Great Lakes Ice Cover: Enriching Database and Improving Forecast

Part A: Extending Data Record Back to the 1890s with Identification of Key Teleconnection Patterns

Client: NOAA Great Lakes Environmental Research Laboratory (GLERL)
Corporate Institute for Great Lakes Research (CIGLR)
Master's Project #P19

School for Environment and Sustainability (SEAS) | University of Michigan, Ann Arbor
Danielle Cohn, Miye Nakashima, Inigo Peng
Dr. Ayumi Fujisaki-Manome | Dr. Philip Chu | Dr. Yuchun Lin
April 21, 2021

A project submitted in partial fulfillment of the requirements for the degree of
Master of Science in Natural Resources and the Environment
at the University of Michigan, Ann Arbor.
April 2021

EXECUTIVE SUMMARY

Great Lakes ice cover has been regularly documented and reported only since the 1970s, while inconsistent records of ice cover began in the 1960s. These ice charts from the 1960s and early 1970s had yet to be digitized from physical prints and preliminary scans into a computer- and web-friendly format. Aerial surveys and satellite imagery of the ice cover started in the 1960s; there were very few records of Great Lakes ice cover before these surveys and ice charts began. Collection of weather data (temperature, wind, precipitation, etc.) began in the 1800s in the region, but ice cover reports were sparse and difficult to estimate. Using the surface air temperature data from 1897 to 1983, ice cover can be estimated and hindcasted back to the start of the weather record for the Great Lakes. Another atmospheric component that influences ice cover on the lakes is that of atmospheric teleconnections, such as the ENSO (El Niño Southern Oscillation), NAO (North Atlantic Oscillation), and newer ABNA (Asian-Bering-North American). While larger, more well-known teleconnections such as the ENSO and NAO have been analyzed next to Great Lakes ice cover, ABNA had yet to be compared to the lakes' annual ice cover.

These gaps in the collection of Great Lakes ice research were filled through this collaborative project between the University of Michigan School for Environment and Sustainability (SEAS) and NOAA's Cooperative Institute for Great Lakes Research (CIGLR). The historical ice charts from 1963 to 1972 are now digitally available through the University of Michigan's Deep Blue repository for this project; the Great Lakes' ice cover has been hindcasted back to the winter of 1898, available here in Appendix II; and the ABNA has been recreated and statistically compared to the Great Lakes' Annual Maximum Ice Cover (AMIC) values, and has proven to be a strong contender in forecasting and hindcasting ice cover on the Great Lakes.

Great Lakes ice cover provides various ecosystem services in the Great Lakes, from tourism, to ice caves, to supporting the spawning of fish by protecting their eggs from wave action. At the same time, ice cover is a significant obstacle for winter navigation in the Great Lakes. Federal and commercial icebreaking operations are thus a vital aspect of wintertime shipping within the Great Lakes. Knowing ice conditions ahead of the winter season, or even with a longer lead time, is critical in planning in all these sectors.

However, seasonal or longer forecasting of Great Lakes ice conditions has been challenging because of the strong year-to-year fluctuations in AMIC and little knowledge in what teleconnection pattern(s) influence the weather across the Great Lakes region. The deliverables of our project – the extended time series of AMIC, and identification of the ABNA as an important teleconnection pattern – well address these needs and provide a strong foundation to improve seasonal Great Lakes ice forecasting at NOAA Great Lakes Environmental Research Laboratory, the client of this project.

CONTENTS

ABSTRACT

The Laurentian Great Lakes are home to millions of Americans and supply drinking water to over 40 million people in both Canada and the United States, as well as recreational opportunities, commerce, and the region's unique climate. An essential part of the Great Lakes' annual water cycle is winter ice cover, which has generally been decreasing in recent decades as anthropogenic climate change advances further. It is vital to understand the historical ice cover of the Great Lakes in order to better understand this and forecast future ice cover on the lakes. Historical ice charts from 1963-1972 were collected and digitized for virtual accessibility and storage, as well as to calculate each lake's annual maximum ice cover (AMIC) for these years. These ice cover values were then used with historical air temperature data to create AMIC models for each of the Great Lakes. Historical air temperature data for the years 1898-1983 were collected and manipulated into two different temperature proxies: cumulative freezing degree-days (CFDD) and net melting degree-days (NMDD). The chosen temperature proxy used for each lake was dependent on the lakes' individual ice cover and temperature trends. The same analysis was applied to the most recent weather and ice data for the period of 2009-2020 to expand the AMIC model. The hindcasted AMIC values for all five lakes for 1980-1983 and 2009-2020 were then compared to various climate indices for teleconnection patterns affecting North American weather, including the Pacific/North American teleconnection pattern, the North Atlantic Oscillation, and the Asian-Bering-North American (ABNA) index. The ABNA is an atmospheric teleconnection that influences temperature and pressure over the Great Lakes, Bering Strait, and Asia. The monthly ABNA index was recreated to ensure its replicability and stability. This study uses historical data while integrating new methods of analyses with traditional ones in order to develop a hindcast of the Great Lakes AMIC that will provide a better understanding of how this lake system develops ice coverage each winter. Percent ice cover values from the historical ice charts were calculated, likely with increased accuracy from the original documents, which were used here and can be used in future analyses. Models for each of the lakes were created to hindcast historical AMIC values, from moderate to high accuracy and R^2 values. The AMIC values from these estimation models can be used in future analyses as well, and were used to determine the correlation between the ABNA index and AMIC on the Great Lakes. A moderate correlation was found between the ABNA index and AMIC for the Great Lakes, indicating that the ABNA index may serve as another way to estimate Great Lakes ice cover annually.

ACKNOWLEDGEMENTS

First, we would like to thank Dr. Ayumi Fujisaki-Manome for creating this project and being an incredible resource and guide throughout our time working on it. Even through the pandemic and having to make everything remote, this research experience has been a wonderful experience; it has taught us so much about the Great Lakes' ice and much more! Thank you, Ayumi, for being such a wonderful and kind mentor.

We would also like to thank Dr. Yuchun Lin for his help in doing the meteorological data analysis. Your expertise in atmospheric modeling was invaluable in helping us to reconstruct the Asian-Bering-North-America Teleconnection Index.

Thanks is also due to the School for Environment and Sustainability (SEAS) for choosing this Master's Project as one of those offered for our cohort to choose from. Additionally, having the pandemic shutting down in-person learning and meetings was difficult to navigate, especially with Master's Projects, so thank you for guiding us and our cohort through the process.

Finally, we wish to thank Dr. Philip Chu and NOAA GLERL, our client contact and client for this project. Thank you for your support and insight throughout the development and completion of our project. We sincerely hope the products of this research will be of great use to GLERL and others in the future!

INTRODUCTION, BACKGROUND, & RESEARCH SIGNIFICANCE

The Laurentian Great Lakes are a vital part of the Midwest and Northeast United States for commerce, recreation, and most importantly, regional climate. Each winter, the Great Lakes develop ice cover, which affects these aspects of human life in the Great Lakes Basin as well as the regional wildlife and seasonal processes (Dempsey et al. 2008). The lakes themselves are affected in turn by a variety of factors, such as atmospheric teleconnections, the lakes' ice cover, lake evaporation in previous seasons (Van Cleave et al. 2014), and, unfortunately, anthropogenic climate change (Kling et al. 2003). It is therefore important to understand the impact of seasonal temperatures on the lakes' ice cover, especially as overall winter temperatures are becoming gradually warmer, and potentially less predictable, with each passing year.

Ice cover on the Great Lakes has been regularly documented since the early 1960s, with improving chart coverage since the late 1960s. Since the start of the satellite era in the 1970s, determining ice cover on the Great Lakes became much more feasible and accurate, especially with minimal cloud cover (Assel, 1972, 1986, 2003; Wang et al. 2018). Having this lengthy record of ice cover from the 1960s to today allows for ice cover models to be compiled for each of the five lakes, based on their influencing factors. The older ice charts, compiled by Donald R. Rondy, Raymond A. Assel, and R. E. Wilshaw in the 1960s and 1970s (Rondy, 1966-1972; Assel, 1972; Wilshaw and Rondy, 1965), have only been available thus far as paper charts, scanned images, or ebooks, and before this study, had yet to be converted to web-friendly, exportable formats that could be analyzed further. Converting these to digital ice charts also aided in estimating historical ice cover for the Great Lakes for the years 1963-1972. It should be noted that the ice cover on these charts is not necessarily representative of the AMIC for each of the lakes, as surveys could not be conducted on a daily basis.

The Great Lakes' ice cover is affected primarily by air temperature, which influences the lake water temperature. When there are several cold days in a row or a few extreme cold days over the Great Lakes, the threshold for cumulative freezing-degree days (CFDDs) may be met for a given lake, and ice cover can start to form and grow. This study uses methods developed and used in previous research, and expands these methods to estimate historical AMIC for each of the Great Lakes. CFDD was used in previous studies (Assel, 1986; 2003) to determine the severity of a given winter season, but was not explicitly used to estimate historical ice cover for the Great Lakes. Similarly, NMDD values were used in a much smaller scope; they were used by Hewer & Gough (2019) to hindcast AMIC values for Lake Ontario, based only on temperature data from Toronto, ON for the years 1840-2019. The work done here widens the scope of these previous studies to estimate historical AMIC, based on CFDD or NMDD values, for all five of the Great Lakes from 1898-1983.

North American winter temperature variations have primarily been attributed to large-scale atmospheric circulation patterns such as Pacific North American (PNA) teleconnection pattern, North Atlantic Oscillation (NAO) pattern, and, indirectly, the El

Niño Southern Oscillation (ENSO). Geographically, the Great Lakes are located on the edge of the two patterns; hence neither indices have shown a strong correlation with AMIC independently. A slight distortion of the pattern and shift of the atmospheric circulation centre may result in different ice cover responses (Assel and Rodionov 1998). Other indices such as Arctic Oscillation (AO), Pacific Decadal Oscillation (PDO), and West Pacific (WP) have also shown to be associated with anomalous ice cover on the Great Lakes (Assel and Rodionov 1998). Bai and Wang (2012) showed negative NAO /AO phase and negative PNA phase is associated with severe ice cover. A positive PNA, El Nino, and NAO/AO are related to mild ice cover. Much of the current studies also focus on how ENSO variability affects the North American climate (Trenberth et al. 1998). Some of the above indices, such as the PNA pattern, are heavily influenced by ENSO events. Improving our understanding of tropical sea surface temperature (SST) variation is important in enhancing NA climate prediction. However, recent studies have also shown that extratropical circulation patterns, snow cover and SST anomalies that are not directly attributable to ENSO have a significant effect on NA climate variability. Multiple studies have found that stationary Rossby waves play an essential role in interannual climate variability from Eurasia to NA (Yu et al. 2018; Ding et al. 2011; Wu et al. 2009).

In the study done by Yu et al. (2018), they showed the third atmospheric teleconnections, termed Asian - Bering - North American (ABNA) patter, that heavily affect NA winter temperature variability. Yu et al. (2016) developed the ABNA index and found that this large atmospheric circulation pattern has an anomalous center over the Great Lakes Region. There are two other anomalous centers, one located over the Bering Strait, and the other over Siberia/Eurasia. With ABNA's strong influence over central NA winter temperature variability and geopotential anomaly centre directly over the Great Lakes, we speculate it may have a strong effect on Great Lakes AMIC. This study component is novel in that it draws a connection between the recently-identified ABNA index and historical AMIC values for the Great Lakes. While many other atmospheric teleconnections have been compared to the Great Lakes ice cover, the ABNA index has yet to be compared; it appears very promising and well-correlated with the AMIC values calculated for this study. If the ABNA shows a strong, significant correlation with AMIC seasonally, it may well improve the seasonal outlook for Great Lakes ice.

MAPPING ICE COVER 1963-1972

NOAA's Great Lakes Environmental Research Laboratory currently holds digitized ice cover records for each lake ranging from 1973 to 2021. This data is available both as time series values and maps of the estimated date of maximum ice cover. These maps contain detailed ice cover concentration percentages to the nearest 5%, and also estimate the maximum ice cover period to one specific date.

The 1963-1972 data digitized with ArcPro in this project from the U.S. Lake Survey Center and National Ocean Survey reports, however, come from less advanced resources and contain ice cover concentration percentages in 20% brackets. Similarly, full-lake surveys could not be done in a single day and were conducted over the course of a week or more, based on predictive estimates of ideal periods to survey the lake ice freezing and melting timeframes.

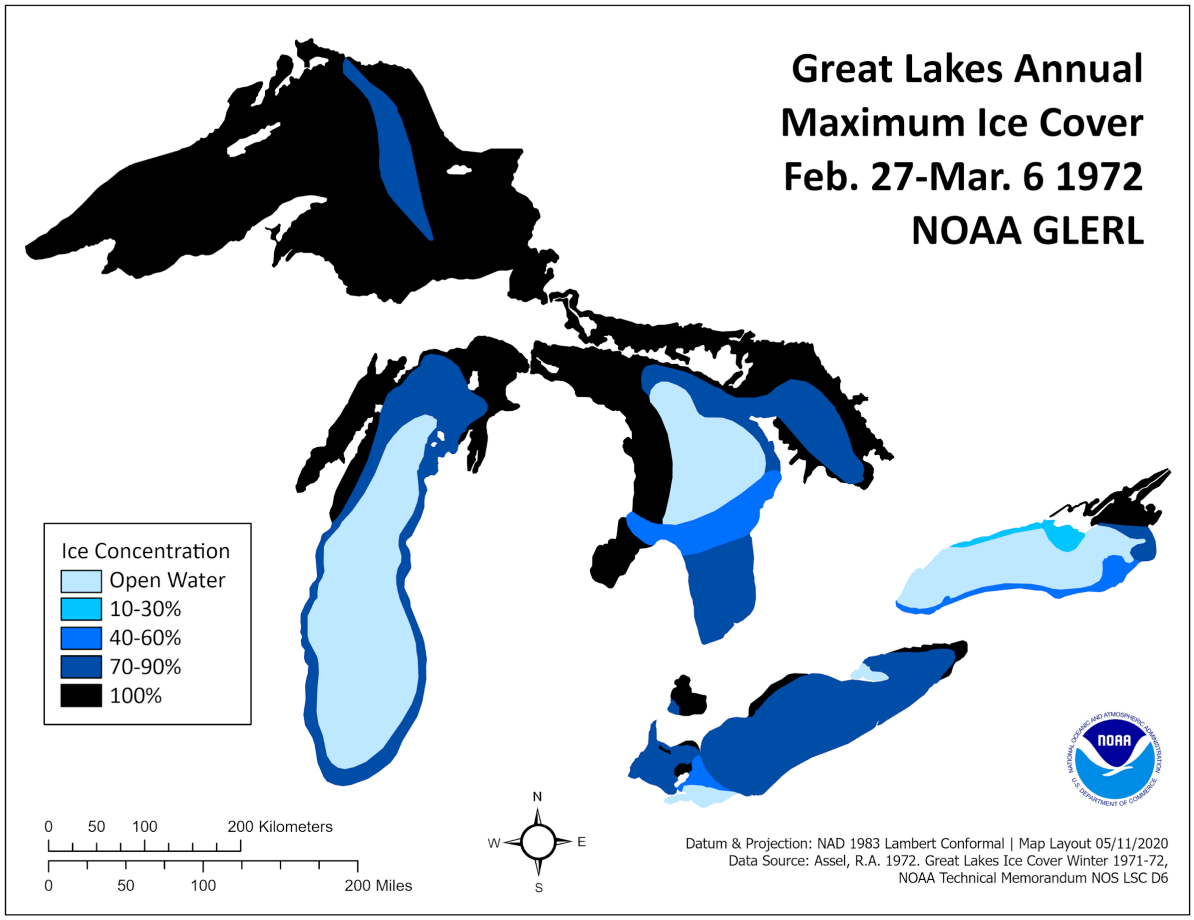


Fig. 1 *Sample of a new map generated from old reports [other years may be found in the Appendix].*

A discrepancy between the older data being added in this report and the current 1973 and beyond data is that the current data selects a single date which represents the overall maximum ice cover for the Great Lakes combined. However, each of the Great Lakes has a slightly different freezing period, and they each reach maximum ice cover at different times. Therefore, this presently recorded data is not a representation of each

individual lake's maximum ice cover, but rather the ice cover during the period of combined maximum lake ice cover. The older data instead chooses two separate periods during which half the lakes are most frozen. This may cause discrepancies in consistency when trying to compare older ice maximums with the current maps.

Unfortunately, even with the old reports gathered, the data remains incomplete. The page with the Northern map for 1966 is missing from the report, which would have been for March 17-21. Additionally, despite pre-1971 maps containing two maps, for each year, 1965 only has one, due to no data being captured for Lake Ontario.

From these digitized maps, total ice cover surface area was calculated and can be found in the Appendix. The results did not precisely match the estimated maximum ice cover described in the reports from which they came (Rogers, 1976). The method used in the reports to generate their numbers is unspecified, but was likely not derived from the calculation of area based on their images.

HINDCASTING HISTORICAL ICE COVER, 1898-1983

A. Data

Historical weather data was collected from the National Snow and Ice Data Center (NSIDC), compiled by Raymond Assel in the 1980s (Assel, 1995), to calculate surface air temperature proxies for 25 stations (see Fig. 9) around the Great Lakes for the years 1897-1983. Unfortunately, some of the data from this source contained duplicate data for a couple of the twenty-five stations due to missing data being replaced by data from a precursory or successive year at the same station. Two different proxies for surface air temperature were calculated using this data: 1) Cumulative freezing-degree days (CFDD), as used by Assel in several studies, were used to approximate the severity of a given winter based on how many days' average temperatures were below-freezing, and how far below freezing they were (see Eq. 1 on following page). 2) Net melting degree days (NMDD), as developed and used by Hewer and Gough (2019), were also used to approximate the severity of a given winter, but are based on daily temperatures both above- and below-freezing temperatures; the mathematical difference between the ice season's melting degree-days (above-freezing temperatures) and freezing degree-days (below-freezing temperatures) results in the NMDD value (Eq. 2). Which temperature proxy was used for each lake was determined based on the correlation between the ice cover estimates (calculated with each of the two proxies) and the actual ice cover reports. Test data was collected for the years 1963-1972, as well as two extra years for Lake Ontario for which there were anecdotal records of full ice cover on the lake. These anecdotal values were published by Hewer & Gough (2019) in their study of Lake Ontario's ice cover trends using NMDD values.

5500	1	1897	44	35	49	37	45	38	42	26
5500	2	1897	30	15	33	19	29	19	31	13
5500	3	1897	27	2	21	12	24	18	31	8
5500	4	1897	42	31	46	26	38	32	40	32
5500	5	1897	38	34	41	35	44	38	68	40
5500	6	1897	58	37	49	43	72	46	65	51
5500	7	1897	72	56	73	62	77	64	98	74
5500	8	1897	73	56	75	60	78	62	71	61
5500	9	1897	65	55	64	50	64	44	75	50
5500	10	1897	63	52	59	51	63	49	76	56
5500	11	1897	48	40	45	38	60	34	60	42
5500	12	1897	22	15	23	11	31	17	36	31
5500	1	1898	17	3	25	16	23	13	33	23

Fig. 2 *Sample of Alpena, MI temperature data from NSIDC (Assel, 1995). Daily high and low temperatures are provided by month.*

$$FDD_j = \sum_{i=Oct.\,1}^{Apr.\,30} (-T_i) \qquad (1)$$

where T_i = surface air temperature departure from freezing (32°F, 0°C) for a given date (Assel, 1980). Days that are colder than freezing have positive FDD values. When the value for CFDD becomes negative, it resets to zero and begins accumulating again.

$$NMDD = MDD - FDD \qquad (2)$$

where MDD is melting degree-days, and FDD is freezing degree-days. In this case, FDD does not use the inverse of the surface air temperature (T_i).

B. Analysis methods

a. Initial analysis

Both Python and Microsoft Excel were used in conjunction to calculate temperature proxies, develop ice cover models based on the temperature proxies, and hindcast historical ice cover back to the winter season of 1897-1898 (referred to by the second year of each winter season). Python was accessed on a MacBook through the programming applications Anaconda and Spyder, base 3.7.6, and through Anaconda3 on Windows. Both CFDD and NMDD values were used to model and estimate historical ice cover for each of the Great Lakes as a whole before determining the stronger option. Different time spans were also tested to determine the most influential months of the winter season on ice cover for each of the lakes. Lastly, different regression models (using lines of best fit) were tested based on each of the lakes' ice cover and temperature proxy scatterplots. These model tests were done for the years 1963-72 ("test data") and 1963-83 ("complete test data"), with the 1963-72 test results considered more heavily in choosing the model for each lake. The 1963-83 test data were considered primarily when the 1963-72 model tests showed low R^2 values, such as in Lake Erie's models.

Based on the scatterplots for each Lake that compare both the CFDD-based and the NMDD-based estimated ice covers to the available reported ice cover data, the models used for each lake are the following: Lake Superior used a linear regression model based on NMDD values from November to February; Lake Michigan used an

exponential regression model based on NMDD values from November to February; Lake Huron used a linear regression model based on CFDD values from October to May; Lake Erie used a piecewise linear model based on CFDD values, with a cutoff value of 352 freezing degree-days, from October to May; and Lake Ontario used an exponential regression model based on NMDD values from January to March.

Prior to finding Hewer and Gough's 2019 study, it was noted that Lakes Erie and Superior had intriguing outliers in their scatterplots comparing AMIC and maximum seasonal CFDD values. In trying to account for these outliers, a method very similar to NMDD was tested to explain these odd values. The CFDD calculation method used so frequently by Assel was modified here to *not* reset when accumulated FDD values became negative (i.e., warm temperatures for several days in a row or extremely warm temperatures occurred), and the modified CFDD values were able to drop below zero to display warming trends in a given winter season. These modified CFDD values were unable to clear Superior or Erie of their outlier data, and therefore were unable to account for it. The only improvement came in Lake Superior's R^2 values in its scatterplot comparing modified CFDD values and maximum ice cover. The main difference between the above method, developed here, and Hewer and Gough's (2019) is that NMDD values have opposite signs to this method; using NMDD, negative values logically indicate colder temperatures, but using this modified CFDD method, negative values indicate warmer temperatures. In realizing the similarity between the two methods, NMDD values were tested for each lake rather than modified CFDD values, based on the peer-reviewed nature of NMDD.

An additional method used in an attempt to improve these ice cover models was to add additional weather station data from Environment Canada. The twenty-five stations used by Assel, available on the NSIDC website, are heavily weighted towards American stations, despite Canada sharing approximately half of the Great Lakes shoreline. This method proved unhelpful in improving the models for ice cover, based on the R^2 values for each lake that this method was tested on. The original data was therefore used for this section of the study.

b. Additional multivariate analysis

This analysis was done to provide additional insight into the connections between some of the variables investigated in the Master's Project, as well as provide other variables to potentially investigate further. Ice cover data (AMIC), temperature proxy data (both CFDD and NMDD), and the new ABNA index values from the main project were used in the secondary analysis. Whether CFDD or NMDD was used was dependent upon which equation was used for each lake (e.g., Lake Erie uses CFDD so that was used; Lake Superior uses NMDD so that data was used). New data used in the analysis include

LakeYear	Lake	CFDDNMDD	IceCover	WindSpd	Evap	ABNA
Ont51	Ontario	-320.875	26.4	7.34	206.48	1.329976557
Ont52	Ontario	-356.375	28.3	7.66	233.32	0.44528
Ont53	Ontario	-81.75	16.3	7.71	278.23	1.179315734
Ont54	Ontario	-397.5	30.7	7.39	231.42	1.217380782
Ont55	Ontario	-537	40.6	7.11	210.58	0.818067501
Ont56	Ontario	-579.75	44.2	7.44	234.68	0.065406875
Ont57	Ontario	-449.125	34.1	7.34	202.13	-1.366526048
Ont58	Ontario	-539.25	40.8	8.18	236.94	1.030471545
Ont59	Ontario	-752.125	62.4	8.27	222.49	-1.690113645
Ont60	Ontario	-701.5	56.4	7.78	257.87	0.05977588
Ont61	Ontario	-537.625	40.7	7.77	237.19	-0.760108493
Ont62	Ontario	-620.875	48	8.17	223	-0.769476047

Fig. 3 *Sample of data used in multivariate analysis.*

average wind speed for December to February over each of the Great Lakes (in meters per second; from NOAA GLERL's Great Lakes Dashboard) and August to October total evaporation data for each of the Great Lakes (millimeters per month; also from NOAA GLERL). Evaporation plays a strong role in ice cover on the Great Lakes every year, but the main project did not look into evaporation as a factor in estimating historical ice cover.

Before running any multivariate analyses, data were compiled into a singular .csv file (Fig. 3), organized by lake and then by year. Data for the years 1951-1983 are used in this analysis to ensure all variables being considered are consistent with one another, and so true correlations can be determined among the variables. The lakes' names are included in the dataset as the singular qualitative variable, and an index for each sample was created using the lakes' first three letters and year for each sample, e.g., Lake Ontario's 1975 data is stored under the index Ont75.

The dataset was previewed in R, and the column of index values (displaying the lake and year each row of data is for) were set as the row index to improve readability and understandability of charts and tables produced from the analyses. A simple correlation matrix was produced before running any analyses to get a sense for any pre-existing correlations among the variables used here. Additionally, a pairs plot was generated to display these correlations graphically, with the points color-coded by which lake the data were for. Different variables offer more distinction between the lakes than others, and certain variables have clear trends with one another, with varying levels of correlation for each of the Great Lakes.

Based on the predominantly quantitative nature of this dataset, a Principal Components Analysis (PCA) was performed to determine which components were most influential and could be retained in a reduced dataset. The PCA was run on the five quantitative variables to standardize them for further analysis: CFDD-NMDD (CFDD *or* NMDD, depending on the lake), Ice Cover, Wind Speed, Evaporation, and ABNA. Six charts were generated to visualize the significance of the first four components derived

from these data, several of which are color-coded to indicate the quality of representation for each data point.

C. Results

a. Initial analysis

Each of the Great Lakes has its own specific ice cover model, based on either CFDD or NMDD values, from the nominal 1898-1983 winter seasons. Some of the lakes were easier to model based on these temperature proxies than the others, likely due to physical factors of each lake, such as depth and volume, as well as properties of each lake's mesoscale weather trends, such as average wind speed and direction. Factors such as these should be investigated in some continuation of this study, as well as future related studies.

Lakes Erie and Ontario were the most difficult to hindcast, with low R^2 values when comparing the CFDD-based ice cover estimates to the test data (1963-72 ice cover). Adding the two anecdotal winter seasons highlighted by Hewer & Gough (2019) with full ice cover to Lake Ontario, though, proved extremely helpful in increasing the R^2 for many of the models tested for this lake. This improvement came after much testing of various models and methods of hindcasting ice cover for Lake Ontario, though. Lake Erie's frequent high ice cover winters are easy to model, but its less frequent low ice cover winters are more difficult to estimate and hindcast.

Considering both tests of accuracy (years 1963-72 and 1963-83), the best-modeled ice cover hindcasts, in order of highest to lowest R^2 value, by lake are: Superior, Michigan, Huron, then Erie and Ontario. The 1963-72 test data scatterplots and correlations showed an impressive R^2 value for Lake Superior of 0.8077, and relatively high values for Huron (0.69), Ontario (when including the anecdotal data, 0.5301), and Michigan (0.4893). Lake Erie, though, had low R^2 values when looking at the 1963-72 scatterplots, with an R^2 value of only 0.1186. When considering the complete test data period (1963-83), two of the lakes saw improvement in their R^2 values: Lakes Michigan (0.7286) and Erie (0.6617). Lake Michigan had an increase in R^2 of nearly 0.24, whereas Lake Erie had an incredible increase in R^2 of 0.54. Of course, this is due largely in part to 1973-83 data being used to develop each of the lakes' models; however, it is important to consider the reliability of each model for these years as well. Lakes Superior (0.671), Huron (0.6349), and Ontario (0.3838) all saw decreases in their R^2 values when including the 1973-83 data.

The model equations used for each lake are provided below, as Equations 3-7; the variable y represents ice cover, and x represents maximum CFDD or minimum NMDD, indicated in the second column. The ice cover hindcast data for each lake from 1898-1983 are provided in Appendix III-A. Additionally, time series plots of hindcasted ice cover are available in Appendix III-B; the hindcasted ice cover is plotted with a ±1 standard deviation ribbon against the CFDD or NMDD values the ice cover is derived from.

Lake Superior	NMDD	$y = -0.0521x - 18.634$	(3)
Lake Michigan	NMDD	$y = 18.438*e^{-0.001x}$	(4)
Lake Huron	CFDD	$y = 0.0627x+3.1391$	(5)
Lake Erie	CFDD	$y = 0.4086*x - 43.534$ (CFDD ≤ 352) $y = 0.0141*x + 84.755$ (CFDD > 352)	(6a) (6b)
Lake Ontario	NMDD	$y = 13.871*e^{-0.002x}$	(7)

b. Additional multivariate analysis

The preliminary covariance matrix indicates there are three sets of variables that are moderately correlated: CFDD-NMDD and Evaporation (R = 0.389), Ice Cover and Evaporation (R = 0.351), and Ice Cover and ABNA (R = -0.357). The variables Wind Speed and Evaporation are the least correlated, with a correlation coefficient (R) of -0.028, indicating little to no correlation. Other combinations of the variables have varying degrees of correlations, but none as strong as CFDD-NMDD and Evaporation. The stronger correlation between these two variables makes sense, as both are related to winter air temperatures; CFDD-NMDD is a direct interpretation of over-lake air temperature over a period of time, and evaporation amounts are strongly influenced by over-lake air temperatures and lake water temperatures - when the air temperature is much colder than the lake water temperature, evaporation is much higher than when there is a smaller difference in temperature. Similarly, the moderate correlation between Evaporation and Ice Cover makes sense as well, as ice cover is related to evaporation and CFDD-NMDD - the more evaporation there is in the fall, the more ice cover there may be in the following winter; once ice cover is present, though, there is greatly decreased evaporation. Ice cover can also be estimated using CFDD-NMDD values, as they indicate the air temperatures for a given winter, and colder air temperatures often lead to more widespread ice cover. Lastly, the moderate correlation between Ice Cover and the ABNA index is promising for the larger Master's Project; the ABNA index may be able to indicate the ice cover for a given year, with some adjustments. This relationship is not heavily studied as of yet; it is one of the goals of the Master's Project to analyze and bring to light this new teleconnection's relationship to Great Lakes ice cover.

The pairs plot (Fig. 4, below) displays these correlations graphically, color-coded by lake. It is much easier to see correlations between variables that depend on the lake the data is for - most notably, the data for CFDD-NMDD vs Ice Cover differ between the five Great Lakes, whereas Wind Speed vs Evaporation has less distinction among the lakes and is more scattered about. One pronounced differentiation between the data for each of the lakes is evaporation: Lake Erie stands out from the four other lakes with

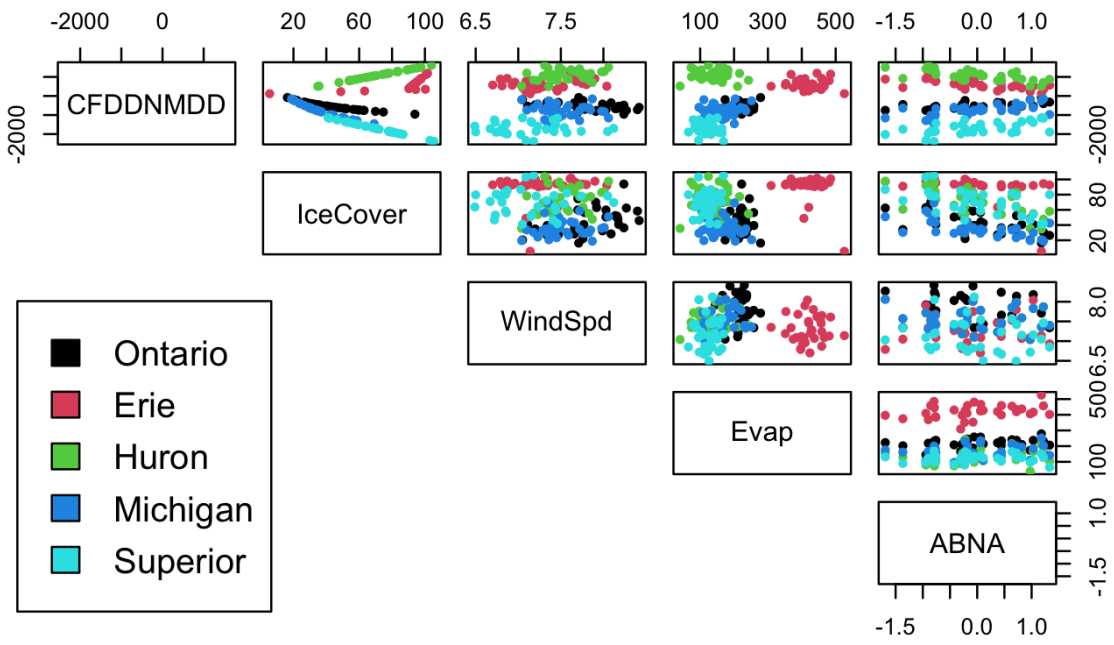


Fig. 4 *Pairs plot of Great Lakes data, from multivariate analysis*

much higher evaporation rates for the months of August through October. This may be due to Lake Erie's much smaller volume, also indicated by its frequent high-ice cover values - Lake Erie becomes completely (or nearly completely) covered by ice more often than the other lakes. Lake Ontario appears to have the highest average wind speed of the five lakes, most visibly in the Wind Speed vs Evaporation or vs ABNA plots. Could this be contributing to Lake Ontario's ice cover values that are often lower than the other four lakes? In the PCA, each of the five quantitative variables were standardized and combined to make five principal components, in varying degrees and concentrations of the original variables. Based on the *summary()* function in R (Fig. 5, below), the first four components account for 92.0% of the variance in the dataset (the first three account for 80.5%), indicating that PCA is a strong method of analysis for these data. The *Loadings* section of the PCA output indicates that all five variables are important in accounting for variability: component 1 is made up of scaled values from CFDD-NMDD, Ice Cover, and Evaporation; component 2 is made up of scaled values for all five variables; component 3 also uses all five variables, with different scales than component 2; lastly, component 4 uses scaled CFDD-NMDD, Wind Speed, and Evaporation for its values. The fifth component uses all five variables at scaled values, however, it accounts for the little remaining 8.0% of the dataset's variance. This information is visualized in the scree plot produced for the PCA, in Fig. 6 - once the first four components are accumulated, there is little change in the variance that is accounted for.

```
summary(gls.pca) # Shows importance of components, and how much variance they account for
```

```
## Importance of components:
##                           Comp.1    Comp.2    Comp.3    Comp.4     Comp.5
## Standard deviation     1.3145671 1.0899046 1.0527151 0.7596193 0.63149889
## Proportion of Variance 0.3456173 0.2375784 0.2216418 0.1154043 0.07975817
## Cumulative Proportion  0.3456173 0.5831957 0.8048375 0.9202418 1.00000000
```

Fig. 5 *R's* summary() *function, which outputs the proportion of variance accounted for in a PCA, by component*

The individuals plot for the PCA shown in Fig. 7 indicates that data from Lake Erie are strongly influenced by the first two principal components (PCs); Lakes Ontario and Superior are also strongly influenced by these PCs, but in the opposite direction. These data points are colored red and orange, indicative of this stronger influence. Lakes Michigan and Huron are less influenced by the first two dimensions, and are thus located closer to the plot origin and are colored in shades of yellow to blue. The two axes of the plot, Dimensions 1 and 2, represent the first two components of the PCA, and also contain the amount of variance they account for – 34.6% and 23.7%, respectively.

The variables correlation circle graphically displays the loadings for the variables that were used to generate the PCs using vector arrows on a compass-like platform. PC1 is determined using CFDD-NMDD, Ice Cover, and Evaporation, each having approximately equal loadings, as well as ABNA to a lesser degree. PC2 uses values from ABNA, CFDD-NMDD, Evaporation, and Ice Cover, with ABNA being the strongest factor of the four, then CFDD-NMDD, Evaporation, and Ice Cover having approximately equal loadings, but Ice Cover values being in the negative direction.

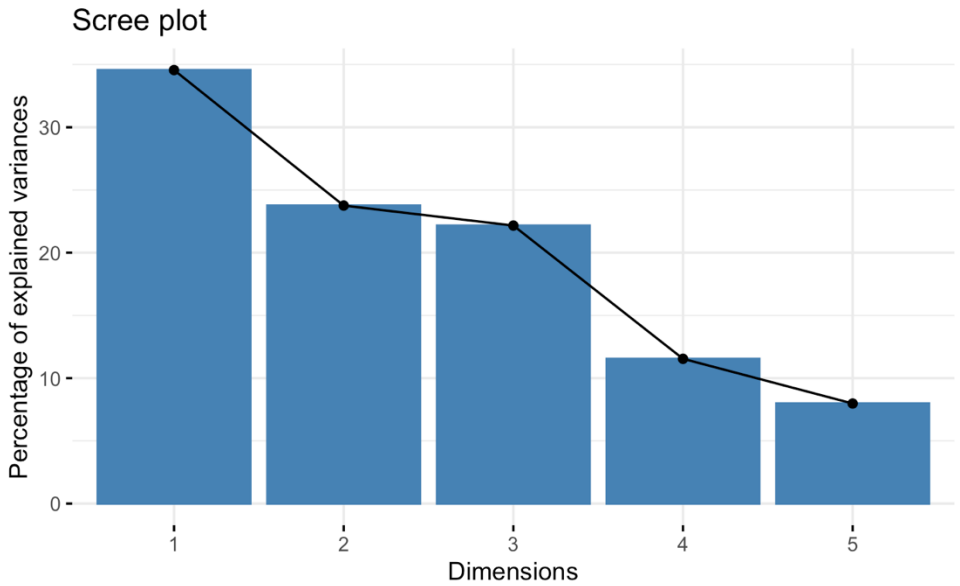


Fig. 6 *Scree plot of percent variance explained in the PCA, by component.*

Lastly, the biplots for the PCA displays both the individuals scatterplot and the variables correlation plot in one encompassing chart for each pair of components. Unfortunately, there are so many individuals in this dataset that their labels crowd out much of the chart and make it quite difficult to read. Therefore, any conclusions to be drawn from this chart can be done more efficiently by looking at the two separate graphs.

D. Conclusion

Modeling Great Lakes ice cover to hindcast and extend the historical record is best done with a combination of models, with each lake using its own specific trendline, timeline, and input data. The two input data types that proved most helpful in hindcasting the Lakes' ice cover were CFDD and NMDD; which one depends on each lake's winter temperatures, depth and volume, and other properties that influence the growth of ice on the lake. Three out of the five Great Lakes showed best results when using linear models based on either CFDD or NMDD - the exceptions being Lakes Michigan and Ontario, which showed the strongest results when using an exponential model. The two most westerly lakes, Superior and Michigan, used NMDD, as well as Lake Ontario; whereas the two middle lakes used CFDD; perhaps this is a factor of continentality: the more land-locked lakes may be able to get colder winters. Lakes Superior and Michigan also have the two largest volumes of the five Great Lakes; it may be that the NMDD accounts for these lakes' vast heat storage properties, sometimes referred to as a lake's "memory". This has to do with the properties of water in terms of its heat capacity;

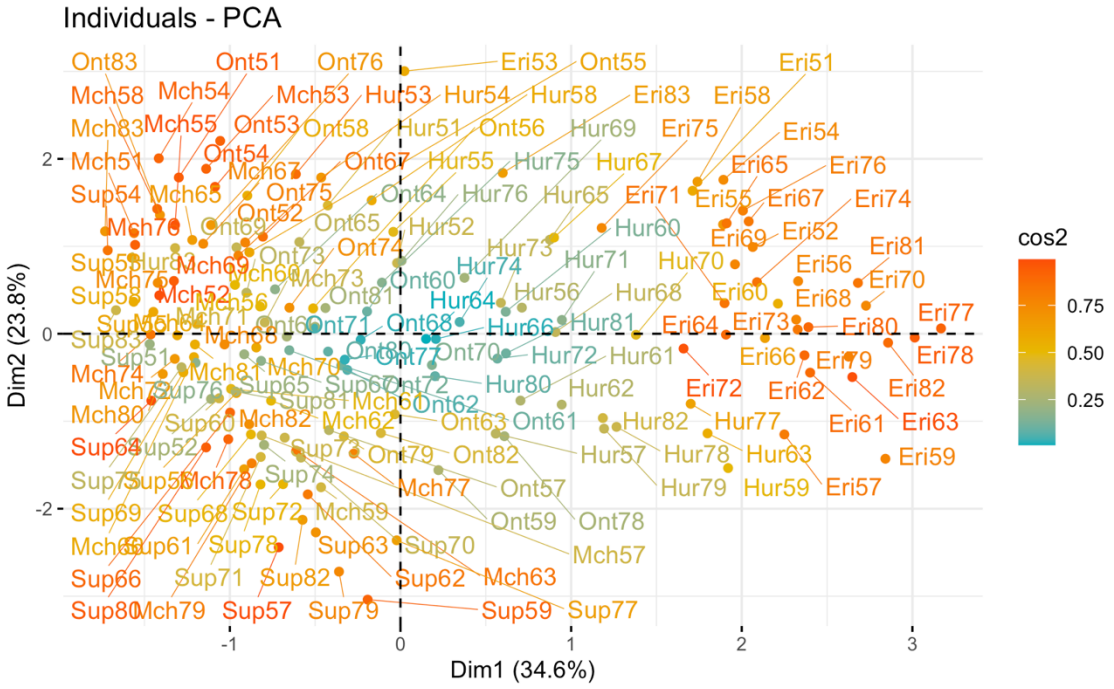


Fig. 7 *Individuals plot from PCA for components 1 and 2. Note that the labels greatly crowd out the data points.*

water has a heat capacity of 4.186 J/g•°C, giving it the ability to retain heat after air temperatures decrease, or to retain its coldness after air temperatures increase. This heat storage is likely why these two lakes did better with NMDD models, which had only a four-month span as compared to the CFDD models, which mostly had an eight-month span.

RECENT ICE COVER 2009-2020

A) Ice Cover

More consistent and higher resolution data started becoming available in 2008, and we have been able to collect fairly detailed information on ice cover progression and melting over the season from November through May. These were plotted for the available years, and it may be notable that, given the uneven distribution in surface area, overall maximum correlates more strongly to periods during which the larger lakes are at their maximums. Therefore, maximum ice cover date tends to neglect freezing patterns of Lake Ontario, which rarely freezes, and Lake Erie, which freezes most frequently and fully.

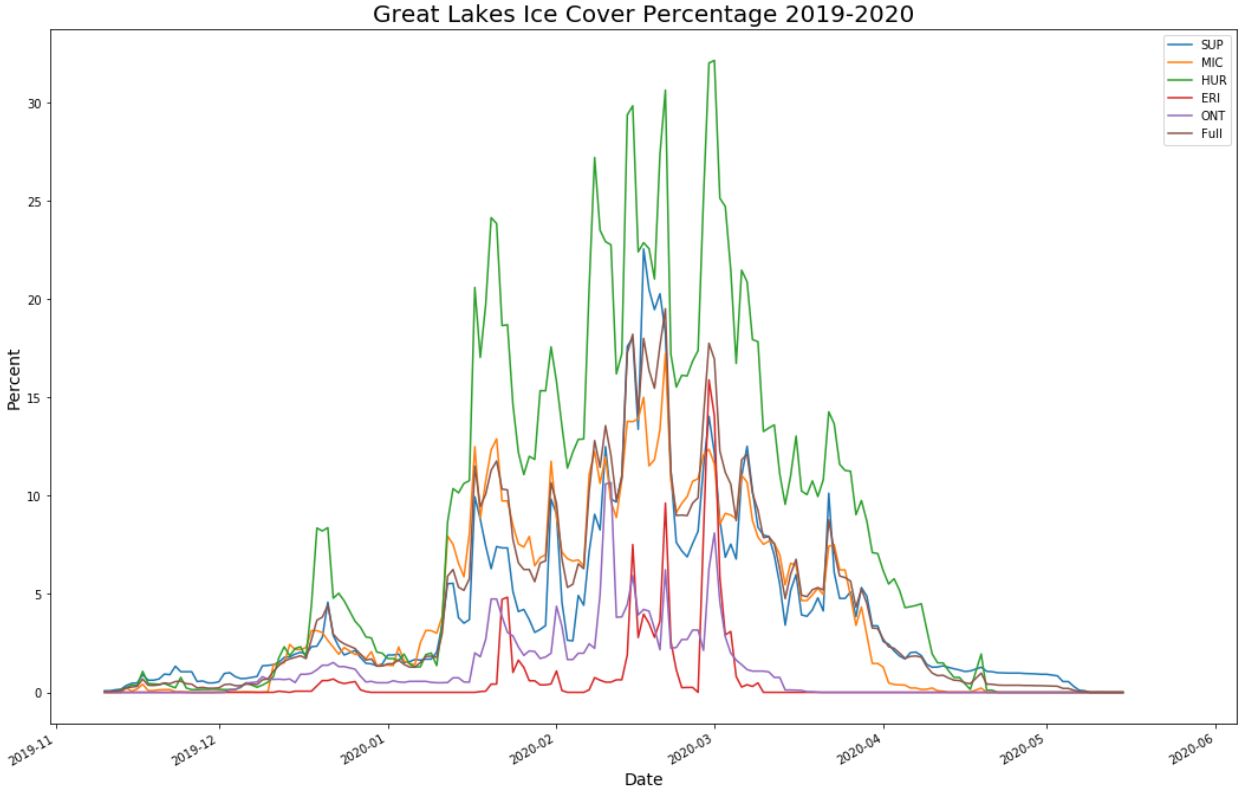


Fig. 8 *Great Lakes ice cover percentage, 2019-2020*

B) CFDD

More recent weather data were also collected from NOAA GLERL CoastWatch for the years 2009-2020, and were used to calculate CFDD values for these years. Twenty-four out of twenty-five of the same stations used in the historical analysis period were examined for this more recent period, as summarized below.

Some of the data were missing for some dates, varying by station; a combination of weather stations from CMAN and Surface Airway Stations was used to enhance temporal coverage. The data from the CoastWatch and CMAN stations had majority equal temperature values, and when unequal, were very similar, and so the more comprehensive station's data were used for a given location. The weather and ice cover data collected for this portion of the study are predominantly continuous data, allowing for season-long analysis, compared to the historical data's snapshot analysis, with maximum CFDD (minimum NMDD) and AMIC.

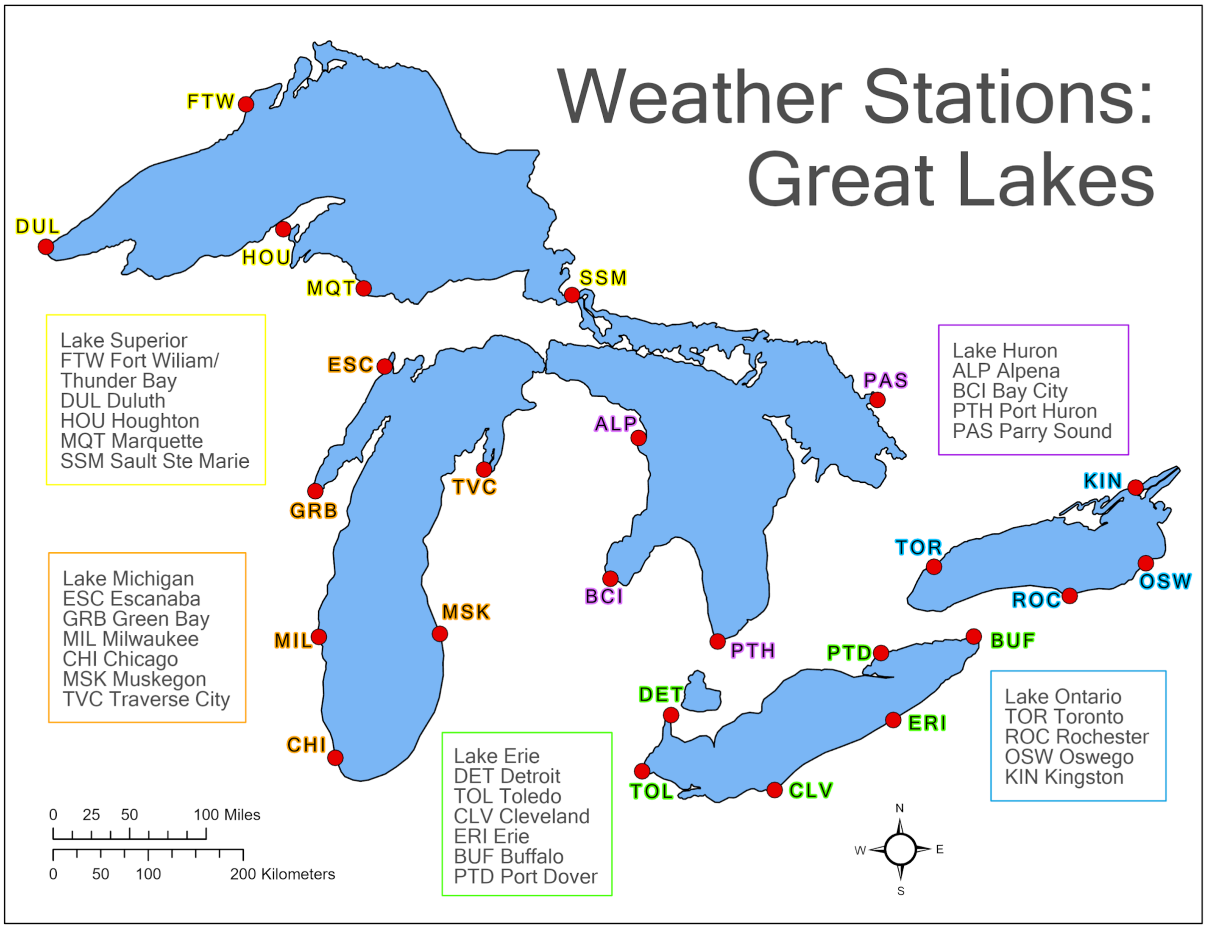


Fig. 9 *Map of Great Lakes weather stations used in this project.*

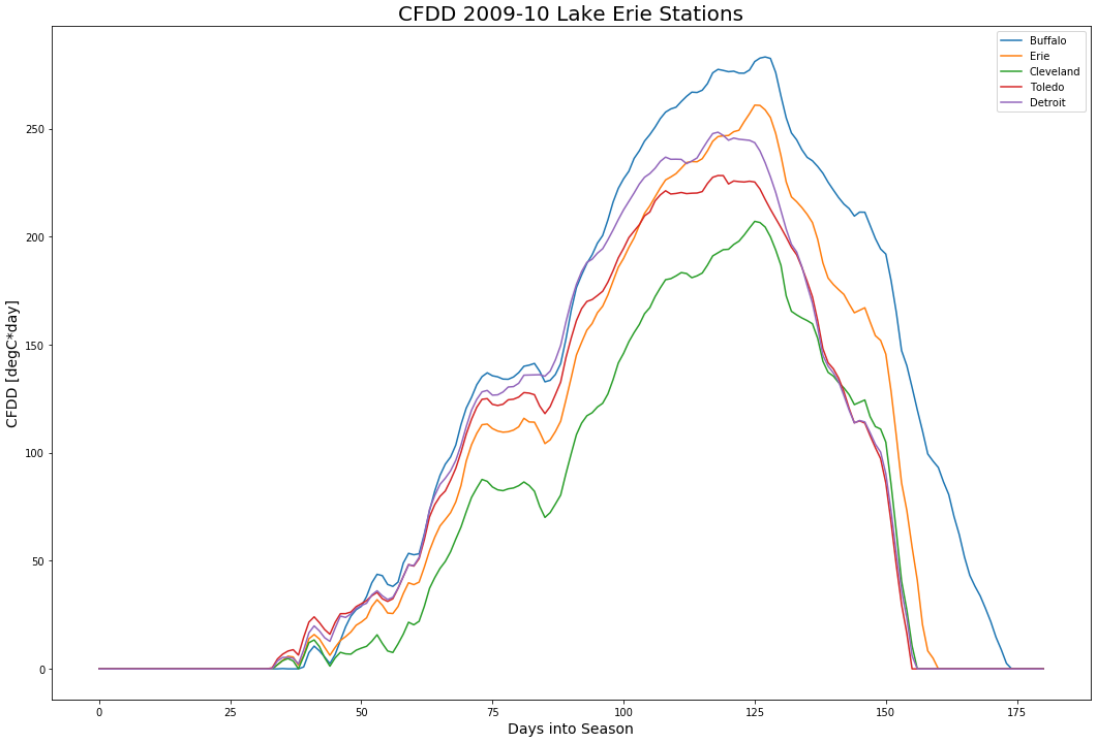


Fig. 10 *CFDD for Lake Erie's stations, 2009-10*

The CFDD calculations here used a slightly narrower winter season than the CFDDs derived from the historical weather data; the season length used here extends from November 1-April 30, whereas the historical data CFDD used a season from October 1-May 31. However, let it be noted that temperatures from October for these years did not contribute to any CFDD value accumulating to November 1st, and beyond April, further temperature data only contributed to a continued decline in CFDD; therefore, this difference should not affect CFDD values for the freezing period of any lake.

These CFDD values were then compiled into spaghetti line plots by year and station for each lake, and were used to estimate how well CFDD correlated to recorded ice cover, as CFDD has been intended for use in historical weather data. Below, we first see a sample of the 2010 freezing period for Lake Erie and then Lake Superior.

From the above, we can see that the accumulation of CFDD progresses similarly in different lakes, but Lake Erie sharply drops after the peak, compared to Lake Superior which slopes down gradually. Also note the difference in magnitude for CFDD, despite Lake Erie freezing more frequently and fully than Lake Superior. Below, we look instead at one station at Lake Erie (Detroit) and Lake Superior (Houghton) for all recent years. Despite interannual variability, we can see the overall shape at each given location will follow a consistent trend.

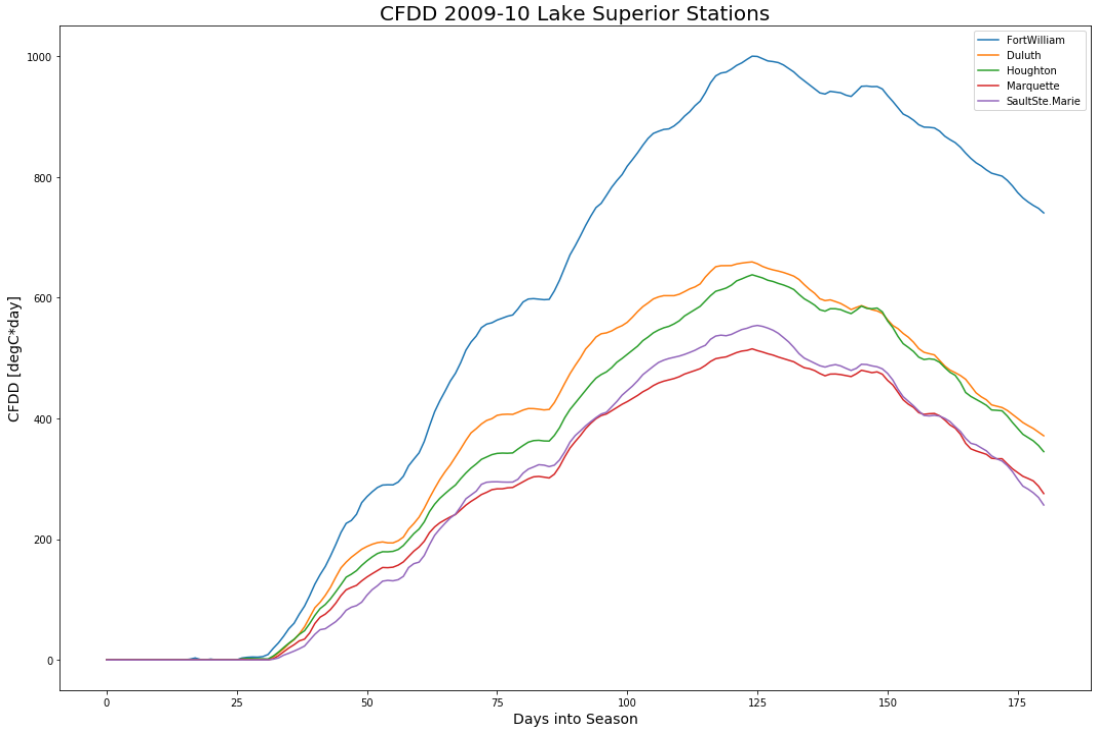


Fig. 11 *CFDD for Lake Superior's stations, 2009-10*

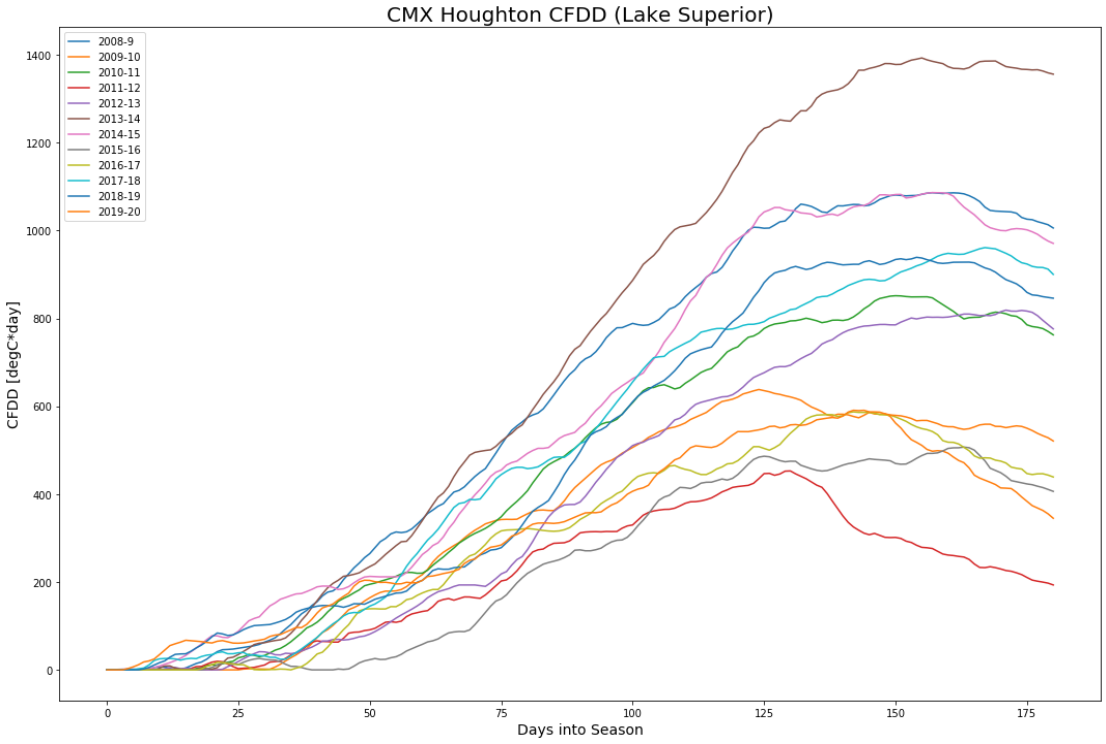


Fig. 12 *CFDD for the Houghton station on Lake Superior from 2008-9 through 2019-20*

C) Ice Cover vs. CFDD

With both the ice cover and CFDD data for these years available, scatter plots were made to analyze their relationship. Below we see the data for the full season period, meaning both their freezing and melting periods. Each lake has its own unique pattern of relationship, some with a more linear growth, while others closer to an S-curve.

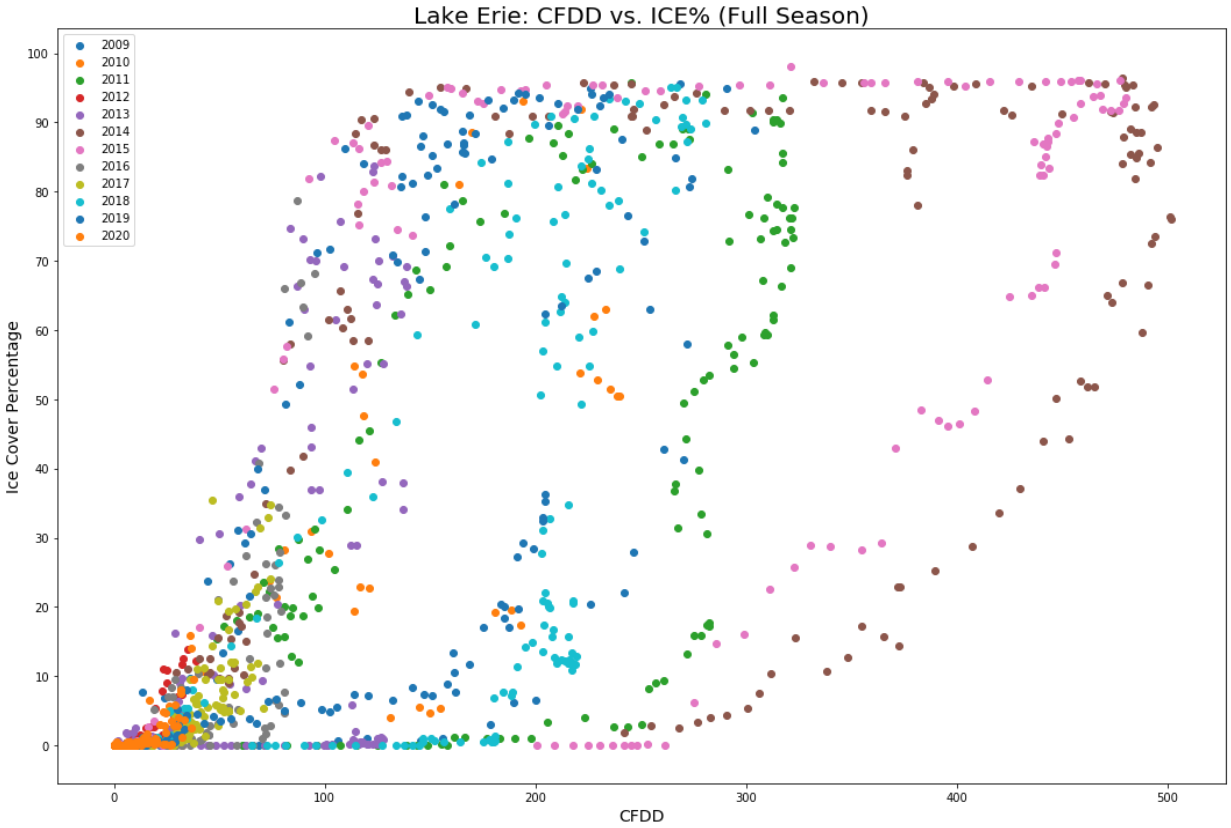


Fig. 13 *Lake Erie CFDD vs ice cover (%) over the full ice season*

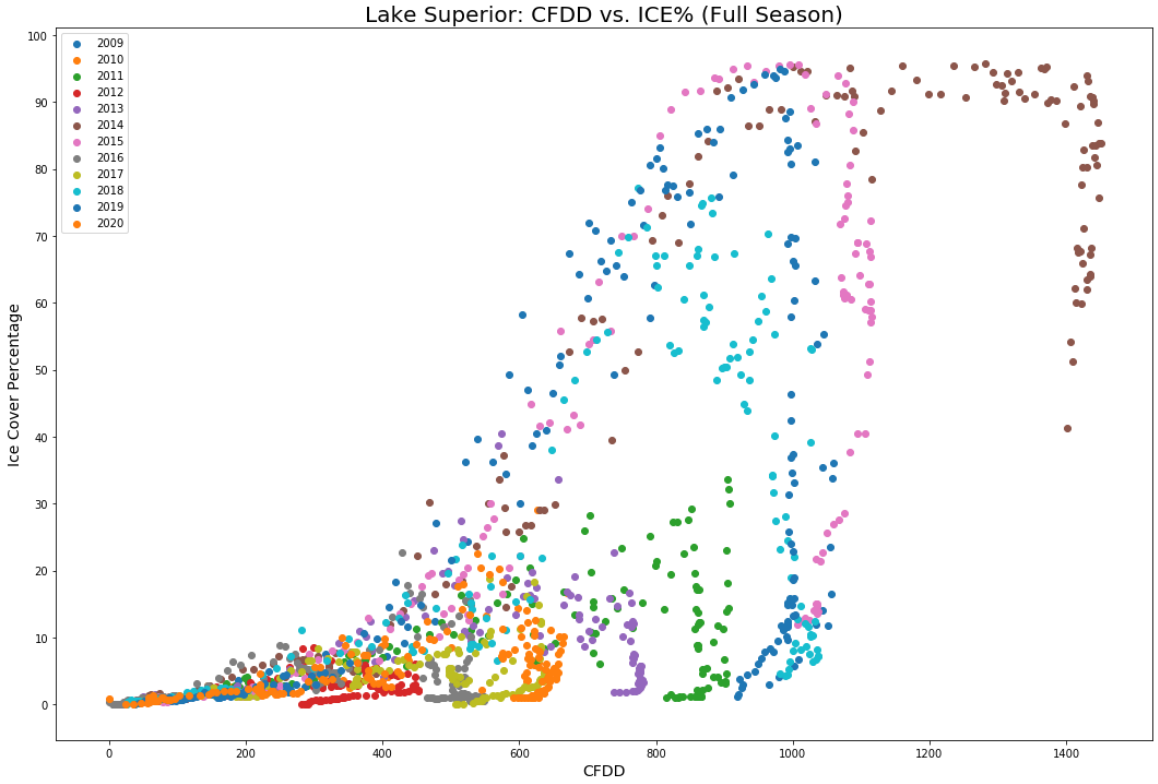


Fig. 14 *Lake Superior CFDD vs ice cover (%) over the full ice season*

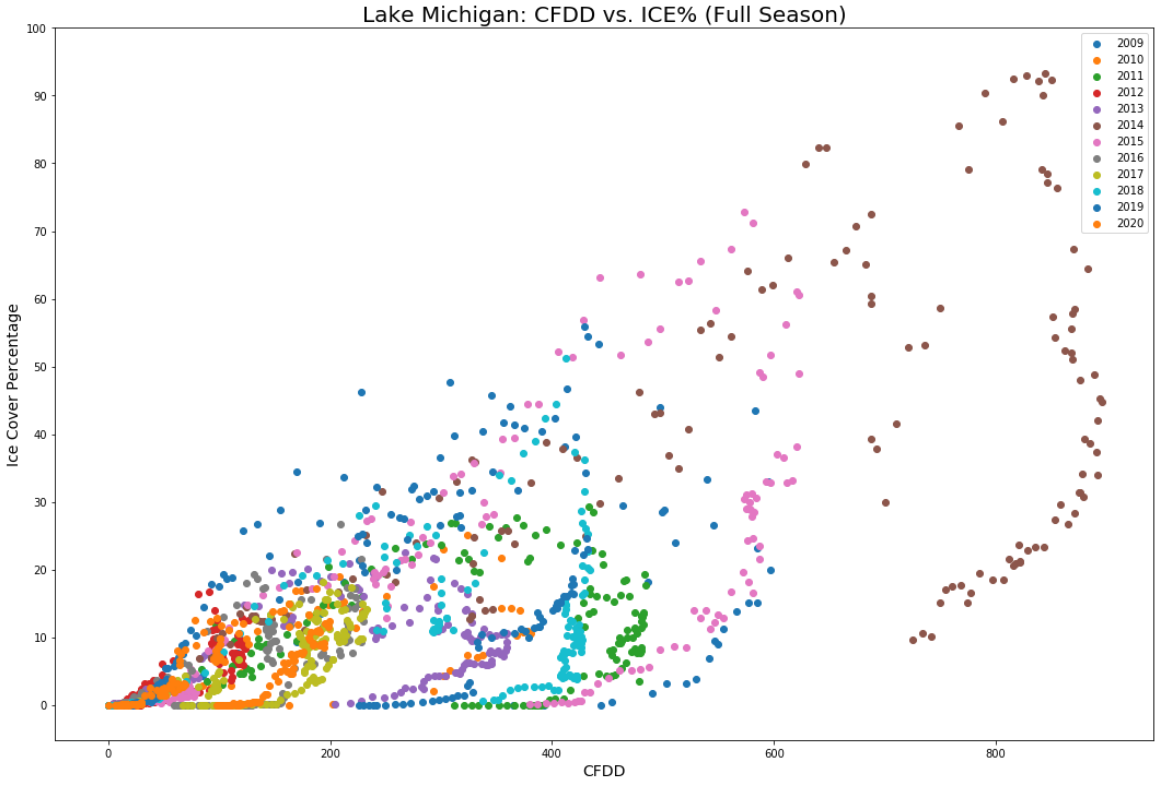


Fig. 15 *Lake Michigan CFDD vs ice cover (%) over the full ice season*

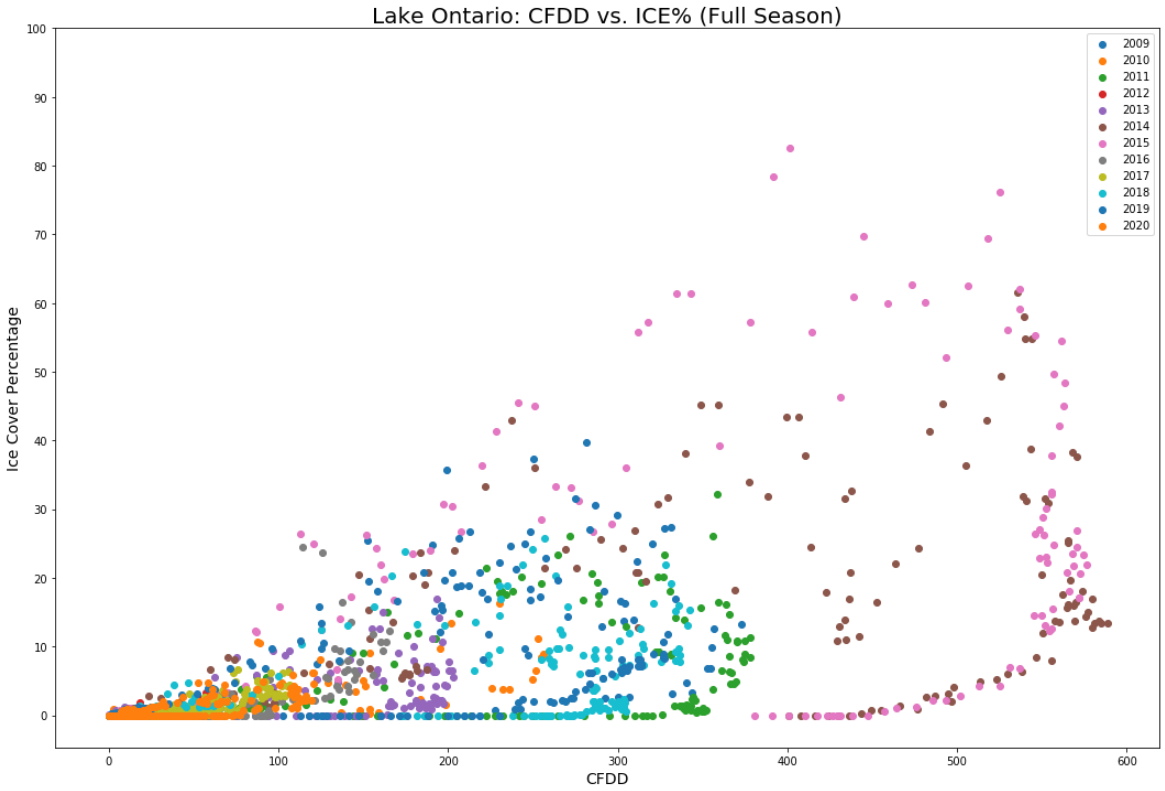


Fig. 16 *Lake Ontario CFDD vs ice cover (%) over the full ice season*

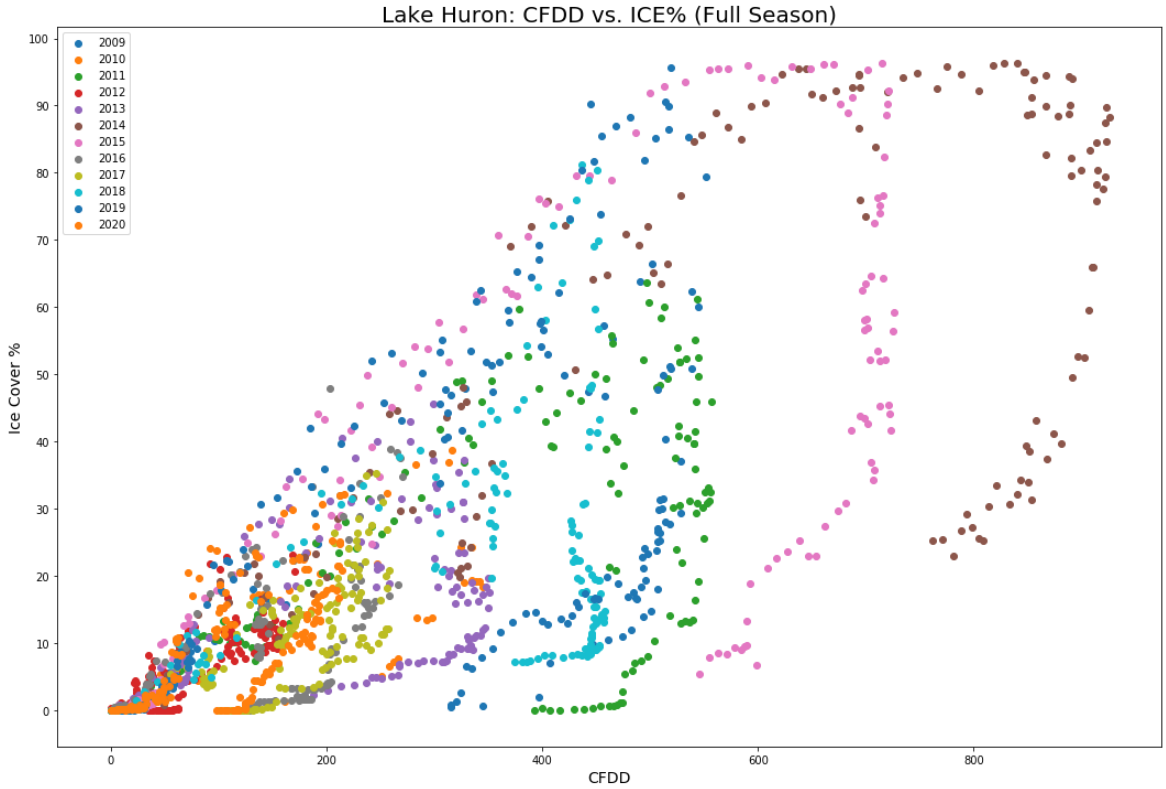


Fig. 17 *Lake Huron CFDD vs ice cover (%) over the full ice season*

Each lake was then broken down to view only its freezing period, defined either by the dates up to the maximum CFDD date or maximum ice cover date. Looking at an example of Lake Huron below, we can see CFDD does not precisely capture the freezing period, as the lake begins to melt even though the cumulative freezing value continues to increase, since this method does not account for NMDD.

