for spatially distributed buildings." *Earthq. Spectra*, 24(4), 889–910.

Goerlandt, F. and Islam, S. (2021). "A Bayesian network risk model for estimating coastal maritime transportation delays following an earthquake in British Columbia." *Reliab. Eng. Syst. Saf.*, 214, 107708.

Grafton, R. Q., Doyen, L., Béné, C., Borgomeo, E., Brooks, K., Chu, L., Cumming, G. S., Dixon, J., Dovers, S., Garrick, D., Helfgott, A., Jiang, Q., Katic, P., Kompas, T., Little, L. R., Matthews, N., Ringler, C., Squires, D., Steinshamn, S. I., Villasante, S., Wheeler, S., Williams, J., and Wyrwoll, P. R. (2019). "Realizing resilience for decision-making." *Nat. Sustainability*, 2(10), 907–913.

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., and Tuzhilina, E. (2022). "Principal component analysis." *Nat. Rev. Methods Primers*, 2(1), 1–21.

Hagenmeyer, V., Kemal Çakmak, H., Düpmeier, C., Faulwasser, T., Isele, J., Keller, H. B., Kohlhepp, P., Kühnapfel, U., Stucky, U., Waczowicz, S., and Mikut, R. (2016). "Information and communication technology in energy lab 2.0: Smart energies system simulation and control center with an open-street-map-based power flow simulation example." *Energy Technol.*, 4(1), 145–162.

Hall, T. M., Kossin, J. P., Thompson, T., and McMahon, J. (2021). "U.S. tropical cyclone activity in the 2030s based on projected changes in tropical sea surface temperature." *J. Clim.*, 34(4), 1321–1335.

Han, Y., Zhang, M.-Z., Xu, Z., and Guo, W. (2022). "Assessing the performance of 33 CMIP6 models in simulating the large-scale environmental fields of tropical cyclones." *Clim. Dyn.*, 58(5-6), 1683–1698.

Harper, B. (2002). "Tropical cyclone parameter estimation in the Australian region: Wind-pressure relationships and related issues for engineering planning and design." *Report No. J0106-PR003E*, Woodside Australian Energy, Australia.

He, P., Ng, T. S., and Su, B. (2019). "Energy-economic resilience with multi-region input–output linear programming models." *Energy Econ.*, 84, 104569.

Hodges, K., Cobb, A., and Vidale, P. L. (2017). "How well are tropical cyclones represented in reanalysis datasets." *J. Clim.*, 30(14), 5243–5264.

Holland, G. J. (1980). "An analytic model of the wind and pressure profiles in hurricanes." *Mon. Weather Rev.*, 108(8), 1212–1218.

Holland, G. J. (2007). "Misuse of landfall as a proxy for Atlantic tropical cyclone activity." *EOS Trans. Trans. Am. Geophys. Union*, 88(36), 349–350.

Holling, C. S. (1973). "Resilience and stability of ecological systems." *Annu. Rev. Ecol. Syst.*, 460–482.

Homeguide. "How much does it cost to frame a house, <https://homeguide.com/costs/cost-to-frame-a-house>. Accessed: 2025-08-28.

Hong, H. P., Li, S. H., and Duan, Z. D. (2016). "Typhoon wind hazard estimation and mapping for coastal region in mainland China." *Nat. Hazard. Rev.*, 17(2), 4016001.

Hong, X., Hong, H. P., and Li, J. (2019). "Solution and validation of a three dimensional tropical cyclone boundary layer wind field model." *J. Wind Eng. Ind. Aerodyn.*, 193, 103973.

HRD/NOAA. "Detailed list of continental United States hurricane impacts/landfalls, <http://www.aoml.noaa.gov/hrd/hurdat/hurdat2.html>. Accessed: 2023-02-22.

Hsu, S. A. and Yan, Z. (1998). "A note on the radius of maximum wind for hurricanes." *J. Coastal Res.*, 14(2), 667–668.

Hu, J. and Ma, H. (2023). "Distributed real-time optimal power flow strategy for DC microgrid under stochastic communication networks." *J. Mod. Power Syst. Clean Energy*, 11(5), 1585–1595.

