3-8-2022 11:30 AM

Characterizing the Concurrent Occurrence of Tornadoes and Flash Floods Across Canada

Yeu Deck Ngui, *The University of Western Ontario*

Supervisor: Najafi, Reza M., *The University of Western Ontario*
Co-Supervisor: Souza, Camila P. E., *The University of Western Ontario*
Co-Supervisor: Sills, David M. L., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Engineering Science degree in Civil and Environmental Engineering

Abstract

Compound weather extreme events, such as tornadoes and flash floods, can significantly impact societies and infrastructure systems. Disaster response agencies provide instructions to the exposed communities to retreat to safety specific to the natural hazard. However, the instructions can become confusing if natural hazards demand conflicting responses. This study characterizes the compound tornado and flash flood (TORFF) events to assess and predict the simultaneous occurrence probability of such hazards across Canada in the long term. We quantify dependencies between the tornadoes and flash floods using ground-based and reanalysis datasets. Tornado data are available based on the recorded Fujita rating for each event, and the corresponding wind speed values are determined through a resampling approach. The TORFF events are clustered and the bivariate probability distributions of the resampled windspeed and precipitation are characterized based on Copula. The corresponding individual and joint return levels are investigated under different scenarios (AND, OR, and conditional) across Canada. Results show positive dependencies between resampled windspeed and associated precipitation in Saskatchewan, followed by Alberta, Manitoba, Ontario, and Quebec (least dependency) regions. Higher dependencies between tornadoes and flash floods over regions such as Saskatchewan suggest that analyzing these events in isolation can underestimate the associated risks. Higher precipitation is also expected during extreme wind speed, as observed in the conditional assessment of precipitation given windspeed. This study provides insight for more realistic recurrence interval estimation for tornadoes and flash floods to aid in the evacuation decision-making process.

Keywords: Compound Events, Copula, Tornado, Precipitation, Joint Probability, Return Periods, NARR, Climate Extremes

Lay Summary

Tornadoes and flash floods are among the most dangerous natural hazards in Canada, especially in populated areas (Environment Canada 2017). Since the 1970s, they have caused a total of more than $CAD 10 billion in damage in Canada. More than 1,800 tornado events are recorded in Environment Canada Tornado Database (Environment Canada, 2020) between 1980-2009. Certain types of thunderstorms generate tornadoes that come with associated precipitation. Intense rainfall in a short period might cause flash floods in certain areas. The combination of multiple extreme events can cause more catastrophic consequences compared to individual extreme occurrences. Unique concerns arise when tornado and flash flood (TORFF) events occur simultaneously due to their specific evacuation instructions: seek underground shelter (tornado) and retreat to high ground (flash floods), which are contradicting with each other. Having a better understanding of TORFF risks and their impacts can help inform emergency response authorities in providing accurate evacuation instructions. Windspeed data were obtained through a unique resampling approach from Fujita rating windspeed interval records, and the corresponding precipitation data were extracted from a reanalysis product (i.e., NARR) based on the location of the tornado event. The dependency between resampled windspeed and associated precipitation is investigated after grouping tornado events into multiple clusters across Canada. The datasets are then analyzed using a multivariate probabilistic model, copula, and their joint variability are assessed under different return period scenarios. Lastly, the areas with higher dependency and subsequently TORFF risks are identified in Canada so that the appropriate response and evacuation instructions can be issued.

Acknowledgement

I will start by acknowledging my supervisors, Dr. Mohammad Reza Najafi, Dr. Camila de Souza, and Dr. David Sills who gave me the opportunity to work under their guidance. I will always be grateful for their tremendous support during my time at Western University, especially their patience when I struggled a lot with my research project during COVID-19, and after I had my surgery. Without their assistance and dedicated involvement in every step throughout the process, the project would have never been accomplished. Thanks Dr. Najafi again for giving his time generously and being patient with me towards the end of the project due to medical complication. I would not complete this thesis without his support and constructive suggestions. I would like to thank you very much for your support and understanding over these past two years. I am also thankful to my research group members, especially Farshad for his patience and guidance to assist me with any technical difficulties I experienced for my research.

I wish to acknowledge the funding provided by IDI multi-hazard initiative that let me work on this challenging but rewarding project. I would like to thank Western University, as an institution, for providing an excellent infrastructure for a graduate student to thrive. In addition, I would also like to thank the open-source community who provide readily available tools and knowledge that make this project possible. I would also like to extend my gratitude to my examination committee for reviewing this dissertation. I appreciate your time, feedback, and valuable advice toward improving this thesis.

Finally, I would like to thank my cat, Toast, for providing me mental health support during this difficult COVID-19 era, where human interaction is reduced to minimal. I would also like to express my gratitude towards my family and friends for supporting me through this journey.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background

Weather-related hazards have caused increasing losses over the years, especially in North America despite the significant advancements in technology during recent decades (Bouwer 2011). Their impacts affect societies greatly in terms of social, economic, and environmental aspects.

Various atmospheric hazards can cause loss of life and severe damage to properties. Tornadoes and flash floods are two of the most impactful, specifically in the United States (NIELSEN, et al. 2015). In Canada, tornadoes and flash floods are included in the top five costliest natural hazards since 1970 (Public Safety Canada 2019). According to Public Safety Canada (PSC), floods are one of the most common natural hazards in Canada that can occur at any time of the year, most often caused by heavy rainfall (Public Safety Canada 2019). The estimated total cost of flooding since the year 1970 sums up to approximately 9.7 billion Canadian Dollars (CAD), of which CAD 6.85 billion is associated with the events between 2010 to 2019 (Public Safety Canada 2019).

Further, Canada is commonly struck by tornadoes with a total of over 1800 events from 1980 till 2009 (Environment Canada, 2020). The total estimated cost for tornado events sums up to approximately CAD 1.1 billion since 1970 (Public Safety Canada 2019). This number is lower than Canada's neighbour in the south, the United States, which is partly associated with differences in population and urbanization between the two countries. In the year 2019, for example, tornadoes in the United States cost a total of over USD 3.1 billion (almost three times more than Canada in 30 years) in property damage. However, tornado impacts are still quite significant in Canada. Even though the population-normalized yearly fatality rate related to

tornadoes has been decreasing steadily over the years, the continuous urbanization could lead to more catastrophic events in the future (Nielsen et al. 2015).

Natural hazards can overwhelm human societies and ecosystems, especially when two or more natural hazards occur simultaneously or successively amplifying the magnitude of the damage and casualties. Concurrent weather extreme events often exceed the impact caused by a single extreme event (Zscheischler et al. 2021). Since tornadoes and floods have the largest impact on urbanized areas, a unique concern arises when these two natural hazards occur simultaneously (known as TORFF events) at the same location. Effective communication is necessary for the communities at risk when it comes to evacuation. Typically, disaster response agencies provide instructions, specific to the natural hazard, to the exposed communities to retreat to safety. For example, the tornado safety protocol suggests taking underground shelter, whereas flood safety protocol recommends retreating to high ground. However, when the community at risk receives both tornado and flash flood warnings at the same time, the instruction could be conflicting and add confusion to lifesaving actions (Nielsen et al. 2015). For example, tornado and flash flood warnings were issued by the same emergency response authority one minute apart in New Jersey, the USA back in September 2021. The tornado warning asked the community at risk to seek underground shelter to avoid getting injured by the debris while the flash flood warning required them to retreat to high ground (Feuerstein 2021). Such conflicting instructions add more confusion to the community especially when it is a life and death situation.

Such concurrent extreme events, also known as compound events, are defined as “a combination of multiple drivers and/or hazards that contribute to societal or environmental risk” (Zscheischler et al. 2020). Hurricane Harvey in 2017 serves as an example of a major compound event. The hurricane developed with 215 km/h of windspeed, generating 52 tornadoes (National Oceanic

and Atmospheric Administration 2017), and resulted in over one trillion gallons of rain over Houston in just four days causing severe flooding (Environmental Protection Agency 2018).

Having a better understanding of the interaction between hazard drivers will help in providing more insight when it comes to the prediction and mitigation of natural hazards. Since compound weather events will amplify the impacts (e.g., property damage, casualties), it is important to understand better the interactions between hazards to prevent the underestimation of societal and environmental risks (Najafi and Singh 2020). In recent years, compound weather events have drawn significant attention in the scientific communities to look more into the mechanisms and assess the resulting risks (Zscheischler et al. 2020).

This thesis studies the dependencies between the drivers of interest: resampled tornado windspeed and associated precipitation in Canada.

1.2 Research Gaps

The majority of populated areas in Canada are in tornado- (Sills, et al. 2012) and flood-prone regions (Ayushi Gaur et al. 2019). Therefore, there might be a high likelihood for the occurrence of TORFF events that can result in more severe damage compared to them occurring individually, especially when human lives are involved.

Not having an explicit evacuation instruction when the two hazards in TORFF event are demanding conflicting procedures can add confusing elements to the community at risk. Such confusion makes it exceptionally challenging for the operational agencies to make the right evacuation decision, mainly due to the complex physical processes and the amplified danger of multiple hazard events that the community is exposed to simultaneously (Rogash and Smith 2000).

Although several studies have been conducted on tornado and flash flood events individually, research on the characteristics of TORFF events and the associated meteorology and climatology is quite limited. For example, Rogash et al.(2000) discussed the meteorological characteristics related to significant tornado events and determined the main meteorological setups that lead to nearby tornado and flash flood events from the previously mentioned literature. However, they did not investigate the amplified hazards of the directly collocated concurrent tornado and flash flood events (Rogash and Smith 2000).

Wind speed and precipitation were typically studied in isolation instead of compound events. In those studies, extreme values (instead of event-based) such as the 99^{th} percentile are commonly utilized on wind speed and precipitation as individual univariate analysis. Furthermore, studies on extreme wind and precipitation are conducted regionally over the Pacific Ocean (Back and Bretherton 2005), topographically, and seasonally (Zscheischler et al. 2021; Owen et al. 2021)

over a specified temporal window but not over Canada. Finally, despite dependency studies on wind speed and precipitation (Zscheischler et al. 2021; Owen et al. 2021; Barcikowska et al. 2018; Waliser and Guan 2017), there is a lack of research on the dependencies between wind speed associated with tornadoes and the concurrent extreme precipitation events, specifically in Canada.

1.3 Research Questions

Considering the research gaps discussed in the previous section, the following research questions are addressed in this thesis:

1. Which areas in Canada experience positive windspeed/precipitation dependency that can potentially lead to TORFF events?
2. What are the characteristics of extreme precipitation and tornado events over Canada considering different scenarios? (See examples below)
    - The probability of BOTH resampled windspeed AND associated precipitation to exceed predefined joint probability.
    - The probability of EITHER resampled windspeed OR associated precipitation to exceed predefined joint probability.
    - The conditional probability of associated precipitation given the predefined probability of resampled windspeed.
3. What are the uncertainties associated with tornado wind speeds in the process of model development?
4. What information can we deduce through the developed model? Does it provide insight into the decision-making process?

1.4 Research Objectives

The overall objective of this study is to investigate the dependencies between tornado windspeed and the associated precipitation and assess their joint return levels and the associated uncertainties under different conditions.

- The proposed research aims to address the following sub-objectives:
    - Investigate the joint variability of resampled tornado wind speed and the associated precipitation, including 30 years of data.
    - Determine areas in Canada with higher windspeed/precipitation dependency that can lead to TORFF events.
    - Investigate the joint return levels corresponding to extreme precipitation and tornado windspeeds across Canada under different scenarios, including but not limited to:
        - AND scenarios- when **both** variables exceeded predefined thresholds
        - OR scenarios- when **either** one of the variables exceeded predefined thresholds
        - Conditional scenarios- the return level of precipitation **conditioned** on wind speed
    - Assess if the developed model can provide helpful information on evacuation instruction decision-making.

1.5 Dissertation Structure

The dissertation contains six major chapters: Introduction, Literature Review, Study Area and Data Availability, Methodology, Results and Discussion, and Conclusion.

- Chapter 1 consists of the broad background and motivation for this study. Research questions and their corresponding objectives are listed in this chapter, including the thesis outline.
- Chapter 2 includes reviewed literature related to the type of thunderstorms that generate heavy precipitation, studies of extreme events, precisely wind speed and precipitation in Canada, and historic TORFF events in North America. Primary methodologies for dependency studies are also reviewed in this section.
- Chapter 3 discusses the study area and data used in this study. Furthermore, historic tornado events were discussed in detail, including average tornadoes per year, active months, and tornado development time.
- Chapter 4 discusses the methodologies for clustering the tornadoes in Canada and the data generation method for the primary analysis based on copula. Investigation of the dependencies between resampled windspeed and associated precipitation through model development and comparison of joint probabilities and return periods under different scenarios were discussed.
- Chapter 5 focuses on the dependencies between resampled windspeed and associated precipitation for each cluster. Copula model selection and its uncertainty were also discussed in this chapter. Different return period scenarios were computed and compared among clusters to assess their dependency. Comparison of bivariate, univariate, and independence hazards are discussed, including their conditional probability assessment.

- Chapter 6 concludes the research results. First, the main findings are highlighted to make sure research objectives are addressed, followed by its implication, research limitations, and recommendations for future research.

Chapter 2 Literature Review

2.1 Weather Extremes in Canada

2.1.1 Extreme Winds, Tornadoes, and Flash Floods across Canada

Before studying compound weather extreme events and their associated dependencies between the variables of interest, it is vital to understand the underlying mechanisms. Understanding their risk and behavior can also provide researchers with the foundation in initiating their studies. Studies such as (Raymond et al. 2020; Zscheischler et al. 2020) helped to provide the big picture for compound events to further their investigation from a different perspective.

As mentioned before, due to the non-linear nature of climate extremes and their potential to cause a significant impact on economic and social human activities, recent research started shifting focuses on climate extremes analysis. Even though extreme events are rare, they still occur and break the historic records regardless. Therefore, it is crucial to understand to what extent these events are characterized by internal variability or anthropogenic effects (Mahmoudi et al. 2021; Najafi, Zwiers, and Gillett 2017).

Extreme Winds and Tornadoes

Extreme winds have been widely studied in Canada by different researchers. They are studied for different purposes, such as forecasting. For example, wind speed information was studied in British Columbia to predict extreme wind recurrences by estimating a Generalized Pareto Distribution (GPD) to represent extreme windspeeds considering the presence of climate variability covariates (Abeysirigunawardena et al. 2009). They concluded that extreme windspeeds respond significantly differently to different scenarios (warm and cold) of El Niño Southern Oscillation (ENSO) modes with the possibility for high extreme windspeed to occur during the cold ENSO phase.

The Tornado database in Canada was updated in 2012 by Environment Canada to define "tornado-prone" regions in Canada. Most of the population areas in Canada are in tornado-prone regions. (Sills, et al. 2012) It is essential to have a better understanding of predicting and tornado-genesis environment in that region to increase the ability to forecast the tornado and prepare for the hazard.

Studies on tornadoes have been conducted since the 1960s (McKay 1960) focused on investigating tornadoes in western Canada and later moving the investigation to eastern Canada in the 1980s. (Asmis 1980) Early studies looked at general occurrence locations and their seasonally and daily climatology. They were often compared to tornadoes in the USA.

With the advancement of technology and the evolution of analysis approaches on natural hazards, probabilistic approaches have become more prevalent. For example, sophisticated methods such as Bayesian modeling were implemented to elucidate the spatiotemporal patterns of tornado activity in North America to infer the likelihood of tornado occurrence in any location

within a specific temporal window (Cheng et al. 2015). Recently, a preferred prediction model for the spatially varying tornado occurrence rate is developed for Canada using statistical methods such as Bayesian inference and maximum likelihood (Huang, Jiang, and Hong 2021).

Extreme Precipitation and Flash Floods

One of the most common factors contributing to flash floods is intense precipitation in a short period. If the surface is impervious such as urban areas or affected by the soil's antecedent moisture content, the effects can be exacerbated (Feng, Zhang, and Bourke 2021). This causes the precipitation flow to travel towards the low ground by gravity instead of infiltrating into the soil.

Studies of extreme precipitation in Canada are essential since flash flood causes the most severe damage to populated areas. Trend analyses dated back to the 1990s were also conducted to estimate changes in the flooding based on heavy precipitation that showed upward trends since the 1950s (Kunkel, Andsager, and Easterling 1999). Seasonal patterns of extreme precipitation were also investigated since precipitation will affect the hydrology of a region coupled with the accumulated snow on the ground. In addition to seasonal investigation, the relationship between precipitation and ENSO was also assessed to improve forecasting skills. (Zeng et al. 2011) In addition to studying extreme precipitation temporally, spatial characteristics of heavy precipitation were investigated to provide a comprehensive characterization across the country. (X. Zhang, Hogg, and Mekis 2001)

The studies mentioned previously were based on ground observation (e.g., rain gauge). However, due to unevenly distributed precipitation gauging stations and incomplete sub-daily precipitation data in Canada, gridded data such as NARR and ERA-Interim were used to study extreme

precipitation for application such as developing intensity-duration-curve (IDF) (Abhishek Gaur, Schardong, and Simonovic 2020).

There are limited studies of flash floods in Canada, where most of them focus on different types of flooding, such as river flooding and coastal flooding (Thompson, Bernier, and Chan 2009; Khalafzai, McGee, and Parlee 2021; 2019; Batchabani, Sormain, and Fuamba 2016).

Lin et al. (2002) analyzed a flash flood in Quebec, Canada in 1996 due to intense rainfall within 48 hours in a basin considered the most devastating flood event in modern days. The flooding destroyed reservoir dikes and the community at risk, which cost $CAD 700 million as a result. The flash flood event was studied to improve the flood warning system to be more accurate and timelier by developing a coupled atmospheric-hydrological modeling system to simulate flash floods. The model was then applied to different case studies and deemed feasible for flash flood forecasts.

Due to Canada's lack of a flash flood database, the precipitation data from reanalysis products are commonly used to estimate flash flood events. Therefore, studies were done on implementing reanalysis products to examine daily precipitation characteristics (Becker, Berbery, and Higgins 2009). In addition, (Dong et al. 2019) utilized NARR to explore the contributions of extreme (threshold of 95^{th} percentile) and non-extreme precipitation in California. Their results suggested increased extreme precipitation due to enhanced extreme intensity. Further, they found that more extreme days dominate the wetter winter, while decreased non-extreme precipitation due to fewer wet days induces the dryer spring and fall.

2.1.2 Historical TORFF Events in North America

- City of Corpus Christi, Texas, November 15^{th}, 2001

Seventeen tornadoes were generated during the afternoon and evening of November 15, 2001, near the northern part of Corpus Christi and New Braunfels. Flash flooding has been happening for 48 hours since the day before. (Bunkers and Doswell 2016)

- City of Superior, Nebraska, June 22nd, 2003

According to Wakimoto et al. (2004) and (Guyer and Ewald 2004), the supercells on June 22nd, 2003, had led to localized but intense TORFF events with at least two high precipitation supercells that generated ten tornadoes and flash floods in two adjacent counties. The TORFF event was determined by looking at National Centers for Environmental Information (NCEI) data. The flash flood was reported within 20 mins and 77 mins of the last tornado in each county, respectively. (Bunkers and Doswell 2016)

- Southern Ontario, Ontario, August 19, 2005

Two confirmed tornadoes were generated due to a supercell thunderstorm that started from Stratford, ON, and Lake Huron as advancing cold front, causing 175mm of rain just the north of Toronto City limits in less than one hour. The intense rainfall caused flooding in less than one hour and overflowed the storm drains. Consequently, it caused severe basement flooding to many thousands of homes. In addition, the flood water also washed out a portion of the parking lot of a specific area. (Environment Canada 2017)

- City of Bennington, Kansas, May 28, 2013

A near stationary high precipitation supercell system-generated flash flooding had occurred in north-central Kansas during the afternoon and early evening of May 28, 2013. Two tornadoes were generated, and flash floods occurred in 35 minutes (Bunkers and Doswell 2016).

2.1.3 Compound Impacts of Severe Weather Events

As mentioned in Chapter 1, compound extreme weather events would ultimately amplify the impacts of the natural hazards compared to their occurrences. When defining compound extreme weather events, understanding complex interactions between various physical processes across multiple spatial and temporal scales is critical. When multiple drivers or hazards combine, their impacts are often amplified, owing to multiple hazards co-occurring.

Zscheischler et al. (2020) summarized the typology of compound events and suggested complementary analytical and modeling approaches from different types of compound events. They characterized compound event types into four major groups: preconditioned, multivariate, temporally compounding, and spatially compounding. Further, the literature provides better insight into their respective mechanisms and impacts by structuring compound events and their respective analysis tools. For example, extreme windspeed and precipitation are multivariate, temporally, and spatially compounding, severely impacting infrastructure and human health. Simultaneous intense supercell thunderstorms across multiple locations generate tornadoes with prolonged associated heavy precipitation, which could damage properties and put exposed communities at risk. During TORFF events, emergency response communication could be crucial when the instruction to safety is conflicting (Nielsen et al. 2015).

Compound extreme weather events are not only associated with property/infrastructure damage and human safety (Najafi, Zhang, and Martyn 2021), but also affect emergency response resources. For example, severe tropical cyclones in the eastern Caribbean caused damage to property and infrastructure due to consecutive storms in 2017. U.S. Federal Emergency Management Agency could not provide an adequate emergency response due to depleted resources by Hurricane Harvey a few weeks back (Murakami et al. 2018; Klotzbach et al. 2018). In time, the inability to provide adequate support to the affected community in the Caribbean had caused them to expose to more risks such as lack of necessities (water, shelter, etc.) and hunger. The unusual cold and wet early spring affected winter-cereal harvests and spring planting across Europe, followed by a hot and dry summer, leading to severe agricultural losses in consecutive cropping seasons. Consequently, the increase in crop prices across the EU affected the continent's government and insurance budget. (European Commission 2018, Faust and Strobl 2018)

Understanding interconnected extremes and collaboration among relevant experts help inform policies toward mitigation strategies to reduce risks and increase resilience to concurrent extremes, which is often overlooked by single extremes. (Raymond et al. 2020) provided examples of hazards with their respective climatic and societal drivers. For instance, coastal flooding could be related to river flow, precipitation, coastal water level, surge, and wind speed, followed by the corresponding impacts on infrastructure and natural coastal barrier (Lentz et al. 2016; Temmerman et al. 2013; Ying Zhang and Najafi 2020; Jalili Pirani and Najafi 2020). Furthermore, the authors also recommended methods to investigate related extreme events and their strengths and weaknesses.

The authors described the statistical approach such as copula as well-developed and straightforward to apply (Sadegh et al. 2017; G. Salvadori and De Michele 2010), specifically for multivariate analyses. However, limited data might affect the data fitting and not identify causal relationships. The weaknesses will be further discussed in the later section of the study.

2.1.4 TORFF Generating Supercell Thunderstorms

One of the most typical extreme weather events that happen almost daily across the globe is a thunderstorm. In North America, thunderstorms usually occur more frequently during the warm weather season. According to National Oceanic and Atmospheric (NOAA) National Weather Service, approximately 1,800 thunderstorms are happening at all times, resulting in about 16-million thunderstorms each year. Lightning and flash floods are considered the two biggest threats associated with most thunderstorms (NOAA 2009).

Four types of thunderstorms could be generated in the cycle of storms: single-cell, multi-cell, squall line, and supercell. Out of the four types of thunderstorms, supercell thunderstorms are the most severe, long-lived, and highly organized. A supercell thunderstorm can grow up to 16 km in diameter and 15 km tall and usually be visible before a tornado.

Supercell thunderstorms can be further characterized into three groups: low precipitation, classic, and high precipitation. High precipitation supercell thunderstorms (Moller et al. 1994) usually generate extremely heavy precipitation (Hitchens and Brooks 2013; Smith et al. 2001) that could cause flash flooding. Besides generating high precipitation, tornadoes that come from it are the most common and often the most dangerous.

When a supercell thunderstorm generates both tornadoes and extreme precipitation at the same time and location, it can be considered a compound extreme weather event. This is because

tornadoes (driver 1) coupled with extreme precipitation (driver 2) could amplify the impacts on the environment.

2.2 Dependency Analysis

2.2.1 Understanding Correlation

There is a preliminary step before the detailed dependency analysis, which investigates the relationship between two variables by conducting a correlation test. However, even though two variables are determined to correlate, it does not necessarily mean their relationship falls under cause and effect. For example, the crime rate and consumption of ice cream are high during summer; the effect of seasons would have more effect on these two variables than the effect of ice cream consumption on the crime rate in an area. Therefore, it is essential to be well equipped with fundamental knowledge about the interaction between variables of interest.

Many studies have been carried out in atmospheric science to study the interaction between variables and their correlations. Understanding the mechanisms and processes of physical atmospheric regimes will clarify the correlation between the variables involved and provide a clearer picture. Multiple researchers have studied the correlation between wind speed and precipitation for different regions using different datasets. For example, in Back and Bretherton (2005), the relationship between wind speed and precipitation was investigated over a large water surface (Pacific Ocean), and the results showed a positive correlation. However, the results also demonstrated the importance of understanding the mechanism of the processes since the relationship is not relatively straightforward for atmospheric processes. Multiple studies have demonstrated that even though wind speed and precipitation are positively correlated globally, they could differ topographically, regionally, and seasonally. Severe natural hazards such as tornadoes and cyclones could affect the correlation between wind speed and precipitation

depending on the study area (Owen et al. 2021; Zhang et al. 2021; Raible et al. 2007; Back and Bretherton 2005).

2.2.2 Multivariate Probability Distributions

The application of multivariate distributions in hydrology is widely documented by multiple studies (Jalili Pirani and Najafi 2020; G. Salvadori, Tomasicchio, and D'Alessandro 2014; Gräler et al. 2013; G. Salvadori and De Michele 2010; Renard and Lang 2007; Favre et al. 2004). One of the advantages is that multivariate distributions identify and show a comparison between measurements and the relationships among them. In addition, there is a more general multivariate distribution for each univariate distribution with a single random variable. For instance, the univariate normal distribution's general counterpart is usually the multivariate normal distribution (Statistics How To 2021). Same goes to other types of distributions such as exponential, log normal and gamma. Multivariate models have been used in climate extreme analysis such as flood risks (discharge volume and discharge peak), drought risks (drought duration and severity), hurricanes (total destruction and maximum landfalling windspeed), infrastructure design (flood volume and peak discharge), and wind-resistant design of engineering structures (windspeed and direction) (Brunner, Seibert, and Favre 2016; Azam et al. 2018; Naz et al. 2019; Leahy 2021; Li, Zheng, and Li 2019).

Hydrological processes typically have heavy tails that need to be investigated to understand their relationships (such as extreme events). Fitting generally used multivariate distributions such as normal distributions is not adequate for capturing these dependencies and model extreme joint hydrologic events (Favre et al. 2004). However, one can implement multivariate gamma distribution, Pareto, or exponential distributions to model hydrological extremes. The downside is that there is a limited option for fitting marginal distributions; the ideal case would be that

marginals are similar to the univariate distribution. In reality, that is not always the case; marginals can appear in different forms for the interested variables. For example, (Brunner, Seibert, and Favre 2016) shows a General Pareto Distribution (GPD) for peak discharges and GEV for the flood volumes.

2.2.3 Marginal Probability Distributions

Selecting the suitable marginal distributions depends on the method used to extract extreme data (also known as sampling). According to Brunner et al. (2016), there are two sampling approaches from a time series: block maxima, and peak-over-threshold (POT). Block maxima samples the highest event over a period of time and POT chooses all values that lie above a predefined threshold. Typically, in the block maxima approach a year is considered as a "block" of time frame, which means only one maximum value can be chosen each year. POT allows for choosing more than one event based on threshold of choice. Sometimes in extreme cases, fewer events are chosen compared to block maxima even though block maxima might miss some other important events.

Datasets generated through block maxima will generally follow a generalized extreme value (GEV) distribution. (Coles 2001) However, according to Naz et al. (2019), multiple continuous marginal distributions were fitted to their variable datasets to identify the most appropriate distribution. In addition to general marginal distributions such as exponential, log-normal, gamma, and normal (Teng and Liu 2001), multiple studies related to marginal distribution fitting on wind speed and precipitation were also studied to determine the best fit. For example, Weibull distribution (Li et al. 2019; Carta et al. 2008) and Gumbel (Zhang 2013) were considered to fit the wind speed data in their studies well.

Marginal distributions are important before proceeding with copula analysis. Once they are defined, the strength of their dependency and their relationship need to be investigated. (Serinaldi 2016).

As discussed before, many bivariate distributions assume similar forms of marginal distributions for both variables and must be characterized by the similar parametric family of univariate distributions. Copula models which are multivariate distribution functions help to overcome the limitation. They are widely used to construct multivariate cumulative distribution functions and conduct multivariate frequency analysis (Singh, Najafi, and Cannon 2021; Singh, Pirani, and Najafi 2020; Singh and Reza Najafi 2020; Gräler et al. 2013; Gianfausto Salvadori 2007).

2.2.4 Copula Analysis

Sklar first introduced copula in 1959 to characterize the dependence structure between two or more variables. Copula is a multivariate statistical analysis method that can overcome the limitations of correlation coefficients and multivariate extensions of univariate parametric distributions in multiple ways. Copula identifies the structure of the multivariate dataset that includes correlation and tail dependence of a dataset. In addition, it is influential in determining dependencies between extreme events. Normal distribution usually fails to capture such rare events and their dependencies.

A major theoretical result that allows for the analysis of dependency independent from marginal distributions is the Sklar's theorem, which states, in plain words, that, for any number of distribution functions, normal or not normal, there is a unique copula function. However, several copula functions can be used depending on the nature of the distribution functions.

The copula is widely used in extreme weather analysis. For example, Leahy (2021) investigated the dependency between the total area of destruction and landfalling windspeed Ingrosso et al. (2020). studied the dependencies of variables presented in the tornado environment.