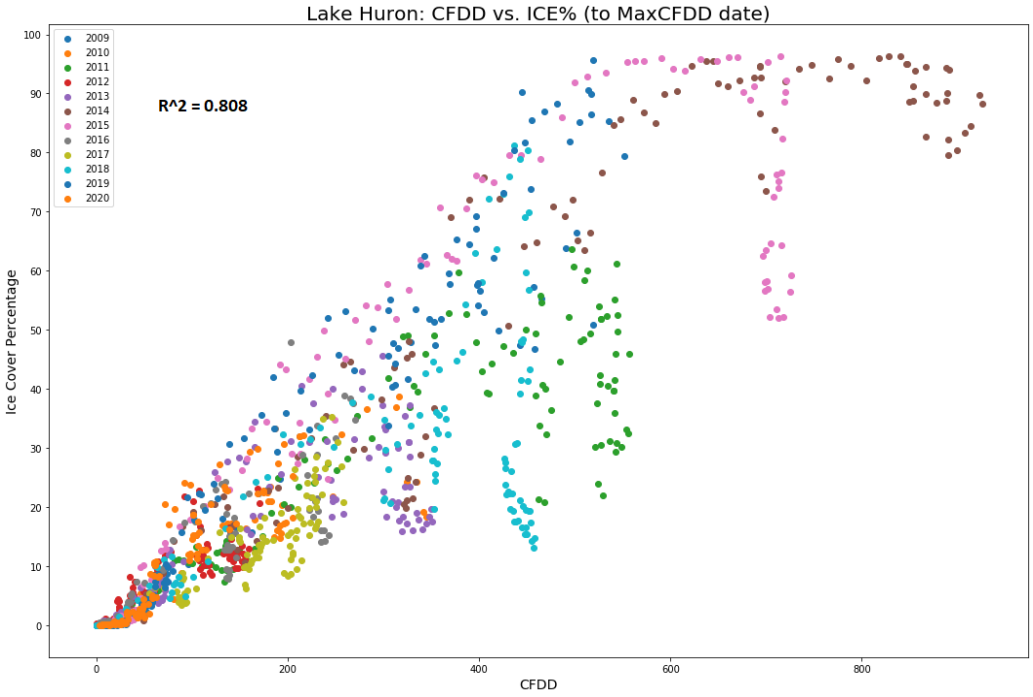


Fig. 18 *Lake Huron CFDD vs ice cover (%), up until date of maximum CFDD*

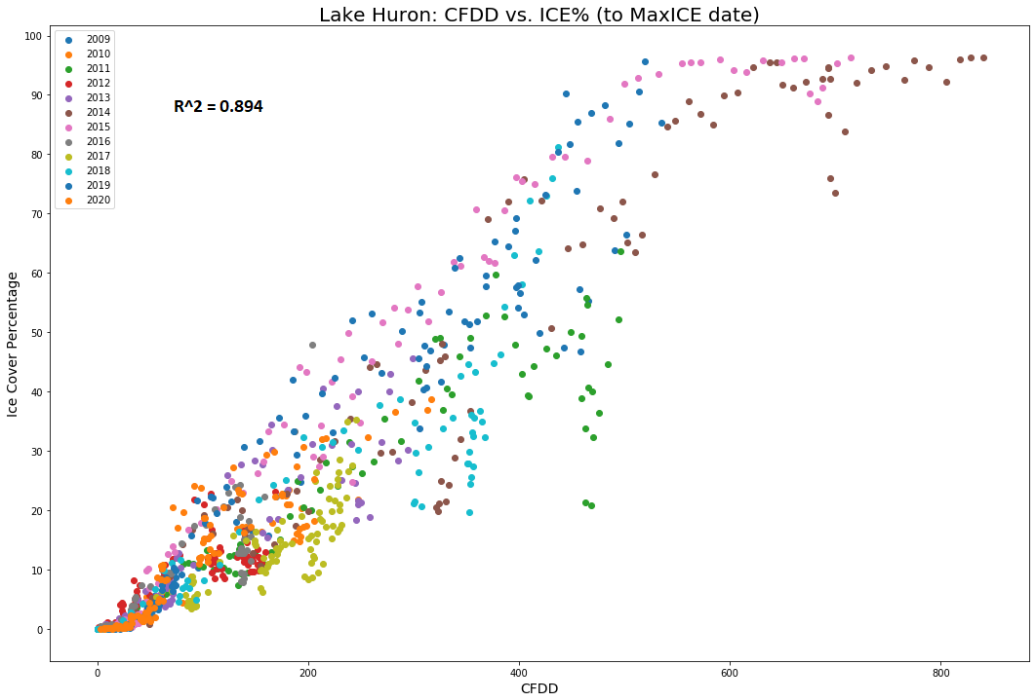


Fig. 19 *Lake Huron CFDD vs ice cover (%), up until date of maximum ice cover*

Lake Michigan and Lake Ontario similarly have a stronger trend with the date capped at the maximum ice cover date instead of the maximum CFDD date, observed below. The R-squared value of the regression line was taken for the resulting points, giving values of 0.984, 0.881, and 0.700 for Lake Huron, Michigan, and Ontario, respectively.

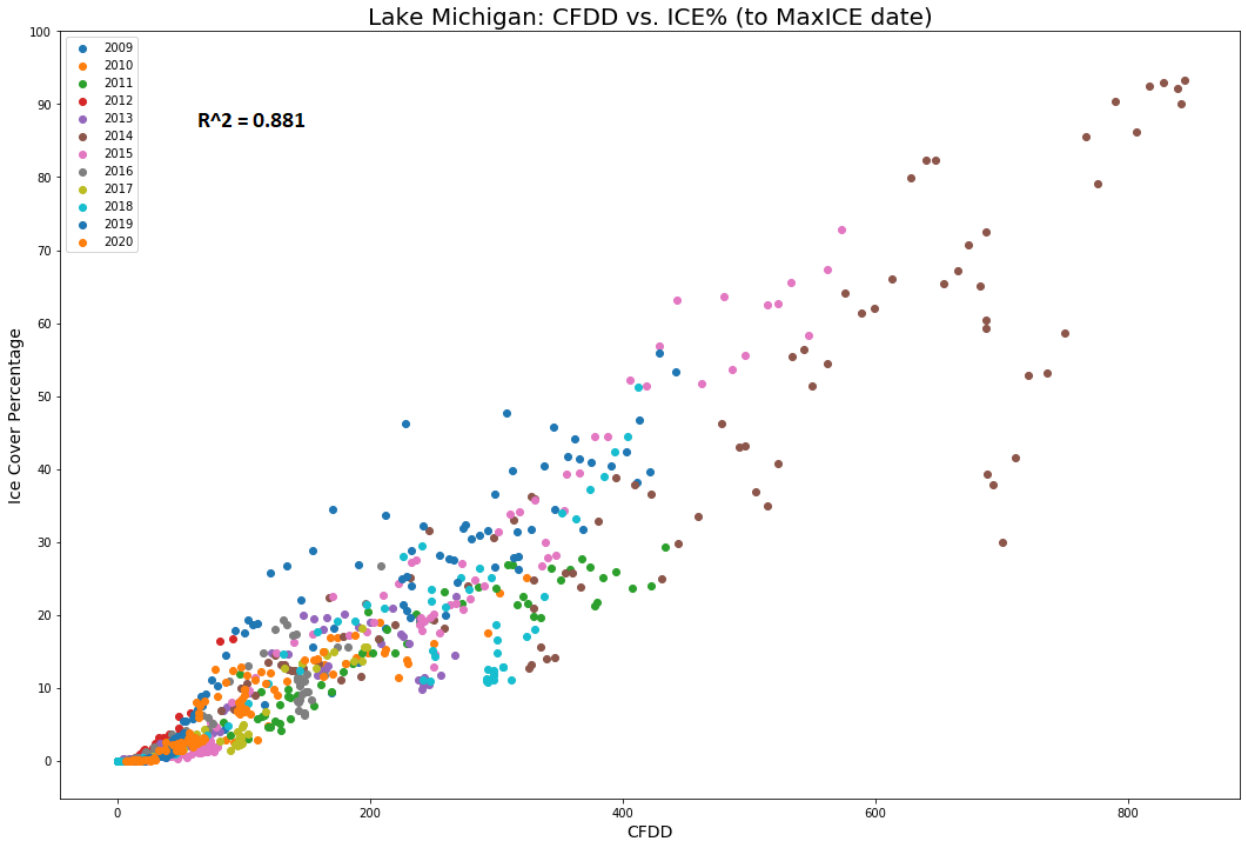


Fig. 20 *Lake Michigan CFDD vs ice cover (%) up until date of maximum ice cover*

Lake Superior and Erie were slightly more complicated. Though they both also correlated best with capping at the maximum ice cover date, they each have unique outliers. Below, we see outlier behavior in 2011 for Lake Superior. By removing this outlier, we improve the R-squared value from 0.672 to 0.798.

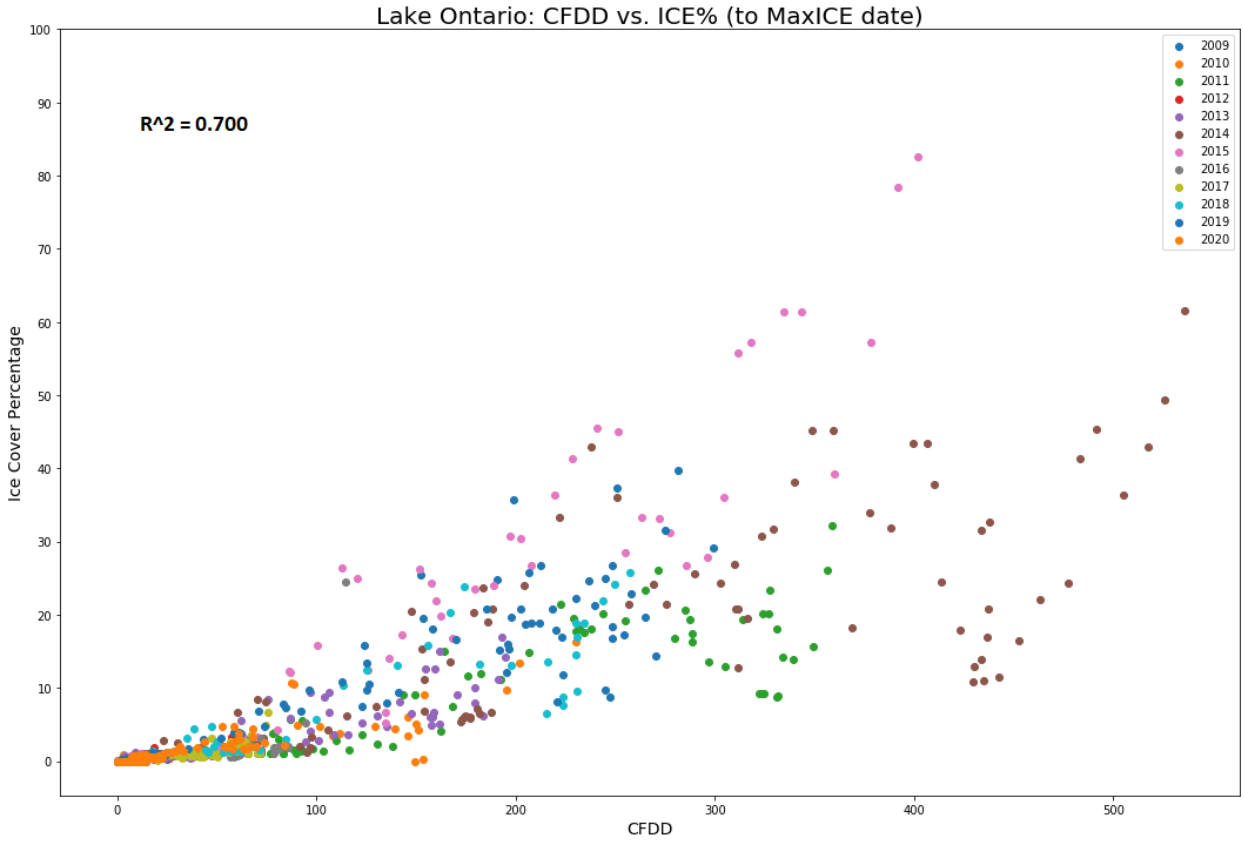


Fig. 21 *Lake Ontario CFDD vs ice cover (%) up until date of maximum ice cover*

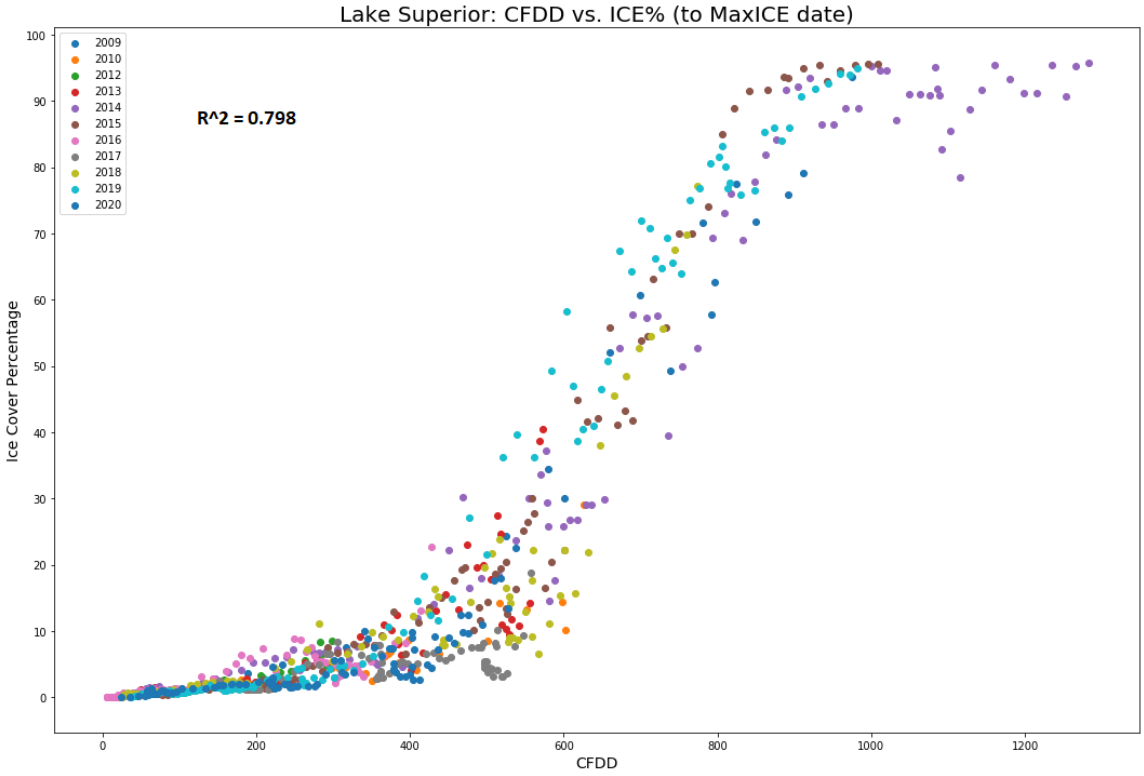


Fig. 22 *Lake Superior CFDD vs ice cover (%) up until date of maximum ice cover*

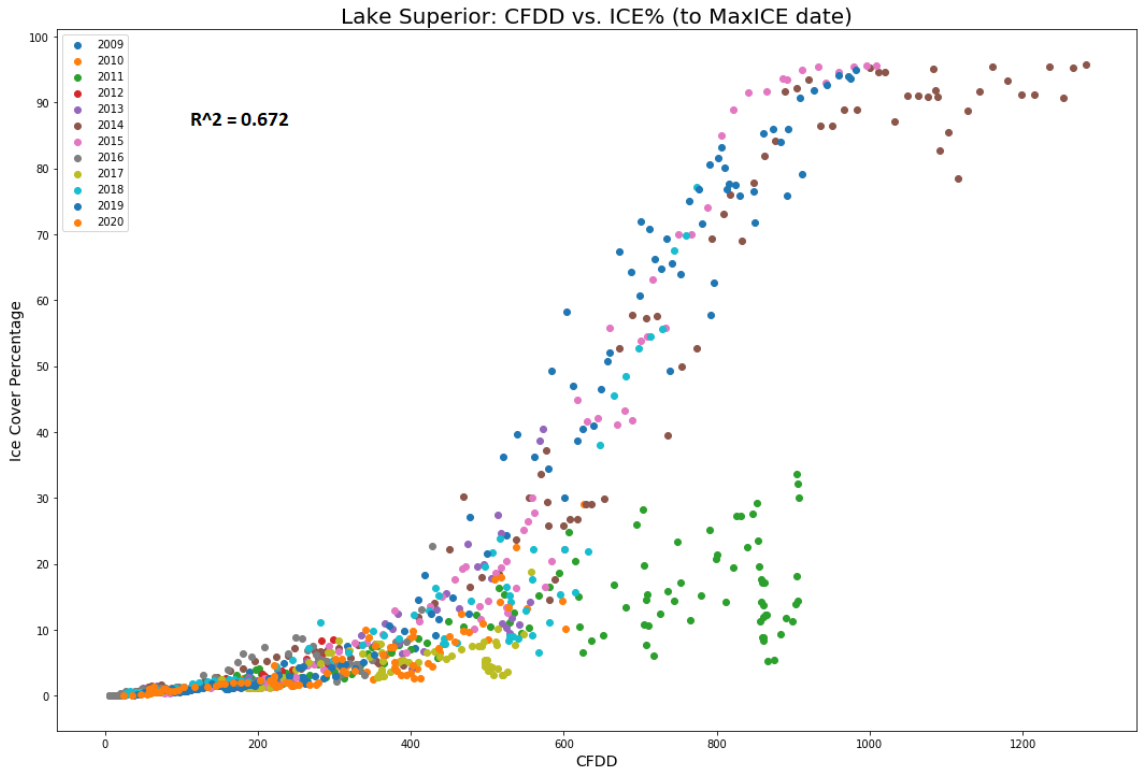


Fig. 23 *Lake Superior CFDD vs ice cover (%) up until date of maximum ice cover, without 2011 data*

Lake Erie has one specific outlier date which extends the time period beyond the beginning of its melt period, meaning that ice cover decreased before increasing beyond its earlier maximum within the same season. It is undetermined whether this is an anomaly in the lake freezing patterns, or whether it is an error in the data collection. It is currently observed only for one year at one lake, so if it was an anomaly, it is an isolated or regional one. By removing this one date of increased ice cover after the melting period initiated (March 6, 2014), the R-squared value improves from 0.797 to 0.867.

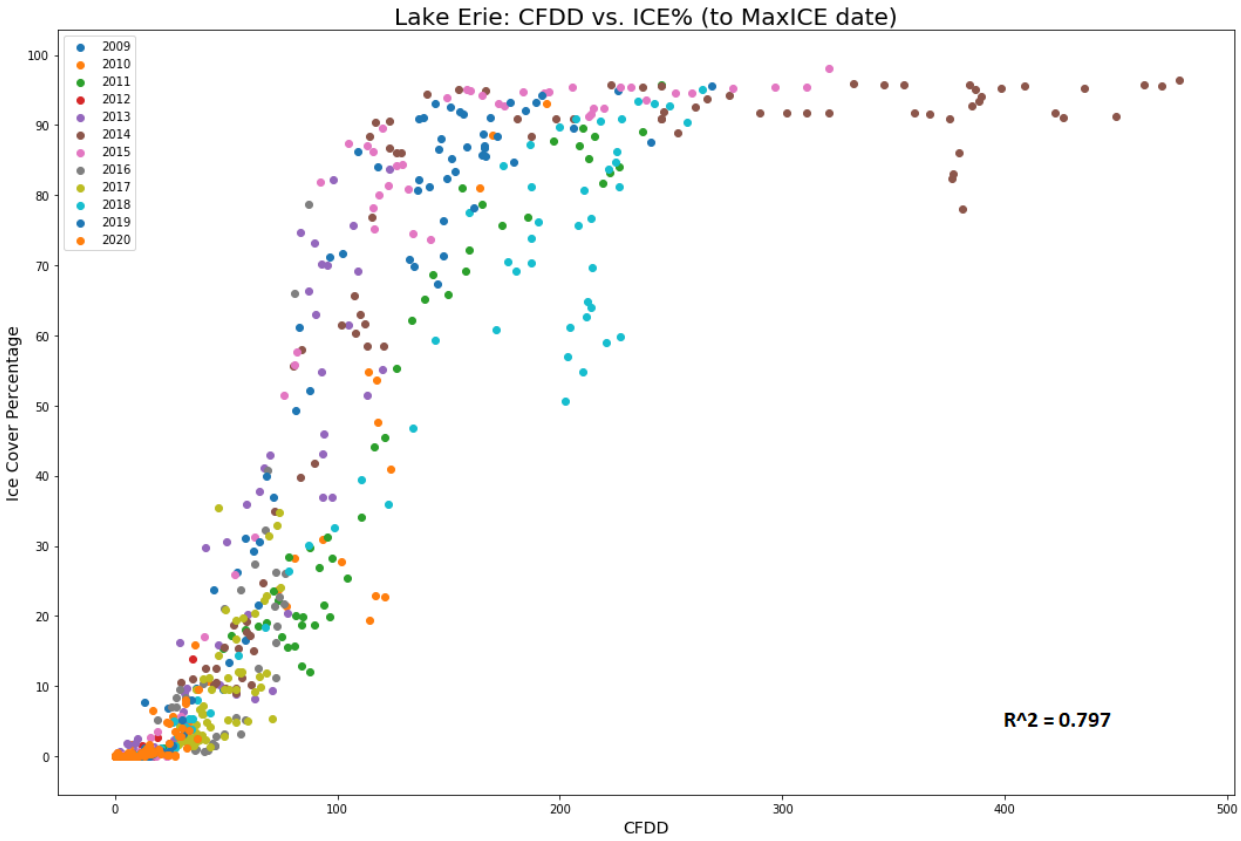


Fig. 24 *Lake Erie CFDD vs ice cover (%) up until date of maximum ice cover*

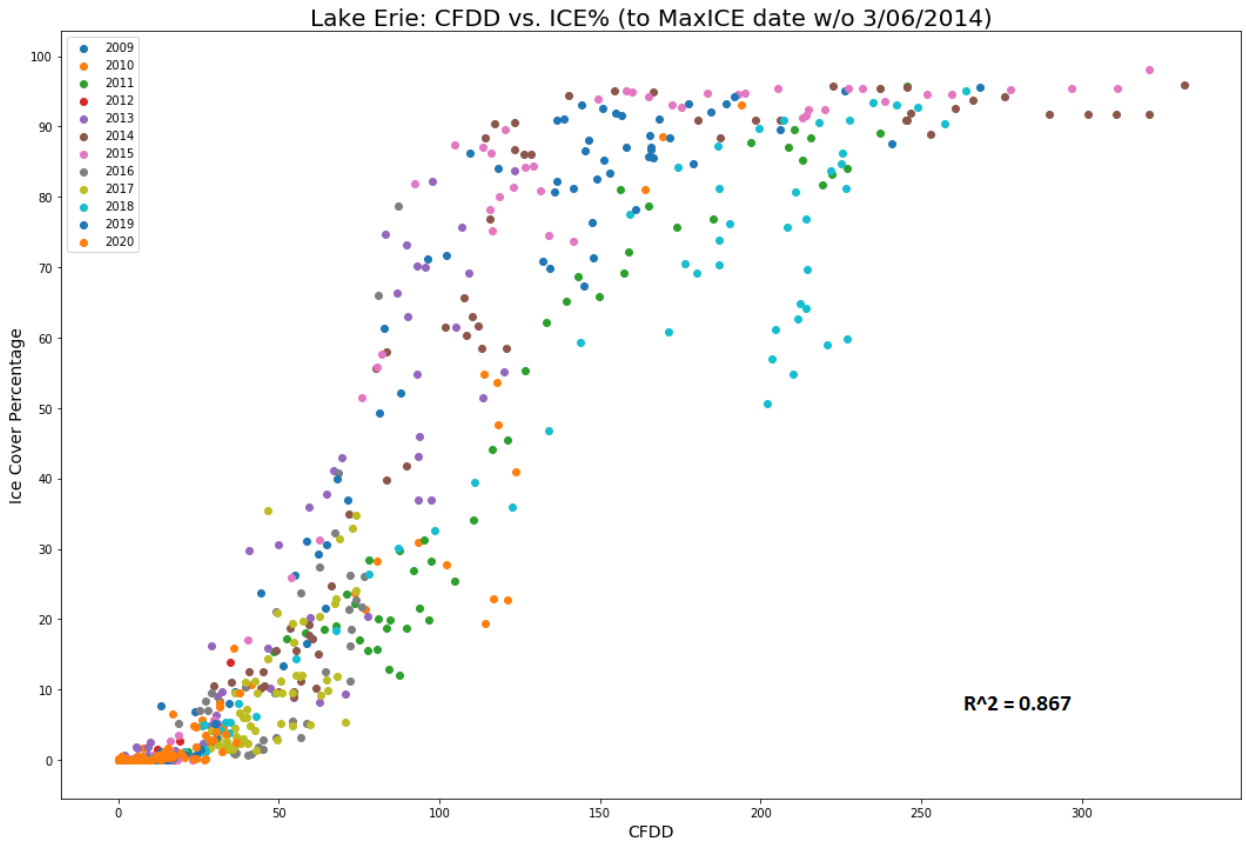


Fig. 25 *Lake Erie CFDD vs ice cover (%) up until date of maximum ice cover, without Mar 6, 2014 data*

These scatter plots show us that the freezing patterns of the lakes are not uniform, and each have unique relationships with the CFDD. This is consistent with observations, as the lakes vary greatly in size and shape, and near-shore regions freeze more easily than open water. As a result, the more elongated, narrow lakes of Ontario, Erie, and Michigan have more linear trends. Lake Erie freezes quicker, and therefore the line is more steep and levels off early on, whereas Lake Ontario never reaches its threshold for full lake freezing, despite reaching CFDDs three times what was necessary to freeze Lake Erie. Lake Huron also has a fairly linear trend despite its irregular shape, because it still avoids large deep water sections, as the Georgian Bay connects almost as a separate section and freezes much more frequently than the southern body of the lake. The Georgian Bay along with the North Channel increases Lake Huron's border-to-area ratio, allowing the majority of its freezing pattern to match the narrow lakes.

Lake Superior has the most distinct S-curve, with a slow initial freezing rate despite rapidly increasing CFDDs. However, this trend is the most cleanly consistent, with little interannual variability. Overall, each lake has a fairly linear relationship between CFDD and ice cover percentage, but with different thresholds of what CFDD is sufficient to initialize the freezing and what CFDD is sufficient for full lake freezing.

D) Coefficient of Determination

Each year's date of maximum ice cover was plotted for its ice cover percentage value versus its corresponding CFDD value. The R-squared value of the regression line was then taken, with the following results for Lake Michigan, Huron, Superior, Ontario, Erie respectively (greatest to least): 0.872, 0.796, 0.745, 0.688, 0.536. For the larger lakes, the value is fair, but less so for Lake Ontario, which hardly freezes, and Lake Erie, which almost always freezes. Lake Erie's especially low R-squared value may be due to unnecessarily high CFDDs being captured due to small increases in ice cover percentage, despite having generally reached its full-freezing threshold. In Lake Erie, the coastal bounds complicates the relation between the CFDD and ice coverage; namely, once the ice coverage reaches ~95%, it no longer increases with the increasing CFDD.

Another R-squared regression was observed between the dates of the maximum ice cover and maximum CFDD. The resulting R-squared values for each of the lakes were all under 0.5, with Lake Ontario the highest. This is an expected result, as CFDD continues to increase even after the lake has fully frozen for all except Lake Ontario, which has not fully frozen since 1979. Even in Lake Ontario though, using the date of the maximum CFDD as a proxy for the date of maximum ice cover would be poor. We can deduce from these results that, though CFDD and ice cover percentage do have a fair relationship, the timing of their maximums do not.

TELECONNECTION

A) Data

We obtained the temperature and atmospheric circulation data from the National Center for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) to recreate the ABNA index. We obtained air temperature at 2 metres (T2M) and 500 hPa geopotential height (Z500) from NCEP-DOE Reanalysis 2 model. It is an improved version of the NCEP Reanalysis I model (NOAA). We removed the linear seasonal trend from the global geopotential height data and the linear seasonal trend from masked North American temperature data. The removal of the monthly means gave us the anomalies for geopotential height and temperature. It was the first step in assessing the interannual variability of the ABNA teleconnection index. The AMIC by Lake is obtained through NOAA-GLERL. This dataset brings the basic units of data together from 1973 to the present. Both the NCEP Reanalysis 2 data and AMIC data could be found on the NOAA website.

B) Methods of Analysis

We used maximum covariance analysis (MCA) to analyze the NCEP geopotential height and temperature data. MCA is similar to Empirical Orthogonal Function Analysis (EOF) as they both deal with the decomposition of the covariance matrix. In EOF, the

covariance matrix is based on a single spatial-temporal field, while in MCA, the cross-covariance matrix is derived from two fields (Björnsson and Venegas, 1997). Through the decomposition of a cross-covariance matrix, we isolated the spatial pattern with the highest squared covariance. We used calculated leading modes of coupled geopotential and temperature spatial patterns to identify the temperature anomalous centres in NA and Z500 anomalous centres in the mid-high latitudes. We also calculated the expansion coefficients(ECS) of T2m and Z500, termed EC1 and EC2. Both expansion coefficients were regressed upon Z500 anomalies. They were also regressed upon Z500 anomalies with PNA signal linearly removed to reveal the significant Z500 anomalous centres. Z500 anomaly was then normalized by its standard deviation and averaged over the three regions of influence: A (45–60 °N, 80–110 °E), B (50–80 °N, 160E–150 °W), and C (40–60 °N, 100–70 °W). The ABNA index was constructed using Eq. 8 below (Yu et al. 2016):

$$\mathrm{ABNA} = \{\Phi'_{500}\}(\mathrm{A}) - \{\Phi'_{500}\}(\mathrm{B}) + \{\Phi'_{500}\}(\mathrm{C}) \qquad (8)$$

Using the reconstructed ABNA, we performed linear regression between the ABNA index and reported AMIC reports and proxies to determine the relationship between the variables.

C) Results

For this paper, we first recreated the ABNA index using the methods described by Yu et al. (2016). We then performed a correlation analysis between the ABNA index and the Great Lakes AMIC. The correlation analysis was based on linear regression. The figures below show the two leading MCA modes of NA 2m temperature and large-scale Z500 geopotential anomalies for NCEP Reanalysis II's DJF over 1980 – 2020. The first MCA model explained 48.3% of the squared covariance, while the second MCA model explained 31.9% of the squared covariance. Together, the two leading modes explain 80.2% of the total squared covariance. This is slightly lower than the 90.8% of the total squared covariance calculated by Yu et al. (2016). We found a similar spatial pattern as Yu et al.(2016) in the leading MCA pattern. It is characterized by warm anomaly over most central NA and cold anomaly over Alaska and Queen Elizabeth Islands. We also found a similar spatial pattern in the Z500 anomalies where the anomalous centres resided over the Bering Strait and NA. However, our anomalous centres had smaller amplitude compared to the results from Yu et al. (2016). The anomalous temperature region is supported by the thermal advection of polar and mid-latitude air exchange (Yu et al. 2016). The second leading MCA T2M pattern showed the above-average temperature in the southeastern United States and below-average temperature in northern Canada and Alaska; this is consistent with Yu et al. (2016). For the Z500 anomalies, we again found a similar spatial pattern with anomalous centres over

mid-latitude Pacific, Atlantic and northern Canada with smaller amplitude in our anomalous centre compared to the results from Yu et al. (2016).

Following the MCA, we calculated the expansion coefficients (ECs) of T2M and Z500, termed EC1 and EC2. Both expansion coefficients were projected onto Z500 anomalies, as seen in the Figures below.

Interestingly, our EC1 only had an 18.6% correlation with PNA while EC2 had -64.2% with PNA (both are significant at 5%) EC1 also had a 36.2% correlation with NAO, while EC2 had a 38.4% correlation with NAO (significant). This shows that an atmospheric circulation other than PNA directly influences the temperature variation in NA. After removing the PNA signal through linear regression, we get a pattern seen in the figure below. We see significant anomalous centres over the Great Lakes, Bering Strait, and Eurasia. Compared to Yu et al.(2016)'s result, we see a stronger signal over South Eastern China and Western Pacific. This could be the signal from a variation of East Asian jet stream (EAJS) connected to the teleconnection pattern (Yang et al. 2002). Ma et al. (2020) proposed that an intensified EAJS is associated with enhanced stationary wave activity from Asia to NA. Furthermore, Song et al. (2016) proposed that an anomalous East Asian trough event could lead to an eastward propagating Rossby wave train from East Asia to NA, affecting the surface temperature.

Using the regression map without the PNA signal, we normalized the significant Z500 anomalies by its standard deviation of the three regions over Eurasia, Bering Strait, and NA to construct the ABNA teleconnection index. We also constructed a second index using Z500 anomalies over just Bering Strait and NA. Using the ABNA index constructed by Yu et al. (2016), our calculated ABNA index, and our new index constructed using only two geopotential anomalous regions, we calculated its respective correlation with recorded AMIC for each lake from 1980 to 2020. We found significant results across all of the indices with various degrees of correlation with the AMIC of each lake. The figure below shows that Dr. Yu's index has stronger correlation values with AMIC of Lake Superior, Lake Michigan, and Lake Ontario. In contrast, our two constructed indices have a stronger correlation with the AMIC of Lake Huron and Lake Erie. Our new index and Dr. Yu's ABNA index have similar strength in correlation with the AMIC of all of the lakes. Furthermore, we ran linear regression with both NMDD and CFDD AMIC proxies and found significant results, as seen below. These high levels of correlation will hopefully create more accurate forecasting for Great Lakes Ice Coverage.

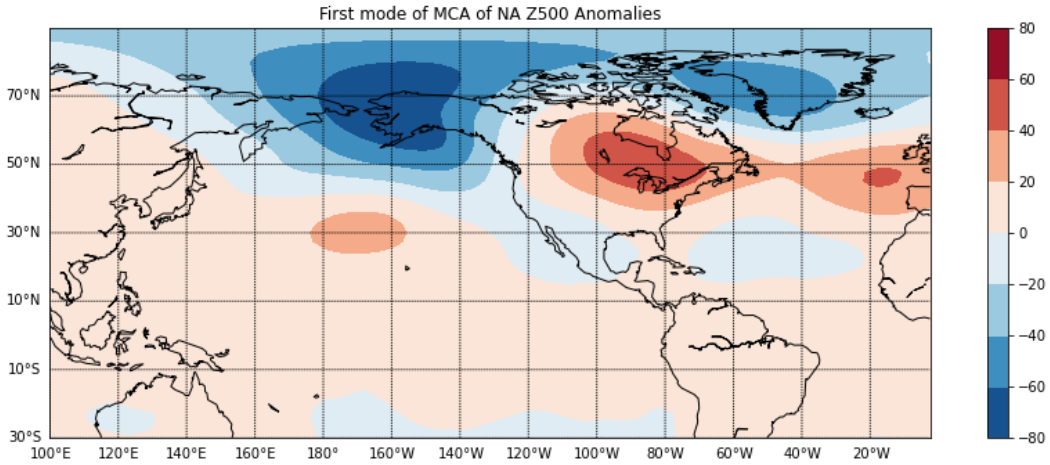


Fig. 26 *First mode of MCA of North America - Anomalies at 500 mb level*

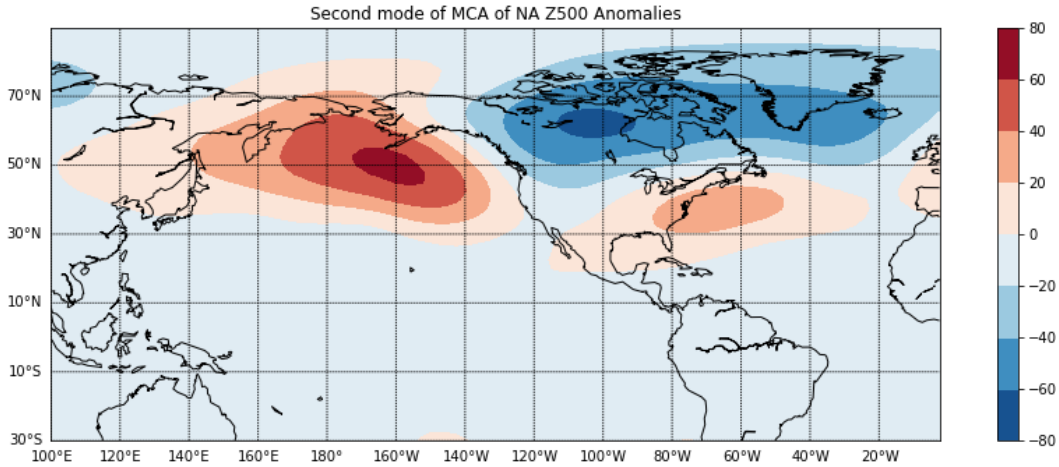


Fig. 27 *Second mode of MCA of North America - Anomalies at 500 mb level*

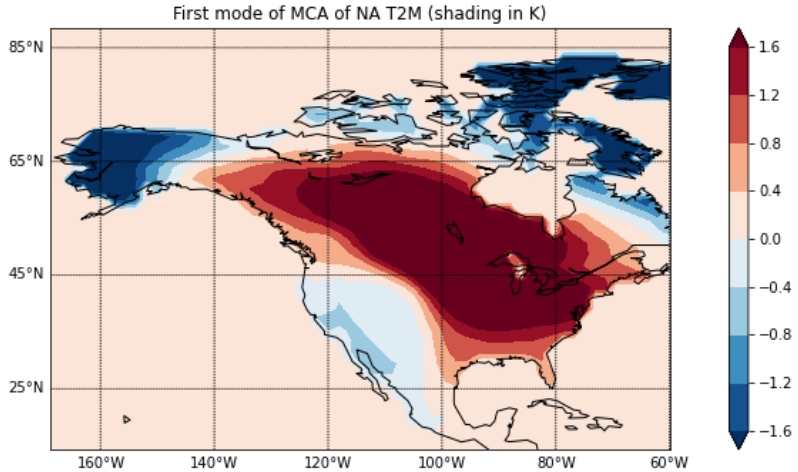


Fig. 28 *First mode of MCA of North America T2M*

Fig. 29 *Second mode of MCA of North America T2M*

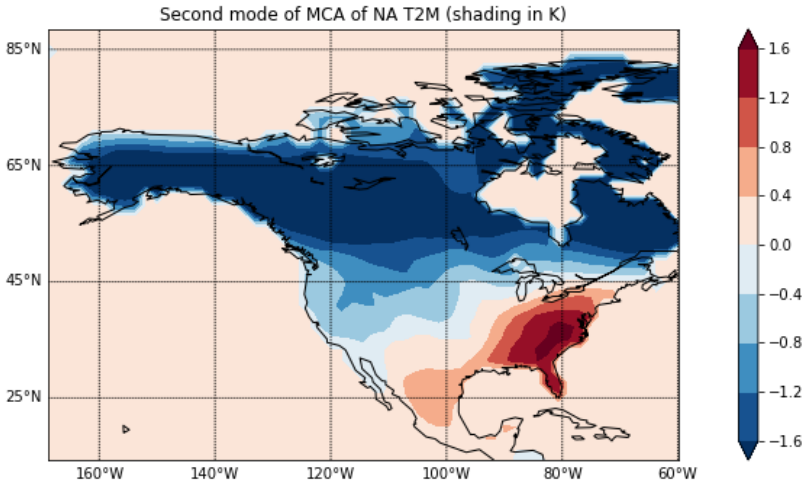


Fig. 30 *Cumulative square covariance explained*

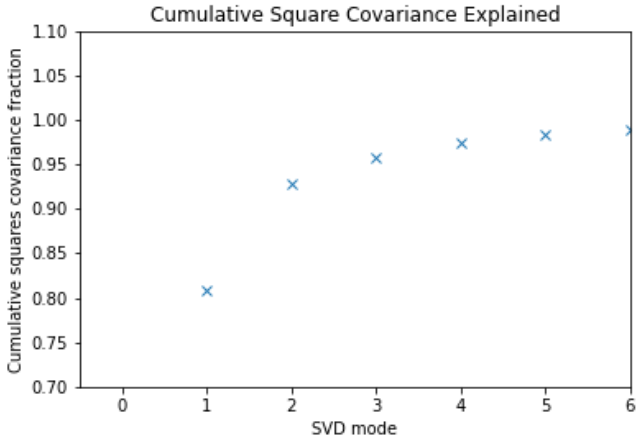
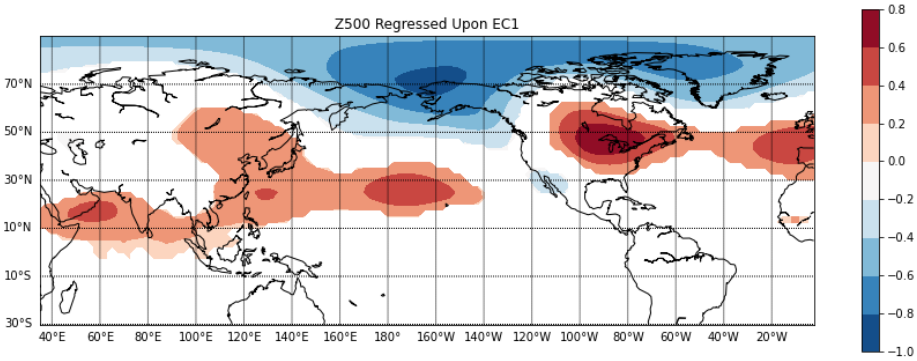


Fig. 31 *500 mb heights regressed upon EC1*



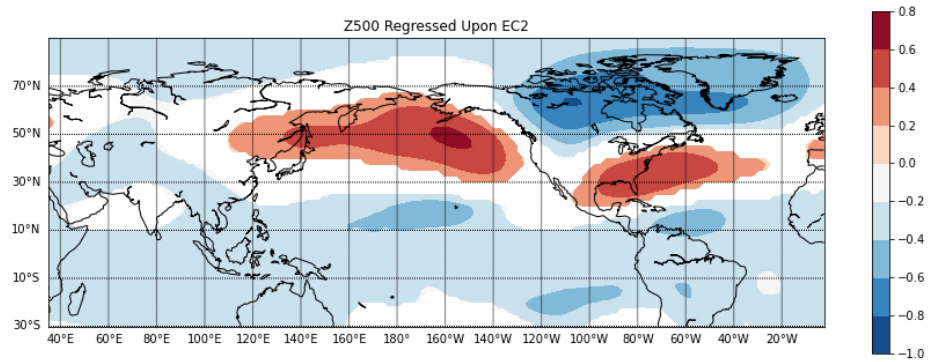


Fig. 32 *500 mb heights regressed upon EC2*

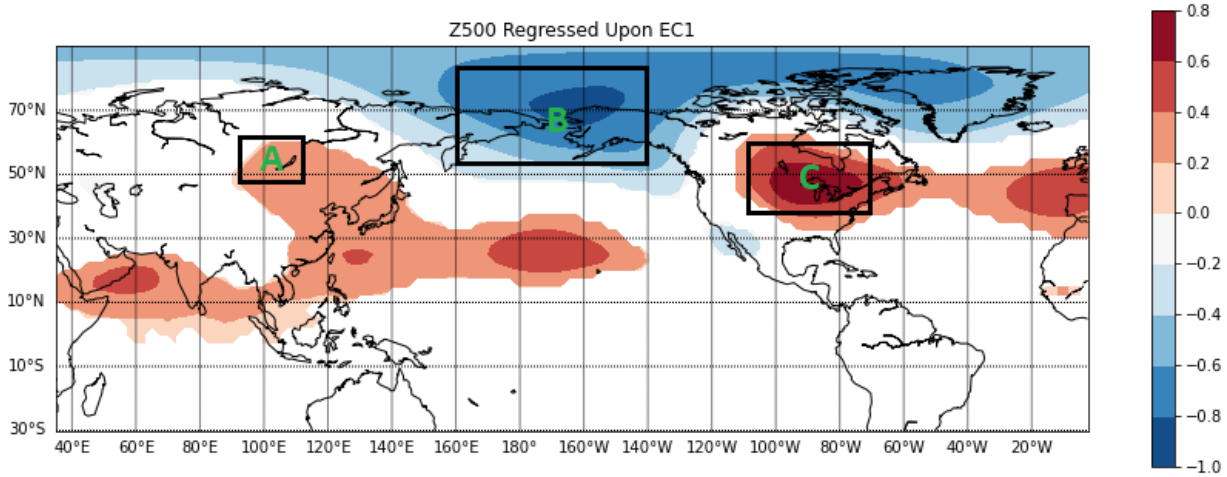


Fig.33 500 mb height regressed upon expansion coefficients from MCA. Significant result at 5%. Highlighted regions A, B, and C are used to construct the ABNA Teleconnection Index. Regions B and C are used to construct the new index.

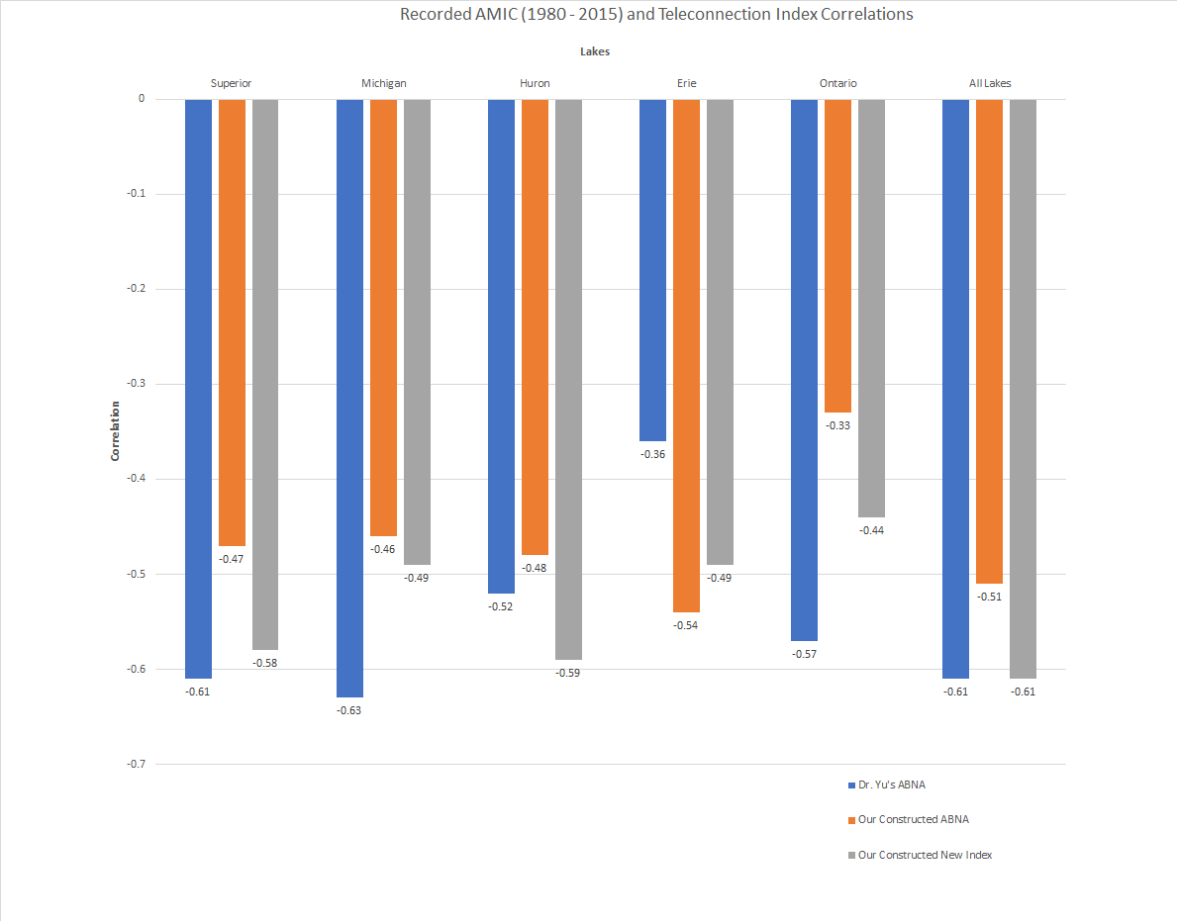


Fig. 34 *Correlation between AMIC and teleconnections*

D) Discussion

In the third part of our project, we used NCEP-Reanalysis II Data to find the leading modes of temperature and atmospheric circulation patterns over NA. After removing the PNA signal, we used the normalized geopotential anomalous fields over Eurasia, Asian-Bering Strait and the Great Lakes, to reconstruct the ABNA teleconnection index. We also constructed a second index using the geopotential anomalous field over the Bering Strait and the Great Lakes. The ABNA teleconnection index is maintained by synoptic eddy forcing. Yu et al. (2016) showed similar geopotential anomalies at Z250 where it exhibited anticyclonic forcing over NA and central Asia, and cyclonic forcing over Bering Sea and Strait. This indicates that the ABNA teleconnection is a zonally elongated synoptic-scale wave train with equivalent barotropic structure in the troposphere (Yu et al. 2016). The high degree of correlation between the ABNA index and the AMIC over the Great Lakes suggests that features from the ABNA index could significantly improve the accuracy of annual ice forecasting.

One of the features that have shown promise in seasonal prediction is looking at the Snow Water Equivalent (SWE). SWE anomalies from the previous season have shown to be associated with the following season of ABNA teleconnection (Yu and Lin, 2019). There has also been studies that showed a significant correlation between snow

cover over the Tibetan Plateau and temperature variation over North America (Lin and Wu 2011, Qian et al. 2019). The anomalous Tibetan Plateau snow cover can persist into the following winter through a positive feedback loop between the atmospheric circulation and the snow cover. The autumn Tibetan Plateau snow cover also showed a weak linkage with ENSO, making it a good predictor for ABNA and NA temperature. Qian et al. (2019) also showed that snow cover over eastern Tibetan Plateau causes perturbation near the core of the East Asian westerly jet. More research is needed to identify the mechanism and relationship between anomalous Tibetan snow cover, East Asian Westerly jet, and ABNA. Further research and quantitative analysis are also required to evaluate the robustness of SWE driving the coherent variability of T2M over North Asia and North America. If we could create a robust seasonal forecasting of the ABNA index using Tibetan Plateau SWE, we could significantly improve the current Great Lakes AMIC forecasting.

CONCLUSION

This portion of the Master's Project — extending the data record to the 1890s and identifying key teleconnections — works in conjunction with its other half, which uses machine learning models, to accomplish the overarching goal of enriching the database and improving the forecast of Great Lakes ice. In this half, we analyzed the past data to connect ice cover patterns with other physical conditions, such as local weather and climate and large-scale teleconnections. In the other half, we worked towards predicting future outcomes using past data. Both are necessary for a holistic evaluation of how ice cover has varied over time and will continue to develop in the near future.

The 2021 winter season, which concludes parallel to this report, has brought us what appears to be the lowest ice cover in decades. Interannual fluctuations in percent ice cover are a natural occurrence, but whether it is a part of a more significant trend remains to be seen. With the most recent data, we hope this project contributes to the extensive ongoing research in improving the linkage to past weather conditions, while also creating insights into future ice cover trends.

REFERENCES

Assel, R. A. (1972). *Great Lakes Ice Cover, Winter 1971-72*. Detroit, Michigan: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Ocean Survey, Lake Survey Center, Limnology Division. Retrieved 2020, from https://www.google.com/books/edition/_/Iiaw-WKMKOAC?hl=en&gbpv=0

Assel, R. A., Environmental Research Laboratories (U.S.). (1986). *Great Lakes degree-day and winter severity index update : 1897- 1983*. Ann Arbor, Mich.: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories.

Assel, R. A. NSIDC: National Snow and Ice Data Center. (1995). *GLERL Great Lakes Air Temperature/Degree Day Climatology, 1897-1983, Version 1*. Boulder, Colorado USA. doi: https://doi.org/10.7265/N5VD6WCJ.

Assel, R. A. (2003). *Great lakes monthly and seasonal accumulations of freezing degree-days, winter 1898-2002* (United States, National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory). Ann Arbor, Michigan: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory.

Assel, R., Cronk, K., & Norton, D. (2003). Recent Trends In Laurentian Great Lakes Ice Cover. *Climatic Change, 57* (1/2), 185-204. doi:10.1023/a:1022140604052

Assel, R. and Rodionov, S. (1998). Atmospheric teleconnections for annual maximum ice cover on the Laurentian Great Lakes. *International Journal of Climatology*. 18(4): 425-442. Doi: 10.1002/(SICI)1097-0088(19980330)18:4<425::AID-JOC258>3.0.CO;2-Q

Bai, X. and Wang. J. (2012). Atmospheric teleconnection patterns associated with severe and mild ice cover on the Great Lakes, 1963 - 2011. *Water Quality Research Journal of Canada*. 47: 421-435.

Bai, X., Wang, J., Sellinger, C., Clites, A., and Assel, R. (2012). Interannual variability of Great Lakes ice cover and its relationship to NAO and ENSO. *Journal of Geophysical Research: Oceans*. 117(C3). doi:10.1029/2010JC006932

Bjornsson, H. and Venegas, S.A. (1997). A manual for EOF and SVD analyses of climate data. McGill University, CCGCR Report No 97-1, Montreal, Quebec, 55 pp.

CoastWatch Great Lakes. (n.d.). Great Lakes Statistics. Retrieved from https://coastwatch.glerl.noaa.gov/statistic/ice/dat/

Dempsey, D. J. Elder, and D. Scavia, 2008: Great Lakes Restoration & the Threat of Global Warming. Valerie Denney Communications, 1–36 pp. https://www.nwf.org/Great-Lakes/More-Resources (Accessed April 16, 2021).

Di Liberto, T. (2018, July 09). Great Lakes ice cover decreasing over last 40 years. Retrieved from https://www.climate.gov/news-features/featured-images/great-lakes-ice-cover-de creasing-over-last-40-years

Ding, Q., Wang, B., Wallace, J.M. and Branstator G. (2011). Tropical - extratropical teleconnections in boreal summer: observed interannual variability. Journal of Climate. 24:1878 - 1896.

Hewer, M. J., & Gough, W. A. (2019). Lake Ontario ice coverage: Past, present and future. *Journal of Great Lakes Research, 45* (6), 1080-1089. doi:10.1016/j.jglr.2019.10.006

Kling, G. W., and Coauthors, 2003: Confronting Climate Change in the Great Lakes Region, Impacts on Our Communities and Ecosystems. 104 pp.

Lang, G. (n.d.). NOAAPORT - Realtime Great Lakes Weather Data and Marine Observations. Retrieved from https://coastwatch.glerl.noaa.gov/marobs/

Lin, H., and Wu.Z. (2012). Contribution of Tibetan Plateau snow cover to extreme winter conditions of 2009/10. *Atmos.Ocean*. 50:86-91.

Ma, T., Chen, W., Graf, H., Ding, S., Xu, P., Song, L., and Lan, X. (2020). Different Impacts of the East Asian Winter Monsoon on the Surface Air Temperature in North America during ENSO and Neutral ENSO Years. *Journal of Climate*. 33(24): 10671 - 10690. Doi: 10.1176/JCLI-D-18-0760.1

NOAA Regional Collaboration. (n.d.). NOAA in the Region: Great Lakes Region. Retrieved from https://www.regions.noaa.gov/great-lakes/index.php/regional-snapshots/

NOAA GLERL Great Lakes Dashboard. (2016, December 19). Index of /data/dashboard/data/hydroIO/wind. Retrieved from https://www.glerl.noaa.gov/data/dashboard/data/hydroIO/wind/

NOAA GLERL. (n.d.). Index of /pubs/tech_reports/glerl-083/UpdatedFiles. Retrieved from https://www.glerl.noaa.gov/pubs/tech_reports/glerl-083/UpdatedFiles/

Qian, Q., Jia, X., and Wu, R. (2019). Changes in the Impact of the Autumn Tibetan Plateau Snow Cover on the Winter Temperature over North America in the mid -1990s. *Journal of Geophysical Research: Atmospheres*. 124(19): 10321-10343. Doi: 10.20939/2019JD030245.

University of Wisconsin SSEC. (n.d.). Great Lakes Weather and Climate - The Stable Seasons. Retrieved from https://www.ssec.wisc.edu/sose/glwx/glwx_module3_summary.html

Rogers, J.C. (1976). Long-Range Forecasting of Maximum Ice Extent on the Great Lakes. *NOAA Technical Memorandum ERL GLERL-7*, Great Lakes Environmental Research Laboratory

Rondy, D. R. (1966). *Great Lakes Ice Cover, Winter 1965-66*. Detroit, Michigan: U.S. Lake Survey. Retrieved 2020, from https://www.google.com/books/edition/_/VpPSru6ApBsC?hl=en&gbpv=0

Rondy, D. R. (1967). *Great Lakes Ice Cover, Winter 1966-67*. Detroit, Michigan: Dept. of the Army, Lake Survey District, Corps of Engineers. Retrieved 2020, from https://www.google.com/books/edition/_/N_dnLLkJO5YC?hl=en&gbpv=0

Rondy, D. R. (1968). *Great Lakes Ice Cover, Winter 1967-68*. Detroit, Michigan: Dept. of the Army, Lake Survey District, Corps of Engineers. Retrieved 2020, from

https://www.google.com/books/edition/Great_Lakes_Ice_Cover_Winter_1967_68/8yK5QoeYfdIC?hl=en&gbpv=0

Rondy, D. R. (1969). *Great Lakes Ice Cover, Winter 1962-63 and 1963-64*. Detroit, Michigan: U.S. Lake Survey. Retrieved 2020, from https://www.google.com/books/edition/Great_Lakes_Ice_Cover_Winter_1962_63_and/4Bb1xQEACAAJ?hl=en&gbpv=0

Rondy, D. R., Lake Survey Center. Limnology Division. (1971). *Great Lakes ice cover, winter 1968-69*. Detroit, Mich.: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Ocean Survey, Lake Survey Center, Limnology Division.

Rondy, D. R. (1972). *Great Lakes ice cover, winter 1969-70*. Detroit, Michigan: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Ocean Survey, Lake Survey Center, Limnology Division. Retrieved 2020, from https://www.google.com/books/edition/Great_Lakes_Ice_Cover_Winter_1969_70/_XQ63PpEj1EC?hl=en&gbpv=0

Song, L. Lin, W., Chen, W. and Zhang, Y. (2016). Intraseasonal Variation of the Strength of the East Asian Trough and Its Climatic Impacts in Boreal Winter. *Journal of Climate*. 29(7): 2557-2577. Doi: 10.1175/JCLI-D-14-00834.1

Trenberth, K. E., Branstator, G., Karoly, D., Kumar, A., Lau N-C. and Ropelewski, C. (1998). Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperature. *Journal of Geophysical Research*. 103:14291-14324.

van Cleave, K., J. D. Lenters, J. Wang, and E. M. Verhamme, 2014: A regime shift in Lake Superior ice cover, evaporation, and water temperature following the warm El Niño winter of 1997-98. Limnology and Oceanography, 59, 1889–1898, https://doi.org/10.4319/lo.2014.59.6.1889.

Wang, J., and Coauthors, 2018: Decadal Variability of Great Lakes Ice Cover in Response to AMO. Journal of Climate, 31, 7249–7268, https://doi.org/10.1175/JCLI-D-17-0283.1.

Wilshaw, R. E., & Rondy, D. R. (1965). *Great Lakes Ice Cover, Winter 1964-65*. Detroit, Michigan: U.S. Lake Survey. Retrieved 2020, from https://www.google.com/books/edition/_/bObIWqsbcrkC?hl=en

Yang, S., Lau, K.M., and Kim, K.M. (2002). Variations of the East Asian Jet Stream and Asian-Pacific-American Winter Climate Anomalies. *Journal of Climate*. 15(3): 306-325. Doi: 10.1175/1520-0442(2002)015<0306:VOTEAJ>2.0.CO;2

Yu, B., Lin, H., Wu, Z.W., & Merryfield, W.J. (2016). Relationship between North American winter temperature and large-scale atmospheric circulation anomalies and its decadal variation. *Environ. Res. Lett*. 11 074001.