Huang, M., Wang, Q., Jing, R., Lou, W., Hong, Y., and Wang, L. (2022). "Tropical cyclone full track simulation in the western North Pacific based on random forests." *J. Wind Eng. Ind. Aerodyn.*, 228, 105119.

Huang, M., Wang, Q., Li, Q., Jing, R., Lin, N., and Wang, L. (2021). "Typhoon wind hazard estimation by full-track simulation with various wind intensity models." *J. Wind Eng. Ind. Aerodyn.*, 218, 104792.

Huang, Z., Rosowsky, D. V., and Sparks, P. R. (2001). "Long-term hurricane risk assessment and expected damage to residential structures." *Reliab. Eng. Syst. Saf.*, 74(3), 239–249.

IPCC (2021). "Climate change 2022: Impacts, adaptation and vulnerability." *Report No. 6*, Intergovernmental Panel on Climate Change, Geneva.

Jarvinen, B. R., Neumann, C. J., and Davis, M. A. S. (1984). "A tropical cyclone data tape for the North Atlantic basin, 1886-1983 : Contents, limitations, and uses." *Report No. NHC 22*, National Hurricane Center, Florida.

Ji, A., He, R., Chen, W., and Zhang, L. (2024). "Computational methodologies for critical infrastructure resilience modeling: A review." *Adv. Eng. Inf.*, 62, 102663.

Jiang, X., Wang, X., Fang, Y., Yang, L., and Tatano, H. (2024). "An agent-based model for the assessment of the ripple effect of disaster economic losses considering firms' adaptive behaviors." *Int. J. Disaster Risk Reduct.*, 111, 104646.

Jing, R. and Lin, N. (2020). "An environment-dependent probabilistic tropical cyclone model." *JAMES*, 12(3), 1–18.

Jun, S. (2021). "Evolutionary algorithm for improving decision tree with global discretization in manufacturing." *Sensors*, 21(8), 2849.

Kaikkonen, L., Parviainen, T., Rahikainen, M., Uusitalo, L., and Lehikoinen, A. (2021). "Bayesian networks in environmental risk assessment: A review." *Integr. Environ. Assess. Manage.*, 17(1), 62–78.

Kalogiannidis, S., Kalfas, D., Chatzitheodoridis, F., and Lekkas, E. (2023). "Role of governance in developing disaster resiliency and its impact on economic sustainability." *J. Risk Financ. Manag.*, 16(3), 151.

Kapucu, N. and Sadiq, A.-A. (2016). "Disaster policies and governance: Promoting community resilience." *Politics Gov.*, 4(4), 58–61.

Kepert, J. (2001). "The dynamics of boundary layer jets within the tropical cyclone core. Part I: Linear theory." *J. Atmos. Sci.*, 58(17), 2469–2484.

Kerber, R. (1992). "ChiMerge: Discretization of numeric attributes." *Proceedings of the Tenth National Conference on Artificial Intelligence*, California, the United States, 123–128.

Kieu, C., Zhao, M., Tan, Z., Zhang, B., and Knutson, T. (2023). "On the role of sea surface temperature in the clustering of global tropical cyclone formation." *J. Clim.*, 36(9), 3145–3162.

Kim, J.-M., Woods, P. K., Park, Y. J., Kim, T., and Son, K. (2016). "Predicting hurricane wind damage by claim payout based on hurricane Ike in Texas." *Geomatics Nat. Hazards Risk*, 7(5), 1513–1525.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J. (2010). "The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data." *Bull. Am. Meteorol. Soc.*, 91(3), 363–376.

Knutson, T., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C.-H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., Walsh, K., and Wu, L. (2020). "Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming." *Bull. Am. Meteorol. Soc.*, 101(3), E303–E322.

Koliou, M., Lindt, J. W., McAllister, T. P., Ellingwood, B. R., Dillard, M., and Cutler, H. (2020). "State of the research in community resilience: Progress and challenges." *Sustain. Resilient Infrastruct.*, 5(3), 131–151.