Chapter 3 Study Area and Data

3.1 Study Area

3.1.1 Overview

Canada is the second largest country globally with an area of 9.984 million square kilometers (Mattyasovszky 2020). Canada is surrounded by the Pacific and Atlantic Oceans on the west and east, respectively, and its mainland is connected to United States on the south as its neighboring country.

The study region covers the entirety of Canada from the longitude of $-141°$ to $-53°$ and latitude of $41°$ to $84°$. Canada has a population of 36 million people as of 2016, which accounts for 0.5% of the total population in the world. According to Sawe (2017), the most populated provinces in Canada are Ontario, Quebec, British Columbia, and Alberta, contributing to 86% of Canada's population. The population is mainly concentrated along the border to the United States.

Canada is positioned on a unique location on the globe geographically, where it hosts a wide range of weather patterns from Artic to moderate, from seemingly endless rains to drought, from numbing cold to heatwaves. With those weather types, severe weather and geological events are a constant possibility.

When a natural hazard affects the communities to the extent that they need assistance dealing with the damage that has occurred to people and the surrounding property and environment, the event is considered a disaster. For example, on June 22, 2007, Canada's first F5 tornado just developed in Elie, Manitoba, with winds exceeding 420 km/h. It destroyed utility poles and carried a house a few hundred meters through the air Mccarthy et al. (2008). In addition, a tornado that developed in Pine Lake, Alberta, on July 14, 2000, resulted in 12 deaths, more than

130 injuries and is considered the deadliest tornado in North America in 2000 (Dupilka and Reuter 2011). Dated back to 1985, fourteen tornadoes hit multiple communities on May 31, causing 12 deaths, injuring hundreds more.

Further, those tornadoes also destroyed or damaged more than 1,000 homes. (National Post 2021) As recent as 2021, tornadoes that hit Barrie, Ontario caused up to CAD 100 million in insured damage on July 15 (Insurance Bureau of Canada, 2021).

The Fujita rating[1] (also known as Fujita scale or Fujita-Pearson scale) is a metric to categorize tornado intensity, based primarily on the damage tornadoes inflict on manmade structures and vegetation. The official Fujita scale category is investigated by experts, typically through ground and/or aerial damage surveys. However, in 2007 and 2013, the enhanced Fujita scale (EF-scale) was adopted in USA, and Canada respectively (Sills 2014). The main reason is that the regular Fujita rating windspeeds were unrealistically high for the upper-end tornadoes, specifically F4 and F5 tornadoes. For example, an EF5 tornado windspeeds of 315 km/hr is more likely compared to F5 tornado windspeeds of 420 km/hr.

Table 1: Fujita Rating Scale & Damage Severity (Source: Sills, 2012 and NWS Alabama Website)

Fujita Rating (F)	**F-Scale Windspeed Rounded to 10 km/h**	**Damage Severity**
0	60-110	• Light damage with some damage to chimneys. • Break branches off trees. • Push over shallow-rooted trees and damage signboards.
1	120-170	• Moderate damage where the lower limit is the beginning of hurricane wind speed.

[1] An Enhanced Fujita scale was adopted in Canada in 2013 that had improved relationships between the observed damage and maximum windspeeds (Sills 2014).

		• Peel surface off roofs and mobile homes pushed off foundations or overturned. • Moving vehicles pushed off the roads.
2	180-240	• Considerable damage where roofs were torn off frame houses. • Mobile homes demolished and boxcars toppled. • Large trees snapped or uprooted, and light-object projectiles were generated.
3	250-320	• Severe damage where Roofs and some walls tore off well-constructed houses. • Trains overturned, and most trees in the forest uprooted. • Heavy cars were lifted off the ground and thrown.
4	330-410	• Devastating damage that caused well-constructed houses leveled. • Sstructures with weak foundations blown off some distance. • Cars were thrown, and large missiles were generated.
5	420-510	• Incredible damage where substantial frame houses lifted off foundations and carried considerable distance to disintegrate. • Automobile sized projectiles fly over 100 meters (109 yds). • Trees debarked.

3.1.2 Historic Tornado Climatology in Canada

Note: All data used to generate figures and tables related to Canada's historic tornado events in this section (3.1.2) is obtained from Environment Canada Tornado Database (Environment Canada 2017).

By creating a heatmap for tornado occurrences in Canada (Figure 3-1), tornadoes mostly developed in Southern Ontario, Alberta, Saskatchewan, and Manitoba.

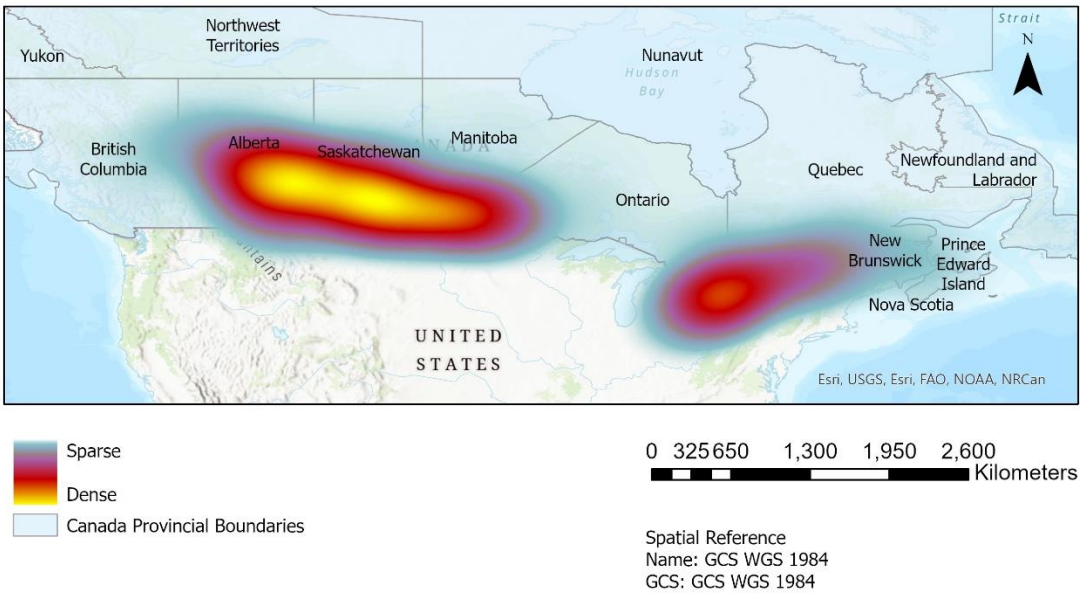


Figure 3-1: Tornado Occurrence Density in Canada from 1980-2009 (Heat density map created using all verified tornado events. Data Source: Environment Canada)

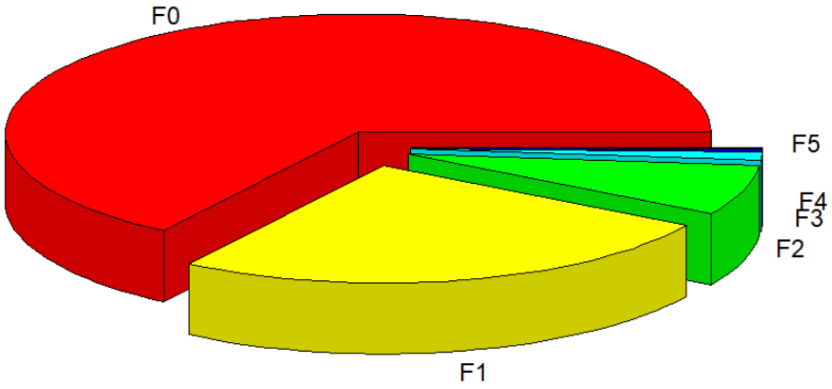


Figure 3-2:Tornado Intensity Composition in Canada form 1980-2009

From 1980 till 2009, a total of 1839 tornado events have occurred, with most of them being F0 and F1 tornadoes (Figure 2), with an average of ~61 tornado events happening every year. According to National Weather Service (NWS), F0 and F1 tornadoes are considered weak. F0 tornadoes will cause light damage to homes such as damaged chimneys, damaged signboards, and push over shallow-rooted trees. On the other hand, F1 tornadoes will begin to tear off roofs,

demolish mobile homes, and cause large trees snap or uproot (NWS Birmingham, Alabama 1981).

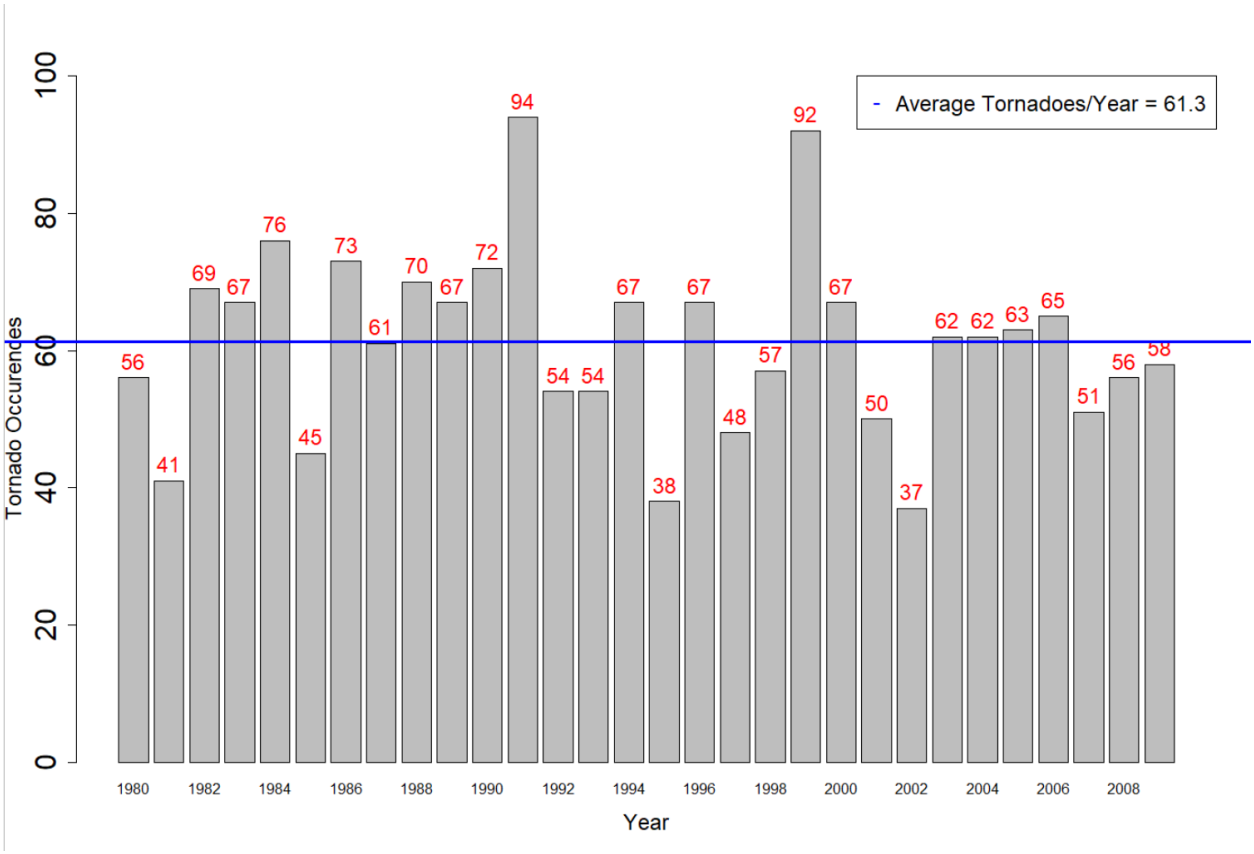


Figure 3-3: Annual Tornado Occurrences in Canada from 1980-2009

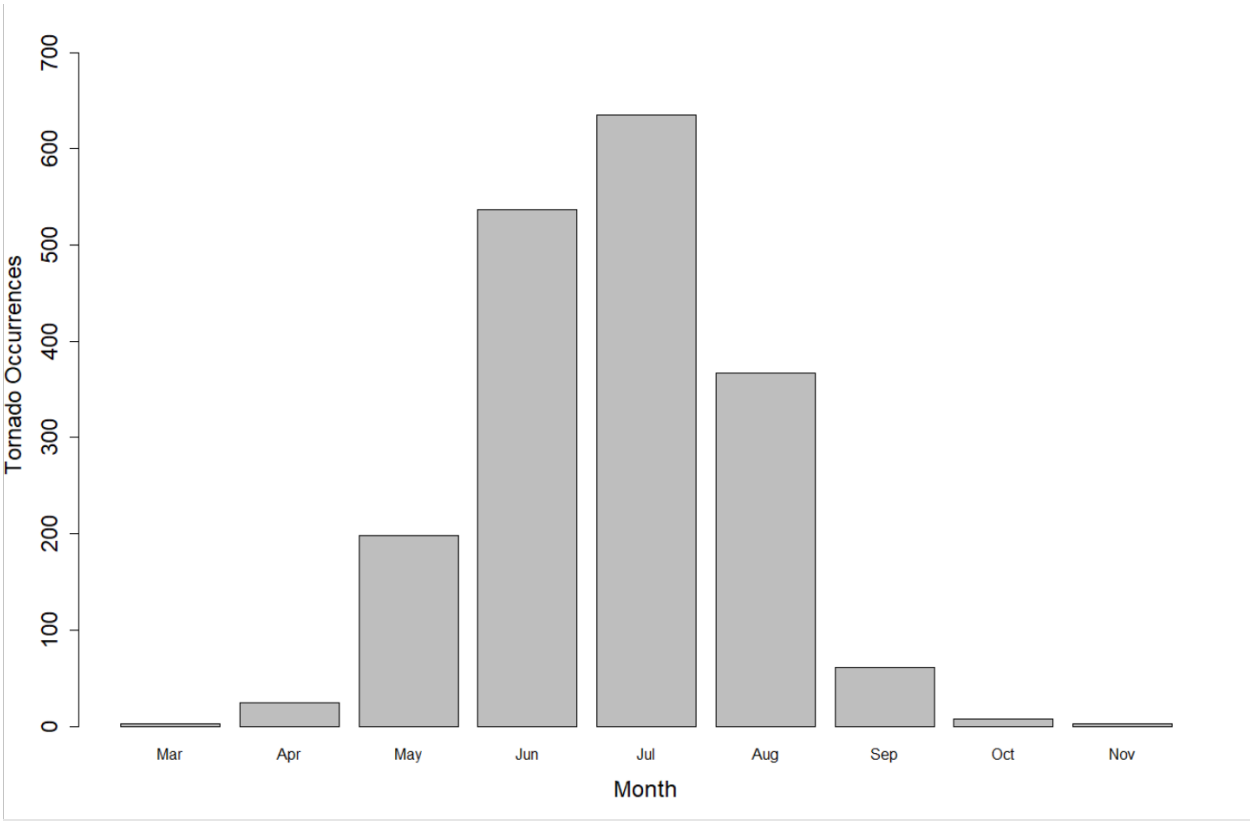


Figure 3-4: Seasonal Tornado Occurrences in Canada from 1980-2009

The top severe tornado events are F4-F5, with only one F5 tornado event recorded in the Canadian Tornado Database (CTD). According to Figure 4, most tornadoes started getting active in May, peaked in July with 635 events, and started to slow down in September.

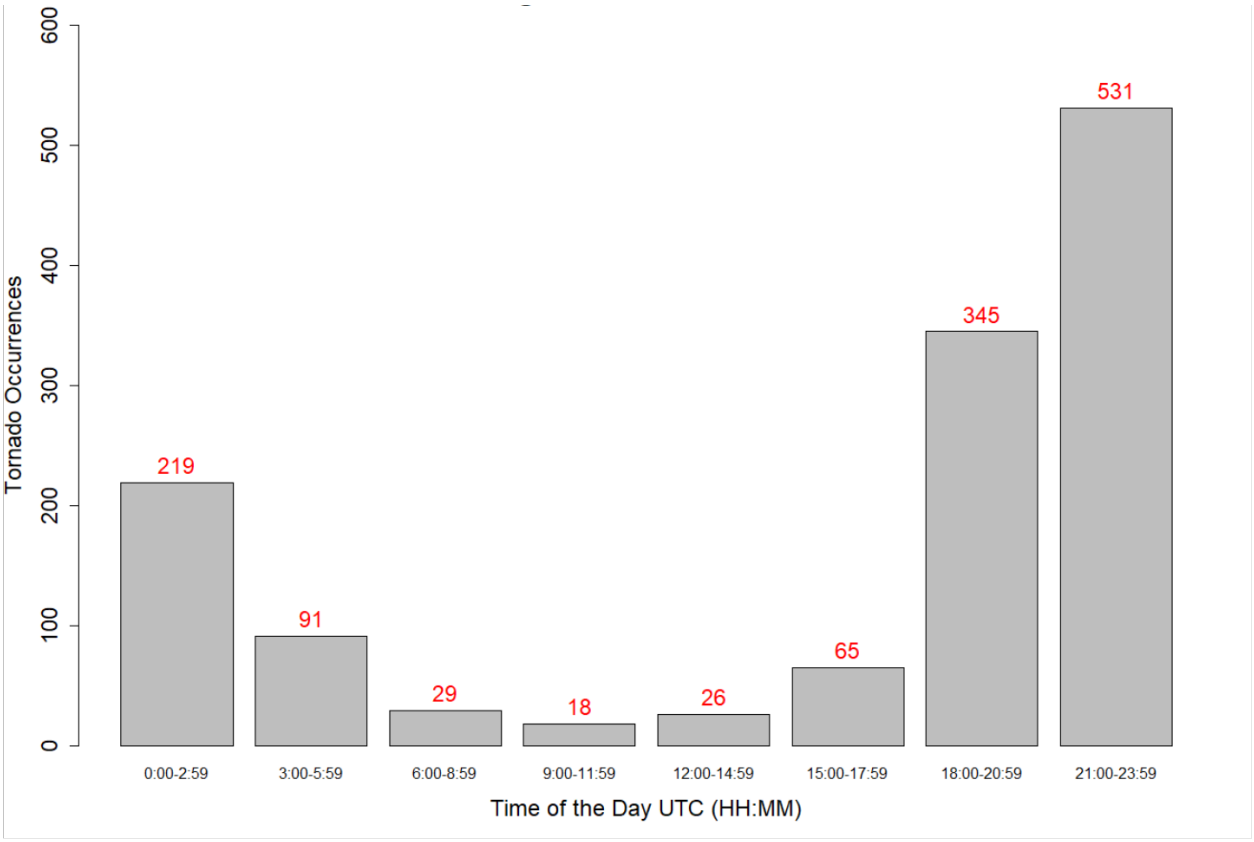


Figure 3-5: Tornado Development Time in Canada from 1980-2009

The tornado development times started to increase in the early afternoon, between 3 and 6 pm, and started to peak between 9 and 11:59 pm UTC (Environment Canada 2017). Please note that all tornado development times are converted to Coordinated Universal Time (UTC) as shown in the CTD. However, it is also important to describe the tornado development times in the context of local times due to the variability of weather system diurnally and nocturnally (e.g., temperature difference). In terms of local time in Canada (specific to Eastern Standard Time (EST)), the development time increased between 2 pm and 5pm, and peaked between 5 and 8:00pm EST. This results agrees with updated climatology as provided by Nielsen et al. (2015). In this literature, the authors concluded that nocturnal tornadoes are responsible for a larger percentage of total events.

Example Historic TORFF Event Specifically in Canada

- Southern Ontario Tornado Outbreak of 2005

On August 19, 2005, severe thunderstorms generated tornadoes that caused damage to homes in Conestoga Lake, Furgus, and Tavistock areas. In addition, the storm cell north of Fergus had spawned two F2-strength tornadoes that caused property damage. The first tornado was tracked through Milverton to Conestogo Lake, and the second tornado moved from Salem to Lake Bellwood, located on the north side of Guelph.

Additionally, a tornado was suspected of developing in Toronto due to the same storm cell but never verified by the Meteorological Service of Canada. The thunderstorm was then transformed into a heavy rain cell that generated 140 mm of rain in part of the city. The extreme rain event caused extensive flooding that washed out roads and damaged essential infrastructures such as storm sewers and electrical systems. According to the Insurance Bureau of Canada, the tornado outbreaks had caused damage of over $CAD 500 million.

The same supercell that developed in Toronto maintained all the characteristics of a tornado producer as it approached the city. The weather radars showed a convective storm and a strong mesocyclone, but the vortex left the ground after the second tornado. Studies were conducted and tried to explain the change in behavior to an extremely high producer of rain. For example, the north side of Toronto had recorded 175 mm of rain in less than one hour.

3.2 Data

Hydrological modeling or developing a statistical model in remote regions is typically challenging due to the incomplete weather station data. Interpolation is often used when the ground observations are sparse and located far away from points of interest. Further, each ground observation station such as rain gauge might have missing data from time to time, which leads to

another problem for continuous modeling. Hence, studies have shown that implementing reanalysis products such as NARR data are a viable approach to complement the weather station data. According to Choi et al. (2009), reanalysis data are considered as a product of an advanced interpolation scheme that takes into account important factors such as topography and land cover with high spatial resolution, which could be a potential advantage of reanalysis data over station data. Additionally, representative areal values are more desirable when a study area encompasses a large domain that includes remote regions (e.g., whole Canada) that might be potentially biased due to its difference from the actual topography condition of the area (Hunter 2016).

The NARR reanalysis product was used in this study. Ground observations such as tornado's Fujita ratings and tornado development times and sites were also implemented for resampling in the later section of the analysis. As mentioned before, the Fujita rating has its corresponding intensity rating with windspeed interval. Precipitation and other affecting environment factors were extracted from the nearest grid cell from the coordinates of the tornado development site. Data from the same source are encouraged to avoid inconsistency during analysis.

Reanalysis products are commonly used to investigate various questions related to past weather events and play a key role in determining how severe weather events may or may not change in a warming climate (King and Kennedy 2019).

3.2.1 Tornado Environment Data

- **3.2.1.1 NARR Reanalysis Product**

Mesinger et al. (2005) describe NARR as a "long-term, dynamically consistent, high-resolution, high-frequency, atmospheric and land surface hydrology dataset for the North American domain."

NARR is chosen due to its ability to distinguish the environment that produces supercells. NARR contains significant environmental factors in determining whether a supercell thunderstorm might generate intense precipitation (King and Kennedy 2019; Gagne et al. 2012).

Brooks et al. (1994) mentioned variables such as lifted condensation level (LCL), mean shear; and a composite of mixing ratio, storm-relative helicity (SRH), and minimum midlevel wind speed were the best discriminators to distinguish between strongly tornadic and other supercells. NARR provides closer representations of the environment for each storm both spatially and temporally.

The dataset is available at a resolution of approximately 32km at the lowest latitude from 1979 to present at 3-hourly intervals. The product is a regional extension of NCEP Global Reanalysis created using the NCEP Eta Model and the Regional Data Assimilation System (RDAS). In this study, the dataset from 1980-2009 was used following the tornado information available with CTD.

3.2.2 Windspeed Data

- **3.2.2.1 Resampled Windspeed Based on Ground Observations**

The windspeeds are generated based on the Fujita Rating identified by the Environment Canada for each tornado event (Environment Canada 2017). 1000 windspeed values were randomly resampled based on the intensity of the tornado event, since Fujita Rating is determined based on the scale of post-event damage, not the actual tornado windspeed. Furthermore, the wind vectors from NARR are too coarse (32km x 32km per grid cell) to be considered as representative for this analysis. Estimated windspeeds were resampled uniformly from its corresponding Fujita rating windspeed interval (Table 1). However, the Fujita windspeeds are also known to be unrealistically high for violent tornadoes, as corrected by the Enhanced Fujita (EF) scale that was adopted by Canada in 2013 (Sills 2014). The reason behind resampling in such a way is to represent surface tornado windspeed and capture any uncertainties. The *u* and *v* wind vector data from the reanalysis products are not considered in this analysis due to its coarse resolution (32km x 32km) to capture tornado windspeeds (NOAA 2003).

3.2.3 Precipitation Data

Precipitation data was extracted from NARR due to the sparsity of ground observations (rain gauges) in Canada corresponding to the development location of historic tornado events. Post tornado precipitation value was extracted for each tornado event. Precipitation data of Canada and Mexico for assimilation come from 1° rain gauge analysis (Bukovsky et al. 2007). NARR shows low negative bias in the summer and outperforms several other reanalysis products in terms of daily error and goodness-of-fit (Choi et al. 2009). The finding from Choi et al. (2009) is in general agreement with Mesinger et al. (2006) where NARR is a viable alternative ground-based observation in the area of interest is sparse.

Flash flood events in this analysis are represented by rainfall as a proxy since there is limited flash floods database in Canada. However, we know that the occurrence of flash floods is not limited to precipitation alone, either resulted from intense precipitation or long duration storm event. Flood-producing mechanism such as antecedent precipitation and moisture content, snowmelt play an important role in resulting flash floods (Holman 2018). Keep in mind that this analysis solely focuses on just precipitation values and positive precipitation values represent possible flood events. Depending on the conditions of the tornado development sites, even low precipitation might result in flash floods occurring due to factors mentioned in Holman (2018).

Chapter 4 Methodology

As discussed in the previous section, the precipitation dataset for this study was extracted from the reanalysis product, NARR. Even though NARR consists of *u* and *v* vectors of the windspeed that could be used to calculate wind speed for each tornado event, values were determined to be too coarse (32 km x 32 km per grid cell) and were deemed to be unrepresentative for tornado windspeed. Hence, a resampling method is implemented in this study. Ground observations of Fujita Ratings (Table 1) recorded by Environment Canada for each tornado event were used as intervals (e.g., F0 tornado's windspeed ranges from 60-110 km/h) in this analysis due as no actual tornado surface wind speed was recorded. Each tornado event was resampled uniformly 1000 times based on its Fujita rating windspeed interval to generate 1000 unique data records to represent the corresponding wind speed. The associated precipitation dataset was then collected a timestamp after each tornado event and paired with resampled windspeed to create 1000 unique pairs of observations per tornado for this study. Since this study focuses on TORFF events, we are looking at precipitation generated from the tornadoes that might eventually lead to flash flood. The precipitation units are given in NARR (kg/m^2) (NOAA 2003) are converted to mm considering the density of water. For example, 40 kg/m^2 of 3-hourly total precipitation corresponds to 40 mm of precipitation in 3 hours for NARR.

4.1 Tornado Event Clustering

The 1839 verified extracted tornado events (Environment Canada 2017) are clustered before extreme analysis based on the Generalized Extreme Value (GEV) distribution. This allows for investigating the TORFF events and the corresponding dependencies between the drivers at different regions across Canada. This will lead to the identification of high TORFF risk areas. In this study, we implement the affinity propagation clustering, similar to Nielsen et al. (2015) with a minor modification. The authors used the geographical coordinates and tornado development time to generate multiple clusters in their analysis. However, they looked at tornadoes generated on a specific date instead of historical data. Therefore, we did not consider tornado development time and clustered the events based on the coordinates.

4.1.1 Affinity Propagation Clustering

The AP analysis was conducted to characterize the tornado events based on longitude and latitude. AP clustering functions by considering all points as cluster centers with each data point viewed as a node in a network that then communicate recursively across the network to minimize an energy function to arrive at data points that are representative of themselves. In this study, Euclidean distance were used as the energy function.

One advantage of AP is that it does not require the data samples to have specific structure and only requires a similarity matrix. Additionally, AP does not require these coordinates to be in specific ranges and can be positive or negative. Furthermore, AP does not even require the similarity matrix to be symmetrical.

After the datapoints are being plotted, the center of each cluster is marked by a small square and all cluster members are connected to their centers with lines.

The performance of regular AP will be measured based on three criteria:

1. Sum of exemplar preferences (input preference such as K-means)
2. Sum of similarities of centers to their cluster members
3. Net fitness: sum of two former

Many clustering algorithms need to know a pre-determined number of clusters, which is often a major challenge, since the exact number of clusters is hard to pre-determine for non-trivial data sets. To overcome this challenge, a search algorithm adjusts input preferences to produce the desired number of clusters in the end. The algorithm reduces the dissimilarity between datasets through an iterative process.

In this study, two conditions were set in creating the usable dataset.

1. At least 30 years of continuous data available for analysis to fulfill the condition of block maxima in GEV.
2. Number of clusters that generates its own corresponding dataset with highest within cluster homogeneity and highest between cluster dissimilarity.

   Note: The purpose of determining highest dissimilarity between clusters is to show the distinction in covariability of dependency across Canada.

4.2 Extreme Values Analysis

Extreme value theory (EVT) is one of the branches of statistics that specifically deals with extreme deviations from the median of probability distributions. Its main purpose is to seek "to assess, from a given ordered sample of a given random variable, the probability of events that are more extreme than any previously observed" (Gabor Melli's Research Knowledge Base 2021). Extreme value analysis is commonly implemented across multiple disciplines including structural engineering, finance, earth sciences, and geological engineering. For example, EVT was used in the field of hydrology to estimate the probability of extreme flooding, such as the 100-year flood (De Paola et al. 2018).

GEV distribution originates from Extreme Value Theory (EVT) and provides a statistical framework to make inferences on extreme/rare events. The GEV distribution combines three types of distributions (Weibull, Gumbel, and Fréchet) into a single family to allow a continuous range of different possible shapes (also known as type I (Gumbel), II (Fréchet), and III (Weibull) extreme value distributions). The type of extreme value distributions (type I, II, and III) are determined when a shape parameter is equal to 0 (exponential tail), greater than 0 (fat tail), and less than 0 (upper finite endpoint), respectively. Additionally, the GEV distribution consists of three main parameters: shape, location, and scale. The cumulative distribution function (CDF) of the GEV distribution is shown as follow:

$$F(x, \mu, \sigma, \xi) = e^{\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}\right)}\right.} \quad (1)$$

where, μ, σ, ξ represent location, scale, and shape of the distribution function, respectively. The scale, σ and the term, $1 + \xi\left(\frac{x-\mu}{\sigma}\right)$ must be greater than zero.

