2023

Joint Probability Analysis of Extreme Precipitation and Water Level for Chicago, Illinois

Anna Li Holey
Michigan Technological University, aeholey@mtu.edu

JOINT PROBABILITY ANALYSIS OF EXTREME PRECIPITATION AND WATER LEVEL FOR CHICAGO, ILLINOIS

By

Anna Li Holey

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Environmental Engineering

MICHIGAN TECHNOLOGICAL UNIVERSITY

2023

This thesis has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Environmental Engineering.

Department of Civil, Environmental, and Geospatial Engineering

Thesis Advisor: *Pengfei Xue*

Committee Member: *William Pringle*

Committee Member: *Jiali Wang*

Committee Member: *Yi Liu*

Department Chair: *Audra Morse*

Table of Contents

Acknowledgements

I would like to thank my advisor Dr. Pengfei Xue for his continuous support and mentorship during my Masters study. I am very grateful for the guidance and opportunities he has provided.

I am thankful to my committee members Dr. William Pringle, Dr. Jiali Wang, and Dr. Yi Liu, for their input and expertise. My sincere thanks to Saumik Mallik for assisting with analyzing data and assisting with the methods.

I would also like to thank Dr. William Asquith, research hydrologist with USGS and adjunct scientist at Texas Tech University, for his continuous assistance with R and willingness to give feedback and guidance throughout the whole research process.

This research was supported by COMPASS-GLM, a multi-institutional project supported by the Earth and Environmental System Science Division of the U.S. Department of Energy's Office of Science.

Abstract

A compound flooding event occurs when there is a combination of two or more extreme factors that happen simultaneously or in quick succession and can lead to flooding. In the Great Lakes region, it is common for a compound flooding event to occur with a high lake water level and heavy rainfall. With the potential of increasing water levels and an increase in precipitation under climate change, the Great Lakes coastal regions could be at risk for more frequent and severe flooding. The City of Chicago which is located on Lake Michigan has a high population and dense infrastructure and is very vulnerable to a compound flooding event, even with the implementation of its water control structures. For this case study, annual maximum precipitation and corresponding lake water level data were analyzed to examine the bivariate return period of a compound flood event using a copula function. The results show that under climate change if the water level were to rise by 0.2, 0.45, or 0.8 m, compound flooding events due to heavy precipitation and a high water level will be more likely in the future. By documenting the joint risk of potential compound flooding in this area, preventative measures and planning can be implemented.

1 Introduction

The Laurentian Great Lakes are one of the largest stores of surface fresh water in the world located in the United States and Canada. The Great Lakes, which consists of five lakes, are considered to be like “inland seas” because of their characteristics including bathymetry gradient, deep depths, and extensive size (Huang et al., 2021). Climate trends of the Great Lakes and surrounding areas have been observed in recent decades, and concerns continue to rise about future climate change impacts to the area. There is strong scientific evidence that climate change is a result of human activities mostly related to the increase in greenhouse gas (GHG) emissions, like burning fossil fuels (IPCC, 2022). Because of the increasing GHG emissions, it is impacting lake evaporation, ice cover, and lake stratification in some of the world’s largest lakes (O’Reilly et al., 2015; Austin et al., 2007; Xue et al., 2022; Hayhoe et al., 2010; Mason et al., 2016; Anderson et al., 2021). Globally, from 1985 to 2009, the mean lake summer surface water temperatures have increased about 0.34°C per decade (O’Reilly et al., 2015). When examining seasonal ice cover, from 1973 to 2010, there has been a decrease in ice cover by 71 percent between the five Great Lakes (Wang et al., 2012). The Great Lakes basin air temperature has also increased and is expected to increase to the end of the century (Wuebbles et al., 2019). With an increase in air and surface water temperature, more evaporation can occur producing moisture laden air, generating an increase in heavy precipitation. Because of the interdependency between a multitude of weather and climate variables, the slightest change in precipitation patterns can lead to long-term impacts in an area.

Precipitation in the Great Lakes region is also expected to increase and become more variable. From 1901 to 2015, the Great Lakes region experienced about a 10 percent increase in annual precipitation, usually in the form of large precipitation events (Wuebbles et al., 2019). With a high GHG emission scenario, Chicago could receive up to 30 percent more precipitation in the spring and winter by the end of the century (Hayhoe et al., 2010). Based on different emission scenarios, it is also predicted that by the mid-21st century (2030–2049) precipitation could increase between 0% to 13% and by the late-21st century (2080–2099) precipitation could increase between 9% to 32% (Xue et al., 2022). Similarly, in the future (2045–2060), the Great Lakes region could experience about a 7 percent increase in average rainfall intensity per ℃ of surface warming based on a high emission scenario (d’Orgeville, 2014). Increase in precipitation could greatly affect flood probabilities in the future.

Amidst escalated precipitation events, there has also been increased lake level variability in the Great Lakes region. Over the last 60 years, the water level in the Great Lakes region was in a high-water level regime followed by a period of record low water level, until around 2013 when water levels started to rise again (Lofgren et al., 2002; Gronewold & Stow, 2014; Gronewold & Rood, 2018). The 2013–2014 water level rise was attributed to increase in spring runoff and persistent precipitation, the continuation of water level rise is driven by a combination of above average overlake precipitation and runoff (Gronewold et al., 2016, 2021). In 2020, the water level was so high that Lake Superior, Lake Michigan-Huron, and Lake Erie all reached new record-high mean water levels multiple times, and numerous areas experienced shore flooding and erosion (2020 Annual Climate Trends and Impacts). Recent studies have suggested a further increase in lake water level with projected climate change (Kayastha et al., 2022).

In addition to the large water level variability on the seasonal and climatic scales, the Great Lakes are also susceptible to surges. Surges are driven by meteorological factors like strong winds or sudden changes in atmospheric pressure, they happen frequently in the Great Lakes and can cause the water level to rise by many feet. A storm surge in the Great Lakes could last all day and can range from 1 to 8 feet high (Wisconsin Sea Grant, 2022). In the future, an increase in storm intensity and duration and a decrease in ice cover could also lead to an increase frequency in storm surges. Increases in storm surge and high-water level could negatively impact aquatic vegetation communities and shoreline wetlands (Wuebbles et al., 2019). With an increase in precipitation and higher water levels, the addition of a surge could greatly increase the potential for a compound flooding event.

The Intergovernmental Panel on Climate Change defines a compound event to be either (1) two or more extreme events occurring simultaneously or successively, (2) combinations of extreme events with underlying conditions that amplify the impact of the events, or (3) combinations of events that are not themselves extreme but lead to an extreme event or impact when combined (Seneviratne et al., 2012). There is a growing concern that due to climate change, compound flooding events will be more likely in the future. This is particularly true for the coastal cities along the Great Lakes coastline. The goal of this study is to examine the probabilities of compound flood events happening in Chicago, Illinois because of heavy rainfalls and high-water levels. Observational annual maximum rainfall data from 1982–2021 with the corresponding daily mean water level will be used and the probability will be examined by applying a copula function and generating return periods. While examining the joint probability of a compound flood event occurring because of surge and precipitation has already been studied, to our knowledge this is the first study to analyze rainfall and water level and look at potential increase under climate change for the Chicago area. For flood risk management and reduction, it is important to understand and analyze the potential cause of flooding. With the dense population and high infrastructure, the downtown Chicago area could be greatly impacted if a compound flood event were to occur. By documenting the joint risk of potential compound flooding in this area, preventative measures and planning can be implemented.