Yu, B., Lin, H., Wu, Z.W., and Merrifield, W.J (2018). The Asian-Bering-North American teleconnection: seasonality, maintenance, and climate impact on North America. *Clim Dyn*.50:2023 -2038. Doi: 10.1007/s00382-017-3734-6.

Great Lakes Ice Cover: Enriching Database and Improving Forecast

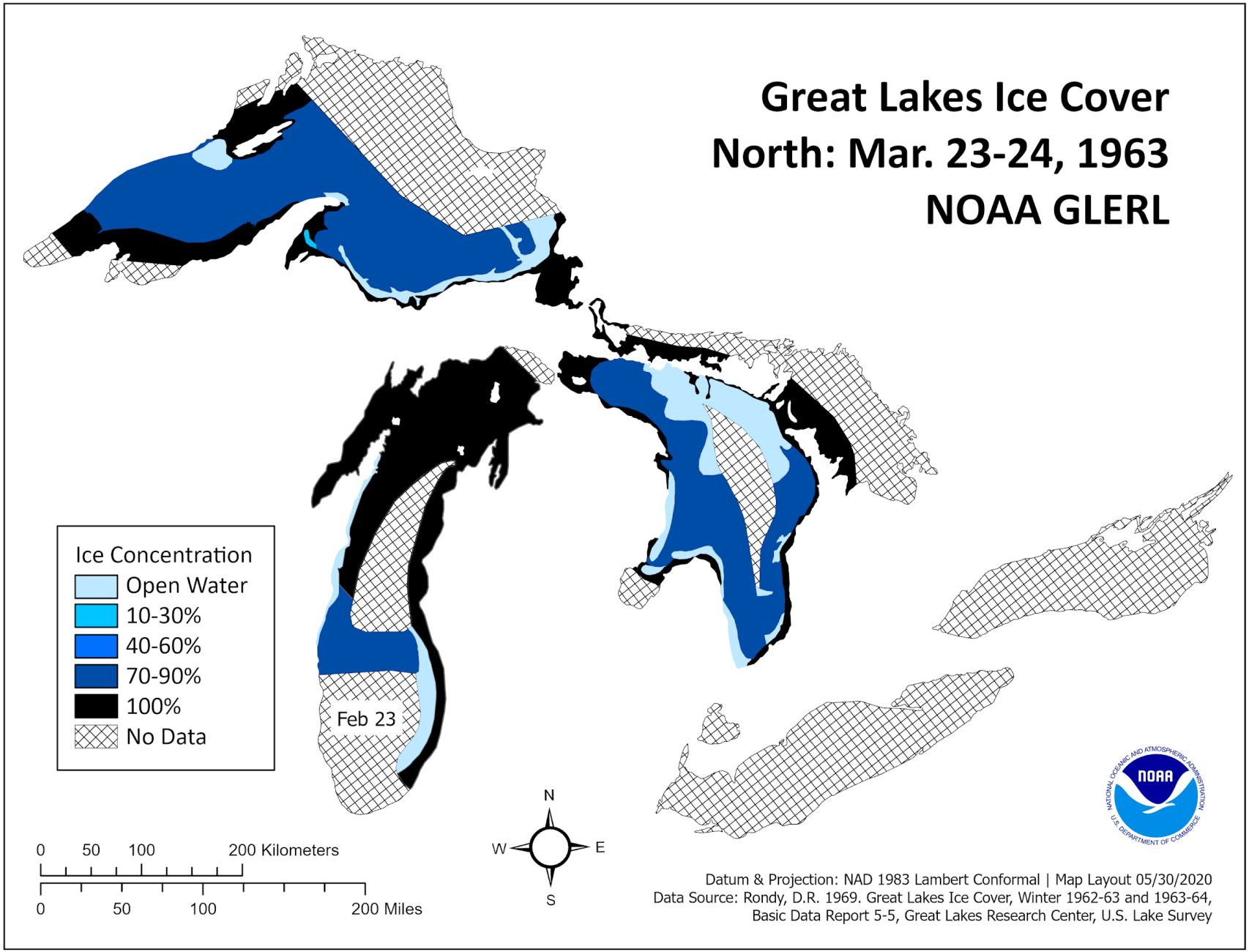
Appendices

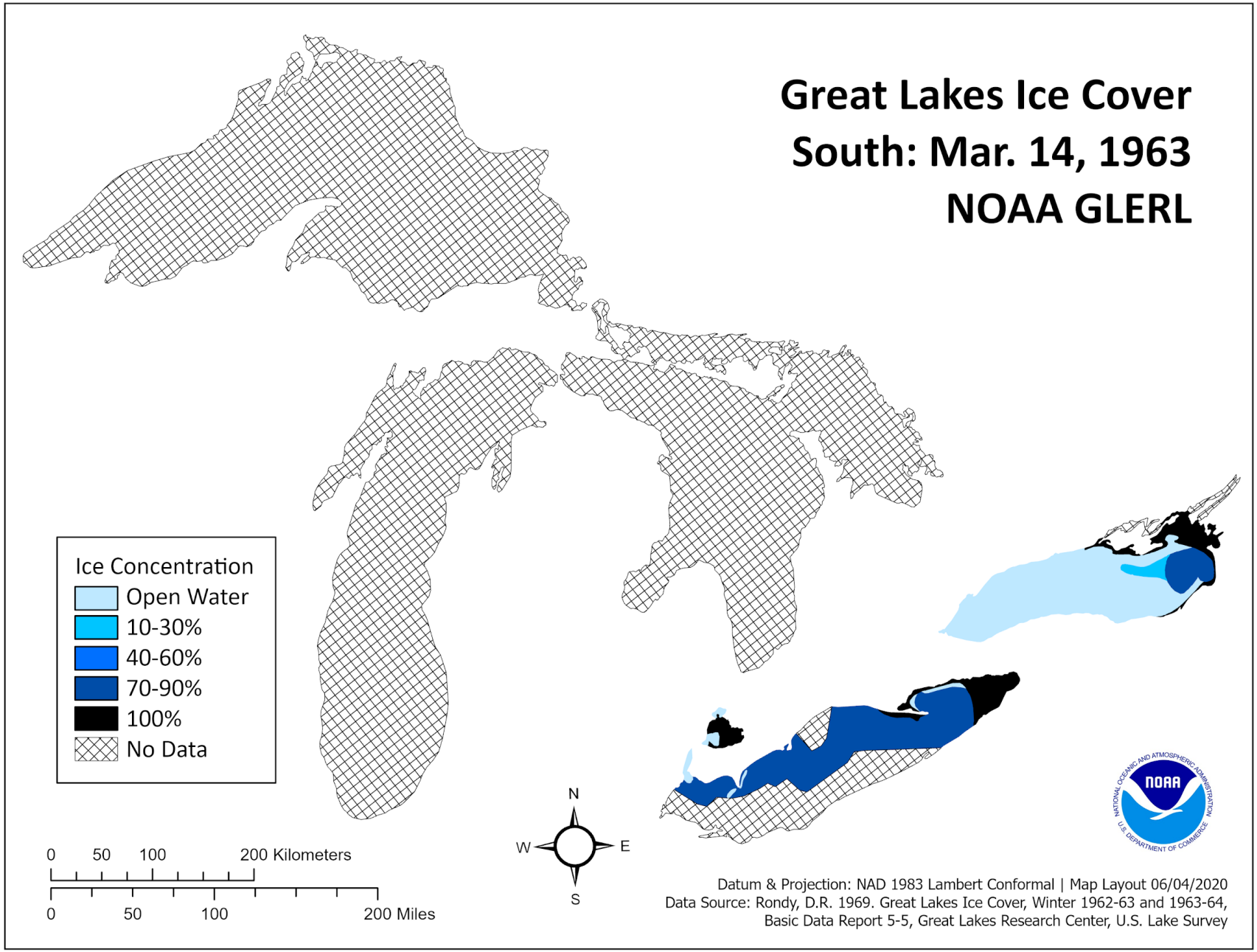
NEW NOAA ICE COVER MAPS: 1963-1972

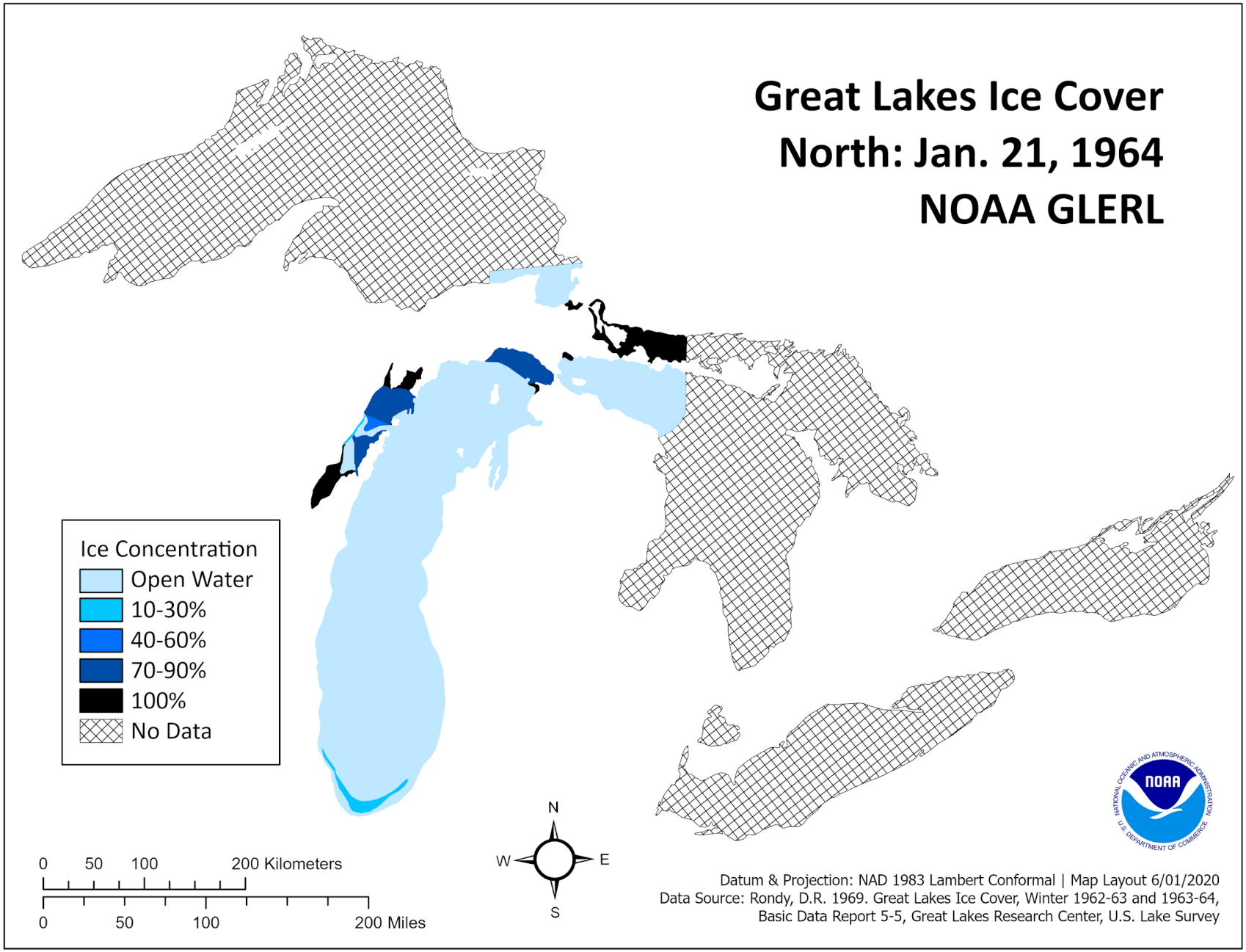
Year	Map	Date	Superior	Huron	Michigan	Erie	Ontario	St. Claire
63	*North*	3/23-3/24	93%	80%	59%	-	-	-
	South	3/14	-	-	-	97%	28%	74%
	Report		95%	97%	63%	98%	51%	-
64	*North*	1/21	NA	NA	10%	-	-	-
	South	1/15, 2/11	-	-	-	96%	2%	80%
	Report		31%	32%	13%	91%	12%	-
65	*Combined*	3/28-4/04	100%	66%	31%	69%	NA	96%
	Report		-	60%	-	-	-	99%
66	*North*	*missing*	NA	NA	NA	-	-	-
	South	2/21-2/25	71%	38%	22%	88%	24%	100%
	Report		60%	20%	15%	85%	10%	-
67	*North*	3/08-3/17	92%	80%	47%	89%	16%	100%
	South	2/21-3/02	92%	78%	64%	94%	23%	100%
	Report		88%	80%	46%	85%	12%	-
68	*North*	3/12-3/18	100%	61%	39%	62%	10%	71%
	South	2/20-2/25	95%	63%	31%	100%	22%	100%
	Report		90%	50%	30%	98%	10%	-
69	*North*	3/10-3/20	54%	60%	15%	55%	13%	100%
	South	2/04-2/13	32%	56%	21%	91%	24%	100%
	Report		40%	50%	15%	80%	10%	-
70	*North*	3/10-3/17	95%	67%	45%	64%	16%	100%
	South	2/16-2/18	48%	71%	20%	100%	16%	100%
	Report		85%	50%	30%	95%	15%	-
*71	*Combined*	2/21-3/02	73%	77%	33%	85%	17%	90%
	Report	-	48%	45%	27%	92%	10%	-
72	*Combined*	2/27-3/06	100%	82%	49%	94%	40%	100%
	Report		95%	70%	40%	95%	20%	-

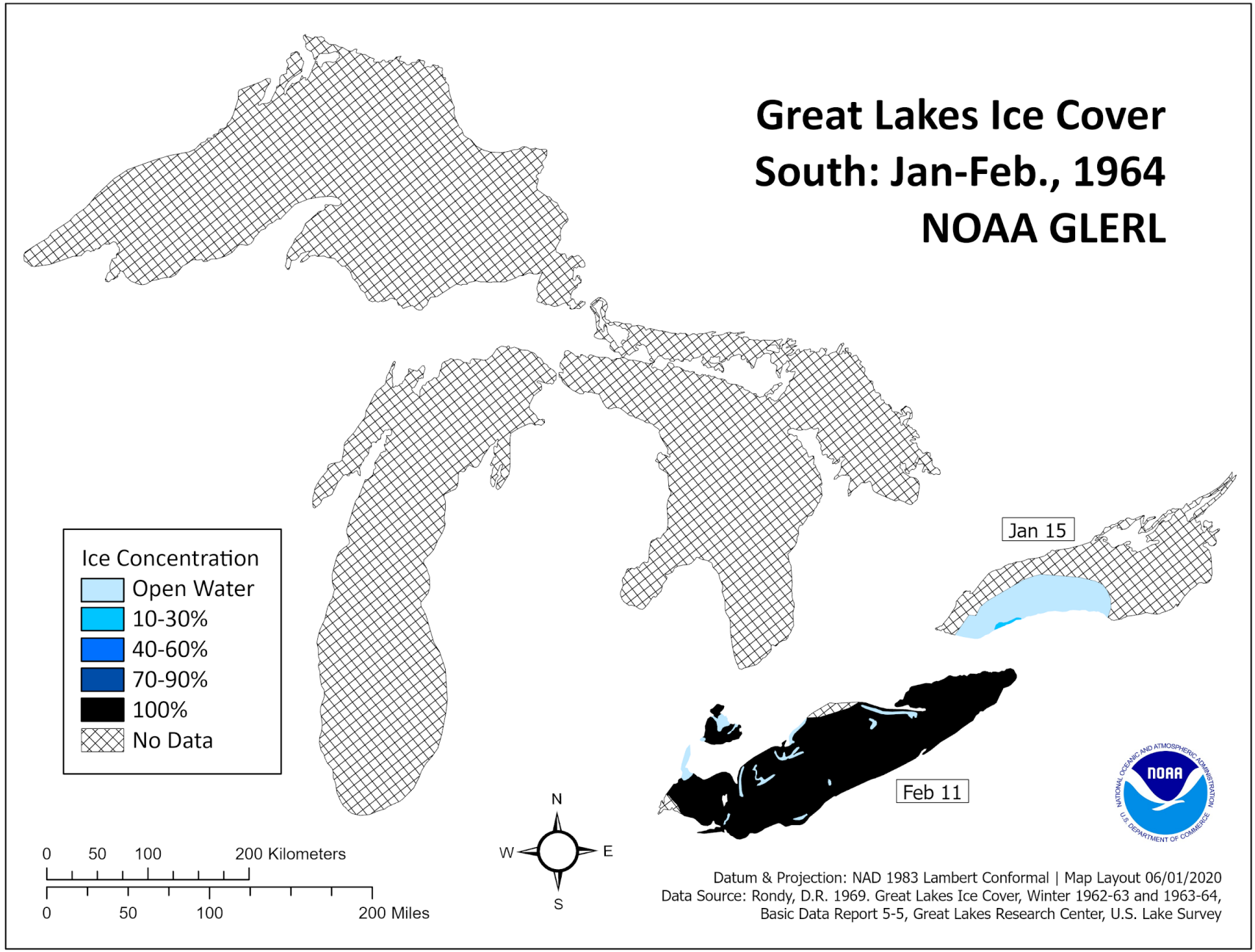
*1971 indicates the year reports switched from Rondy to Assel, with one ice cover maximum period now recorded instead of two (one for the Northern lakes and one for the Southern lakes).

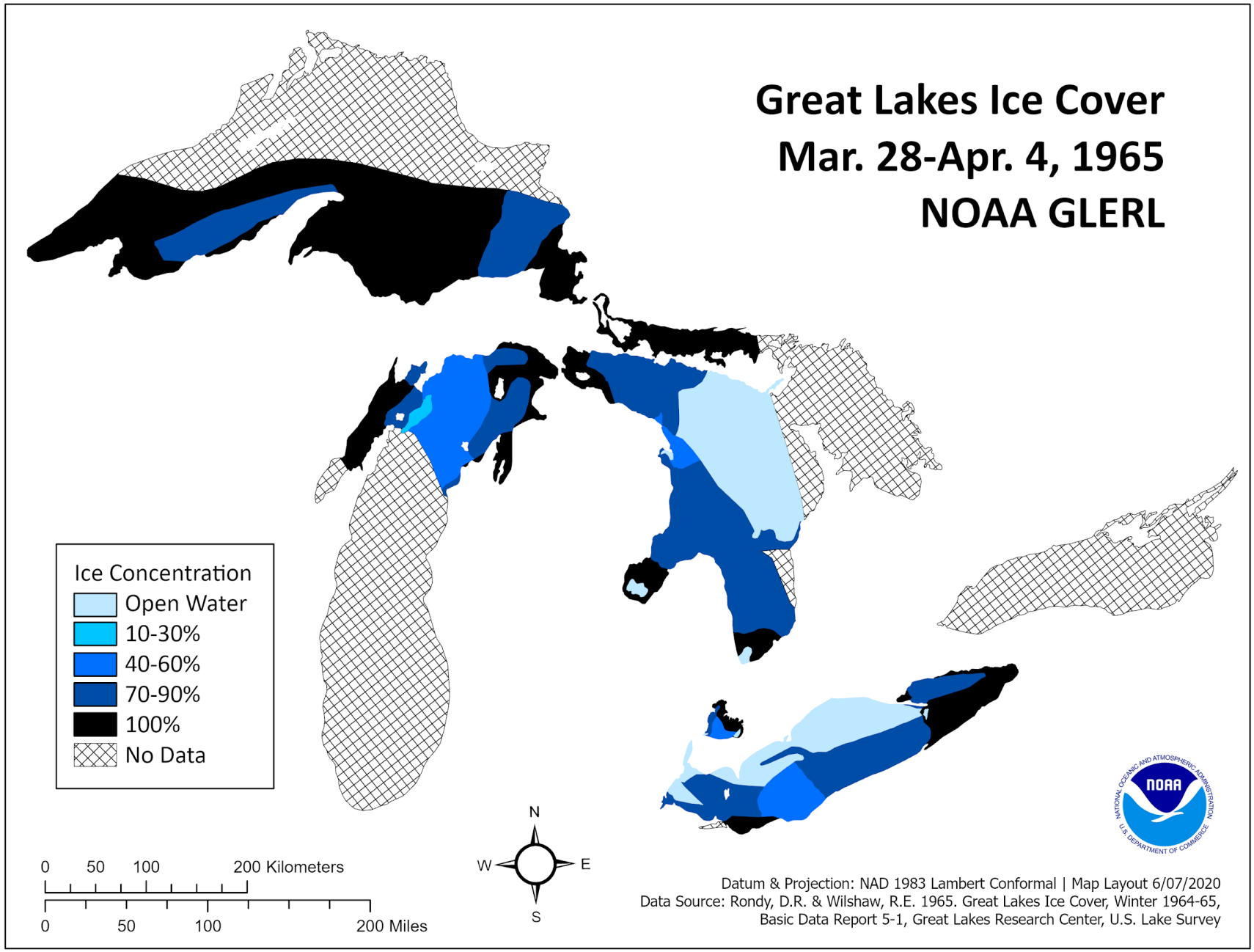
The chart above shows the original ice cover percentage recorded in the official lake reports, compared with the calculated values of ice cover percentage based on the area of ice in the newly generated maps below:

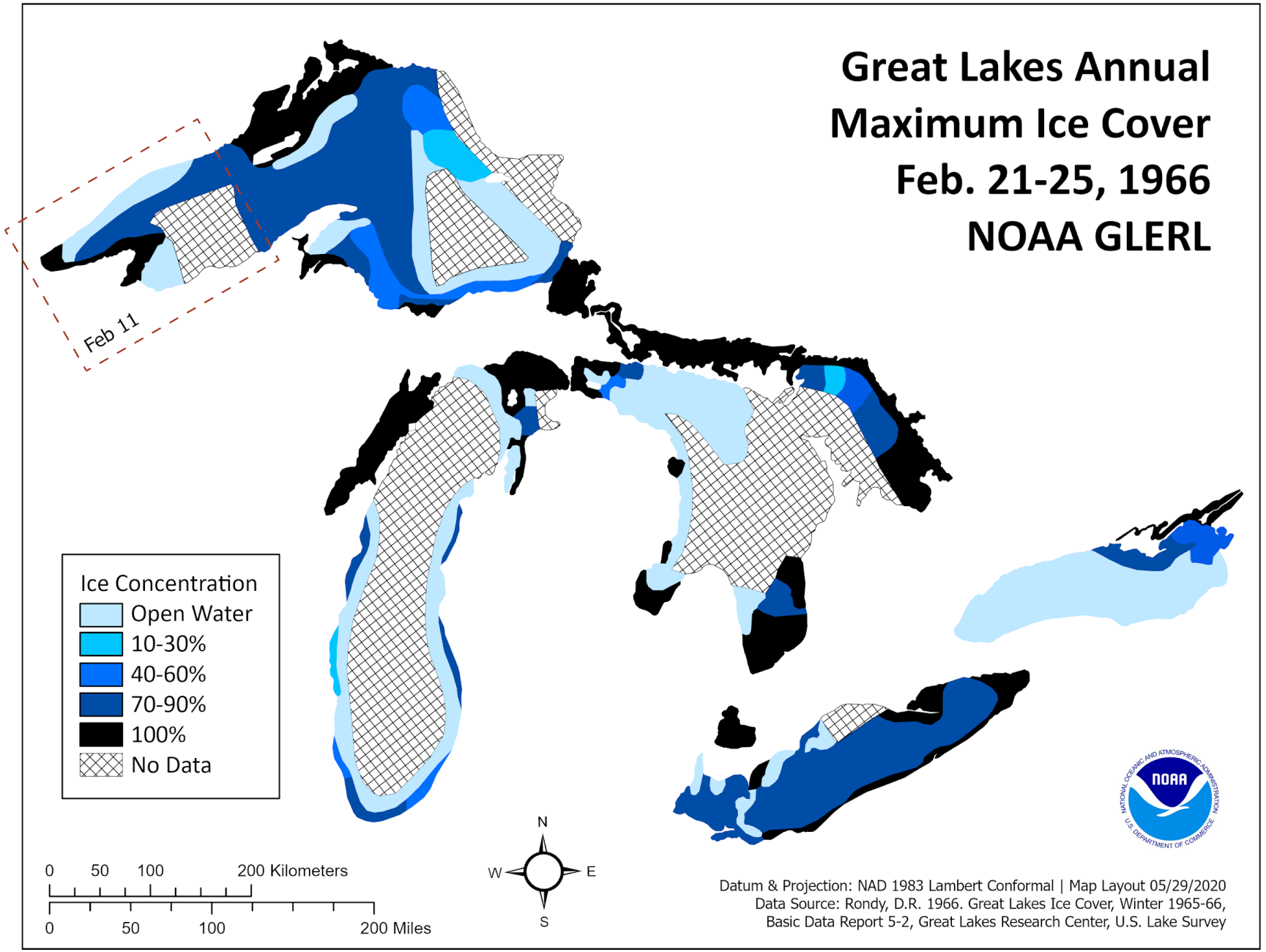

Great Lakes Ice Cover
North: Mar. 23-24, 1963
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
No Data
Feb 23
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/30/2020
Data Source: Rondy, D.R. 1969. Great Lakes Ice Cover, Winter 1962-63 and 1963-64,
Basic Data Report 5-5, Great Lakes Research Center, U.S. Lake Survey

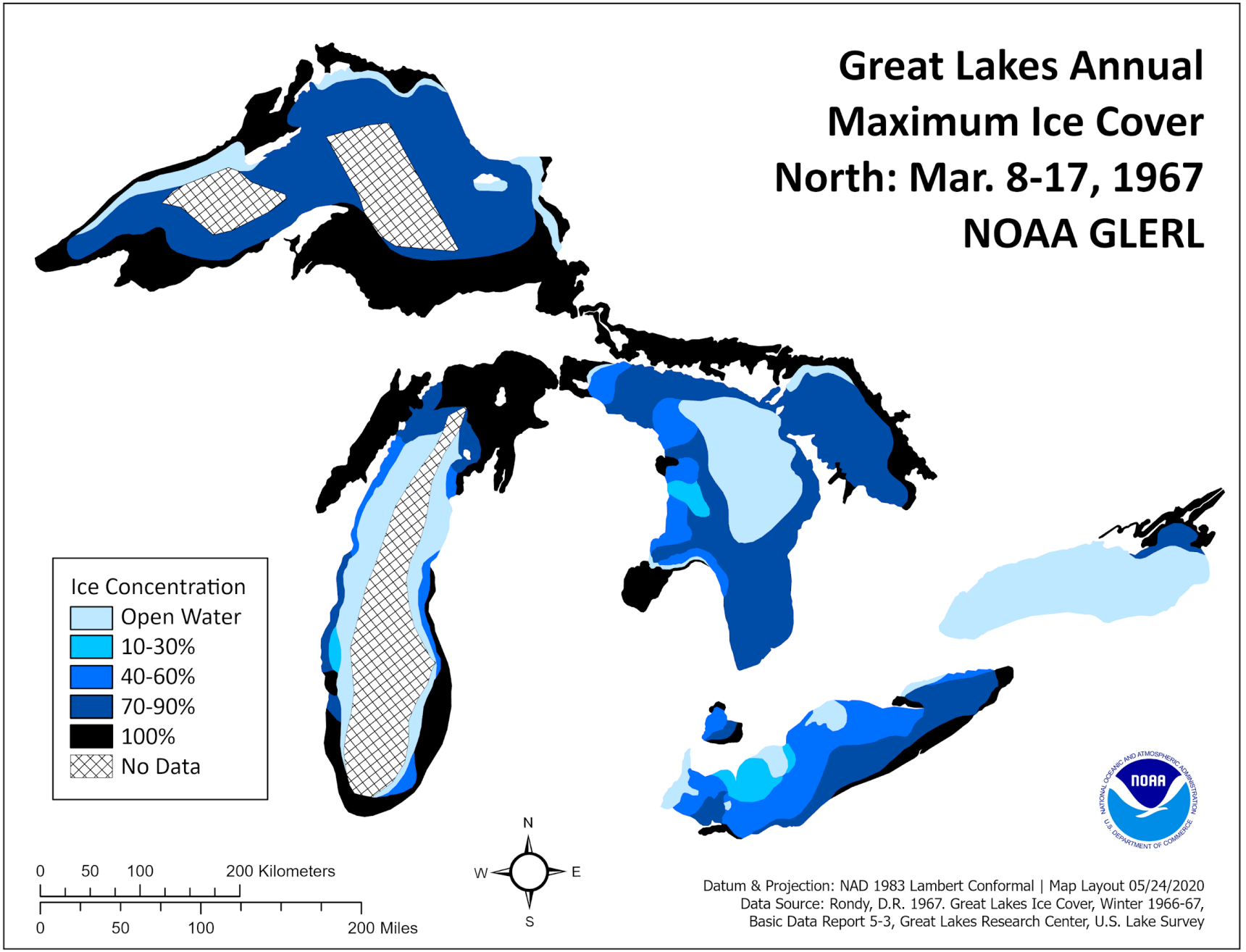

Great Lakes Ice Cover
South: Mar. 14, 1963
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
No Data
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 06/04/2020
Data Source: Rondy, D.R. 1969. Great Lakes Ice Cover, Winter 1962-63 and 1963-64,
Basic Data Report 5-5, Great Lakes Research Center, U.S. Lake Survey

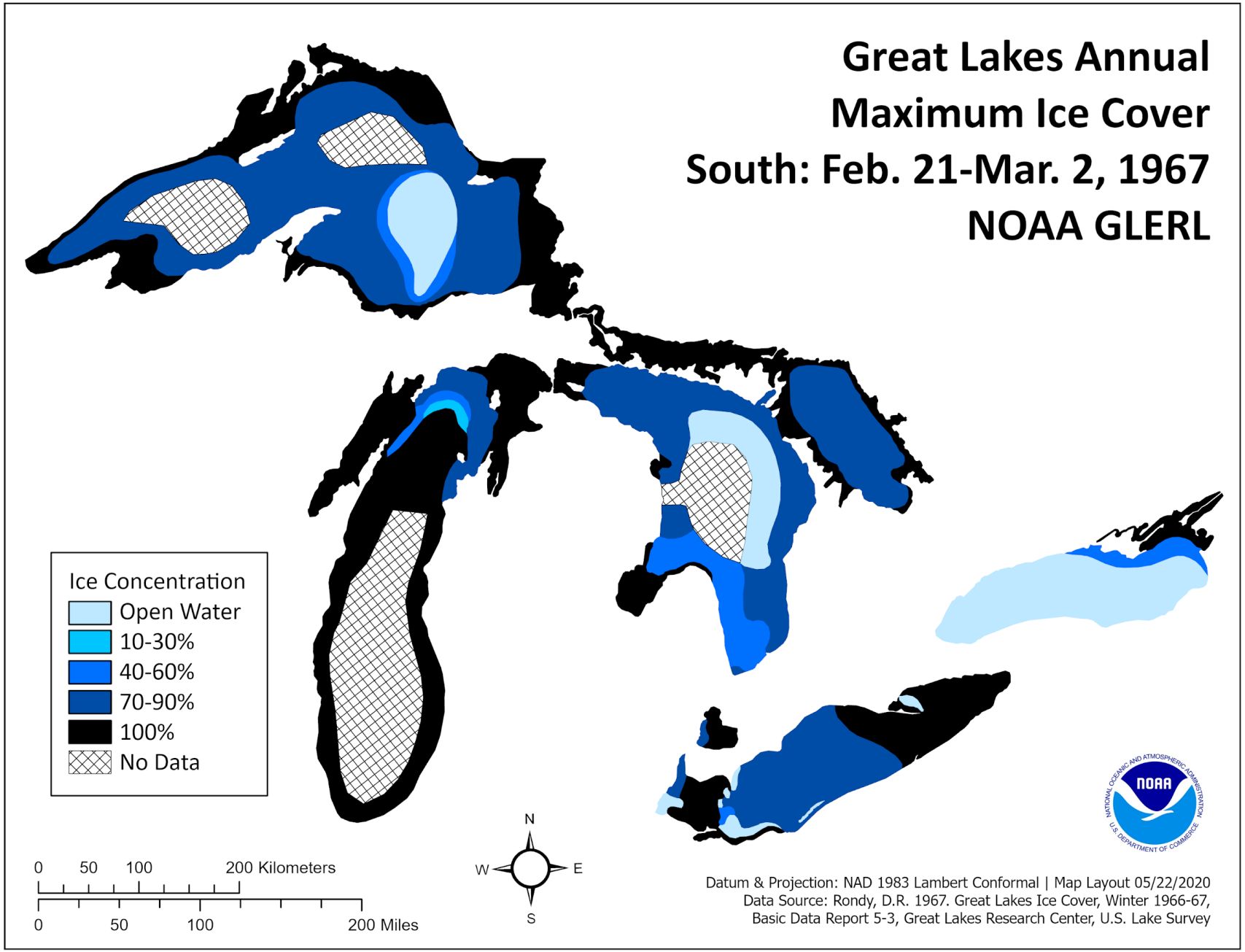

Great Lakes Ice Cover
North: Jan. 21, 1964
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
No Data
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 6/01/2020
Data Source: Rondy, D.R. 1969. Great Lakes Ice Cover, Winter 1962-63 and 1963-64,
Basic Data Report 5-5, Great Lakes Research Center, U.S. Lake Survey

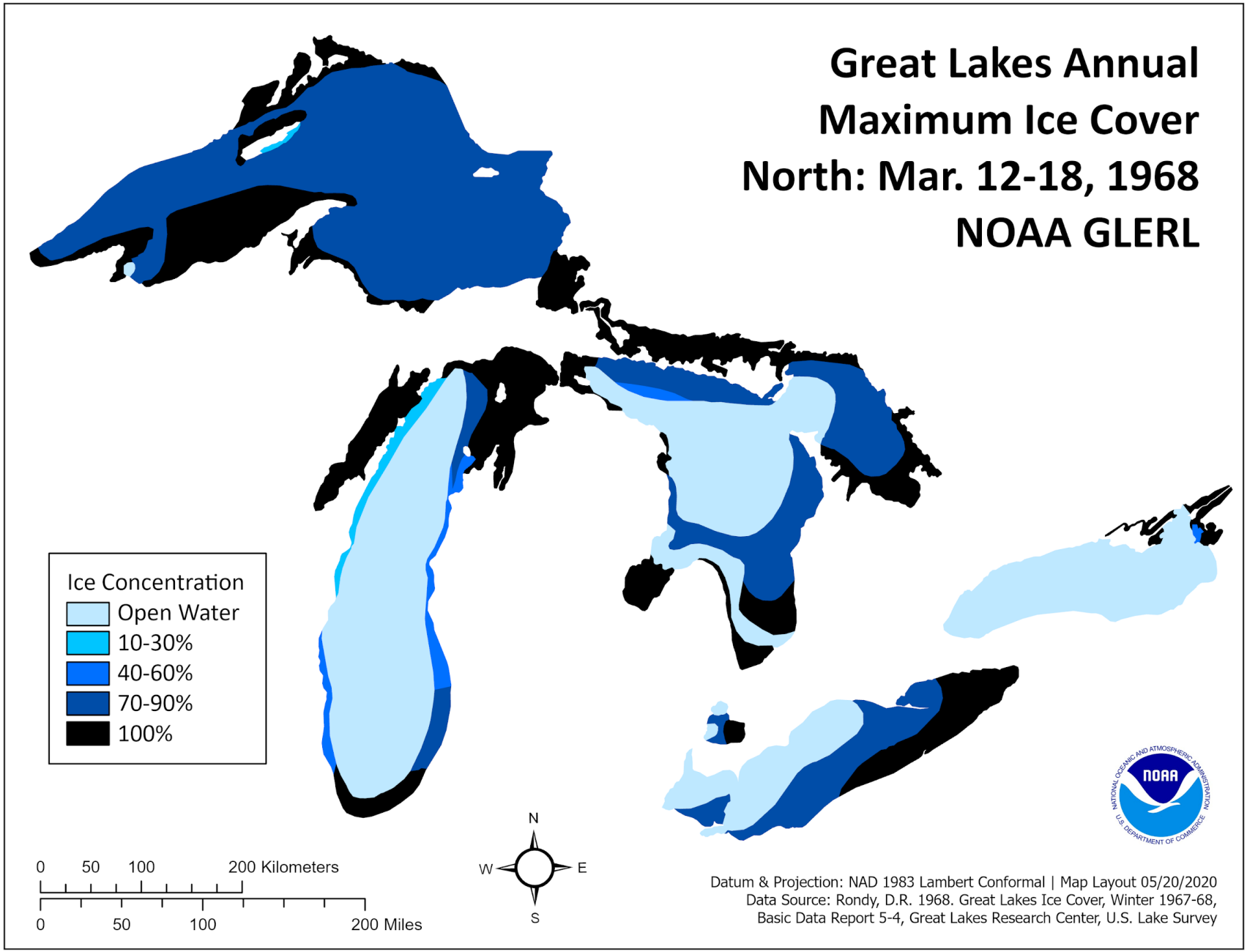

Great Lakes Ice Cover
South: Jan-Feb., 1964
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
No Data
Jan 15
Feb 11
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 06/01/2020
Data Source: Rondy, D.R. 1969. Great Lakes Ice Cover, Winter 1962-63 and 1963-64,
Basic Data Report 5-5, Great Lakes Research Center, U.S. Lake Survey

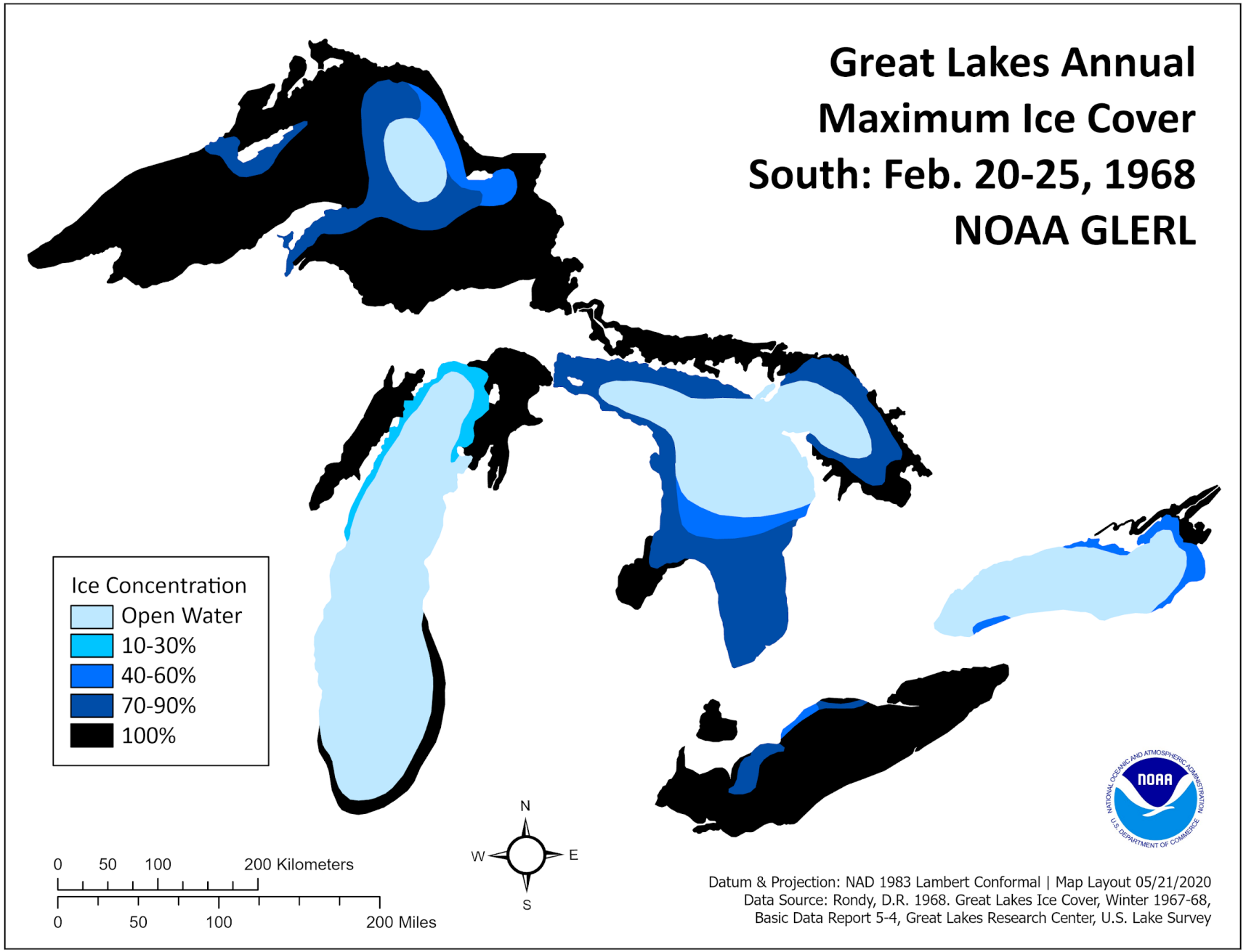

Great Lakes Ice Cover
Mar. 28-Apr. 4, 1965
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
No Data
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 6/07/2020
Data Source: Rondy, D.R. & Wilshaw, R.E. 1965. Great Lakes Ice Cover, Winter 1964-65,
Basic Data Report 5-1, Great Lakes Research Center, U.S. Lake Survey

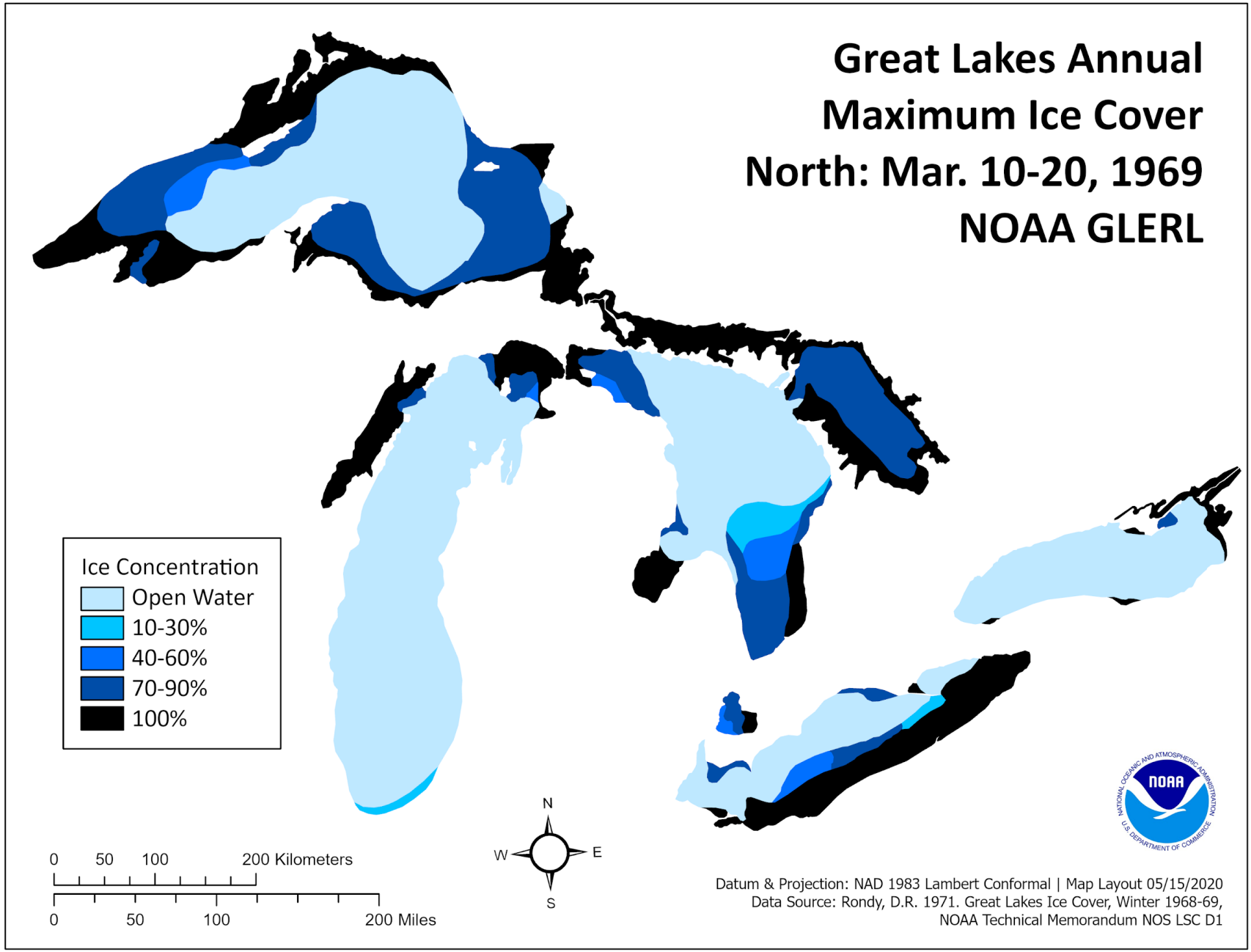

Great Lakes Annual
Maximum Ice Cover
Feb. 21-25, 1966
NOAA GLERL
Feb 11
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
No Data
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/29/2020
Data Source: Rondy, D.R. 1966. Great Lakes Ice Cover, Winter 1965-66,
Basic Data Report 5-2, Great Lakes Research Center, U.S. Lake Survey

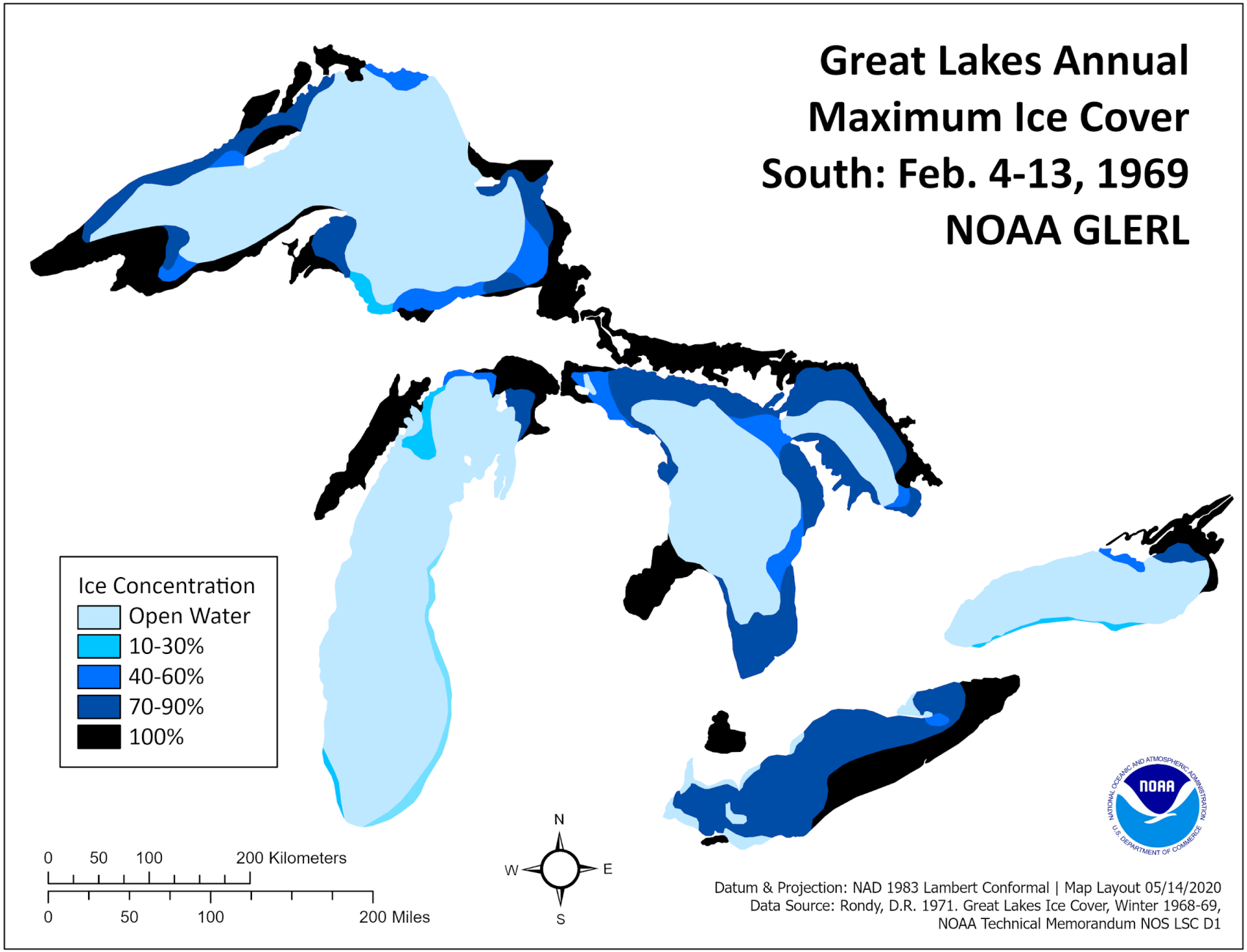

Great Lakes Annual
Maximum Ice Cover
North: Mar. 8-17, 1967
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
No Data
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/24/2020
Data Source: Rondy, D.R. 1967. Great Lakes Ice Cover, Winter 1966-67,
Basic Data Report 5-3, Great Lakes Research Center, U.S. Lake Survey

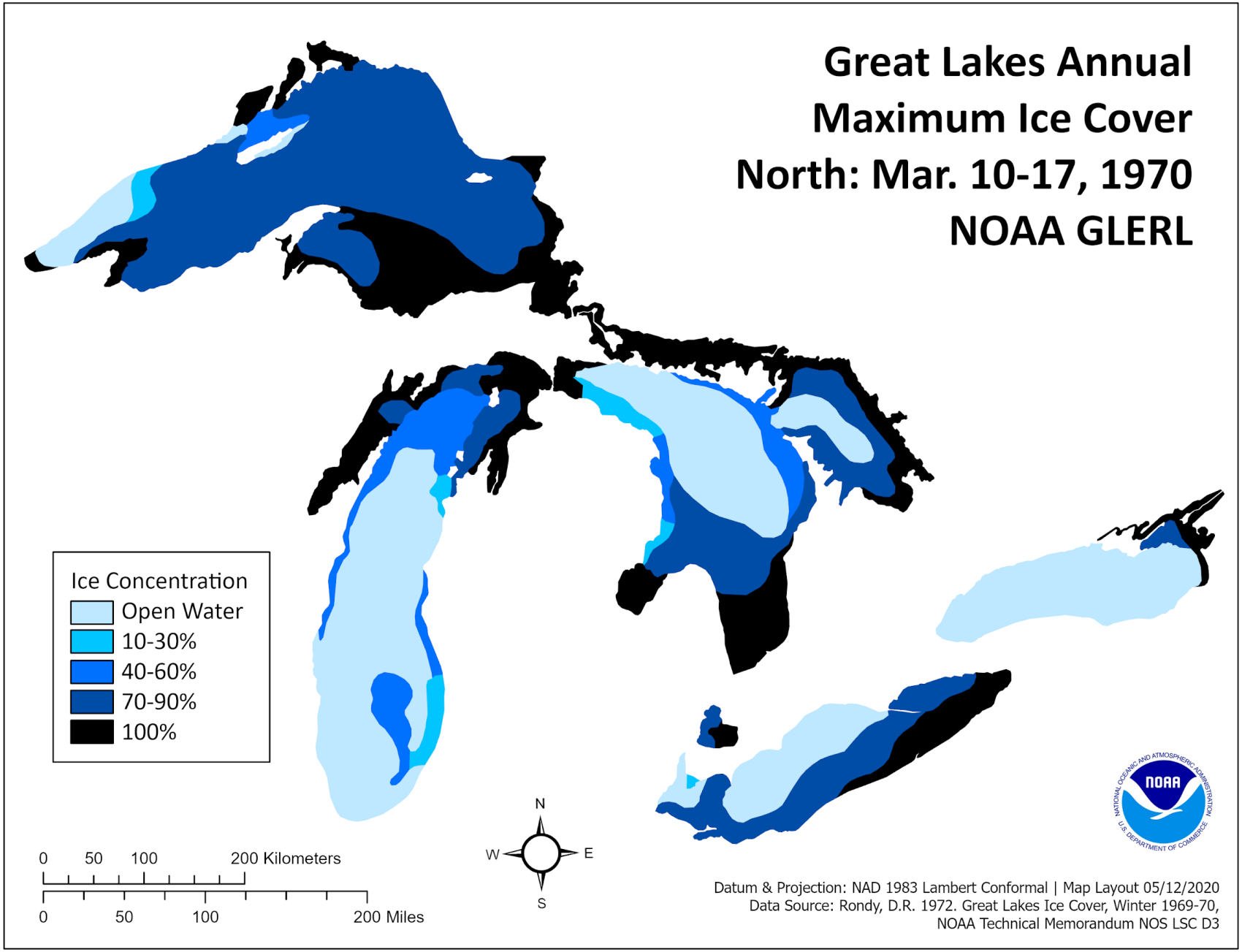

Great Lakes Annual
Maximum Ice Cover
South: Feb. 21-Mar. 2, 1967
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
No Data
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/22/2020
Data Source: Rondy, D.R. 1967. Great Lakes Ice Cover, Winter 1966-67,
Basic Data Report 5-3, Great Lakes Research Center, U.S. Lake Survey

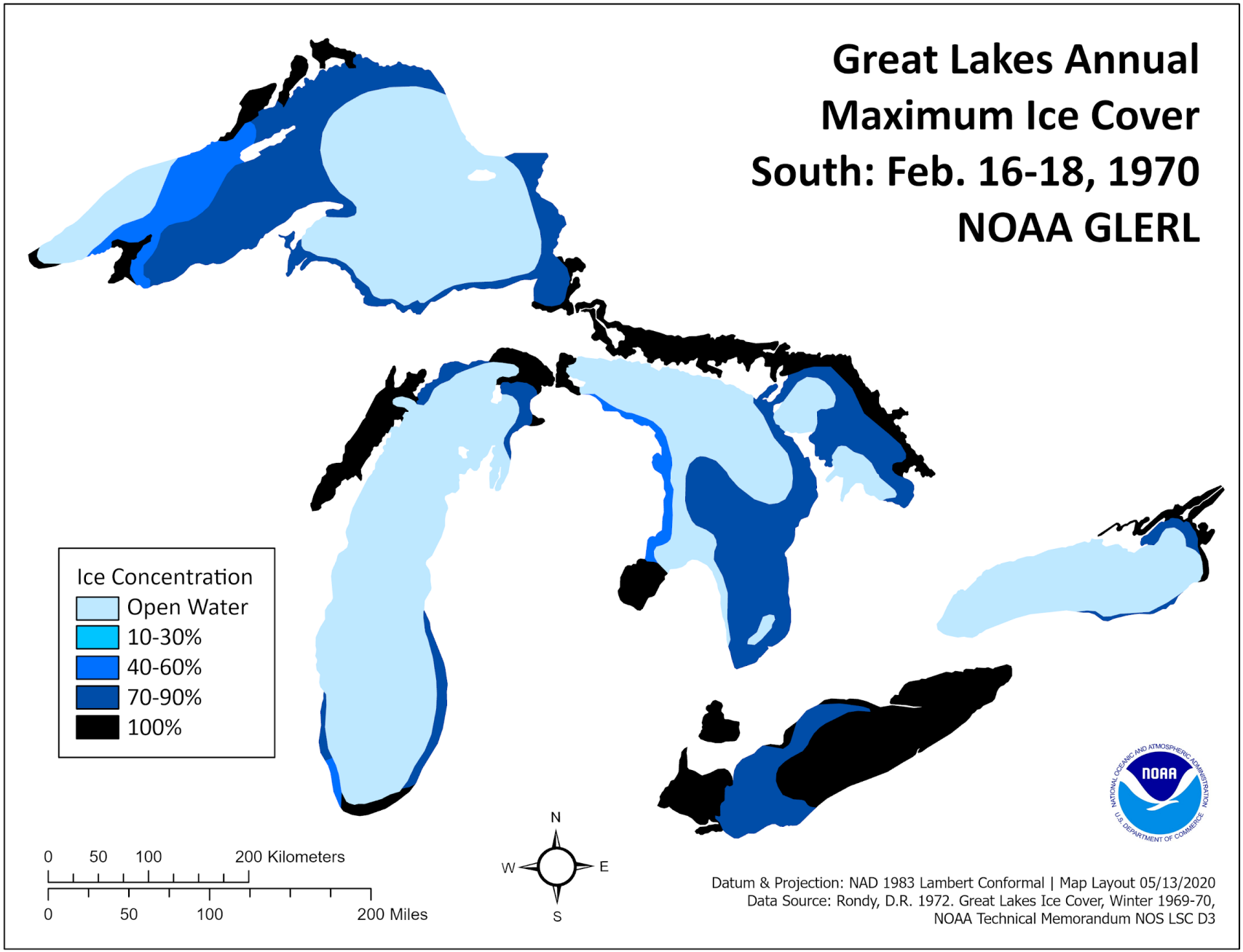

Great Lakes Annual
Maximum Ice Cover
North: Mar. 12-18, 1968
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/20/2020
Data Source: Rondy, D.R. 1968. Great Lakes Ice Cover, Winter 1967-68,
Basic Data Report 5-4, Great Lakes Research Center, U.S. Lake Survey

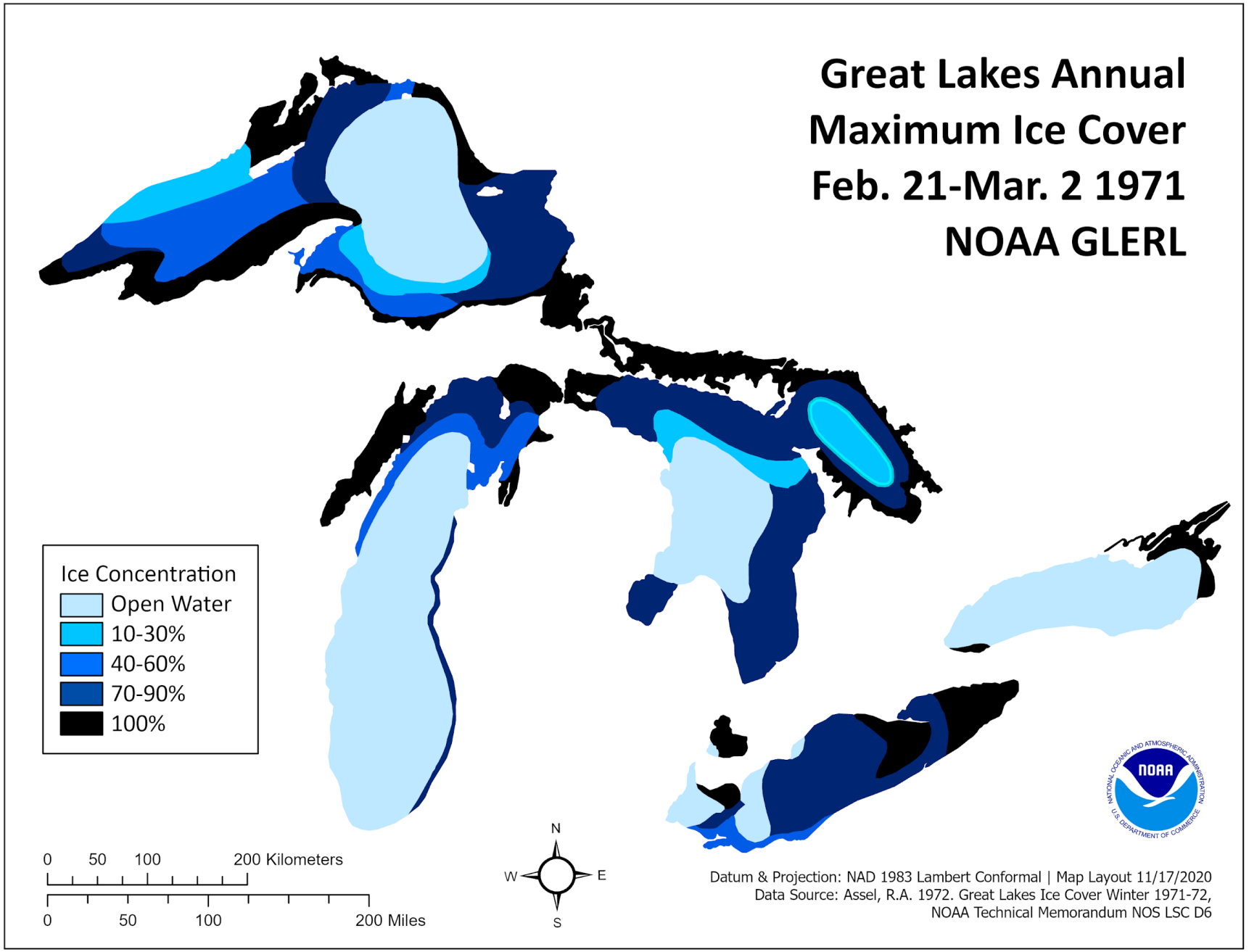

Great Lakes Annual
Maximum Ice Cover
South: Feb. 20-25, 1968
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/21/2020
Data Source: Rondy, D.R. 1968. Great Lakes Ice Cover, Winter 1967-68,
Basic Data Report 5-4, Great Lakes Research Center, U.S. Lake Survey

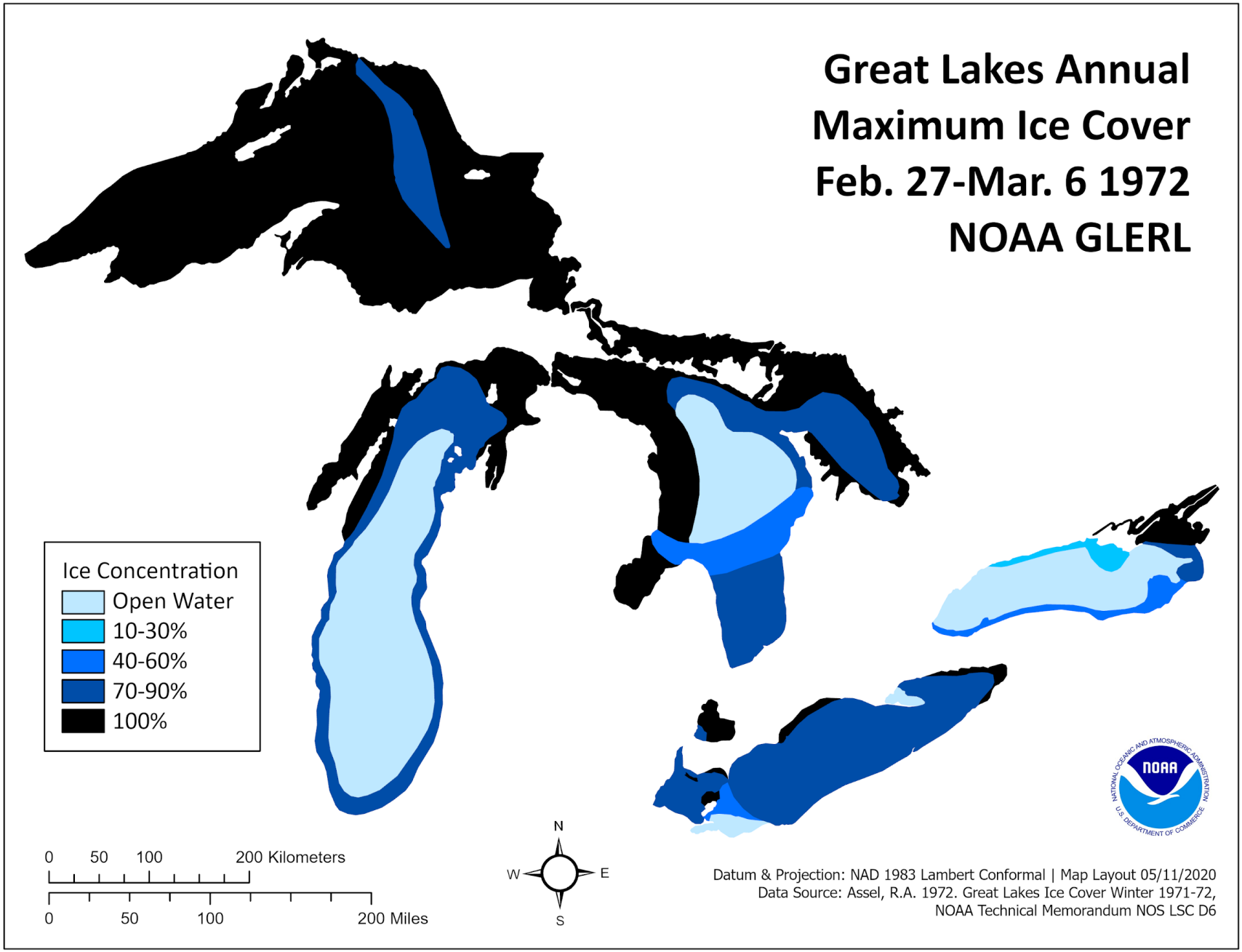

Great Lakes Annual
Maximum Ice Cover
North: Mar. 10-20, 1969
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/15/2020
Data Source: Rondy, D.R. 1971. Great Lakes Ice Cover, Winter 1968-69,
NOAA Technical Memorandum NOS LSC D1


Great Lakes Annual
Maximum Ice Cover
South: Feb. 4-13, 1969
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/14/2020
Data Source: Rondy, D.R. 1971. Great Lakes Ice Cover, Winter 1968-69,
NOAA Technical Memorandum NOS LSC D1


Great Lakes Annual
Maximum Ice Cover
North: Mar. 10-17, 1970
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/12/2020
Data Source: Rondy, D.R. 1972. Great Lakes Ice Cover, Winter 1969-70,
NOAA Technical Memorandum NOS LSC D3


Great Lakes Annual
Maximum Ice Cover
South: Feb. 16-18, 1970
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/13/2020
Data Source: Rondy, D.R. 1972. Great Lakes Ice Cover, Winter 1969-70,
NOAA Technical Memorandum NOS LSC D3


Great Lakes Annual
Maximum Ice Cover
Feb. 21-Mar. 2 1971
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 11/17/2020
Data Source: Assel, R.A. 1972. Great Lakes Ice Cover Winter 1971-72,
NOAA Technical Memorandum NOS LSC D6


Great Lakes Annual
Maximum Ice Cover
Feb. 27-Mar. 6 1972
NOAA GLERL
Ice Concentration
Open Water
10-30%
40-60%
70-90%
100%
0 50 100 200 Kilometers
0 50 100 200 Miles
N
W
E
S
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
Datum & Projection: NAD 1983 Lambert Conformal | Map Layout 05/11/2020
Data Source: Assel, R.A. 1972. Great Lakes Ice Cover Winter 1971-72,
NOAA Technical Memorandum NOS LSC D6