In this study, annual maximum series (AMS) was generated from each cluster using the GEV distribution with the highest resampled windspeed per year according to CTD. The distributions of windspeed datapoints are then represented by GEV for univariate and bivariate analyses. The 1000 resampled windspeed dataset generated from each tornado event is then paired with the associated precipitation dataset for copula analysis.

4.3 Probabilistic Modelling

4.3.1 Copula Theory

Copula was first introduced by Sklar (1959) for multivariate probability modelling. Copula is a statistical method that allows the researcher to identify the interdependence structure between continuous random variables of interest (X, Y). Copula theorem is a function that links any pairs of continuous/discrete random variables, $X\ and\ Y$, into a joint Cumulative Density Function (CDF) $F_{xy}(x, y)$ independent from their marginal distributions. The equation can be simplified as:

$$F_{XY}(x, y) = C[F_X(x), F_Y(y)] \quad (2)$$

$F_X(x)$ and $F_Y(y)$ are the marginal distributions of variables *X* and *Y*, and *C* is the "copula function built over uniform marginals, which are the quantile transformations of *X* and *Y*."

Copula functions are multivariate distributions with uniform marginal functions, which can be model interdependent if there is more than one variable. Marginal distributions typically work well if two marginal distributions in multivariate analysis are identical. However, if there is no dependency, these two distributions can be considered as two individual marginal distributions that are independent of each other. One main advantage of using copula functions is that it allows the researchers to determine the interdependence structure between the variables even

though the marginal functions are different. Through copula theorem, appropriate marginal distribution functions can be selected for setting multivariable functions as well as defining nonlinear and asymmetric relationship between variables (Mesbahzadeh et al. 2019).

There are many copula families and each one has their unique properties that allow the models to identify the appropriate bivariate structure. In other words, there will be a suitable and unique copula representation as long as the marginal distributions are continuous. (Favre et al. 2004) In hydrology, different families of copula have been implemented. (Favre et al. 2004; G. Salvadori and De Michele 2004) In this study, we apply all families of copulas to each of the 1000 simulated datasets, and the best one is selected using the AIC criterion similar to (Singh et al. 2020).

Table 2 below includes the copula functions included in the model selection process.

Table 2:Copula Families Included in the Model Selection Process.

Copulas	Bivariate Copula, C_θ (u,v)	Parametes, θ
Archimedean Copulas		
Clayton	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1}{\theta}}$	$\theta \in [-1,\infty) \backslash\{0\}$
Frank	$-\frac{1}{\theta} log\left[1+\frac{\left(e^{-\theta_u}-1\right)\left(e^{-\theta_v}-1\right)}{e^{-\theta}-1}\right]$	$\theta \in [-\infty,\infty) \backslash\{0\}$
Gumbel	$exp\left[-\left(u^{-\theta}+v^{-\theta}\right)^{\frac{1}{\theta}}\right]$	$\theta \in [1,\infty)$
Joe	$1-[(1-u)^{\theta}+(1-v)^{\theta} -(1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$	$\theta \in [1,\infty)$
Elliptical Copulas		
Student's t	$\int_{-\infty}^{t_\vartheta^{-1}(u)}\int_{-\infty}^{t_\vartheta^{-1}(v)}\frac{1}{2\pi\sqrt{(1-r^2)}}\left[1+\frac{x^2-2rxy+y^2}{\vartheta(1-r^2)}\right]^{\frac{\vartheta+2}{2}}dxdy$ $t_\vartheta(x)=\int_{\infty}^{x}\frac{\Gamma\left((\vartheta+1)/2\right)}{\sqrt{(\pi\vartheta)}\Gamma\left(\frac{\vartheta}{2}\right)}(1+y^2/\vartheta)^{-(\vartheta+1)/2}dy,$ $\vartheta \neq 0$	$\vartheta > 2, r \in (0,1]$
Gaussian	$\Phi_2(\Phi^{-1}(u),\Phi^{-1}(v),\rho)$	$-1 \leq \rho \leq 1$

Copulas	Bivariate Copula, C_θ (u,v)	Parametes, θ
Archimedean Copulas		
Clayton	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1}{\theta}}$	$\theta \in [-1,\infty)\backslash\{0\}$
Frank	$-\frac{1}{\theta} log\left[1+\frac{\left(e^{-\theta_u}-1\right)\left(e^{-\theta_v}-1\right)}{e^{-\theta}-1}\right]$	$\theta \in [-\infty,\infty)\backslash\{0\}$
Gumbel	$exp\left[-\left(u^{-\theta}+v^{-\theta}\right)^{\frac{1}{\theta}}\right]$	$\theta \in [1,\infty)$
Joe	$1-[(1-u)^{\theta}+(1-v)^{\theta}-(1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$	$\theta \in [1,\infty)$
Elliptical Copulas		
Student's t	$\int_{-\infty}^{t_\vartheta^{-1}(u)}\int_{-\infty}^{t_\vartheta^{-1}(v)}\frac{1}{2\pi\sqrt{(1-r^2)}}\left[1+\frac{x^2-2rxy+y^2}{\vartheta(1-r^2)}\right]^{\frac{\vartheta+2}{2}}dxdy$ $t_\vartheta(x)=\int_{\infty}^{x}\frac{\Gamma\left((\vartheta+1)/2\right)}{\sqrt{(\pi\vartheta)}\Gamma\left(\frac{\vartheta}{2}\right)}(1+y^2/\vartheta)^{-(\vartheta+1)/2}dy,$ $\vartheta \neq 0$	$\vartheta > 2, r \in (0,1]$
Gaussian	$\Phi_2(\Phi^{-1}(u),\Phi^{-1}(v),\rho)$	$-1 \leq \rho \leq 1$

4.3.2 Investigation of the Dependencies

Since copula provides the joint distribution of correlated variables, it is important to analyze the dependence between the variables. Visualization of the dependence structure can be conducted through generating a scatterplot of the data such as Chi-plots and K-plots.

Chi-plot is generated by using a rank-based measure of the location of each of the observations. It explains the dependency (positive or negative) by observing the transformed data scattered on the region, which is defined by the confidence interval of the Chi-plot. (Fisher and Switzer 1985)

A K-plot (Kendall's plot) is a plot "between the order statistics of the data and the values of these statistics expected in case of independence" (Genest et al. 2003). In the application of copula, K-plots can be implemented as the bivariate copula equivalent to QQ-plots. If the observation points lie near the 45-degree diagonal line, the variables are considered independent. In case of positive dependence, the observations should be located above the diagonal line, and below for negative dependence. The more dependent the two random variables will as the observations deviate further from the diagonal line. If the observations lie perfectly on the curve (known as $K_0(\omega)$) located above the diagonal, this indicates perfect positive dependence. However, if observation points $(W_{i:n}, H_i)$ lie on the x-axis, this indicates a perfect negative dependence.

Other than that, there are other dependence measures such as:

- Pearson's correlation coefficient, r
- Kendall's rank correlation coefficient (also known as Kendall's tau), τ, and
- Spearman's rho, ρ

All three dependence measures are computed to compare the correlations and associations for each cluster. These measures can be computed to quantify the dependence between variables.

The null hypothesis of no correlation can be rejected if *p*-values are deemed to be statistically significant that is, less than the significance level of 0.05. Null and alternative hypotheses can be set up as follow:

- Null hypothesis: TORFF driving mechanisms have no correlation (correlation coefficient equals to zero).
- Alternative hypothesis: TORFF variables have either positive/negative correlation (non-zero correlation coefficient).

Kendall's tau is considered as a robust dependence measure in the theory of copulas out of the three previously mentioned methods. (Naz et al. 2019). We note that the correlation metric is a crude measure of dependency (e.g., does not characterize the dependencies in high/low values) and that the rejection/acceptance of the null hypothesis is bound by the selected significance level. That said we deem the correlation metric appropriate for initial analysis of the dependencies between different drivers.

4.3.3 Copula Model Selection Criterion

The copula model is selected based on Akaike Information Criterion (AIC) that explains how good the model is in explaining relationships, in other words, it is a mathematical method for evaluating how well a model fits the data. It can be generated by comparing different possible models and determining which one is the best fit for the data. (Akaike 1974)

The equation for AIC can be defined as follows:

$$AIC = 2k - 2\ln(\mathcal{L}) \tag{3}$$

where, $k = number\ of\ estimated\ parameter(s)$, and $\mathcal{L} = maximum\ value\ of\ likelihood$

The preferred copula model is commonly the one with the minimum AIC value. The equation includes a penalty that prevents overfitting since a higher number of parameters, *k*, will always improve the model's goodness of fit.

Given the small sample size of the dataset (25-30 years of annual maxima in our case), the corrected version of the AIC (AIC_c) is considered in this analysis. (Burnham and Anderson 2004) The formula can be defined as follows:

$$AIC_c = -2\sum_{i=1}^{n} \ln\left[C\left(u_{i,1}, v_{i,2}\middle|\theta\right)\right] + 2k + \frac{2k^2 + 2k}{n - k - 1} \tag{4}$$

where n is the sample size. The –2 times likelihood function term is shown in the right-hand side of the equation above, followed by model complexity penalty, and a small-sample correction.

By implementing AIC_c, the copula with the lowest score would be selected since it shows the minimum loss in information between considered copula models. In addition, marginal distributions for windspeed and precipitation will also be determined through AIC_c. Since the windspeed dataset is resampled one thousand times, selected copula model composition will be evaluated by using the goodness of fit test. (Genest, Quessy, and Remillard 2006)

4.3.4 Marginal Distributions

According to Naz et al. (2019), copula analysis can be conducted by splitting into two different parts: marginal distribution and dependence structure.

In each cluster, TORFF variables (total accumulated precipitation (APCP) & resampled wind speed (WS)) are analyzed as two univariate distributions. Different probability density functions were applied on each variable individually that include: gamma, Weibull, exponential, logistic, normal, lognormal, Cauchy, and Generalized extreme value (GEV) distributions (Naz et al. 2019; G. Salvadori and De Michele 2004; Favre et al. 2004; Leahy 2021; Mesbahzadeh et al. 2019). This study utilizes precipitation data from reanalysis product, windspeed dataset was created by generating one thousand unique sets of windspeed data through resampling based on the recorded Fujita Scale for each tornado event. Probability density functions were fitted to each distinct dataset (*associated APCP* and resampled *WS*) to determine the most appropriate distribution function for modeling. However, with one thousand sets of resampled windspeed data, there is a possibility that there will be multiple determined marginal distributions for modeling.

4.5 Return Periods

4.5.1 Bivariate Return Periods

Previous studies have discussed the limitation in risk analysis if climate events are analyzed in isolation (Zscheischler et al. 2018; AghaKouchak et al. 2014). Investigating single variable might not fully capture the full underlying risk, as impacts are always amplified if there are dependencies between the driving mechanisms. Hence, in compound event analysis we study the interdependence of multiple drivers of extreme events (Raymond et al. 2020). Raymond et al. (2020) summarized the general climatic drivers for significant hazards such as drought, concurrent wind and precipitation extremes, coastal flooding, etc, and concluded that isolation of analyses for compound event can potentially underestimate the actual impacts and risks.

Therefore, a bivariate analysis is advisable when two variables of interest play a significant role in the behavior of a climate extreme event such as compound flooding (Salvadori et al. 2014). In bivariate frequency analysis, the definition of an event with a given return period is not unique; however, it is determined by the problem at hand. (Serinaldi 2015) The general definition of return period is not exclusive to univariate setting but also a multivariate setting and helps in transitioning the study emphasis from univariate to multivariate framework (Salvadori et al. 2011).

Furthermore, Salvadori et al. (2014) defined the return period of a "dangerous" event as

$$T_D = \frac{\mu}{Pr[X \epsilon D]} \tag{5}$$

where *D* is a set collecting all the values judged to be dangerous/extreme based on appropriate criterion, μ is the average interarrival time of two realizations of *X*, and $Pr[X \epsilon D]$ is the probability of a random variable *X* to lie in the dangerous region D. For example, in single

significant variable setting, a critical value x is used to identify the dangerous region D consisting of all values exceeding x. In a bivariate setting, the dangerous region D can be defined in multiple ways allowing for different return period definitions based on the problem at hand.

In (Brunner, Seibert, and Favre 2016), the authors mentioned that the return period used to assess bivariate events can be determined through three types of approaches.

- Conditional probability to determine a conditional return period
- Joint probability distributions to calculate join return periods
- Kendall's distribution or survival function

In extreme compound weather events such as TORFF, the conditional probability can be used to describe the probability of associated precipitation exceeding a given threshold (e.g., 50 or 100-year event) given the windspeed exceeds a specific threshold, or vice versa. The information can also be used to quantify return level of precipitation with set windspeed threshold for different return periods.

The joint probability distribution can describe the TORFF in two scenarios (AND/OR). For AND scenario, we look at the probability that both variables in TORFF, *APCP* and *WS*, exceed thresholds during a TORFF event. Secondly, is the probability that either the *APCP*, or *WS* exceed given thresholds (OR scenario).

4.5.2 Conditional Return Period

The conditional return period approach is commonly applied in situations in which one of the interested variables is considered to be more impactful than the others. (G. Salvadori, Tomasicchio, and D'Alessandro 2014) In this study, we quantify the return level of the precipitation, with given windspeed threshold for the different return periods. The purpose is to assess the relationship and the impact of the associated precipitation (dependent) with tornado events. Depending on the issues we are going to focus on, the conditional return period depends on a conditional probability distribution function of a variable provided with set condition. For example, in this analysis the probability of associated precipitation is conditioned on various resampled windspeed to assess the return levels.

The two events investigated in this research include tornadoes and flash floods which can be described as

$$Event_{X|Y} = \{X > x \mid Y > y\} \text{ and}$$

$$Event_{Y|X} = \{Y > y \mid X > x\},$$

with their probabilities defined as $Pr[X > x|Y > y]$ and $Pr[Y > y|X > x]$, respectively. By referring to our study as example, the return level of associated precipitation exceeding 100-year event, with probability P(X> *x*) = 0.99 given the 20-year resampled windspeed event, with probability P(Y>*y)* = 0.95 can provide us with the risk of extreme rainfall chances accompanied by relatively high wind events. It is also possible to estimate the return level with the higher windspeed exceedance probability, where P(Y>*y)* = 0.99 (100-year event). This event is considered to investigate the dependencies between the drivers of TORFF and compare bivariate, univariate, and independence hazard scenarios to improve future risk managements.

The conditional distribution functions for these events are given by:

$$Pr(X > x \mid Y > y) = 1 - \frac{F_X(x) - F_{XY}(x,y)}{1 - F_Y(y)}, \tag{6}$$

and

$$Pr(Y > y \mid X > x) = 1 - \frac{F_Y(y) - F_{XY}(x,y)}{1 - F_X(x)}. \tag{7}$$

Thus, the conditional return period of these two conditional events are:

$$T(x|y) = \frac{\mu}{1 - \frac{F_X(x) - F_{XY}(x,y)}{1 - F_Y(y)}} \tag{8}$$

and

$$T(y|x) = \frac{\mu}{1 - \frac{F_Y(y) - F_{XY}(x,y)}{1 - F_X(x)}} \tag{9}$$

where μ is defined as the mean time interval between two situations of exceedance of X given Y exceedance or vice versa. In our case, the μ is equal to 1 since extreme events are extracted based on are annual maxima.

4.5.3 Conditional Return Period Using Copulas

In this study, the conditional distribution is analyzed using copula (G. Salvadori 2004; G. Salvadori and De Michele 2004; Renard and Lang 2007; Gianfausto Salvadori 2007; Vandenberghe et al. 2011; G. Salvadori, De Michele, and Durante 2011; Gräler et al. 2013).

Considering Equations 8 and 9 and replacing F_X and F_Y with random variables *u* and *v* (with uniform distribution), the conditional return periods can be described as follows in terms of copula:

$$T(u|v) = \mu \frac{1 - v}{1 - u - v + C(u, v)} \tag{10}$$

And

$$T(v|u) = \mu \frac{1 - u}{1 - u - v + C(u, v)} \tag{11}$$

In our case, *u* and v could represent associated precipitation and conditional windspeed, respectively, while *C(u,v)* represents their joint probability.

4.5.4 Joint Return Period

The joint return period of one or more variables (multivariate) event can be computed through joint probability distribution functions. There are multiple ways to define the joint scenarios based on the joint probability distribution function that can be divided into four quadrants (Brunner, Seibert, and Favre 2016). (see Figure 4-6)

- Quadrant I: $Pr[X > x, Y > y] = 1 - F_X(x) - F_Y(y) + F_{XY}(x, y) = S_{XY}(x, y)$
- Quadrant II: $Pr[X \leq x, Y > y]$
- Quadrant III: $Pr[X \leq x, Y \leq y] = F_{XY}(x, y)$
- Quadrant IV: $Pr[X \geq x, Y \leq y]$

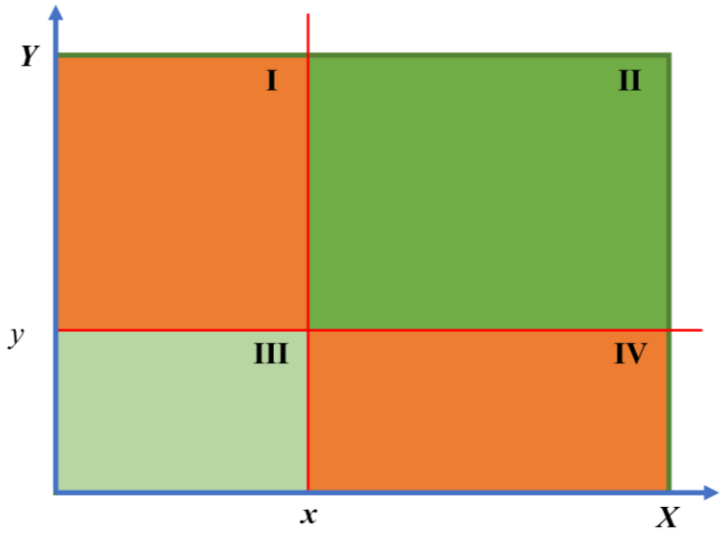


Figure 4-6: Joint Probability Quadrant (Source: Brunner 2013)

In compound extreme weather events, there is generally more interest in working with events situated in (1) Quadrant I, where *X* exceeds *x* **and** *Y* exceeds *y* or (2) Quadrant I, II and IV, where

X exceeds *x*, *Y* exceeds *y*, **or** both *X* exceeds *x* and *Y* exceeds *y*. (Yue and Rasmussen 2002) These are usually considered as "AND" and "OR" scenarios, respectively.

For TORFF analysis, we are interested in studying the concurrent tornado and flash flood events; hence, Quadrant I is our main focus. For example, we analyze the joint return-level of extreme cases where both *X* and Y exceed the corresponding 99 percentiles (100-year event). Besides, we investigate the "OR" scenario, Events from Quadrant I, II and IV, where associated precipitation, *Y*, exceeds a given threshold but not the windspeed in Quadrant II. In contrast, events in Quadrant IV describes the TORFF events where windspeed, *X*, exceeds specific threshold but not the associated precipitation. Lastly, "OR" scenarios could also include Quadrant I, where both wind speed, *X*, and associated precipitation, *Y*, exceed specific threshold *x* and *y*.

The return period of events located in Quadrants I, II, or IV where either associated precipitation and tornado windspeed (or both) exceeds a given threshold can be expressed by the joint OR scenario return period and denoted as follow:

OR Scenario

$$T^{\vee}(x,y) = \frac{\mu}{Pr[X > x \vee Y > y]} = \frac{\mu}{1 - F_{XY}(x,y)} \tag{12}$$

AND Scenario (Brunner, Seibert, and Favre 2016)

$$T^{\wedge}(x,y) = \frac{\mu}{Pr[X > x \wedge Y > y]} = \frac{\mu}{1 - F_X(x) - F_Y(y) + F_{XY}(x,y)} \tag{13}$$

Implementation of equations above in copula will be discussed in the next section.

4.5.5 Joint Return Period Using Copulas

Our objective is to investigate the dependency between TORFF driving mechanisms using copula, therefore the bivariate joint distribution of associated precipitation and windspeed is obtained through a bivariate copula model. (Requena, Mediero, and Garrote 2013) The joint distribution function used for the calculation of return period is expressed in the form of copula.

$$\{U > u\} \vee \{V > v\}$$

And

$$\{U > u\} \wedge \{V > v\}$$

Where *U* represents $F_X(X)$, associated precipitation, and *V* represents $F_Y(Y)$, resampled tornado windspeed, are transformed via the probability integral transform, respectively.

In this study, both joint copula return periods, and conditional return periods are computed and analyzed to investigate the risks of TORFF events across Canada.

Chapter 5 Results and Discussion

5.1 Introduction

Tornado events can occur in remote or populated areas. However, tornado events are reported more frequently when they affect populated areas mainly due to human observations (Cheng et al. 2015), which also agrees with Canada's tornado prone map results generated in Sills (2012). In fact, similar rationale could be applied to flash floods. It is certainly more concerning that two extreme weather events occur concurrently in the same location since the impacts can be amplified and cause more damage and property loss compared to single hazard events. Additionally, historic TORFF events have severely affected the exposed communities and caused devastations, specifically when important infrastructures are damaged. In this study, copula models have been developed to predict the return levels and return periods of TORFF events in different regions (clusters) across Canada. The correlations and dependencies of the two variables of interest in TORFF events (windspeed and precipitation) have been analyzed before the development of the copula model to assess the TORFF driving mechanisms probabilistically.

Tornado events in Canada were grouped into several clusters using Affinity Propagation (AP) based on the dissimilarities within clusters. The relationship and dependency between precipitation and wind speed were investigated before their corresponding marginal and joints distributions were developed. Each cluster shows different characteristics due to the nature of the dataset (precipitation dataset is paired with one thousand resampled windspeed dataset corresponding to the tornado intensity on Fujita scale). The determined marginal distributions of both variables for each cluster are then incorporated into copula analysis. 32 copula families were considered when determining the most suitable model for all one thousand datasets through

AIC. Joint distributions were then generated according to their marginal distributions and corresponding models.

By studying the dependencies between the TORFF variables, precipitation (APCP), and resampled wind speed (WS), the results demonstrated that some clusters have higher dependencies than others. Clusters with higher dependency show higher tendency of TORFF events happening in that region.

Note: All calculation conducted in this analysis were done by using R. Main analytical packages include: “APcluster”, “VineCopula”, and “copula”.

5.2 Affinity Propagation Clustering

Following the procedures described in the methodology section, the analyses began by determining appropriate number of clusters across Canada. Originally, we determined the best number of clusters with minimum dissimilarities within each cluster, which resulted in a total number of 22 clusters with net similarity score of -6,959.254 (the lower the better). Keep in mind that the net similarity score does scale according to the values of data points. Higher within cluster net similarity score is desired so that tornadoes that are close to each other can be clustered based on geographic information. However, more clusters indicate less data points are available per cluster. Ultimately, we identified the optimal number of clusters that meet the dataset requirements (maximum dissimilarity between clusters and at least 30 data points per cluster). Dissimilarity is determined by evaluating longitude and latitude coordinates. The algorithm determines if a point belongs to a specific cluster based of its distance to the center of the cluster. The main reason to set a dataset requirement is to facilitate the marginal distribution analysis in the later section.

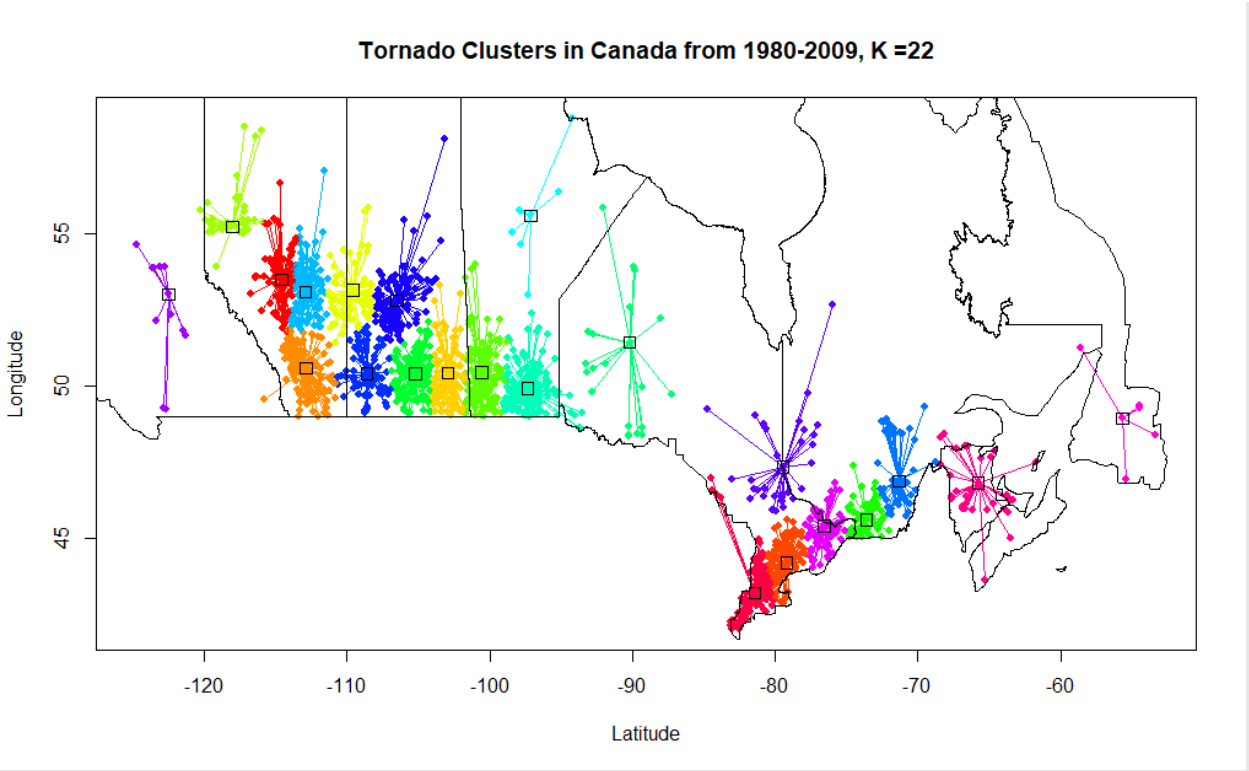


Figure 5-7: Generated Clusters based on the AP Clustering method. AP cluster Algorithm determined 22 clusters (K=22) to be the optimal number of clusters based on net similarity score.

The optimal number of clusters (K=22) generated automatically by AP cluster algorithm not only had resulted in significantly lower number of observations for the analysis, but also unable to generate an AMS of 30 years as required. To overcome this issue, AP K-mean clustering is implemented to determine the best number of clusters for this analysis. Firstly, the net similarity scores were compared across multiple clustering configurations (from three to six clusters). The changes in net similarity score, as shown in Figure 5-8, significantly decreased (by approximately 280%) when the predefined number of clusters went from four to three, which will significantly affect the quality of the dataset within the cluster and unrepresentative due to

larger spatial coverage per cluster, resulting in more extreme cases getting ignored. However, the score changes are not as drastic from six to four clusters.