2 Study Area and Data

Chicago is located in the northeastern corner of Illinois and lies on Lake Michigan, one out of the five Great Lakes. It is home to 2.69 million people and is the third-largest city in the United States (U.S. Census Bureau, 2020). Chicago is one of the hotspots that receives growing concern of flood impacts because of the high population density and the high amount of impervious surface that makes up the bustling city. Chicago has a long history of flooding since the 1800s. Chicago has been built in a flat area of prairie wetlands, and is naturally prone to flooding because the City sits on an impervious layer of clay leading to poor drainage (Brosnan, 2020). In the early 1800's to the mid-1900's Chicago faced many severe floods and water quality issues because of the combined sewer and storm water system. Since 1945 Chicago has had more damaging floods caused by suburban sprawl and climate change (Brosnan, 2020). From 2019 to 2021 the National Weather Service has listed 17 flooding events for the area of Chicago. While FEMA has identified about 1,500 properties located in a flood hazard area, a new report estimated that 77,000 properties might be at risk (Grunderson, 2020).

The Chicago River runs through the city and is 28 miles long, it previously discharged into Lake Michigan before its reversal in the 1900s, and now drains into the Mississippi River. The Chicago River is part of the Chicago Area Waterway System (CAWS), that consists of more than 100 miles of man-made channels and former natural streams that have been modified by dredging, straightening, widening and realigning (MWRDGC). The heavily engineered CAWS was constructed from 1892 to 1965 to accommodate growing population and public health concerns (Lanyon et al., 2013). CAWS consists of locks, dams, and ports and is used for recreational and commercial transportation. Through CAWS, the Metropolitan Water Reclamation District of Greater Chicago (MWRD) can readily regulate the water levels. In the past, Chicago flooding has caused environmental and economic damage, from 2007 to 2014 there was at least $2.319 billion in documented damages for northeastern Illinois according to the 2015 Illinois Department of Natural Resources. Under climate change, Chicago could face more unpredicted storms that would result in flooding (Markus et al., 2012).

To mitigate flooding risk, Chicago has an underground reservoir system, implemented under the Tunnel and Reservoir Plan (TARP) that can store excess water for heavy rainfall events that might cause flooding. The MWRD is in charge of regulating the water level and can reverse the flow of the Chicago River to discharge into Lake Michigan during major flood events. The Chicago River water level height is extensively monitored by MWRD and has four different control structures to regulate the water level. The Chicago River water level usually stays around 2 feet below the Chicago City Datum (0 ft). Backflows into Lake Michigan occur if the river water level reaches about 3.5 feet above the city datum, which is about 5.5 feet higher than normal. However, the MWRD is only able to reverse the flow if the water level in the river is higher than the lake level.

On 17 May 2020, MWRD was unable to reverse the river flow due to a high amount of precipitation received leading up to the event and a high lake level. It was not until the evening of 17 May 2020, MWRD was finally able to reverse the flow. However, because the lake water level was so high, MWRD had to open and close the sluice gate multiple times until the river water level decreased. This event caused flooding downtown and electrical power outages throughout the city. Some factors that might cause backflows are timing, hydraulic properties of the channels, distribution of precipitation, and the operation of the control structures (Duncker & Johnson, 2016). Figure 1, shows the increase in backflows per year from 1985–2021. When backflows occur, untreated sewage flows back into Lake Michigan, causing negative environmental impacts and public health concerns to Lake Michigan.

2.1 Data Collection

The rainfall data for this study was obtained from 1982 to 2021(excluding 1996) from the National Oceanic and Atmospheric Administration Online Weather Data (NOWData). The data set consisted of the annual precipitation maxima from the Chicago Midway Airport 3 SW (111577) rain gauge station. The Chicago Midway Airport is South of O'Hare International Airport located in Cook County (Figure 2). This rain gauge station offered the most complete data set, where about 16% of the daily observational record was marked as a "T", for trace amounts of precipitation that wet the rain gauge but was less than 0.01 inches and there were no missing data.

The water level dataset was extracted using water level data from NOAA at the Calumet Harbor, IL Station (9087044) shown in Figure 2. The corresponding daily mean water levels were extracted using the observational hourly water level data and obtaining the mean water level for the same day as the annual maximum precipitation event. This dataset consists of 39 years because the year of 1996 is sufficiently incomplete at the Calumet Harbor station.

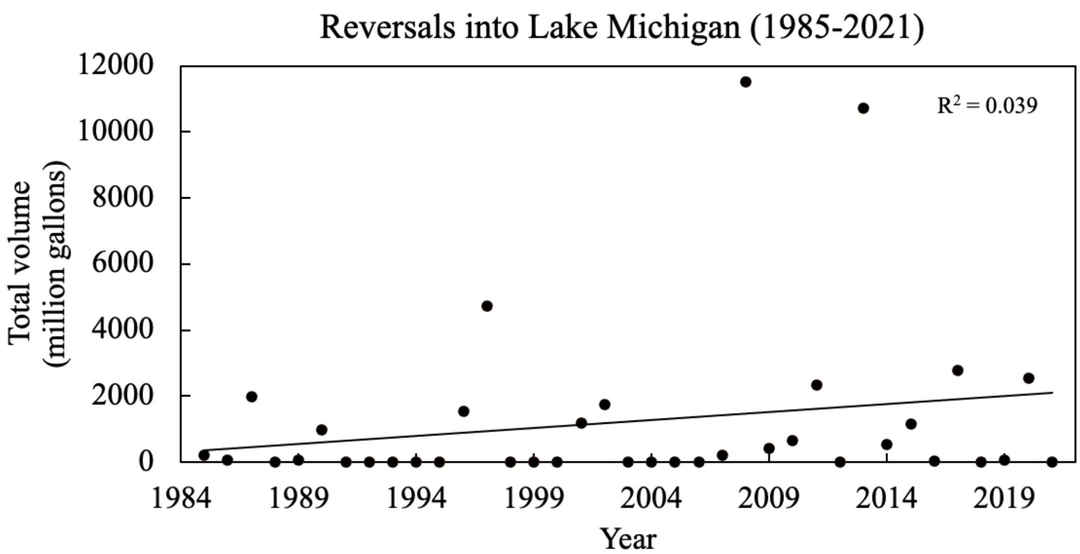

Figure 1. Reversals from the Chicago Area Waterway System to Lake Michigan from 1985–2021 with a linear regression trend line ($y = 47.73x - 9437$, R^2=0.039). Data provided by the Metropolitan Water Reclamation District of Greater Chicago.