Kong, J., Zhang, C., and Simonovic, S. P. (2021). "Optimizing the resilience of interdependent infrastructures to regional natural hazards with combined improvement measures." *Reliab Eng Syst Safe*, 210, 107538.

Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M.-J., Subramanian, S. V., and Carson, R. (2002). "Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter." *Am. J. Epidemiol.*, 156(5), 471–482.

Kropf, C. M., Vaterlaus, L., Bresch, D. N., and Pellissier, L. (2025). "Tropical cyclone risk for global ecosystems in a changing climate." *Nat. Clim. Change*, 15(1), 92–100.

Kunitsugu, M. (2012). "Tropical cyclone information provided by the RSMC Tokyo - Typhoon Center." *Trop. Cyclone Res. Rev.*, 1(1), 51–59.

Kyne, D. and Aldrich, D. P. (2020). "Capturing bonding, bridging, and linking social capital through publicly available data." *Risk Hazards Crisis Public Policy*, 11(1), 61–86.

Landsea, C. W., Vecchi, G. A., Bengtsson, L., and Knutson, T. R. (2010). "Impact of duration thresholds on Atlantic tropical cyclone counts." *J. Clim.*, 23(10), 2508–2519.

Law, A. M. (2013). *Simulation modeling and analysis*. McGraw-Hill Education, Dubuque, the United States.

Lee, C.-Y., Camargo, S. J., Sobel, A. H., and Tippett, M. K. (2020). "Statistical–dynamical downscaling projections of tropical cyclone activity in a warming climate: Two diverging genesis scenarios." *J. Clim.*, 33(11), 4815–4834.

Lee, J. Y. and Ellingwood, B. R. (2017). "A decision model for intergenerational life-cycle risk assessment of civil infrastructure exposed to hurricanes under climate change." *Reliab. Eng. Syst. Saf.*, 159, 100–107.

Lee, K. H. and Rosowsky, D. V. (2005). "Fragility assessment for roof sheathing failure in high wind regions." *Eng. Struct.*, 27(6), 857–868.

Lee, K. H. and Rosowsky, D. V. (2007). "Synthetic hurricane wind speed records: Development of a database for hazard analyses and risk studies." *Nat Hazards Rev*, 8(2), 23–34.

Lee, R. and Kiremidjian, A. S. (2007). "Uncertainty and correlation for loss assessment of spatially distributed systems." *Earthq. Spectra*, 23(4), 753–770.

Li, B., Chen, Y., Huang, S., Guan, H., Xiong, Y., and Mei, S. (2020). "A Bayesian network model for predicting outages of distribution system caused by hurricanes." *IEEE Power & Energy Society General Meeting*, Quebec, Canada, 1–5.

Li, Y. (2005). "Fragility methodology for performance-based engineering of wood-frame residential construction." Ph.D. thesis, Georgia Institute of Technology.

Li, Y. and Ellingwood, B. R. (2006). "Hurricane damage to residential construction in the US: Importance of uncertainty modeling in risk assessment." *Eng. Struct.*, 28(7), 1009–1018.

Lin, N., Emanuel, K., Oppenheimer, M., and Vanmarcke, E. (2012). "Physically based assessment of hurricane surge threat under climate change." *Nat. Clim. Change*, 2(6), 462–467.

Lin, P. and Wang, N. (2016). "Building portfolio fragility functions to support scalable community resilience assessment." *Sustain. Resilient Infrastruct.*, 1(3-4), 108–122.

Lin, Y.-S. (2009). "Development of algorithms to estimate post-disaster population dislocation—A research-based approach." Ph.D. thesis, Texas A&M University.

Liu, C., Sun, J., Wang, F., Ning, S., and Xu, G. (2020a). "Bayesian network method for fault diagnosis of civil aircraft environment control system." *Proc. Inst. Mech. Eng. I*, 234(5), 662–674.

Liu, W., Shan, M., Zhang, S., Zhao, X., and Zhai, Z. (2022). "Resilience in infrastructure systems: a comprehensive review." *Buildings*, 12(6), 759.