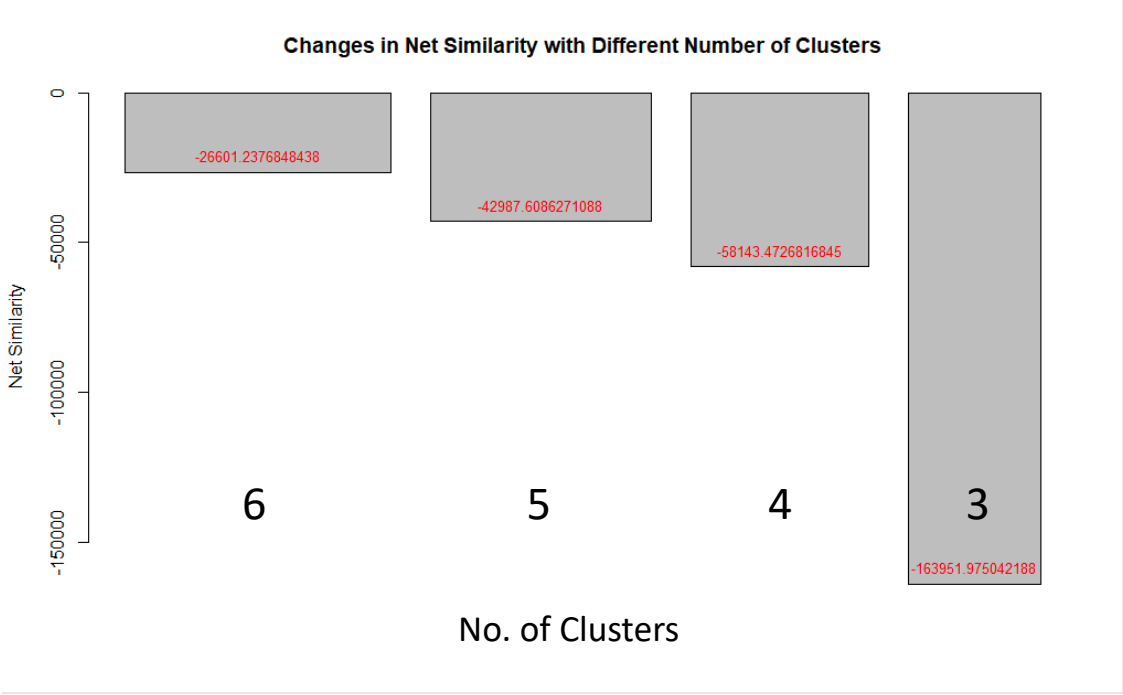
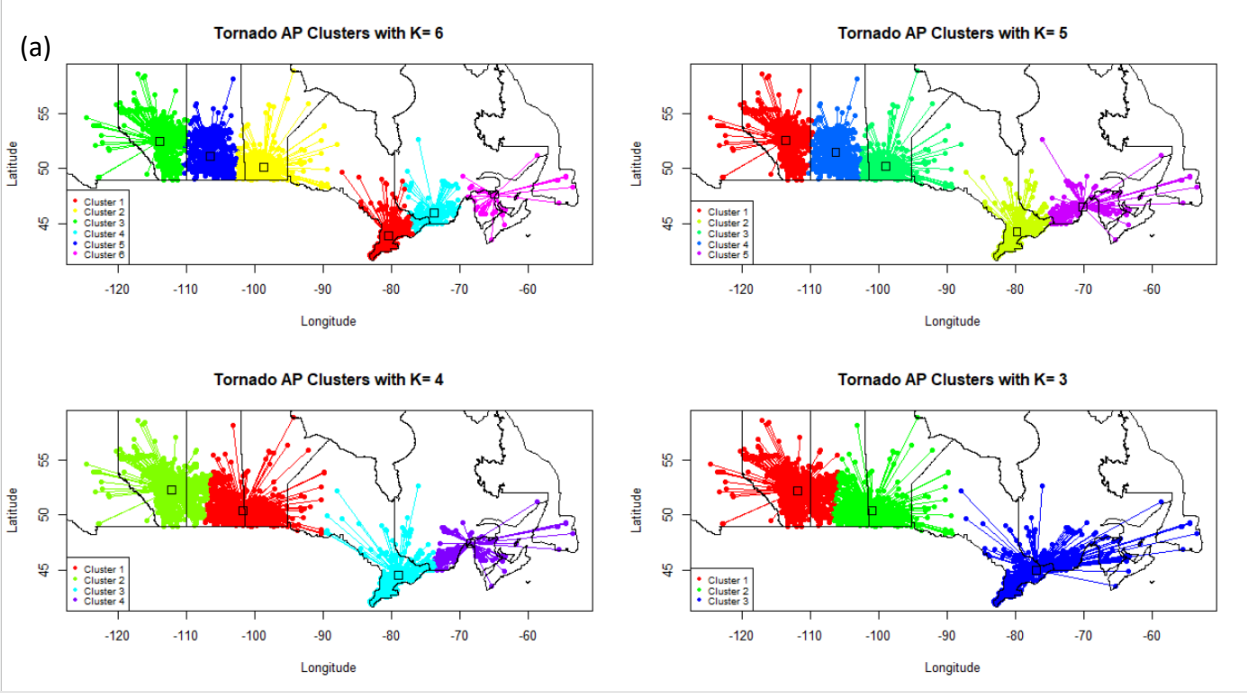


Figure 5-8: Changes in net similarity scores for different clustering configurations (higher score is preferred)



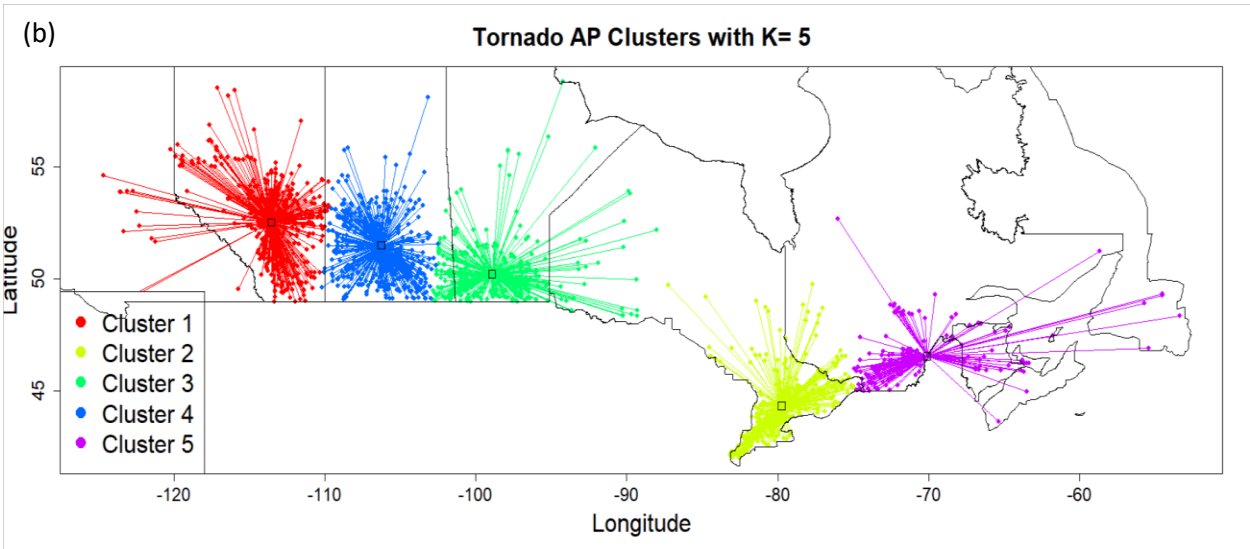


Figure 5-9: Determined clusters with different configurations. Five clusters are determined to be the optimal number for this study. Tornado cluster with K=5 (Figure 5-9b) is zoomed in to show a clearer picture of the clusters.

Before proceeding to the next step, the statistical characteristics of TORFF variables (resampled windspeed, *WS*, and associated precipitation, *APCP*) should be investigated. The characteristics of the highest values of windspeed each year (annual maxima) and the corresponding associated precipitation are shown in Table 3 and Table 4. Pairs of tornado wind speed and precipitation

data from multiple locations within each cluster are combined to develop the corresponding joint distributions. Such data-generating approach is used due to the limitation of return period calculation, where one of the parameters in Equations 12 and 13, μ, (mean-inter-arrival time of two consecutive events) is needed to calculate the return periods under "AND" and "OR" scenarios. Since APCP and WS intensities could vary widely when multiple tornadoes happen on the same day within the same cluster, it is not preferable to average the values as they could affect the results of the analysis.

Table 3: Statistical Characteristics of Resampled Windspeed. Clusters correspond to Figure 5-9(b).

Statistical Measures for Simulated Windspeed (km/hr)	**Cluster**				
	1	**2**	**3**	**4**	**5**
	Mean (LL-UL)*	**Mean (LL-UL)**	**Mean (LL-UL)**	**Mean (LL-UL)**	**Mean (LL-UL)**
Minimum	62.859(62.692-63.026)	63.268(63.076-63.461)	62.952(62.779-63.126)	62.169(62.042-62.295)	62.74(62.575-62.906)
Median	103.248(103.024-103.471)	110.851(110.688-111.014)	105.537(105.349-105.724)	93.808(93.519-94.097)	97.413(97.111-97.715)
Maximum	293.717(292.536-294.899)	226.449(225.756-227.141)	228.563(227.968-229.158)	279.823(278.385-281.262)	163.607(163.251-163.964)
Mean	124.724(124.539-124.908)	121.259(121.074-121.444)	121.5(121.327-121.672)	110.877(110.702-111.052)	103.841(103.651-104.031)
St. Deviation	59.78(59.546-60.013)	45.116(44.914-45.318)	47.87(47.668-48.072)	52.585(52.353-52.817)	30.563(30.372-30.754)
Skewness	1.298(1.286-1.309)	0.701(0.69-0.713)	0.726(0.716-0.735)	1.658(1.645-1.671)	0.482(0.473-0.491)
Kurtosis	0.972(0.931-1.013)	-0.459(-0.484-(-0.434))	-0.613(-0.632-(-0.593))	2.189(2.123-2.256)	-1.001(-1.018-(-0.983))

*LL = The lower limit of 95% confidence interval. UL = upper limit of 95% confidence interval.

Table 4: Statistical Characteristics of associated precipitation. Clusters correspond to Figure 5-9(b).

Statistical Measures for 3-Hourly Precipitation, *PR* (mm/3hr)	**Cluster**				
	1	**2**	**3**	**4**	**5**
Minimum	0.414	0.000	0.003	0.086	0.000
Median	4.938	4.243	4.207	3.219	4.961
Maximum	30.594	20.148	17.359	15.328	8.992
Mean	6.727	5.869	6.133	5.045	4.261
St. Deviation	6.522	5.429	4.695	3.975	3.163
Skewness	1.937	1.152	0.921	0.957	-0.042
Kurtosis	4.028	0.406	-0.217	0.029	-1.606

Cluster 1 shows the highest values across all statistical measures for PR in general and cluster 5 shows the minimum except for median. For resampled WS, the highest maximum resampled WS corresponds to Cluster 1 at the mean value of 293.717 km/s, whereas the lowest maximum WS is associated with Cluster 5 at 163.607 km/h.

TORFF variables (*PR, WS*) in majority clusters are positively skewed (except cluster 5 for *PR*) indicates heavy tail distributions might be required to represent the corresponding variabilities (Naz et al. 2019).

5.3 Determining the Dependence of TORFF Driving Mechanisms

To measure the statistical dependence between the TORFF variables, three correlation methods were used: Pearson's (linear), Kendall's (rank-based), and Spearman's (rank-based) correlations. The purpose of including Pearson's correlation method is to examine linear dependency between the TORFF variables, with the assumption of linear relationship; however, the results might not be as reliable due to its sensitivity to outliers. Further, Kendall's and Spearman's coefficients can determine the monotonic increasing or decreasing trend. They are used explicitly in copula due to their "invariancy under monotonic non-linear transformation" (Naz et al. 2019).

Table 5 shows the dependency between *PR* and *WS* of a TORFF event with their corresponding *p*-values. Note that since *WS* data were resampled, a confidence interval of 95% is shown to quantify the corresponding uncertainties.

Table 5: Dependence Measures for APCP and WS. Clusters correspond to Figure 5-9(b).

	Cluster				
	1	**2**	**3**	**4**	**5**
	Pearson's coefficient, r Mean (LL- UL)**				
CC*	0.457 (0.45-0.46)	0.325 (0.321-0.329)	0.296 (0.292-0.300)	0.310 (0.307-0.313)	0.233 (0.226-0.239)
P-value	0.016 (0.015-0.017)	0.101 0.096-0.106)	0.134 (0.128-0.139)	0.110 (0.106-0.114)	0.297 (0.284-0.310)
	Kendall's Tau, τ Mean (Lower Limit- Upper Limit)				
CC*	0.214 (0.210-0.218)	0.21 (0.206-0.213)	0.198 (0.195-0.202)	0.230 (0.225-0.235)	0.156 (0.151-0.161)
P-value	0.137 (0.129-0.144)	0.142 (0.135-0.150)	0.165 (0.157-0.174)	0.136 (0.126-0.146)	0.333 (0.318-0.349)
	Spearman's Rho,ρ Mean (Lower Limit- Upper Limit)				
CC*	0.305 (0.300-0.310)	0.273 (0.269-0.278)	0.287 (0.283-0.292)	0.33 (0.323-0.337)	0.227 (0.221-0.234)
P-value	0.135 (0.128-0.142)	0.179 (0.172-0.179)	0.151 (0.144-0.158)	0.133 (0.124-0.143)	0.315 (0.301-0.329)

*CC = correlation coefficient
**LL = The lower boundary of 95% confidence interval. UL = upper boundary of 95% confidence interval.

The correlation coefficient (CC) and their respective p-values in all three methods (see Table 5) for APCP and WS dataset in each cluster were generated and analyzed. All clusters show positive correlation; however, the majority of the clusters show correlations that are not statistically at the 5% level except for Cluster 1 for Pearson's correlation method. Clusters 1, 2 and 4 show significant correlation at the 15% level based on Kendall's Tau.

When the null hypothesis is not rejected, it only indicates the incompatibility of data with a specific statistical model that typically involves the null hypothesis (Wasserstein and Lazar 2016) and acts as supplemental information to corroborate our rationale in this study. As discussed before, correlation metrics can provide initial measures of the overall dependencies

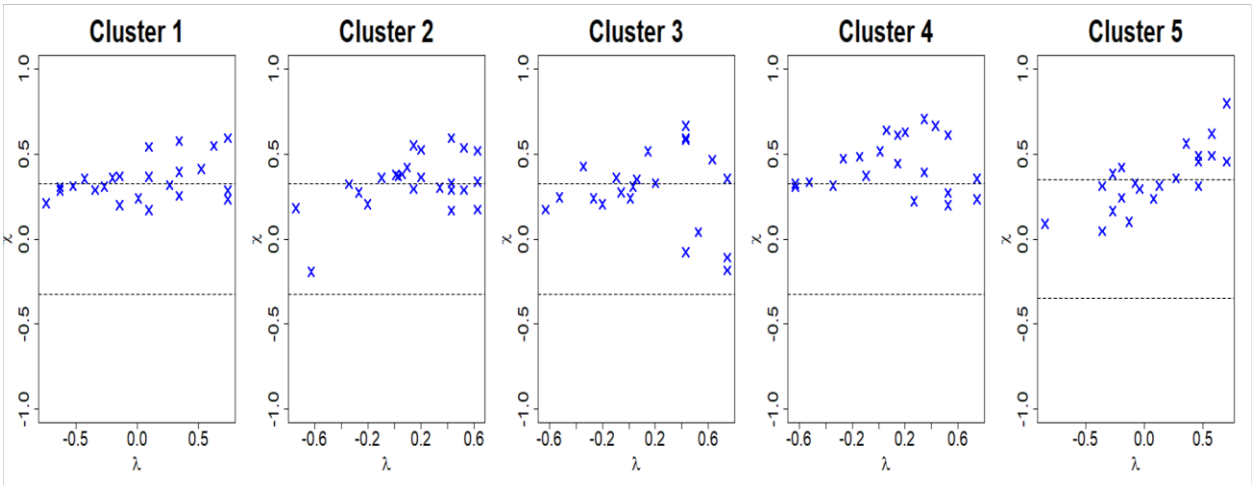
however this information is incomplete and does not fully represent the dependence structure (including the dependencies at the tails). Therefore, further analyses are performed to study the complex underlying mechanisms of TORFF event.

5.4 Dependence structure between Precipitation and Resampled Windspeed Using Copula Function

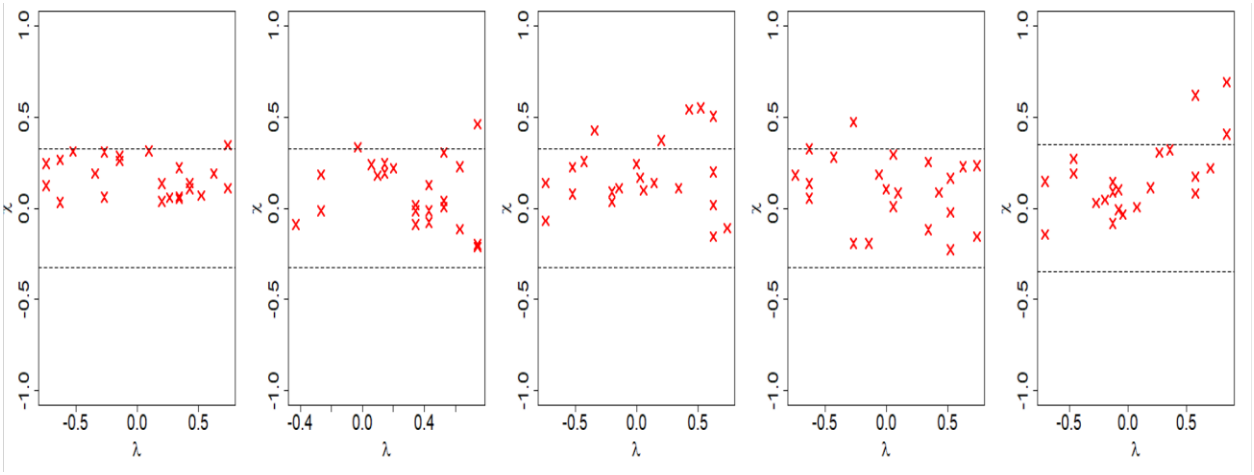
Dependency between the variables of interest (APCP and WS) can be visualized through the scatterplot of standardized ranks. An example out of one thousand unique sets of WS resampled data coupled with precipitation data is shown to visualize the dependency between the variables in Figure 5-10. Other results such as Chi-plot and K-plot are displayed in Figures 5-10 and 5-11, as an example, respectively. The purpose of this step is to further investigate the dependencies of resampled windspeed, and associated precipitation.

For interpretation of Chi-plot, the dependencies are considered weak if most of the observation points lied within the confidence band (also known as an asymptotic confidence interval, ACI). Fisher and Switzer (1985) explained that n bivariate numerical datapoints, (X_i, Y_i) are transformed into n pairs $(\lambda_i,\ \chi_i)$ to reveal more detailed and explicit information regarding the nature of association between *X* and *Y*. The χ (y-axis) transformation provides an indication of *X*-*Y* association separately for each data point. χ_i represents a correlation coefficient between dichotomized *X* values and dichotomized *Y* values, which results in all values of χ_i lying in the interval [-1,1]. The λ (x-axis) transformation help position that point with respect to *X* and *Y* marginal distributions. The values of λ_i should lie in the interval of [-1,1]. If the datapoints are independent and show no correlation, the values of λ_i should be uniformly distributed. On the other hand, if *X* and *Y* are associated, then the values of λ_i may show clustering. More detailed explanation of transformation can be found in Fisher and Switzer (1985).

Due to the amount of resampling dataset, the Chi-plots with the highest dependence and weakest dependence are compared for each cluster (See Figure 5-10(a) and (b)).

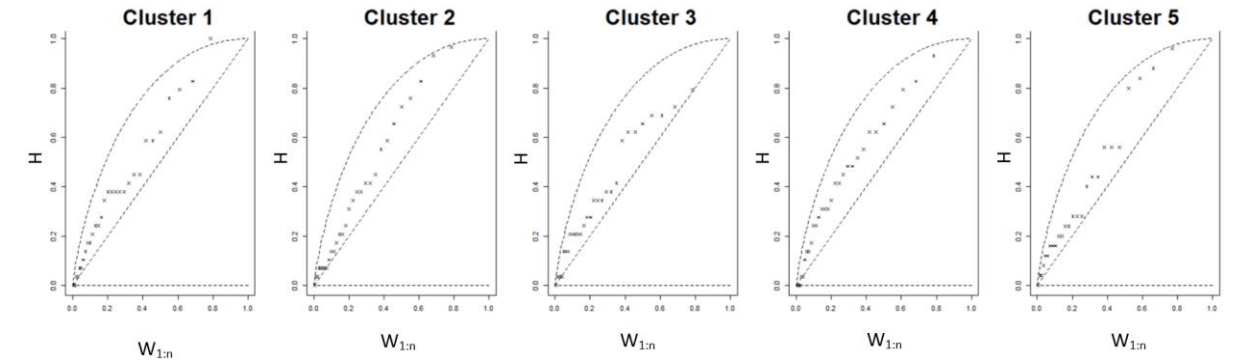


(a)

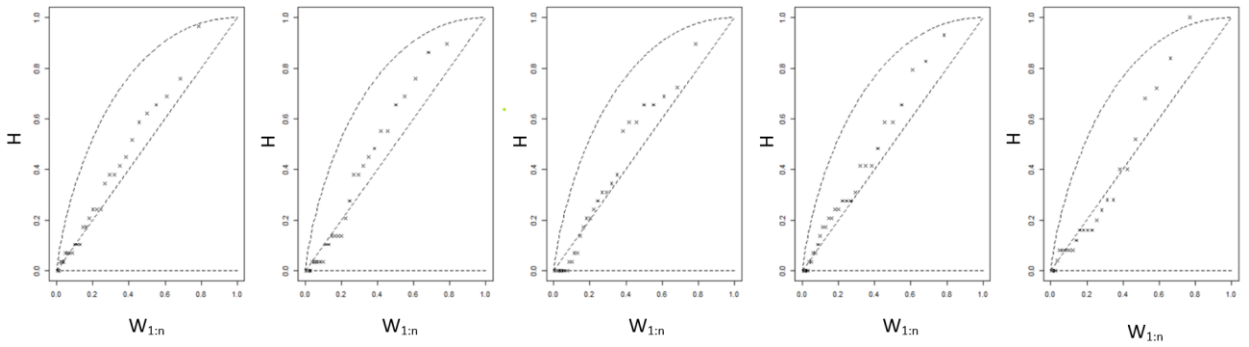


(b)

Figure 5-10: Chi-plot for each cluster with for resampled dataset with a) highest Kendall's Tau and b) lowest Kendall's Tau. The clusters correspond to Figure 5-9(b). All clusters are consistently showing weak dependency fall between the ACI.



(a)



(b)

Figure 5-11: K-plots for resampled dataset with highest (top row) and lowest tau (bottom). The clusters correspond to Figure 5-9(b). Cluster 1,3, and 4 in top row show more deviation from the diagonal line. Data points in each cluster in bottom row show weaker dependence. The data has been transformed to make the margins uniform on the interval [0,1] for x and y axis.

The strongest and weakest dependence Chi-plots were shown using Tau. Resampled pair with the highest Tau are chosen and plotted for each cluster as shown in Figure 5-10 above. The same procedures apply to the resampled pair with the weakest dependence (based on Tau) for each cluster. From Figure 5-10, we can see that Cluster 4 shows higher dependency compared to the other clusters. Meanwhile, Chi-plots with the weakest dependence show that majority of the observation points fell between the ACI, which are consistent across all clusters.

As for K-plots, the dependencies are analyzed by observing the plot of the data points along the diagonal line. Observations above the diagonal line indicate positive dependence, and observations below the diagonal line indicate negative dependence. In addition, the further away

the data points deviate from the diagonal line, the higher the association between the variables. (Genest and Boies 2003). From Figure 5-11 resampled datasets with the highest and lowest correlation from each cluster were chosen to generate the K-plots. From the plots, we can see that cluster 2 shows higher dependency at the upper tail, and cluster 4 shows higher dependency in general compared to other clusters. As for the K-plots generated with the lowest correlation, most data points are close to the diagonal lines, with some even falling onto the x-axis (Azam et al. 2018).

5.5 Marginal Distribution Analysis

By implementing the distribution fitting for each unique set of data, we notice that for *WS*, most of the marginal distributions are GEVs. However, some clusters consist of other distributions such as lognormal and Cauchy. This could be due to the random resampling nature of the windspeed that resulted in generating different types of distributions in the same cluster. Since the dataset for *WS* were generated using annual maxima and majority of the fitted marginal distribution for all clusters are GEV, all 1000 resampled windspeed data will be treated as GEV distributions. On the other hand, for APCP, two out of five clusters were determined to be gamma, followed by lognormal and Weibull.

Table 6: Marginal Distribution Composition and their Corresponding Parameters and Statistical Measures

Cluster	Variable	Determined Marginal Distribution Composition (n/1000)	Parameters, mean (LL-UL)**				Statistical Measures, mean (LL-UL)**		
			Location	Scale	Shape	Rate	K-S* Statistic	K-S Statistic P-Values	Loglikelihood
1	WS	GEV (1000)	92.014(91.757-92.271)	29.287(29.121-29.453)	0.447(0.439-0.455)	-	0.162(0.161-0.163)	0.391(0.381-0.401)	-155.956(-156.064--155.847)
	APCP	Lognormal	-	-	6.727	6.413	0.697	3.77E-15	-85.904
2	WS	GEV (533)	97.167(96.805-97.529)	29.687(29.5-29.874)	0.213(0.204-0.222)	-	0.129(0.127-0.131)	0.656(0.639-0.673)	-152.632(-152.813--152.451)
		Lognormal (427)	-	-	121.378(121.112-121.645)	44.22(43.93-44.51)	0.995(0.995-0.995)	~0	-154.006(-154.18--153.833)
	APCP	Gamma	-	-	0.88	0.15	0.106	0.855	-82.92
3	WS	GEV (799)	94.789(94.481-95.097)	29.822(29.655-29.988)	0.294(0.285-0.303)	-	0.143(0.142-0.145)	0.537(0.523-0.551)	-154.045(-154.176--153.914)
		Lognormal (213)	-	-	121.811(121.445-122.177)	46.123(45.71-46.536)	0.994	~0	-155.125(-155.364--154.887)
	APCP	GEV	3.715	2.96	0.218	-	0.121	0.726	-83.71
4	WS	GEV (997)	83.38(83.136-83.625)	21.479(21.335-21.622)	0.513(0.504-0.522)	-	0.215(0.214-0.217)	0.125(0.12-0.13)	-147.779(-147.902--147.656)
		Cauchy (3)	105.577(101.876-109.278)	24.548(23.004-26.091)	-	-	0.231(0.197-0.266)	0.083(0.01-0.156)	-148.531(-149.655-147.406)
	APCP	Weibull	-	1.261	5.417	-	0.152	0.446	-77.404
5	WS	GEV (745)	88.664(88.296-89.033)	22.147(21.965-22.329)	0.103(0.087-0.119)	-	0.138(0.136-0.14)	0.652(0.636-0.668)	-123.003(-123.177--122.829)
		Lognormal (241)	-	-	103.959(103-602-104.316)	30.56(30.286-30.833)	0.999	~0	-124.602(-124.824--124.381)
		Gamma (14)	-	-	11.913(11.241-12.585)	0.115(0.107-0.122)	0.125(0.114-0.135)	0.755(0.673-0.837)	-124.739(125.719-123.759)
	APCP	Gamma	-	-	0.491	0.115	0.257	0.054	-57.721

*K-S = Kolmogorov-Smirnov

**LL = The lower limit of 95% confidence interval. UL = upper limit of 95% confidence interval.

Note: The parameters correspond to the dominant distribution.

Table 6 summarizes the marginal distribution parameters and Kolmogorov-Smirnov (K-S) statistics with a *p*-value for *WS*, and *APCP* of each cluster. Higher K-S statistic (with a maximum value of 1) indicate that the distribution well-represent the variability of the data. Additionally, higher loglikelihood indicates that the model fit well with the datasets. Highest score were analyzed to determine whether the marginal distributions are a good fit for *APCP* and *WS* in each cluster. The determined marginal distributions for both variables were considered to be the best fit after comparing with other distributions. The statistical measures for APCP and WS in each cluster, show that some distributions fit overwhelmingly well (K-S statistic value closer to 1) and some not as well (lower than 0.2 K-S statistics value). For example, WS in cluster 2 and cluster 5 shows K-S statistic values of almost 1, while WS in the remaining clusters shows less of a good fit ranging approximately from 0.12 to 0.25. Further, associated precipitation (APCP) varies widely across all clusters for its K-S statistic value ranging from 0.2 to 0.7. The K-S statistic computed for resampled windspeed and associated precipitation in Table 6 has the highest score compared to other fitted marginal distributions.

Figures 5-17 to 5-21 are examples of determining marginal distribution plots for each cluster. Marginal distributions for *WS* shown in the figures below were randomly generated based on their pool of marginal distributions. As observed in Table 6, 1000 out of 1000 resampled windspeed data followed the GEV distribution. However, Cluster 2's results for *WS* consist of two major marginal distributions: GEV and Lognormal. Even though datasets are generated using AMS and GEV distributions are expected, other distributions such as lognormal in Cluster 2 and 3 were determined. The same procedures were applied to the remaining clusters.

From Section 4.2 we learned that there are three types of extreme distributions depending on the shape parameter, ξ. Shape parameter values from Table 10 shows positive values ranging from

approximately 0.1 to 0.5 for all clusters. As mentioned previously, shape parameter values greater than 0 were deemed to be Type II (or Fréchet) extreme distributions that consist of fat tails. Fat tail distributions indicate that towards extreme cases, the return levels are higher compared to other distributions, it may lead to overestimation of return levels for resampled windspeeds (Willems et al. 2007) when extrapolating for longer period of time.

Figures 5-12 to 5-16 compare the empirical and theoretical probability distribution functions (PDFs) and their corresponding cumulative distribution functions (CDFs) for the resampled windspeed and associated precipitation. CDFs and PDFs were computed using GEVs for all clusters for most of the resampled windspeed. Since the values of associated precipitation is dependent on annual maxima of resampled windspeed, its corresponding marginal distribution will not be the same. Cluster 1 to Cluster 5 were computed using lognormal, Gamma, GEV, Weibull, and Gamma, respectively. The derived theoretical CDFs and PDFs show good agreement with the empirical ones, which indicates that these probability distributions fit well the observed data.

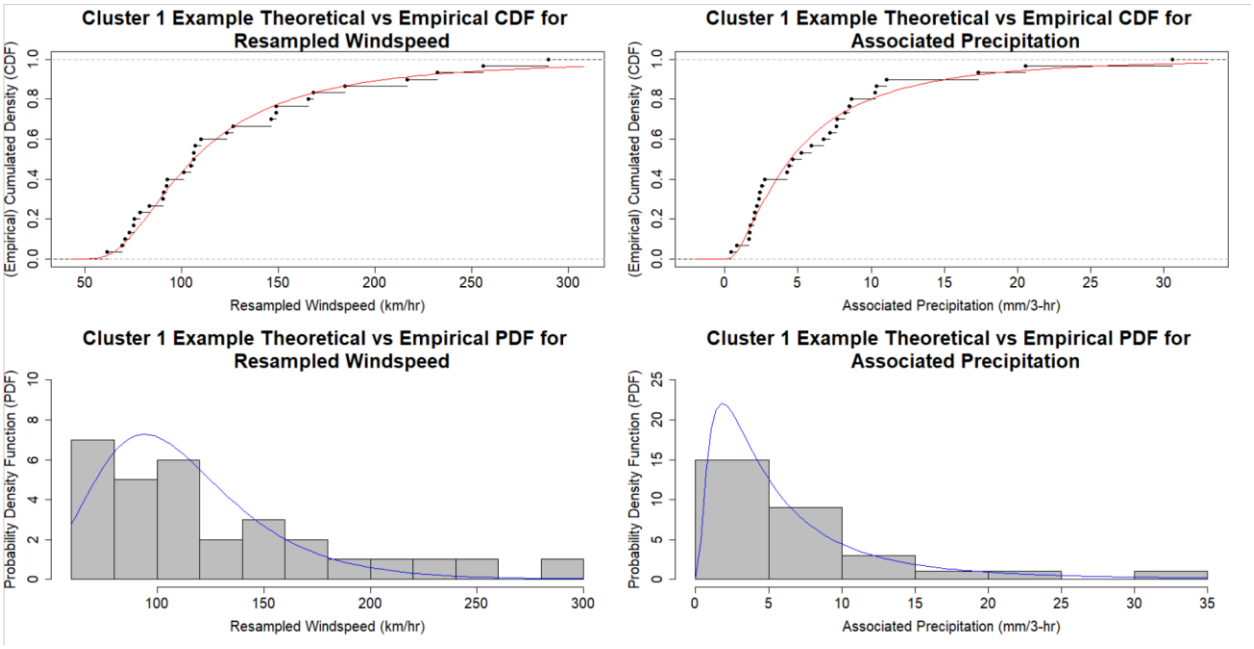


Figure 5-12: CDF (top row) plots and PDF (bottom row) plots for resampled windspeed and associated precipitation for Cluster 1. Red curve in CDF plots and blue curve in PDF plots show theoretical CDF and theoretical PDF, respectively.