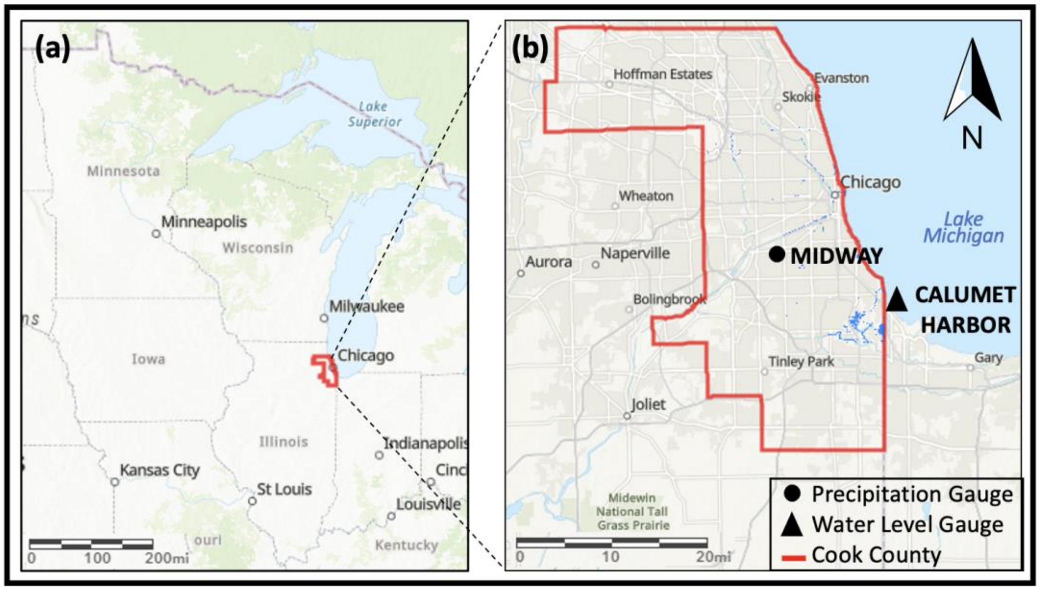

Figure 2. Study site location (a) Cook County in the state of Illinois, USA. (b) Close up of Cook County with the precipitation and water level gauge stations used for this study.

3 Methodology

3.1 Marginal Distributions

The marginal distributions that suitably fit the data need to be obtained. The distributions tested in this study were Normal (NORM), Lognormal (LNORM), Weibull (WEIB), Gamma (GAMM), Gumbel (GUMB) and Generalized Extreme Distribution (GEV). The parameters for the distributions were obtained using the maximum likelihood method (MLE). For each data value MLE maximizes the summation of logarithms of the probability density to estimate the parameters for a chosen distribution (Asquith et al., 2017). MLE was chosen to estimate parameters because of its consistency and tendency to show less biases compared to other methods (Chen et al. 2017). As guidance to the best distribution, the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson-Darling (A-D), Kolmogorov-Smirnov (K-S) test and Cramer Von Mises (CVM) were used. When distributions had similar goodness of fit, a P-P plot was used to visually complement the aforementioned metrics.

AIC and BIC are standard goodness of fit tests. They can be expressed as:

$$AIC = 2k - 2\ln(L) \tag{1}$$

$$BIC = k\ln(n) - 2\ln(L) \tag{2}$$

where k represents the number of independent variables, L is the log-likelihood estimate and n are the number of observations. AIC and BIC are both ways to score models based on their complexity and their log-likelihood. There is slight variation between the two tests, BIC penalizes model parameters more than AIC, which results in fewer complex models being favored. Both AIC and BIC were implemented in this study because neither one is better than the other and they should both produce similar results. AIC and BIC are both standard goodness of fit tests, they may not be the best for fitting data that has high asymmetrical distributions or a small sample size (Liao et al., 2009), but in some cases each can be helpful for identifying the best model (Chen et al., 2017).

K-S, A-D, and CVM are all non-parametric hypothesis-based model selection techniques. All three tests produce a test statistic that represents the difference between the empirical and theoretical cumulative distribution functions, which results in a smaller test statistic being more favorable. The null hypothesis being tested for the K-S, A-D and CVM is that the data follows a specified distribution. To reject the null hypothesis, the test statistic produced would have to be greater than the critical value. For K-S and CVM the critical value is determined by the significance level (α = 0.05) and the number of data points. For the A-D test, the critical value is determined by the significance level (α = 0.05) and the type of distribution (Pettit, 1976; Stephens, 1979). The A-D test is similar to the K-S test but gives more weight to the tails, which is important to note when analyzing hydrologic frequency analysis (Zeng et al., 2015). A-D and CVM are alternatives to K-S and tend to be more powerful than the K-S test.

3.2 Copula

Copulas have become increasingly popular, used for examining the joint probability of hydrologic data (Chen & Guo, 2019; Salvadori & De Michele, 2007). A copula is a multivariate distribution function that is used to show the relationship between two or more variables based on their marginal distributions. The notion of copula is based on Sklar's theorem, which states that any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula which describes the dependence structure between the variables (Sklar, 1959). If X and Y represent two random variables, then their marginal cumulative distribution functions would be

$$u = F_X(x) \text{ and } v = F_Y(y) \tag{3}$$

$$F(x,y) = C(u,v) = C(F_X(x), F_Y(y)) \tag{4}$$

Where $C(u,v)$ is the copula function when u and v are continuous. Copulas offer a more flexible way to explore nonlinear correlation and dependencies with multiple variables and distributions (Zellou et al., 2019; Xu et al., 2019). There are many different families of copulas but for hydrologic analysis Archimedean copulas, which only require one parameter, are more often being employed (Wang et al., 2017). Copulas allow variables to transform from the probability state space of [0,1] to $[0, \infty]$ and then using an inverse generator function transform the variable back to [0,1]. This allows copulas to show the relations between multiple variables and their joint probability.

Using copulas for multivariate analysis for compound events has been widely implemented. For example, analyzing the joint probability of storm surge and heavy rainfall for major U.S coastal cities (Wahl et al., 2015) and similarly, heavy rainfall and storm tides (Xu et al., 2014; Xu et al., 2019). There have also been numerous case studies conducted on specific sites looking at multivariate compound flood events using copulas, such as in south Florida (Jane et al., 2020), Buffalo, New York (Saharia et al. 2021), Ravenna, Italy (Bevacqua et al., 2017), Fuzhou City, China (Lian et al., 2013), Netherlands (Hurk et al., 2015), and Canada's Coast (Pirani et al., 2020).

In this study, four Archimedean copulas are tested, Gumbel, Clayton, Frank, and Joe. The common bivariate copula functions are listed in Table 1. The maximum likelihood method was used to estimate the parameters for the copulas.