ICE COVER HINDCASTS, 1898-1983

A. Hindcasted ice cover (percent), by lake and year. Based on regression analysis.

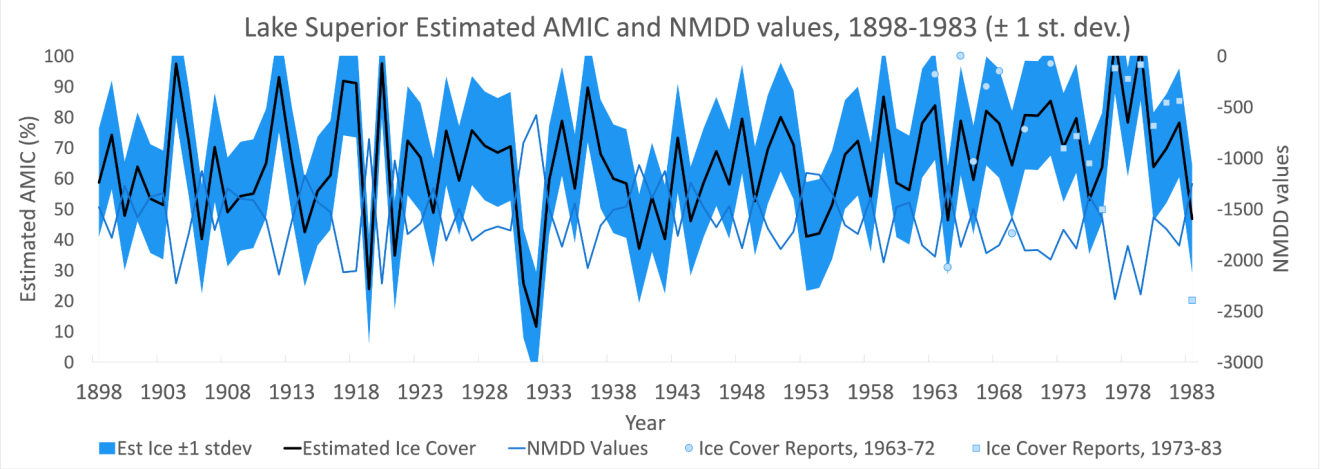
Ice Cover Hindcasts, by Lake and Year

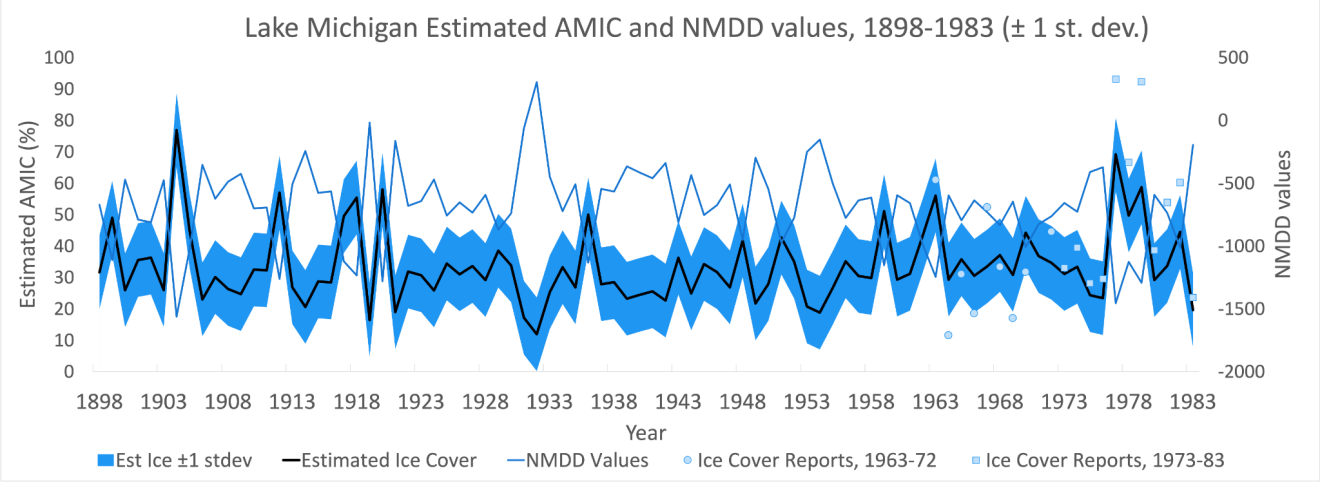
Year	Lake Superior	Lake Michigan	Lake Huron	Lake Erie	Lake Ontario
1898	58.6	31.6	56.3	90.4	21.0
1899	74.2	49.0	97.8	95.6	57.4
1900	47.8	25.9	78.7	94.8	67.9
1901	63.8	35.5	75.8	93.7	64.3
1902	53.4	36.3	64.0	95.5	34.7
1903	51.3	25.9	63.3	92.4	21.5
1904	97.5	76.9	124.9	102.3	136.8
1905	71.6	44.9	107.4	100.0	98.0
1906	40.2	22.9	60.3	90.0	36.9
1907	70.2	30.1	80.4	94.1	59.2
1908	49.0	26.3	70.5	91.8	55.0
1909	54.2	24.7	56.8	51.8	32.7
1910	55.0	32.5	66.5	95.1	28.5
1911	65.0	32.2	66.6	91.7	42.7
1912	93.0	56.9	109.0	100.6	144.3
1913	65.5	26.8	62.4	90.5	26.9
1914	42.4	20.6	72.0	93.2	71.8
1915	55.9	28.7	66.3	93.3	35.9
1916	61.0	28.4	81.8	94.4	59.6
1917	91.8	49.5	96.2	97.2	63.6
1918	91.1	55.4	115.7	101.4	88.4
1919	23.9	16.4	37.3	17.4	19.7
1920	97.6	58.0	111.8	99.9	82.0
1921	34.8	19.0	35.6	31.1	14.9
1922	72.3	31.8	62.5	90.9	34.2
1923	66.9	30.7	92.6	92.5	84.1
1924	48.8	25.8	69.2	92.4	50.1
1925	75.6	34.4	79.3	92.8	31.5
1926	59.3	31.0	85.3	94.7	59.0
1927	75.7	33.6	67.4	91.1	27.1
1928	70.6	29.1	60.3	90.6	36.9
1929	68.5	38.5	71.2	91.9	31.0
1930	70.5	33.8	68.6	91.4	34.1
1931	25.8	17.1	50.5	74.1	28.7
1932	11.6	11.9	33.2	18.5	17.5
1933	59.4	25.2	47.0	29.5	17.5
1934	78.8	33.2	96.2	95.2	92.4
1935	56.7	26.8	69.8	92.4	43.7
1936	89.7	50.0	91.6	98.9	53.4
1937	67.9	27.8	53.0	46.1	24.0
1938	59.9	28.4	66.5	90.7	27.0
1939	58.4	23.2	63.7	99.6	45.3
1940	37.1	24.4	74.2	96.2	93.8
1941	53.9	25.5	67.2	91.7	72.9
1942	40.2	22.6	52.8	91.5	38.6
1943	73.2	36.2	82.3	94.1	57.2
1944	46.1	24.9	55.2	91.0	38.0
1945	58.5	34.2	74.0	96.0	36.7
1946	68.8	31.6	62.4	91.7	22.7
1947	58.0	26.8	63.5	91.0	42.5
1948	79.5	41.7	82.7	95.9	68.4

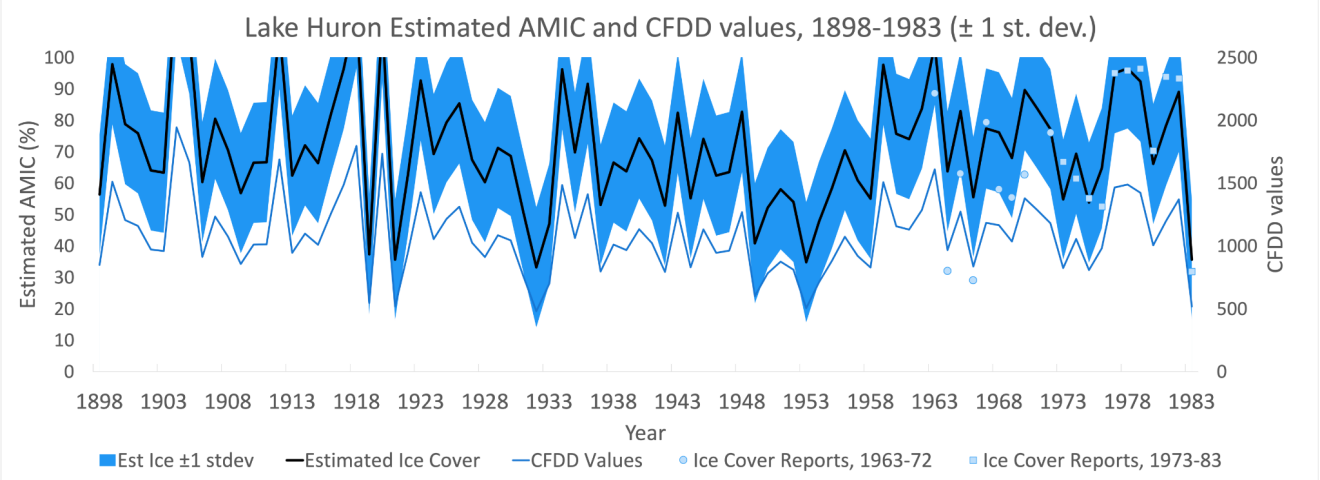
Ice Cover Hindcasts, by Lake and Year

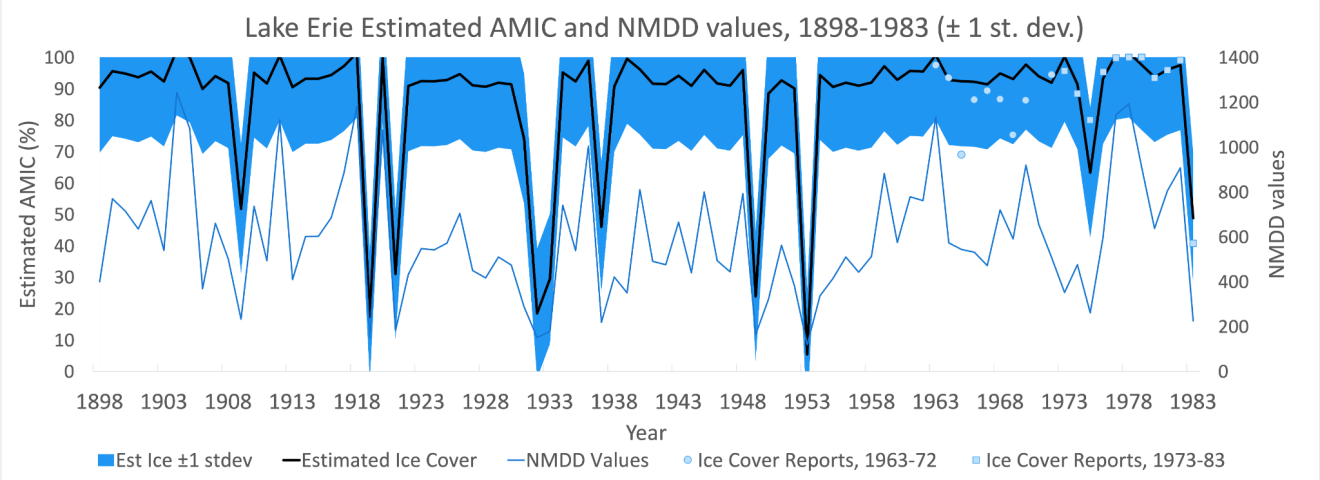
Year	Lake Superior	Lake Michigan	Lake Huron	Lake Erie	Lake Ontario
1949	52.6	21.7	40.8	24.0	21.5
1950	69.5	27.8	52.1	88.5	36.6
1951	80.0	42.6	58.0	92.7	26.4
1952	71.0	35.0	54.0	90.2	28.3
1953	41.1	20.7	34.8	5.5	16.3
1954	42.0	18.8	47.7	94.4	30.7
1955	51.4	26.7	58.3	90.6	40.6
1956	67.8	35.1	70.4	92.0	44.2
1957	72.2	30.4	60.8	91.0	34.1
1958	53.7	29.8	55.0	92.0	40.8
1959	86.7	51.0	97.6	97.2	62.4
1960	58.6	29.2	75.6	92.9	56.4
1961	56.2	31.1	74.0	95.7	40.7
1962	78.0	43.3	83.7	95.5	48.0
1963	83.8	56.0	104.0	100.7	59.9
1964	46.4	29.3	63.7	92.8	28.4
1965	78.8	35.7	82.9	92.4	52.4
1966	59.5	30.5	55.5	92.3	34.8
1967	82.0	33.5	77.4	91.4	36.1
1968	78.0	37.1	76.1	94.9	59.8
1969	64.3	30.8	68.0	93.1	36.4
1970	80.7	44.2	89.6	97.7	74.7
1971	80.5	36.8	83.5	94.0	58.4
1972	85.3	34.6	77.2	91.9	51.7
1973	70.1	31.1	54.8	100.3	22.2
1974	79.7	33.3	69.3	91.5	42.6
1975	53.0	24.3	53.8	63.3	27.2
1976	63.7	23.4	64.7	93.2	32.4
1977	105.5	69.1	95.0	100.9	45.9
1978	78.3	49.7	96.5	101.6	93.9
1979	103.1	58.7	92.3	97.6	63.0
1980	63.8	29.2	66.1	93.7	50.6
1981	69.7	33.6	78.3	96.1	39.4
1982	78.2	44.4	89.0	97.6	69.9
1983	46.8	19.6	35.6	48.8	20.4

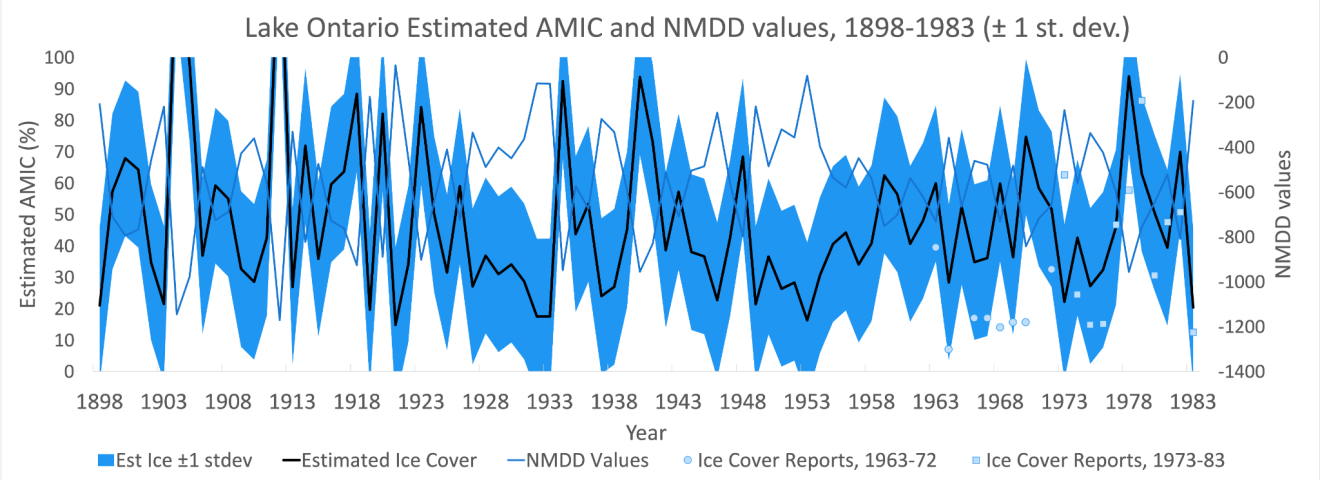
B. Hindcasted ice cover (± 1 standard deviation) and CFDD or NMDD time series



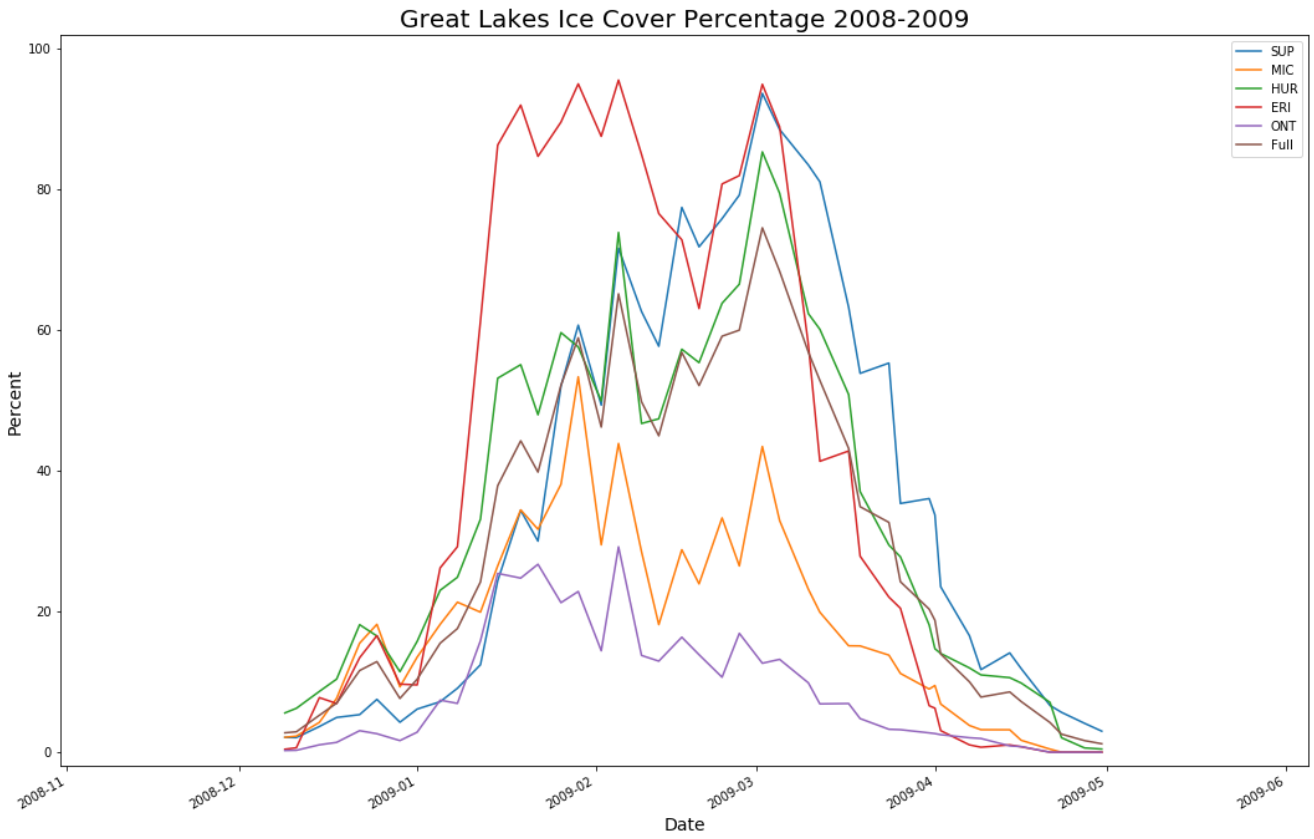


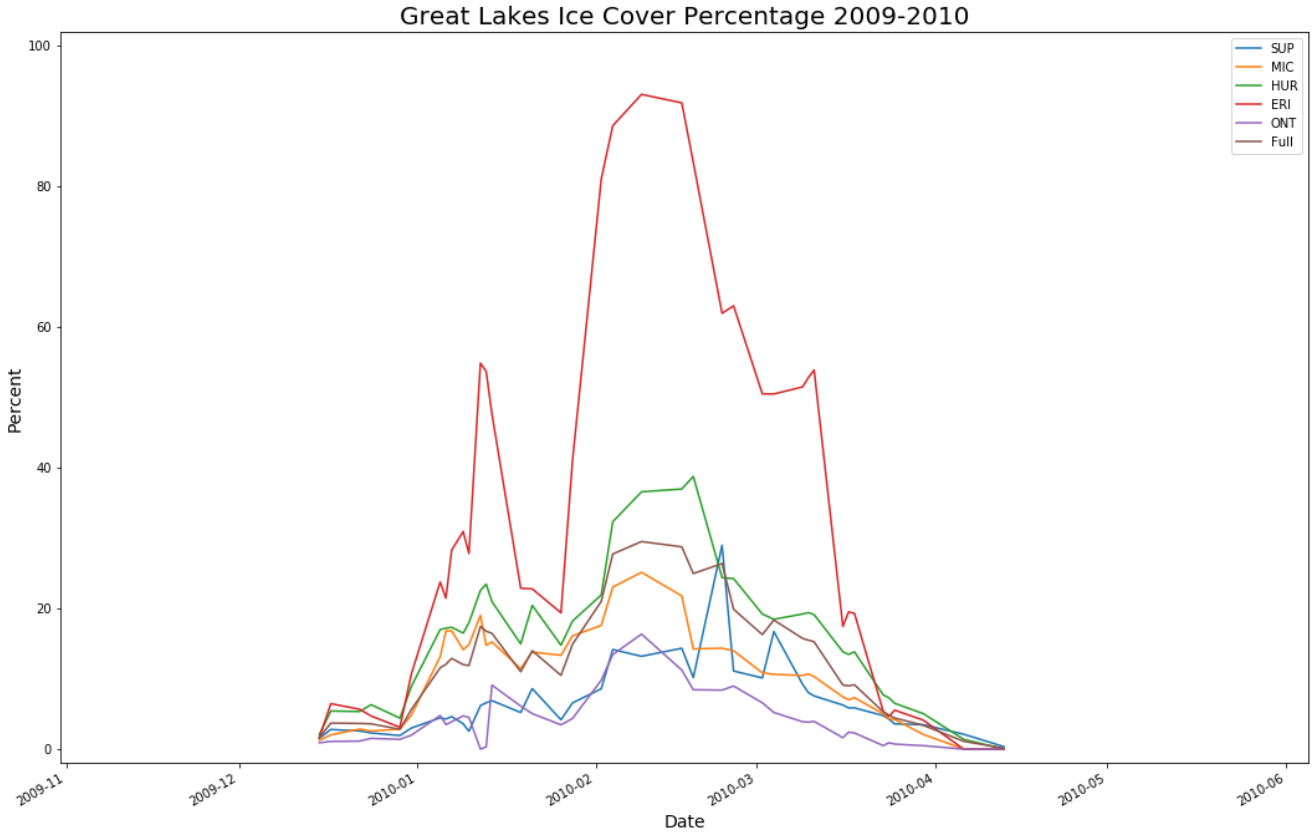


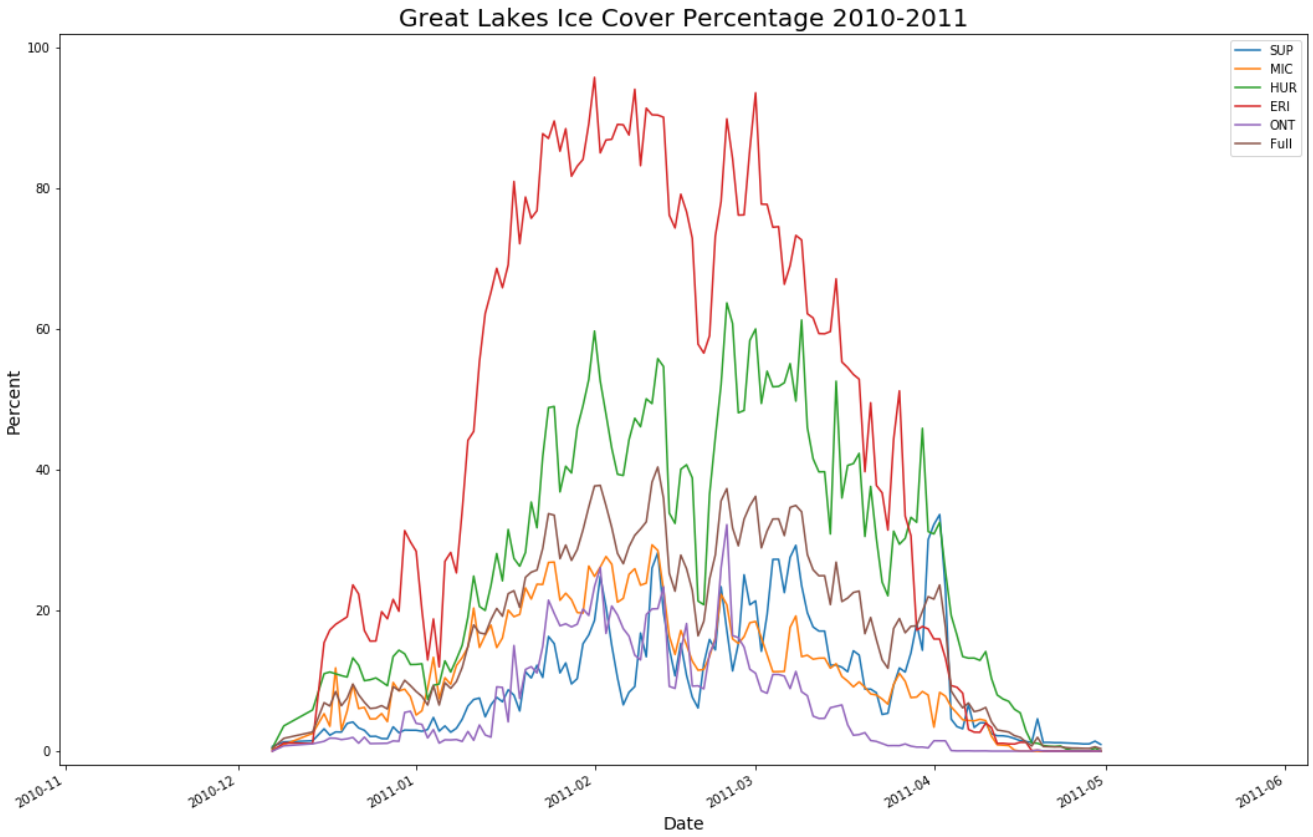

Lake Erie Estimated AMIC and NMDD values, 1898-1983 (± 1 st. dev.)
Estimated AMIC (%)
NMDD values
Year
Est Ice ±1 stdev
Estimated Ice Cover
NMDD Values
Ice Cover Reports, 1963-72
Ice Cover Reports, 1973-83

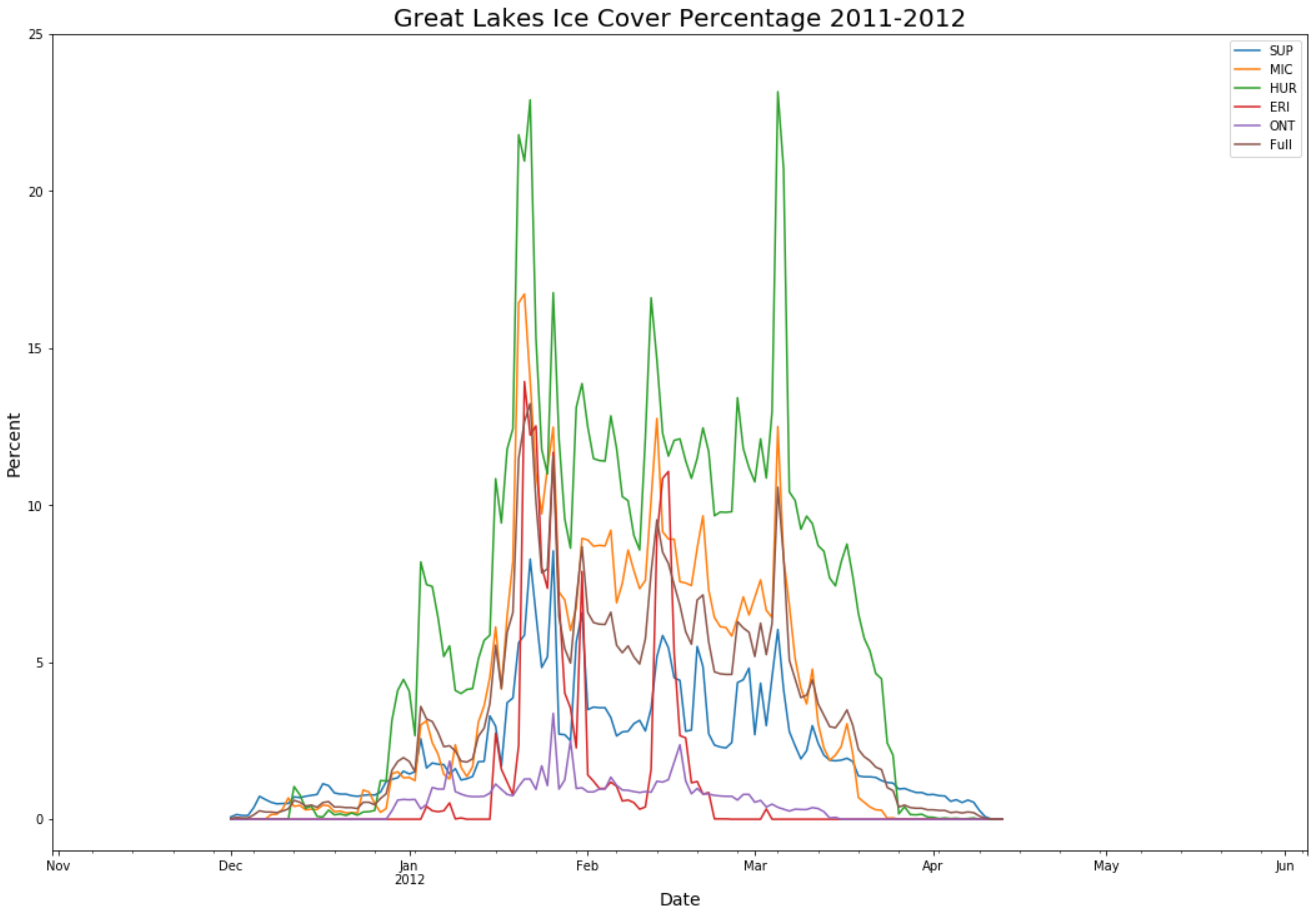

Lake Ontario Estimated AMIC and NMDD values, 1898-1983 (± 1 st. dev.)
Estimated AMIC (%)
NMDD values
Year
Est Ice ±1 stdev
Estimated Ice Cover
NMDD Values
Ice Cover Reports, 1963-72
Ice Cover Reports, 1973-83

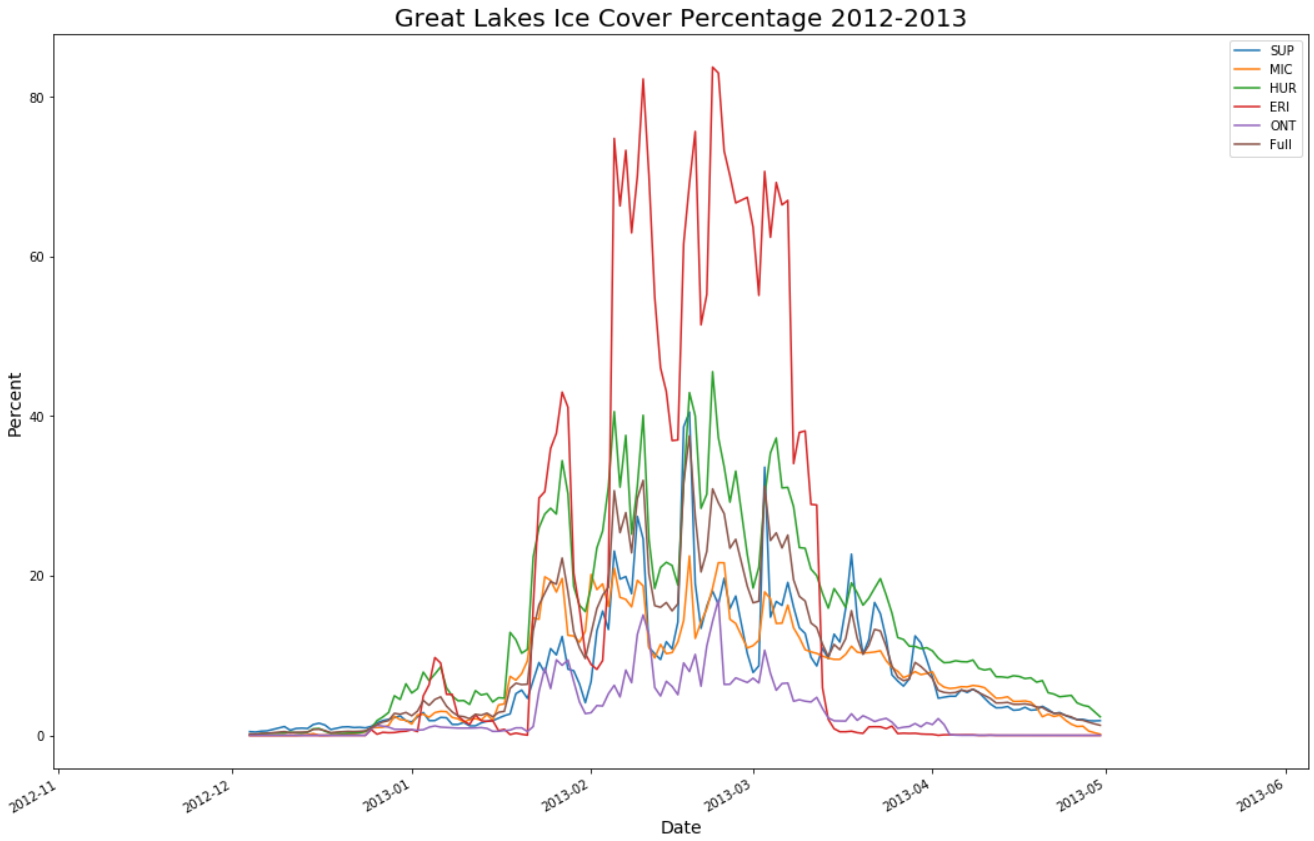
ICE COVER TIME SERIES 2009-2020

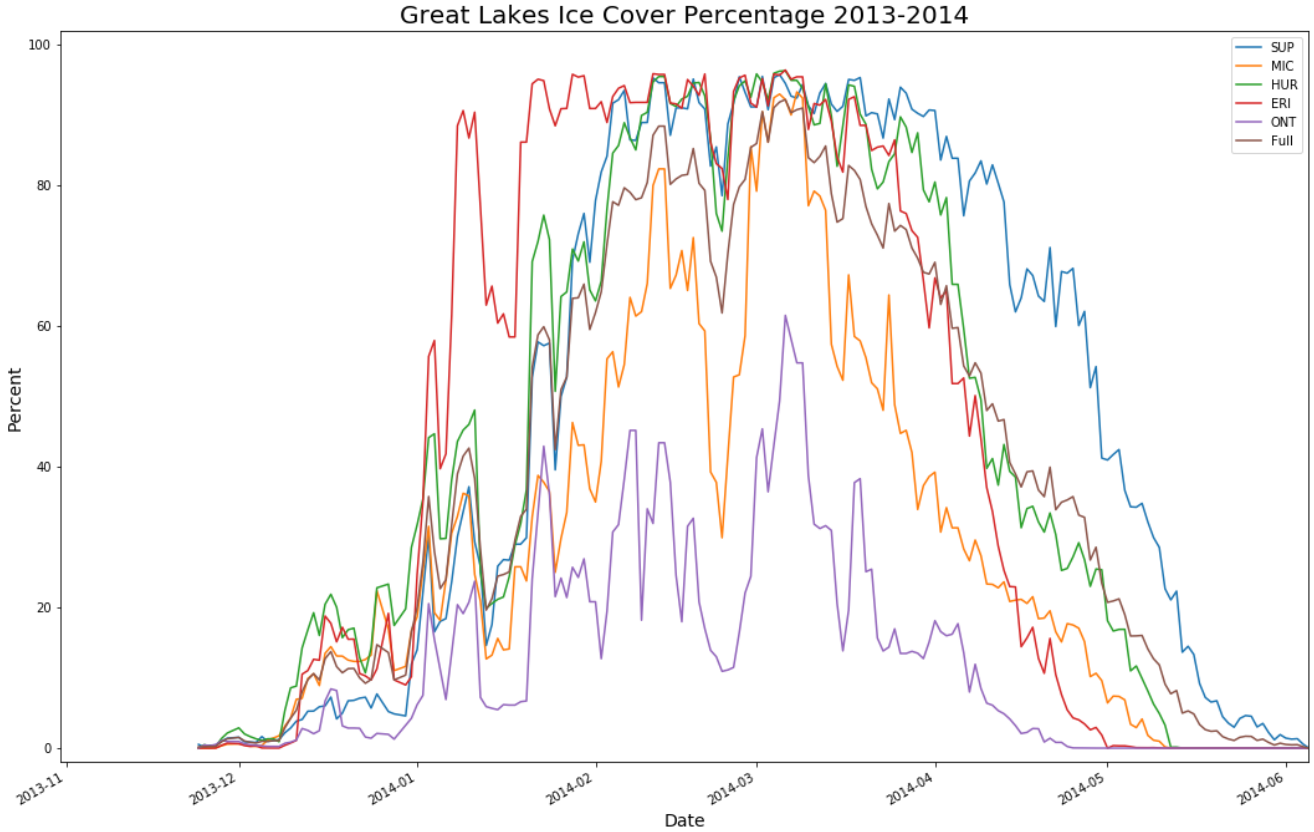


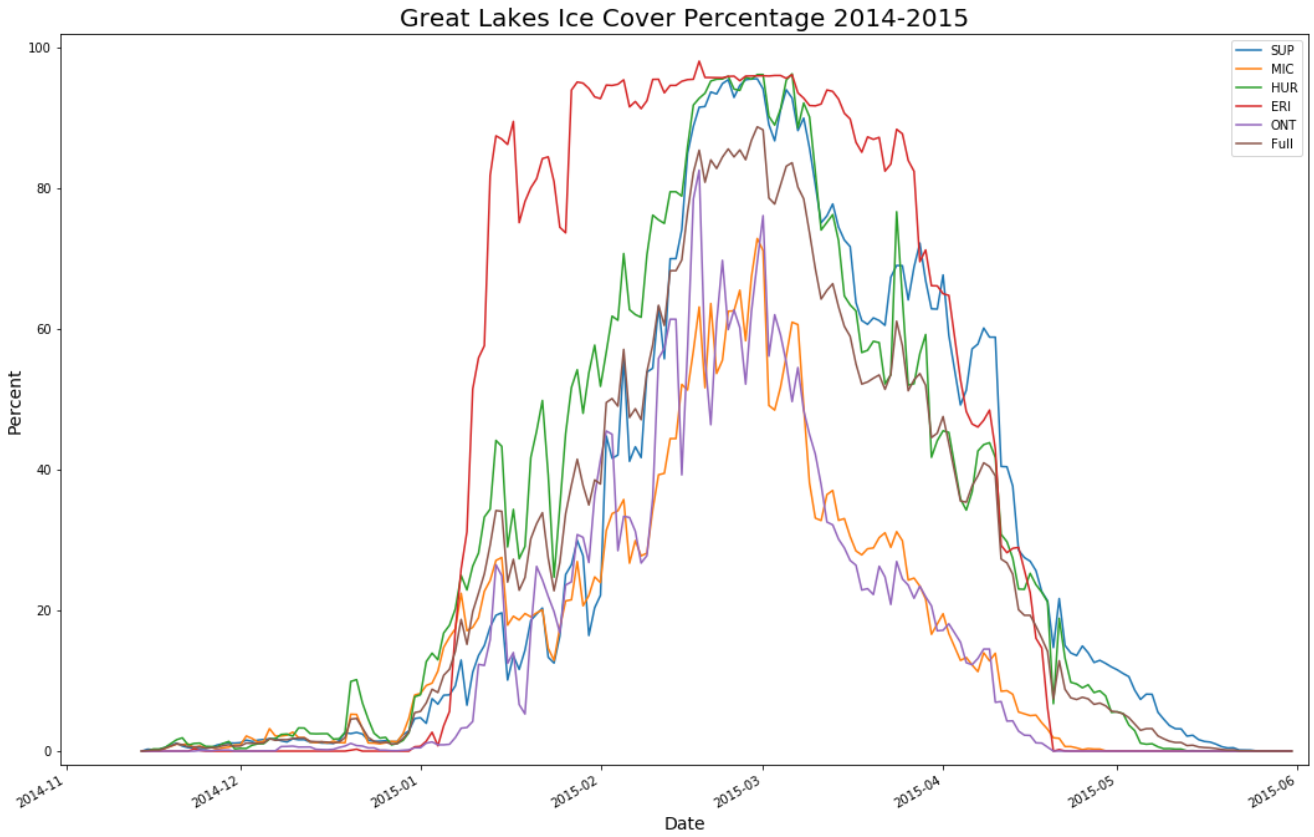


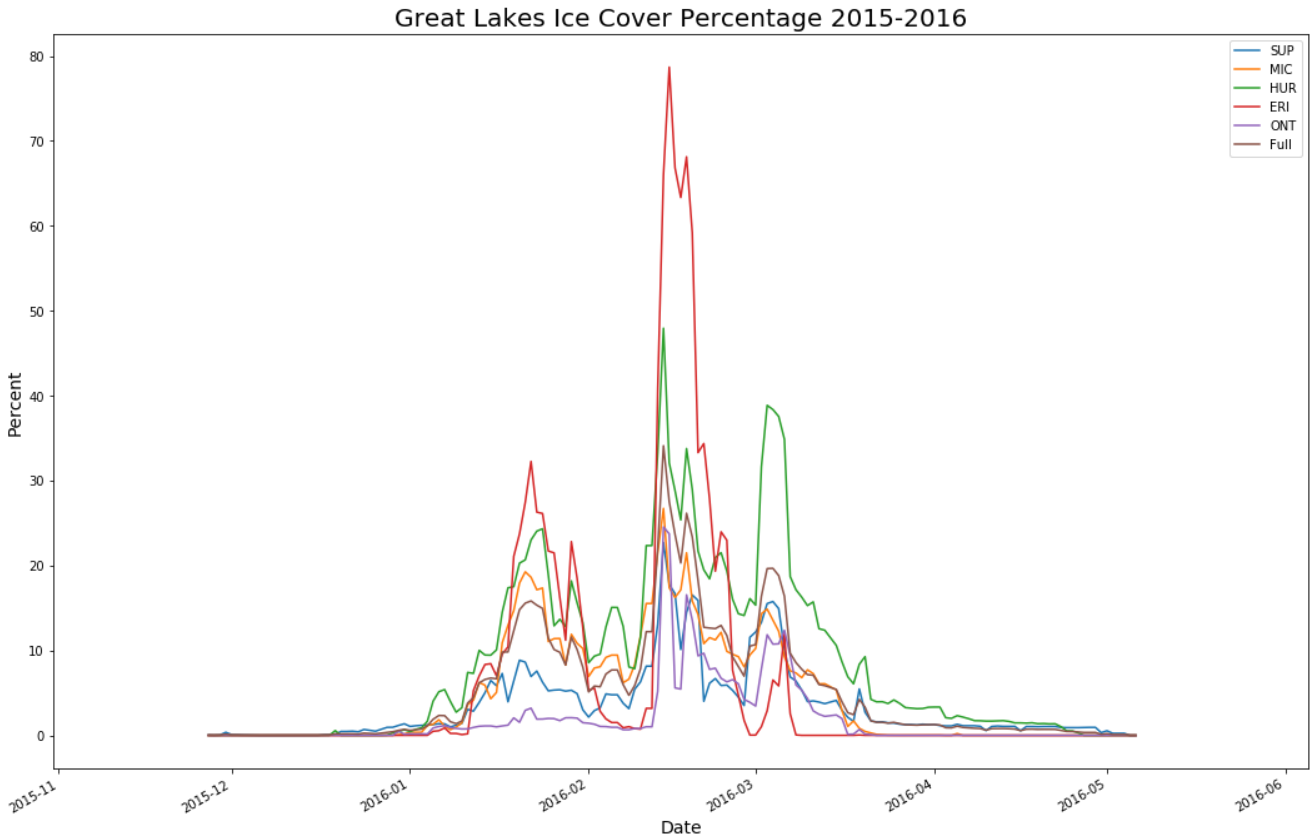

Great Lakes Ice Cover Percentage 2010-2011
SUP
MIC
HUR
ERI
ONT
Full
Percent
Date
2010-11
2010-12
2011-01
2011-02
2011-03
2011-04
2011-05
2011-06

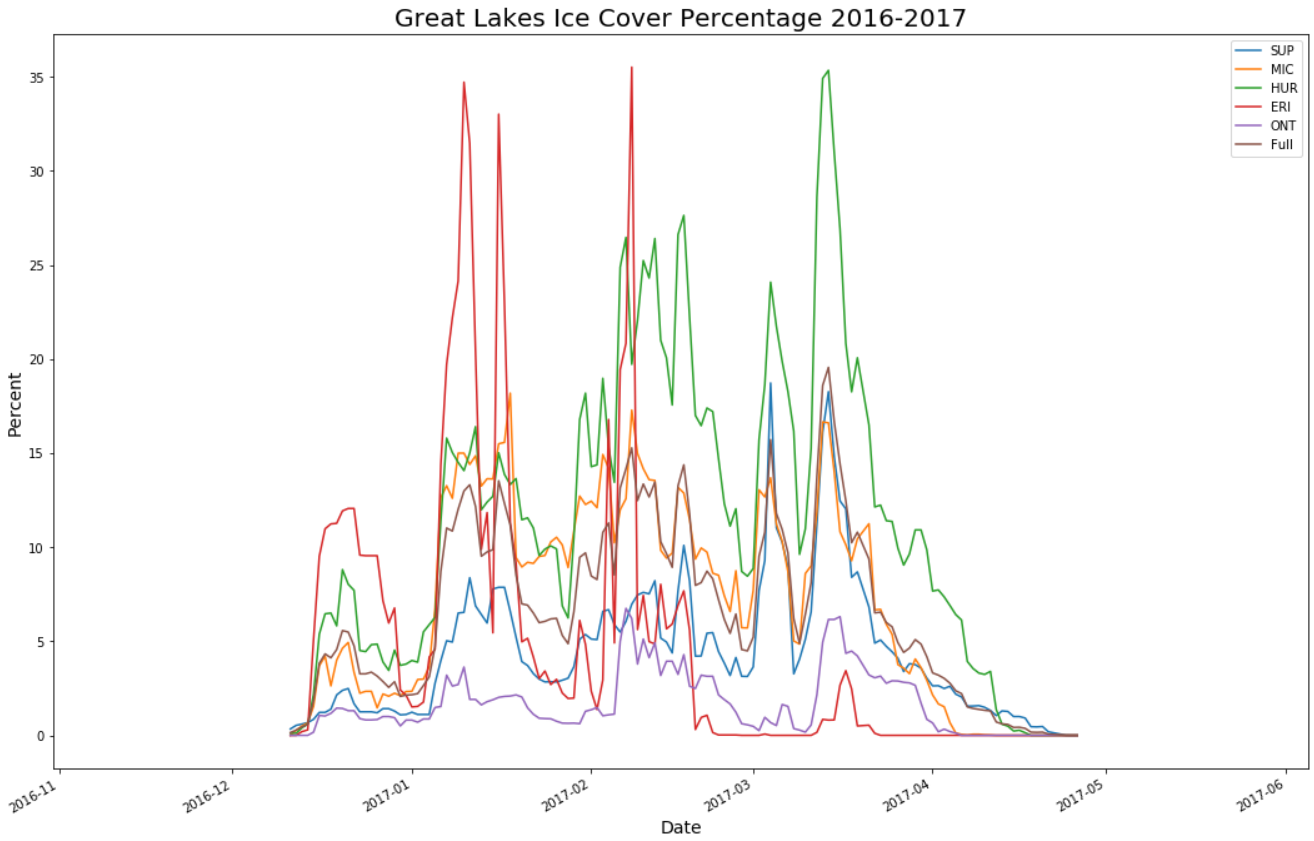

Great Lakes Ice Cover Percentage 2011-2012
SUP
MIC
HUR
ERI
ONT
Full
Percent
Date
Nov
Dec
Jan
2012
Feb
Mar
Apr
May
Jun

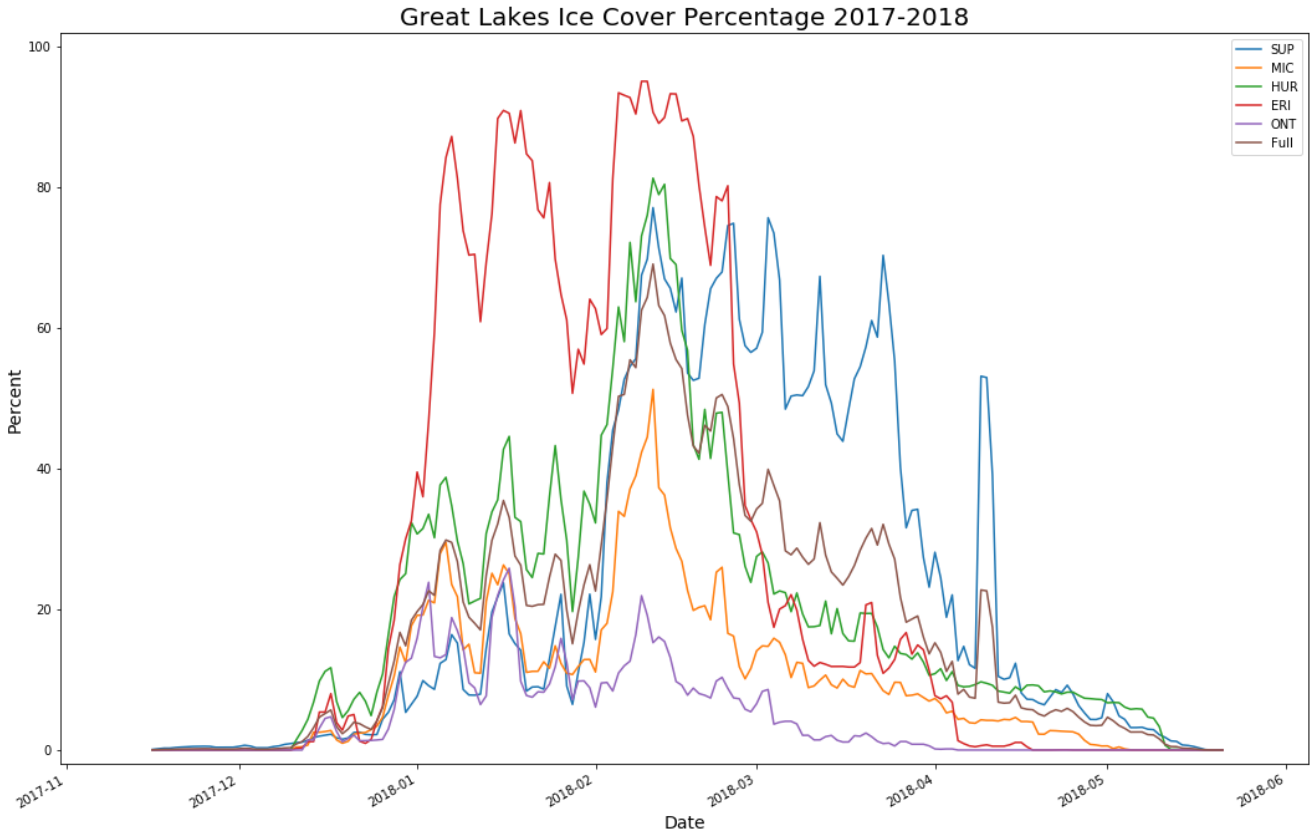

Great Lakes Ice Cover Percentage 2012-2013
SUP
MIC
HUR
ERI
ONT
Full
Percent
0
20
40
60
80
2012-11
2012-12
2013-01
2013-02
2013-03
2013-04
2013-05
2013-06
Date

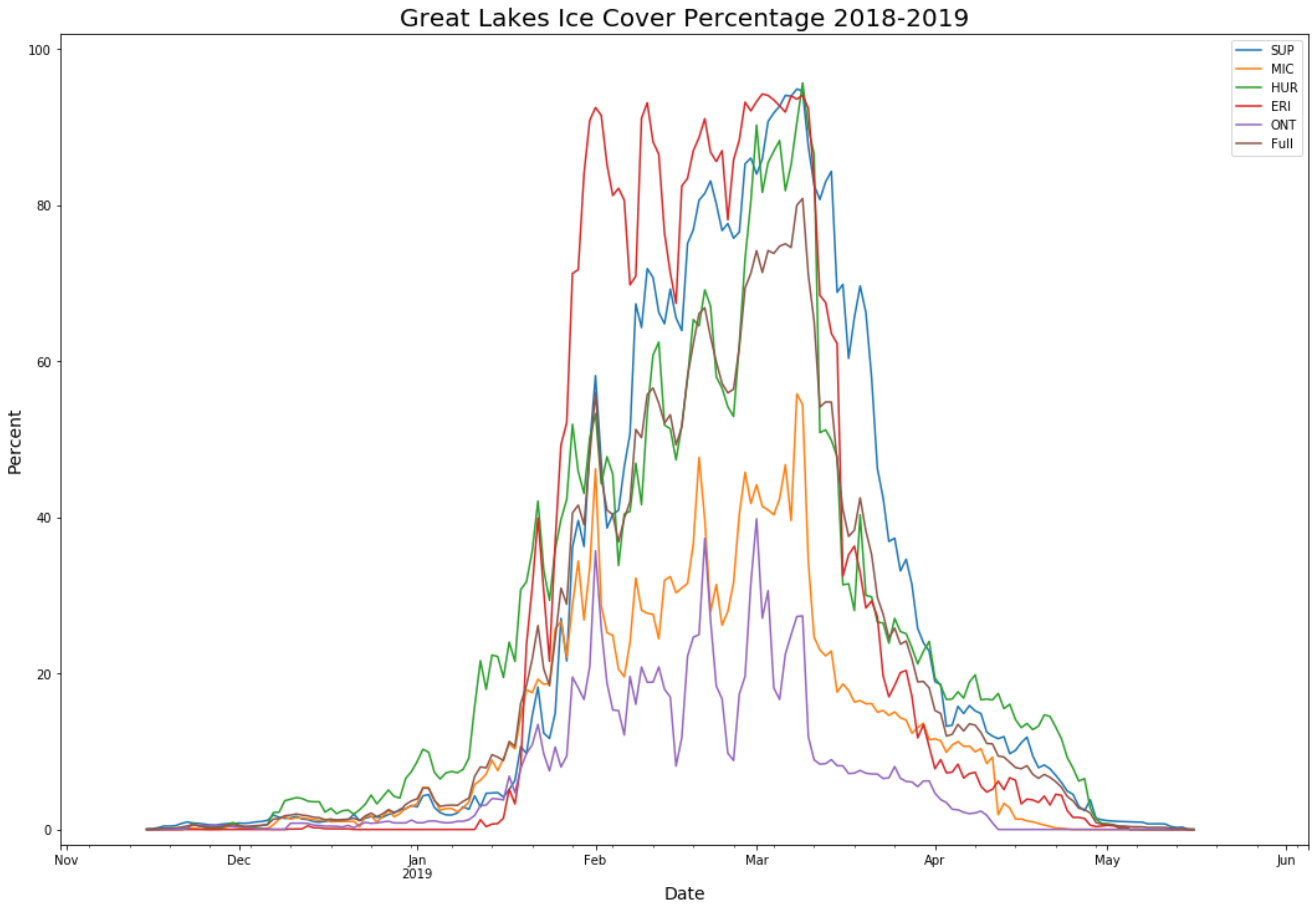

Great Lakes Ice Cover Percentage 2013-2014
SUP
MIC
HUR
ERI
ONT
Full
Percent
0
20
40
60
80
100
2013-11
2013-12
2014-01
2014-02
2014-03
2014-04
2014-05
2014-06
Date

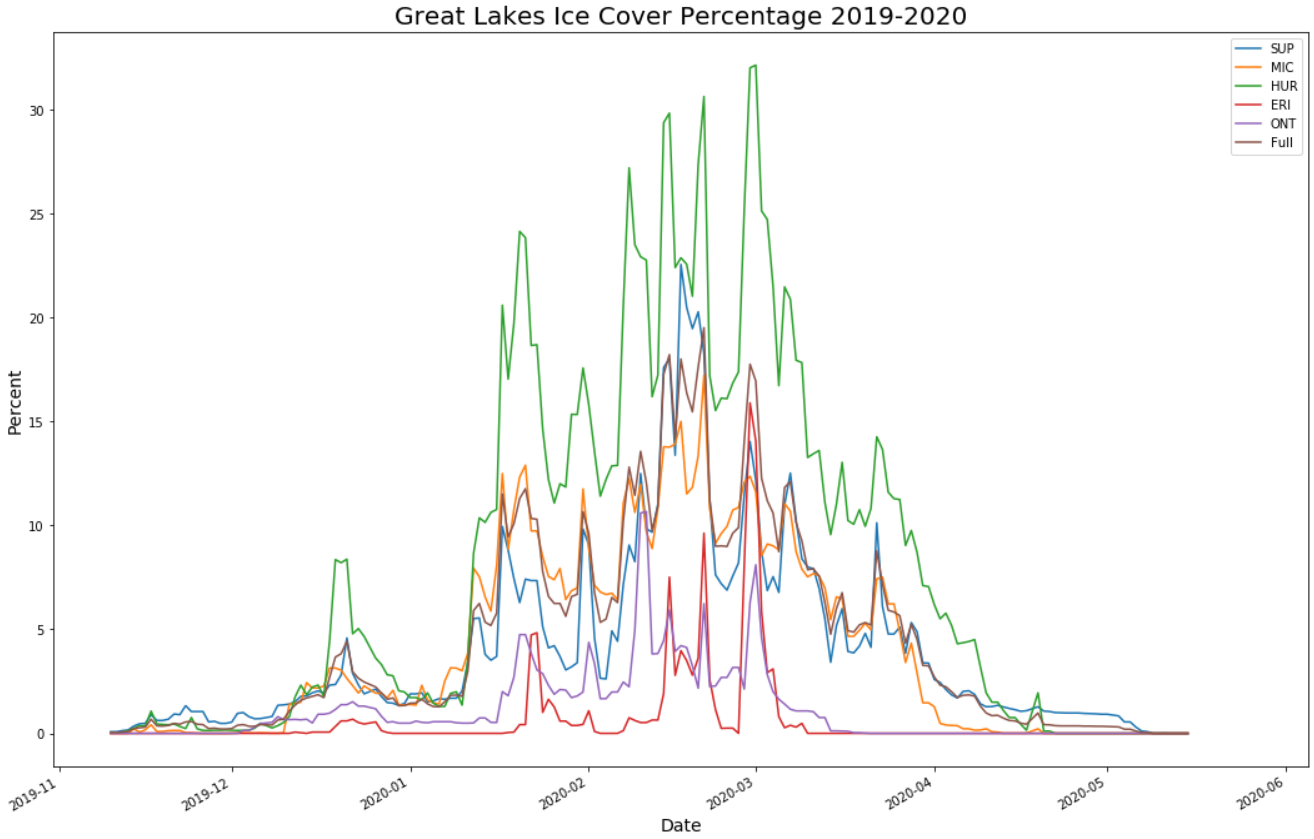

Great Lakes Ice Cover Percentage 2014-2015
Percent
Date
SUP
MIC
HUR
ERI
ONT
Full
0
20
40
60
80
100
2014-11
2014-12
2015-01
2015-02
2015-03
2015-04
2015-05
2015-06

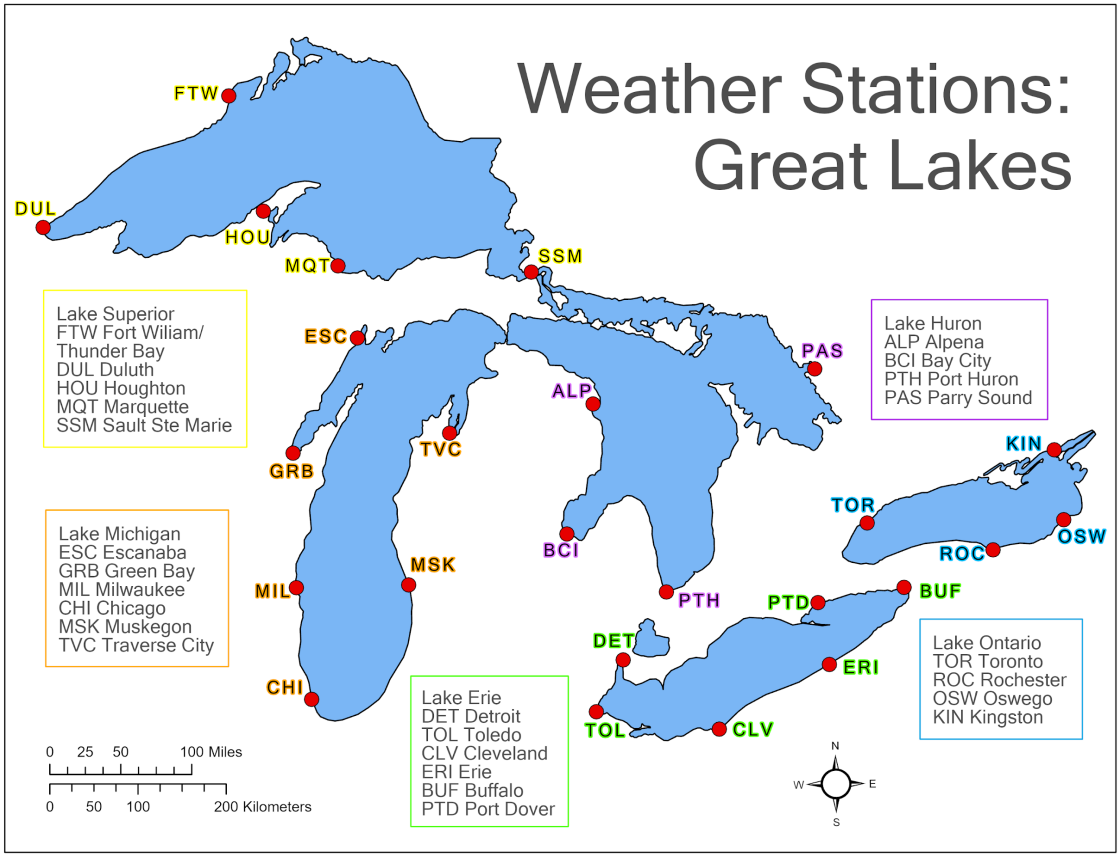

Great Lakes Ice Cover Percentage 2015-2016
Percent
Date
SUP
MIC
HUR
ERI
ONT
Full
0
10
20
30
40
50
60
70
80
2015-11
2015-12
2016-01
2016-02
2016-03
2016-04
2016-05
2016-06


Great Lakes Ice Cover Percentage 2016-2017
SUP
MIC
HUR
ERI
ONT
Full
Percent
Date
2016-11
2016-12
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06

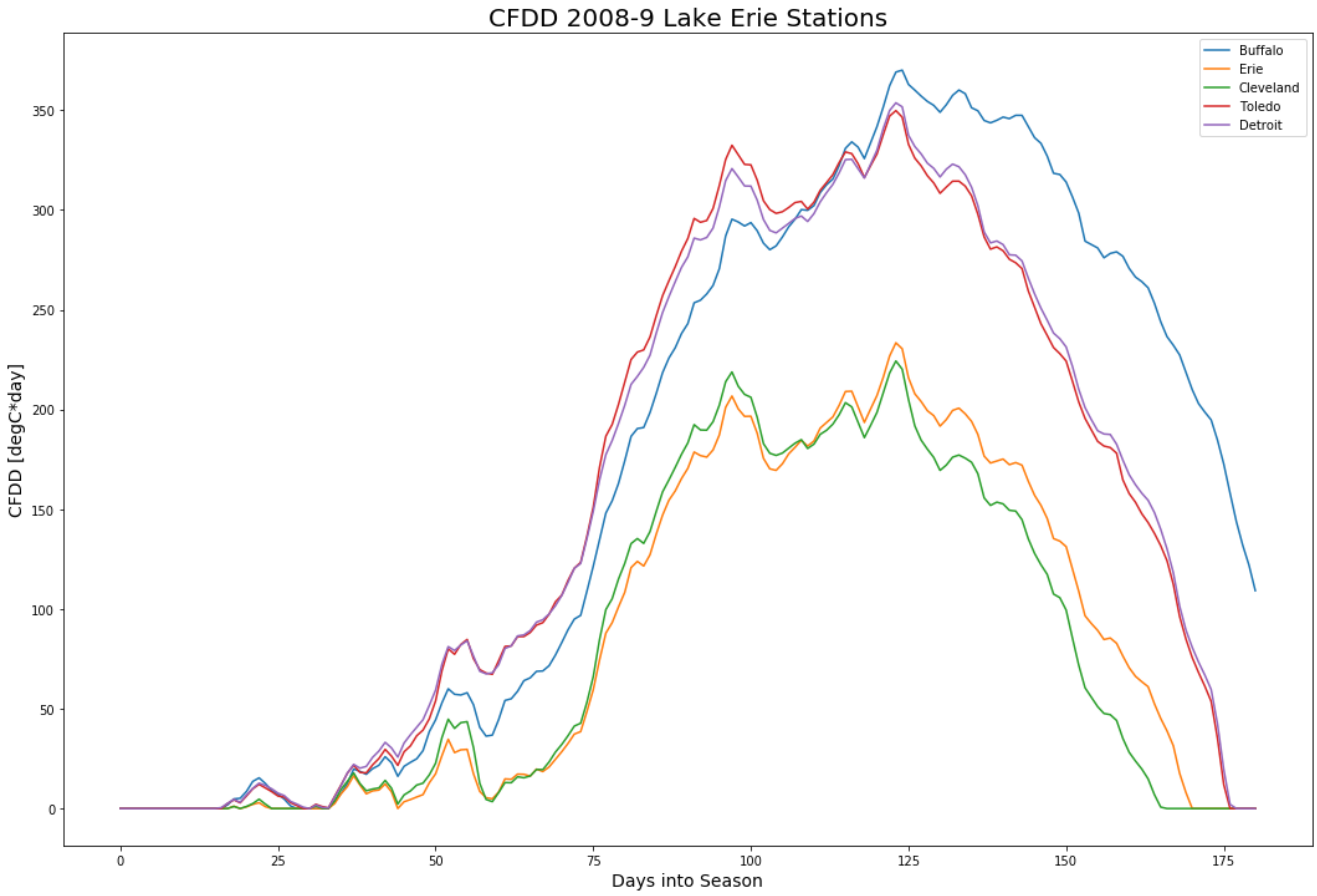

Great Lakes Ice Cover Percentage 2017-2018
SUP
MIC
HUR
ERI
ONT
Full
Percent
Date
2017-11
2017-12
2018-01
2018-02
2018-03
2018-04
2018-05
2018-06

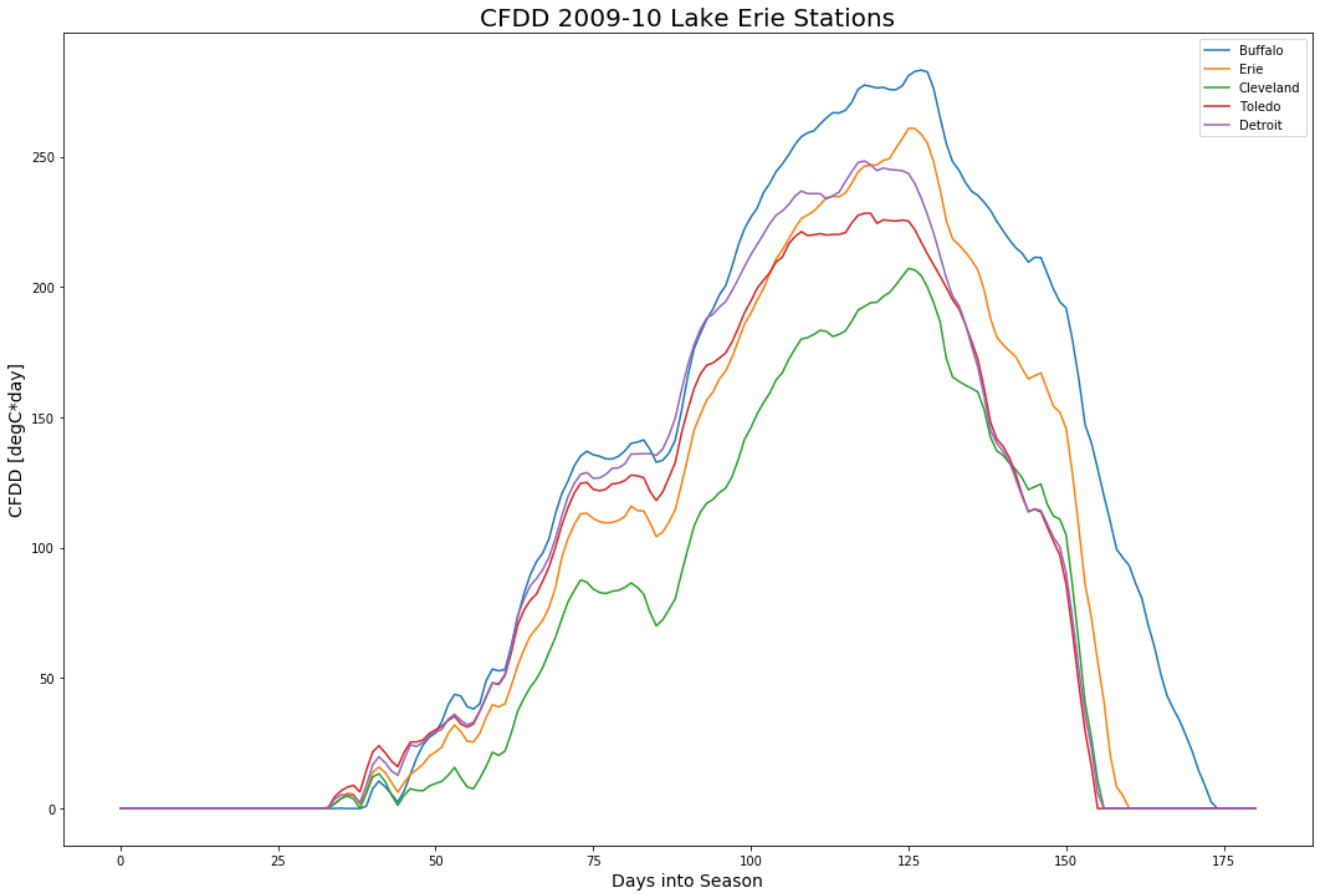

Great Lakes Ice Cover Percentage 2018-2019
SUP
MIC
HUR
ERI
ONT
Full
Percent
Date
Nov
Dec
Jan
2019
Feb
Mar
Apr
May
Jun