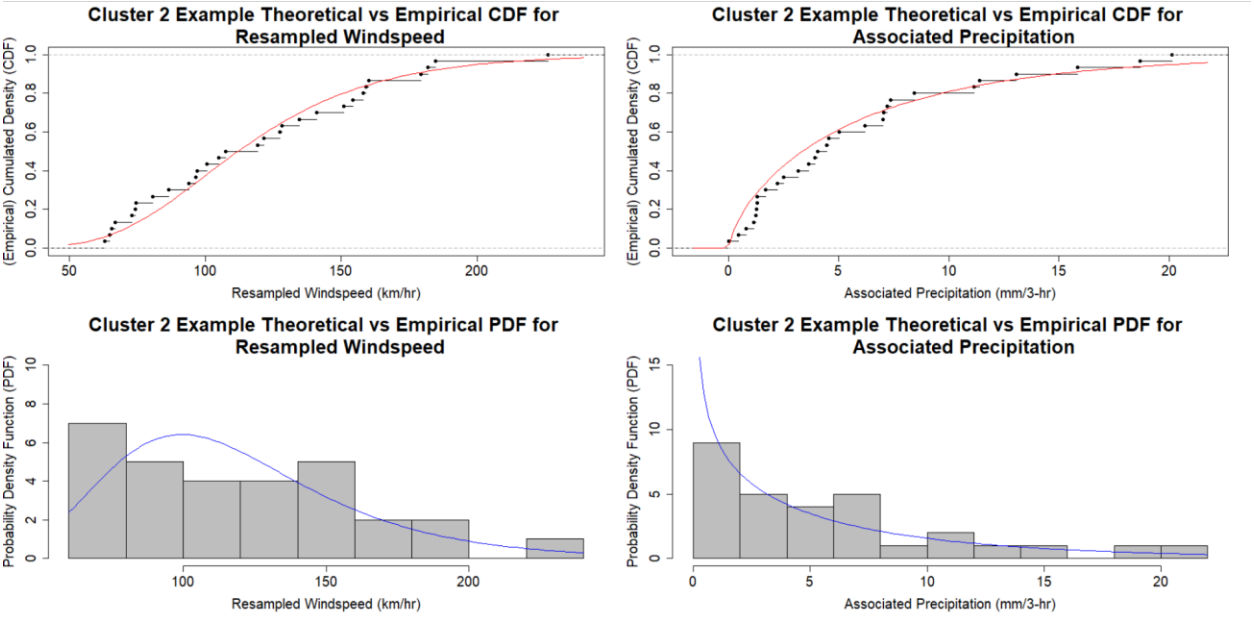


Figure 5-13: CDF (top row) plots and PDF (bottom row) plots for resampled windspeed and associated precipitation for Cluster 2. The red curve in CDF plots and blue curve in PDF plots show theoretical CDF and theoretical PDF, respectively.

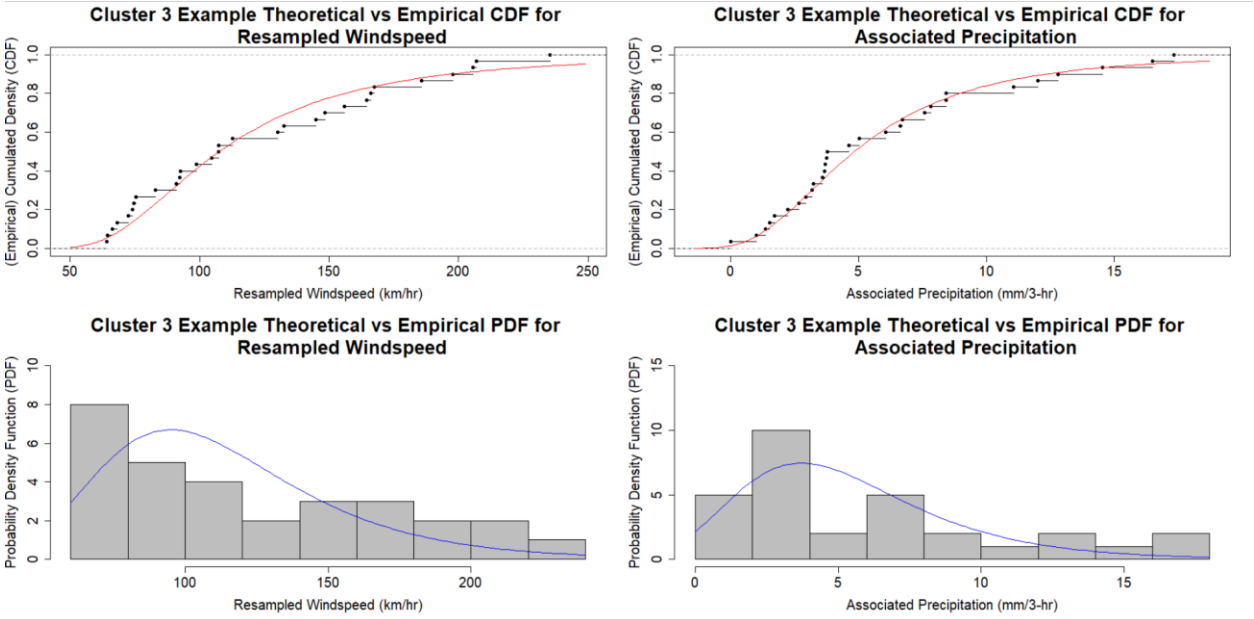


Figure 5-14: CDF (top row) plots and PDF (bottom row) plots for resampled windspeed and associated precipitation for Cluster 3. Red curve in CDF plots and blue curve in PDF plots show theoretical CDF and theoretical PDF, respectively.

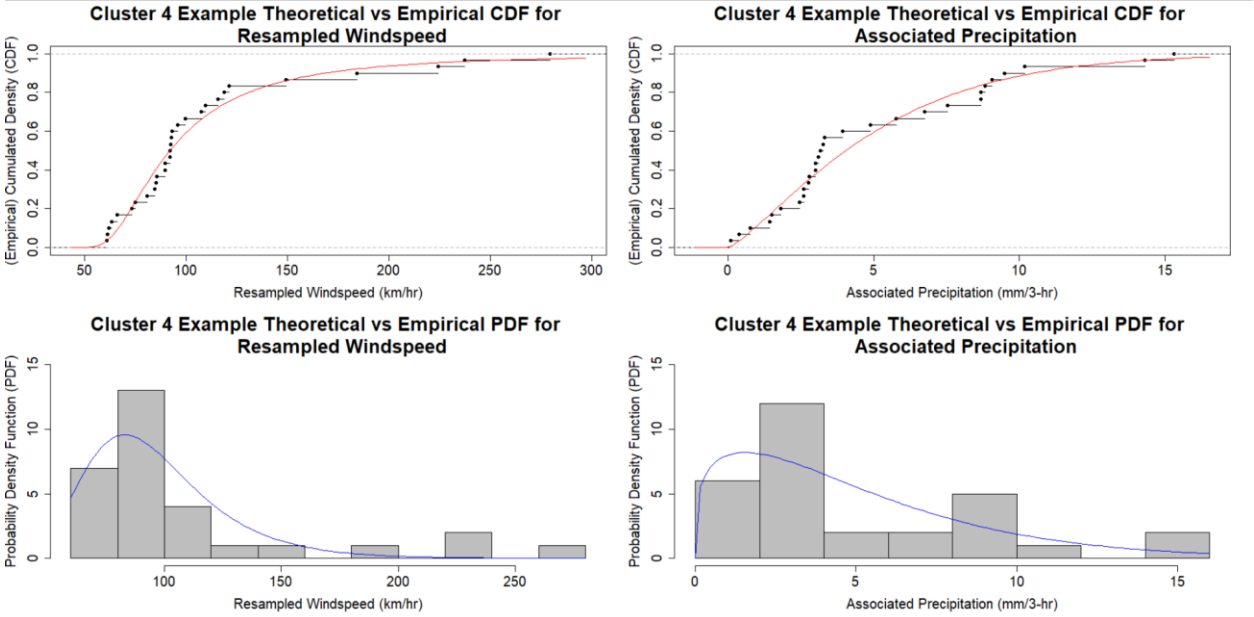


Figure 5-15: CDF (top row) plots and PDF (bottom row) plots for resampled windspeed and associated precipitation for Cluster 4. Red curve in CDF plots and blue curve in PDF plots show theoretical CDF and theoretical PDF, respectively.

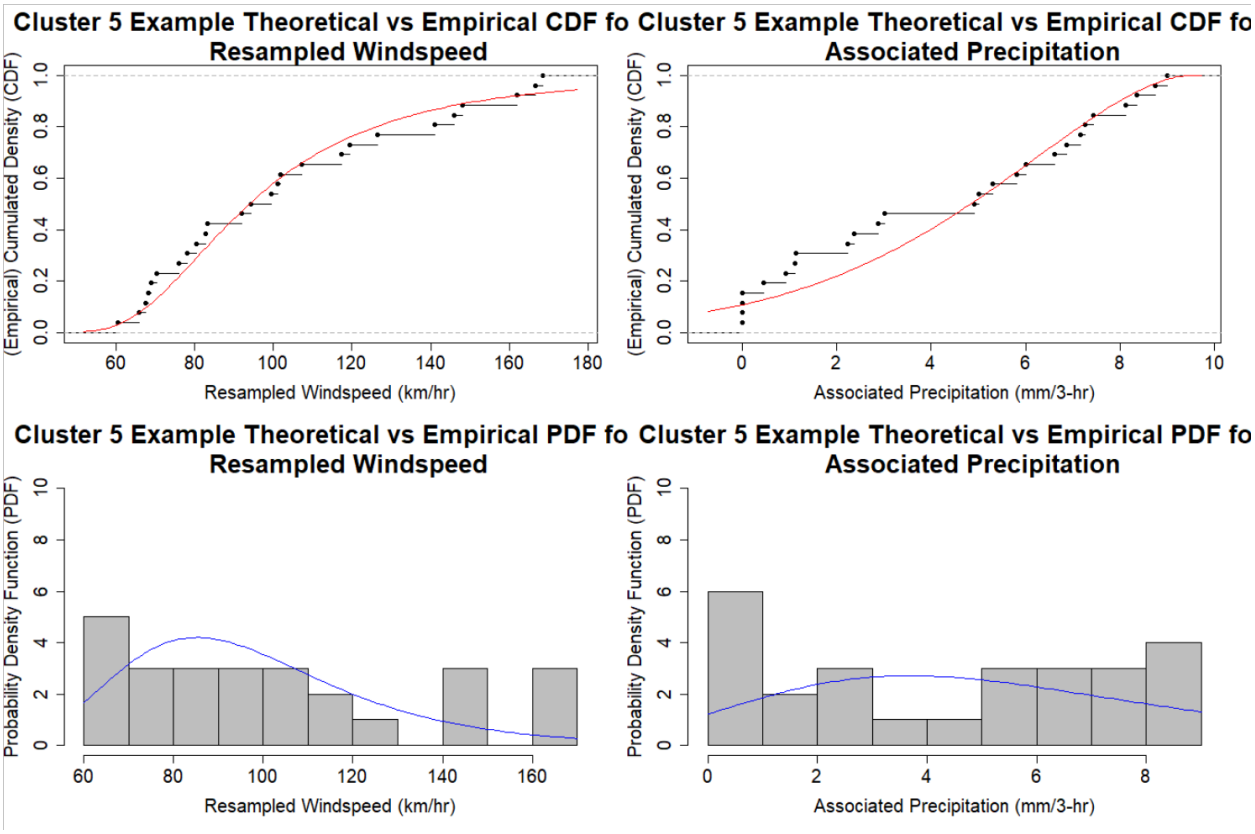


Figure 5-16: CDF (top row) plots and PDF (bottom row) plots for resampled windspeed and associated precipitation for Cluster 5. Red curve in CDF plots and blue curve in PDF plots show theoretical CDF and theoretical PDF, respectively.

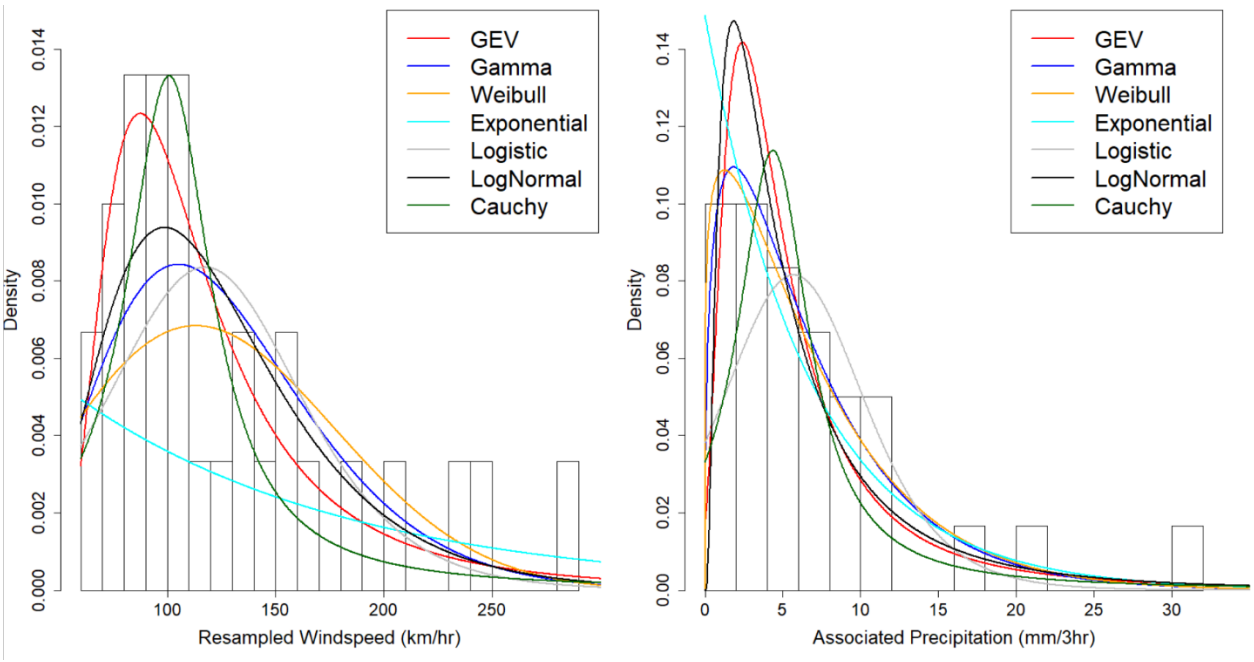


Figure 5-17: Cluster 1 APCP and WS histogram and marginal distribution.

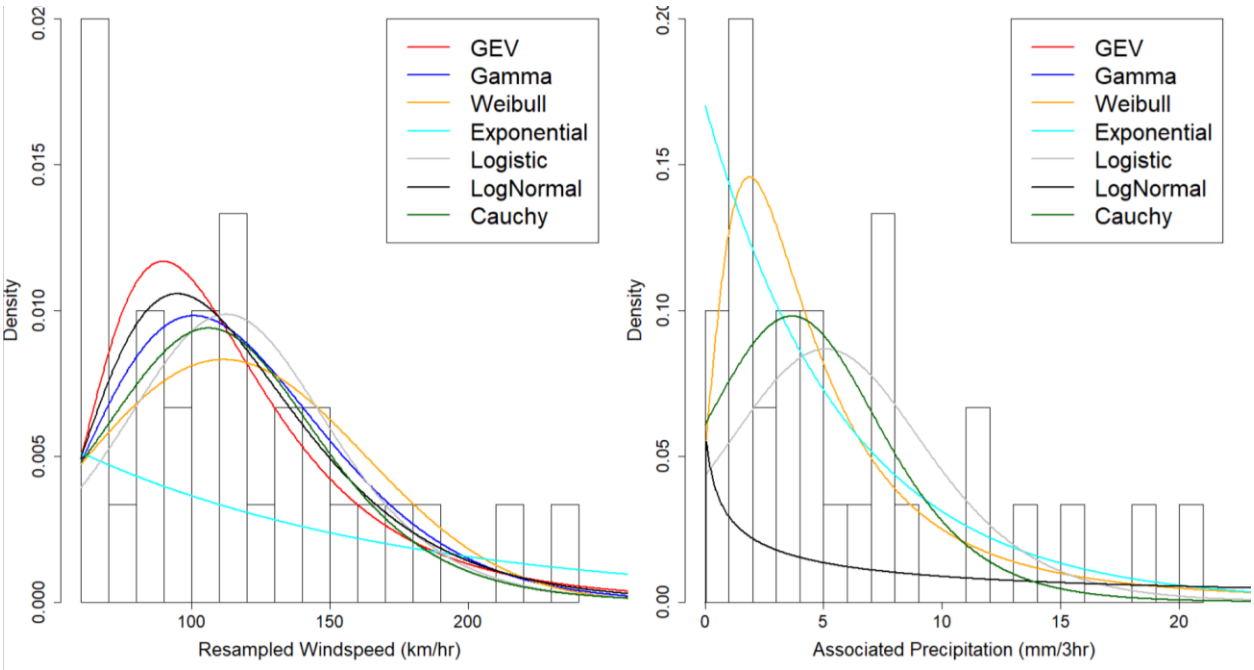


Figure 5-18: Cluster 2 APCP and WS histogram and marginal distribution. The resampling nature of the windspeed resulted in different marginal distributions.

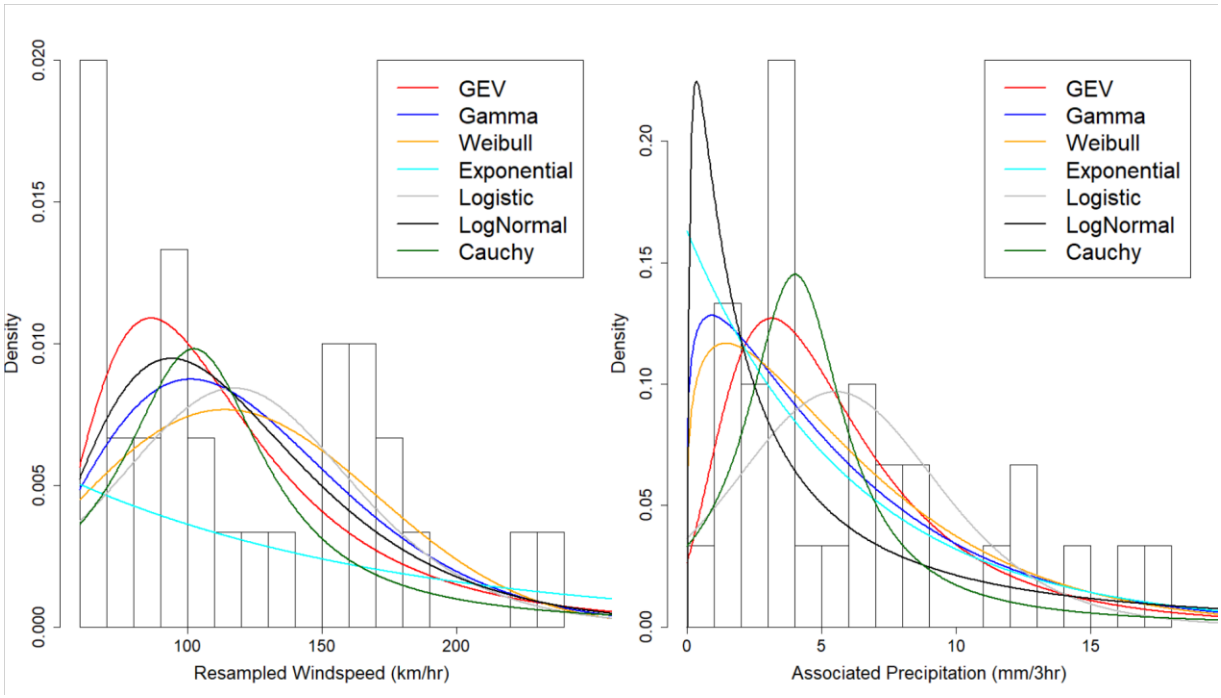


Figure 5-19: Cluster 3 APCP and WS histogram and marginal distribution. The resampling nature of the windspeed resulted in different marginal distributions.

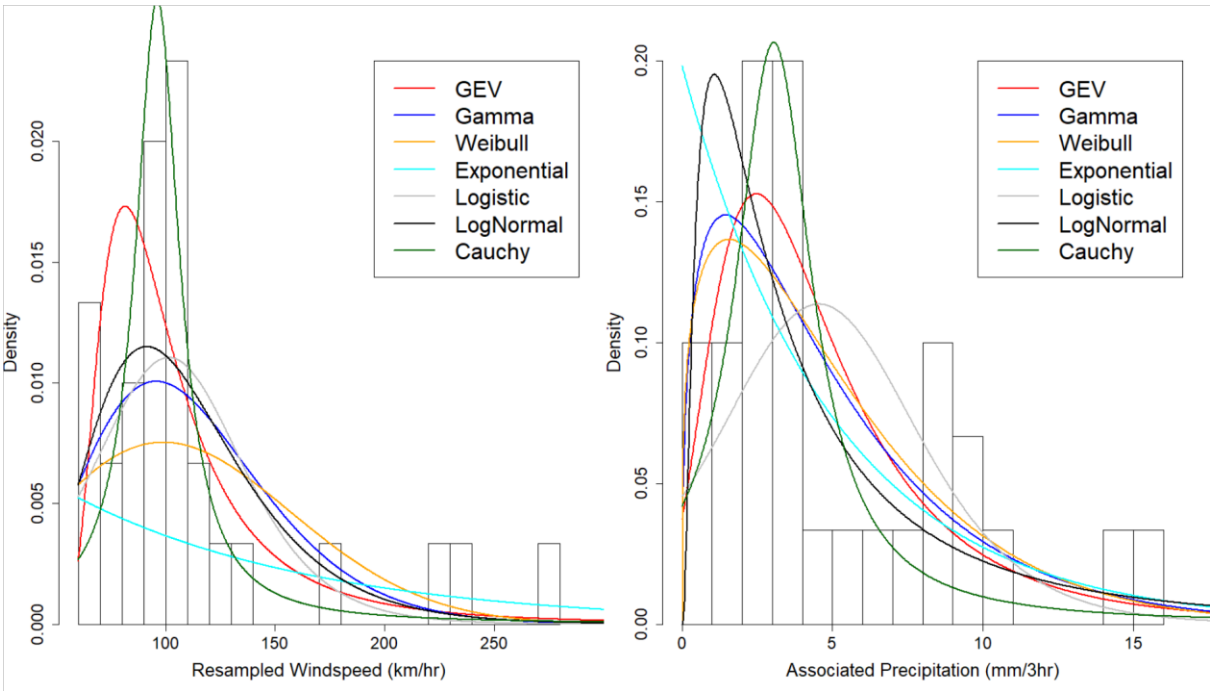


Figure 5-20: Cluster 4 APCP and WS histogram and marginal distribution. The resampling nature of the windspeed resulted in different marginal distributions.

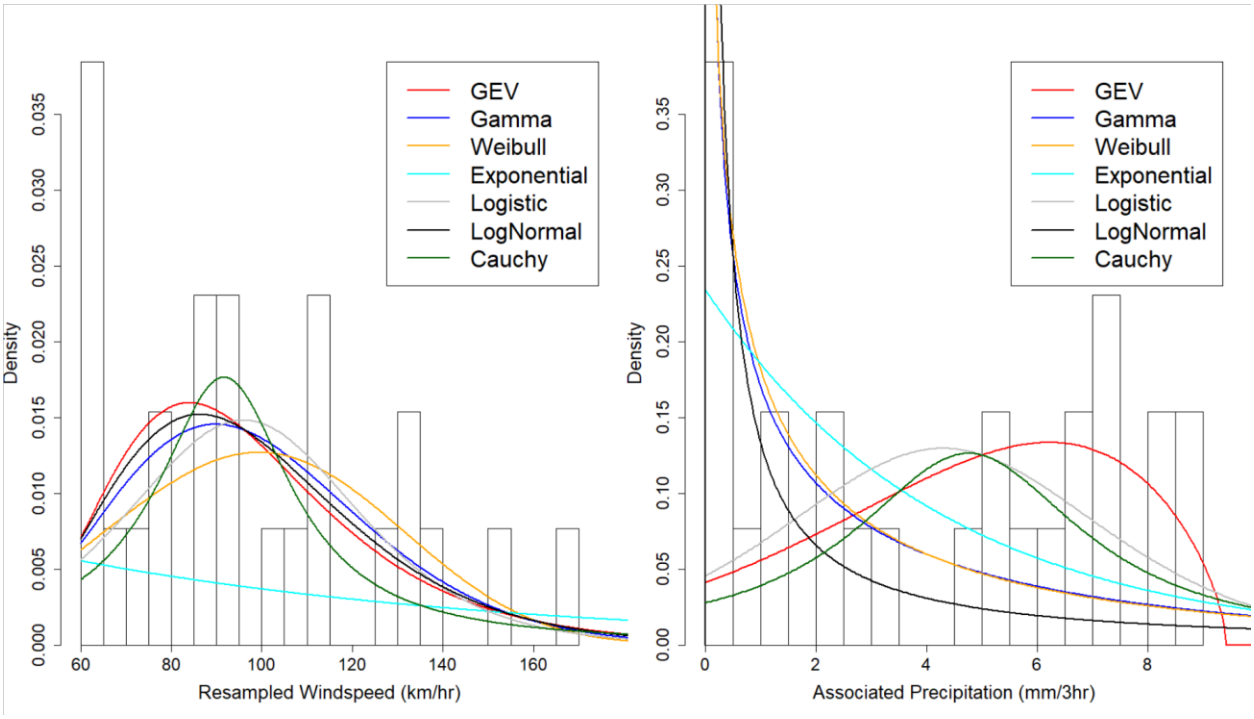


Figure 5-21: Cluster 5 APCP and WS histogram and marginal distribution. Red and blue curves show their corresponding fitted marginal distributions. The resampling nature of the windspeed resulted in different marginal distributions.

5.6 Copula Modelling

As mentioned in the previous sections, one thousand unique datasets consisting of APCP and WS were generated through resampling. First, WS was resampled one thousand times according to the corresponding Fujita Rating and its windspeed range. The resampled windspeed was then paired with their associated precipitation. Next, each of the one thousand paired datasets went through the copula model selection process through AIC_c.

As observed in Figures 5-22 to 5-24, each cluster consists of 50% to 90% independent copulas followed by gaussian, clayton, or t copulas. Note that since tornado windspeeds are estimates based on the Fujita rating wind speed interval and may not represent the actual windspeed

accurately. Therefore, 1000 samples are extracted to represent the wind speed and the corresponding uncertainties.

If a dataset is determined to be independent means, there is no dependency between the variables. The probability of WR and APCP does not affect each other. The probability of extreme WS and APCP happening at the same time in the same location can be interpreted as $p(WS) * p(APCP)$ since they are both independent in certain clusters. However, we do not rely on the datasets for their individual behavior, we are going to treat them as one big dataset.