Table 1. Common bivariate copula functions (Xu et al., 2019, Saharia et al., 2021).

Copulas	$C(u,v)$
Gumbel	$C(u,v) = \exp\left\{-\left[(\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{1/\theta}\right\}$
Clayton	$C(u,v) = (u^{-\theta} + v^{-\theta})^{-1/\theta}$
Frank	$C(u,v) = -\frac{1}{\theta}\ln\left[1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1)}\right]$
Joe	$C(u,v) = 1 - [(1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$

3.3 Return Period

The general equation for return period is as follows:

$$RP = \frac{T}{E} \tag{5}$$

where the return period (*RP*) for an event is the expected time interval (*T*) over the annual exceedance probability of an event occurring (*E*). For annual maximum events *T* would be one year. Flooding could occur when there is either a high amount of precipitation (*R)* or high water level (*S)* that might exceed a certain threshold. To look at the joint probability using the copula function, AND, and OR return period events were estimated. The AND joint probability, $P \cap (r,s)$, refers to the probability of heavy rainfall and high water level event that would happen simultaneously. The OR joint probability, $P \cup (r,s)$, refers to the probability that either rainfall or water level would exceed its threshold value and cause flooding. The AND joint return period, $T \cap (r,s)$ equation and the OR joint return period, $T \cup (r,s)$ are as follows:

$$T \cap (r,s) = \frac{1}{P\cap(r,s)} = \frac{1}{P\big((R>r)\cap(S>s)\big)} = \frac{1}{1-F(r)-F(s)+F(r,s)} \quad (6)$$

$$T \cup (r,s) = \frac{1}{P\cup(r,s)} = \frac{1}{P\big((R>r)\cup(S>s)\big)} = \frac{1}{1-F(r,s)} \quad (7)$$

Where, F(r,s) is the joint distribution of rainfall and water level based on the copula output and *F(r)* and *F(s)* are the marginal distributions for precipitation and water level respectively.

4 Results

Kendall's tau correlation coefficient is a non-parametric measure of the relationship between two variables based their ordinal association, where values range from 0 to 1 and a value closer to 1 is ideal. The Kendall's tau between annual maximum rainfall and corresponding water level is 0.114 (n = 39). The results reveal that there is a weak but statistically significant (α = 0.05) correlation between the variables. The dependence between the two hydrologic variables should be taken into account when looking at flood probability to ensure a more accurate return period output.

4.1 Marginal Distributions

Table 2 lists the output for testing goodness of fit for the candidate distributions. The parameters for each distribution were fit using MLE. The best distribution would display the lowest AIC, BIC value and pass the K-S, A-D and CVM tests. As stated above, the K-S, A-D and CVM are all non-parametric tests that compare the difference between the empirical and theoretical cumulative distribution functions; a small value represents how close the distributions are to one another. To pass all of the tests, the test statistic had to be smaller than the critical value. For the K-S test, the critical value with a 95% confidence level is 0.217 for a sample size of 39. All of the distributions pass the K-S test with the highest K-S D statistic value being 0.119. GEV was the chosen distribution for rainfall. The GEV distribution produced the lowest K-S D statistic for rainfall with a K-S D statistic of 0.076. The CVM critical value was 0.461, based on a significance level of α=0.05 and 39 data points. All of the distributions passed the CVM test, the Weibull distribution for water level produced the largest test statistic of 0.123, which is still fairly below the threshold. The critical values for A-D are based on significance (alpha = 0.5) and the chosen distribution, the critical values used are shown in Table 3. For the selected distributions, they all pass the A-D test, with the exception of the Weibull and GEV distribution for the water level data.

When distributions had similar goodness of fit, a P-P plot was used to visually identify the best fit. Figure 3, shows the P-P plots of the distributions chosen by comparing the empirical and theoretical probabilities. The root mean square error (RMSE) is also shown, which displays the difference between the observed and predicted values. The

RMSE results confirm that GEV and Gumbel (a special case of the GEV) are sufficient choices for the rainfall and water level data respectively. The RMSE for the rainfall data was 0.0265 and the RMSE for the water level data was 0.0389.

While many papers only use only two or three goodness of fit tests, there are some drawbacks because some goodness of fit tests tend to favor certain distributions. The hope was to eliminate any underlying bias, by using multiple goodness of fit tests for this study. Based on the results, it was common for a particular distribution to have the lowest AIC and BIC values, but did not yield the lowest A-D, K-S and CVM values. For example, when looking at the GEV distribution for water level, it produced one of the lowest A-D, K-S and CVM values, but one of the highest AIC and BIC values. Laio et al. (2009) also mentions similar results, where they found that AIC and BIC produce similar model selection when compared to Anderson-Darling Criterion, which is based on the Anderson-Darling test statistic. They also found that BIC and AIC tend to favor distributions with fewer parameters. This might be the reason that some of the distribution results favored AIC and BIC but contrasted with the A-D, K-S, and CVM. With that being said, for choosing the best distribution, there was more emphasis on passing the A-D, K-S and CVM, while also examining which test produced the smaller test statistic. In this study, multiple goodness fit tests were performed by aiming to minimize any underlying bias.

Table 2. Marginal Distribution goodness of fit tests for rainfall and water level.

Variable	Distribution	AIC	BIC	AD	K-S	CVM
Rainfall	NORM	358	361.3	0.187	0.088	0.032
	LNORM	362.3	365.6	0.445	0.119	0.07
	WEIB	358.1	364.6	0.207	0.086	0.033
	GEV	**359.7**	**364.6**	**0.176**	**0.076**	**0.029**
	GAMM	359.4	362.7	0.26	0.097	0.043
	GUMB	361.9	365.2	0.445	0.116	0.067
Water Level	NORM	50.27	53.6	0.433	0.097	0.064
	LNORM	50.24	53.57	0.431	0.096	0.063
	WEIB	57.45	60.78	0.845	0.116	0.123
	GEV	50.44	55.43	0.353	0.092	0.051
	GAMM	50.25	53.57	0.431	0.0969	0.0639
	GUMB	**49.15**	**52.48**	**0.366**	**0.087**	**0.052**

Table 3. Critical values for K-S, CVM and A-D goodness of fit tests.