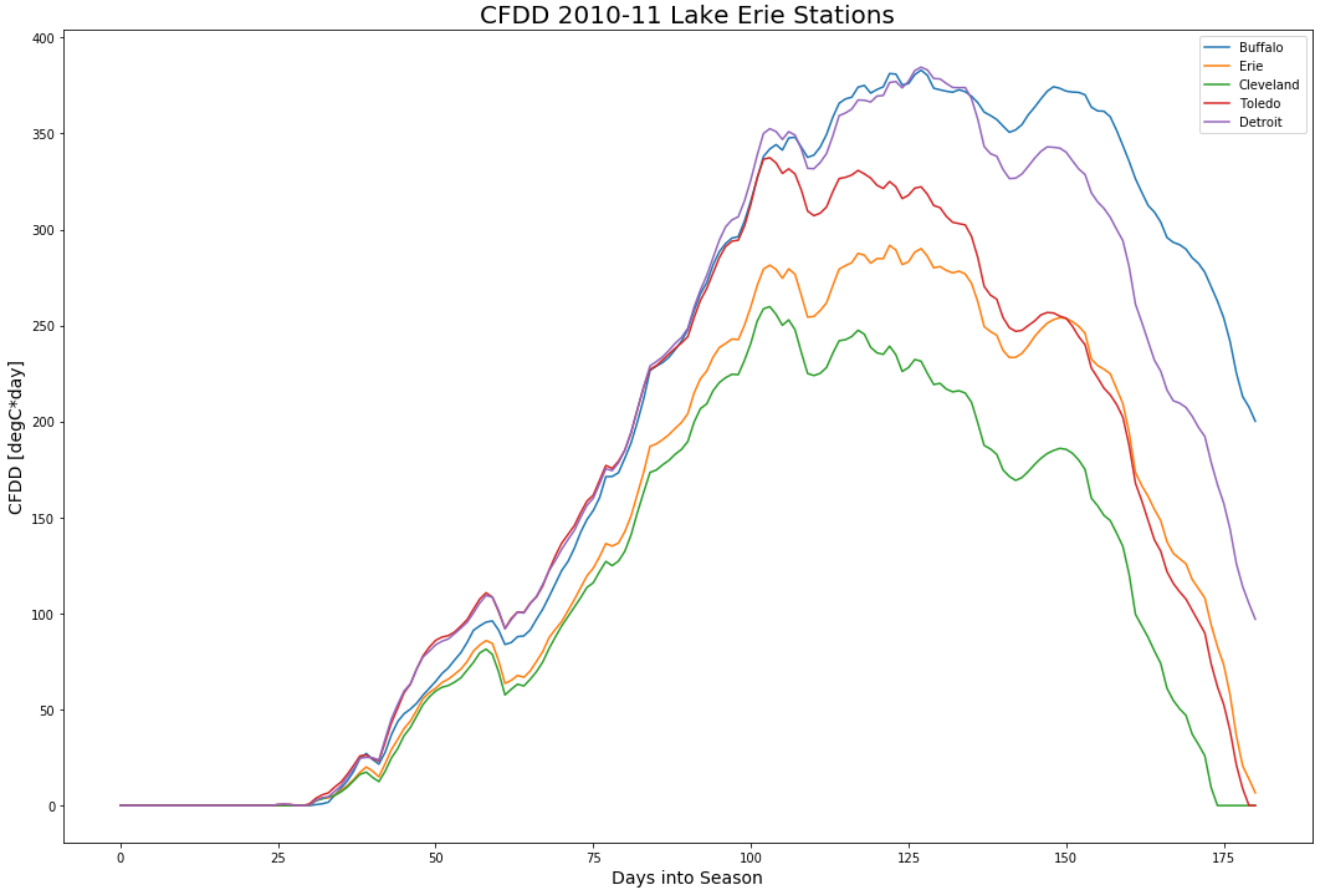

Great Lakes Ice Cover Percentage 2019-2020
SUP
MIC
HUR
ERI
ONT
Full
Percent
Date
2019-11
2019-12
2020-01
2020-02
2020-03
2020-04
2020-05
2020-06

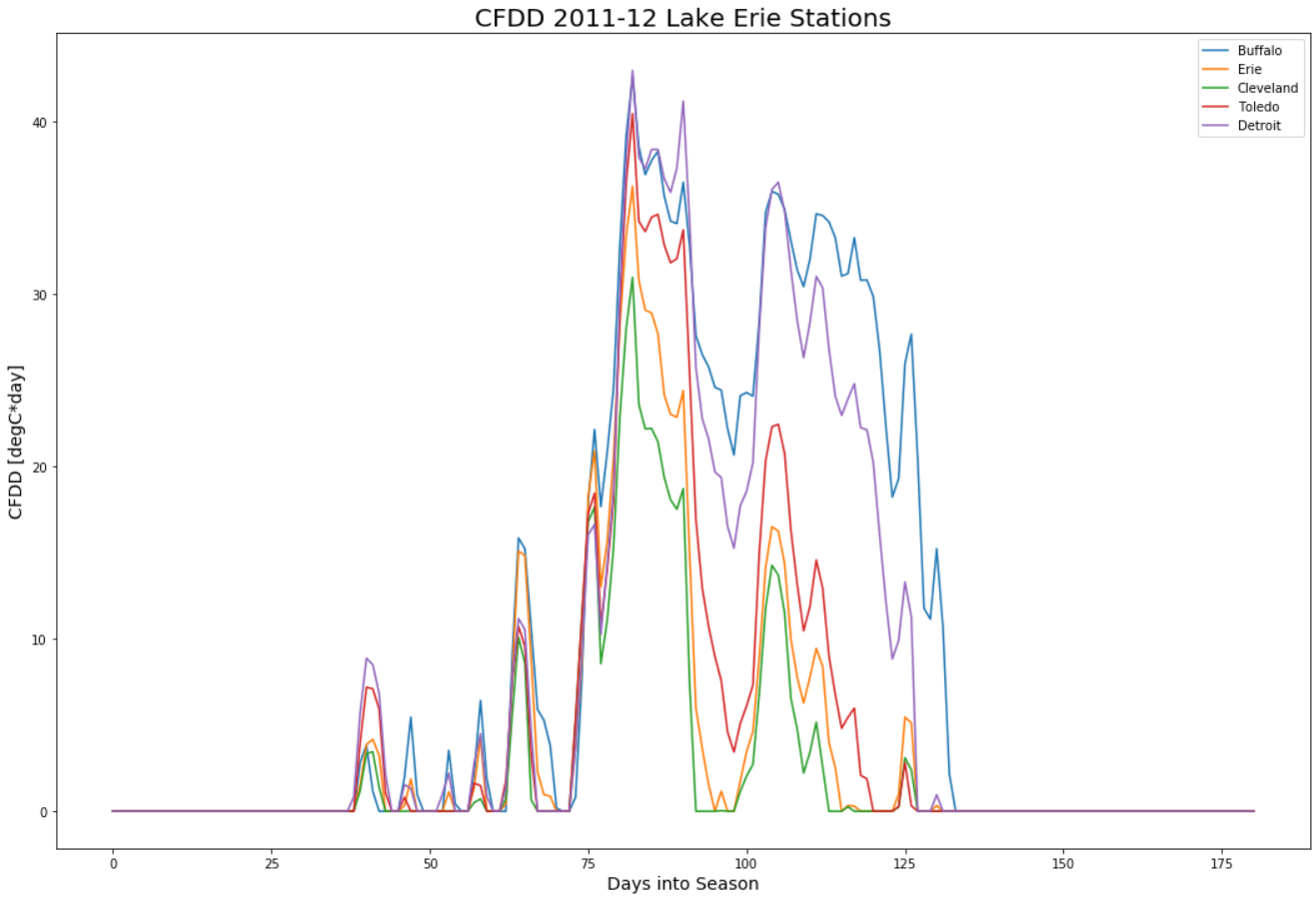

Weather Stations:
Great Lakes
FTW
DUL
HOU
MQT
SSM
Lake Superior
FTW Fort Wiliam/
Thunder Bay
DUL Duluth
HOU Houghton
MQT Marquette
SSM Sault Ste Marie
ESC
PAS
ALP
TVC
GRB
Lake Huron
ALP Alpena
BCI Bay City
PTH Port Huron
PAS Parry Sound
KIN
TOR
OSW
ROC
BCI
MSK
MIL
PTH
PTD
BUF
Lake Michigan
ESC Escanaba
GRB Green Bay
MIL Milwaukee
CHI Chicago
MSK Muskegon
TVC Traverse City
DET
ERI
Lake Ontario
TOR Toronto
ROC Rochester
OSW Oswego
KIN Kingston
CHI
TOL
CLV
Lake Erie
DET Detroit
TOL Toledo
CLV Cleveland
ERI Erie
BUF Buffalo
PTD Port Dover
0 25 50 100 Miles
0 50 100 200 Kilometers
N
W
E
S

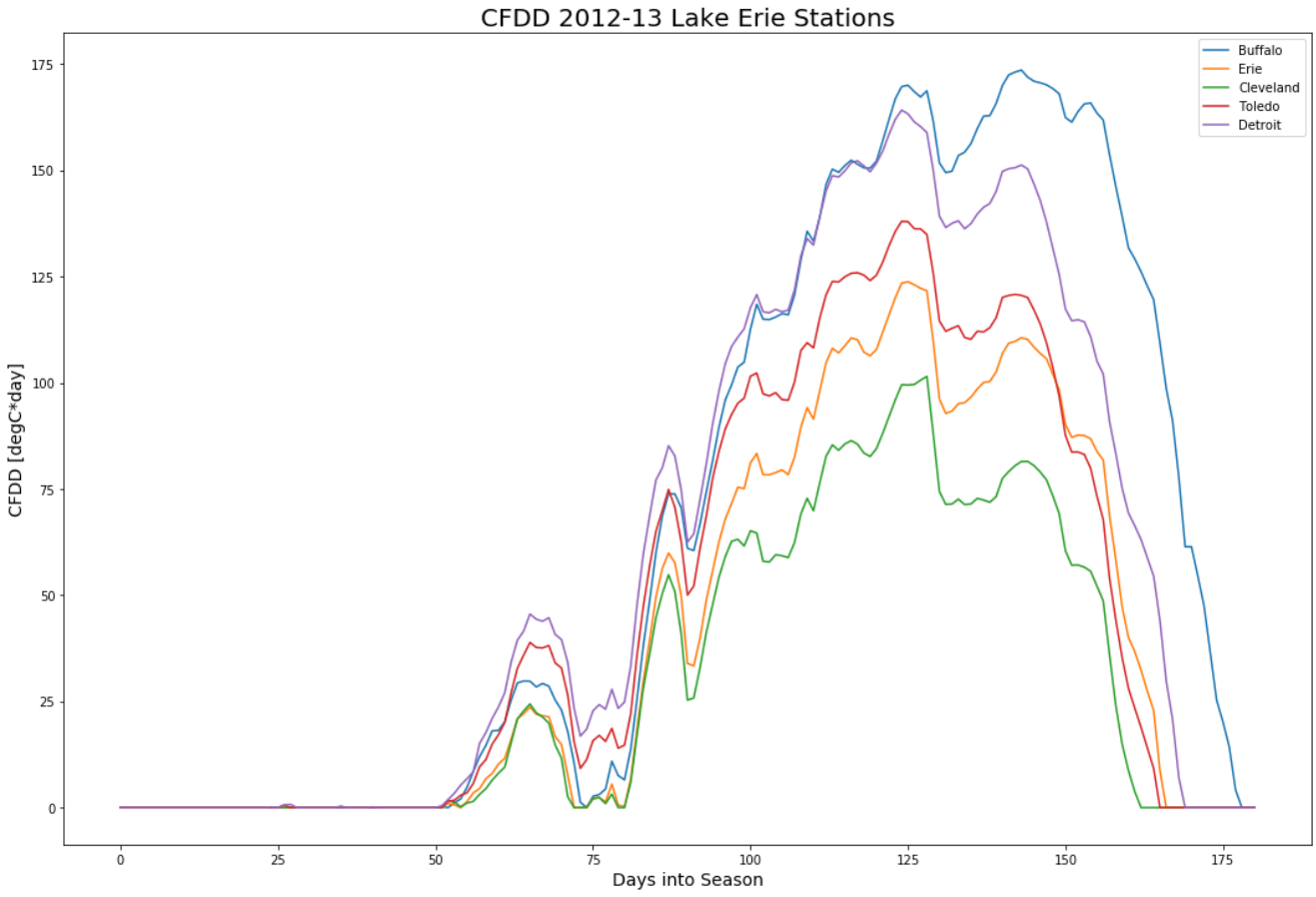
a. **Lake Erie**

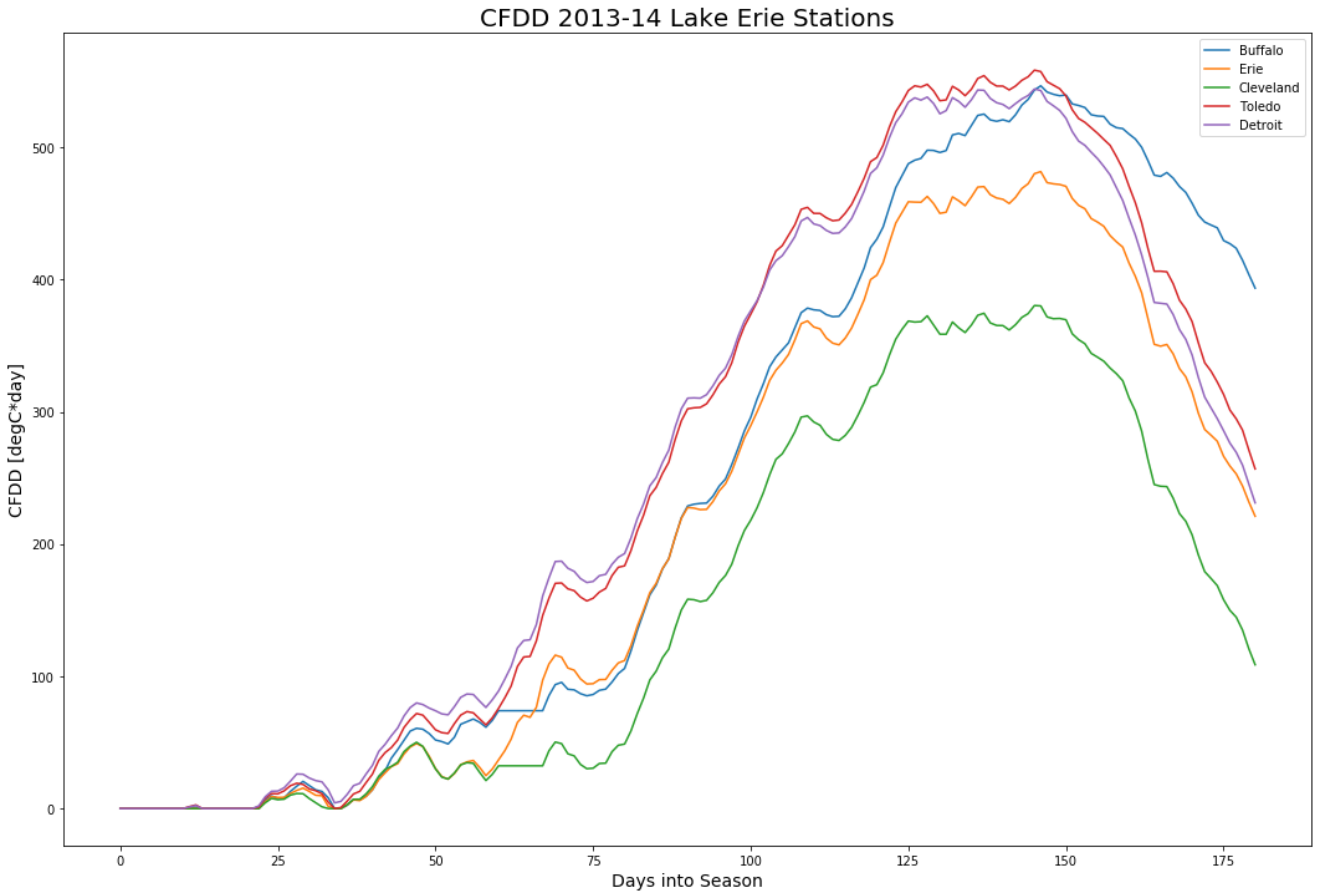

CFDD 2008-9 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
300
350
0
25
50
75
100
125
150
175

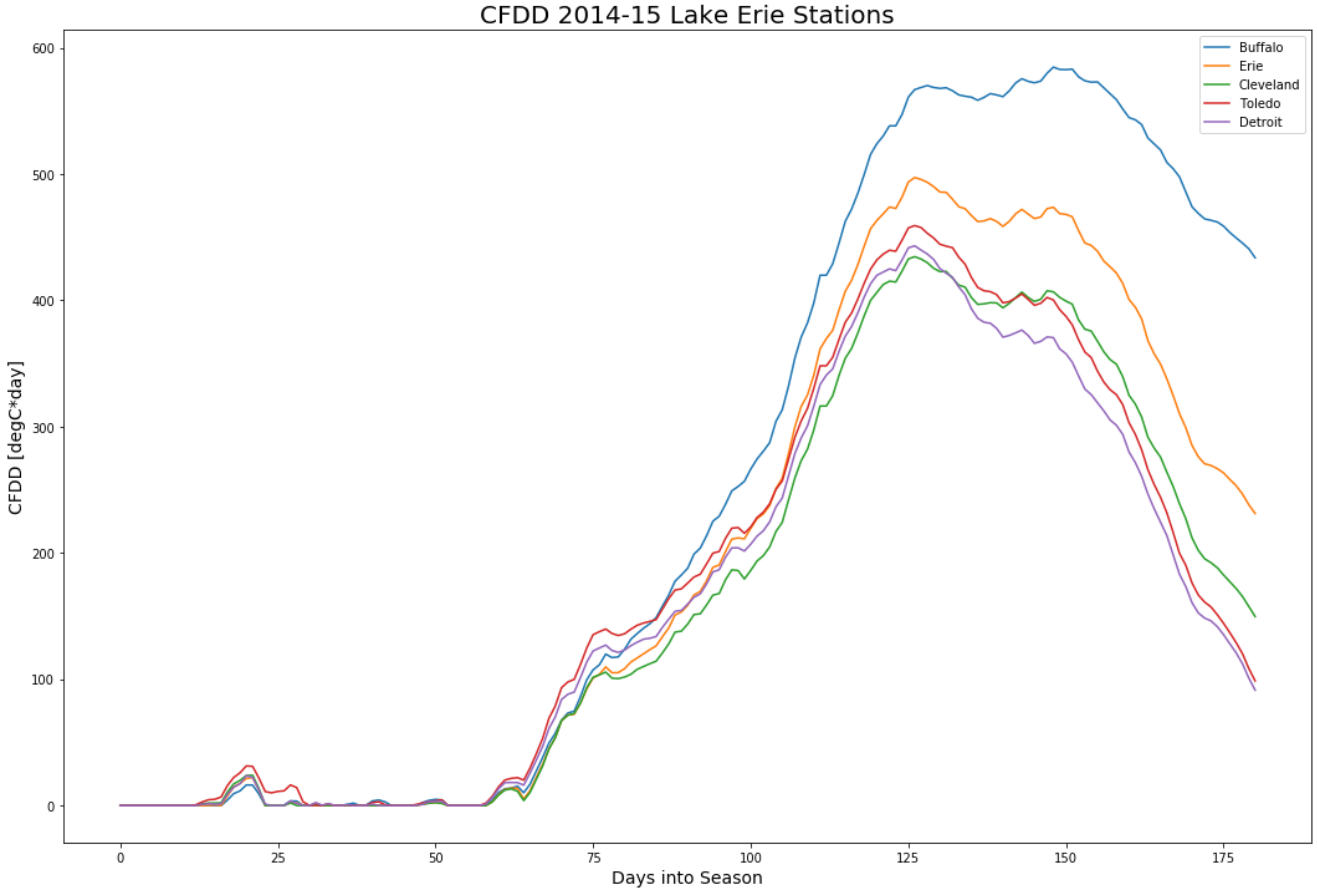

CFDD 2009-10 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
0
25
50
75
100
125
150
175

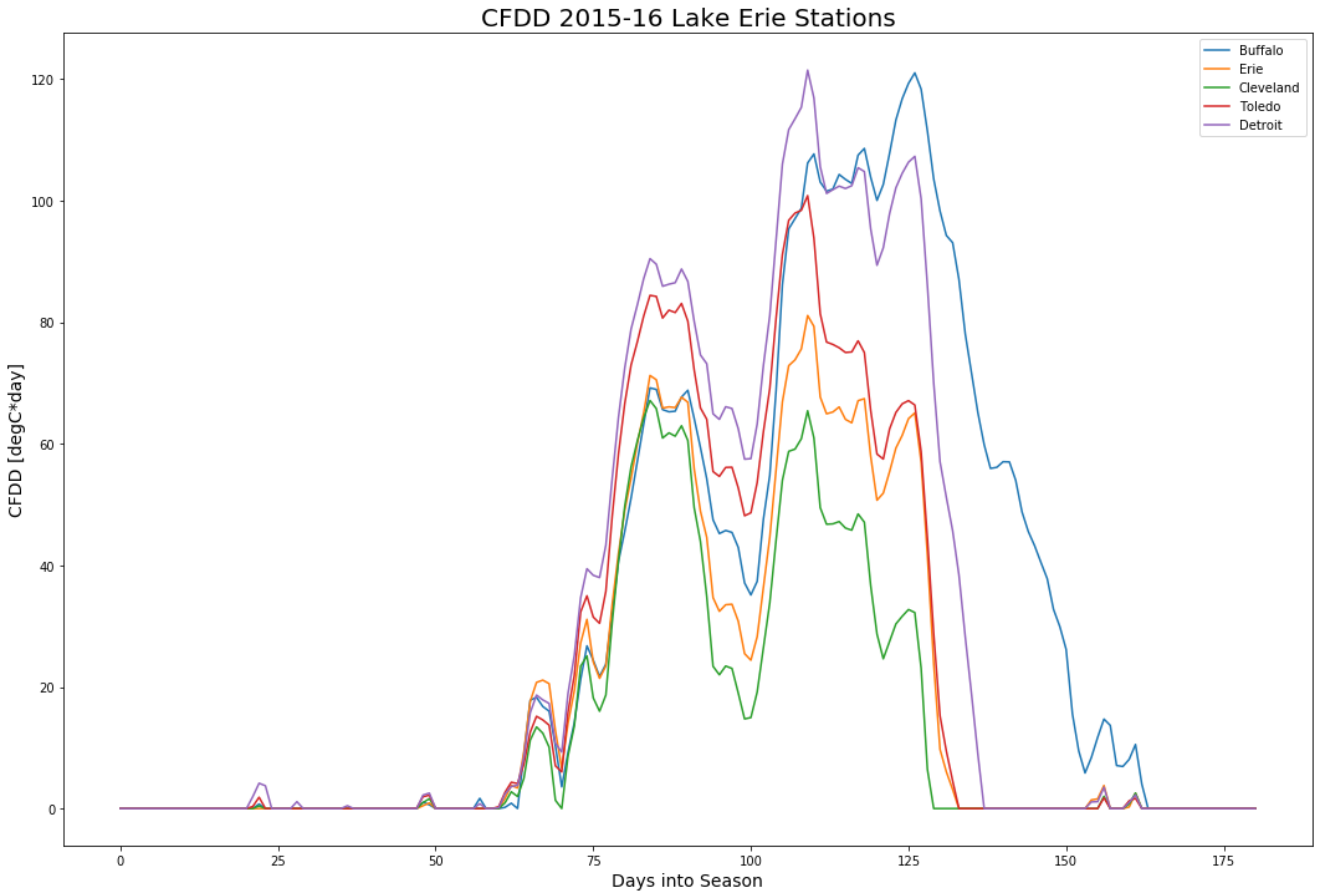

CFDD 2010-11 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
300
350
400
0
25
50
75
100
125
150
175

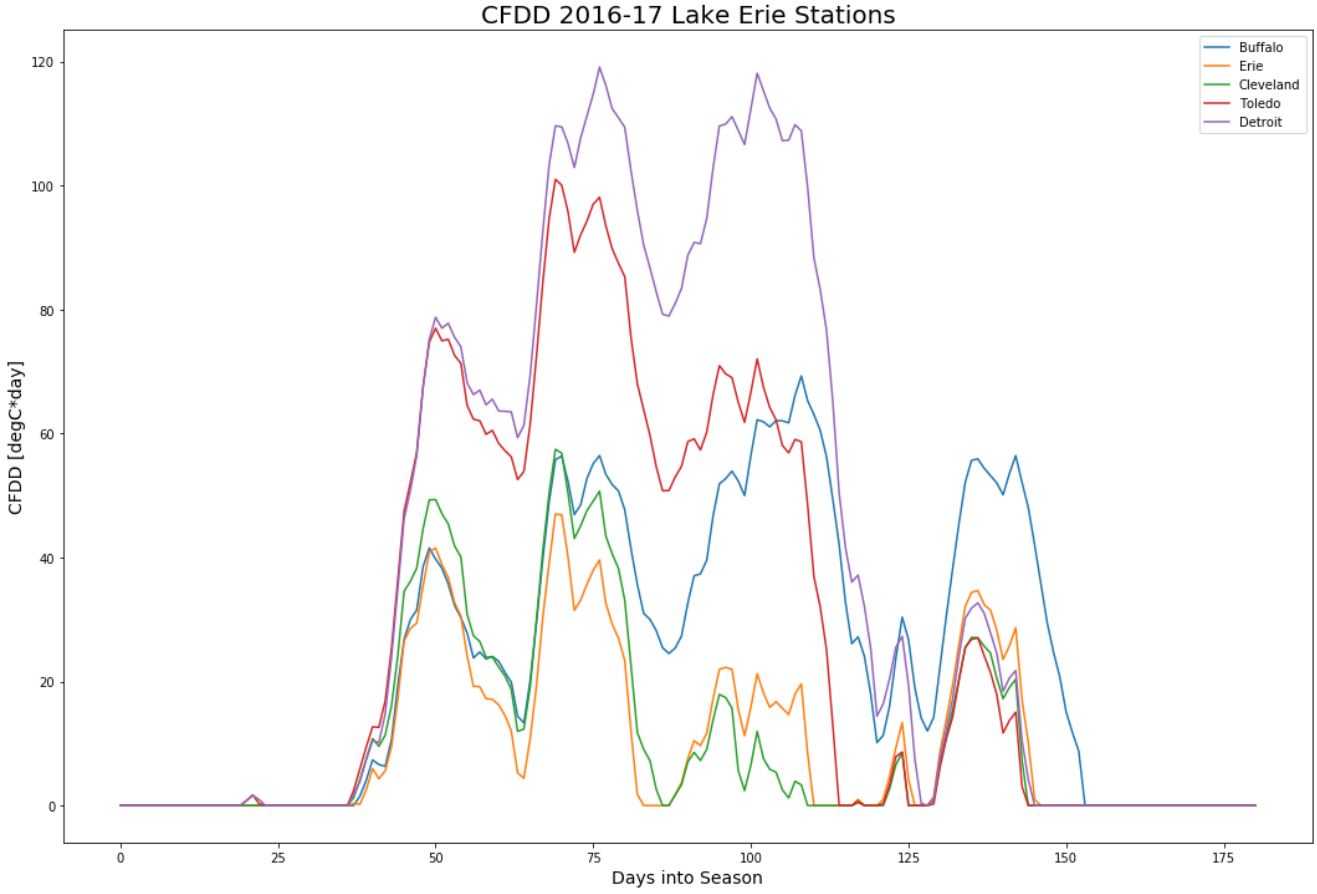

CFDD 2011-12 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
10
20
30
40
0
25
50
75
100
125
150
175

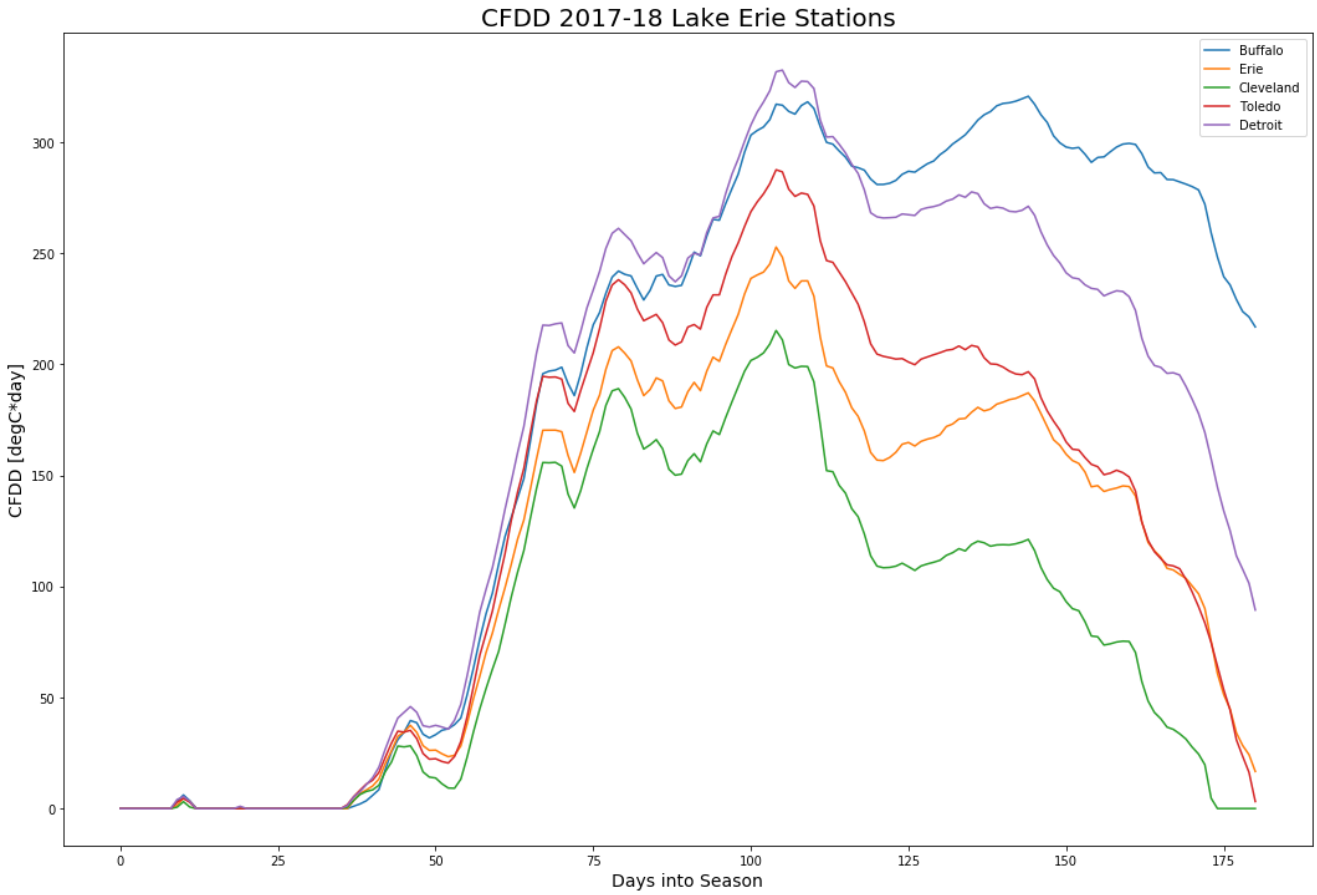

CFDD 2012-13 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
25
50
75
100
125
150
175
0
25
50
75
100
125
150
175

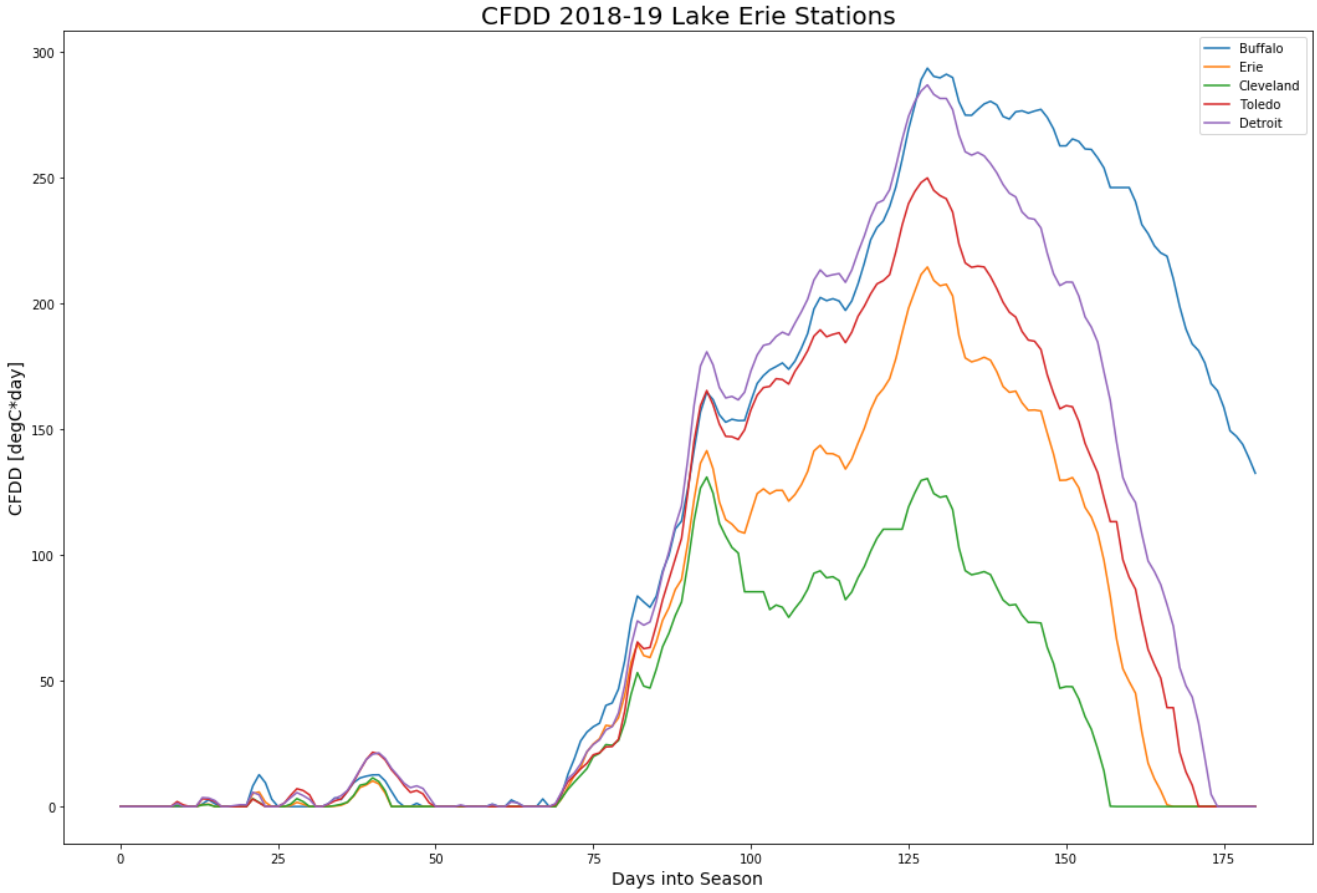

CFDD 2013-14 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
0
25
50
75
100
125
150
175

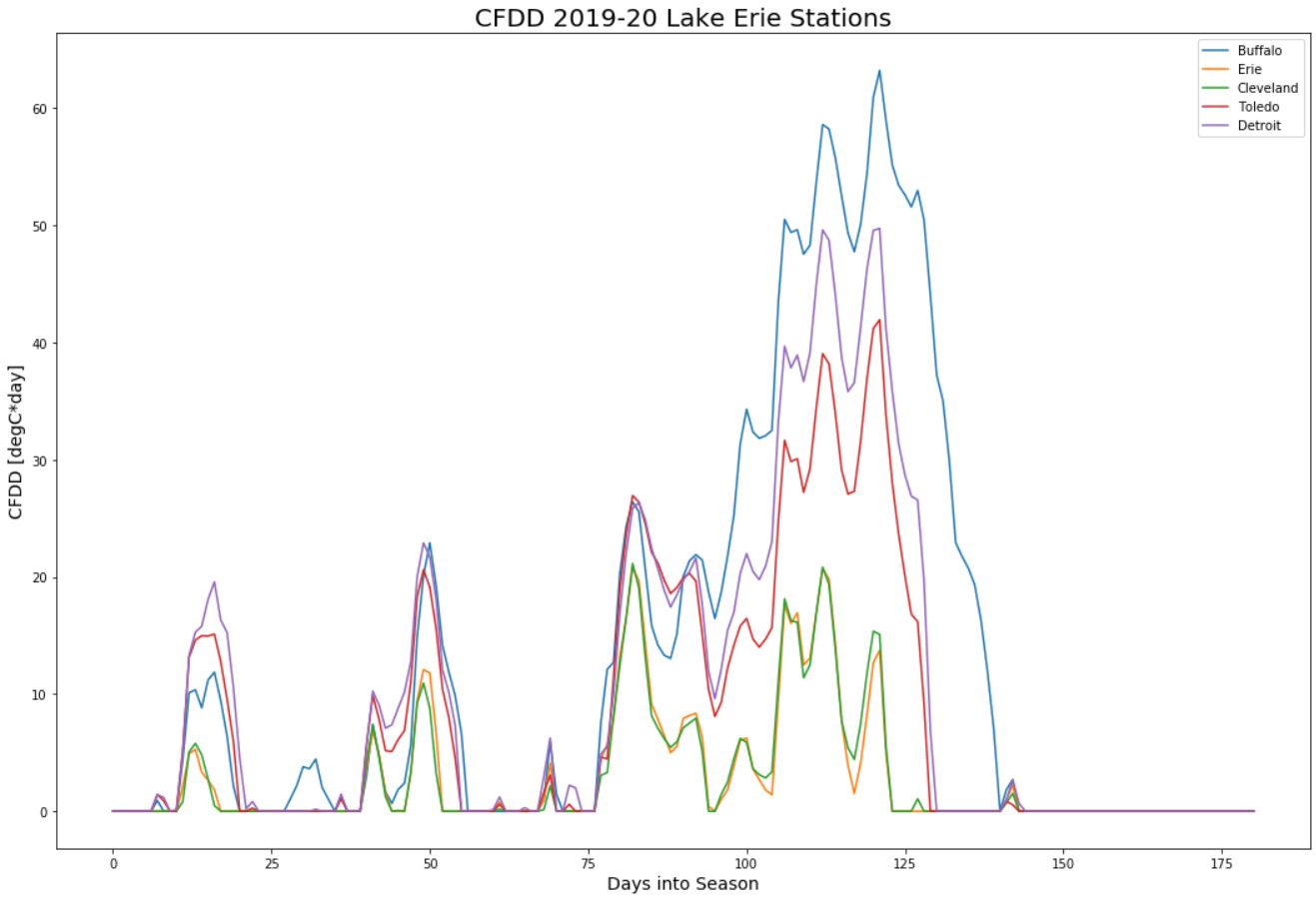

CFDD 2014-15 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
0
25
50
75
100
125
150
175


CFDD 2015-16 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
20
40
60
80
100
120
0
25
50
75
100
125
150
175

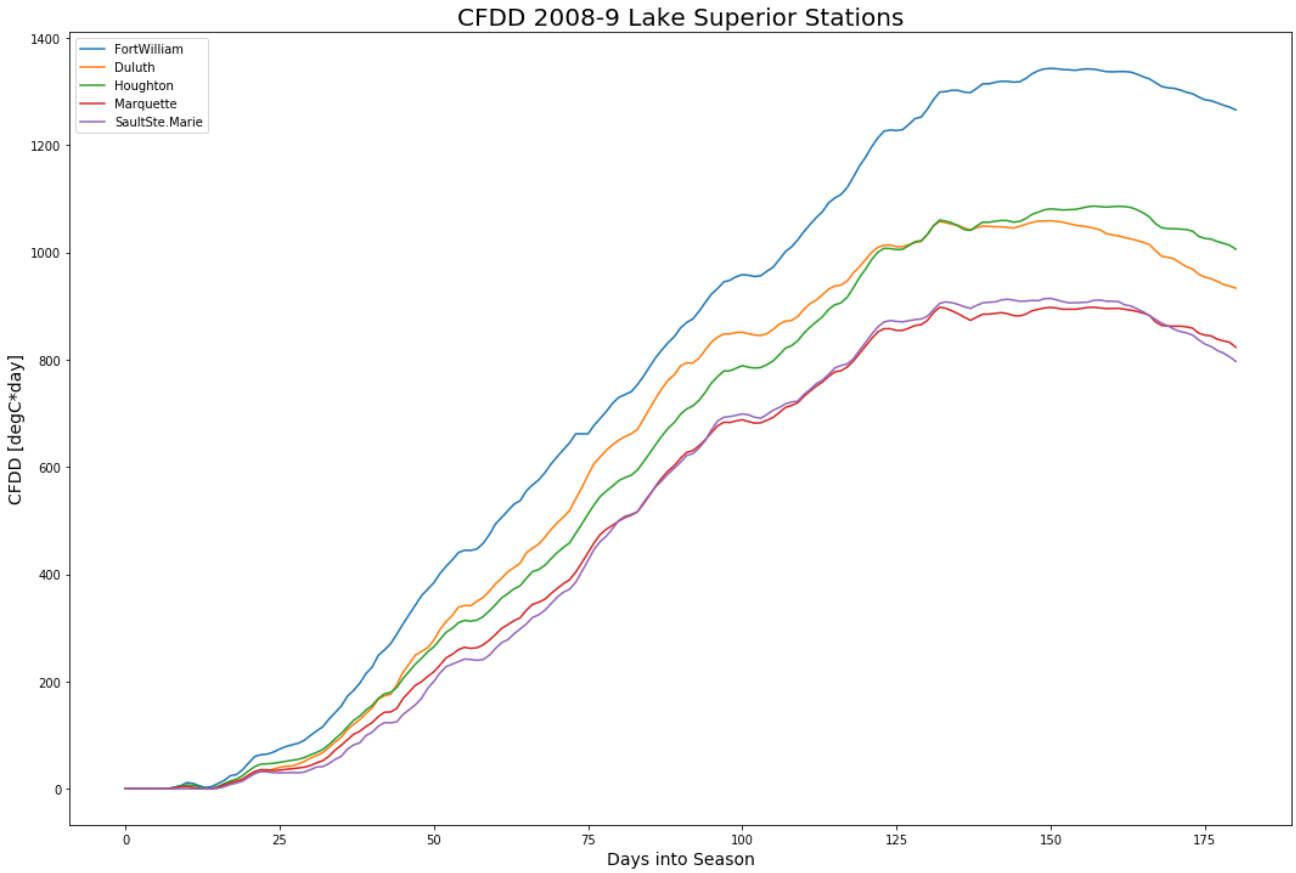

CFDD 2016-17 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
20
40
60
80
100
120
0
25
50
75
100
125
150
175

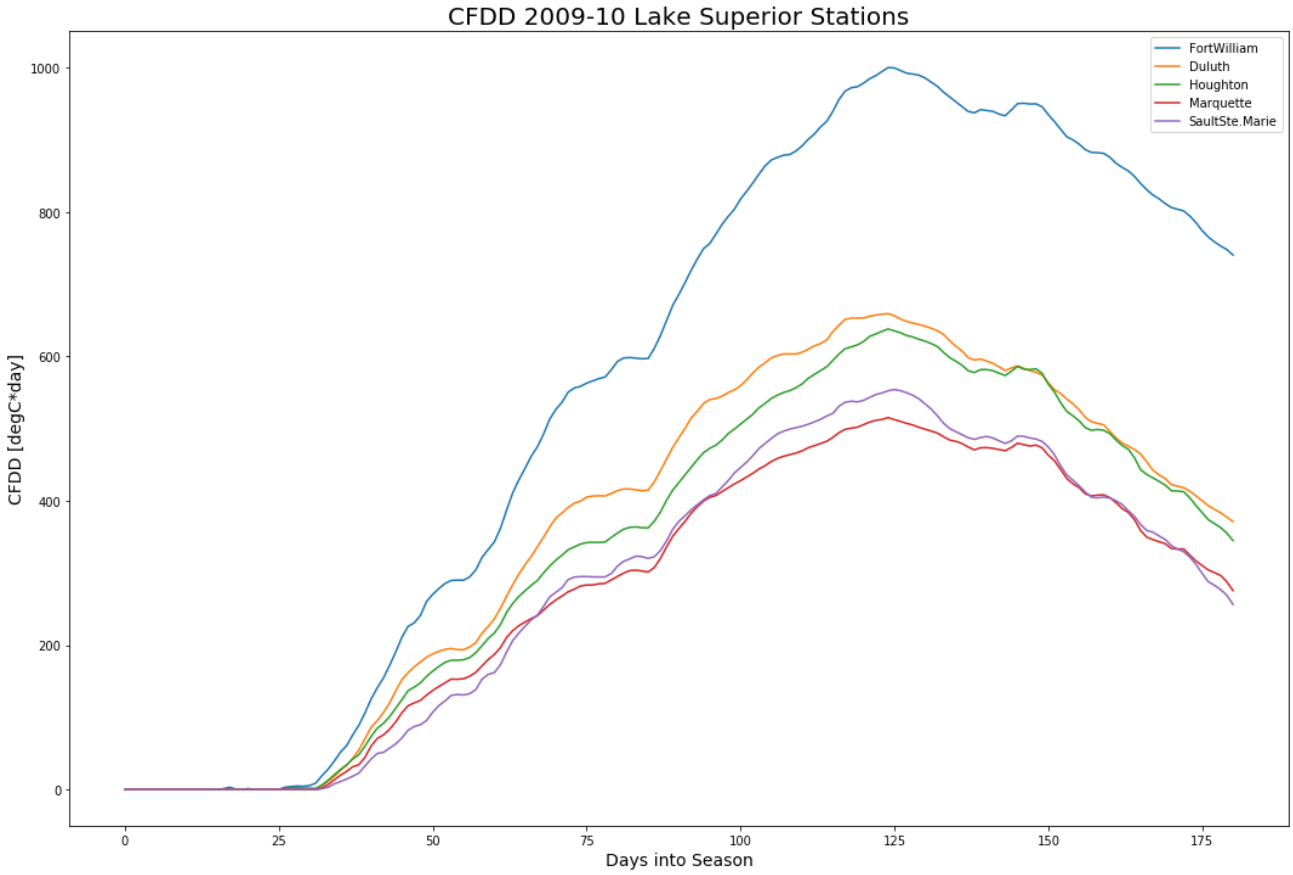

CFDD 2017-18 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
25
50
75
100
125
150
175
0
50
100
150
200
250
300

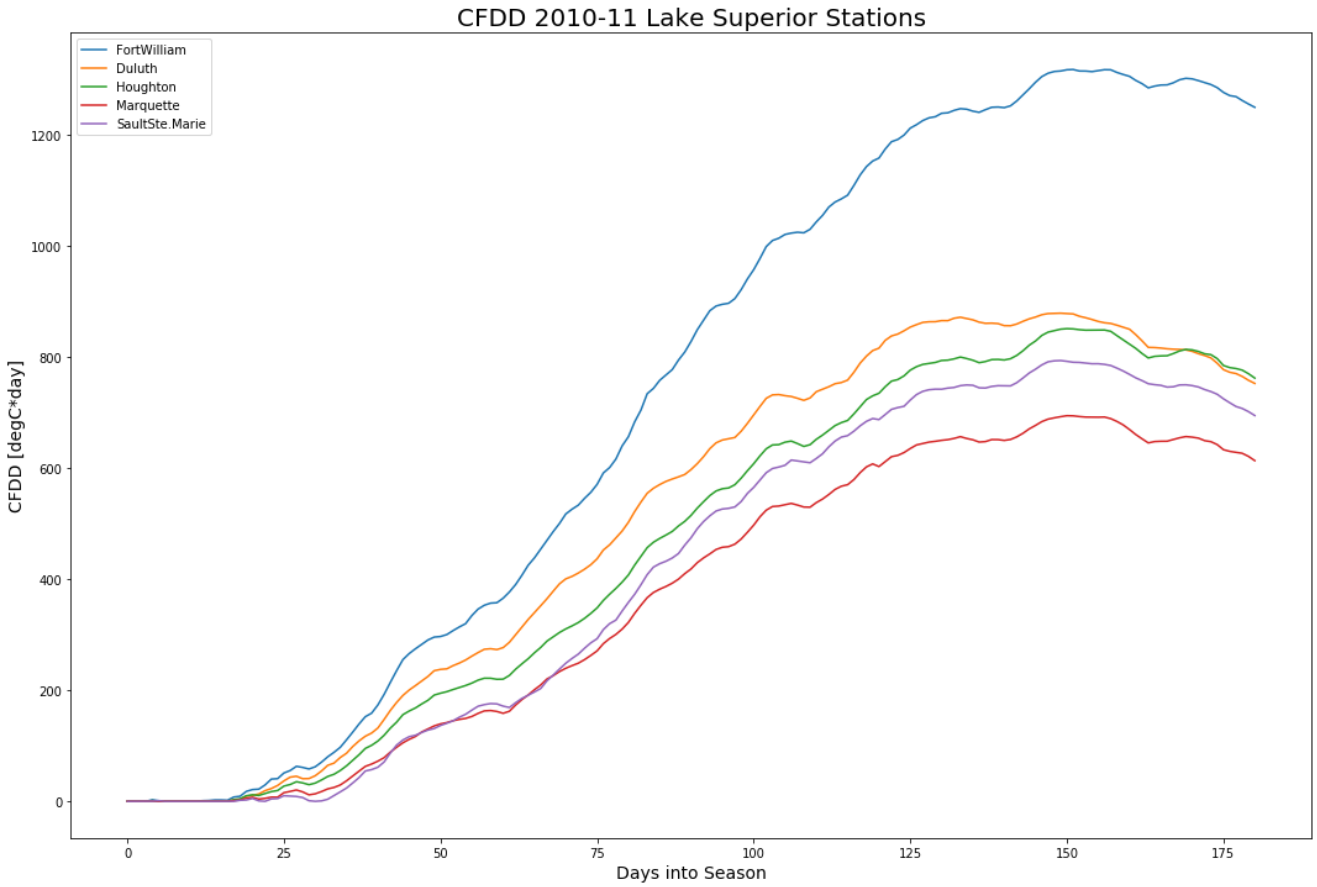

CFDD 2018-19 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
25
50
75
100
125
150
175
0
50
100
150
200
250
300

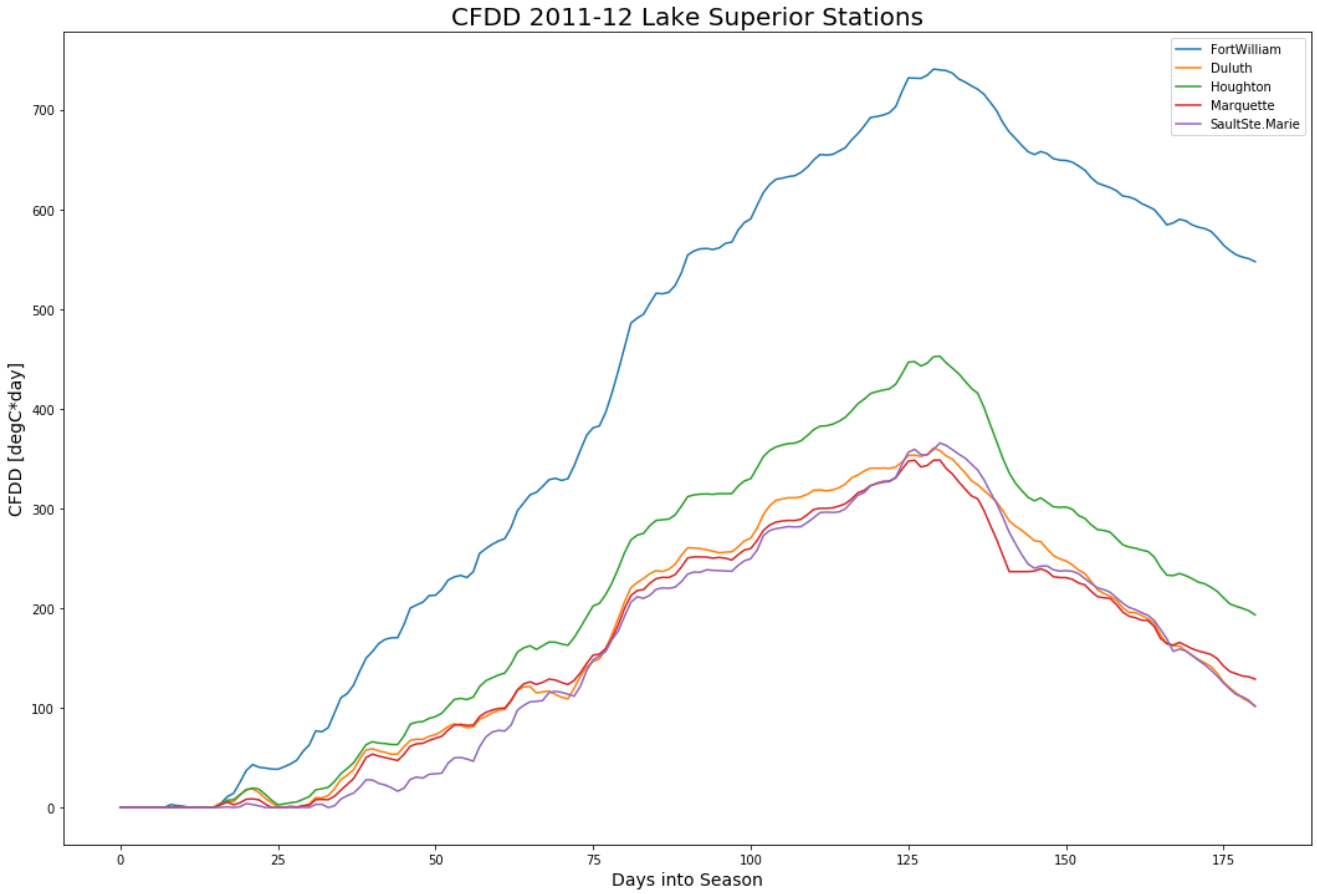

CFDD 2019-20 Lake Erie Stations
Buffalo
Erie
Cleveland
Toledo
Detroit
CFDD [degC*day]
Days into Season
0
10
20
30
40
50
60
0
25
50
75
100
125
150
175

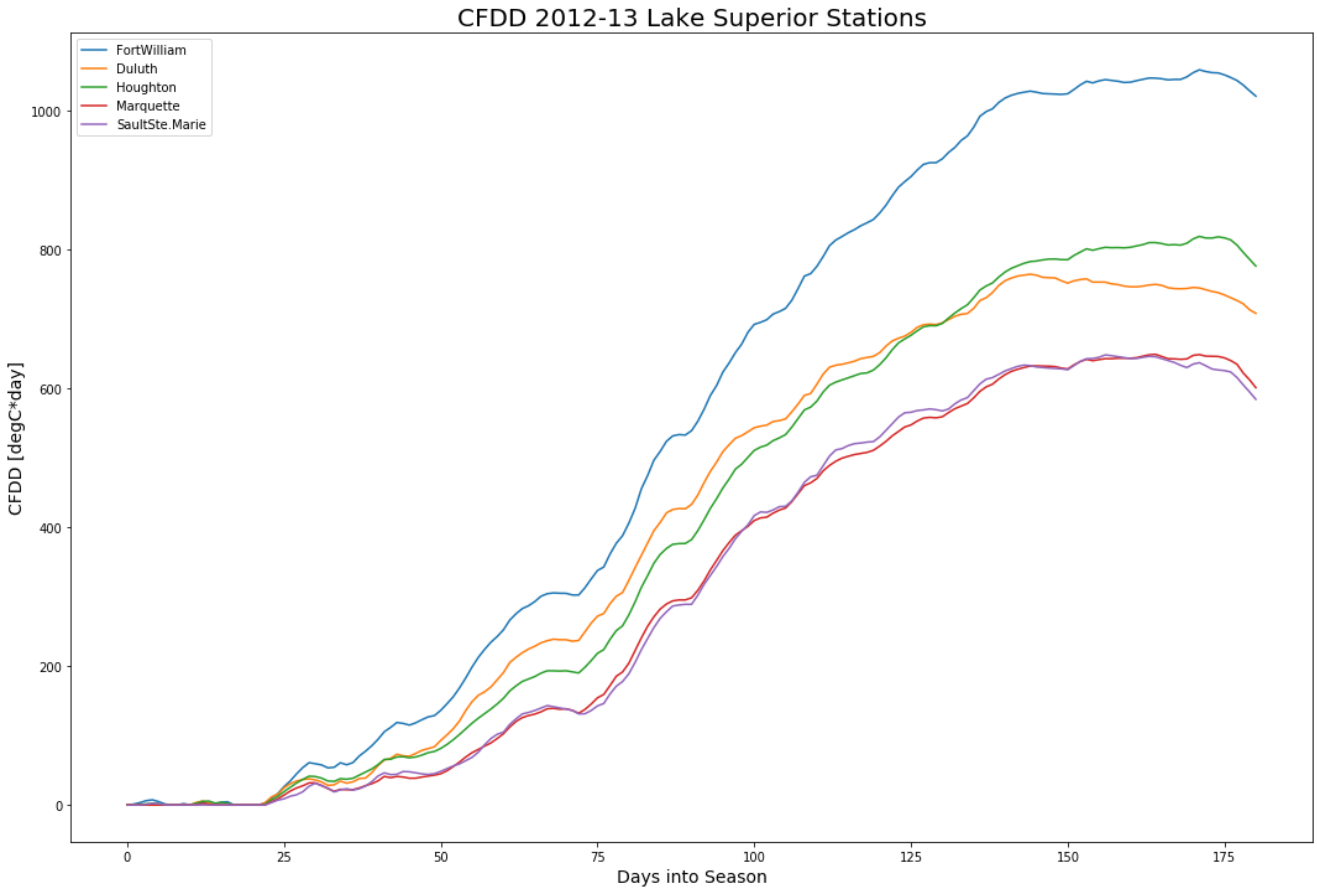
b. **Lake Superior**

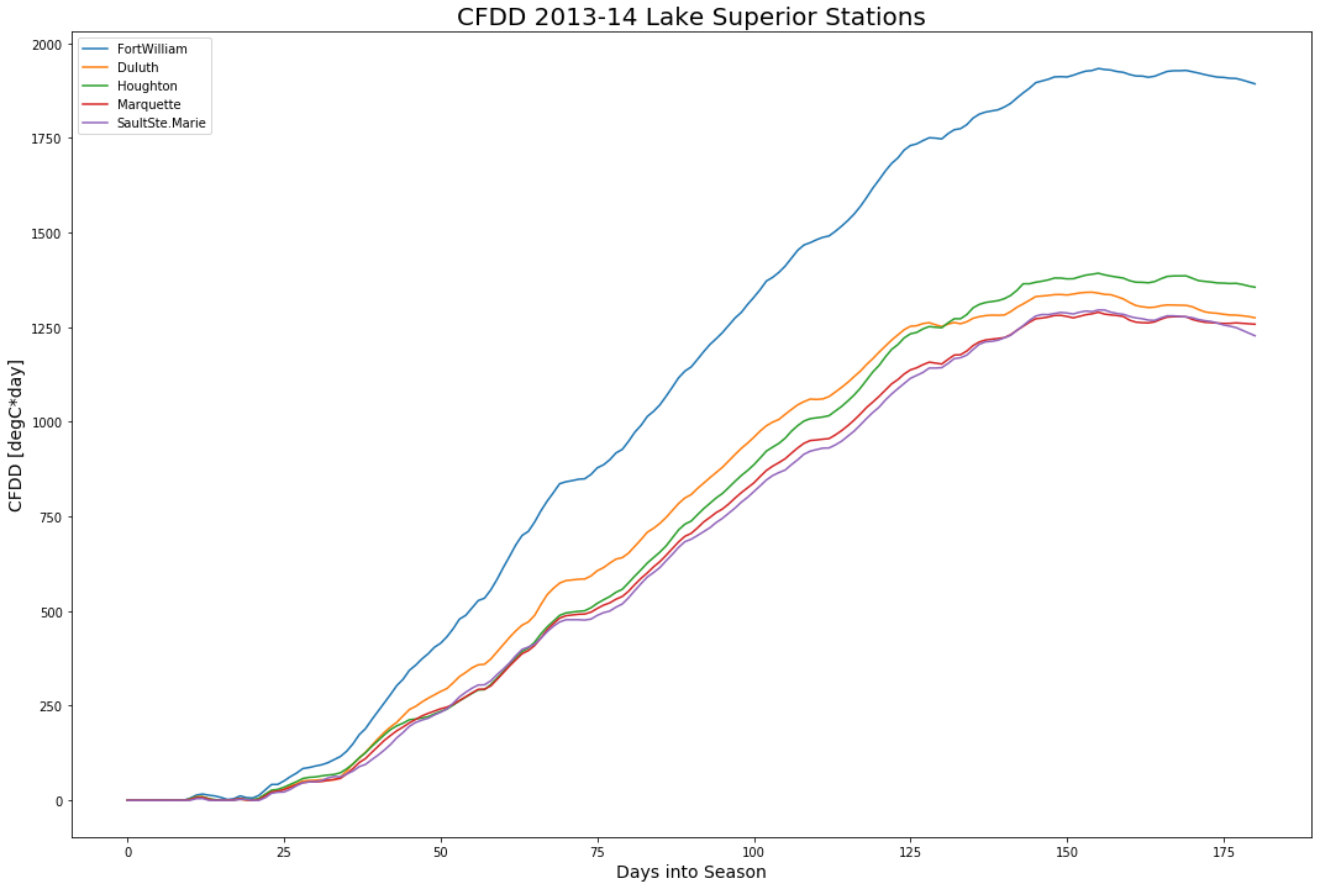


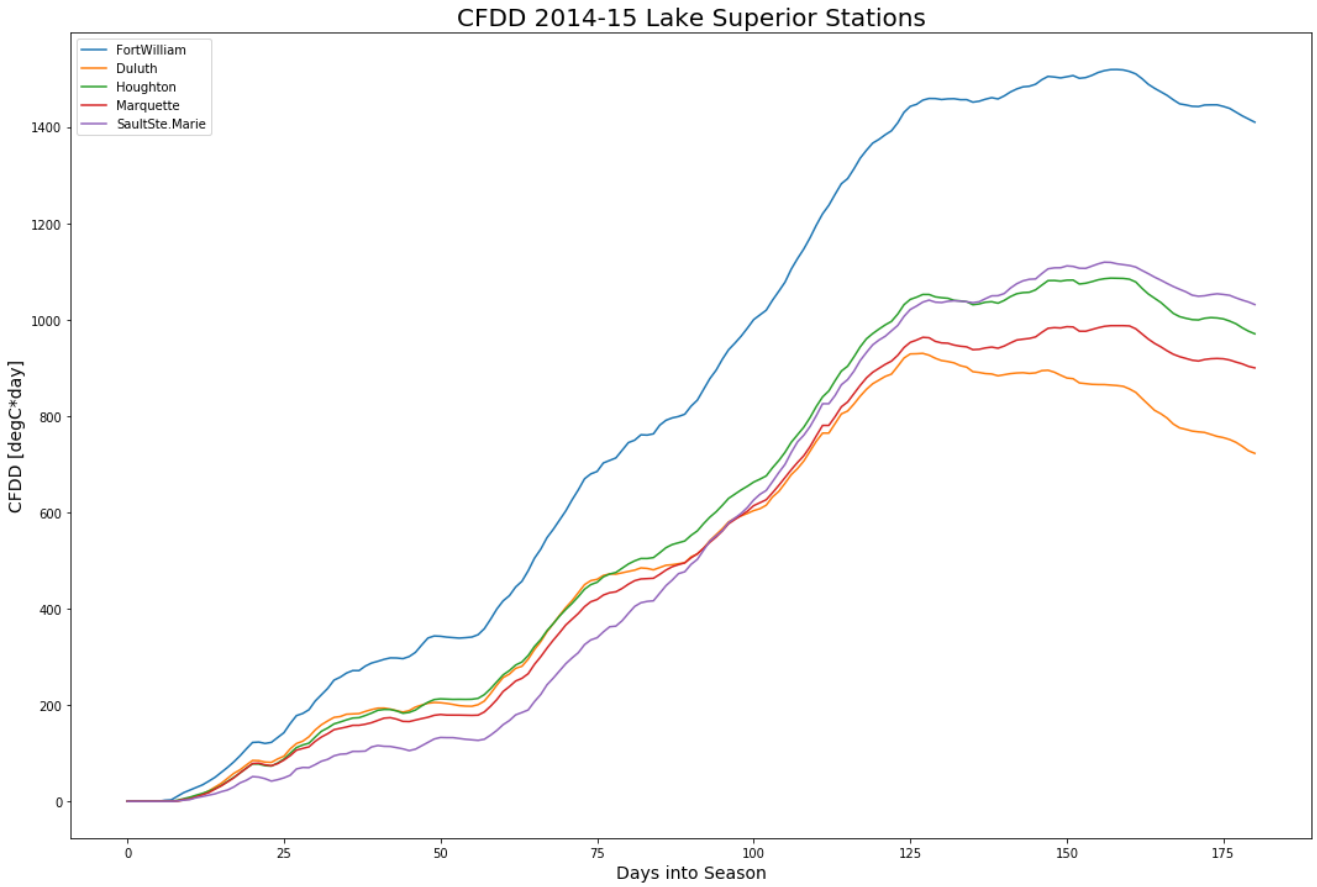


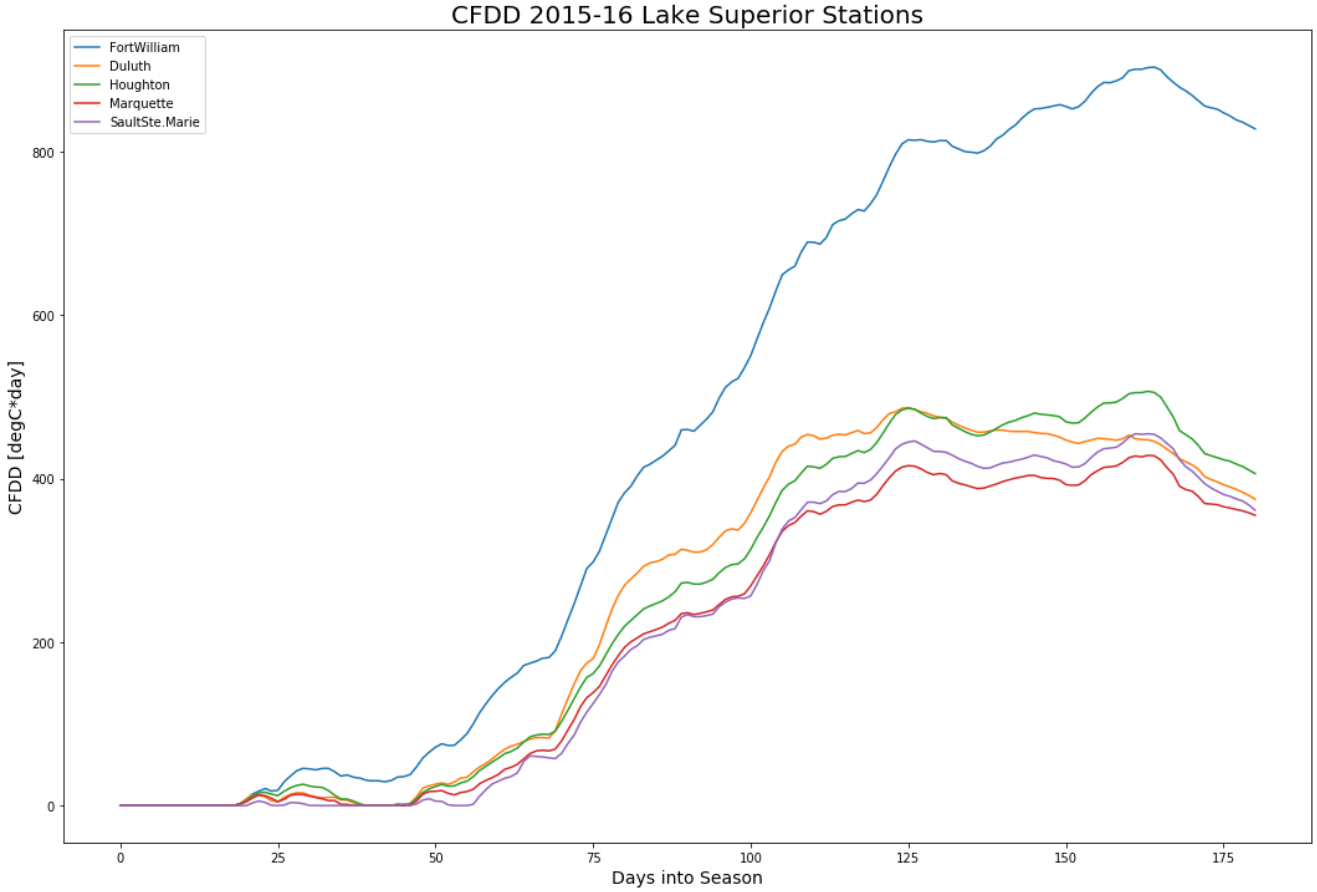

CFDD 2010-11 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season
0
200
400
600
800
1000
1200
0
25
50
75
100
125
150
175