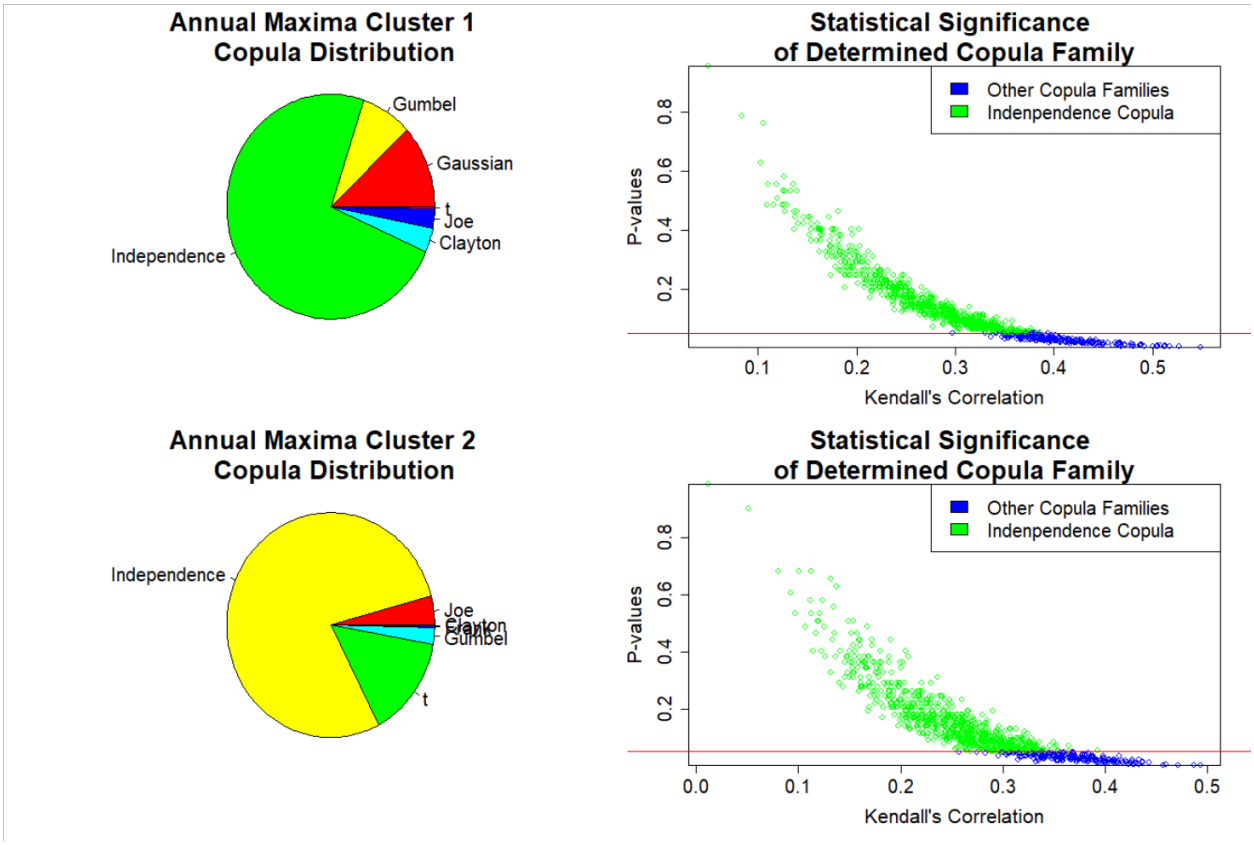


Figure 5-22: Copula family composition and statistical significance for cluster 1 and cluster 2. Pie charts (left) show the composition of determined copula functions (out of 1000 resampled datasets). Scatterplots (right) show the relationship between p-values of each determined copula function and its corresponding Kendall's correlation.

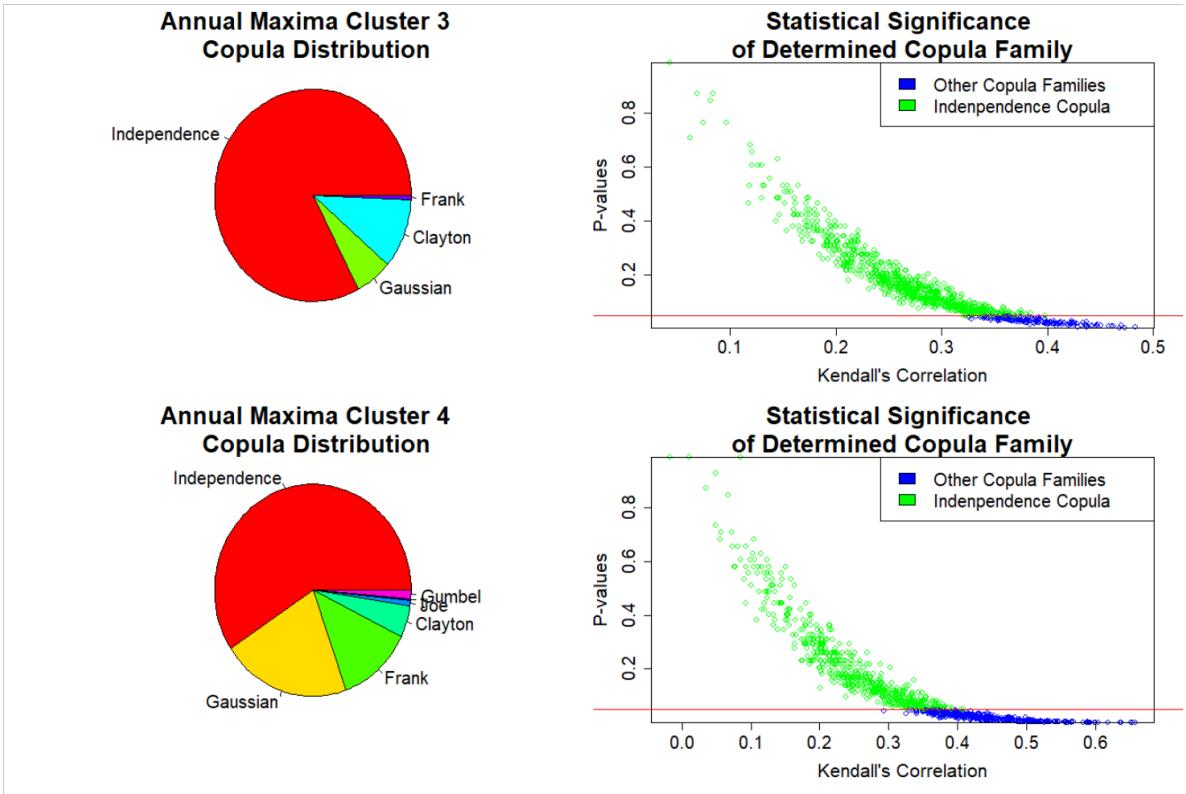


Figure 5-23: Copula Family Composition and Statistical Significance for Cluster 3 and Cluster 4. Pie charts (left) show the composition of determined copula functions (out of 1000 resampled datasets). Scatterplots (right) show the relationship between p-values of each determined copula function and its corresponding Kendall's correlation.

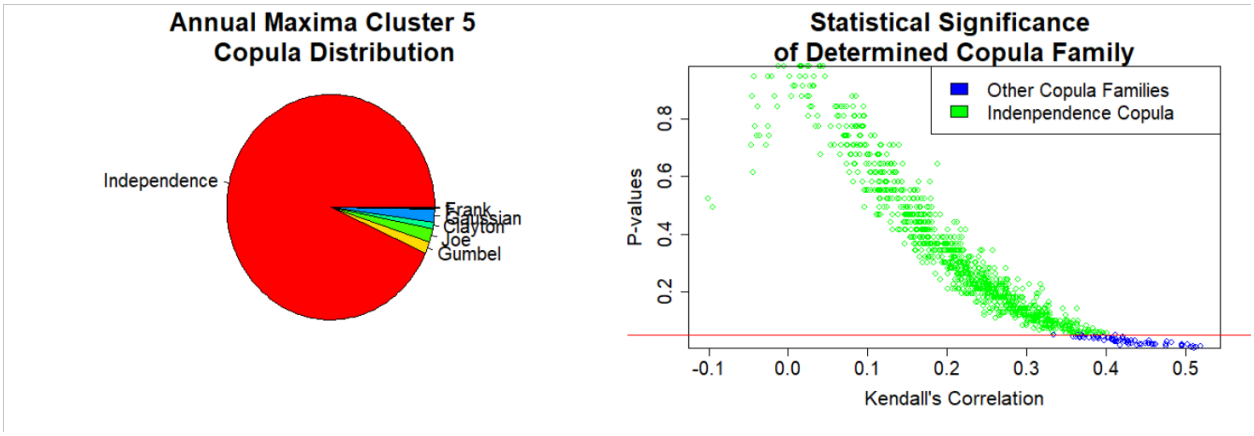


Figure 5-24: Copula Family Composition and Statistical Significance for Cluster 5. Pie charts (left) show the composition of determined copula functions (out of 1000 resampled datasets). Scatterplots (right) show the relationship between p-values of each determined copula function and its corresponding Kendall's correlation.

The pie charts in Figures 5-22 to 5-24 result from applying a different copula function to each resampled dataset (e.g., an F1 tornado event are resampled 1000 times from its corresponding windspeed range on the Fujita scale, which is 120-170 km/hr). Using such a method, results in different dependencies between resampled windspeed and associated precipitation. Detailed distribution of copula families is shown in Table 7. Cluster 4 showed the least amount of independence copula.

Table 7: Identified Copula Function Composition for Each Cluster. Frequencies are out of 1000 resampled datasets.

1		**2**		**3**	
Copula Functions	**Frequency**	**Copula Functions**	**Frequency**	**Copula Functions**	**Frequency**
Independence	736	Independence	784	Independence	826
Gaussian	119	t	147	Clayton	106
Gumbel	79	Joe	42	Gaussian	61
Clayton	35	Gumbel	24	Frank	7
Joe	28	Frank	2	Joe	0
t	3	Clayton	1	Gumbel	0
Frank	0	Gaussian	0	t	0

4		**5**	
Copula Functions	**Frequency**	**Copula Functions**	**Frequency**
Independence	592	Independence	932
Gaussian	213	Joe	19
Frank	123	Gumbel	18
Clayton	48	Gaussian	18
Gumbel	13	Clayton	10
Joe	9	Frank	2
t	2	t	1

In this study, each resampled dataset was identified as the best-suited copula through AIC and Max-Loglikelihood. Table 8 removed all statistically insignificant independence copula (due to all parameters values being zero) from the 1000 resampled datasets and only kept the ones with statistically significant copula functions. The estimated copula parameter, θ, represents the strength of the dependence for copula models. We can see that Cluster 4 has the strongest dependency among other clusters. As observed in the P-value column from the table, all datasets with copula structures other than independence were statistically significant, indicating those datasets were dependent on each other. Each model was selected by choosing the lowest AIC score and highest Max-loglikelihood out of seven copula candidates as shown in Table 8. However, AIC does not provide information related to the quality of a model (only the quality relative to other models). It will still show the model with the best fit even though all models fit poorly to the dataset. Therefore, it is essential to validate the absolute quality of the model in multiple ways including checking the model's residuals or applying the model on a test dataset.

Table 8: Estimated Copula Parameters and Goodness of Fit Measures for Resampled Windspeed and Associated Precipitation for All Resampled Datasets

	Estimated Copula Parameter, θ	**AIC**	**P-value**	**Max-Loglik**
Cluster	**Mean (LL-UL)***	**Mean (LL-UL)***	**Mean (LL-UL)***	**Mean (LL-UL)***
1	0.98(0.92-1.04)	-5.08(-5.28-(-4.876))	0.028(0.0268-0.0298)	3.55(3.44-3.65)
2	0.89(0.82-0.98)	-6.68(-7.05-(-6.31))	0.0293(0.0277-0.031)	5.02(4.8-5.23)
3	0.78(0.72-0.85)	-4.29(-4.57-(-4.02))	0.0314(0.0297-0.0331)	3.14(3-3.28)
4	1.42(1.3-1.54)	-4.92(-5.14-(-4.7))	0.022(0.02-0.023)	3.46(3.3-3.57)
5	1.38(1.22-1.53)	-6.03(-6.57-(-5.48))	0.03(0.026-0.032)	4.0(3.75-4.3)

*LL = The lower limit of 95% confidence interval. UL = upper limit of 95% confidence interval.

5.7 Dependence Measures

Next, the dependency measures among the clusters were evaluated and shown in Table 9. Cluster 4 showed the highest dependency and association compared to the remaining clusters. This is also indicated in Figure 5-22, where the most resampled datasets were identified with non-dependence copula structures, where Gaussian and Frank copulas made up approximately 50% of the non-dependence copula in that cluster. Cluster 2 shows the strongest lower and upper tail dependence compared to other clusters, even though it has a weak dependency and below-average correlation. Interestingly, Cluster 3 showed no upper tail dependence with the score of zero for all 1000 resampled datasets, but the second-highest lower tail dependence score. This might imply the dependency is higher towards the lower values of both variables. However, the tail dependence is considered low across all clusters in comparison to Poulin et al. (2007) where datasets with adequate tail dependence showed significantly higher value.

Table 9: Dependence Measures for Each Cluster

Cluster	$\boldsymbol{Tau, \tau}$	$\boldsymbol{rho, \rho}$	$\boldsymbol{Lower\ Tail\ Dependence, \lambda_L}$	$\boldsymbol{Upper\ Tail\ Dependence, \lambda_U}$
	Mean(LL-UL)*	**Mean(LL-UL)***	**Mean(LL-UL)***	**Mean(LL-UL)***
1	0.085(0.076-0.094)	0.213(0.21-0.217)	0.023(0.016-0.029)	0.045(0.037-0.053)
2	0.071(0.063-0.08)	0.209(0.206-0.213)	0.058(0.049-0.067)	0.089(0.078-0.1)
3	0.051(0.044-0.058)	0.198(0.194-0.201)	0.047(0.038-0.056)	~0
4	0.132(0.122-0.142)	0.229(0.224-0.235)	0.033(0.026-0.041)	0.033(0.026-0.039)
5	0.024(0.018-0.03)	0.156(0.151-0.161)	0.007(0.003-0.011)	0.045(0.037-0.053)

**LL = The lower boundary of 95% confidence interval. UL = upper boundary of 95% confidence interval.

5.8 Bivariate Return Period

TORFF frequency analysis is performed using the best-fitted copula functions corresponding to each cluster. AND, and OR scenarios are investigated for resampled windspeed and the associated precipitation with univariate return periods of 10, 20, 50, 100 years for the five clusters using Equations 12 and 13, respectively. The results are then compared with the conventional approach if the two drivers have no dependency (i.e., the independence scenario). Since the "AND" and "OR" return periods explain different joint scenarios, the preference of the joint return period may change as discussed in the methodology section. The different levels of TORFF risks can be used to assess the severity of the possible tornado events and the corresponding risk of intense rainfall. The extrapolation of univariate resampled windspeed for extreme cases such as a 100-year event actually agree well with a F5 tornado event (580 km/hr vs 510 km/hr).

Figures 5-27 and 5-28 show the results corresponding to the "AND" and OR scenarios respectively. The cluster colors on the map are corresponding with the color of the graphs and each scenario is represented by different symbols. Besides, Table ST8 shows the difference in return period, according to the type of risk. For example, in Cluster 1, considering a univariate return period of 100 years, the values of mean resampled windspeed and associated precipitation is 585 km/hr and 40.5 mm/3-hr, respectively. If dependency between the resampled windspeed and associated precipitation is considered to determine the probability of both variables exceeding the same return level of 100-year event (585 km/hr and 40.5 mm/3-hr), the estimated mean joint return period is 741 years. However, the mean joint return period for the same threshold (100-year event) would be 945 years if the data are assumed to be independent. This indicates an underestimation of actual risks and impacts if we analyze the hazards in isolation.

Each cluster shows different return levels with different return periods, respectively representing the dependencies of TORFF events. As shown in Table 10, the joint probability and return period of TORFF events increases with the increase in risk factor thresholds (10 vs 100-years event for APCP and WS). Take the univariate 10-year events' mean WS and APCP as example, their corresponding return levels are determined to be 206.9 km/hr and 15.06 mm/3hr for Cluster 1. The "AND" and "OR" return periods given these thresholds are approximately 4.3 years and 56.37 years, respectively. Considering the unrealistic independence scenario, the return period is 68.4 years. The return periods in "OR" scenario reduced significantly compared to univariate return period, which indicates the higher possibility for one of the variables to exceed the threshold (4.3 years vs 10 years) compared to their analyses individually. The impact amplification of the TORFF events will be greater under the "AND" scenario compared to "OR" due to its rarity/extremity (longer return periods).

The independence assumption has resulted in different levels of underestimation of the risks for different clusters (see underestimation percentage in Table 10). Cluster 4 has the highest underestimation ranging from 26% to 31% amongst all clusters and Cluster 5 has the lowest underestimation percentage of 3%. Cluster 1 and 2 also show significant risk underestimation ranging from 13% to 22%, with Cluster 3 on the lower side of risk underestimation at 9%. Higher percentage of risk underestimation indicates higher level of dependency between the TORFF events.

Table 10: Comparison of Return Periods Under Different Scenarios in Each Cluster

Cluster	Return Period, T	Resampled Windspeed, WS (km/hr)	Associated Precipitation, APCP (mm/3hr)	AND Return Period (Year)	Independent Return Period (Year)	Underestimation %	OR Return Period
		Mean(LL-UL)*					
1	10	206.9(206.2-207.7)	15.06	56.37(54.99-57.76)	68.4(67.7-69.1)	18%	4.3(4.3-4.3)
	20	278.8(276.4-281.1)	21.24	217.8(211.2-224.446)	273.1(269.7-276.6)	20%	8.6(8.5-8.6)
	50	421(413.6-428.3)	31.29	556.9(538-575.9)	708.2(697.1-719.3)	21%	12.91(12.8-13)
	100	585(569.1-600.9)	40.5	741.3(715.4-77)	945.2(929.7-960.7)	22%	14.9(14.7-15)
2	10	184(183.4-184.7)	14.57	37.39(36.5-38.2)	42.9(42.4-43.4)	13%	3.1(3.1-3.1)
	20	220.3(218.867-221.9)	19.65	66.2(64.6-67.8)	76.9(76-77.8)	14%	3.7(3.7-3.8)
	50	277.3(273.5-281.2)	26.48	91.8(89.4-94.1)	107.4(105.9-108.8)	15%	4.1(4-4.1)
	100	330.2(323.3-337.2)	31.72	102.1(99.3-104.9)	119.5(117.6-121.5)	15%	4.19(4.1-4.2)
3	10	191.8(191.0-192.6)	27.32	47.84(47.-48.6)	51.91(51.3-52.4)	8%	3.544(3.5-3.5)
	20	240.2(237.8-242.6)	47.86	77.35(76-78.6)	84.76(83.7-85.7)	9%	4.291(4.3-4.304)
	50	325.5(316.8-334.2)	89.95	100.7(98.9-102.6)	111(109.4-112.6)	9%	4.711(4.6-4.73)

	100	415.9(393-438.8)	136.9	109(106.9-111.2)	120.3(118.4-122.3)	9%	4.82(4.8-4.84)
	10	174.9(174.3-175.5)	14.64	72.19(70.1-74.29)	97.71(96.62-98.82)	26%	5.227(5.2-5.25)
	20	236.9(235.1-238.8)	18.4	263.9(254.7-273.2)	374(368.87-379.16)	29%	9.39(9.33-9.46)
	50	366.5(360.5-372.6)	23.19	626.3(601.58-651.2)	903.6(887.13-920.2)	31%	14.79(14.65-14.94)
4	100	523.7(510.3-537.1)	26.71	785.1(752.25-818.01)	1137(1113-1161.36)	31%	16.7(16.52-16.89)
	10	145.7(145.24-146.27)	11.57	68.5(67.53-69.48)	70.4(69.62-71.19)	3%	4.2 (4.19-4.2)
	20	167.8(166.5-169.2)	16.47	100.1(98.22-102.16)	102.9(101.15-104.67)	3%	4.77 (4.75-4.79)
	50	202.2(198.3-206.14)	23.26	131.8(128.4-135.3)	135.4(132.18-138.69)	3%	5.11(5.1-5.13)
5	100	235(226.68-243.36)	28.54	143.9(139.61-148.34)	147.8(143.64-152.02)	3%	5.22(5.19-5.24)

*LL = The lower limit of 95% confidence interval. UL = upper boundary of 95% confidence interval.

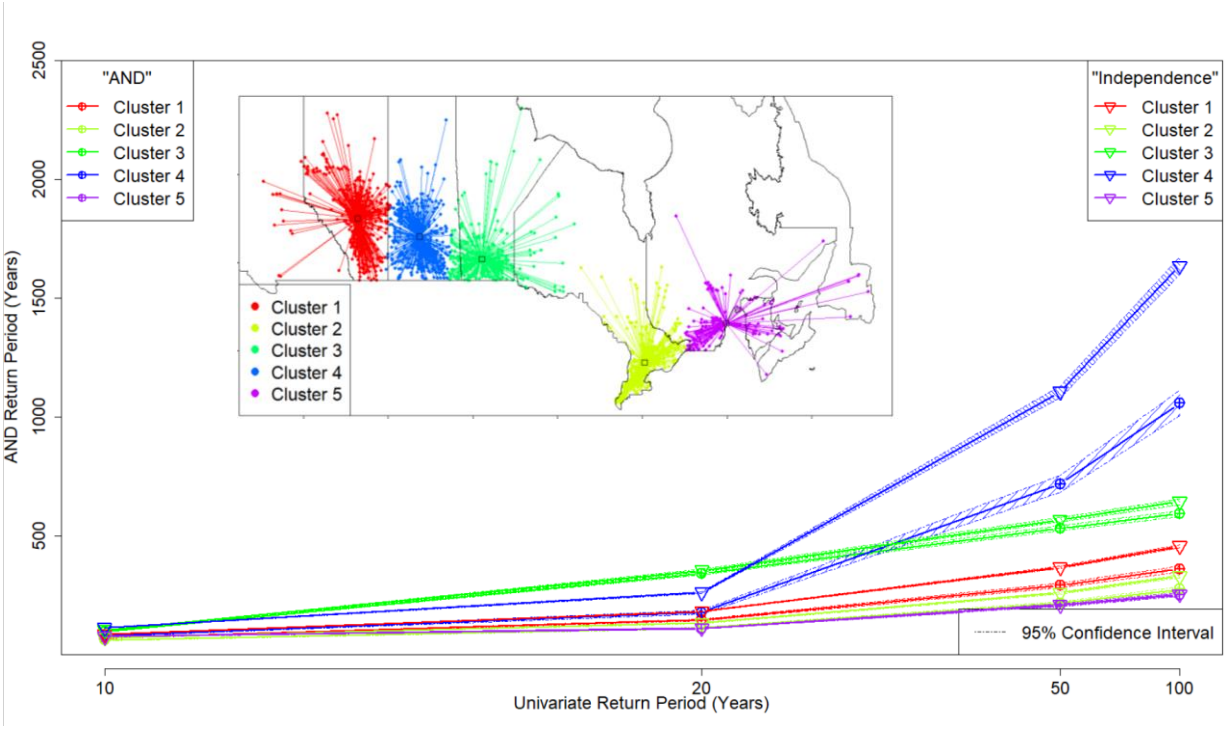


Figure 5-25: Comparison between the AND joint return period (Table 10 column 5) and independence scenario (column 6). X-axis represents the univariate return periods of which the return levels are presented in Table ST8 for resampled windspeed, WS (column 3), and associated precipitation, APCP (column 4), respectively. The dashed line for each cluster represents the 95% confidence interval of mean values from 1000 resampled datasets.

Figure 5-25 above shows the comparison chart between AND scenario with joint bivariate return periods compared to a scenario without dependency. From the chart, it can be noticed that scenarios without dependencies contain higher return periods for all clusters compared to AND scenarios. Out of all clusters, Cluster 4 showed the highest difference when compared to the two scenarios. This could indicate Cluster 4 has higher dependency and more sensitive when dependency is involved.

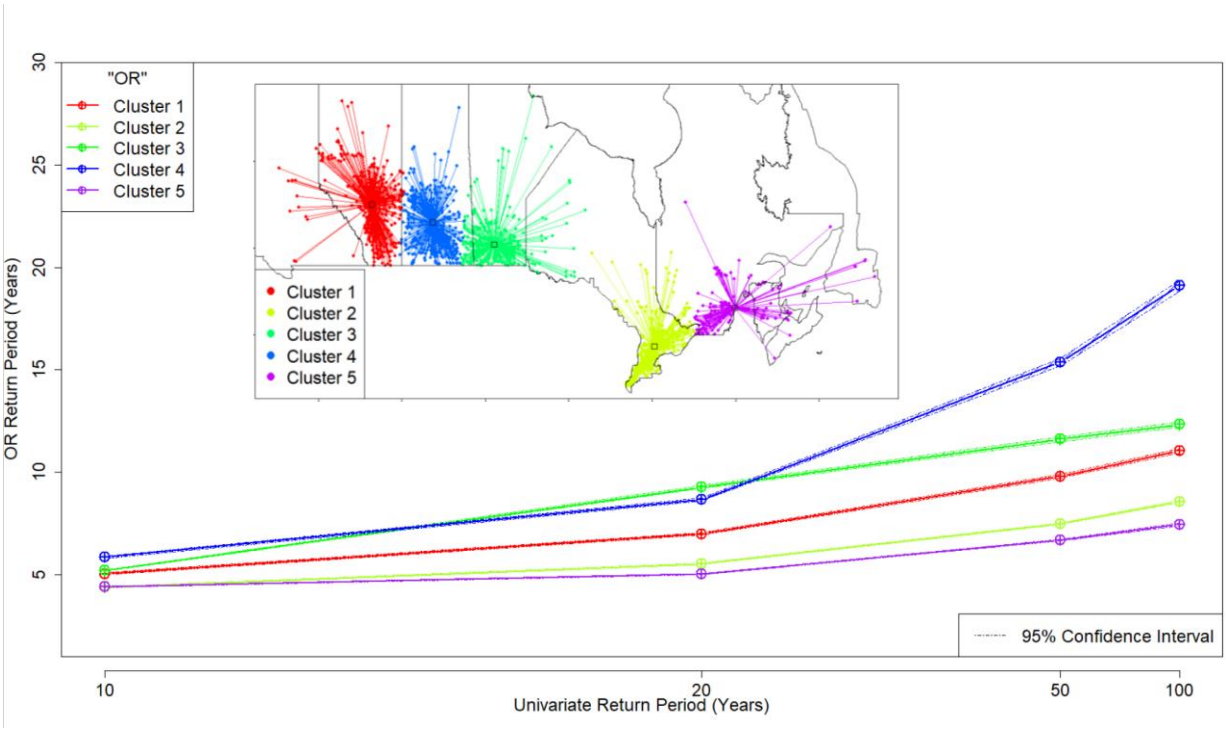


Figure 5-26: OR joint return period (values are shown in Table 10 column 8). Top left legends show the joint return periods with dependency. X-axis represents the univariate return periods of which the return levels are presented in Table ST8 for resampled windspeed, WS (column 3), and associated precipitation, APCP (column 4), respectively. The dashed line for each cluster represents the 95% confidence interval of mean values from 1000 resampled datasets.

Figure 5-26 shows the joint return periods for OR scenario where either of the events exceed the threshold of set probability. As mentioned previously, the more holistic OR scenario will lead to a lower return period. As we can see from Figure 5-26, Cluster 4's OR joint return period has increased significantly after we extended the univariate return period to 100 years.

5.9 Conditional Return Periods

The conditional return periods of the associated precipitation given resampled windspeed have been calculated using equations 7 & 8 and are shown in Figures 5-27 to 5-31. Each value represents the mean conditional return level of 1000 resampled data. The bivariate conditional return periods behave differently for each cluster, based on the predefined probability for resampled windspeed. For example, Cluster 5 shows less variation between bivariate, and independence and univariate compared to Cluster 1. Interestingly, as the predefined resampled windspeed probability became more extreme (moving towards 100-year events), the difference between the scenarios reduces. In addition, as we can see for all clusters, the bivariate scenario (considering the dependency) shows a higher return level compared to the independence and univariate scenario.

Knowing the conditional probability for associated windspeed on resampled windspeed is essential in knowing the possible intensity of the precipitation given the windspeed when a tornado event occurs. A bigger difference in return levels for 10-year resampled windspeed suggests higher dependency between associated precipitation and lower wind speed.

Additionally, the bivariate hazard scenarios also show narrow 95% confidence intervals. One of the reasons could be the number of independence copula that was determined during the model selection phase. Same return levels were generated from the independence copula that was determined in the model selection phase. With 60-90% of the same values could affect the values from the resampled dataset that has dependencies. The calculated values, in the end, could end up being less than just considering resampled data with dependency.

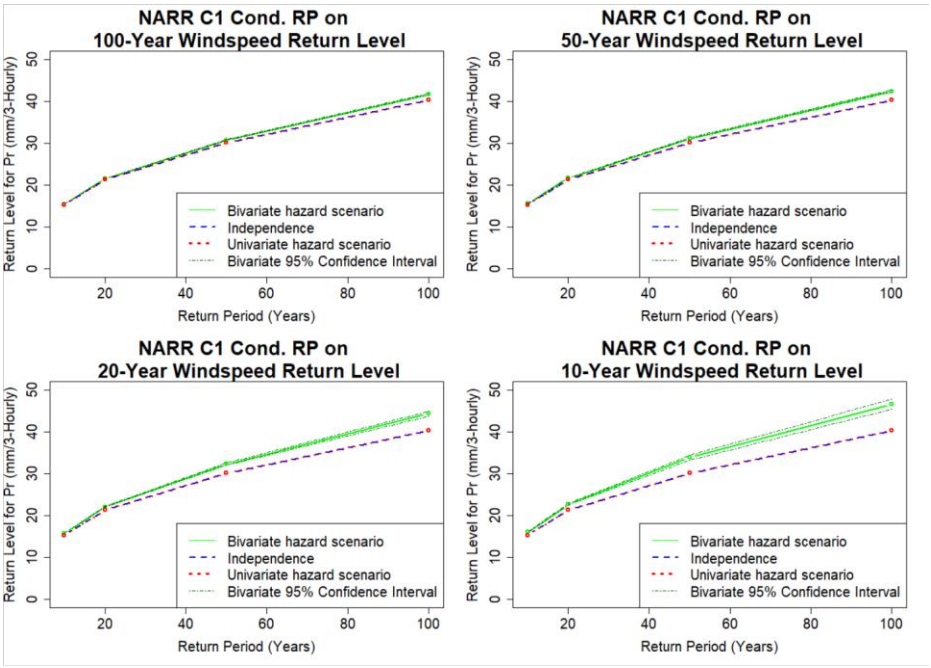


Figure 5-27: Associated precipitation condition return level on different resampled windspeed for Cluster 1. The green line shows the bivariate return level of associated precipitation given resampled windspeed at different return period, while blue and red dashed lines represent univariate and independence return periods.