Liu, X., Chen, B., Chen, C., and Jin, D. (2020b). "Electric power grid resilience with interdependencies between power and communication networks – A review." *IET Smart Grid*, 3(2), 182–193.

Lu, Q. and Zhang, W. (2022). "Integrating dynamic Bayesian network and physics-based modeling for risk analysis of a time-dependent power distribution system during hurricanes." *Reliab. Eng. Syst. Saf.*, 220, 108290.

Mann, M. E. and Emanuel, K. A. (2006). "Atlantic hurricane trends linked to climate change." *EOS Trans. Trans. Am. Geophys. Union*, 87(24), 233–241.

Masoomi, H., Lindt, J. W., Ameri, M. R., Do, T. Q., and Webb, B. M. (2019). "Combined wind-wave-surge hurricane-induced damage prediction for buildings." *J. Struct. Eng.*, 145(1), 4018227.

Matellini, D., Wall, A., Jenkinson, I., Wang, J., and Pritchard, R. (2013). "Modelling dwelling fire development and occupancy escape using Bayesian network." *Reliab. Eng. Syst. Saf.*, 114, 75–91.

McAllister, T. (2016). "Research needs for developing a risk-informed methodology for community resilience." *J. Struct. Eng.*, 142(8), C4015008.

Mehrotra, R. and Sharma, A. (2015). "Correcting for systematic biases in multiple raw GCM variables across a range of timescales." *J. Hydrol.*, 520, 214–223.

Meiler, S., Ciullo, A., Kropf, C. M., Emanuel, K., and Bresch, D. N. (2023). "Uncertainties and sensitivities in the quantification of future tropical cyclone risk." *Commun. Earth Environ.*, 4(1), 371.

Melchers, R. E. and Beck, A. T. (2018). *Structural reliability analysis and prediction.* John Wiley & Sons, New Jersey, the United States.

Memari, M., Attary, N., Masoomi, H., Mahmoud, H., Van De Lindt, J. W., Pilkington, S. F., and Ameri, M. R. (2018). "Minimal building fragility portfolio for damage assessment of communities subjected to tornadoes." *J. Struct. Eng.*, 144(7), 04018072.

Meng, Y., Matsui, M., and Hibi, K. (1995). "An analytical model for simulation of the wind field in a typhoon boundary layer." *J. Wind Eng. Ind. Aerodyn.*, 56(2), 291–310.

Mohebbi, S., Zhang, Q., Christian Wells, E., Zhao, T., Nguyen, H., Li, M., Abdel-Mottaleb, N., Uddin, S., Lu, Q., Wakhungu, M. J., Wu, Z., Zhang, Y., Tuladhar, A., and Ou, X. (2020). "Cyber-physical-social interdependencies and organizational resilience: A review of water, transportation, and cyber infrastructure systems and processes." *Sustainable Cities Soc.*, 62, 102327.

Moriguchi, N., Ito, L., and Tokai, A. (2023). "Risk assessment of chemical release accident triggered by landslide using Bayesian network." *Sci. Total Environ.*, 890, 164321.

Nofal, O. M., Lindt, J. W., Do, T. Q., Yan, G., Hamideh, S., Cox, D. T., and Dietrich, J. C. (2021). "Methodology for regional multihazard hurricane damage and risk assessment." *J. Struct. Eng.*, 147(11), 4021185.

Onisko, A., Druzdzel, M. J., and Austin, R. M. (2019). "Application of Bayesian network modeling to pathology informatics." *Diagn. Cytopathol.*, 47(1), 41–47.

Ouyang, M. (2013). "Comparisons of purely topological model, betweenness based model and direct current power flow model to analyze power grid vulnerability." *Chaos: Interdiscip. J. Nonlinear Sci.*, 23(2), 023114.

Ouyang, M. and Duenas-Osorio, L. (2014). "Multi-dimensional hurricane resilience assessment of electric power systems." *Struct. Saf.*, 48, 15–24.