Test	Signficance level	Distribution	Critical value	Source
Kolmogrov-Smirnov (K-S)	0.05	All	0.217	Massey, 1951
Cramer Von Mises (CVM)	0.05	All	0.461	Choulakian et al. 1994
Anderson-Darling (A-D)	0.05	NORM	0.752	D'Agostino and Stephens, 1986
		LNORM	0.752	
		WEIB	0.757	
		GAMM	0.752	
		GUMB	0.757	
		GEV	0.277	Shin et al. 2011

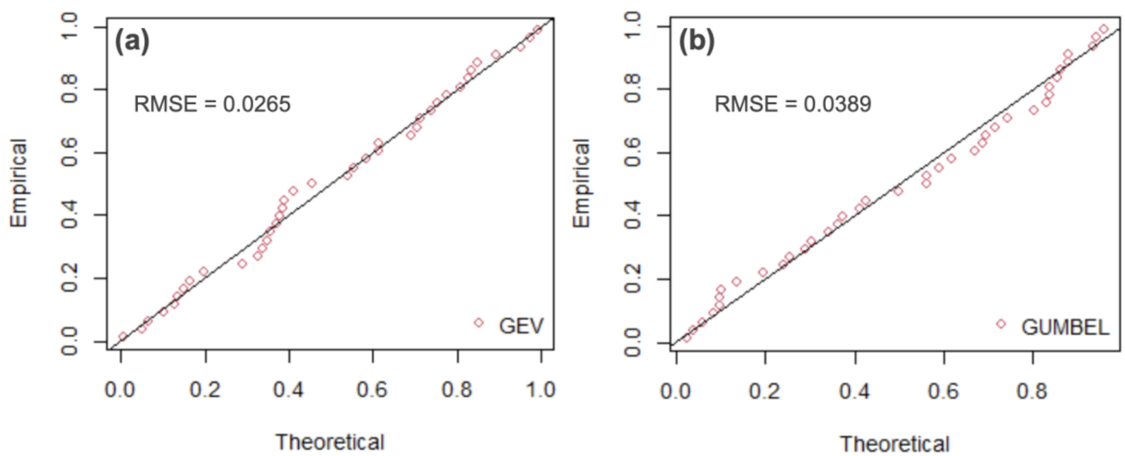

Figure 3. P-P plot of best marginal distribution for (a) rainfall and (b) water level.

4.2 Copula

Archimedean copulas are very suitable for modeling extreme events because of their ability to capture various dependence structures and only require one parameter (Hofert, 2008). Four Archimedean copulas were tested, Gumbel, Clayton, Frank, and Joe to construct the joint distribution between rainfall and water level. The parameters for each copula were calculated using MLE. AIC, BIC, and LogLik were used to test the goodness of fit for the different copulas (Table 4). The best fit was determined by the lowest AIC and BIC value and the highest LogLik value. Frank copula was the best fit and resulted in the lowest AIC and BIC values of 1.19 and 2.86 respectively. Frank copula also had the largest Loglik value of 0.4.

Table 4. Joint distribution goodness of fit for rainfall and water level.

Model Selection Criteria	Copula Family			
	Gumbel	Clayton	Frank	Joe
AIC	1.64	1.27	**1.19**	1.85
BIC	3.3	2.94	**2.86**	3.51
LogLik	0.18	0.36	**0.4**	0.07

4.3 Return Period

The T-year joint return period for rainfall and water level can be useful for flood planning and mitigation strategies for the future. The "OR" Return Period represents the likelihood that either rainfall or water level will meet the threshold shown in Table 5 for the designated univariate return period. The "AND" Return Period represents the likelihood that both rainfall and water level will meet the predicted rainfall and water level threshold value.

The univariate RP is larger than the OR RP and smaller than the AND RP. This indicates that the univariate RP does not provide complete information about the different factors in flood risk (Xu et al., 2019). For example, looking at rainfall and water level for a 10-year univariate RP, the OR RP is 5.35 years and the AND RP is 72 years. This means that when using a copula function, the probability of water level or rainfall exceeding the threshold, there is a greater chance of "flooding" than when examining just one variable.

When looking at the daily observational data from 1982 to 2021, the predicted 50-year rainfall of 123.6 mm with a water level of 177.83 m, has not occurred jointly in the past. For reference, the long-term average water level in Lake Michigan is around 176.5 m (USACE). In fact, out of the 39 years of reliable data available, there have only been five days of rainfall over 100 mm. It is easy to observe how a slight change in rainfall can impact the OR and AND Return Period. For instance, when looking at the 5-year RP, rainfall is 95.82 mm and a water level of 176.94 m, predicts AND RP of 19.3 years. With an increase of 10.49 mm (0.41 inches) of rain and an increase of 0.28 m of water level, the AND RP jumps to 72 years. Figure 4 shows the contour plots for OR RP and AND RP. The OR case is important because it establishes that flooding could be more frequent due to either heavy precipitation or high water levels.

Table 5. Rainfall and water level OR and AND return periods for 2, 5, 10, 20, and 50 years.

Return Period	Rainfall (mm)	Water Level (m)	OR RP (Years)	AND RP (Years)
2	75.82	176.52	1.38	3.6
5	95.82	176.94	2.86	19.3
10	106.31	177.22	5.35	72
20	114.71	177.49	10.4	279
50	123.6	177.83	25.1	1673

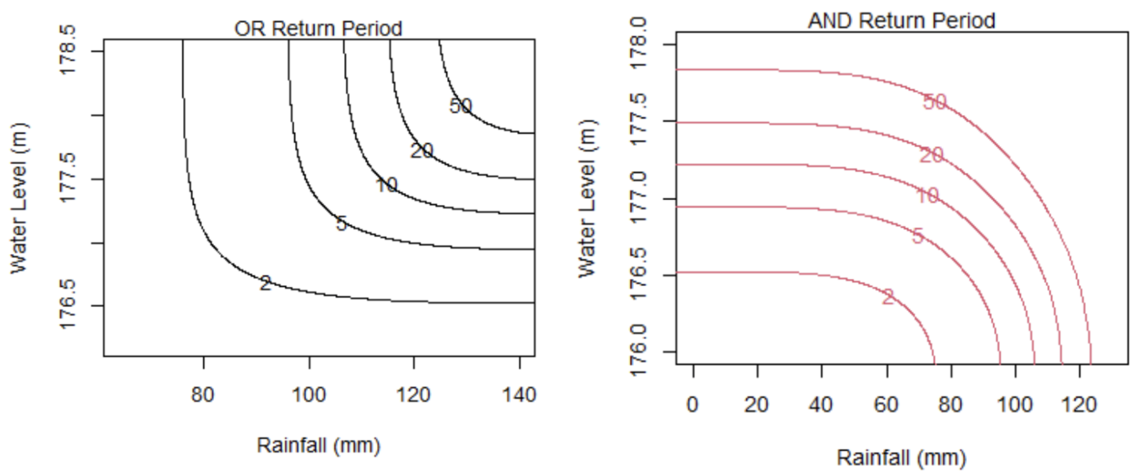

Figure 4. Contour plot of the joint return periods for the OR scenario (left) and the AND scenario (right).

4.3.1 May 17 Event

Table 6 shows the total amount of daily rainfall and daily mean water level leading up to and after the May 17th event. When looking at the 17th of May event, there was 100.33 mm of rain and a mean lake water level of 177.55 m, this falls between the 10 to 20-year univariate RP shown in Table 5. The univariate RP for 100.33 mm of rainfall is about 14 years, while the univariate RP for 177.55 m water level is around 24 years. When examining the RP for both of these values using the copula function the OR RP was 6.11 years and the RP for them to occur jointly was 318 years. When looking at the historical observational data, the May 17th event is also the only event with water level and rainfall at that magnitude to have occurred over the course of 39 years, indicating it was a highly extreme and rare occurrence.