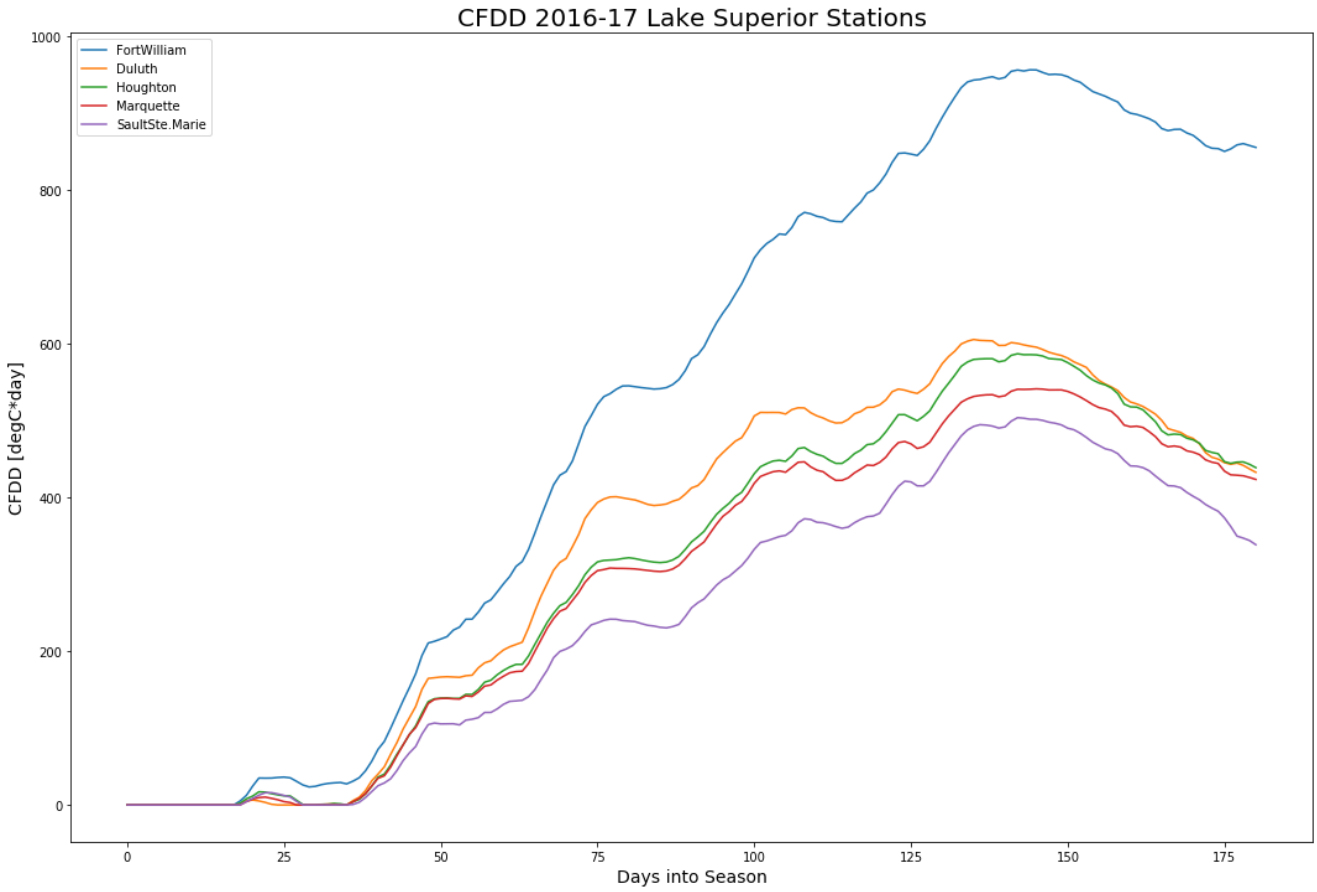

CFDD 2011-12 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
700
0
25
50
75
100
125
150
175

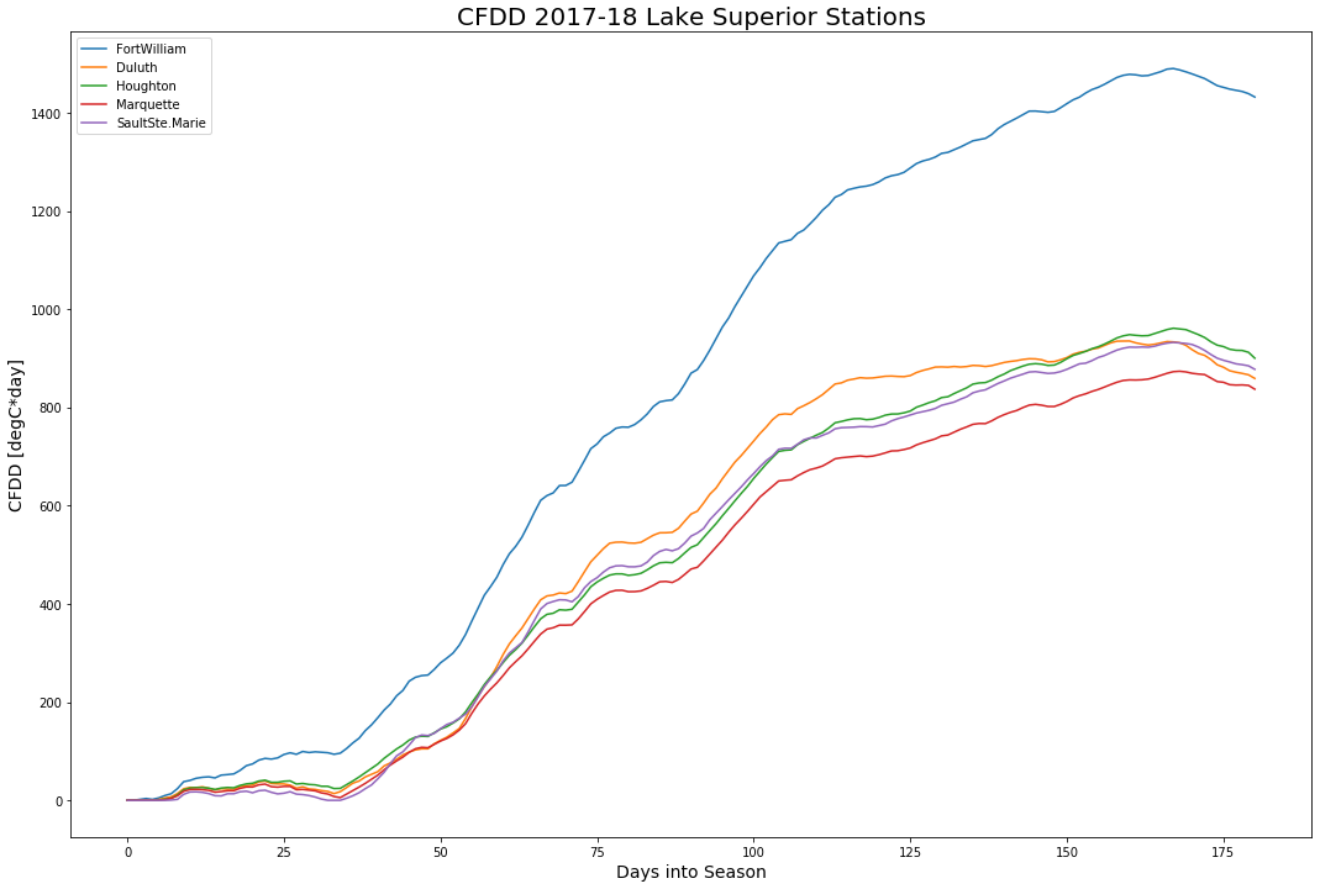

CFDD 2012-13 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season
0
200
400
600
800
1000
0
25
50
75
100
125
150
175

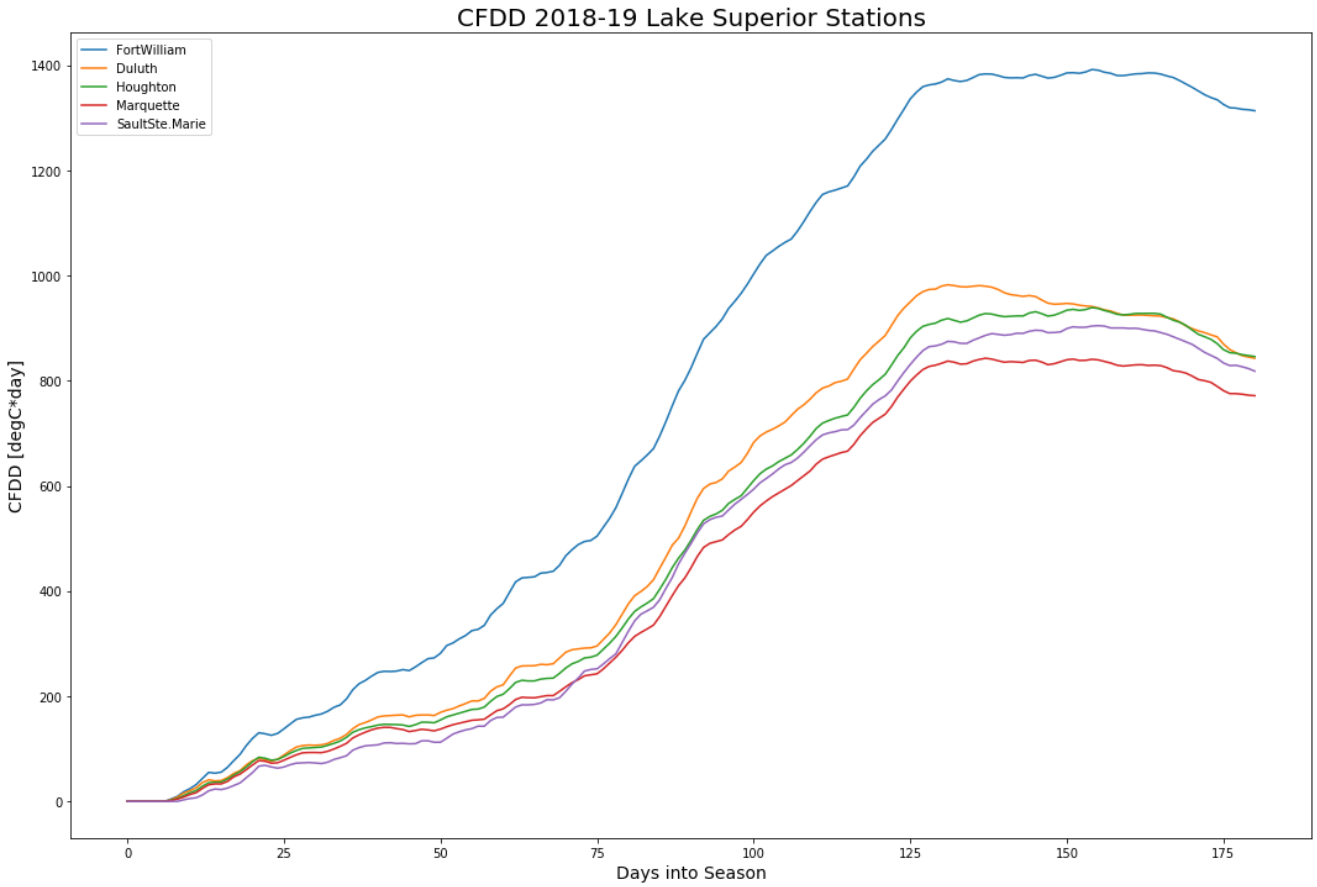

CFDD 2013-14 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season
0
250
500
750
1000
1250
1500
1750
2000
0
25
50
75
100
125
150
175

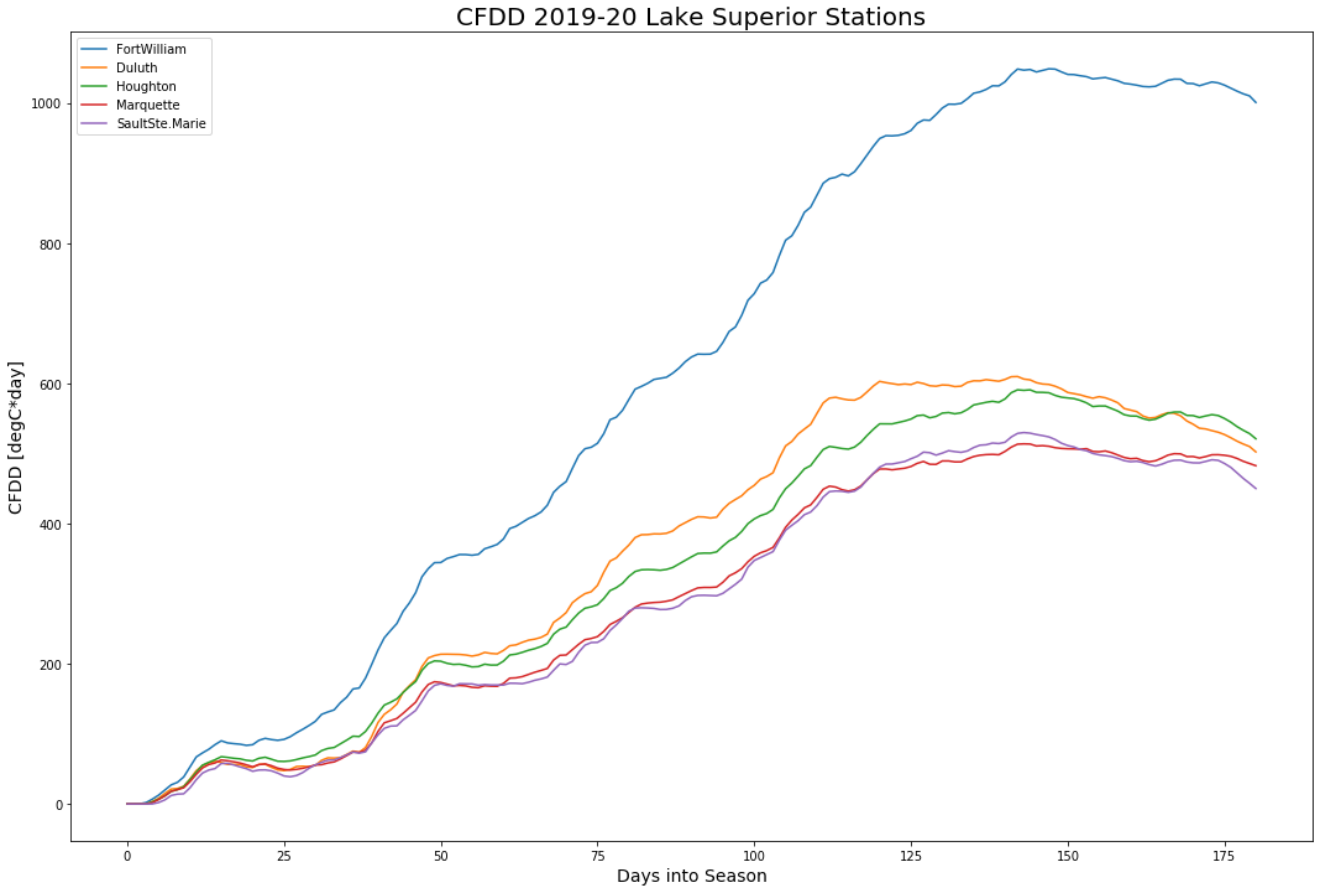

CFDD 2014-15 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season
0
200
400
600
800
1000
1200
1400
0
25
50
75
100
125
150
175


CFDD 2015-16 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season
0
200
400
600
800
0
25
50
75
100
125
150
175

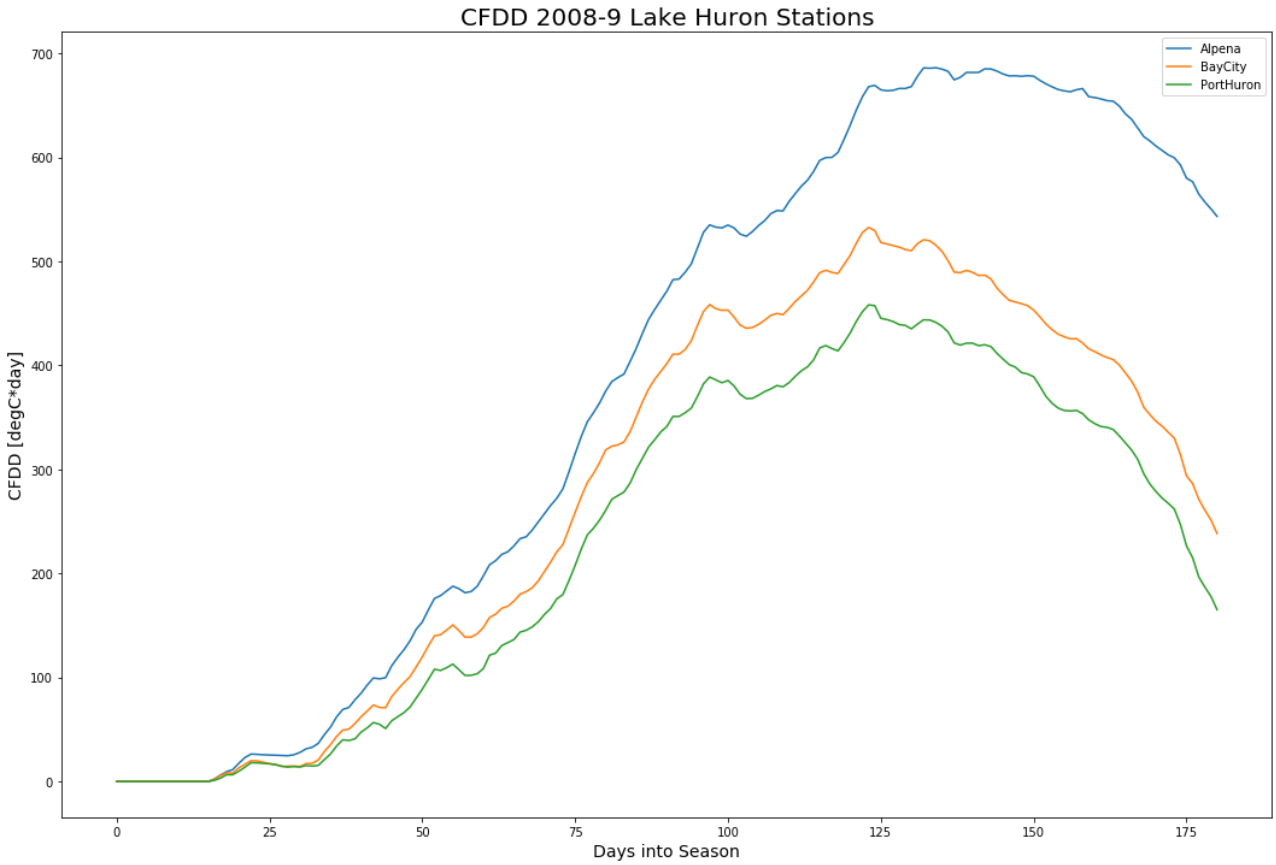

CFDD 2016-17 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season
0
200
400
600
800
1000
0
25
50
75
100
125
150
175

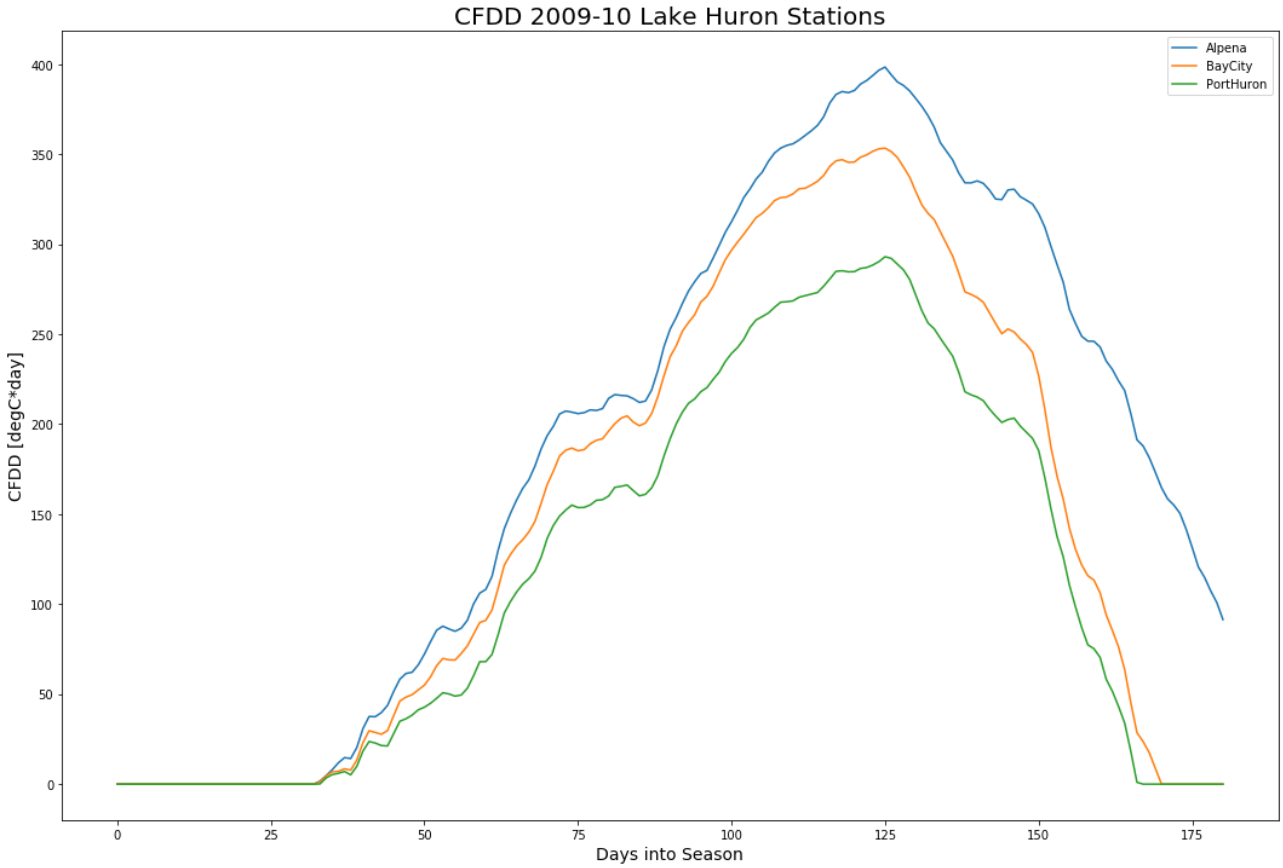

CFDD 2017-18 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season
0
200
400
600
800
1000
1200
1400
0
25
50
75
100
125
150
175

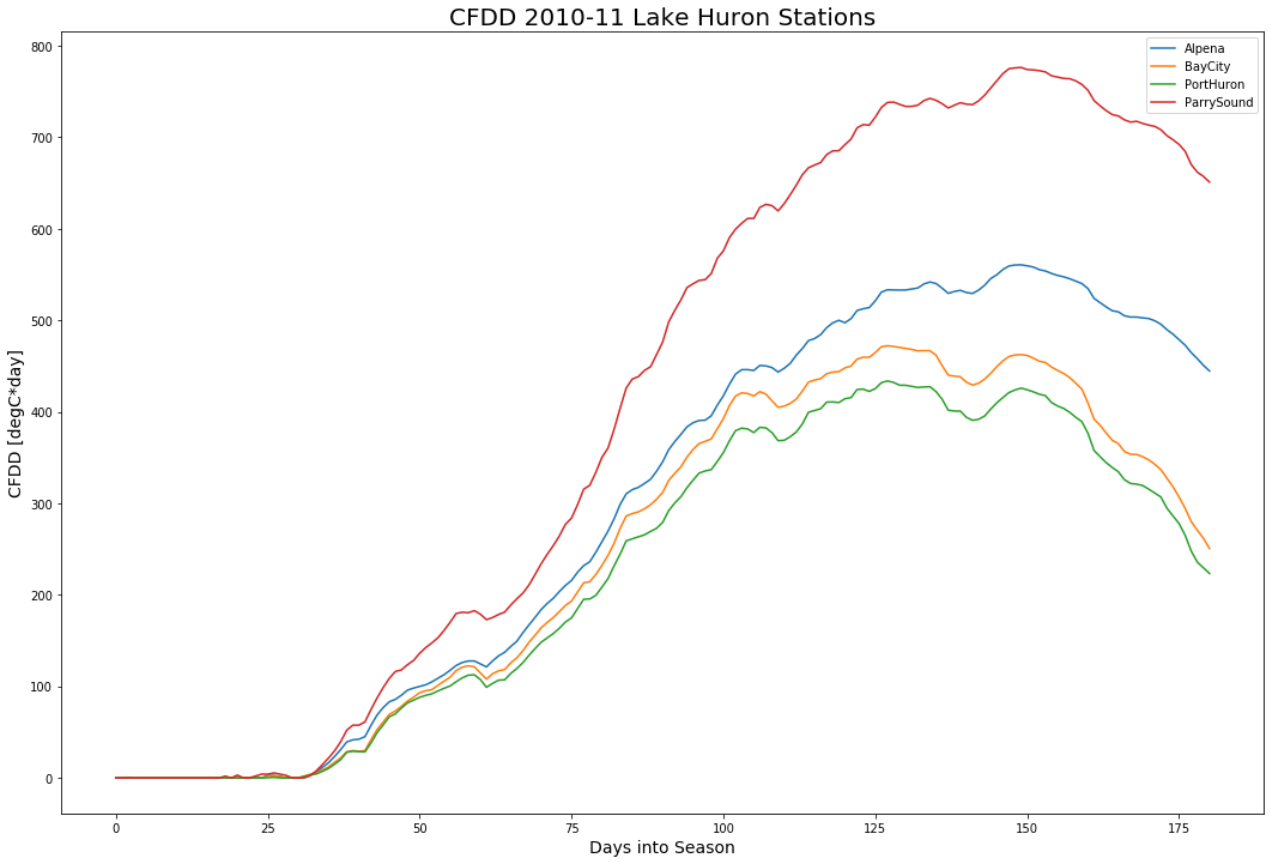

CFDD 2018-19 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season

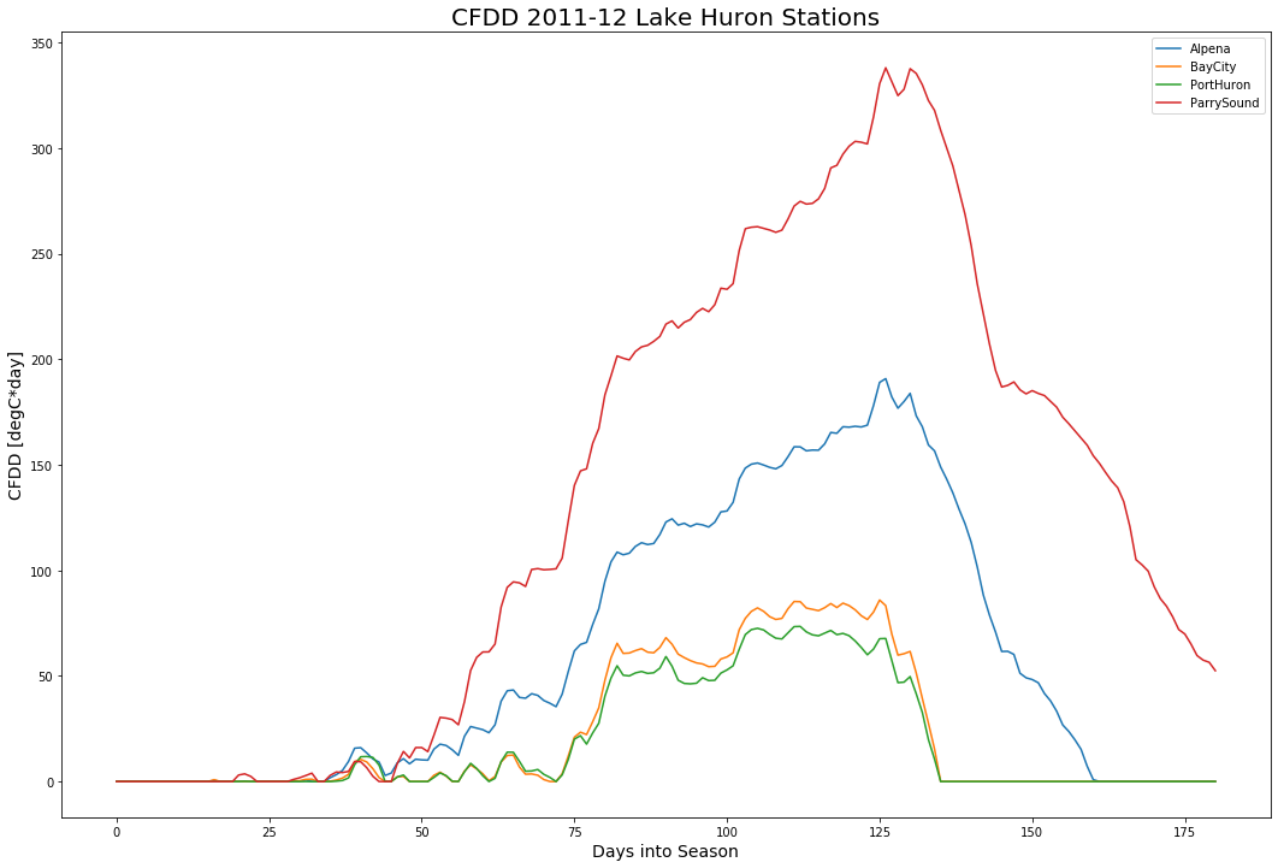

CFDD 2019-20 Lake Superior Stations
FortWilliam
Duluth
Houghton
Marquette
SaultSte.Marie
CFDD [degC*day]
Days into Season

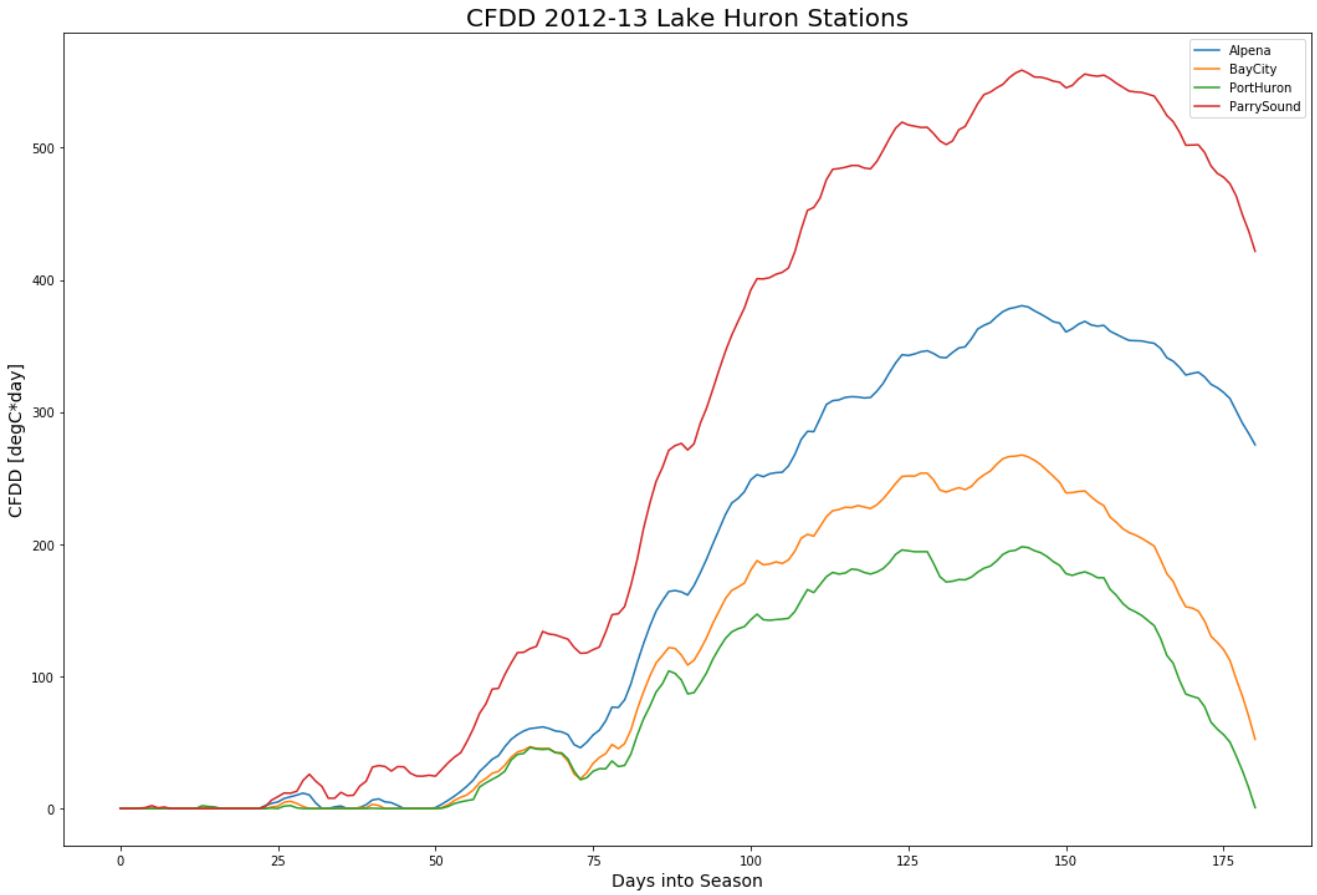
c. **Lake Huron**

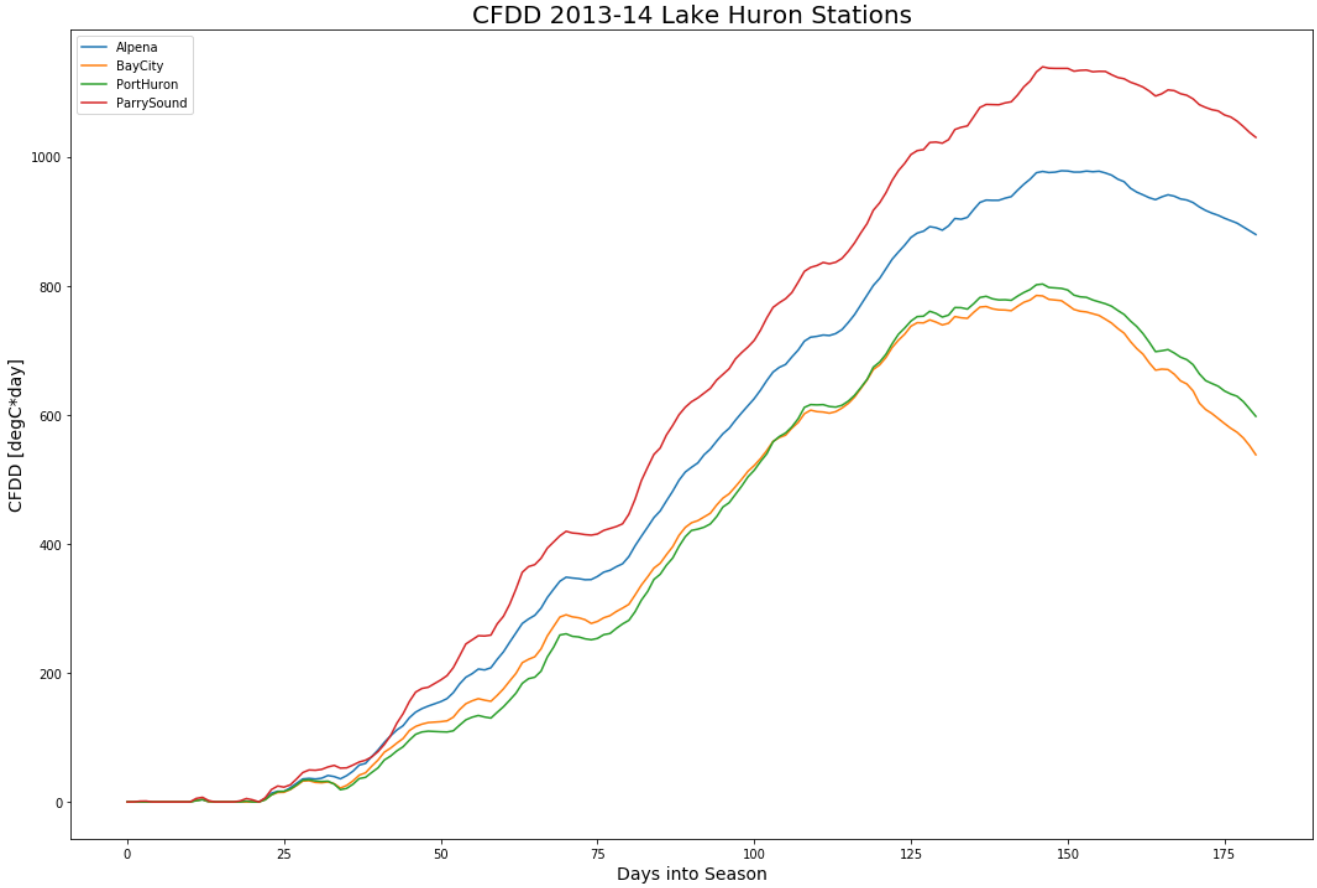

CFDD 2008-9 Lake Huron Stations
Alpena
BayCity
PortHuron
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
700
0
25
50
75
100
125
150
175

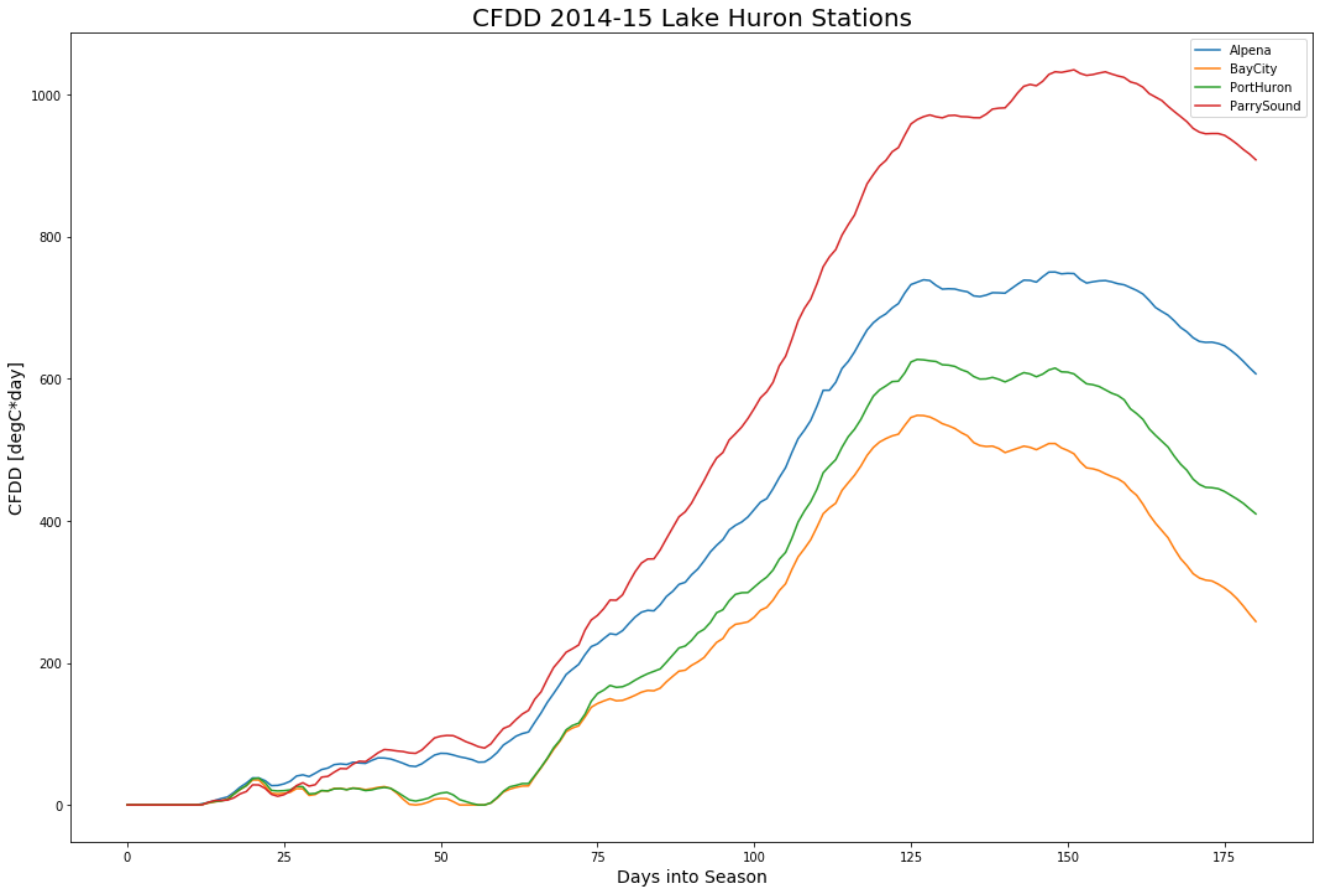

CFDD 2009-10 Lake Huron Stations
Alpena
BayCity
PortHuron
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
300
350
400
0
25
50
75
100
125
150
175

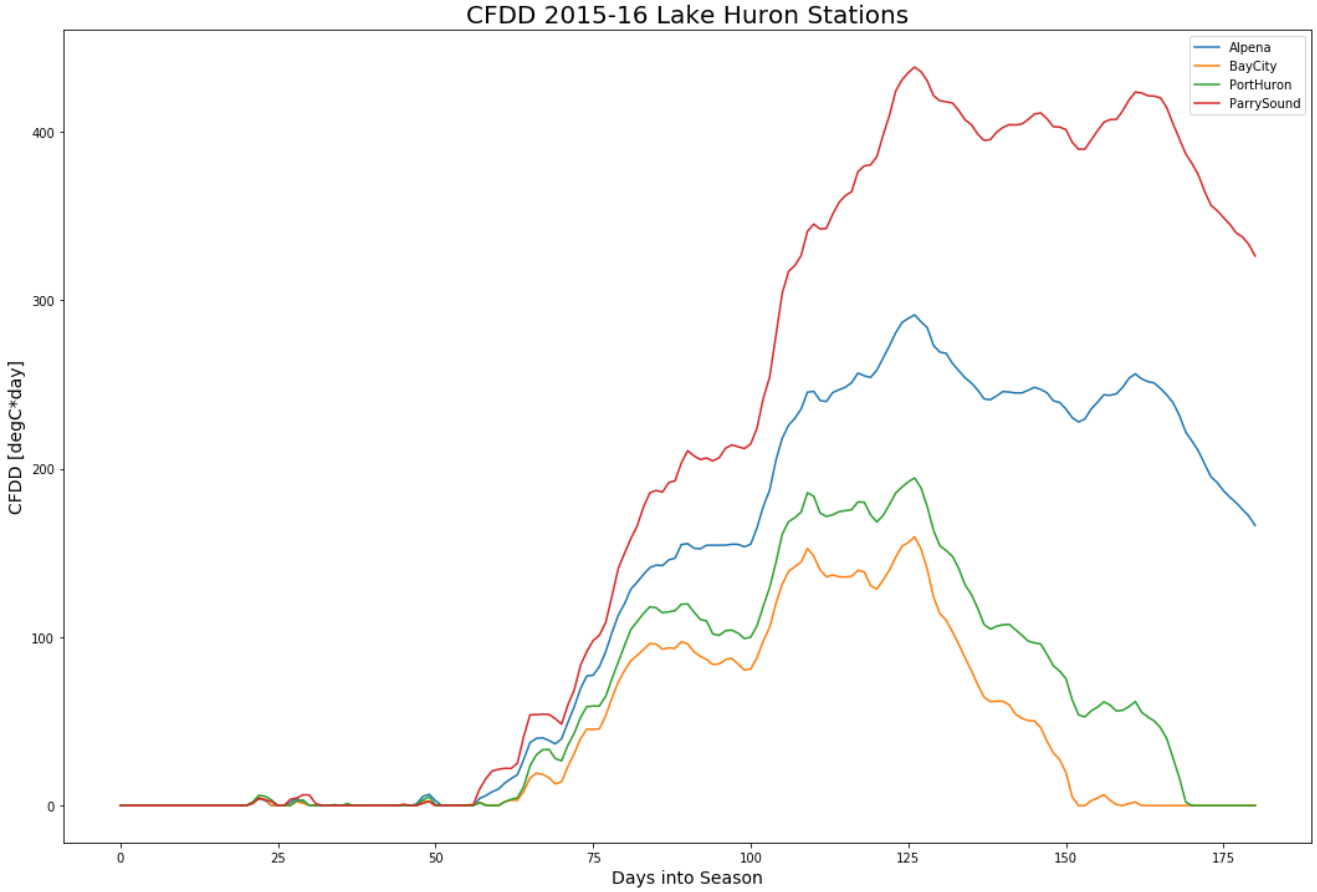

CFDD 2010-11 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
700
800
0
25
50
75
100
125
150
175

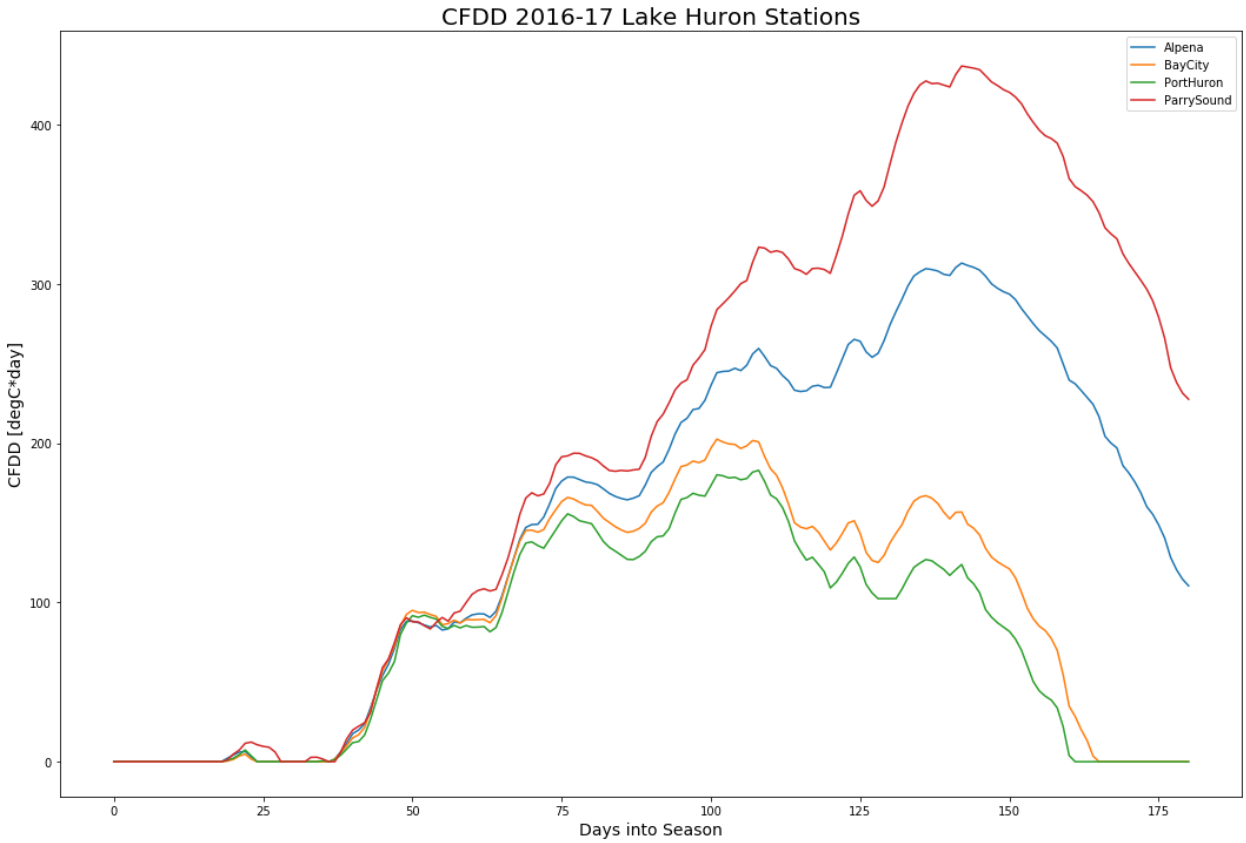

CFDD 2011-12 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
300
350
0
25
50
75
100
125
150
175

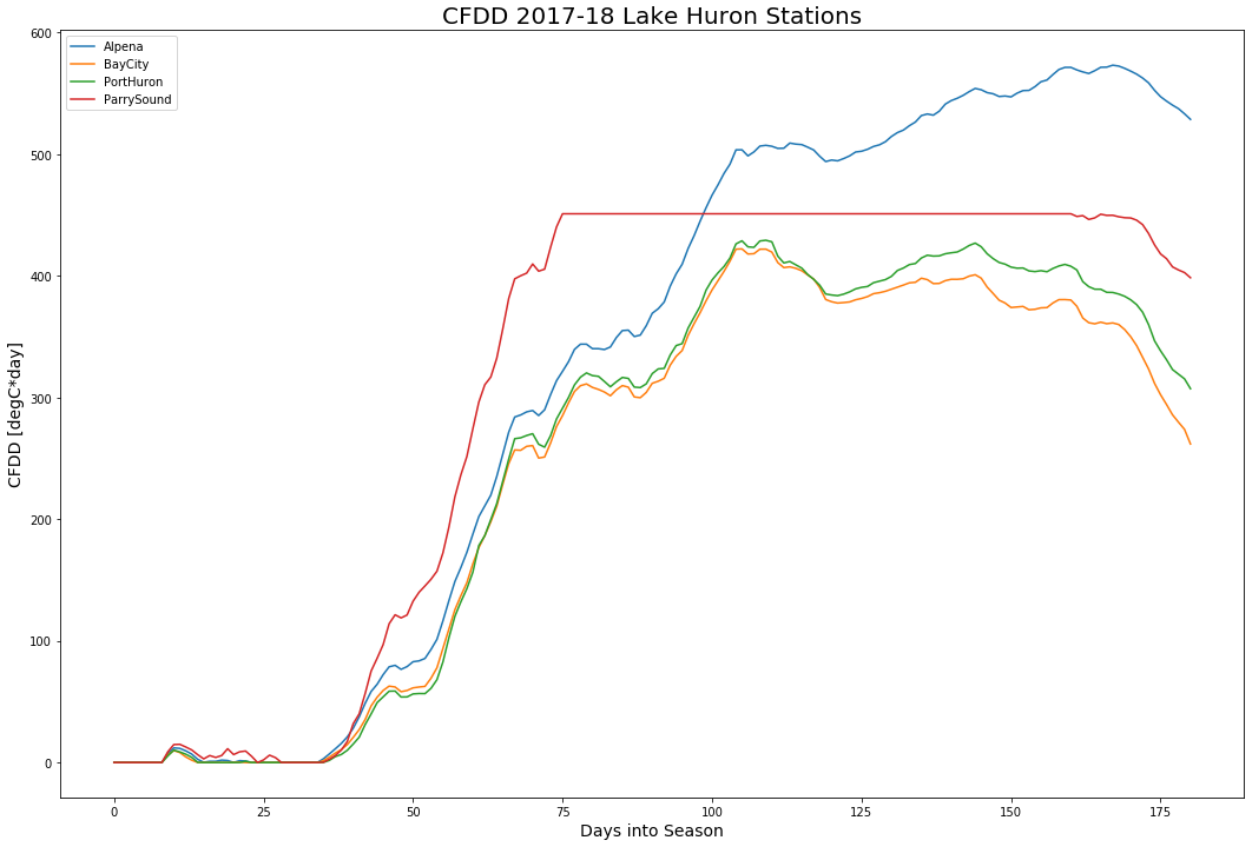

CFDD 2012-13 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
0
25
50
75
100
125
150
175

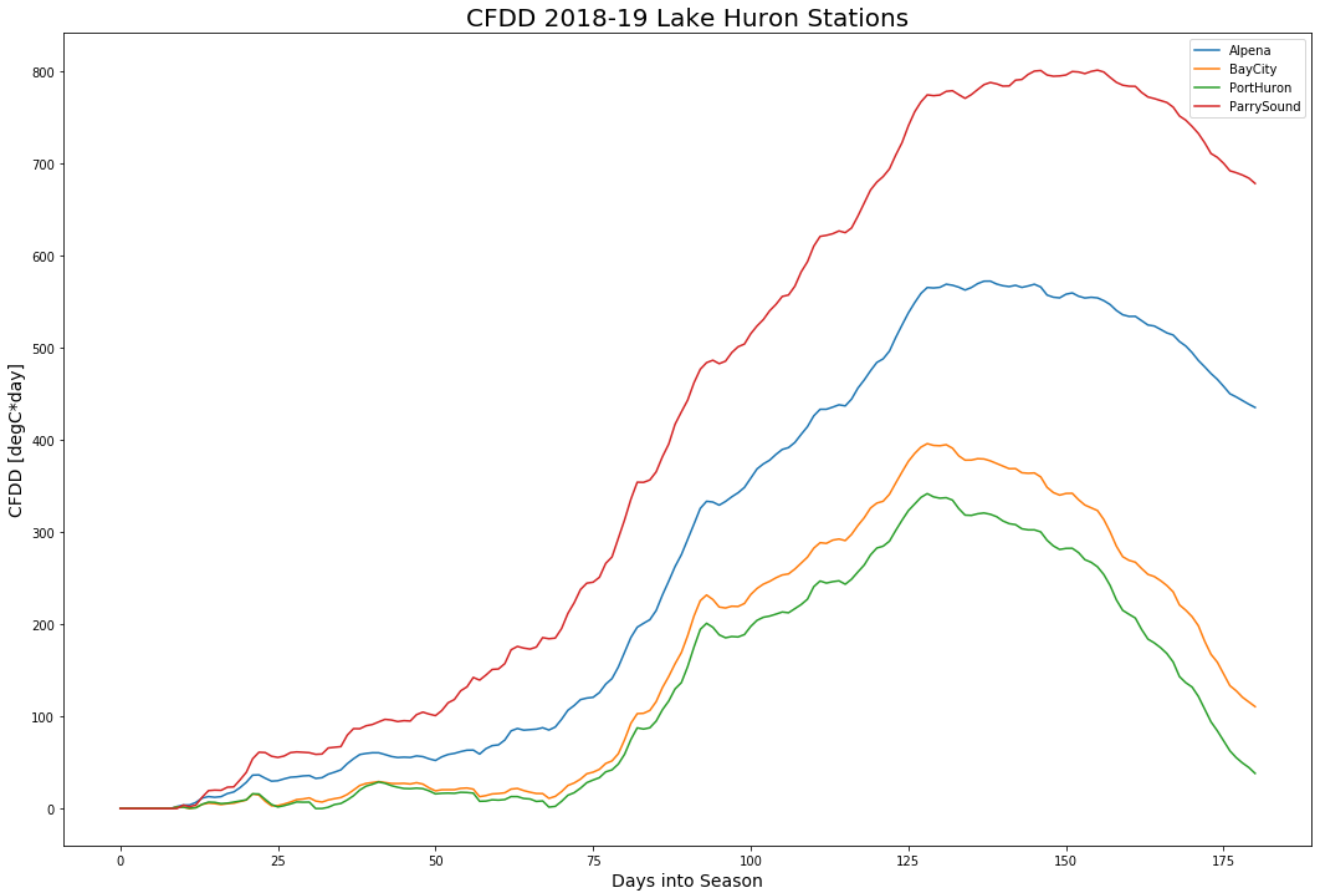

CFDD 2013-14 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
200
400
600
800
1000
0
25
50
75
100
125
150
175

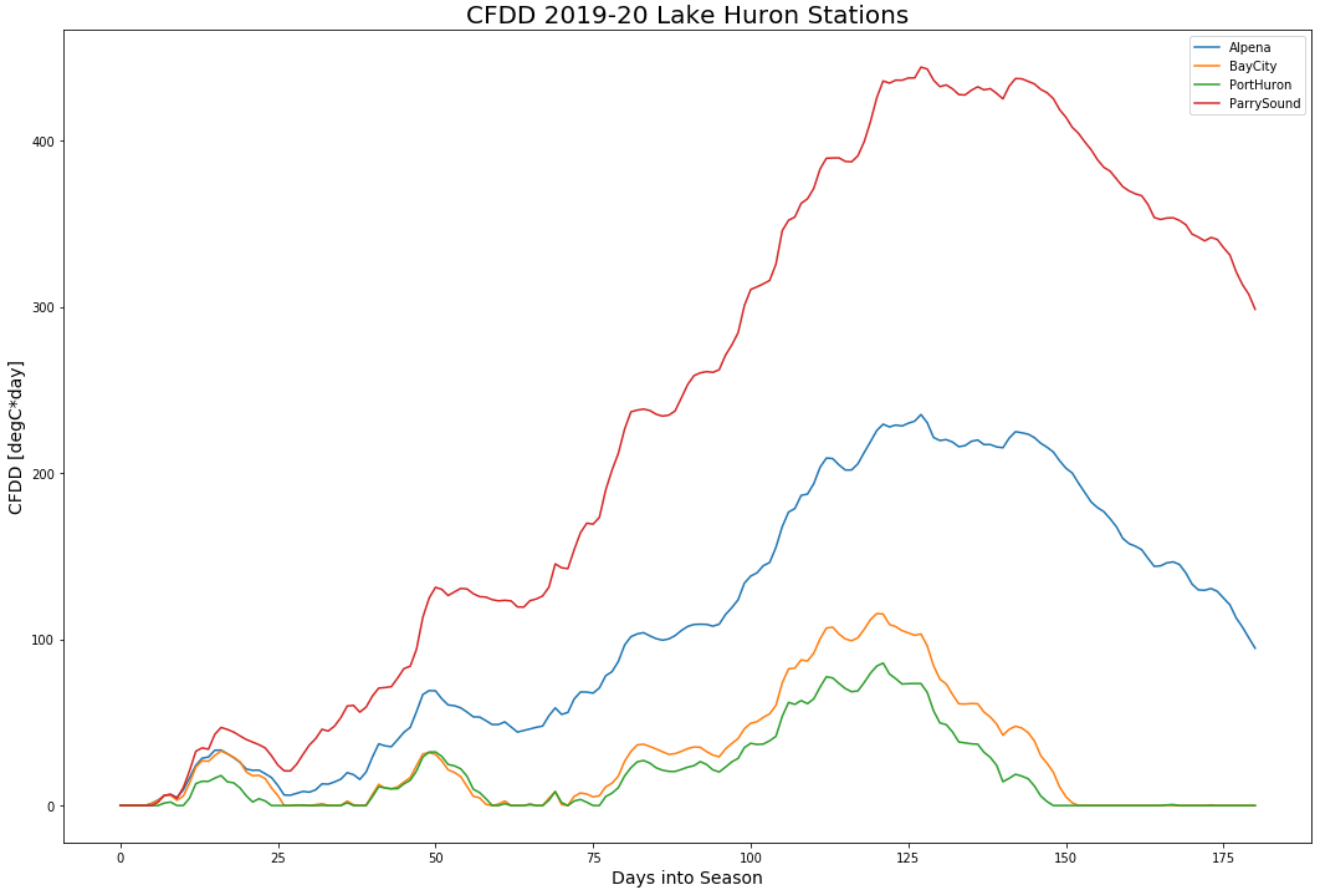

CFDD 2014-15 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
200
400
600
800
1000
0
25
50
75
100
125
150
175


CFDD 2015-16 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
100
200
300
400
0
25
50
75
100
125
150
175

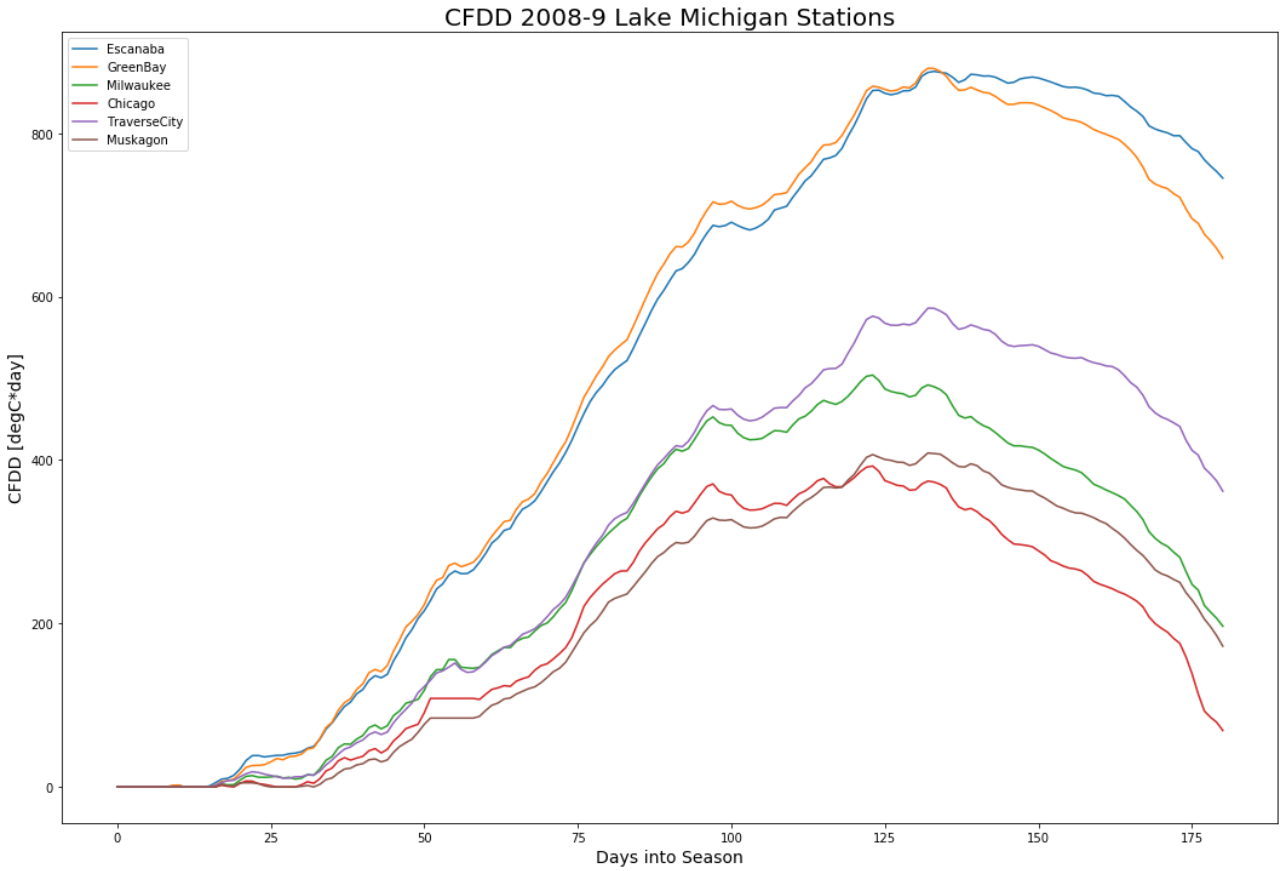

CFDD 2016-17 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
100
200
300
400
0
25
50
75
100
125
150
175