Ouyang, M., Hong, L., Mao, Z.-J., Yu, M.-H., and Qi, F. (2009). "A methodological approach to analyze vulnerability of interdependent infrastructures." *Simul. Model. Pract. Theory*, 17(5), 817–828.

Ouyang, M. and Wang, Z. (2015). "Resilience assessment of interdependent infrastructure systems: With a focus on joint restoration modeling and analysis." *Reliab. Eng. Syst. Saf.*, 141, 74–82.

Pan, X., Zuo, D., Zhang, W., Hu, L., Wang, H., and Jiang, J. (2021). "Research on human error risk evaluation using extended Bayesian networks with hybrid data." *Reliab. Eng. Syst. Saf.*, 209, 107336.

Pang, W., Liu, F., Fang, S., and Li, Y. (2012). "Spatial correlation and wind speed uncertainties of hurricane wind field model." *Joint Conference of the Engineering Mechanics Institute and the 11th ASCE Joint Specialty Conference on Probabilistic Mechanics and Structural Reliability*, Indiana, the United States.

Pant, S. and Cha, E. J. (2019). "Potential changes in hurricane risk profile across the United States coastal regions under climate change scenarios." *Struct. Saf.*, 80, 56–65.

Peacock, W. G., Lin, Y.-S., Lu, J.-C., and Zhang, Y. (2008). "Household dislocation algorithm 2: An OLS through the origin approach." *Report No. 08-04R*, Hazard Reduction and Recovery Center, Texas.

Pita, G., Pinelli, J. P., Cocke, S., Gurley, K., Mitrani-Reiser, J., Weekes, J., and Hamid, S. (2012). "Assessment of hurricane-induced internal damage to low-rise buildings in the Florida public hurricane loss model." *J. Wind Eng. Ind. Aerodyn.*, 104–106, 76–87.

Powell, M. D., Houston, S. H., Amat, L. R., and Morisseau-Leroy, N. (1998). "The HRD real-time hurricane wind analysis system." *J. Wind Eng. Ind. Aerodyn.*, 77–78, 53–64.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A. (2003). "Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century." *J. Geophys. Res.*, 108(D14), 1–29.

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods.* Springer, New York, the United States.

Roberts, L. M., Kahn, C. E., and Haddawy, P. (1995). "Development of a Bayesian network for diagnosis of breast cancer." *International Joint Conference on AI*, Quebec, Canada.

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., Caron, L.-P., Chauvin, F., Terray, L., Valcke, S., Moine, M.-P., Putrasahan, D., Roberts, C., Senan, R., Zarzycki, C., and Ullrich, P. (2020). "Impact of model resolution on tropical cyclone simulation using the HighResMIP–PRIMAVERA multimodel ensemble." *J. Clim.*, 33(7), 2557–2583.

Rosowsky, D. V. (2021). "Projecting the effects of a warming climate on the hurricane hazard and insured losses: Methodology and case study." *Struct. Saf.*, 88, 102036.

Rosowsky, D. V. and Ellingwood, B. R. (2002). "Performance-based engineering of wood frame housing: Fragility analysis methodology." *J. Struct. Eng.*, 128(1), 32–38.

Rouhana, F., Zhu, J., Bagtzoglou, A. C., and Burton, C. G. (2025). "Analyzing structural inequalities in natural hazard-induced power outages: A spatial-statistical approach." *Int. J. Disaster Risk Reduct.*, 117, 105184.

Russell, L. R. (1969). "Probability distributions for Texas Gulf coast hurricane effects of engineering interest." Ph.D. thesis, Stanford University Stanford.

Saint-Martin, D., Geoffroy, O., Voldoire, A., Cattiaux, J., Brient, F., Chauvin, F., Chevallier, M., Colin, J., Decharme, B., Delire, C., Douville, H., Guérémy, J.-F., Joetzjer, E., Ribes, A., Roehrig, R., Terray, L., and Valcke, S. (2021). "Tracking changes in climate sensitivity in CNRM climate models." *JAMES*, 13(6), 1–19.