Table 6. Daily total rainfall at Midway Airport Station and daily mean water level for May 14, 2020 to May 19, 2020.

Day	Rainfall (mm)	Daily Mean Water Level (m)
5/14/20	59.94	177.44
5/15/20	8.89	177.42
5/16/20	10.16	177.46
5/17/20	100.33	177.55
5/18/20	0.2	177.59
5/19/20	5.58	177.61

4.3.2 Climate Change

Based on Kayastha et al. (2022), there is a predicted increase of water level due to an increase in over-lake precipitation and basin runoff with a potential increase in lake evaporation. Using the Great Lakes-Atmosphere Regional Model (GLARM; Xue et al., 2017, 2022), the mean water level of Lake Michigan is projected to increase by 0.44 m to possibly 0.8 m from 2040–2049 in Lake Michigan. Using the May 17th event values as an indicator that a reversal event may be able to happen when a major flooding might occur, the daily mean water level of 177.55 m can be looked at as a threshold, when the water level is at or above 177.55 m, a flooding event might occur. When considering climate change, if the climatological mean water level were to increase by 0.2 m while interannual and seasonal variability is assumed to remain the same as shown in historic observational data, an adjusted threshold of 177.35 m would need to be reached for a flood event to occur in the future. Table 7 lists the predicted OR RP and AND RP under climate change with the assumption that water levels could increase by 0.2, 0.45, and 0.8 m, respectively. As water level increases the AND RP decreases, implicating that flooding is more likely to occur in the future under climate changes. Based on historical conditions, such an event is estimated to have a return period of 318 years, however, with a changing climate, such a compound event may occur much more frequently, and return periods may be as short as 15.2 to 66.9 years depending on the future water level rise.

Table 7. Predicted OR RP and AND RP under climate change.

Water Level Increase (m)	Water Level (m)	Rainfall (mm)	OR RP (Years)	AND RP (Years)
0.2	177.35	100.3	4.78	66.9
0.45	177.1	100.3	3.84	36.2
0.8	176.7	100.3	2.3	15.2

5 Discussion and Conclusion

There are two types of reversals into Lake Michigan, gate reversals and lock reversals. Gate reversals are when there is a smaller volume of water, while lock reversals allow a greater volume of water flow into Lake Michigan and usually occur during severe storm events (MWRD). Table 8 lists the reversals from the Chicago Area Waterway system (CAWS) into Lake Michigan, with the amount of effluent discharge from the Chicago River Controlling Works (CRCW) control structure. There are four different control structures in CAWS. The CRCW control structure was chosen because of its central location in downtown Chicago and is the location of interest for the May 17th event. During the 17th of May Chicago flood event, MWRD was unable to reverse the river flow due to a high amount of precipitation and a high lake level, the precipitation received on May 17th was 110.33 mm and the daily mean water level was 177.55 m. One article stated that in addition, if a two-foot storm surge (0.609 m) were to happen, the Chicago lock itself might be useless (Egan, 2021). This event was used to drive the research question and examine the joint probability of this event happening again or more frequently under climate change. Which can even be applied to examining if reversal events will increase in the future. While in communication with a public affairs specialist at Metropolitan Water Reclamation District of Greater Chicago (MWRD), he made it clear that the length of a rainfall event and water level of Lake Michigan does not trigger a reversal, it is based on how much water enters the CAWS. However, it can be inferred that perhaps longer rainfall events do cause more water to enter the CAWS and with the fluctuations of lake water level, there is the great potential for more frequent flooding and reversals in the future. With a higher lake water level MWRD might be unable to reverse the flow of the river causing overflow of the CAWS contributing to more extreme flooding events.

For this study, annual maximum rainfall and corresponding mean daily water levels were examined to estimate the joint probability of the occurrence of high water level and rainfall event. Analysis on two-, three- and four-day cumulative rainfall was also examined, because many severe storm events happen over a course of more than 24 hours. However, when looking at annual maximum cumulative rainfall and corresponding water level, the correlations were weak and resulted in a poor RP output when applying the copula function. Cumulative rainfall should still be considered for future analysis with a larger data set, which may lead to enhanced reliability. The MWRD (who control the water levels in CAWS) have their own set of rain gauges that they use, with data only available from 2016 to present. For a more accurate RP output, it would be helpful to use their rain gauge data when there is more extensive observational data available.

Table 8. Reversals to Lake Michigan from the Chicago Area Waterway System from 1985–2021 (Data from MWRD).

Reversal Events		Rainfall (mm)	Water Level (m)	CRCW Discharged (Million Gallons)
Year	Date			
1987	8/13-8/14	62.992	177.02	986
1990	5/9-5/10	62.992	176.34	208
1990	11/27-11/28	84.836	176.33	86
1996	7/17-7/18	205.232	---	519
1997	2/20-2/22	116.84	176.93	1947
1997	8/16-8/19	77.978	177.22	402
2001	8/2	90.932	176	883
2002	8/22	46.736	176.3	1296.4
2008	9/13-9/16	145.034	176.25	5438.2
2010	7/24	162.306	176.32	5784.6
2011	7/23	58.674	176.29	1716.2
2013	4/18-4/19	114.808	175.83	6104.7
2014	6/30-7/1	70.866	176.4	362
2015	6/15-6/16	77.216	176.78	997
2017	10/14-10/15	137.414	176.93	2456.4
2020	5/17-5/18	110.49	177.57	1731.6

Overall, copulas are a powerful tool that can be used to analyze joint probability for extreme hydrological events. The methods outlined could be applied to other areas to evaluate flooding potential. Chicago is a large metropolitan area that is of high concern of flood risk due to the amount of infrastructure and the high population density. While the TARP project is still underway, Chicago still can face major flood risk in the future, especially under climate change. By using observational data to analyze joint probability, a better understanding can be developed for the area in terms of high precipitation and high water level. While it is still uncertain what the future will hold under climate change, copulas can be used as a simple predictive tool. Chicago has an underground reservoir system, implemented under the Tunnel and Reservoir Plan (TARP) that can store excess water for heavy rainfall events that might cause flooding. There are two phases to this project; the first phase was completed in 2006. Phase II is currently underway with two out of the three reservoirs complete and the third is expected to be completed in 2029 (MWRD). With the completion of Phase II, it would add 15.15 billion gallons of storage for combined sewer overflows. But MWRD states that it could still be possible for the system to reach capacity during an extreme storm event.

6 Reference List

Anderson, E.J., Stow, C.A., Gronewold, A.D., Mason, L.a., McCormick, M.J., Qian, S.S., Ruberg, S.A., Beadle, K., Constant, S.A., and Hawley, N. (2021). Seasonal overturn and stratification changes drive deep-water warming in one of Earth's largest lakes. *Nature Communications 12*. https://doi/org/10.1038/s41467-021-21971-1.