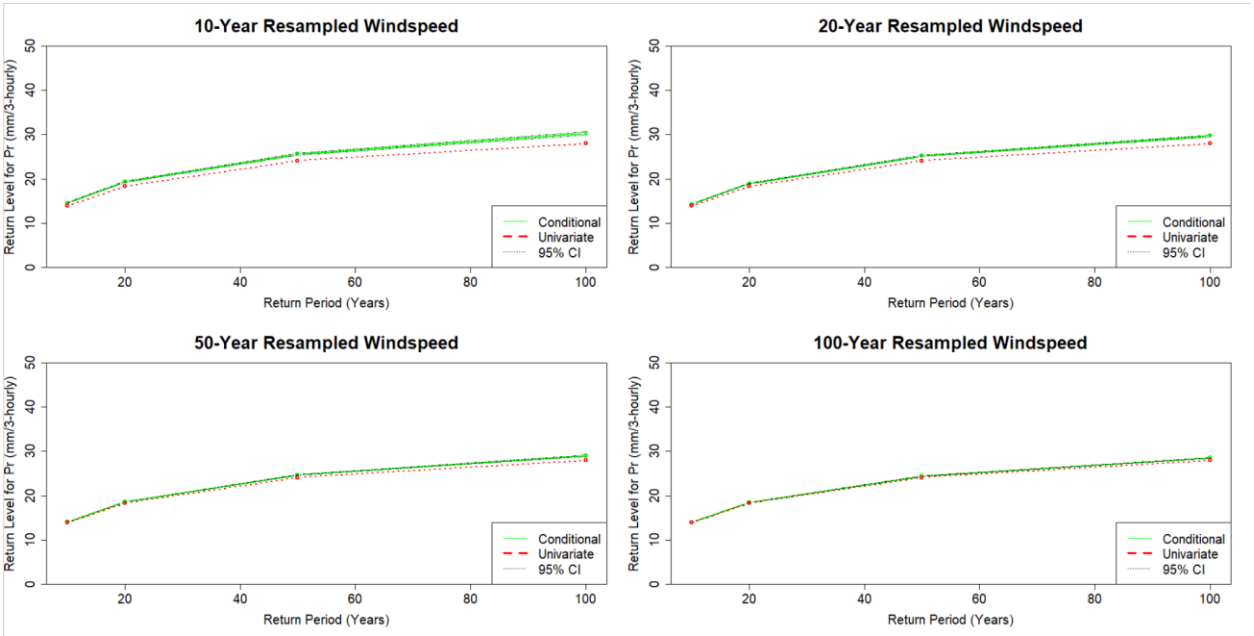


Figure 5-28: Associated precipitation condition return level on different resampled windspeed for Cluster 2. The green line shows the bivariate return level of associated precipitation given resampled windspeed at different return period, while blue and red dashed lines represent univariate and independence return periods.

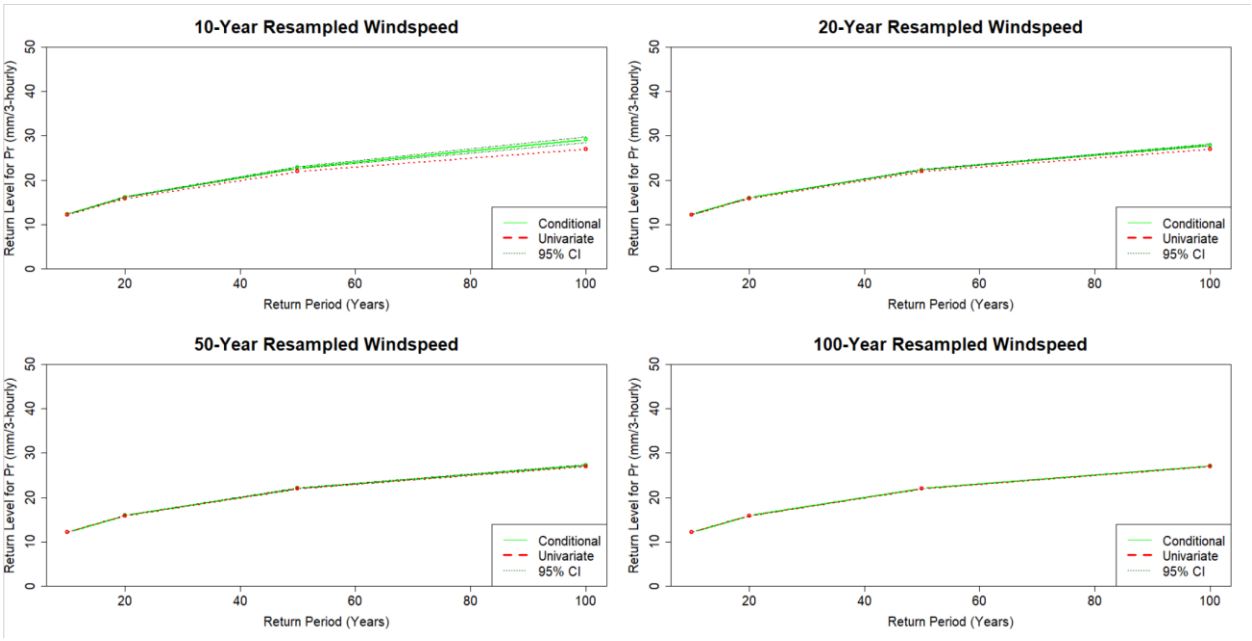


Figure 5-29: Associated precipitation condition return level on different resampled windspeed for Cluster 3. The green line shows the bivariate return level of associated precipitation given resampled windspeed at different return period, while blue and red dashed lines represent univariate and independence return periods.

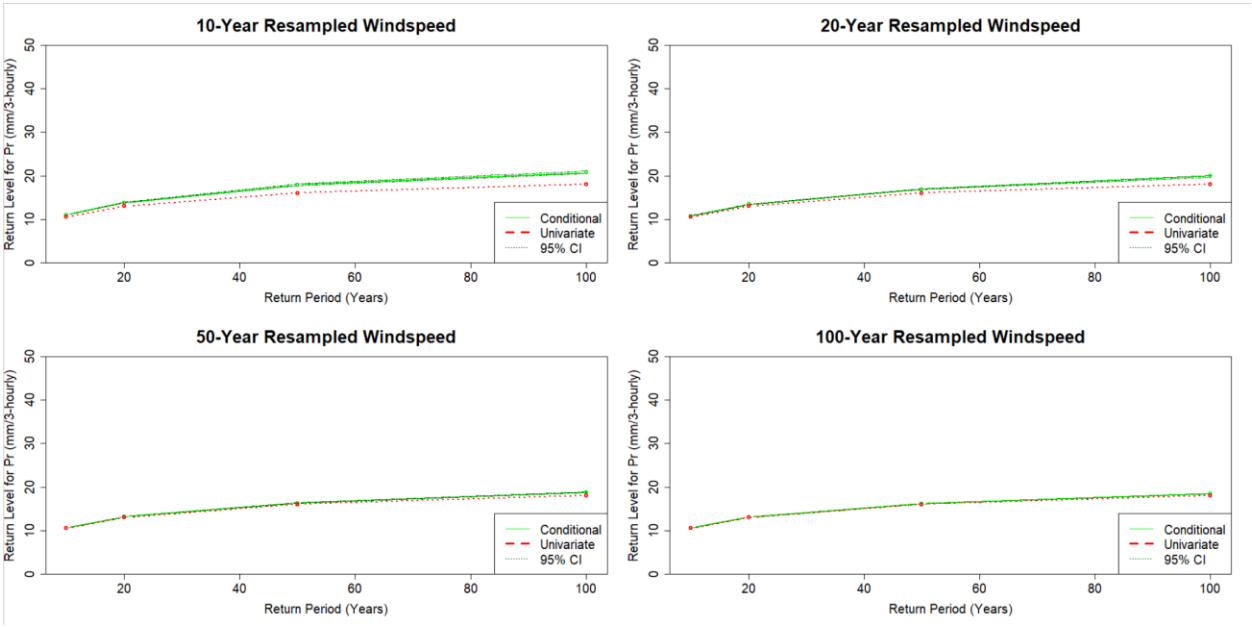


Figure 5-30: Associated precipitation condition return level on different resampled windspeed for Cluster 4. The green line shows the bivariate return level of associated precipitation given resampled windspeed at different return periods, while blue and red dashed lines represent univariate and independence return periods.

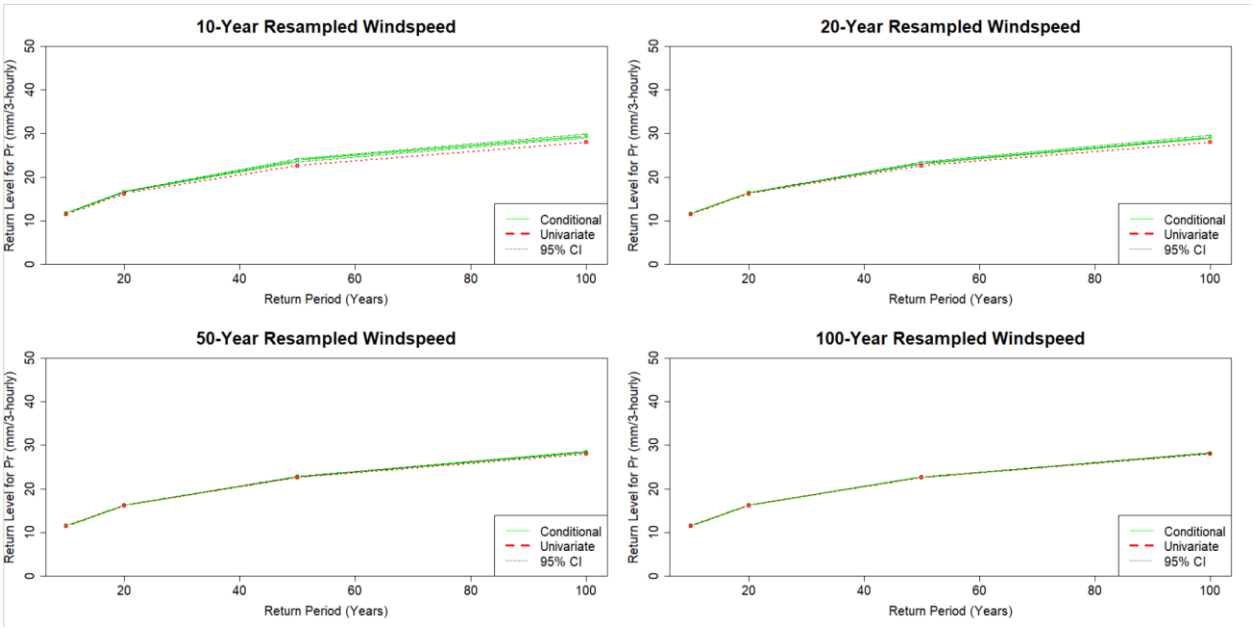


Figure 5-31: Associated precipitation condition return level on different resampled windspeed for Cluster 5. The green line shows the bivariate return level of associated precipitation given resampled windspeed at different return period, while blue and red dashed lines represent univariate and independence return periods.

Table 11: Conditional bivariate return levels with different fixed resampled windspeed return periods for all clusters.

Conditional Bivariate Return Level (RL) with Fixed WS Return Period (RP)							
Resampled Windspeed Return Period							
100		**50**		**20**		**10**	
Cluster 1							
APCP Return Period (Years)	**Mean (mm) (LL-UL)***	**APCP Return Period (Years)**	**Mean (mm) (LL-UL)***	**APCP Return Period (Years)**	**Mean (mm) (LL-UL)***	**APCP Return Period (Years)**	**Mean (mm) (LL-UL)***
100	41.60(41.42 -41.78)	100	42.38(42.1-42.67)	100	44.32(43.7-44.9)	100	46.58(45.45 -47.7)
50	30.67(30.59 -30.75)	50	31.12(30.98 -31.26)	50	32.28(31.98 -32.58)	50	33.81(33.24 -34.39)
20	21.52(21.49 -21.54)	20	21.71(21.66 -21.76)	20	22.11(22.02 -22.219)	20	22.67(22.5-22.83)
10	15.46(15.45 -15.48)	10	15.55(15.53 -15.57)	10	15.77(15.72 -15.81)	10	16.08(15.99 -16.16)
Cluster 2							
APCP Return Period (Years)	**Mean (mm) (LL-UL)***	**APCP Return Period (Years)**	**Mean (mm) (LL-UL)***	**APCP Return Period (Years)**	**Mean (mm) (LL-UL)***	**APCP Return Period (Years)**	**Mean (mm) (LL-UL)***
100	28.57(28.49 -28.66)	100	28.99(28.86 -29.13)	100	29.7(29.46-29.94)	100	30.28(29.95 -30.61)
50	24.44(24.40 -24.48)	50	24.67(24.6-24.74)	50	25.17(25.03 -25.31)	50	25.62(25.41 -25.83)
20	18.48(18.46 -18.50)	20	18.64(18.59 -18.68)	20	18.95(18.87 -19.04)	20	19.33(19.2-19.46)
10	13.98(13.97 -13.99)	10	14.06(14.04 -14.08)	10	14.26(14.21 -14.30)	10	14.55(14.47 -14.63)
Cluster 3							
Return Period (Years)	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***
100	27.09(27.05 -27.12)	100	27.26(27.18 -27.35)	100	27.91(27.64 -28.18)	100	29.13(28.45 -29.81)
50	22.01(21.99 -22.03)	50	22.09(22.06 -22.13)	50	22.32(22.22 -22.41)	50	22.82(22.59 -23.06)

20	15.91(15.90 -15.92)	20	15.94(15.93 -15.95)	20	16.03(16- 16.06)	20	16.20(16.14 -16.27)
10	12.20(12.19 -12.20)	10	12.21(12.21 -12.22)	10	12.25(12.24 -12.27)	10	12.32(12.3- 12.35)
Cluster 4							
Return Period (Years)	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***
100	18.46(18.42 -18.51)	100	18.78(18.7- 18.85)	100	29.2(28.86- 29.54)	100	20.78(20.48 -21.08)
50	16.19(16.17 -16.2)	50	16.34(16.31 -16.37)	50	23.22(23.05 -23.4)	50	17.93(17.72 -18.14)
20	13.11(13.1- 13.12)	20	13.2(13.18- 13.21)	20	16.42(16.36 -16.48)	20	13.82(13.75 -13.89)
10	10.58(10.57 -10.59)	10	10.63(10.62 -10.65)	10	11.62(11.59 -11.65)	10	11.01(10.97 -11.05)
Cluster 5							
Return Period (Years)	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***	**Return Period (Years)**	**Mean (mm) (LL-UL)***
100	28.21(28.14 -28.29)	100	28.49(28.35 -28.64)	100	29.2(28.86- 29.54)	100	29.44(29.05 -29.84)
50	22.70(22.66 -22.73)	50	22.82(22.75 -22.89)	50	23.22(23.05 -23.4)	50	23.89(23.53 -24.25)
20	16.24(16.23 -16.26)	20	16.29(16.26 -16.32)	20	16.42(16.36 -16.48)	20	16.65(16.53 -16.77)
10	11.53(11.53 -11.54)	10	11.55(11.54 -11.57)	10	11.62(11.59 -11.65)	10	11.73(11.67 -11.79)

*LL = The lower boundary of 95% confidence interval. UL =

Chapter 6 Conclusions and Future Works

6.1 Conclusions

In this study, historic tornado events in Canada from 1980-2009 were studied together with the associated precipitation (that might cause flash floods) to investigate the corresponding dependencies. The purpose is to provide insight to emergency response authorities regarding evacuation procedures when tornado and flash flood (TORFF) events happen simultaneously for their longer-term emergency planning rather than immediate events. TORFF events demand different evacuation procedures, and they affect the community's safety at risk if chosen incorrectly. Flash floods data (represented by the intensity of associated precipitation) were obtained from reanalysis data, NARR at its nearest grid cell to post-tornado event and paired with the corresponding tornado event. Since the actual tornado windspeed is not available, an alternative resampling method was proposed by randomly resampling windspeed from the tornado's corresponding Fujita rating windspeed interval for each tornado event. This method generated 1000 sets of unique wind speed datasets paired with associated precipitation. Wind resampling from the Fujita rating is expected to represent the corresponding wind speed, considering the associated uncertainties.

The generated dataset is then separated into multiple clusters using affinity propagation (AP) based on longitude and latitude coordinates (geographic locations). The resampled windspeed and associated precipitation dataset were then processed followed by copula analysis considering the annual maximum series. Each year, the highest resampled windspeed was selected and paired with its corresponding precipitation in each cluster to form one thousand continuous observations (30 annual maximum records) for each cluster. First, dependency investigation was conducted for all resampled datasets for each cluster and appropriate marginal distributions were fitted

accordingly. Then, a copula model was identified for each resampled dataset by the AIC selection criterion. A joint probability distribution was then generated for each resampled dataset in each cluster by using the copula function and its corresponding marginal probability distribution functions. Besides, dependency measures such as Kendall's tau and Spearman's rho were also computed to quantify the interrelationships.

There is evidence for positive dependencies between the tornadoes and flash floods, especially in Saskatchewan region (Cluster 4), which demand the emergency response authorities to pay more attention when there is a possibility of a TORFF event happening to provide a more appropriate evacuation procedure accordingly. Saskatchewan region shows higher dependency than other regions, despite having lower precipitation rates than other clusters. Dependency measures such as Kendall's tau and Spearman's rho were also computed to show that Saskatchewan region has higher dependencies. Bivariate joint return periods were computed and compared under different scenarios: AND, OR, conditional, and the unrealistic independence assumption. The results showed that different clusters behave differently due to spatial variability and their climate. For example, the univariate 100-year return period for precipitation in Manitoba region (Cluster 3) is estimated to be 137 mm/3-hr. In comparison, precipitation for the same return period for Saskatchewan region is estimated to be 27 mm/3-hr, which is almost five times the amount of precipitation for the same return period. While comparing with a scenario without dependency, the estimated joint return periods were higher for all clusters. Saskatchewan region showed the most significant difference in estimated joint return periods compared to a scenario without dependency. This indicates Saskatchewan region has a higher dependency compared to other clusters. As expected, the joint return periods under the OR scenario are lower than the univariate estimates. Higher precipitation rates are expected during extreme wind speed,

according to the conditional assessment of precipitation given wind speed. The return levels for associated precipitation are higher when compared to univariate and independent scenarios. The statement remains valid for all clusters in conditional assessment even though the magnitude of return levels varies between each cluster. For instance, 100-year conditional precipitation return levels for 100-year wind speed in the Saskatchewan region are lower than the Alberta region (41mm vs. 18mm).

The different levels of TORFF risks can be used to assess the severity of the possible tornado event and the risk of intense rainfall. For example, the Saskatchewan region has lower precipitation and relatively high wind speed than other clusters. Therefore, the emergency response authority should focus more on tornado events and provide evacuation accordingly so the community at risk could retreat to safety. On the other hand, the Manitoba region showed significantly higher associated precipitation for the same 100-year threshold at approximately 137 mm/3-hr compared to the Saskatchewan region at 26.71 mm/3-hr, which indicates the emergency response authority in the Manitoba region should focus on improving their flashfloods evacuation, and mitigation plan since the return level for precipitation is significantly higher.

However, the occurrences of flash flood are not limited to just intensity of precipitation. One of the main uncertainties for this study could come from no direct wind speed measurement from the tornadoes. Hence, the windspeed used in this study was resampled from the recorded Fujita rating corresponding to each tornado event. In addition, most determined copula models were considered independent, which may lead to underestimating actual return levels that are genuinely associated with dependency. In addition, limited observation points (30 values) per dataset add more uncertainties when predicting more extensive return periods such as from 100

years onwards. Furthermore, the resolution of NARR could also contribute to the uncertainty in this study. Its spatial resolution of 32 km x 32 km might not capture precipitation reliably if the location of the tornado event is too close to the edge of the grid cell.

This study provides insight for more realistic recurrence interval estimation for tornadoes and flash floods to aid in the evacuation decision-making process.

6.2 Recommendations for Future Works

Future studies are recommended to extend the analyses:

- To assess TORFF events, we considered the resampled windspeed annual maxima and its corresponding associated precipitation to represent actual tornado surface windspeed in this study. Datasets generated using a resampling method may not fully represent the actual windspeed which is a source of uncertainty when fitting marginal distribution. Even though best fitted marginal distributions were identified, they do not necessarily mean the "true" distribution. Resampled datasets with fat tails might overestimate return levels for resampled windspeed and associated precipitation. This results in a conservative assessment of TORFF risks.
- The clustering method in this analysis were done based on geographic location (longitude and latitude) without considering any environmental variables. This is due to the high number of events happening on the same day across Canada. Nielsen et al. (2015) conducted the clustering based on the events that happened on the same day within specific spatial window. Then, the tornado events were clustered based on local development time. In case of Canada, including more environmental variables might provide a more accurate clustering results due to more tornado environment information

being considered. For example, including tornado environment variables such as dewpoint temperature, convective available potential energy (CAPE), and vertical shear velocity in better clustering alternatives such as K-mean clustering or principal component analysis (PCA) could generate more representative results. This way the physical characteristics of tornadoes can be derived from statistical analysis. Characterization of tornado climatology (e.g., looking at actual events in the NARR dataset. For example, conduct further analysis on 2005 Southern Ontario TORFF events) in Canada can set a foundation for TORFF events studies in the future.

- Even though Saskatchewan region shows the highest dependency among other clusters, however, the dependencies are purely based on two major TORFF variables and their geographic locations. The physical drivers that caused the high dependency should be look further into to fully understand the interaction between them.
- Climate model simulations can be applied to characterize the physical behavior of TORFF events. Good simulation results can be used to validate developed TORFF predictive models to assess their accuracy. Simulation results can be improved by considering physical phenomenon and their environment variables as model inputs. In addition, a complete historic hazards database (e.g., GIS shapefiles of events or hazard warnings) such as flash flood events are helpful to compare and capture the actual TORFF events in Canada.
- Future analyses are encouraged to analyze the dependencies between different tornado environment variables that result in TORFF events in our model better to characterize the risks of the compound events in Canada. The developed predictive model can be validated using accurate historical databases.

- TORFF associated precipitation was assumed to be representative when extracted from nearest NARR grid cell coordinates. Reanalysis products with higher resolution can improve the datasets generated for the study.
- Future analyses can consider other reanalysis products to reduce the uncertainty in pre-tornado environments and improve dry biases to represent the associated precipitation better. For example, ERA-5 with higher spatial and temporal resolution (1-hourly with 0.1-degree x 0.1-degree compared to 3-hourly and 0.3-degree x 0.3-degree from NARR) with dry bias correction might provide a more representative precipitation values compared to NARR (Tarek et al. 2020).

Bibliography

Asmis, G.J.K. 1980. *The tornado risk in eastern Canada.* Conference Paper, Montreal, QC: Atomic Energy Control Board.

Bouwer, Laurens M. 2011. "Have Disaster Losses Increased Due to Anthropogenic Climate Chnage." *American Meteorological Society* 39-46.

Burnham, Kenneth P, and David R Anderson. 2004. "Model Selection and Multimodel Inference." *A Practical Information-Theoretic Approach, Second Edition* 63.

Environment Canada. 2017. *Canadian National Tornado Database: Verified Events (1980-2009) - Public.* 05 17. Accessed 12 18, 2020. https://open.canada.ca/data/en/dataset/f314a39f-009d-430b-97b9-d6e0cae22340.

Environmental Protection Agency. 2018. *Hurricane Harvey: The State and Federal Response.* June. Accessed December 18, 2020. https://www.epa.gov/sites/production/files/2018-06/documents/hurricane_harvey_the_state_and_federal_response_tceq_0.pdf.

European Commission. 2018. "Short-term Outlook for EU Agricultural Markets in 2018 and 2019. Report 22." Report.

Faust, Eberhard, and Maximilian Strobl. 2018. *Heatwaves, drought and forest fires in Europe: Billions of dollars in losses for agricultural sector.* 07 31. Accessed 11 29, 2021. https://www.munichre.com/topics-online/en/climate-change-and-natural-disasters/climate-change/heatwaves-and-drought-in-europe.html.

Feuerstein, Jacob. 2021. "PARTS OF BUCKS COUNTY ARE IN BOTH A TORNADO EMERGENCY AND FLASH FLOOD EMERGENCY. THIS IS AN EXTREMELY RARE WORST CASE SCENARIO." *Twitter.* September 1. Accessed December 20, 2021. https://twitter.com/Jacob_Feuer/status/1433205056443138050.

Gabor Melli's Research Knowledge Base. 2021. *Extreme Value Analysis Task.* June 17. Accessed December 22, 2021. http://www.gabormelli.com/RKB/Extreme_Value_Analysis_Task.

Guyer, Jared, and Rick Ewald. 2004. *RECORD HAIL EVENT – EXAMINATION OF THE AURORA, NEBRASKA SUPERCELL OF 22 JUNE 2003.* Conference Paper, Hyannis, MA: 22nd Conf. on Severe Local Storms.

King, Austin T, and Aaron D Kennedy. 2019. "North American Supercell Environments in Atmospheric Reanalyses and RUC-2." *Journal of Applied Meteorology and Climatology* 71-92.

Mattyasovszky , Miklos. 2020. *The Largest Countries In The World.* August 14. Accessed September 20, 2021. https://www.worldatlas.com/articles/the-largest-countries-in-the-world-the-biggest-nations-as-determined-by-total-land-area.html.

McKay, G. A., and A. B. Lowe. 1960. "“The Tornado in Western Canada.”." *Bulletin of the American Meteorological Society 41, no. 1* 1–8.

Najafi, Mohammad Reza, and Harsimrenjit Singh. 2020. "Evaluation of gridded climate datasets over Canada using univariate and bivariate approaches: Implications for hydrological modelling." *Journal of Hydrology* 584.

National Oceanic and Atmospheric Administration. 2017. *National Hurricane Center Tropical Cyclone REport.* Post Natural Disaster Report, Eric S. Blake and David A. Zelinsky National Hurricane Center.

National Post. 2021. *Some of Canada's worst tornadoes over the years.* July 15. Accessed December 14, 2021. https://nationalpost.com/pmn/news-pmn/canada-news-pmn/some-of-canadas-worst-tornadoes-over-the-years.

NIELSEN, ERIK R., GREGORY R. HERMAN, ROBERT C. TOURNAY, JOHN M. PETERS, and AND RUSS S. SCHUMACHER. 2015. "Double Impact: When Both Tornadoes and Flash Floods Threaten the Same Place." *American Meteorological Society* Volume 30.

NOAA. 2003. "Contents of NARR output AWIPS GRIB files." *NORTH AMERICAN REGIONAL REANALYSIS.* December 31. Accessed November 28, 2021. https://www.emc.ncep.noaa.gov/mmb/rreanl/narr_archive_contents.pdf.

—. 2009. *Why are there more thunderstorms during the summer?* May 2001. Accessed December 11, 2021. https://science.howstuffworks.com/nature/climate-weather/storms/question646.htm.

NWS Birmingham, Alabama. 1981. *Fujita Scale.* Accessed December 20, 2021. https://www.weather.gov/bmx/fujitascale.

Public Safety Canada. 2019. *Hazards and Emergencies.* 10 30. Accessed 09 16, 2021. https://www.getprepared.gc.ca/cnt/hzd/flds-en.aspx#a1.

—. 2019. *The Canadian Disaster Database.* 09 19. Accessed December 18, 2020. https://www.publicsafety.gc.ca/cnt/rsrcs/cndn-dsstr-dtbs/index-en.aspx.

Sills, David, Vincent Cheng, Patrick McCarthy, Brad Rousseaul, James Waller, Lesley Elliott, Joan Klaassen, and and Heather Auld. 2012. "USING TORNADO, LIGHTNING AND POPULATION DATA TO IDENTIFY TORNADO PRONE AREAS IN CANADA ." *Amer. Meteorol. Soc* 59.

Singh, Harsimrenjit. 2019. *Spatial and Temporal Characteristics of Temperature-Precipitation Covariability across Canada in a Changing Climate.* Masters Thesis, London: Western University.

Statistics How To. 2021. *Multivariate Distribution: Definition.* November 11. Accessed December 13, 2021. https://www.statisticshowto.com/multivariate-distribution/.

Zhang, Qinghe. 2013. "Joint probability distribution of winds and waves from wave simulation of 20 years (1989-2008) in Bohai Bay." *Water Science and Engineering 6(3)* 296-307.

Abeysirigunawardena, Dilumie S., Eric Gilleland, David Bronaugh, and Pat Wong. 2009. "Extreme Wind Regime Responses to Climate Variability and Change in the Inner South Coast of British Columbia, Canada." *Atmosphere-Ocean* 47 (1): 41–62. https://doi.org/10.3137/AO1003.2009.

AghaKouchak, Amir, Linyin Cheng, Omid Mazdiyasni, and Alireza Farahmand. 2014. "Global Warming and Changes in Risk of Concurrent Climate Extremes: Insights from the 2014 California Drought: Global Warming and Concurrent Extremes." *Geophysical Research Letters* 41 (24): 8847–52. https://doi.org/10.1002/2014GL062308.

Akaike, H. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* 19 (6): 716–23. https://doi.org/10.1109/TAC.1974.1100705.

Azam, Muhammad, Seung Maeng, Hyung Kim, and Ardasher Murtazaev. 2018. "Copula-Based Stochastic Simulation for Regional Drought Risk Assessment in South Korea." *Water* 10 (4): 359. https://doi.org/10.3390/w10040359.

Back, Larissa E., and Christopher S. Bretherton. 2005. "The Relationship between Wind Speed and Precipitation in the Pacific ITCZ." *Journal of Climate* 18 (20): 4317–28. https://doi.org/10.1175/JCLI3519.1.

Barcikowska, Monika J., Scott J. Weaver, Frauke Feser, Simone Russo, Frederik Schenk, Dáithí A. Stone, Michael F. Wehner, and Matthias Zahn. 2018. "Euro-Atlantic Winter Storminess and Precipitation Extremes under 1.5 °C vs. 2 °C Warming Scenarios." *Earth System Dynamics* 9 (2): 679–99. https://doi.org/10.5194/esd-9-679-2018.