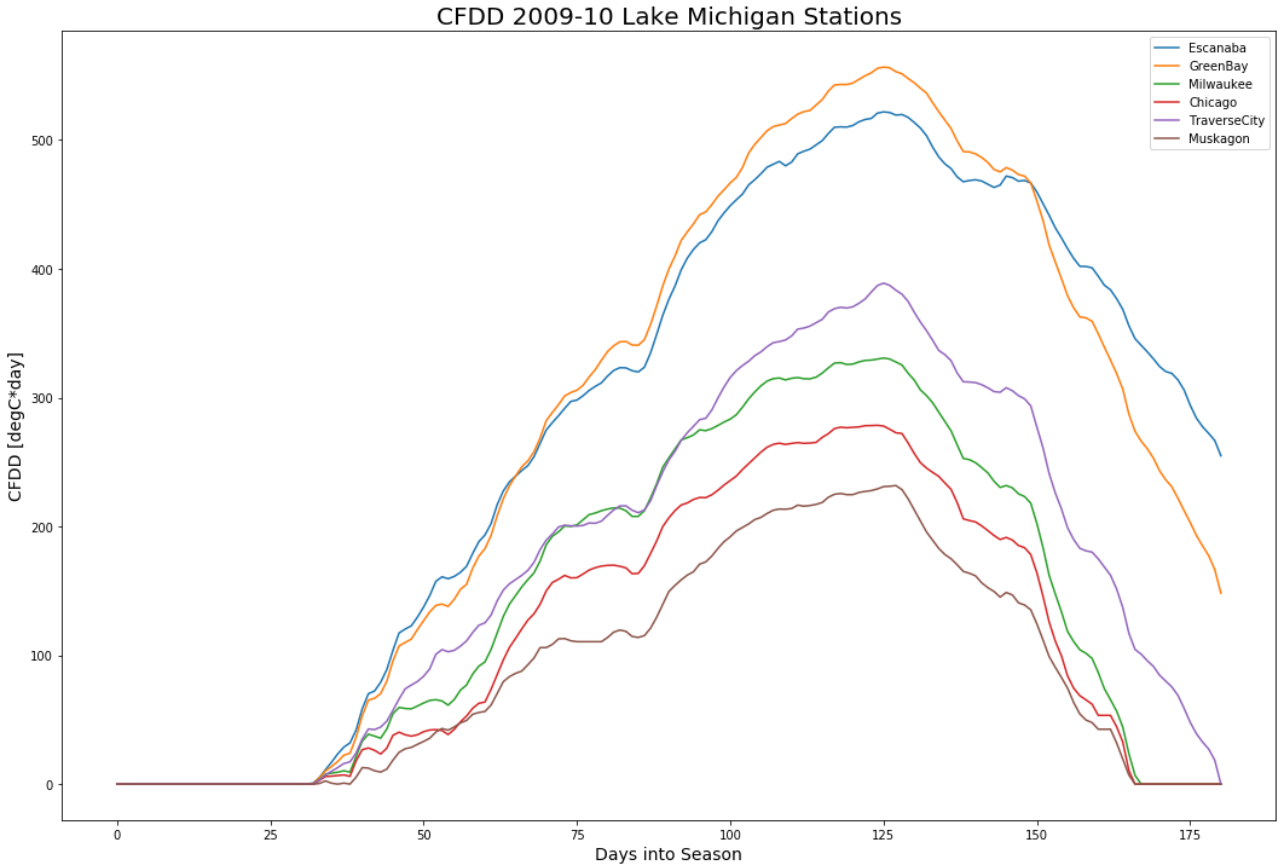

CFDD 2017-18 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
0
25
50
75
100
125
150
175

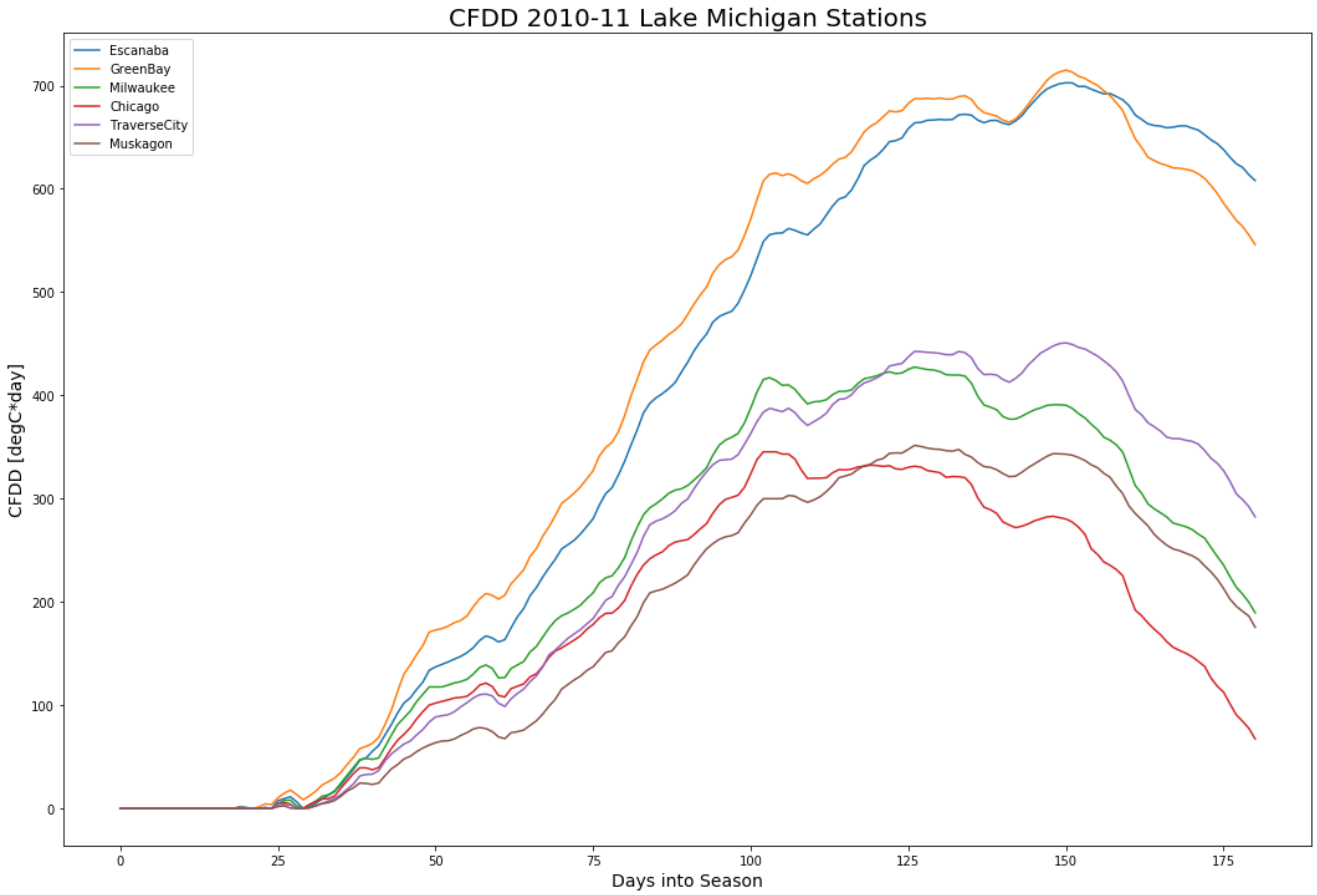

CFDD 2018-19 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
700
800
0
25
50
75
100
125
150
175

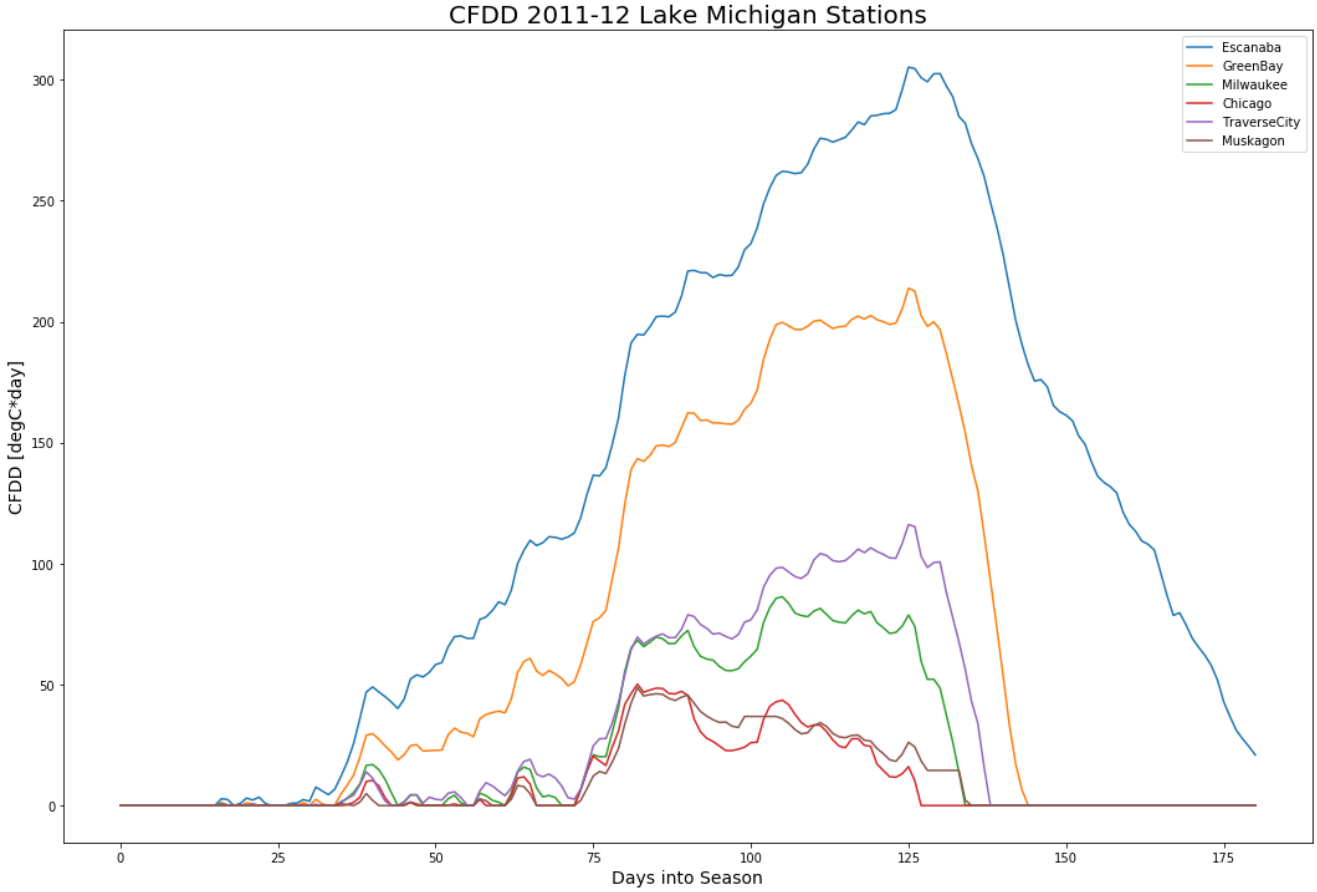

CFDD 2019-20 Lake Huron Stations
Alpena
BayCity
PortHuron
ParrySound
CFDD [degC*day]
Days into Season
0
100
200
300
400
0
25
50
75
100
125
150
175

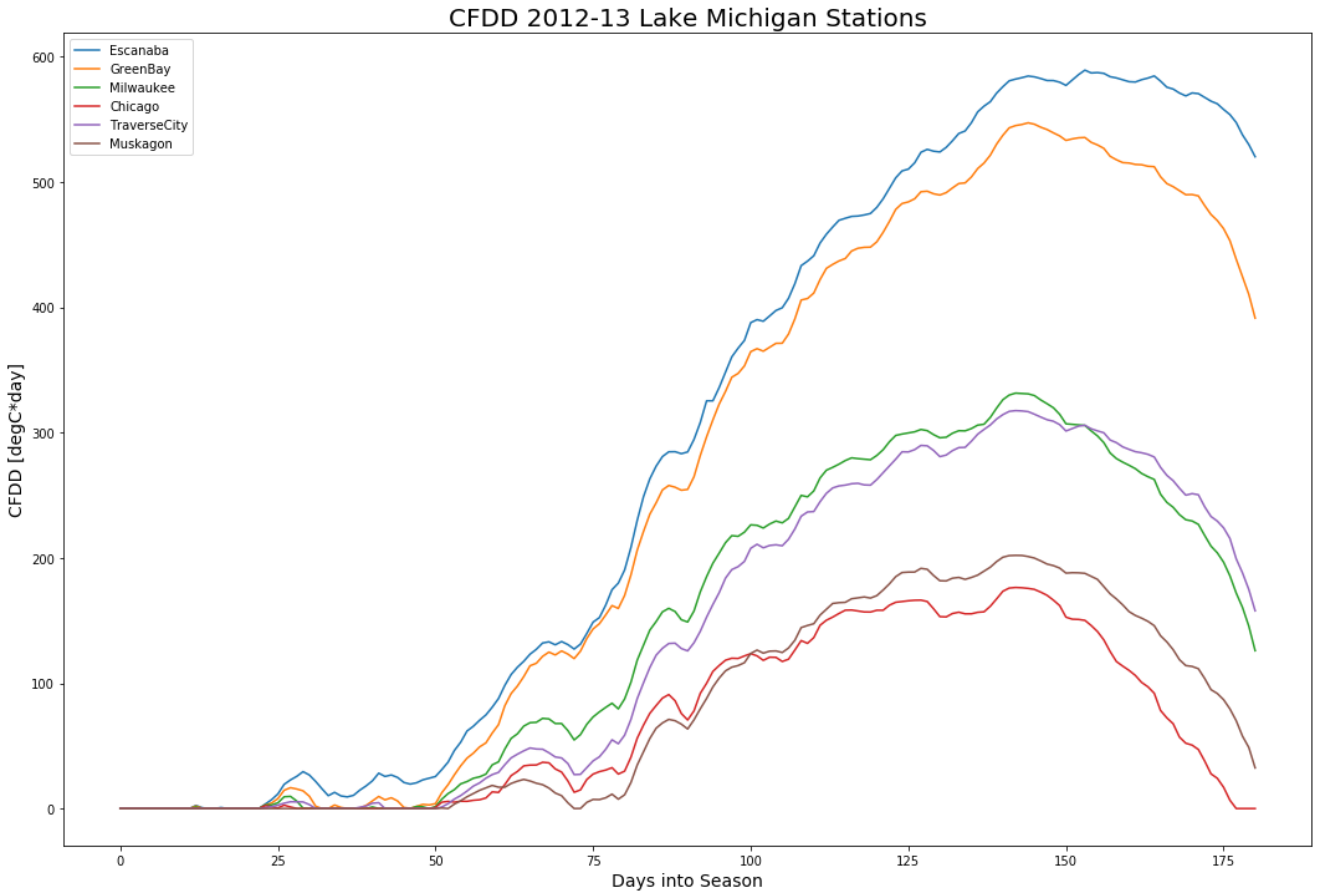
d. **Lake Michigan**

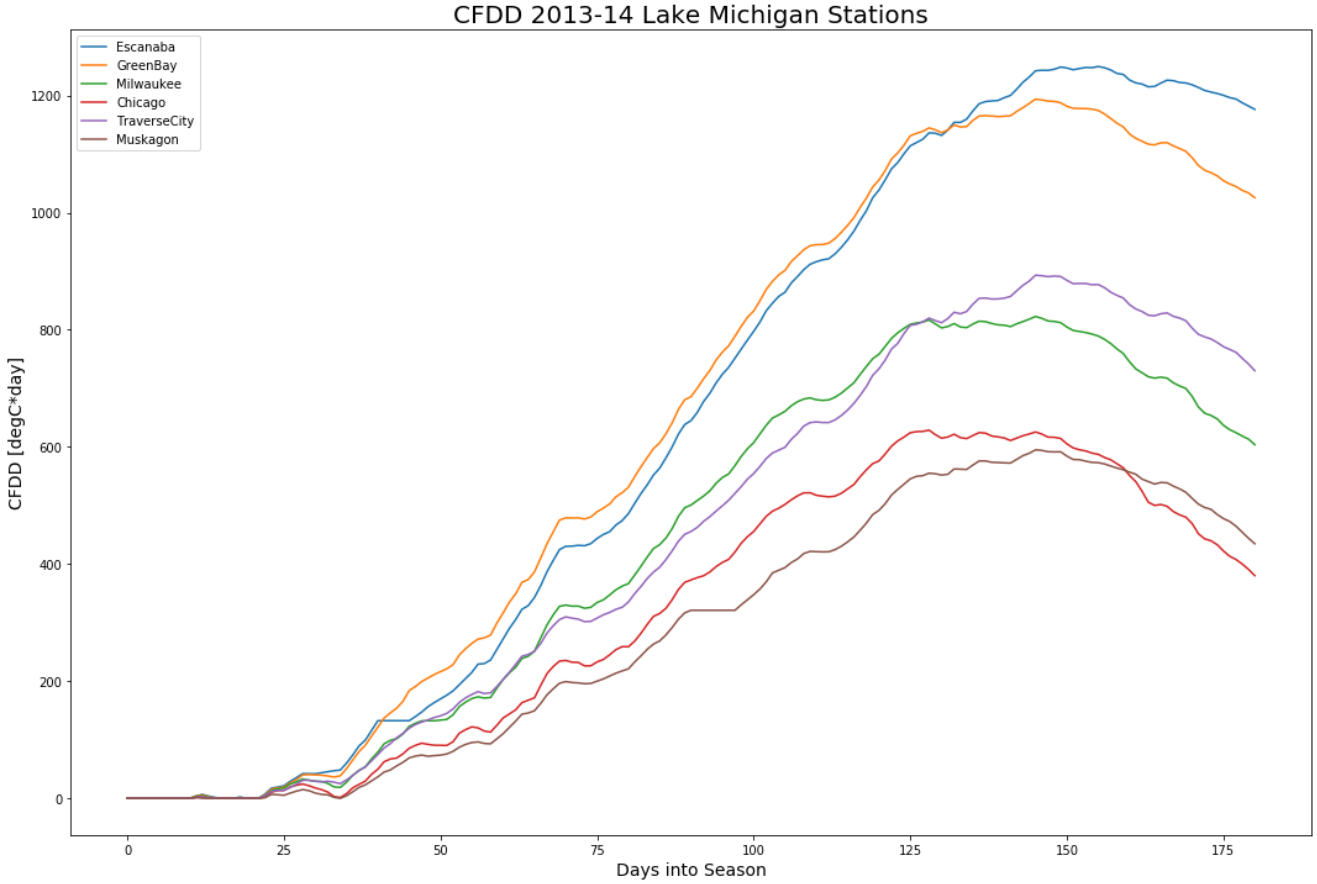


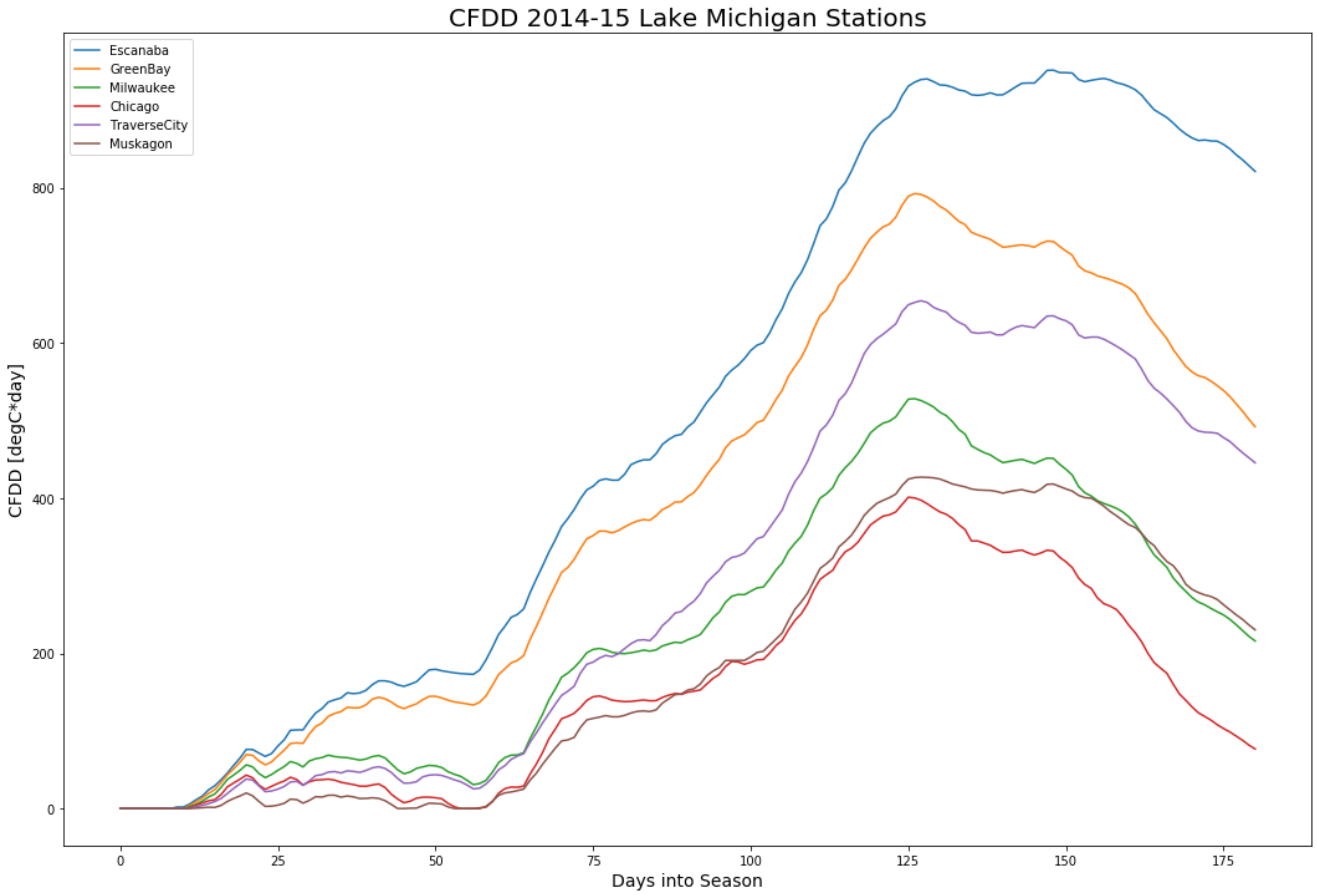


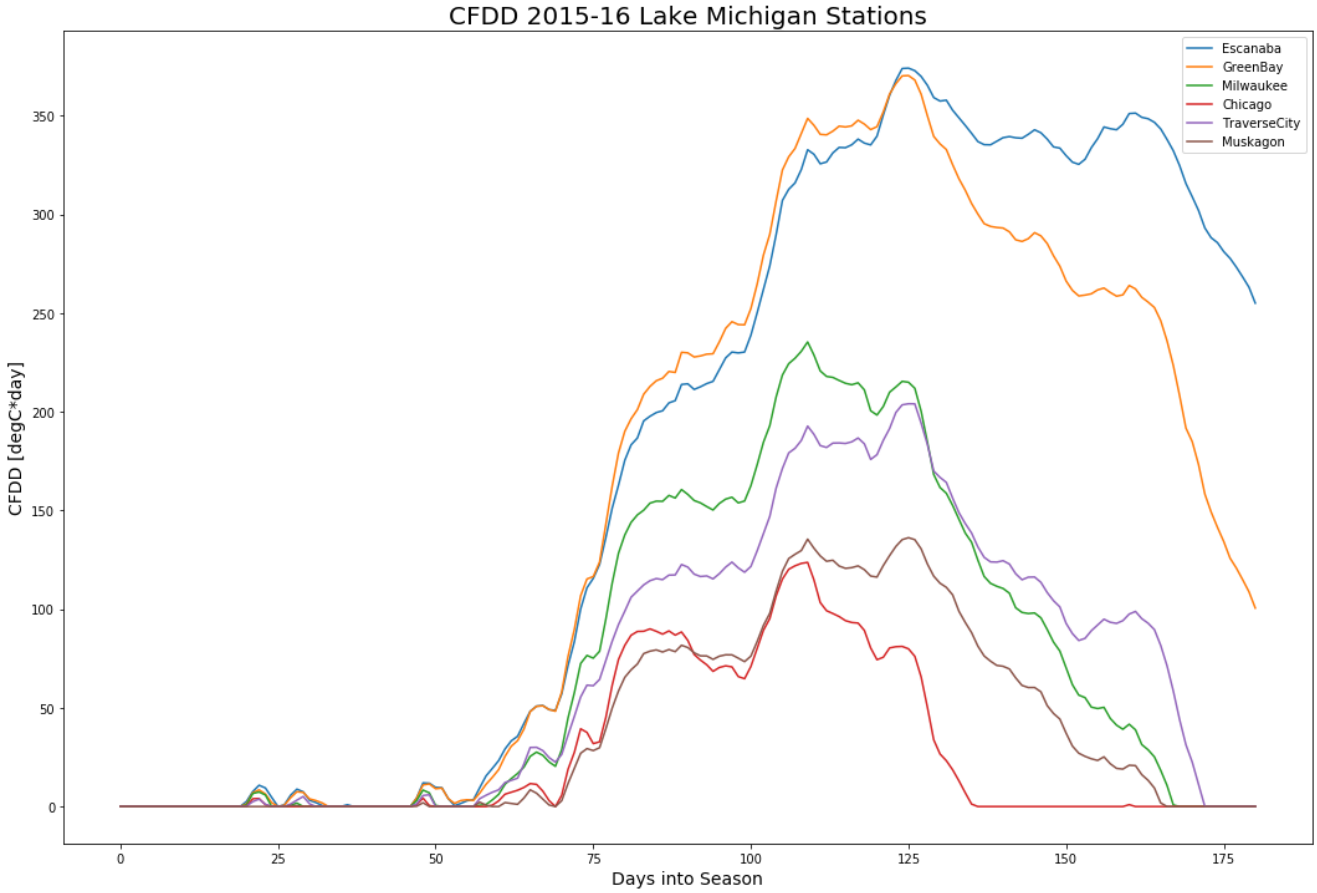

CFDD 2010-11 Lake Michigan Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season

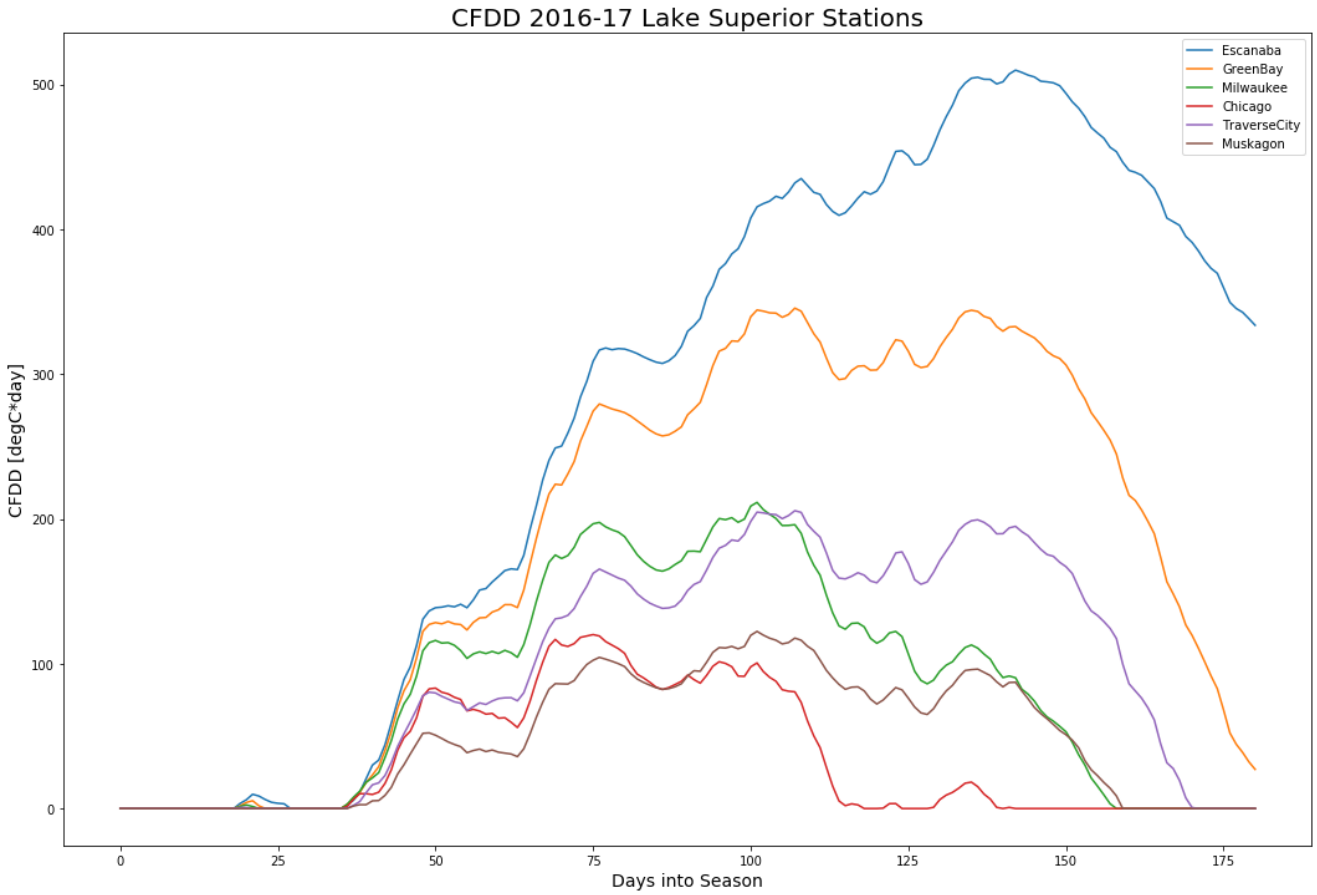

CFDD 2011-12 Lake Michigan Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season

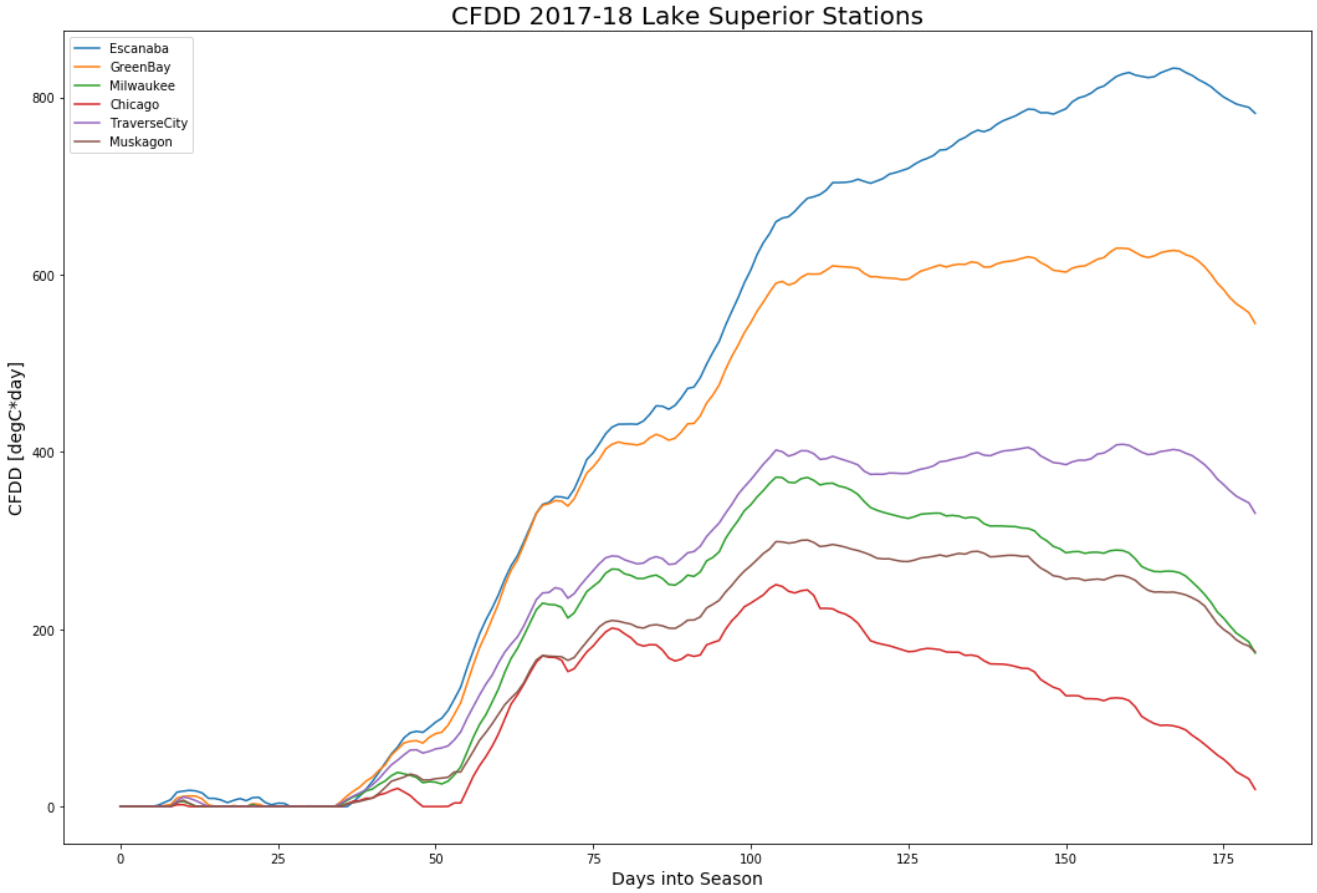

CFDD 2012-13 Lake Michigan Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
0
25
50
75
100
125
150
175

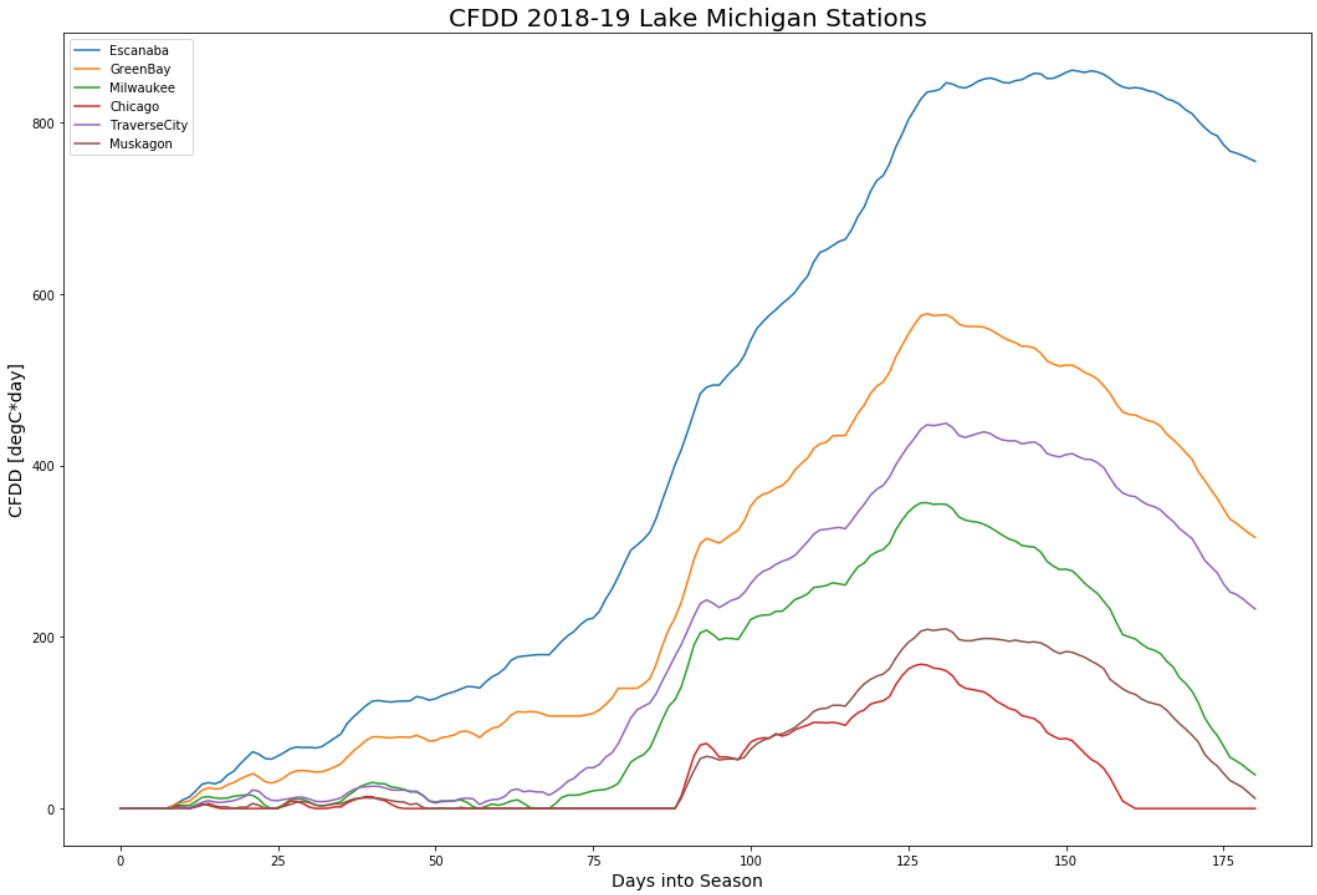

CFDD 2013-14 Lake Michigan Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season
0
200
400
600
800
1000
1200
0
25
50
75
100
125
150
175

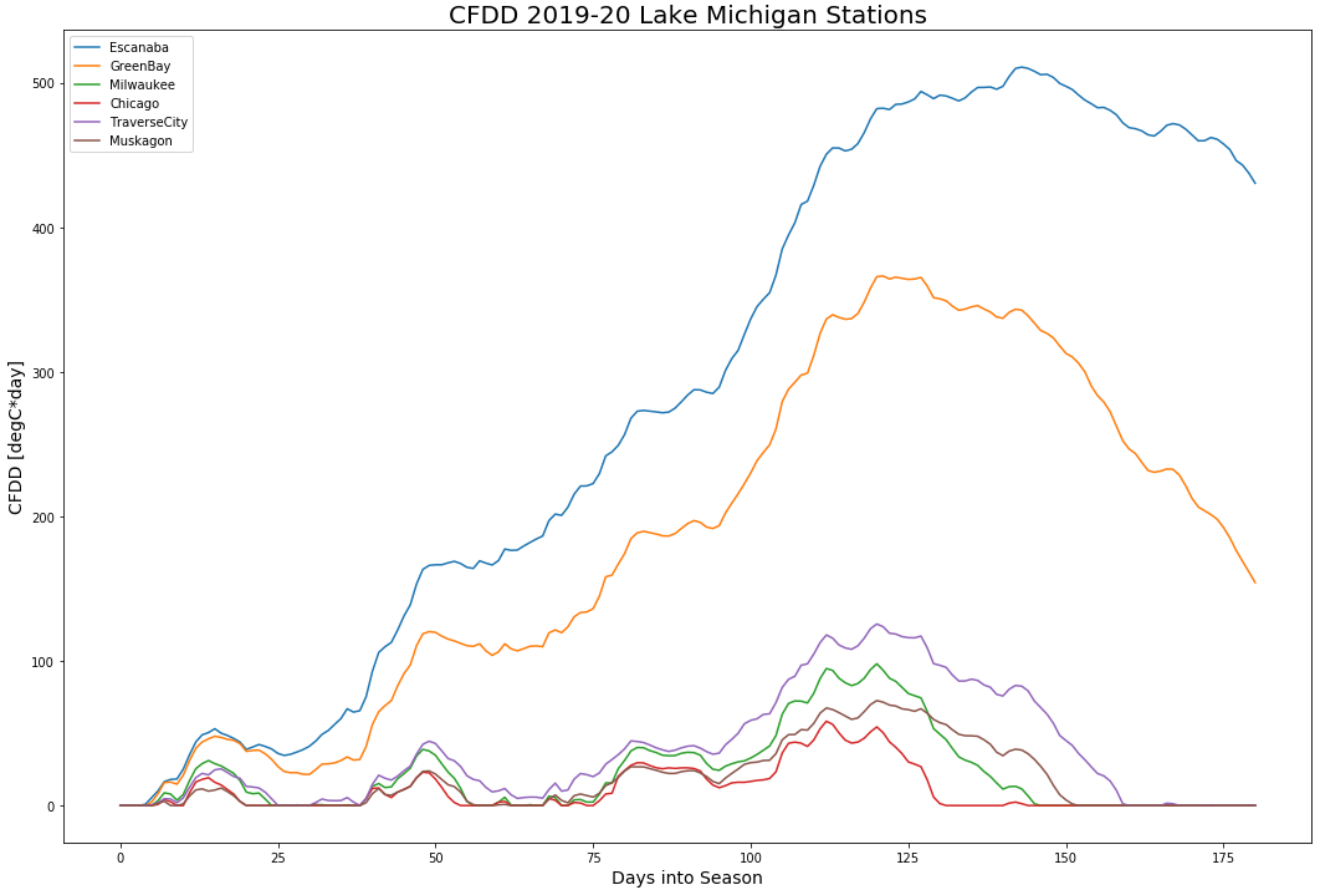

CFDD 2014-15 Lake Michigan Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season
0
200
400
600
800
0
25
50
75
100
125
150
175


CFDD 2015-16 Lake Michigan Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
300
350
0
25
50
75
100
125
150
175

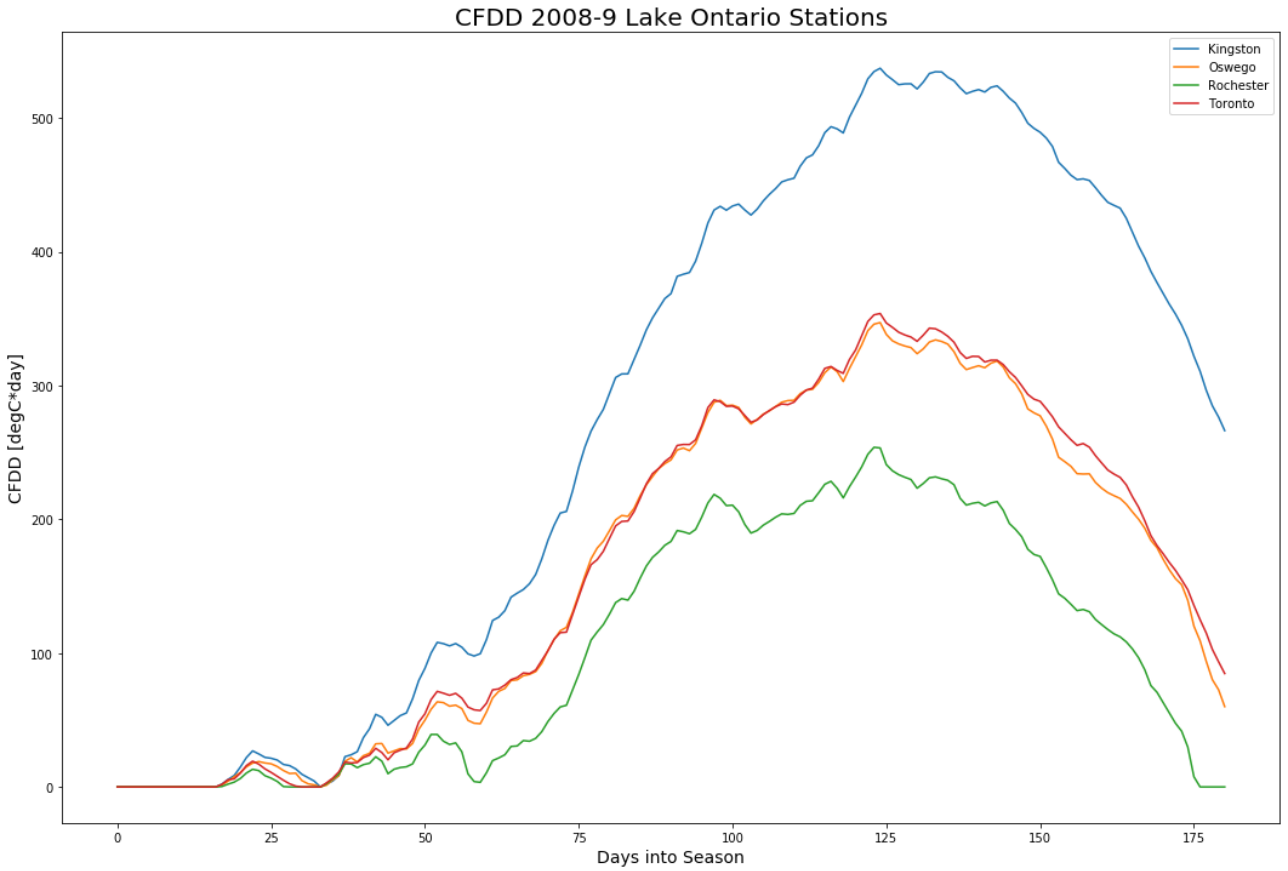

CFDD 2016-17 Lake Superior Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
0
25
50
75
100
125
150
175

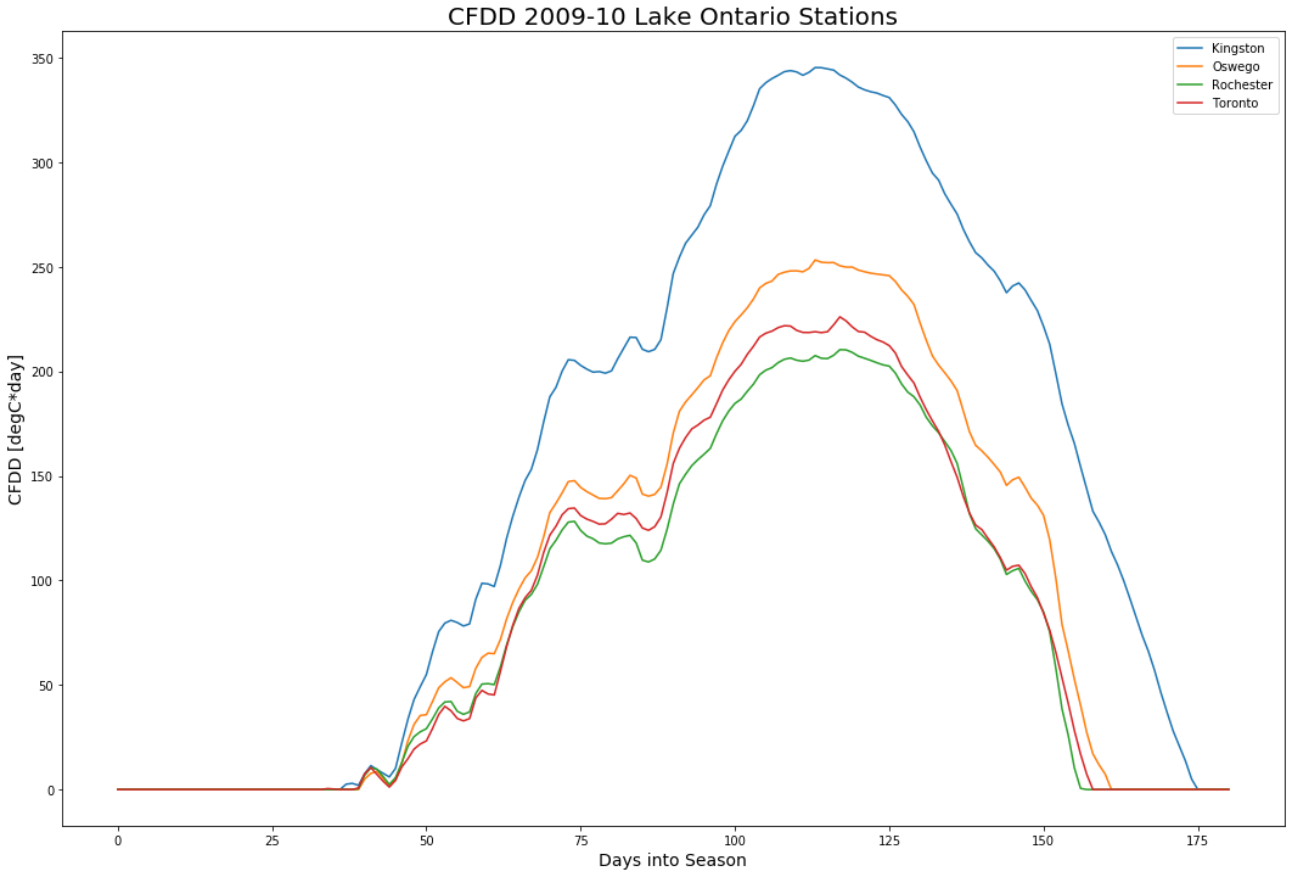

CFDD 2017-18 Lake Superior Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season
0
200
400
600
800
0
25
50
75
100
125
150
175

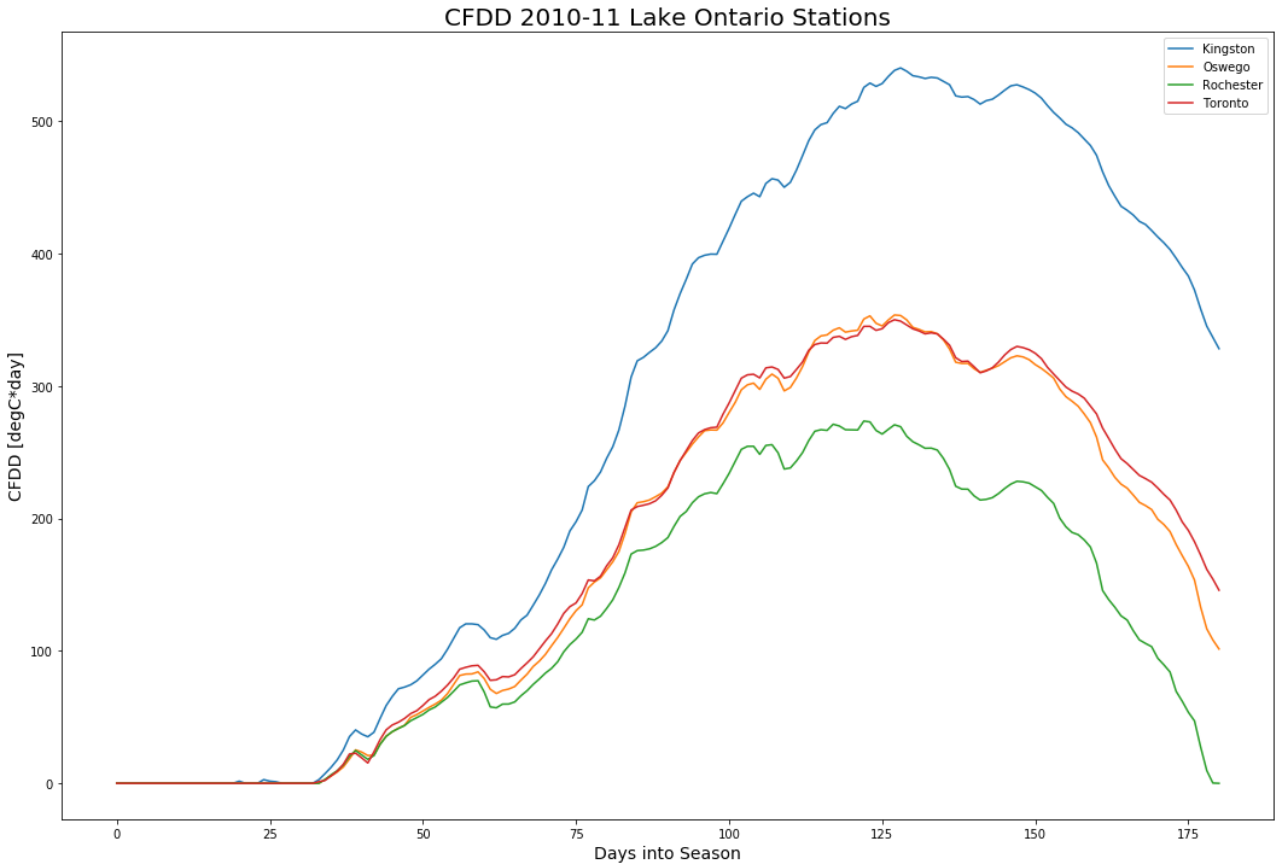

CFDD 2018-19 Lake Michigan Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season
0
200
400
600
800
0
25
50
75
100
125
150
175

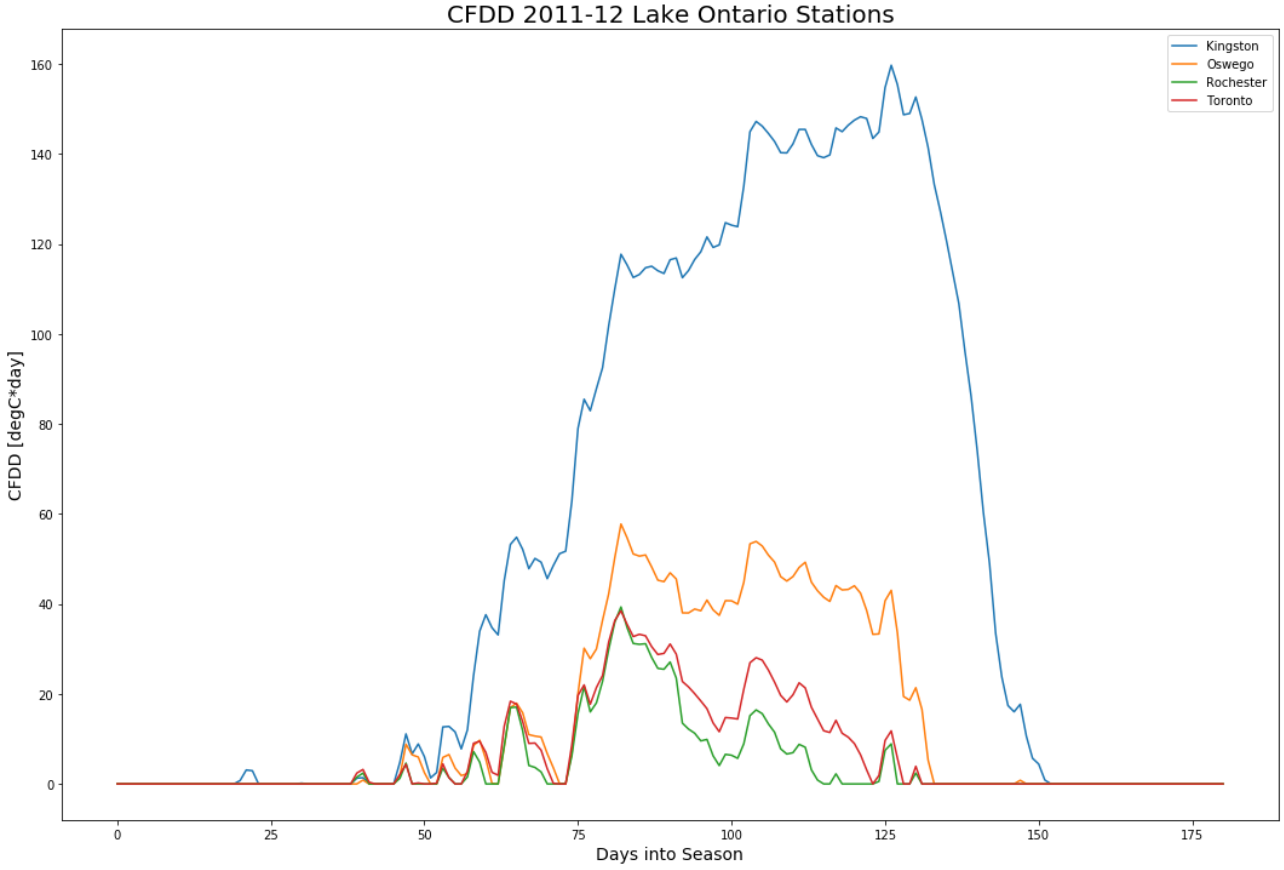

CFDD 2019-20 Lake Michigan Stations
Escanaba
GreenBay
Milwaukee
Chicago
TraverseCity
Muskagon
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
0
25
50
75
100
125
150
175

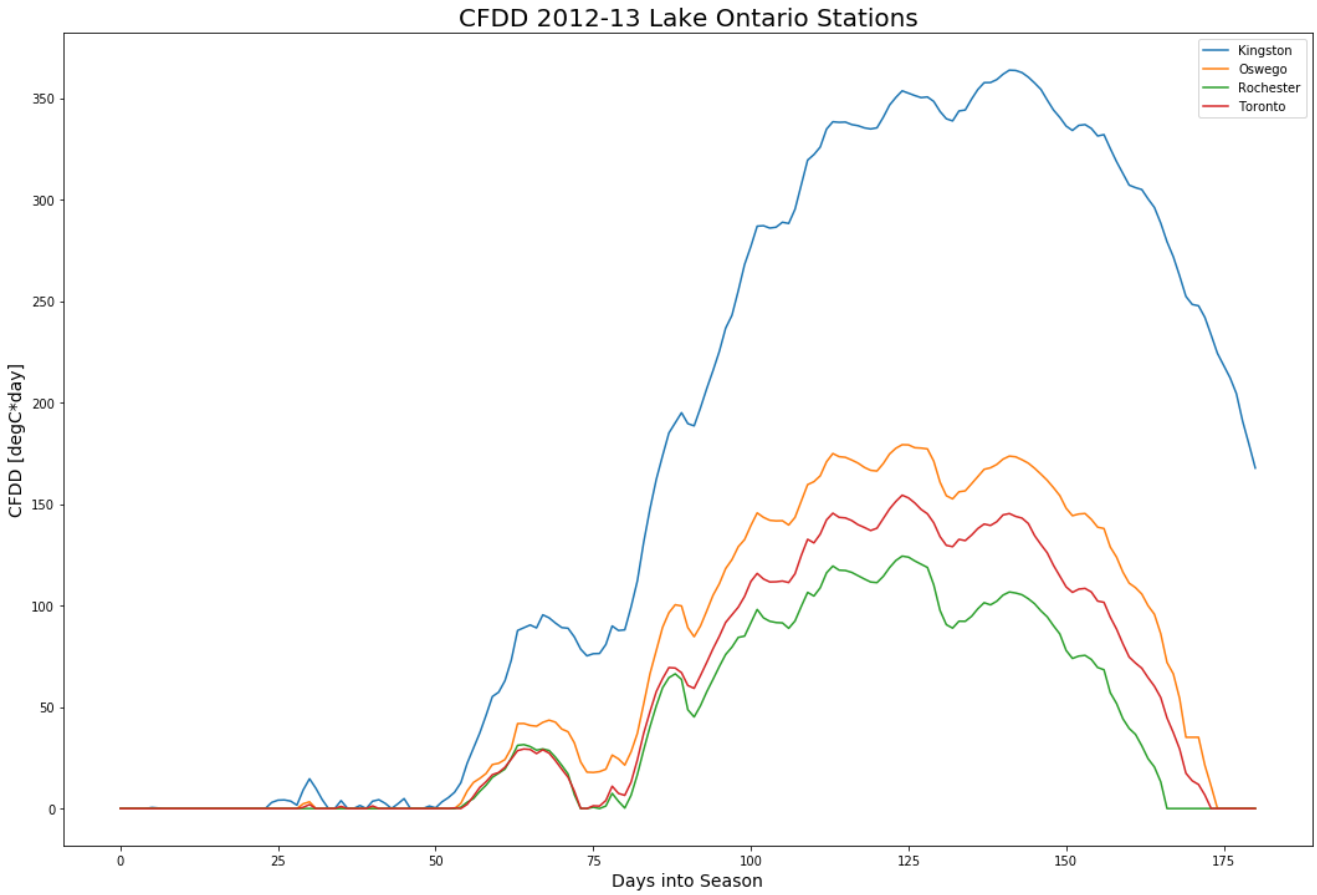
e. **Lake Ontario**

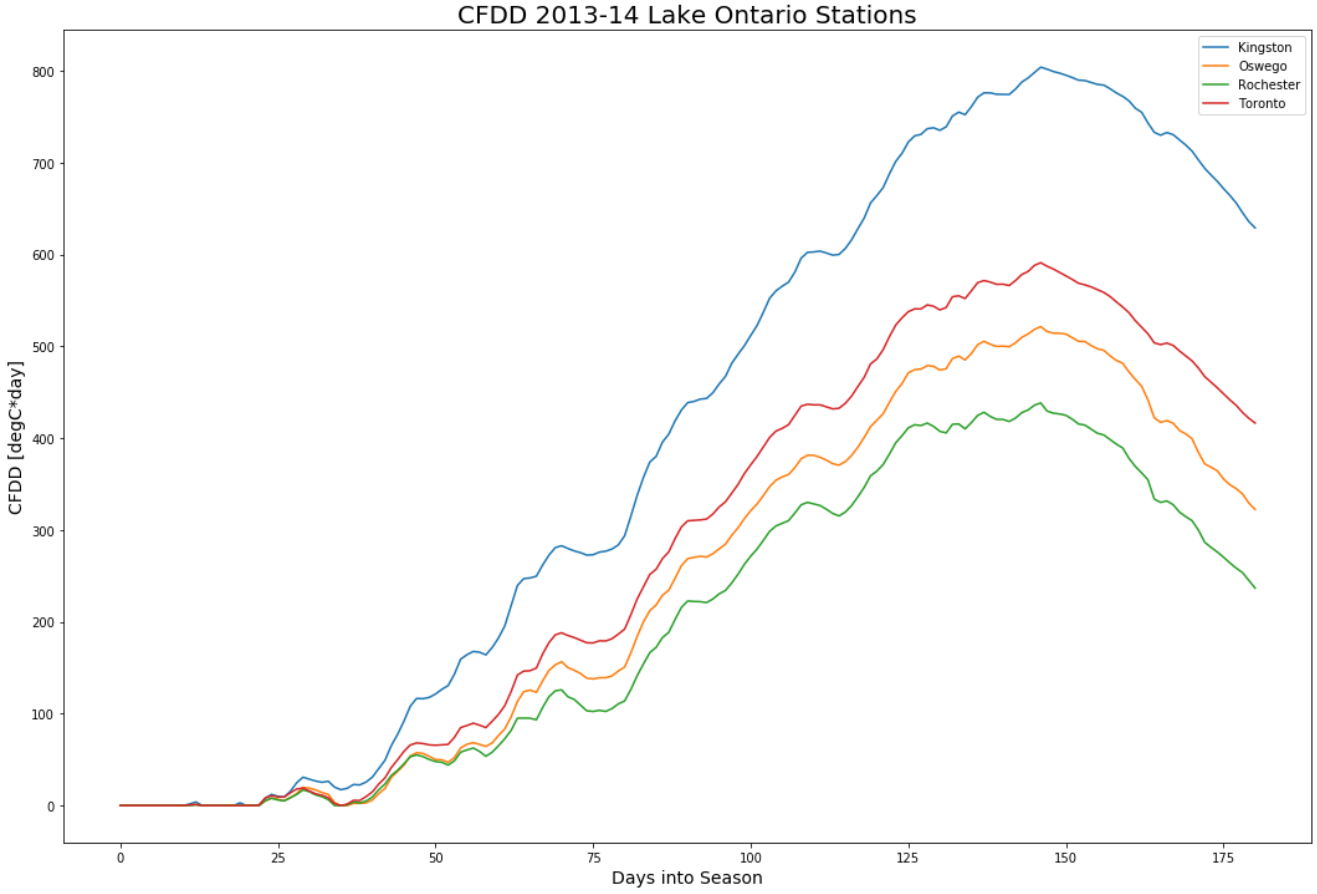


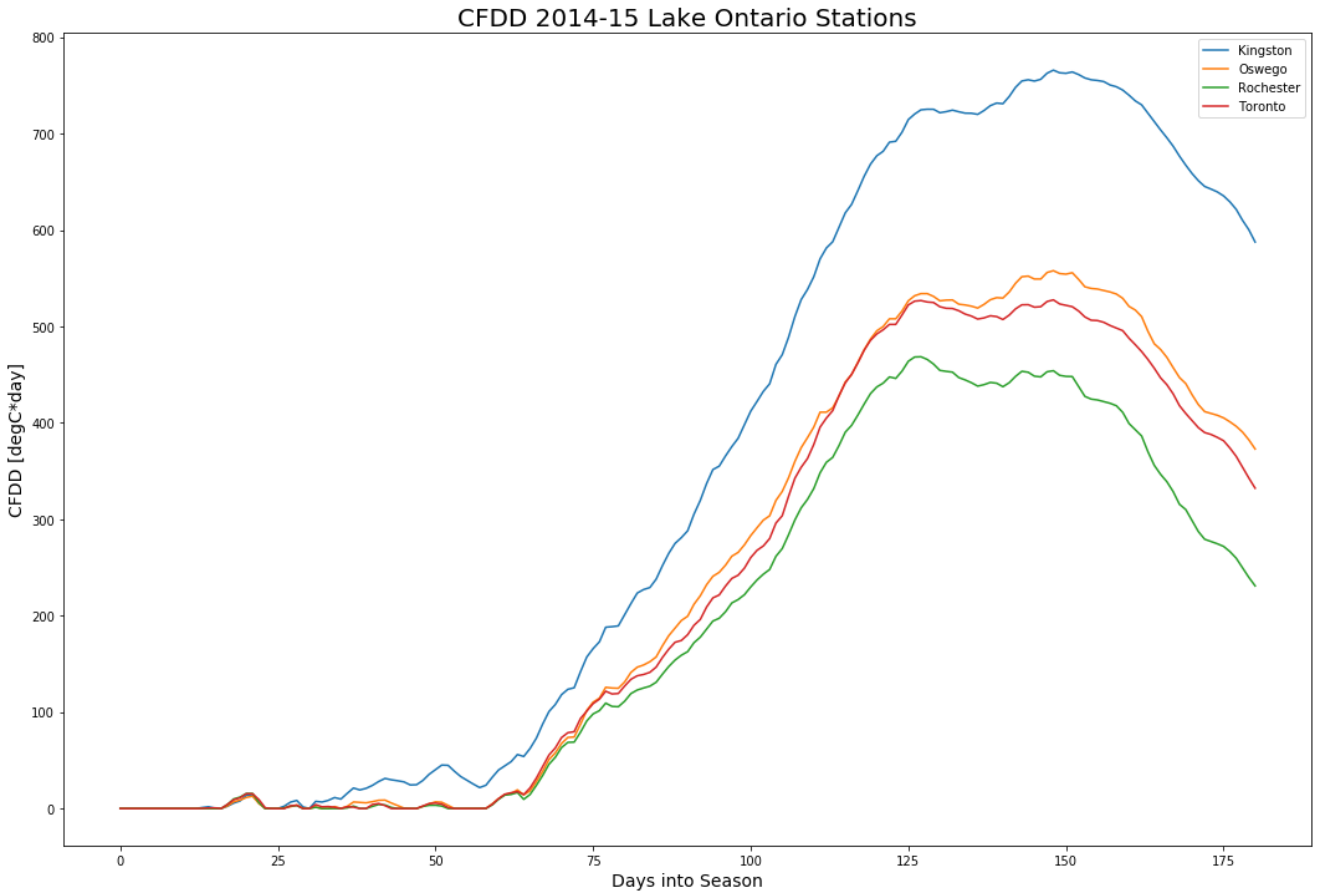


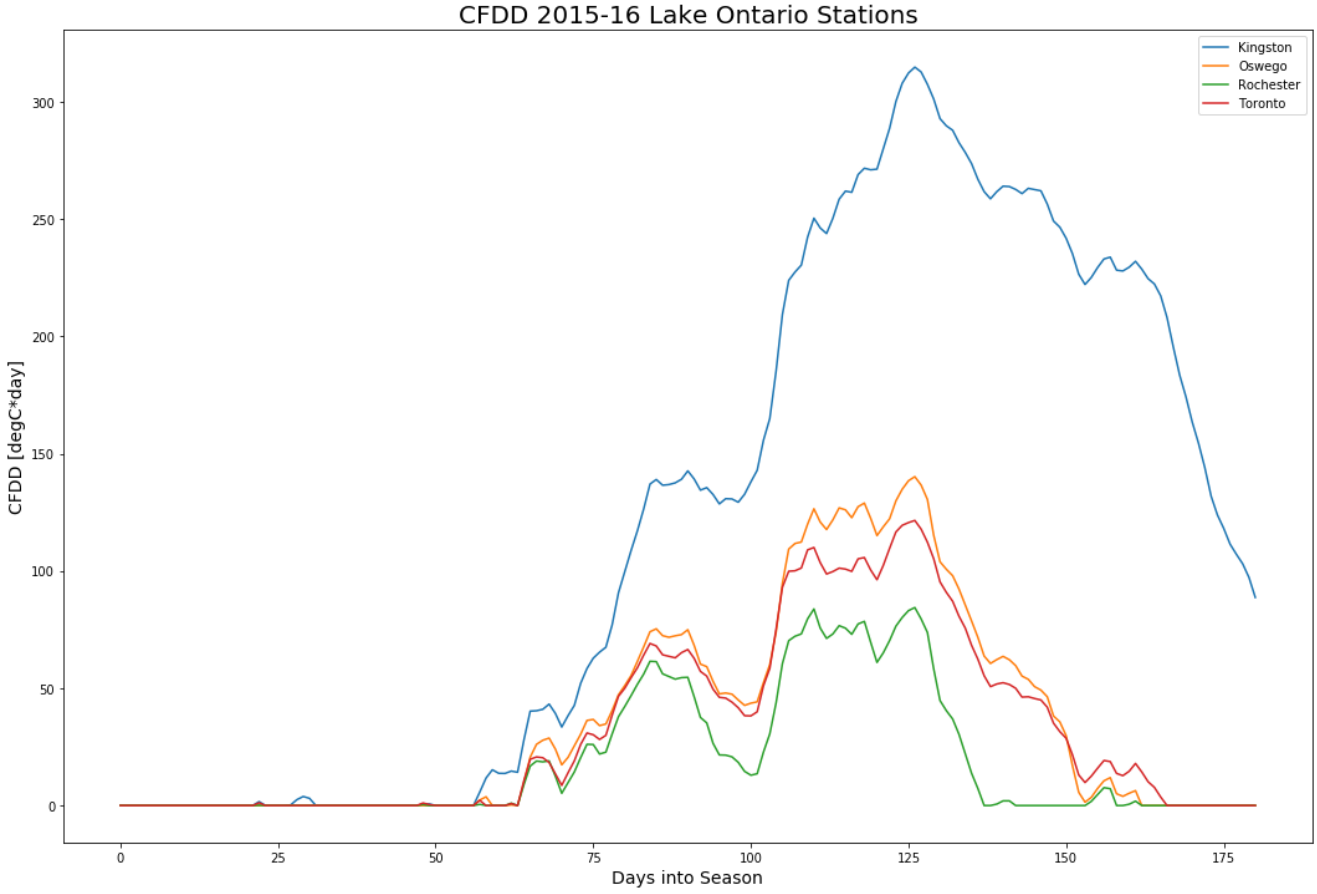

CFDD 2010-11 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
0
25
50
75
100
125
150
175