Asquith, W., Kiang, J.K., and Cohn, T.A. (2017). Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigations Report 2017–5038, 93 p.5 appendices, https://doi.org/10.3133/sir20175038.

Austin, J.A., and Colman, S.M. (2007). Lake Superior summer water temperatures are increasing more rapidly than regional air temperatures: A positive ice-albedo feedback. *Geophysical Research Letters 34*. https://doi/org/10.1029/2006GL029021.

Bevacqua, E., Maraun, D., Hoaek Haff, I., Wimann, M., Vrac, M. (2017). Multivariate statistical modeling of compound events via pair-copula constructions; analysis of floods in Ravenna (Italy). *Hydrological Earth Systems Science 21(6),* 271–272. https://doi.org/10.5194/hess-21-2701-2017.

Brosnan, K. A., Barnett, W. C., & Keating, A. D. (2020). *City of Lake and Prairie: Chicago's Environmental History*. University of Pittsburgh Press. https://doi.org/10.2307/j.ctv1595nc9

Chen, X., Shao, Q., Xu, C-Y., Zhang, J., Zhang, L., and Ye, C. (2017). Comparative Study on the Selection Criteria for Fitting Flood Frequency Distribution Models with Emphasis on Upper -Tail Behavior. *Water 9*(5), 320. https://doi.org/10.3390/w9050320.

Chen, L., and Guo, S. (2019). *Copulas and Its Application in Hydrology and Water Resources*. Springer. https://link.springer.com/content/pdf/10.1007/978-981-13-0574-0.pdf.

Choulakain, V., Lockhart, R.A., and Stephens, M.A. (1994). Cramer-von Mises Statistics for Discrete Distributions. *The Canadian Journal of Statistics 22*(1), 125–137. https://doi.org/10.2307/3315828.

Duncker, J.J., and Johnson, K.K. (2016). Hydrology of and Current Monitoring Issues for the Chicago Area Waterway System, Northeastern Illinois (ver,1.1 March 2016): U.S Geological Survey Scientific Investigations Report 2015–5115, 48 p., https://doi.org/10.3133/sir20155115.

D'Agostino, R.B., and Stephens, M.A. (1986). *Goodness of Fit Techniques*. Marcel Dekker, Inc.

d'Orgeville, M. Peltier, W.R., Erler, A.R., and Gula, J. (2014). Climate change impacts on Great Lakes Basin precipitation extremes. *Journal of Geophysical Research: Atmospheres 119* (18), 10779–10812. http://doi.org/10.1002/2014JD021855.

Egan, D. (2021). The climate crisis haunts Chicago's future. A Battle Between a Great City and a Great Lake. *New York Times.* Retrieved 16 March 2022, from https://www.nytimes.com/interactive/2021/07/07/climate/chicago-river-lake-michigan.html.

Environment and Climate Change Canada and the U.S. National Oceanic and Atmospheric Administration. 2020 Annual climate Trends and Impacts Summary for the Great Lakes Basin. 2021. Available at https://binational.net.

Gronewold, A.D., Do, H.S., Mei, Y., and Stow, C.A. (2021). A tug-of-war within the hydrologic cycle of a continental freshwater basin. *Geophysical Research Letters 48* (4). https://doi.org/10.1029/2020GL090374

Gronewold, A.D., Bruzer, J., Durnford, D., Smith, J.P., Clites, A.H., Segienieks, F., Qian, S.S., Hunter, T.S., and Fortin, V. (2016). Hydrological drivers of record-setting water level rise on Earth's largest lake system. *Water Resource Research* 52 (5), 4026–4042. https://doi.org/10.1002/2015WR018209.

Gronewold, A., Fortin, V., Lofgren, B., Clites, A., Stow, C.A., and Quinn, F. (2013). Coasts, water levels and climate change: A Great Lakes perspective. *Climatic Change* 120:697–711. http://doi.org/10.1007/s10584-013-0840-2.

Gronewold, A.D., and Rood, R.B. (2018). Recent water level changes across Earth's largest lake system and implications for future variability. *Journal of Great Lakes Research* 45 (1), 1–3. http://doi.org/10.1016/j.jglr.2018.10.012.

Gronewold, A.D., and Stow, C.A. (2014). Water Loss from the Great Lakes. *Science 343* (6175), 1084–1085. http://doi.org/10.1126/science.1249978

Grunderson, Erica. (2020, July 6). New Analysis Shows Large Swaths of Chicago at High Flood Risk.Window to the Word (WTTW) https://news.wttw.com/2020/07/06/new-analysis-shows-large-swaths-chicago-high-flood-risk

Hayhoe, K., VanDorn, J., Croley II, T., Schlegal, N., and Wuebbles, D. (2010). Regional climate change projects for Chicago and the U.S Great Lakes. *Journal of Great Lakes Research* 36: 7–21. http://doi.org/10.1016/j.jglr.2010.03.012.

Hofert, M. (2008). Sampling Archimedean copulas. *Computational Statistics & Data Analysis 52*(12), 5163–5174. https://doi.org/10.1016/j.csda.2008.05.019

Huang, C., Anderson, E., Liu, Y., Ma, Gangfeng, Mann, G., and Xue, P. (2021). Evaluating essential processes and forecast requirements for meteotsunami-induced coastal flooding. *Natural Hazards* 110, 1693–1718. https://doi.org/10.1007/s11069-021-05007-x.

Hurk, B.v.d., Meijgaard, E.v., Vald, d.V., Herringen, KJ. v., and Gooijer, J. (2015). Analysis of compounding surge and precipitation event in the Netherlands. *Environmental Research Letters* 10. https://doi.org/10.1088/1748-9326/10/3/035001.

Illinois Department of Natural Resources. (2015 June). Report for the Urban Flooding Awareness Act. https://www2.illinois.gov/dnr/WaterResources/Documents/Final_UFAA_Report.pdf

Intergovernmental Panel on Climate Change. (2022). *Report on Climate Change 2022: Impacts, Adaptations, and Vulnerability*. Cambridge University Press. https://report.ipcc.ch/ar6/wg2/IPCC_AR6_WGII_FullReport.pdf

Jane, R., Cadavid, L., Obeysekera, J., Wahl,T. (2020). Multivariate statistical modeling of the drivers of compound flood events in south Florida. *National Hazards Earth System Science* 20(10), 2681–2699. https://doi.org/10.5194/nhess-20-2681-2020

Kayastha, M.B., Ye, X., Huan, C., and Xue, P. (2022). Future rise of the Great Lakes water level under climate change. *Journal of Hydrology* 612: Part B. https://doi.org/10.1016/j.jhydrol.2022.128205.

Laio, F., Baldassarre, G.D., and Montanari, A. (2009). Model selection techniques for the frequency analysis of hydrological extremes. *Water Resource Research 45.* http://doi.org/10.1029/2007WR00666666.