Salarieh, B., Ugwu, I. A., and Salman, A. M. (2023). "Impact of changes in sea surface temperature due to climate change on hurricane wind and storm surge hazards across US Atlantic and Gulf coast regions." *SN Appl. Sci.*, 5(8), 205.

Salim, M. Z., Qiang, Y., Dixon, B., and Collins, J. (2024). "A disparate disaster: Spatial patterns of building damage caused by hurricane Ian and associated socio-economic factors." *Remote Sens.*, 16(20), 3792.

Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). "Finding a Kneedle in a haystack: Detecting knee points in system behavior." *International Conference on Distributed Computing Systems Workshops*, Washington, D.C., the United States, 166–171.

Schloemer, R. W. (1954). *Analysis and synthesis of hurricane wind patterns over lake Okeechobee, Florida*. US Department of Commerce, Washington, D.C., the United States.

Seixas, F. L., Zadrozny, B., Laks, J., Conci, A., and Muchaluat Saade, D. C. (2014). "A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment." *Comput. Biol. Med.*, 51, 140–158.

Sesen, M. B., Nicholson, A. E., Banares-Alcantara, R., Kadir, T., and Brady, M. (2013). "Bayesian networks for clinical decision support in lung cancer care." *PLOS One*, 8(12), e82349.

Shapiro, D., Rojahn, C., Reaveley, L. D., Smith, J. R., and Morelli, U. (2000). "NEHRP guidelines and commentary for the seismic rehabilitation of buildings." *Earthq. Spectra*, 16(1), 227–239.

Sheng, C., Fan, Q. Y., and Hong, H. P. (2024). "Estimating intraevent and interevent variability and spatial correlation of tropical cyclone wind fields and their use for the risk assessment of a portfolio of structures." *Struct. Saf.*, 108, 102440.

Smith, A. B. "An active year of U.S. billion-dollar weather and climate disasters, <https://www.climate.gov/news-features/blogs/beyond-data/2024-active-year-us-billion-dollar-weather-and-climate-disasters>. Accessed: 2025-03-15.

Snaiki, R. and Wu, T. (2020). "Revisiting hurricane track model for wind risk assessment." *Struct. Saf.*, 87, 102003.

Suess, E. A. and Trumbo, B. E. (2010). *Introduction to probability simulation and Gibbs sampling with R.* Springer, New York, the United States.

Teixeira, L. V., Assuncao, R. M., and Loschi, R. H. (2015). "A generative spatial clustering model for random data through spanning trees." *IEEE International Conference on Data Mining*, New Jersey, the United States, 997–1002.

Timmerman, P. (1981). "Vulnerability, resilience and the collapse of society." *Report No. 1*, Institute for Environmental Studies, Toronto, Canada.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., and Reed, K. A. (2021). "TempestExtremes v2.1: A community framework for feature detection, tracking, and analysis in large datasets." *Geosci. Model Dev.*, 14(8), 5023–5048.

United Nations Office for Disaster Risk Reduction. "Disaster resilience scorecard for cities, <https://mcr2030.undrr.org/disaster-resilience-scorecard-cities>. Accessed: 2025-03-17.

U.S. Department of Energy. "Presidential Policy Directive 21, <https://www.energy.gov/ceser/presidential-policy-directive-21>. Accessed: 2025-03-17.

Valamanesh, V., Myers, A. T., Arwade, S. R., Hajjar, J. F., Hines, E., and Pang, W. (2016). "Wind-wave prediction equations for probabilistic offshore hurricane hazard analysis." *Nat. Hazard.*, 83(1), 541–562.

Van Verseveld, H. C. W., Van Dongeren, A. R., Plant, N. G., Jäger, W. S., and Den Heijer, C. (2015). "Modelling multi-hazard hurricane damages on an urbanized coast with a Bayesian network approach." *Coast. Eng.*, 103, 1–14.

Vecchi, G. A., Landsea, C., Zhang, W., Villarini, G., and Knutson, T. (2021). "Changes in Atlantic major hurricane frequency since the late-19th century." *Nat. Commun.*, 12(1), 4054.