"Barrie-and-Area Tornadoes Insured Damage Increases to $100 Million." n.d. Accessed April 22, 2022. http://www.ibc.ca/on/resources/media-centre/media-releases/barrie-and-area-tornadoes-insured-damage-increases-to-100-million.

Batchabani, Essoyeke, Elsa Sormain, and Musandji Fuamba. 2016. "Potential Impacts of Projected Climate Change on Flooding in the Riviere Des Prairies Basin, Quebec, Canada: One-Dimensional and Two-Dimensional Simulation-Based Approach." *Journal of Hydrologic Engineering* 21 (12): 05016032. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001461.

Becker, Emily J., Ernesto Hugo Berbery, and R. Wayne Higgins. 2009. "Understanding the Characteristics of Daily Precipitation over the United States Using the North American Regional Reanalysis." *Journal of Climate* 22 (23): 6268–86. https://doi.org/10.1175/2009JCLI2838.1.

Brunner, Manuela Irene, Jan Seibert, and Anne-Catherine Favre. 2016. "Bivariate Return Periods and Their Importance for Flood Peak and Volume Estimation." *WIREs Water* 3 (6): 819–33. https://doi.org/10.1002/wat2.1173.

Bukovsky, Melissa S., and David J. Karoly. 2007. “A Brief Evaluation of Precipitation from the North American Regional Reanalysis.” *Journal of Hydrometeorology* 8 (4): 837–46. https://doi.org/10.1175/JHM595.1.

Bunkers, Matthew J., and Charles A. Doswell. 2016. “Comments on ‘Double Impact: When Both Tornadoes and Flash Floods Threaten the Same Place at the Same Time.’” *Weather and Forecasting* 31 (5): 1715–21. https://doi.org/10.1175/WAF-D-16-0116.1.

“Canadian National Tornado Database: Verified Events (1980-2009) - Public - Open Government Portal.” n.d. Accessed April 21, 2022. https://open.canada.ca/data/en/dataset/f314a39f-009d-430b-97b9-d6e0cae22340.

Carta, José A., Penélope Ramírez, and Celia Bueno. 2008. “A Joint Probability Density Function of Wind Speed and Direction for Wind Energy Analysis.” *Energy Conversion and Management* 49 (6): 1309–20. https://doi.org/10.1016/j.enconman.2008.01.010.

Cheng, Vincent Y.S., George B. Arhonditsis, David M.L. Sills, William A. Gough, and Heather Auld. 2015. “A Bayesian Modelling Framework for Tornado Occurrences in North America.” *Nature Communications* 6 (1): 6599. https://doi.org/10.1038/ncomms7599.

Choi, Woonsup, Sung Joon Kim, Peter F Rasmussen, and Adam R Moore. 2009. “Use of the North American Regional Reanalysis for Hydrological Modelling in Manitoba.” *Canadian Water Resources Journal* 34 (1): 17–36. https://doi.org/10.4296/cwrj3401017.

Coles, Stuart. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London: Springer London. https://doi.org/10.1007/978-1-4471-3675-0.

De Paola, Francesco, Maurizio Giugni, Francesco Pugliese, Antonio Annis, and Fernando Nardi. 2018. “GEV Parameter Estimation and Stationary vs. Non-Stationary Analysis of Extreme Rainfall in African Test Cities.” *Hydrology* 5 (2): 28. https://doi.org/10.3390/hydrology5020028.

Dong, Lu, L. Ruby Leung, Jian Lu, and Yang Gao. 2019. “Contributions of Extreme and Non-Extreme Precipitation to California Precipitation Seasonality Changes Under Warming.” *Geophysical Research Letters* 46 (22): 13470–78. https://doi.org/10.1029/2019GL084225.

Dupilka, Max L., and Gerhard W. Reuter. 2011. “AN EXAMINATION OF THREE SEVERE CONVECTIVE STORMS THAT PRODUCED SIGNIFICANT TORNADOES IN CENTRAL ALBERTA.” In .

Favre, Anne-Catherine, Salaheddine El Adlouni, Luc Perreault, Nathalie Thiémonge, and Bernard Bobée. 2004. “Multivariate Hydrological Frequency Analysis Using Copulas: MULTIVARIATE FREQUENCY ANALYSIS USING COPULAS.” *Water Resources Research* 40 (1). https://doi.org/10.1029/2003WR002456.

Feng, Boyu, Ying Zhang, and Robin Bourke. 2021. “Urbanization Impacts on Flood Risks Based on Urban Growth Data and Coupled Flood Models.” *Natural Hazards* 106 (1): 613–27. https://doi.org/10.1007/s11069-020-04480-0.

Fisher, N. I., and P. Switzer. 1985. “Chi-Plots for Assessing Dependence.” *Biometrika* 72 (2): 253–65. https://doi.org/10.1093/biomet/72.2.253.

Gagne, David John, Amy McGovern, Jeffrey B. Basara, and Rodger A. Brown. 2012. “Tornadic Supercell Environments Analyzed Using Surface and Reanalysis Data: A Spatiotemporal Relational Data-Mining Approach.” *Journal of Applied Meteorology and Climatology* 51 (12): 2203–17. https://doi.org/10.1175/JAMC-D-11-060.1.

Gaur, Abhishek, Andre Schardong, and Slobodan P. Simonovic. 2020. “Gridded Extreme Precipitation Intensity–Duration–Frequency Estimates for the Canadian Landmass.” *Journal of Hydrologic Engineering* 25 (6): 05020006. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001924.

Gaur, Ayushi, Abhishek Gaur, Dai Yamazaki, and Slobodan P. Simonovic. 2019. “Flooding Related Consequences of Climate Change on Canadian Cities and Flow Regulation Infrastructure.” *Water* 11 (1): 63. https://doi.org/10.3390/w11010063.

Genest, Christian, and Jean-Claude Boies. 2003. “Detecting Dependence With Kendall Plots.” *The American Statistician* 57 (4): 275–84. https://doi.org/10.1198/0003130032431.

Genest, Christian, Jean-Francois Quessy, and Bruno Remillard. 2006. “Goodness-of-Fit Procedures for Copula Models Based on the Probability Integral Transformation.” *Scandinavian Journal of Statistics* 33 (2): 337–66. https://doi.org/10.1111/j.1467-9469.2006.00470.x.

Gräler, B., M. J. van den Berg, S. Vandenberghe, A. Petroselli, S. Grimaldi, B. De Baets, and N. E. C. Verhoest. 2013. “Multivariate Return Periods in Hydrology: A Critical and Practical Review Focusing on Synthetic Design Hydrograph Estimation.” *Hydrology and Earth System Sciences* 17 (4): 1281–96. https://doi.org/10.5194/hess-17-1281-2013.

Hitchens, Nathan M., and Harold E. Brooks. 2013. “Preliminary Investigation of the Contribution of Supercell Thunderstorms to the Climatology of Heavy and Extreme Precipitation in the United States.” *Atmospheric Research* 123 (April): 206–10. https://doi.org/10.1016/j.atmosres.2012.06.023.

Holman, Kathleen D. 2018. “Characterizing Antecedent Conditions Prior to Annual Maximum Flood Events in a High-Elevation Watershed Using Self-Organizing Maps.” *Journal of Hydrometeorology* 19 (11): 1721–30. https://doi.org/10.1175/JHM-D-17-0229.1.

Huang, Q., W.J. Jiang, and H.P. Hong. 2021. “Statistical Assessment of Spatial Tornado Occurrences in Canada: Modelling and Estimation.” *Journal of Applied Meteorology and Climatology*, October. https://doi.org/10.1175/JAMC-D-20-0141.1.

Hunter, Cameron. 2016. “An Evaluation of the North American Regional Reanalysis Precipitation Fields in a Topographically Complex Domain, British Columbia, Canada.” https://doi.org/10.14288/1.0308672.

Ingrosso, Roberto, Piero Lionello, Mario Marcello Miglietta, and Gianfausto Salvadori. 2020. “A Statistical Investigation of Mesoscale Precursors of Significant Tornadoes: The Italian Case Study.” *Atmosphere* 11 (3): 301. https://doi.org/10.3390/atmos11030301.

Jalili Pirani, F., and M. R. Najafi. 2020. “Recent Trends in Individual and Multivariate Compound Flood Drivers in Canada’s Coasts.” *Water Resources Research* 56 (8). https://doi.org/10.1029/2020WR027785.

Khalafzai, Muhammad-Arshad K., Tara K. McGee, and Brenda Parlee. 2019. “Flooding in the James Bay Region of Northern Ontario, Canada: Learning from Traditional Knowledge of Kashechewan First Nation.” *International Journal of Disaster Risk Reduction* 36 (May): 101100. https://doi.org/10.1016/j.ijdrr.2019.101100.

Khalafzai, Muhammad-Arshad K. 2021. “Spring Flooding and Recurring Evacuations of Kashechewan First Nation, Northern Ontario, Canada.” *International Journal of Disaster Risk Reduction* 63 (September): 102443. https://doi.org/10.1016/j.ijdrr.2021.102443.

King, Austin T., and Aaron D. Kennedy. 2019. “North American Supercell Environments in Atmospheric Reanalyses and RUC-2.” *Journal of Applied Meteorology and Climatology* 58 (1): 71–92. https://doi.org/10.1175/JAMC-D-18-0015.1.

Klotzbach, Philip J., Carl J. Schreck III, Jennifer M. Collins, Michael M. Bell, Eric S. Blake, and David Roache. 2018. “The Extremely Active 2017 North Atlantic Hurricane Season.” *Monthly Weather Review* 146 (10): 3425–43. https://doi.org/10.1175/MWR-D-18-0078.1.

Kunkel, Kenneth E., Karen Andsager, and David R. Easterling. 1999. “Long-Term Trends in Extreme Precipitation Events over the Conterminous United States and Canada.” *Journal of Climate* 12 (8): 2515–27. https://doi.org/10.1175/1520-0442(1999)012<2515:LTTIEP>2.0.CO;2.

Leahy, Thomas Patrick. 2021. “Copula Approach for Simulated Damages Caused by Landfalling US Hurricanes.” *Advances in Statistical Climatology, Meteorology and Oceanography* 7 (1): 1–11. https://doi.org/10.5194/ascmo-7-1-2021.

Lentz, Erika E., E. Robert Thieler, Nathaniel G. Plant, Sawyer R. Stippa, Radley M. Horton, and Dean B. Gesch. 2016. “Evaluation of Dynamic Coastal Response to Sea-Level Rise Modifies Inundation Likelihood.” *Nature Climate Change* 6 (7): 696–700. https://doi.org/10.1038/nclimate2957.

Li, Hong-Nan, Xiao-Wei Zheng, and Chao Li. 2019. “Copula-Based Joint Distribution Analysis of Wind Speed and Direction.” *Journal of Engineering Mechanics* 145 (5): 04019024. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001600.

Lin, Charles A. 2002. “A Coupled Atmospheric-Hydrological Modeling Study of the 1996 Ha! Ha! River Basin Flash Flood in Québec, Canada.” *Geophysical Research Letters* 29 (2): 1026. https://doi.org/10.1029/2001GL013827.

Mahmoudi, Mohammad Hasan, Mohammad Reza Najafi, Harsimrenjit Singh, and Markus Schnorbus. 2021. “Spatial and Temporal Changes in Climate Extremes over Northwestern North America: The Influence of Internal Climate Variability and External Forcing.” *Climatic Change* 165 (1–2): 14. https://doi.org/10.1007/s10584-021-03037-9.

Mccarthy, Patrick, Dave Carlsen, and Jim Slipec. 2008. “Elie, Manitoba, Canada, June 22, 2007: Canada’s First F5 Tornado.” In .

Mesbahzadeh, T., M. M. Miglietta, M. Mirakbari, F. Soleimani Sardoo, and M. Abdolhoseini. 2019. “Joint Modeling of Precipitation and Temperature Using Copula Theory for Current and Future Prediction under Climate Change Scenarios in Arid Lands (Case Study, Kerman Province, Iran).” *Advances in Meteorology* 2019 (June): 1–15. https://doi.org/10.1155/2019/6848049.

Mesinger, Fedor, Geoff DiMego, Eugenia Kalnay, Kenneth Mitchell, Perry C. Shafran, Wesley Ebisuzaki, Dušan Jović, et al. 2006. “North American Regional Reanalysis.” *Bulletin of the American Meteorological Society* 87 (3): 343–60. https://doi.org/10.1175/BAMS-87-3-343.

Moller, Alan R., Charles A. Doswell, Michael P. Foster, and Gary R. Woodall. 1994. “The Operational Recognition of Supercell Thunderstorm Environments and Storm Structures.” *Weather and Forecasting* 9 (3): 327–47. https://doi.org/10.1175/1520-0434(1994)009<0327:TOROST>2.0.CO;2.

Murakami, H., E. Levin, T. L. Delworth, R. Gudgel, and P.-C. Hsu. 2018. “Dominant Effect of Relative Tropical Atlantic Warming on Major Hurricane Occurrence.” *Science* 362 (6416): 794–99. https://doi.org/10.1126/science.aat6711.

Najafi, Mohammad Reza, Francis W. Zwiers, and Nathan P. Gillett. 2017. “Attribution of Observed Streamflow Changes in Key British Columbia Drainage Basins.” *Geophysical Research Letters* 44 (21). https://doi.org/10.1002/2017GL075016.

Naz, Saba, Muhammad Ahsanuddin, Syed Inayatullah, Tanveer Siddiqi, and Muhammad Imtiaz. 2019. “Copula-Based Bivariate Flood Risk Assessment on Tarbela Dam, Pakistan.” *Hydrology* 6 (3): 79. https://doi.org/10.3390/hydrology6030079.

Nielsen, Erik R., Gregory R. Herman, Robert C. Tournay, John M. Peters, and Russ S. Schumacher. 2015. “Double Impact: When Both Tornadoes and Flash Floods Threaten the Same Place at the Same Time.” *Weather and Forecasting* 30 (6): 1673–93. https://doi.org/10.1175/WAF-D-15-0084.1.

Owen, Laura E., Jennifer L. Catto, David B. Stephenson, and Nick J. Dunstone. 2021. “Compound Precipitation and Wind Extremes over Europe and Their Relationship to

Extratropical Cyclones." *Weather and Climate Extremes* 33 (September): 100342. https://doi.org/10.1016/j.wace.2021.100342.

Poulin, Annie, David Huard, Anne-Catherine Favre, and Stéphane Pugin. 2007. "Importance of Tail Dependence in Bivariate Frequency Analysis." *Journal of Hydrologic Engineering* 12 (4): 394–403. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(394).

Raible, C. C., M. Yoshimori, T. F. Stocker, and C. Casty. 2007. "Extreme Midlatitude Cyclones and Their Implications for Precipitation and Wind Speed Extremes in Simulations of the Maunder Minimum versus Present Day Conditions." *Climate Dynamics* 28 (4): 409–23. https://doi.org/10.1007/s00382-006-0188-7.

Raymond, Colin, Radley M. Horton, Jakob Zscheischler, Olivia Martius, Amir AghaKouchak, Jennifer Balch, Steven G. Bowen, et al. 2020. "Understanding and Managing Connected Extreme Events." *Nature Climate Change* 10 (7): 611–21. https://doi.org/10.1038/s41558-020-0790-4.

Renard, B., and M. Lang. 2007. "Use of a Gaussian Copula for Multivariate Extreme Value Analysis: Some Case Studies in Hydrology." *Advances in Water Resources* 30 (4): 897–912. https://doi.org/10.1016/j.advwatres.2006.08.001.

Requena, A. I., L. Mediero, and L. Garrote. 2013. "A Bivariate Return Period Based on Copulas for Hydrologic Dam Design: Accounting for Reservoir Routing in Risk Estimation." *Hydrology and Earth System Sciences* 17 (8): 3023–38. https://doi.org/10.5194/hess-17-3023-2013.

Rogash, Joseph A., and Richard D. Smith. 2000. "Multiscale Overview of a Violent Tornado Outbreak with Attendant Flash Flooding." *Weather and Forecasting* 15 (4): 416–31. https://doi.org/10.1175/1520-0434(2000)015<0416:MOOAVT>2.0.CO;2.

Sadegh, Mojtaba, Elisa Ragno, and Amir AghaKouchak. 2017. "Multivariate C Opula A Nalysis T Oolbox (MvCAT): Describing Dependence and Underlying Uncertainty Using a B Ayesian Framework." *Water Resources Research* 53 (6): 5166–83. https://doi.org/10.1002/2016WR020242.

Salvadori, G. 2004. "Bivariate Return Periods via 2-Copulas." *Statistical Methodology* 1 (1–2): 129–44. https://doi.org/10.1016/j.stamet.2004.07.002.

Salvadori, G., and C. De Michele. 2004. "Frequency Analysis via Copulas: Theoretical Aspects and Applications to Hydrological Events: FREQUENCY ANALYSIS VIA COPULAS." *Water Resources Research* 40 (12). https://doi.org/10.1029/2004WR003133.

Salvadori, G. 2010. "Multivariate Multiparameter Extreme Value Models and Return Periods: A Copula Approach." *Water Resources Research* 46 (10): 2009WR009040. https://doi.org/10.1029/2009WR009040.

Salvadori, G., C. De Michele, and F. Durante. 2011. “On the Return Period and Design in a Multivariate Framework.” *Hydrology and Earth System Sciences* 15 (11): 3293–3305. https://doi.org/10.5194/hess-15-3293-2011.

Salvadori, G., G.R. Tomasicchio, and F. D’Alessandro. 2014. “Practical Guidelines for Multivariate Analysis and Design in Coastal and Off-Shore Engineering.” *Coastal Engineering* 88 (June): 1–14. https://doi.org/10.1016/j.coastaleng.2014.01.011.

Salvadori, Gianfausto, ed. 2007. *Extremes in Nature: An Approach Using Copulas*. Water Science and Technology Library 56. Dordrecht: Springer.

Serinaldi, Francesco. 2015. “Dismissing Return Periods!” *Stochastic Environmental Research and Risk Assessment* 29 (4): 1179–89. https://doi.org/10.1007/s00477-014-0916-1.

Serinaldi, Francesco. 2016. “Can We Tell More than We Can Know? The Limits of Bivariate Drought Analyses in the United States.” *Stochastic Environmental Research and Risk Assessment* 30 (6): 1691–1704. https://doi.org/10.1007/s00477-015-1124-3.

Sills, David M. L. 2012. “Using Tornado, Lightning, and Population Data to Identify Tornado Prone Areas in Canada.” In .

Sills, David M.L. 2014. “Implementation and Application of the EF-Scale in Canada.” In .

Singh, Harsimrenjit, Mohammad Reza Najafi, and Alex J. Cannon. 2021. “Characterizing Non-Stationary Compound Extreme Events in a Changing Climate Based on Large-Ensemble Climate Simulations.” *Climate Dynamics* 56 (5–6): 1389–1405. https://doi.org/10.1007/s00382-020-05538-2.

Singh, Harsimrenjit, Farshad Jalili Pirani, and Mohammad Reza Najafi. 2020. “Characterizing the Temperature and Precipitation Covariability over Canada.” *Theoretical and Applied Climatology* 139 (3–4): 1543–58. https://doi.org/10.1007/s00704-019-03062-w.

Singh, Harsimrenjit, and Mohammad Reza Najafi. 2020. “Evaluation of Gridded Climate Datasets over Canada Using Univariate and Bivariate Approaches: Implications for Hydrological Modelling.” *Journal of Hydrology* 584 (May): 124673. https://doi.org/10.1016/j.jhydrol.2020.124673.

Smith, James A., Mary Lynn Baeck, Yu Zhang, and Charles A. Doswell. 2001. “Extreme Rainfall and Flooding from Supercell Thunderstorms.” *Journal of Hydrometeorology* 2 (5): 469–89. https://doi.org/10.1175/1525-7541(2001)002<0469:ERAFFS>2.0.CO;2.

Tarek, Mostafa, François P. Brissette, and Richard Arsenault. 2020. “Evaluation of the ERA5 Reanalysis as a Potential Reference Dataset for Hydrological Modelling over North America.” *Hydrology and Earth System Sciences* 24 (5): 2527–44. https://doi.org/10.5194/hess-24-2527-2020.

Temmerman, Stijn, Patrick Meire, Tjeerd J. Bouma, Peter M. J. Herman, Tom Ysebaert, and Huib J. De Vriend. 2013. “Ecosystem-Based Coastal Defence in the Face of Global Change.” *Nature* 504 (7478): 79–83. https://doi.org/10.1038/nature12859.

Teng, Chung-Chu, and Paul C. Liu. 2001. “Estimating Wave Height Distributions from Wind Speed Distributions.” In *Coastal Engineering 2000*, 310–19. Sydney, Australia: American Society of Civil Engineers. https://doi.org/10.1061/40549(276)24.

Thompson, Keith R., Natacha B. Bernier, and Paul Chan. 2009. “Extreme Sea Levels, Coastal Flooding and Climate Change with a Focus on Atlantic Canada.” *Natural Hazards* 51 (1): 139–50. https://doi.org/10.1007/s11069-009-9380-5.

Vandenberghe, S., N. E. C. Verhoest, C. Onof, and B. De Baets. 2011. “A Comparative Copula-Based Bivariate Frequency Analysis of Observed and Simulated Storm Events: A Case Study on Bartlett-Lewis Modeled Rainfall: COPULA-BASED EVALUATION OF RAINFALL MODELS.” *Water Resources Research* 47 (7). https://doi.org/10.1029/2009WR008388.

Wakimoto, Roger M., Huaqing Cai, and Hanne V. Murphey. 2004. “The Superior, Nebraska, Supercell During BAMEX.” *Bulletin of the American Meteorological Society* 85 (8): 1095–1106. https://doi.org/10.1175/BAMS-85-8-1095.

Waliser, Duane, and Bin Guan. 2017. “Extreme Winds and Precipitation during Landfall of Atmospheric Rivers.” *Nature Geoscience* 10 (3): 179–83. https://doi.org/10.1038/ngeo2894.

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. “The ASA Statement on *p* -Values: Context, Process, and Purpose.” *The American Statistician* 70 (2): 129–33. https://doi.org/10.1080/00031305.2016.1154108.

Willems, Patrick, Armelle Guillou, and Jan Beirlant. 2007. “Bias Correction in Hydrologic GPD Based Extreme Value Analysis by Means of a Slowly Varying Function.” *Journal of Hydrology* 338 (3–4): 221–36. https://doi.org/10.1016/j.jhydrol.2007.02.035.

Yue, Sheng, and Peter Rasmussen. 2002. “Bivariate Frequency Analysis: Discussion of Some Useful Concepts in Hydrological Application.” *Hydrological Processes* 16 (14): 2881–98. https://doi.org/10.1002/hyp.1185.

Zeng, Z., W. W. Hsieh, A. Shabbar, and W. R. Burrows. 2011. “Seasonal Prediction of Winter Extreme Precipitation over Canada by Support Vector Regression.” *Hydrology and Earth System Sciences* 15 (1): 65–74. https://doi.org/10.5194/hess-15-65-2011.

Zhang, Xuebin, W. D. Hogg, and Éva Mekis. 2001. “Spatial and Temporal Characteristics of Heavy Precipitation Events over Canada.” *Journal of Climate* 14 (9): 1923–36. https://doi.org/10.1175/1520-0442(2001)014<1923:SATCOH>2.0.CO;2.

Zhang, Ying, and Mohammad Reza Najafi. 2020. “Probabilistic Numerical Modeling of Compound Flooding Caused by Tropical Storm Matthew Over a Data-Scarce Coastal Environment.” *Water Resources Research* 56 (10). https://doi.org/10.1029/2020WR028565.

Zhang, Yuqing, Xiubao Sun, and Changchun Chen. 2021. “Characteristics of Concurrent Precipitation and Wind Speed Extremes in China.” *Weather and Climate Extremes* 32 (June): 100322. https://doi.org/10.1016/j.wace.2021.100322.

Zscheischler, Jakob, Olivia Martius, Seth Westra, Emanuele Bevacqua, Colin Raymond, Radley M. Horton, Bart van den Hurk, et al. 2020. “A Typology of Compound Weather and Climate Events.” *Nature Reviews Earth & Environment* 1 (7): 333–47. https://doi.org/10.1038/s43017-020-0060-z.

Zscheischler, Jakob, Philippe Naveau, Olivia Martius, Sebastian Engelke, and Christoph C. Raible. 2021. “Evaluating the Dependence Structure of Compound Precipitation and Wind Speed Extremes.” *Earth System Dynamics* 12 (1): 1–16. https://doi.org/10.5194/esd-12-1-2021.

Zscheischler, Jakob, Seth Westra, Bart J. J. M. van den Hurk, Sonia I. Seneviratne, Philip J. Ward, Andy Pitman, Amir AghaKouchak, et al. 2018. “Future Climate Risk from Compound Events.” *Nature Climate Change* 8 (6): 469–77. https://doi.org/10.1038/s41558-018-0156-3.

YEU DECK (TED) NGUI

ENGINEER-IN-TRAINING (TX AND NY, USA)

PROFESSIONAL EXPERIENCE

TEXAS COMMISSION ON ENVIRONMENTAL QUALITY, AUSTIN, TX, JUNE 2016 TO DECEMBER 2018
ENVIRONMENTAL ENGINEERING SPECIALIST II, STORMWATER & PRETREATMENT TEAM

- Led yearly EPA Audit Risk Based Assessments for 25 pretreatment programs to evaluate their operations and performance.
- Mobilized and supervised 8-person team in ensuring quality project management, hosting team meetings regularly to monitor progress and track task completion; successfully shortened project duration by 50%.
- Managed 12 pretreatment programs in Texas to protect water quality and municipal wastewater treatment plants and conveyance infrastructure from conventional and toxic industry pollutants.
- Conducted and headed National Pollutant Discharge Elimination System (NPDES) pretreatment audits to assess compliance with regulatory requirements.
- Inspected more than 50 manufacturing industries wastewater treatment system to ensure they are properly regulated by the municipalities.
- Assessed Technically Based Local Limits (TBLLs) for Control Authorities (e.g., City, County, MUD) by calculating complex WWTP removal efficiency and effluent limits for WWTPs.
- Reviewed NPDES permit applications from municipalities, defining requirements for the development and implementation of a pretreatment program during in the regulatory permitting process.
- Documented technical reports and industrial sites reviews for compliance during audits to comply with approved pretreatment programs.
- Provided excellent customer services to municipal's managerial officials and consultants regarding regulatory permitting processes, question/interpretation for wastewater rules and regulations.

COLLEGIATE COURSEWORK AND PROJECTS

STORMWATER MANAGEMENT FACILITY DESIGN, UNIVERSITY OF WESTERN ONTARIO, SEPTEMBER 2019 TO DECEMBER 2019

- Led 7-person team to design stormwater wet pond for a 50ha subdivision development site in London for quality and quantity control of surface runoff before discharging to receiving water bodies.
- Developed stormwater models for different design storms and successfully simulated pre/post development runoff condition by using Visual Otthymo 2.0 as the target flows for the wet pond design.
- Performed complex engineering calculations to identify wet pond's dimensions in compliance with City of London's and Ontario's SWM requirements.
- Successfully designed a wet pond that reduces post-development surface runoff (up to 100-year storm events) and guaranteed 80% Total Suspended Solids (TSS) removal down to pre-development condition.
- Generated a 30-page technical design report that encompasses assumptions, justifications, design calculations, site due diligence, CAD drawings, and stormwater modelling processes.

STATISTICAL ANALYSIS ON FLOOD RISKS AND IMPACT OF WILDFIRE, SEPTEMBER 2019 TO DECEMBER 2019

- Conducted statistical analyses on flood risks and extreme precipitations after wild fires in British Columbia.
- Collected over 8,000 data from climate and stream stations and processed them efficiently using RStudio and ArcGIS.
- Utilized advanced statistical models such as multivariate regressions and variables correlation to analyze the relationship between burned area, rainfall events, and flood events.

AUTOMATED WATERING SYSTEM DESIGN FOR COMMUNITY GARDEN, UNIVERSITY AT BUFFALO, SEPTEMBER 2015 TO DECEMBER 2015

- Coordinated with four engineering colleagues from different fields in designing an automated watering system for a community garden at University Height.
- Monitored the purification and filtration system, water collection, and distribution for the garden, comparing and recording alternatives based on feasibility, safety, and efficiency.
- Performed cost analysis on the parts to build the automated watering system.

FEASIBILITY STUDY OF REMEDIATION ALTERNATIVES IN LACKAWANNA, UNIVERSITY AT BUFFALO, JANUARY 2015 TO MAY 2015

- Headed a two-part project, including remedial investigation, treatability study, and overall feasibility study.
- Tracked wells and piezometers to generate chemical profile within the water table, performed data analysis, and created a contour map of potentiometric surface using Surfer8 and VisualAEM.
- Directed site characterization analysis on Bethlehem Steel Site, using data provided by NYSDEC, to recommended remediation alternatives in compliance with NYS standards, criteria, and guidelines.
- Evaluated parameters to ensure the model was correct on hydraulic conductivity, aquifer thickness, existence of pumps, heads, etc.
- Determined the groundwater flow direction in the basin before and after the recharge (storm, discharge, etc.) including addition of pump.

EDUCATION

MASTERS OF ENGINEERING SCIENCE (M.E.SC.) IN CIVIL AND ENVIRONMENTAL ENGINEERING, 2022; *Western University, ON;*

BACHELOR OF SCIENCE (B.SC.) IN ENVIRONMENTAL ENGINEERING, 2016; *State University of New York, Buffalo, NY;*

AWARDS AND HONORS

- Western University Civil and Environment Engineering 3-Minute Thesis, First Place
- Texas Commission on Environmental Quality, Water Quality Division, Outstanding Employee, July 2018

ADDITIONAL INFORMATION

Technical Proficiencies: RStudio, MATLAB, Python, HEC-RAS, HEC-HMS, Visual Otthymo 2.0, PCSWMM, ArcGIS (Coursera Certified), AutoCAD, Microsoft Office Suite