Lanyon, R. (2013). Building the Canal to Save Chicago. *World Environmental and Water Resources Congress* 21–31

Lian, J., Xu, K., and Ma, C.(2013). Joint impact of rainfall and tidal level on flood risk in a coastal city with a complex river network: a case study of Fuzhou City, China. Hydrol Earth Syst Sci 17(2): 679–689. https://doi.org/10.5194/hess-17-679-2013

Lofgren, B.M., Quinn, F.H., Clites, A.H., Assel, R.A., Eberhardt, A.J., and Luukkonen, C.L. (2002). Evaluation of Potential Impacts on Great Lakes Water Resources Based on Climate Scenarios of Two GCMS. *Journal of Great Lakes Research 28*(4), 537–554. https://doi.org/10.1016/S0380-1330(02)70604-7

Markus, M., Wuebbles D.J., Liang, X-X., Hayhoe, K., and Kristovich, D.A.R. (2012). Diagnostic analysis of future climate scenarios applied to urban flooding in the Chicago metropolitan area. *Climatic Change* 111:879–902. https://doi.org/10.1007/s10584-011-0172-z

Massey, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statstical Association, 4*(6), 70. https://doi.org/10.2307/2280095.

Mason, L.A., Riseng, C.M., Gronewold, A.D., Rutherford, E.S., Wang, J., Clites, A., Smith, S.D., and McIntyre, P.B. (2016). Fine-scale spatial variation in ice cover and surface temperature trends across the source of the Laurentian Great Lakes. *Climate Change* 138, 71–83. http://doi.org/10.1007/s10584-016-1721-2.

Metropolitan Water Reclamation District of Greater Chicago. (n.d). *About Chicago Area Waterway System.* https://mwrd.org/chicago-area-waterway-system.

Metropolitan Water Reclamation District of Greater Chicago. (31 Oct. 2022). *TARP at 50: How one of the world's largest public works projects for water has protected the Chciago Region.* https://mwrd.org/tarp-50-how-one-worlds-largest-public-works-projects-water-has-protected-chicago-region

Metropolitan Water Reclamation District of Greater Chicago.(n.d). *Tunnel and Reservoir Plan (TARP).* https://mwrd.org/tunnel-and-reservoir-plan-tarp

Metropolitan Water Reclamation District of Greater Chicago.(31 Dec. 2022). *TARP Status Report as of December 31, 2022. https://mwrd.org/sites/default/files/documents/TARP_Status_Report_Dec2022.pdf*

O'Reilly, C.M., Sharma, S., Gray, D.K., and Hampton, S.E. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters 42*(24), 10773–10781. https://doi.org/10.1002/2015GL066235.

Pettitt, A.N. (1976). A Two-Sample Anderson-Darling Rank Statistic. *Biometrika 63*(1), 161–168. https://doi.org/10.2307/2335097.

Pirani, F.J., and Najafi, M.R. (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.

Saharia, A.M., Zhu, Z., and Atkinson, J.F. (2021). Compound flooding from lake seiche and river flow in a freshwater coastal river. *Journal of Hydrology* 608(B). https://doi.org/10.1016/j.jhydrol.2021.126969.

Salvadori, G., and De Michele, C. (2007). On the Use of Copulas in Hydrology: Theory and Practice. *Journal of Hydrologic Engineering 12*(4). https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(369).

Seneviratne, S.I., N. Nicholls, D. Easterling, C.M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, and X. Zhang, 2012: Changes in climate extremes and their impacts on the natural physical environment. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY , USA, pp. 109–230

Shin, H., Jung, Y., Jeong, C., and Heo, Jun-Haeng. (2012). Assessment of modified Anderson-Darling test statistics for the generalized extreme value and generalized logistic distribution. *Stochastic Environmental Research and Risk Assessment 26*(1), 105-114. http://doi.org/10.1007/s00477-011-0463-y.

Sklar, A. (1959). Fonctions de répartition à n dimensions de leurs marges. *Publ. Inst. Stat. Univ. Paris* 8: 229–231.

Stephens, M.A. (1979). The Anderson-Darling statistic. Technical Report No. 39. Stanford University.

Wang, J., Bai, X., Hu, H., Clites, A., Colton, M., Lofgren, B. (2012). Temporal and Spatial Variability of Great Lakes Ice Cover, 1973–2010. *Journal of Climate 25*, 1318-1329. https://doi.org/10.1175/2011JCLI4066.1.

Wang, Y., Li, C., Liu, J., Yu, F., Qiu, Q., Tian J., and Zhang, M. (2017). Multivariate Analysis of Joint Probability of Different Rainfall Frequencies Based on copulas. *Water* 9(3), 198. https://doi.org/10.3390/w9030198.

Wahl, T., Jain, S., Bender, J., Meyers, S.D., Luther, M.E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *National Climate Change 5(12)*, 1093–1097.https://doi.org/10.1038/nclimate2736.

Wisconsin Sea Grant. *Storm Surges, Seiches and Edge Waves.* University of Wisconsin. https://www.seagrant.wisc.edu/our-work/focus-areas/coastal-processes-and-engineering/coastal-processes/storm-surges-seiches-and-edge-waves/. (accessed 5 July 2022).

Wuebbles, D., Cardinale, B., Cherkauer, K., Davidson-Arnott, R., Hellmann, J., Infante, D., and Ballinger, A.(2019). An assessment of the impacts of climate change on the Great Lakes, Environmental Law & Policy Center, https://elpc.org/wp-content/uploads/2020/04/2019-ELPCPublication-Great-Lakes-Climate-Change-Report.pdf (last access: 5 July 2022).

Xue, P., Ye, X., Pal, J.S., Chu, P.Y., Kayastha, M.B., and Huang, C. (2022). Climate projections over the Great Lakes Region: using two-way coupling of a regional climate model with a 3-D lake model. *Geoscientific Model Development 15*(11), 4425-4446. https://doi.org/10.5194/gmd-15-4425-2022.

Xue, P., Pal, J.S., Ye, X., Lenters, J.D., Huang, C., and Chu, P.Y. (2017). Improving the Simulation of Large Lakes in Regional Climate Modeling: Two-Way Lake-Atmosphere Coupling with a 3D Hydrodynamic Model of the Great Lakes. *Journal of Climate 30*(5), 1605–1627. https://doi.org/10.1175/JCLI-D-16-0225.1

Zeng, X., Wang, D., and Wu, J. (2015). Evaluating the Three Methods of Goodness of Fit Test for Frequency Analysis. *Journal of Risk analysis and Crisis Response 5*(3), 178-187. https://doi.org/10.2991/jrarc.2015.5.3.5.

Zellou, B., and Rahali, H. (2019). Assessment of the joint impact of extreme rainfall and storm surge on the risk of flooding in a coastal area. *Journal of Hydrology* 569, 647–665. https://doi.org/10.1016/j.jhydrol.2018.12.028.