Vertommen, I., Mitrović, D., Van Laarhoven, K., Piens, P., and Torbeyns, M. (2022). "Optimization of water network topology and pipe sizing to aid water utilities in deciding on a design philosophy: A real case study in Belgium." *Water*, 14(23), 3973.

Vickery, P. J. (2005). "Simple empirical models for estimating the increase in the central pressure of tropical cyclones after landfall along the coastline of the United States." *J. Appl. Meteorol.*, 44(12), 1807–1826.

Vickery, P. J., Masters, F. J., Powell, M. D., and Wadhera, D. (2009a). "Hurricane hazard modeling: The past, present, and future." *J Wind Eng Ind Aerod*, 97(7), 392–405.

Vickery, P. J. and Skerlj, P. F. (2005). "Hurricane gust factors revisited." *J. Struct. Eng.*, 131(5), 825–832.

Vickery, P. J., Skerlj, P. F., and Twisdale, L. A. (2000). "Simulation of hurricane risk in the U.S. using empirical track model." *J. Struct. Eng.*, 126(10), 1222–1237.

Vickery, P. J. and Wadhera, D. (2008). "Statistical models of Holland pressure profile parameter and radius to maximum winds of hurricanes from flight-level pressure and H*Wind data." *J. Appl. Meteorol. Climatol.*, 47(10), 2497–2517.

Vickery, P. J., Wadhera, D., Twisdale, L. A., and Lavelle, F. M. (2009b). "U.S. hurricane wind speed risk and uncertainty." *J. Struct. Eng.*, 135(3), 301–320.

Vitoontus, S. (2012). "Risk assessment of building inventories exposed to large scale natural hazards." Ph.D. thesis, Georgia Institute of Technology.

Voldoire, A. "CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 HighResMIP, <https://doi.org/10.22033/ESGF/CMIP6.1387>. Accessed: 2024-08-02.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas Y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R. (2019). "Evaluation of CMIP6 DECK experiments with CNRM-CM6-1." *JAMES*, 11(7), 2177–2213.

Wang, C., Li, Q., Zhang, H., and Ellingwood, B. R. (2017). "Modeling the temporal correlation in hurricane frequency for damage assessment of residential structures subjected to climate change." *J. Struct. Eng.*, 143(5), 4016224.

Wang, C., Liu, Y., Zhang, X., Li, X., Paramygin, V., Sheng, P., Zhao, X., and Xu, S. (2024). "Scalable and rapid building damage detection after hurricane Ian using causal Bayesian networks and InSAR imagery." *Int. J. Disaster Risk Reduct.*, 104, 104371.

Wang, C., Teh, L. H., and Feng, K. (2023). "A probabilistic framework for resilience quantification of residential building portfolios exposed to tropical cyclone winds." *ASCE-ASME J. Risk Uncertain. Eng. Syst. A: Civ. Eng.*, 9(4), 4023040.

Wang, H. S. (2023). "A unified dynamic model of electrodermal activity." Ph.D. thesis, Northeastern University.

Wang, Y., Rahimi-Golkhandan, A., Chen, C., Taylor, J. E., and Garvin, M. J. (2019). "Measuring the impact of transportation diversity on disaster resilience in urban communities: Case study of hurricane harvey in Houston, TX." *International Conference on Computing in Civil Engineering*, Georgia, the United States, 555–562.

Wang, Y. and Rosowsky, D. V. (2018). "Hazard-based regional loss estimation considering hurricane intensity, size and sea surface temperature change." *Sustainable Resilient Infrastruct.*, 3(4), 151–164.

Wang, Y., Wang, N., Lin, P., Ellingwood, B., and Mahmoud, H. (2020). "Lifecycle analysis (LCA) to restore community building portfolios by building back better II: Decision formulation." *Struct. Saf.*, 84, 101921.

Wang, Z., Chen, G., Hill, D. J., and Dong, Z. Y. (2016). "A power flow based model for the analysis of vulnerability in power networks." *Phys. A: Stat. Mech. Appl.*, 460, 105–115.

Weaver, M. M. and Garner, A. J. (2023). "Varying genesis and landfall locations for North Atlantic tropical cyclones in a warmer climate." *Sci. Rep.*, 13(1), 5482.

Western, J., Stimson, R., Baum, S., and Gellecum, Y. V. (2005). "Measuring community strength and social capital." *Regional Studies*, 39(8), 1095–1109.

Wikipedia. "List of costliest Atlantic hurricanes, <https://en.wikipedia.org>. Accessed: 2025-03-17.

Wikipedia. "List of costliest tropical cyclones, <https://en.wikipedia.org>. Accessed: 2025-07-17.

Wikipedia. "List of the deadliest tropical cyclones, <https://en.wikipedia.org>. Accessed: 2025-03-17.

Wilkins, R. and McCarthy, M. (2011). "National strategy for disaster resilience." *Report No. 1*, National Emergency Management Committee, Canberra, Australia.

Xie, L., Wang, N., Lin, P., and Mahmoud, H. (2025). "Spatiotemporal forecasting of functionality states of community building portfolios under flood evolution." *J. Struct. Eng.*, 151(2), 4024218.

Xie, W., Li, N., Wu, J.-D., and Hao, X.-L. (2014). "Modeling the economic costs of disasters and recovery: Analysis using a dynamic computable general equilibrium model." *Nat. Hazards Earth Syst. Sci.*, 14(4), 757–772.

Yan, D. and Zhang, T. (2022). "Research progress on tropical cyclone parametric wind field models and their application." *Reg. Stud. Mar. Sci.*, 51, 102207.

Yu, X., Wu, Y., Meng, F., Zhou, X., Liu, S., Huang, Y., and Wu, X. (2024). "A review of graph and complex network theory in water distribution networks: Mathematical foundation, application and prospects." *Water Res.*, 253, 121238.

Zeng, D., Li, Q.-S., and Chan, P. (2024). "A geographically weighted regression full-track model of tropical cyclones for typhoon hazard assessment in coastal regions of China." *J. Wind Eng. Ind. Aerodyn.*, 246, 105660.

Zeng, D., Zhang, H., Dai, H., and Beer, M. (2023). "Scalable risk assessment of large infrastructure systems with spatially correlated components." *Struct. Saf.*, 101, 102311.

Zeng, D., Zhang, H., Li, Q., and Ellingwood, B. R. (2021). "Tropical cyclone damage assessment of distributed infrastructure systems under spatially correlated wind speeds." *Struct. Saf.*, 91, 102080.

Zeng, D., Zhang, H., and Wang, C. (2020). "Modelling correlated damage of spatially distributed building portfolios under scenario tropical cyclones." *Struct. Saf.*, 87, 101978.

Zhang, J., Li, G., and Zhang, M. (2022). "Multi-objective optimization for community building group recovery scheduling and resilience evaluation under earthquake." *Comput.-Aided Civ. Infrastruct. Eng.*, 38(12), 1657–1676.

Zhang, W., Lin, P., Wang, N., Nicholson, C., and Xue, X. (2018). "Probabilistic prediction of postdisaster functionality loss of community building portfolios considering utility disruptions." *J. Struct. Eng.*, 144(4), 4018015.

Zhang, W. and Nicholson, C. (2016). "A multi-objective optimization model for retrofit strategies to mitigate direct economic loss and population dislocation." *Sustainable Resilient Infrastruct.*, 1(3-4), 123–136.

Zhao, L., Cui, W., Fang, G., Cao, S., Zhu, L., Song, L., and Ge, Y. (2024). "State-of-the-art review on typhoon wind environments and their effects on long-span bridges." *Adv. Wind Eng.*, 1(1), 100007.

Zheng, Y., Xie, Y., and Long, X. (2021). "A comprehensive review of Bayesian statistics in natural hazards engineering." *Nat. Hazard.*, 108(1), 63–91.

Zhou, R., Fang, W., and Wu, J. (2020). "A risk assessment model of a sewer pipeline in an underground utility tunnel based on a Bayesian network." *Tunnelling Underground Space Technol.*, 103, 103473.