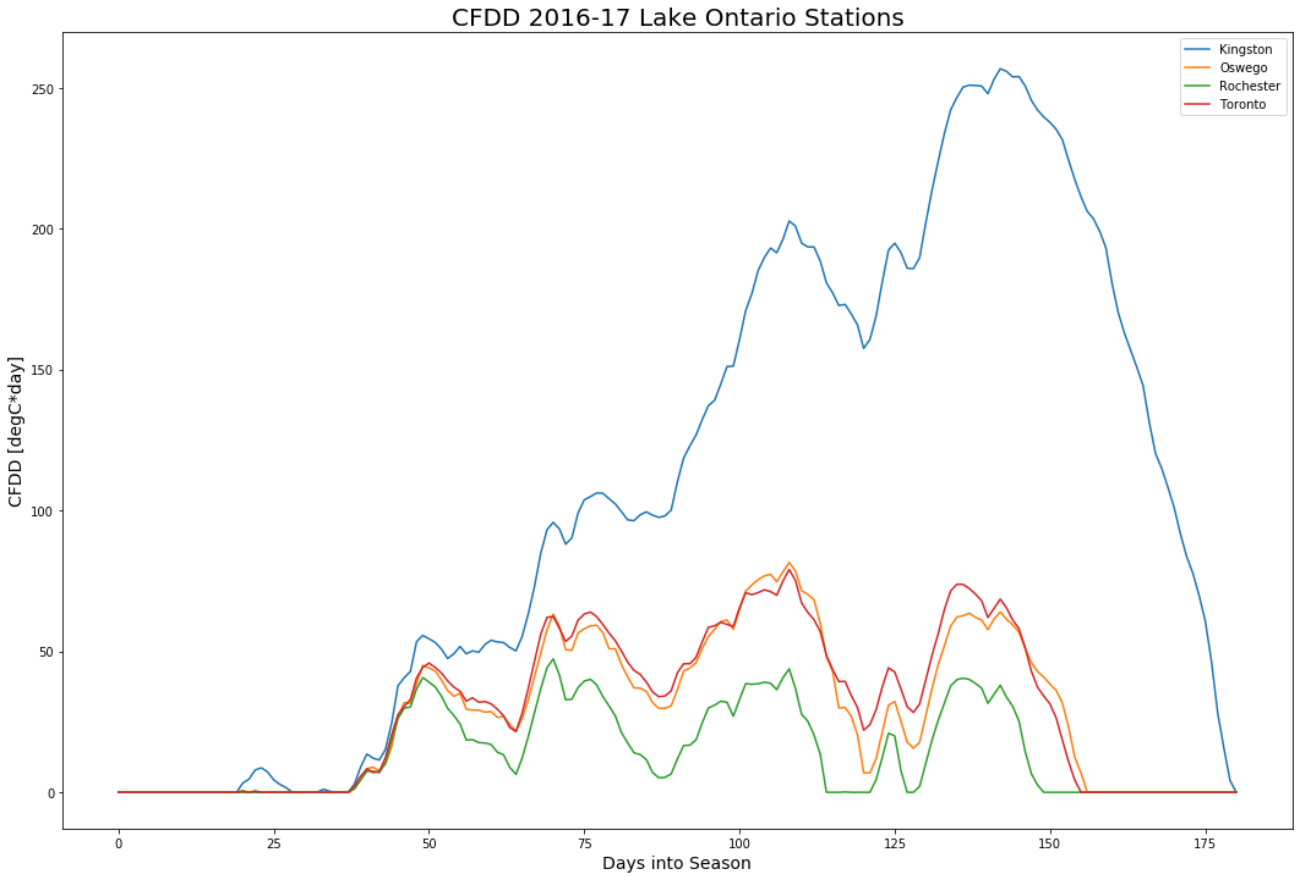

CFDD 2011-12 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
20
40
60
80
100
120
140
160
0
25
50
75
100
125
150
175

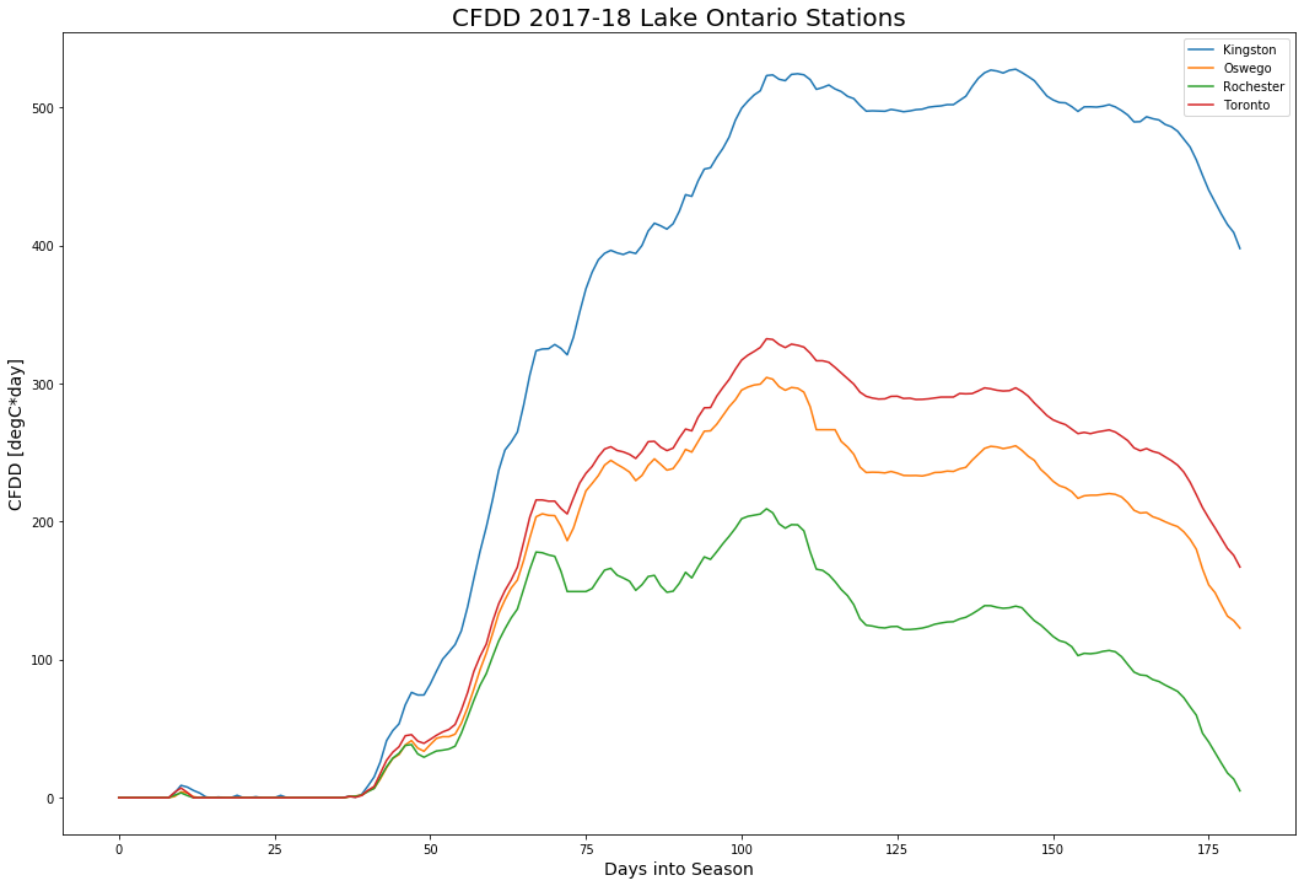

CFDD 2012-13 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
300
350
0
25
50
75
100
125
150
175

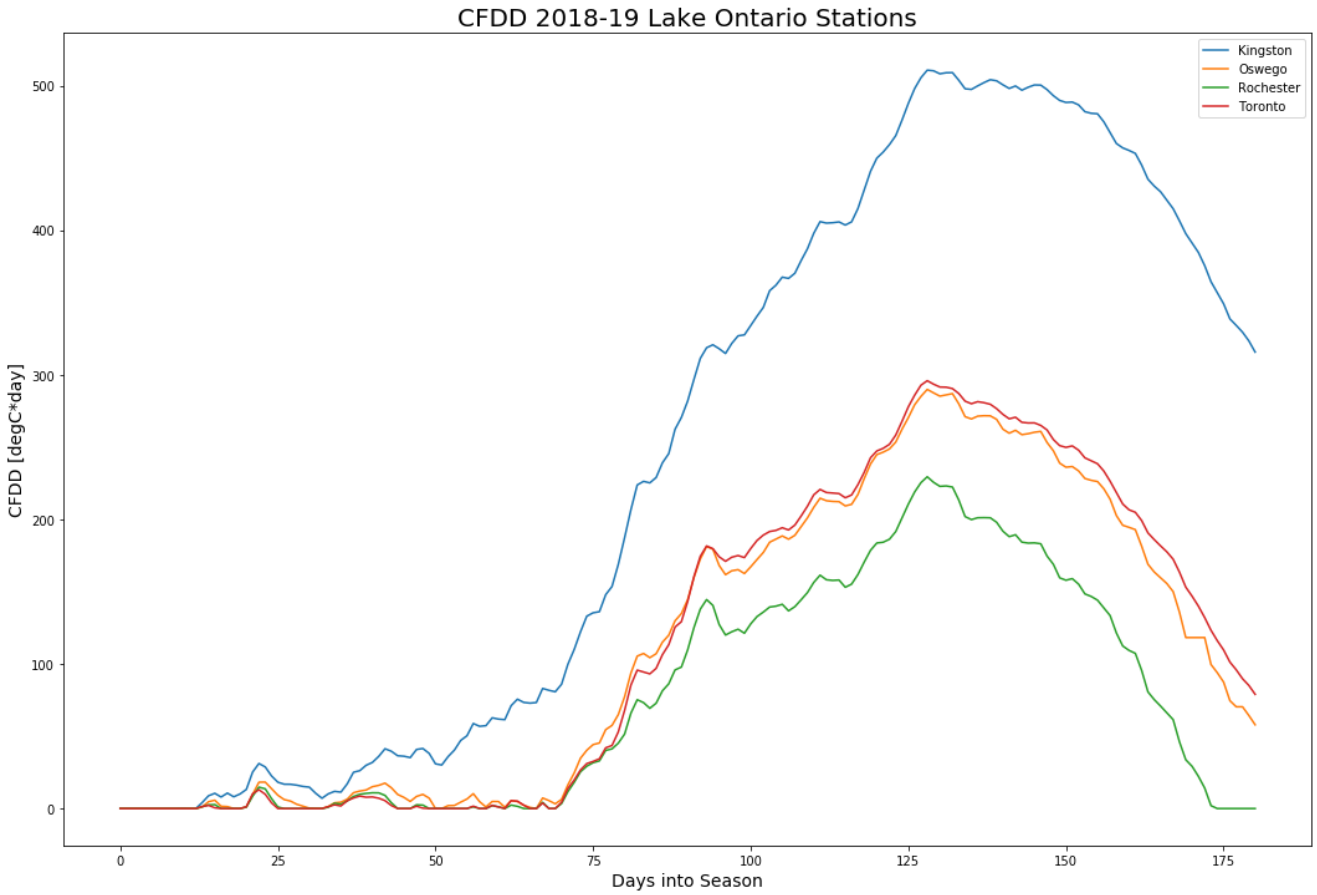

CFDD 2013-14 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
700
800
0
25
50
75
100
125
150
175

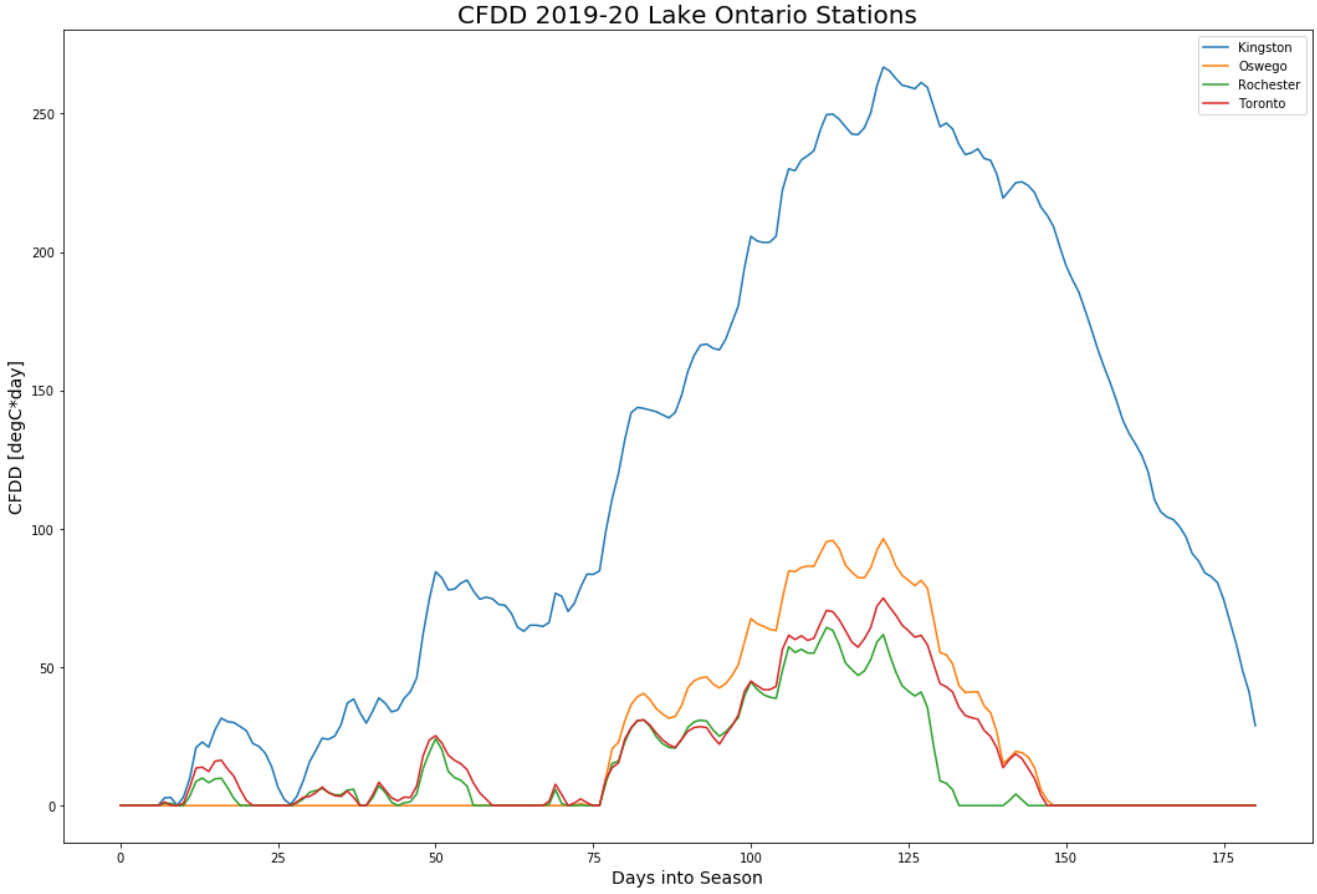

CFDD 2014-15 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
600
700
800
0
25
50
75
100
125
150
175


CFDD 2015-16 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
300
0
25
50
75
100
125
150
175


CFDD 2016-17 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
0
25
50
75
100
125
150
175


CFDD 2017-18 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
0
25
50
75
100
125
150
175


CFDD 2018-19 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
0
25
50
75
100
125
150
175


CFDD 2019-20 Lake Ontario Stations
Kingston
Oswego
Rochester
Toronto
CFDD [degC*day]
Days into Season
0
50
100
150
200
250
0
25
50
75
100
125
150
175

CFDD BY STATION

A. Lake Erie

BUF Sensor Height: 178m

	Avg.Temp	Max.CFDD	Missing
2008-2009	0.248592	370.102746	0.0
2009-2010	1.782958	283.152684	0.0
2010-2011	-0.129435	383.049905	1.0
2011-2012	3.845597	42.600996	0.0
2012-2013	1.535369	173.598869	2.0
2013-2014	-1.556760	546.541660	8.0
2014-2015	-1.521630	584.719497	1.0
2015-2016	3.258187	121.030154	2.0
2016-2017	2.900397	69.303365	0.0
2017-2018	-0.180312	320.677830	1.0
2018-2019	0.068451	293.426384	3.0
2019-2020	1.981461	63.232877	0.0

CLV Sensor Height: NA m

	Avg.Temp	Max.CFDD	Missing
2008-2009	2.048821	224.320748	0.0
2009-2010	3.011398	207.080608	0.0
2010-2011	1.513832	259.856621	0.0
2011-2012	5.441383	30.963156	1.0
2012-2013	2.793112	101.503097	0.0
2013-2014	0.312720	380.395457	8.0
2014-2015	0.308824	434.648454	1.0
2015-2016	4.701969	67.134712	4.0
2016-2017	4.632021	57.469433	0.0
2017-2018	1.456197	215.224138	1.0
2018-2019	2.455782	130.919276	9.0
2019-2020	3.410934	21.153965	2.0

ERI Sensor Height: 223m

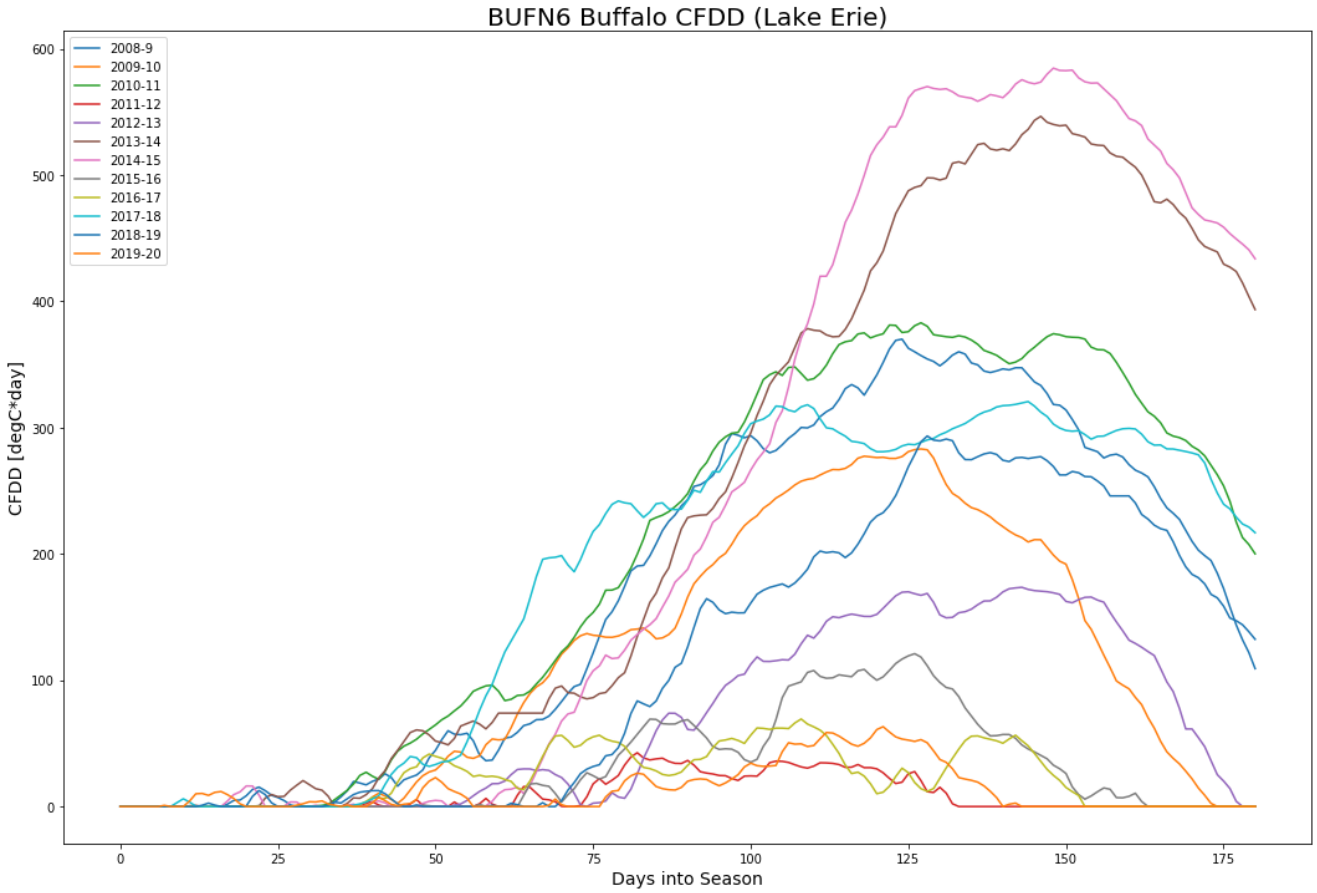
	Avg.Temp	Max.CFDD	Missing
2008-2009	1.795012	233.445329	0.0
2009-2010	2.515928	260.858857	0.0
2010-2011	0.956381	291.806458	0.0
2011-2012	5.025413	36.250000	0.0
2012-2013	2.397888	123.775538	0.0
2013-2014	-0.335230	481.667568	0.0
2014-2015	-0.149610	497.294931	0.0
2015-2016	4.545627	81.108974	0.0
2016-2017	4.392964	47.019615	0.0
2017-2018	1.023448	252.783961	0.0
2018-2019	1.884337	214.397493	0.0
2019-2020	3.429046	20.791667	0.0

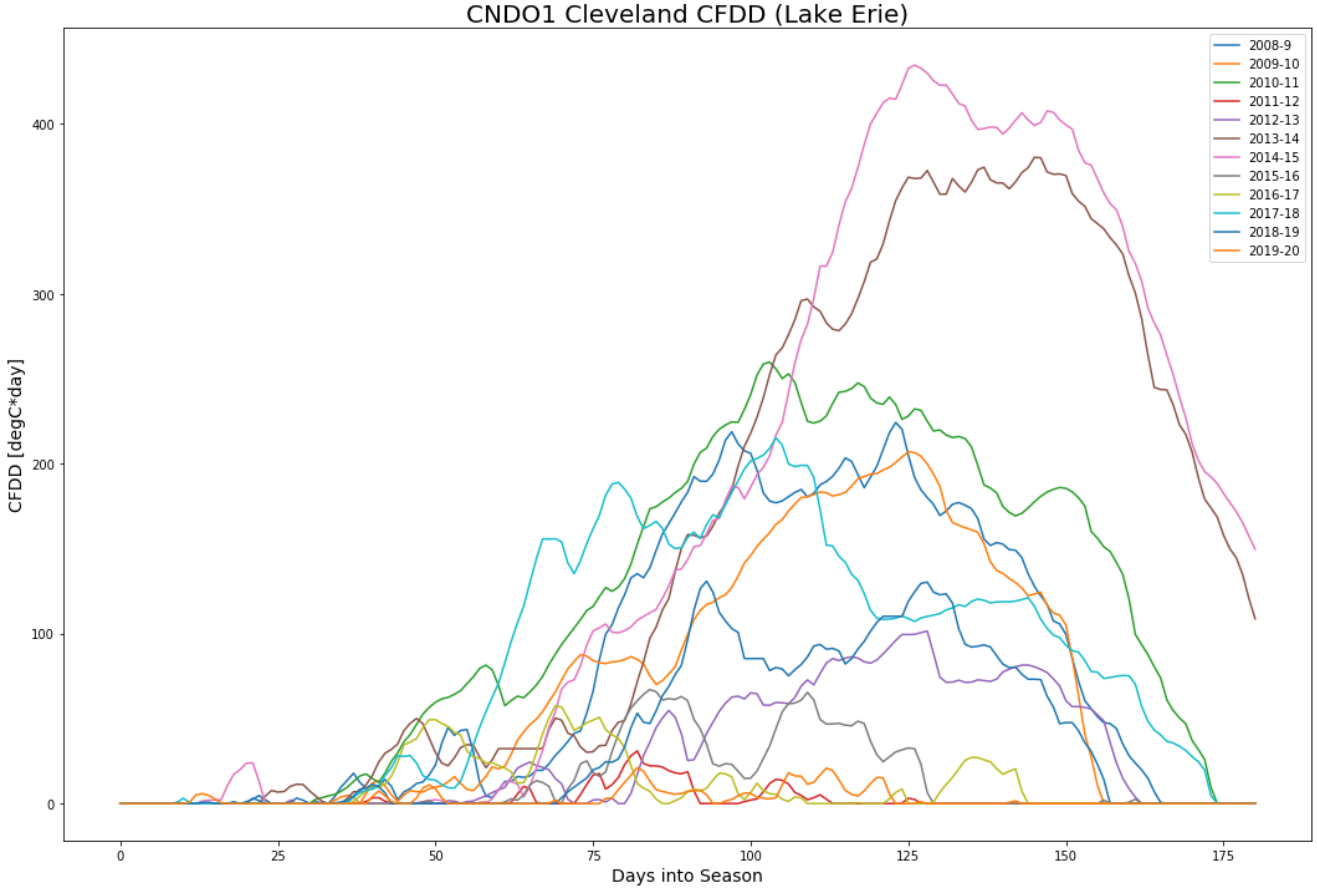
TOL Sensor Height: NA m

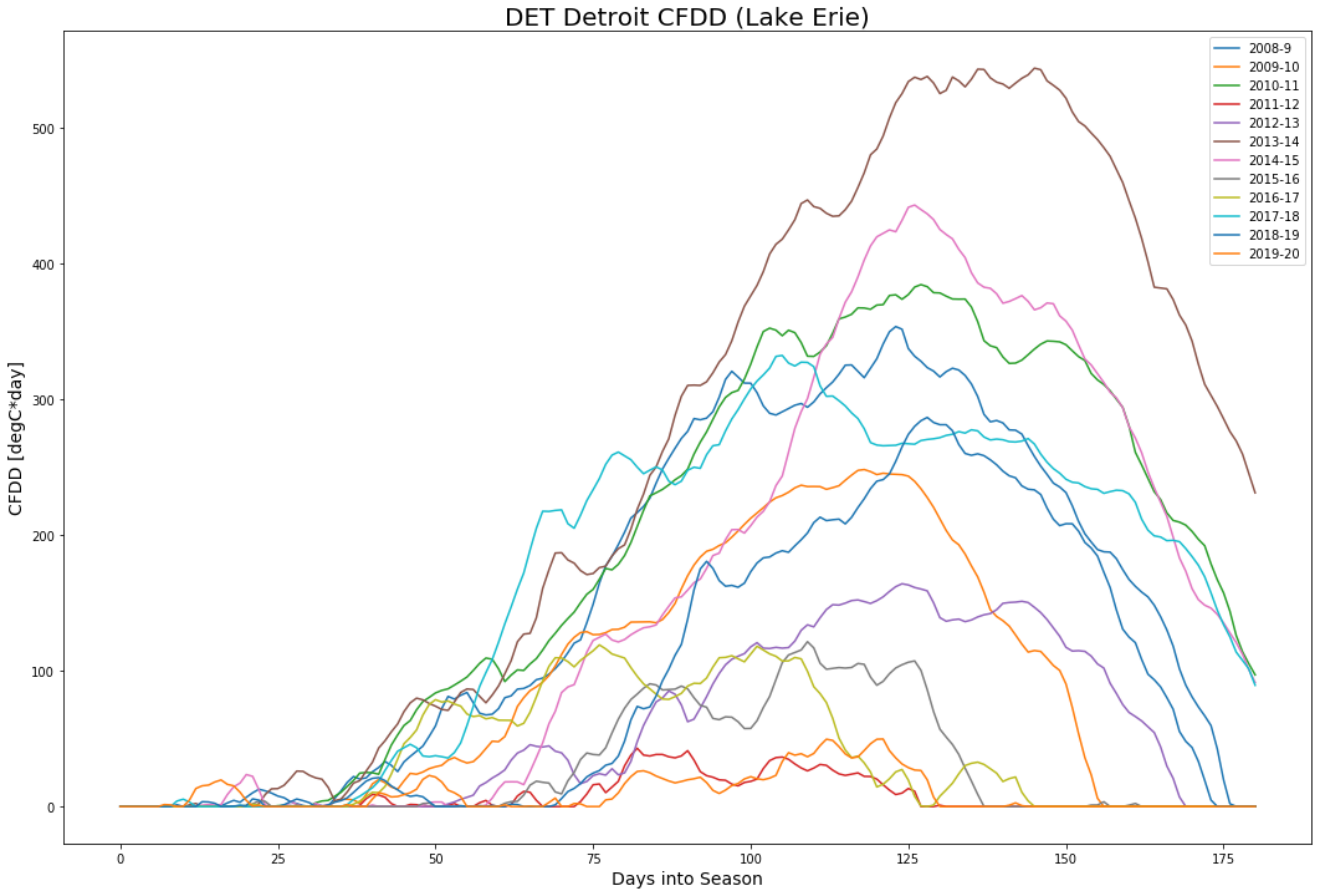
	Avg.Temp	Max.CFDD	Missing
2008-2009	1.194593	349.803118	0.0
2009-2010	3.069574	228.275505	0.0
2010-2011	0.986101	337.430529	0.0
2011-2012	5.137768	40.446411	0.0
2012-2013	2.202195	138.014964	0.0
2013-2014	-0.722849	558.402865	1.0
2014-2015	0.229007	459.255074	1.0
2015-2016	4.222674	100.835878	2.0
2016-2017	4.159300	101.017526	0.0
2017-2018	0.844056	287.580250	12.0
2018-2019	1.342836	249.803594	2.0
2019-2020	2.868742	41.951256	0.0

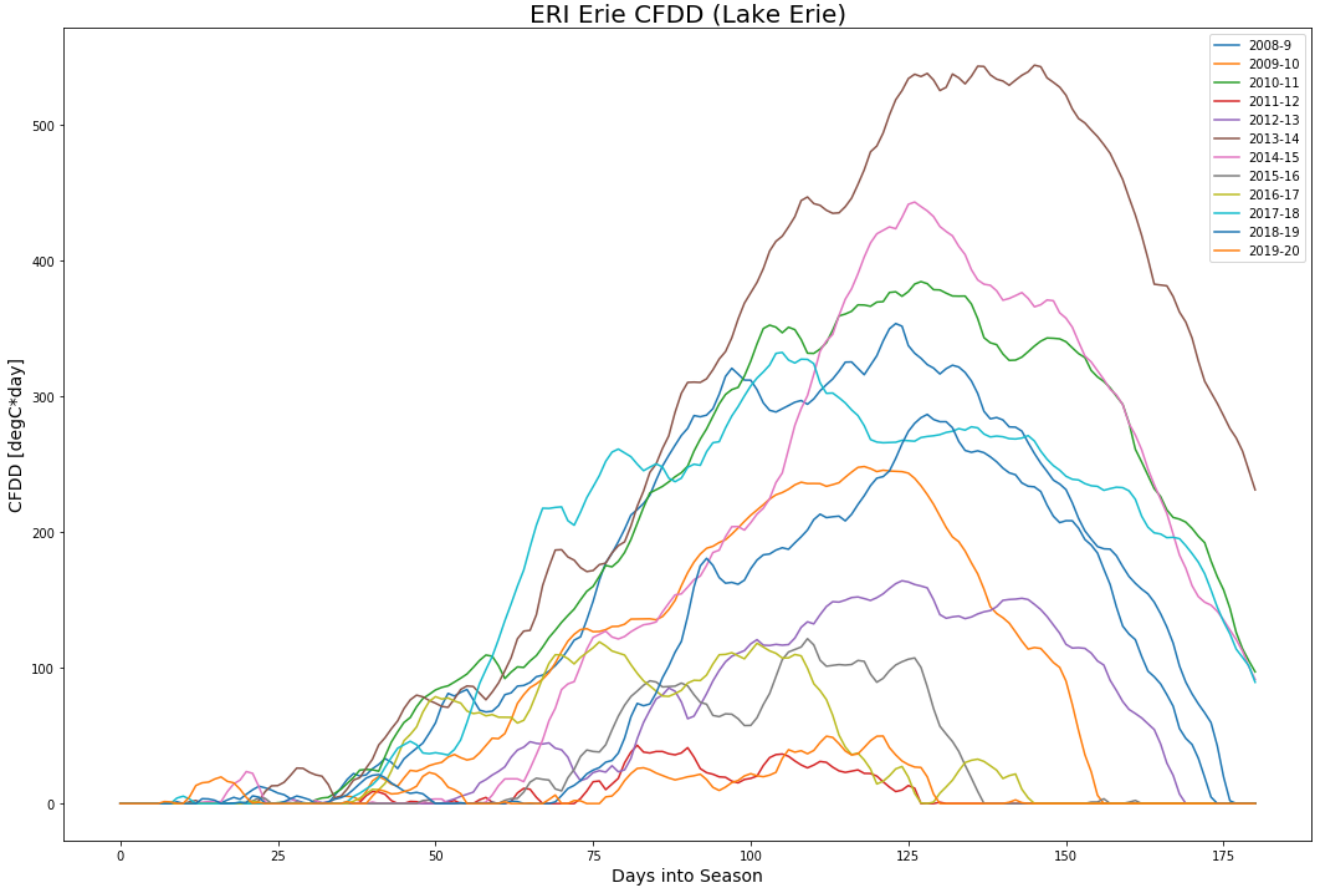
DET Sensor Height: 190m

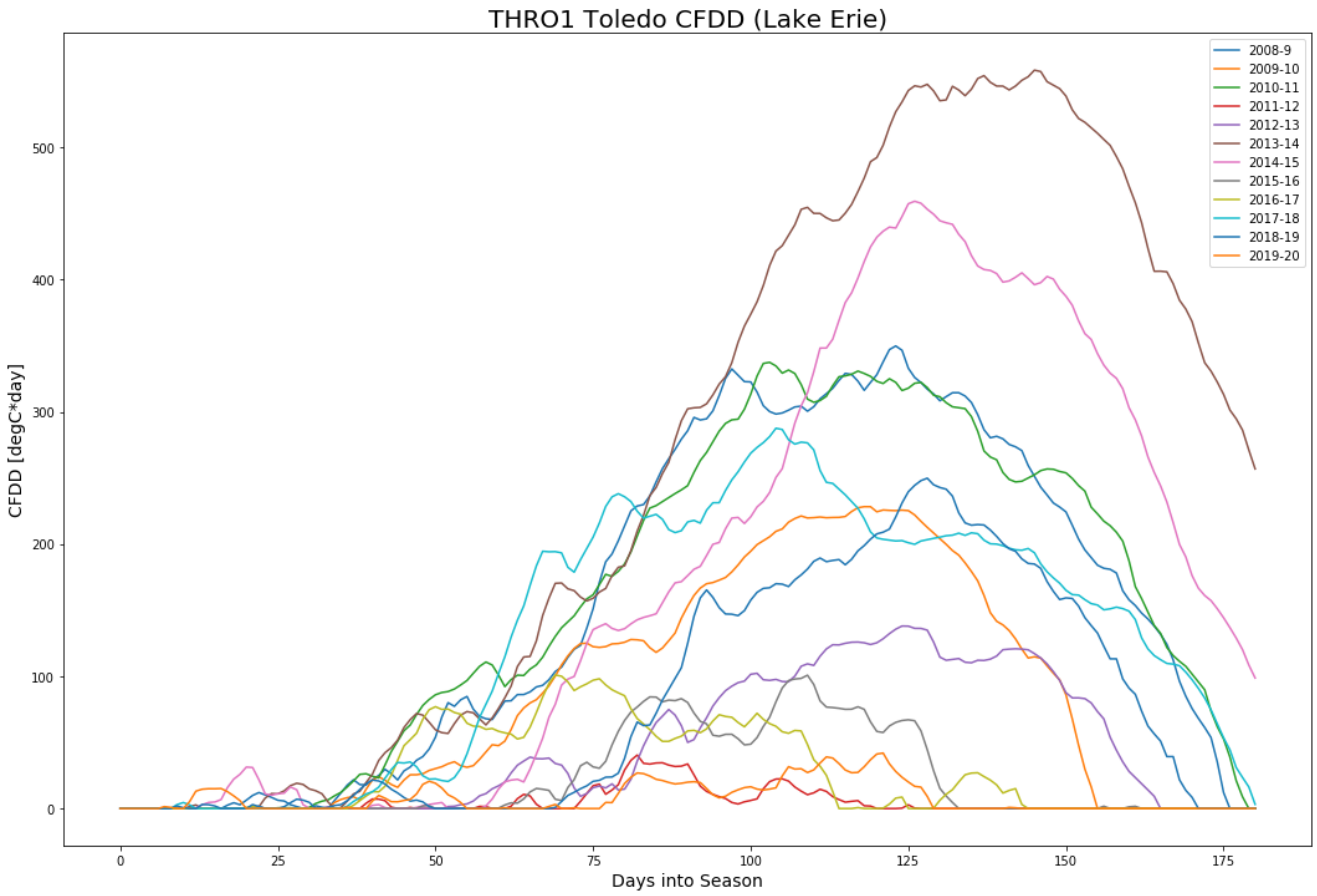
	Avg.Temp	Max.CFDD	Missing
2008-2009	1.027413	353.715682	0.0
2009-2010	2.851372	248.326608	1.0
2010-2011	0.339347	384.602819	0.0
2011-2012	4.699394	42.960145	0.0
2012-2013	1.773480	164.166831	0.0
2013-2014	-0.534059	544.139032	0.0
2014-2015	0.493412	443.244290	0.0
2015-2016	3.651708	121.478483	0.0
2016-2017	3.511172	119.116172	0.0
2017-2018	0.495258	332.433685	0.0
2018-2019	0.952036	286.757058	0.0
2019-2020	2.514111	49.740568	0.0


BUFN6 Buffalo CFDD (Lake Erie)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


CNDO1 Cleveland CFDD (Lake Erie)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


DET Detroit CFDD (Lake Erie)
CFDD [degC*day]
Days into Season
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20


ERI Erie CFDD (Lake Erie)
CFDD [degC*day]
Days into Season
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20


THRO1 Toledo CFDD (Lake Erie)
CFDD [degC*day]
Days into Season
0
100
200
300
400
500
0
25
50
75
100
125
150
175
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20

B. **Lake Superior**

FTW Sensor Height: 199 m

	Avg.Temp	Max.CFDD	Missing
2008-2009	-6.795001	1342.854169	2.0
2009-2010	-3.657142	1000.444529	0.0
2010-2011	-6.636331	1318.504239	0.0
2011-2012	-2.755730	740.733696	0.0
2012-2013	-5.518797	1058.673623	0.0
2013-2014	-10.322941	1933.744008	0.0
2014-2015	-7.740228	1519.461367	0.0
2015-2016	-4.054415	903.836959	0.0
2016-2017	-4.183734	956.823129	0.0
2017-2018	-7.907900	1489.934997	0.0
2018-2019	-7.181557	1391.879472	0.0
2019-2020	-5.544016	1048.758302	0.0

HOU Sensor Height: 333m

	Avg.Temp	Max.CFDD	Missing
2008-2009	-5.183243	1086.136749	0.0
2009-2010	-1.234691	638.161086	0.0
2010-2011	-3.833799	851.565472	0.0
2011-2012	-0.737128	452.771225	0.0
2012-2013	-4.010262	818.772945	0.0
2013-2014	-7.300873	1392.805609	1.0
2014-2015	-5.262083	1086.346504	0.0
2015-2016	-1.470989	507.070817	1.0
2016-2017	-1.569038	587.226950	0.0
2017-2018	-4.944479	961.130685	0.0
2018-2019	-4.566046	939.045091	0.0
2019-2020	-2.878967	590.785549	0.0

DUL Sensor Height: 185m

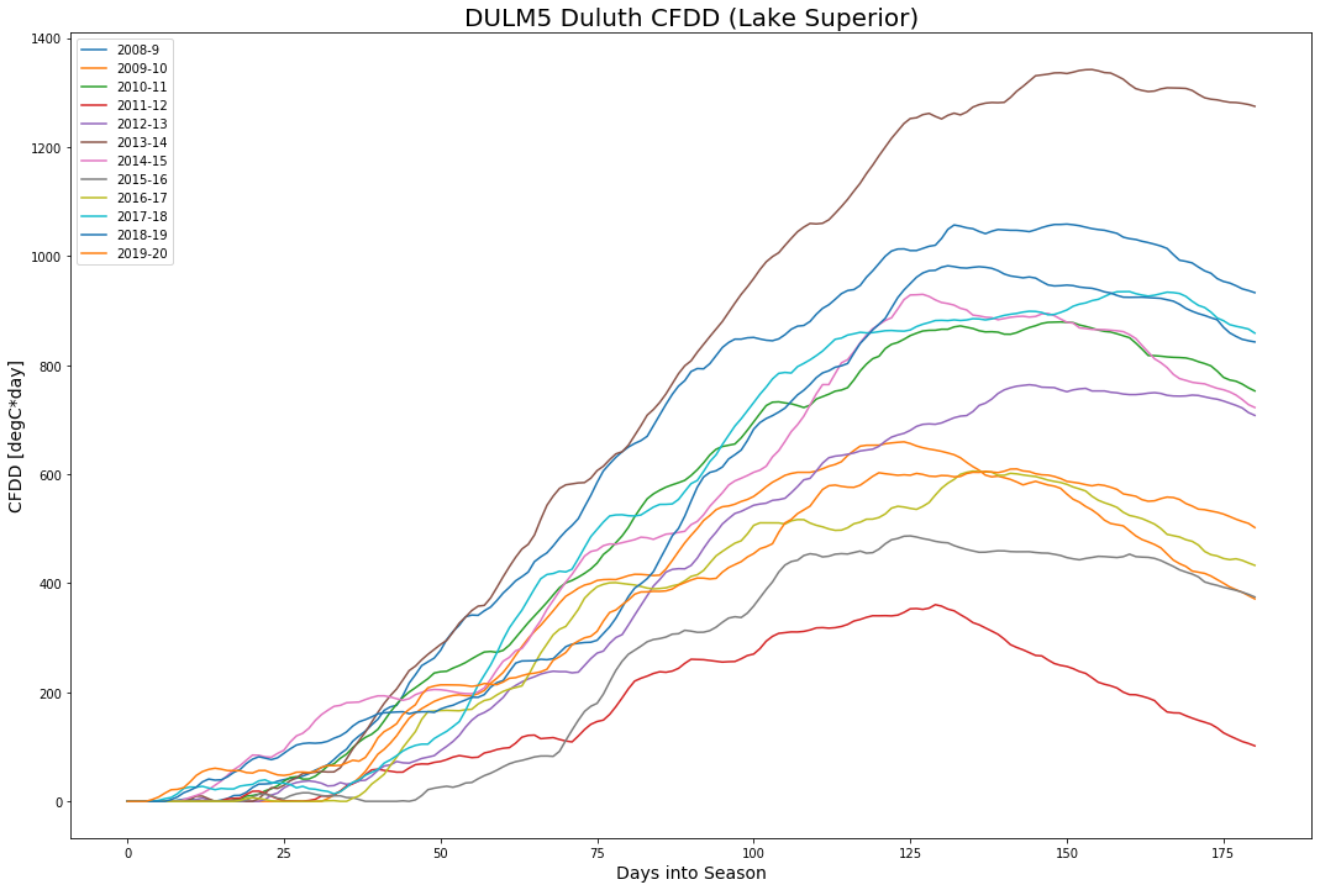
	Avg.Temp	Max.CFDD	Missing
2008-2009	-4.751956	1058.921121	0.0
2009-2010	-1.136175	659.478032	0.0
2010-2011	-3.696224	879.457737	0.0
2011-2012	-0.102233	360.793214	0.0
2012-2013	-3.513843	764.316110	0.0
2013-2014	-6.811946	1342.585571	1.0
2014-2015	-3.832183	930.104414	1.0
2015-2016	-1.248062	486.934506	3.0
2016-2017	-1.348378	605.835211	0.0
2017-2018	-4.738567	935.161795	1.0
2018-2019	-4.513280	982.479427	0.0
2019-2020	-2.749226	609.779706	0.0

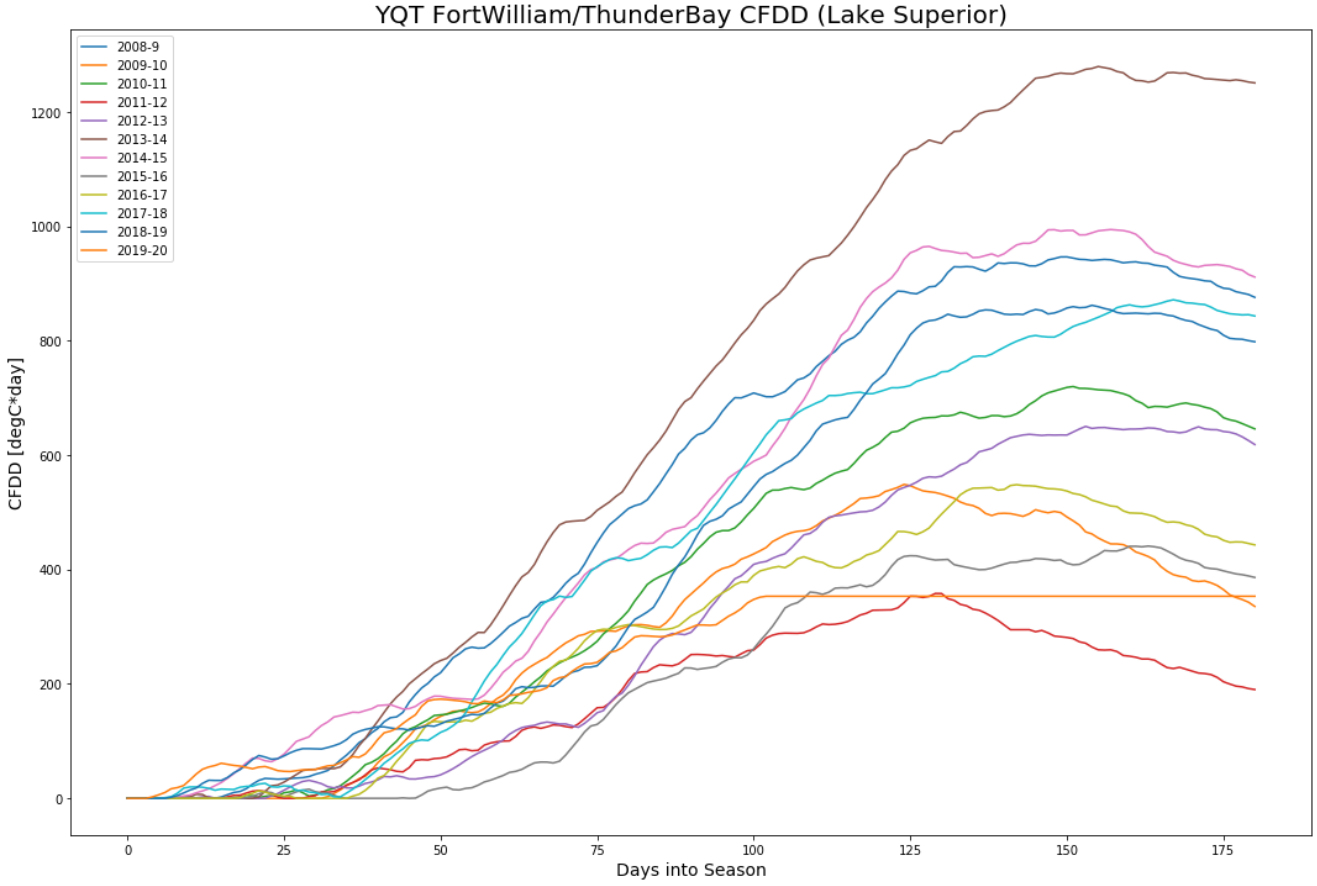
MQT Sensor Height: 188m

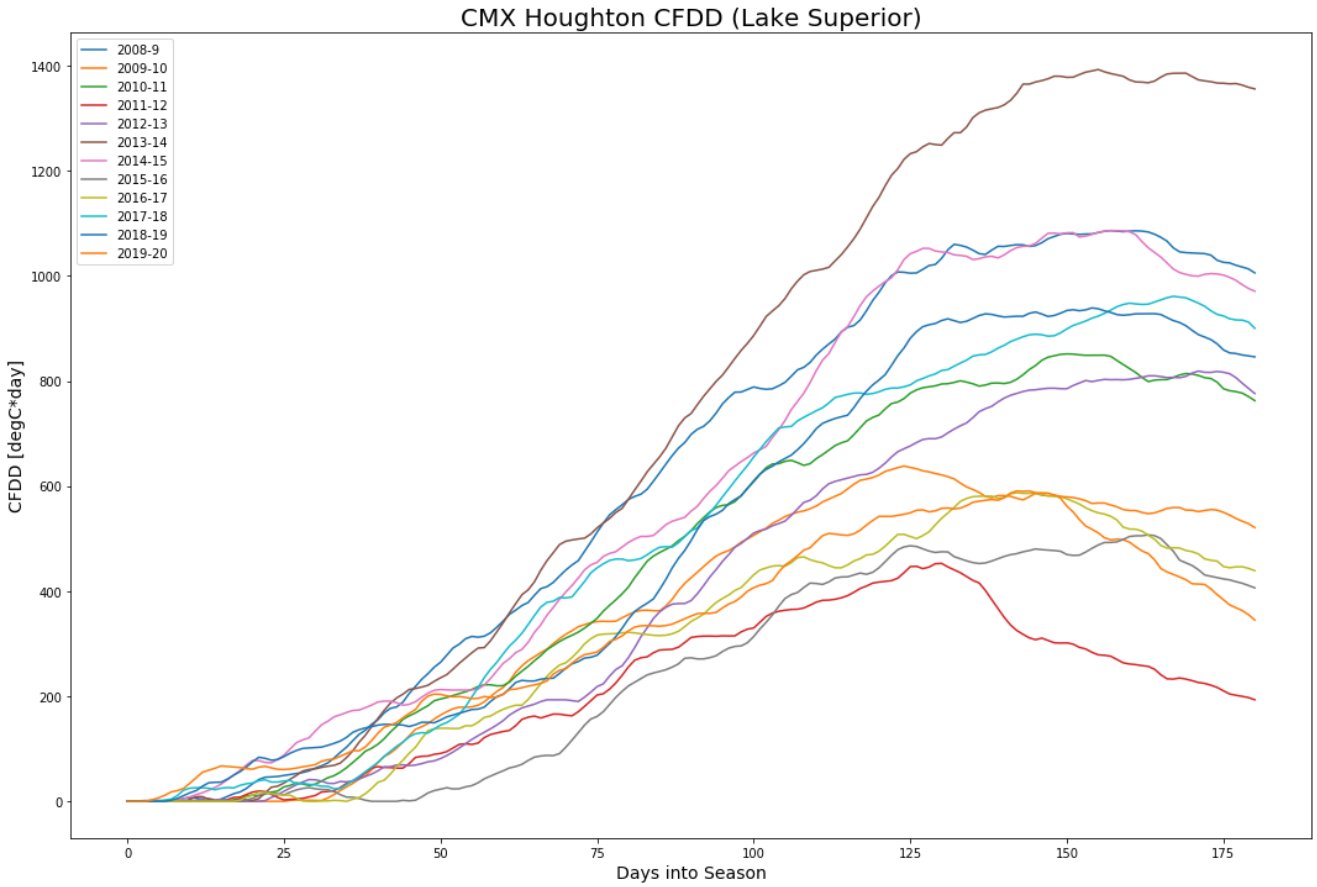
	Avg.Temp	Max.CFDD	Missing
2008-2009	-4.097038	897.575481	0.0
2009-2010	-0.741127	515.437448	0.0
2010-2011	-2.914083	694.751323	0.0
2011-2012	-0.241594	348.671787	3.0
2012-2013	-2.883796	648.921059	0.0
2013-2014	-6.736080	1289.650490	1.0
2014-2015	-4.856609	987.592171	1.0
2015-2016	-1.106429	428.715755	3.0
2016-2017	-1.340550	541.562266	0.0
2017-2018	-4.593346	873.557091	1.0
2018-2019	-4.123614	842.720522	0.0
2019-2020	-2.654283	513.708098	0.0

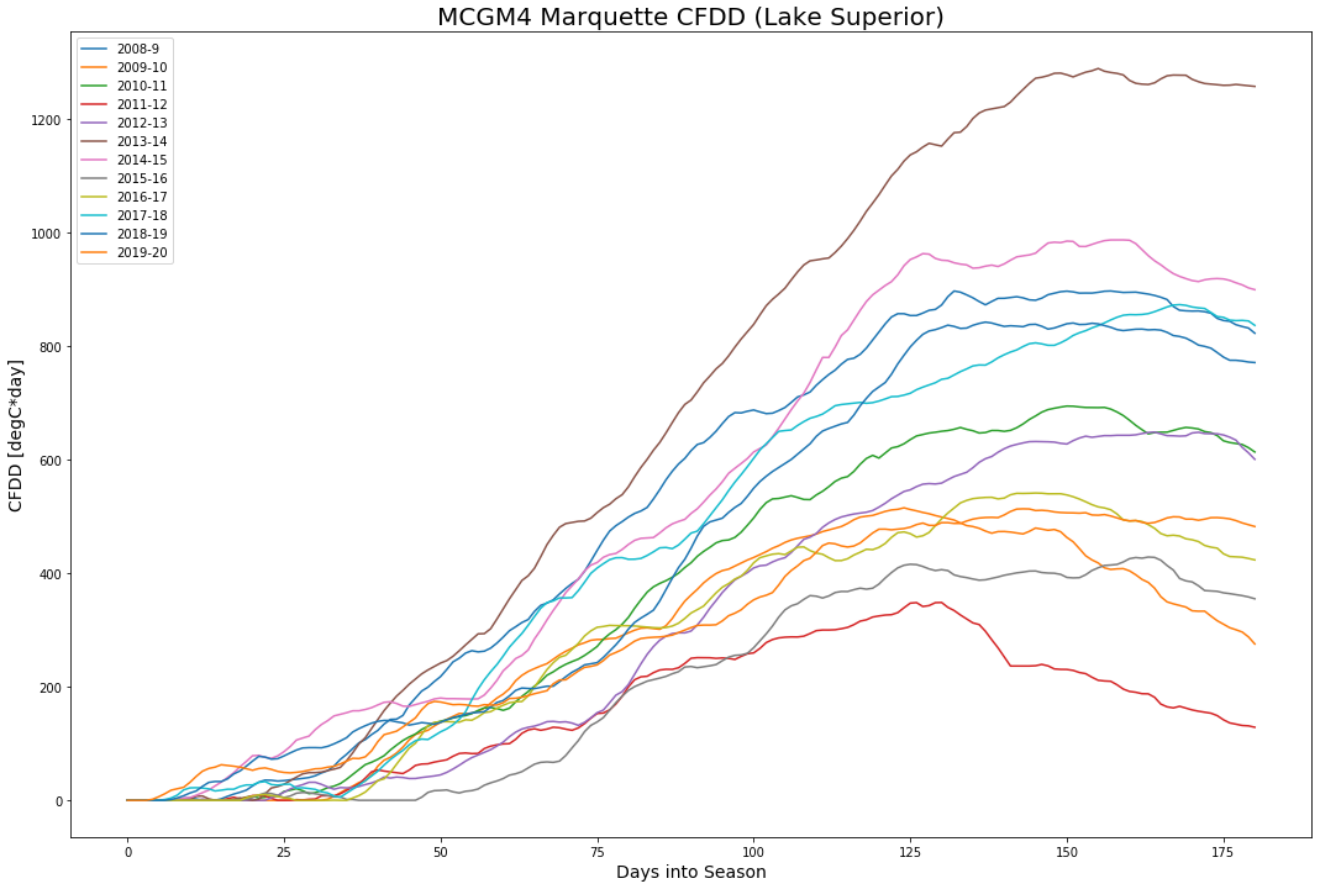
SSM Sensor Height: 186m

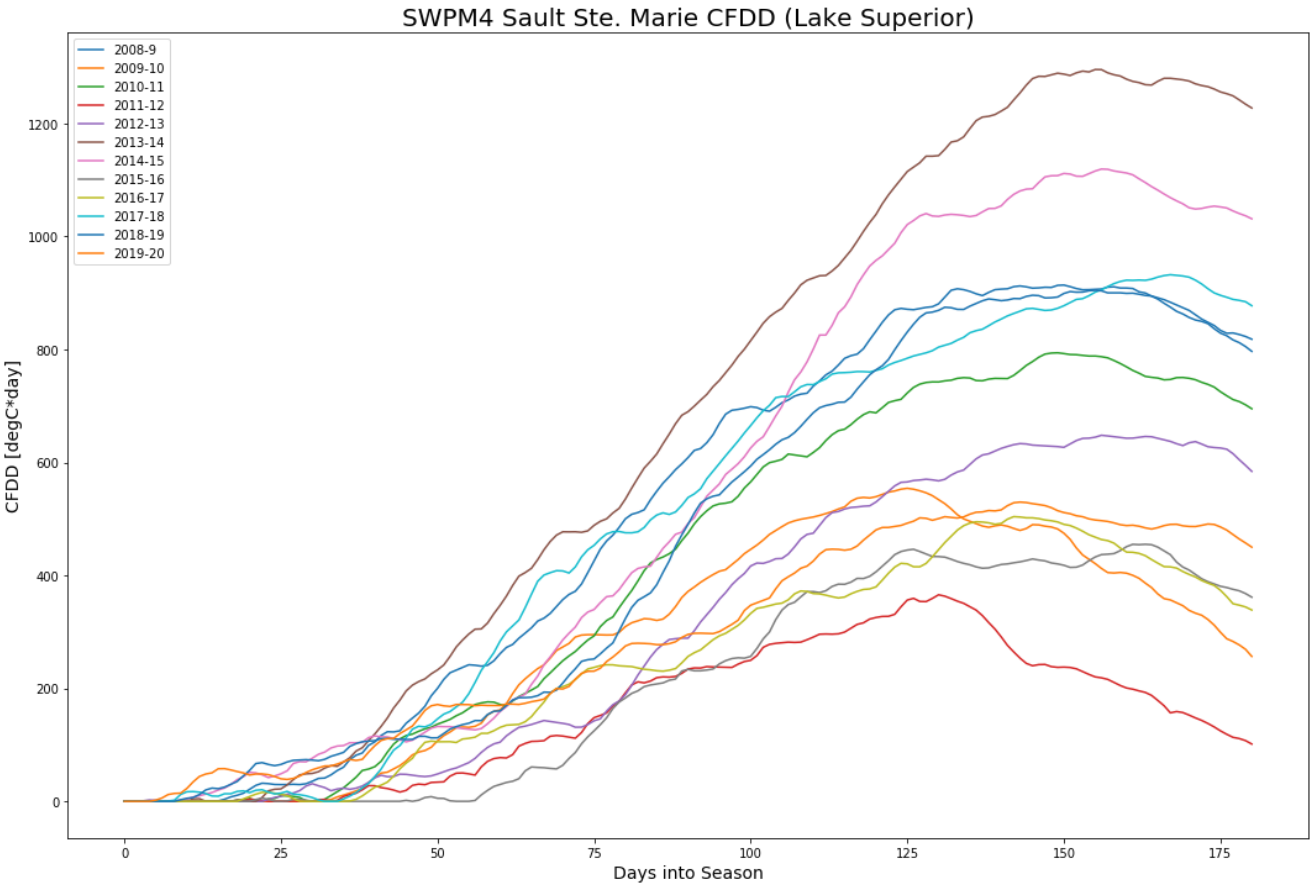
	Avg.Temp	Max.CFDD	Missing
2008-2009	-3.892146	914.065454	0.0
2009-2010	-0.574795	554.207803	0.0
2010-2011	-3.375236	794.108245	0.0
2011-2012	0.118515	365.700434	0.0
2012-2013	-2.757634	648.093539	0.0
2013-2014	-6.528079	1295.718624	1.0
2014-2015	-5.598225	1119.456538	1.0
2015-2016	-0.909927	455.076908	3.0
2016-2017	-0.902851	504.080994	0.0
2017-2018	-4.685896	932.403984	1.0
2018-2019	-4.398129	904.762498	2.0
2019-2020	-2.437095	529.814727	0.0


DULM5 Duluth CFDD (Lake Superior)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


YQT FortWilliam/ThunderBay CFDD (Lake Superior)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


CMX Houghton CFDD (Lake Superior)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


MCGM4 Marquette CFDD (Lake Superior)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


SWPM4 Sault Ste. Marie CFDD (Lake Superior)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
0
200
400
600
800
1000
1200
0
25
50
75
100
125
150
175
Days into Season

C. **Lake Huron**

ALP Sensor Height: 179m

	Avg.Temp	Max.CFDD	Missing
2008-2009	-2.472989	686.205898	0.0
2009-2010	0.468178	398.450432	0.0
2010-2011	-1.754066	560.702598	0.0
2011-2012	1.562104	190.885682	0.0
2012-2013	-0.885928	380.475477	0.0
2013-2014	-4.442292	978.390224	1.0
2014-2015	-3.127038	750.180983	1.0
2015-2016	0.459222	291.293662	6.0
2016-2017	0.548174	313.142240	0.0
2017-2018	-2.471747	573.236203	1.0
2018-2019	-2.185152	572.058077	0.0
2019-2020	-0.451124	235.119803	0.0

PTH Sensor Height: 198m

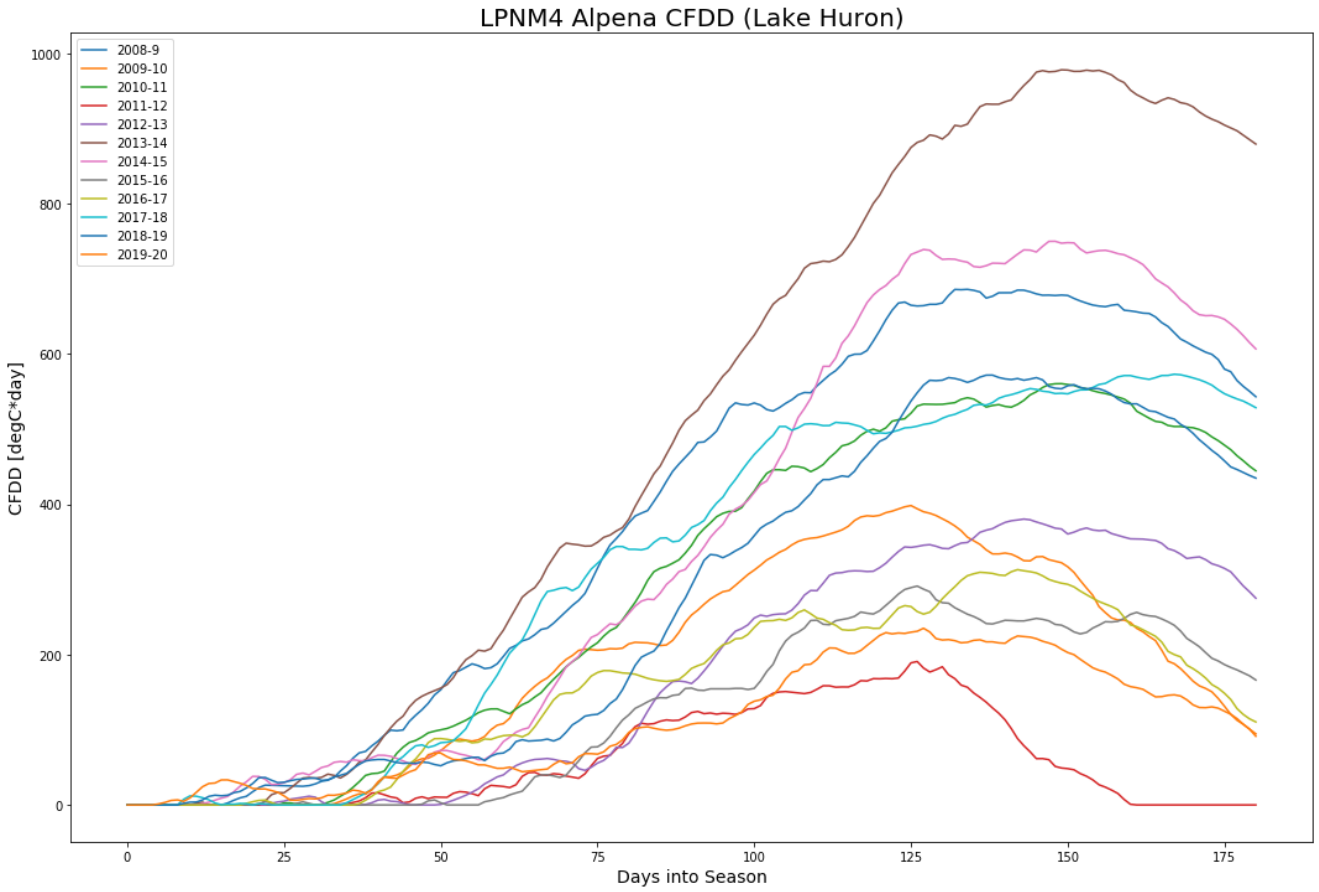
	Avg.Temp	Max.CFDD	Missing
2008-2009	-0.325474	458.206040	0.0
2009-2010	1.601228	293.073311	0.0
2010-2011	-0.639986	433.801895	0.0
2011-2012	3.336376	73.488967	0.0
2012-2013	0.968217	198.066209	0.0
2013-2014	-2.723007	802.760784	0.0
2014-2015	-1.886424	627.136115	0.0
2015-2016	2.156716	194.451746	0.0
2016-2017	2.297256	182.961179	3.0
2017-2018	-1.027469	429.309487	0.0
2018-2019	0.077156	341.563106	0.0
2019-2020	1.516333	85.675434	0.0

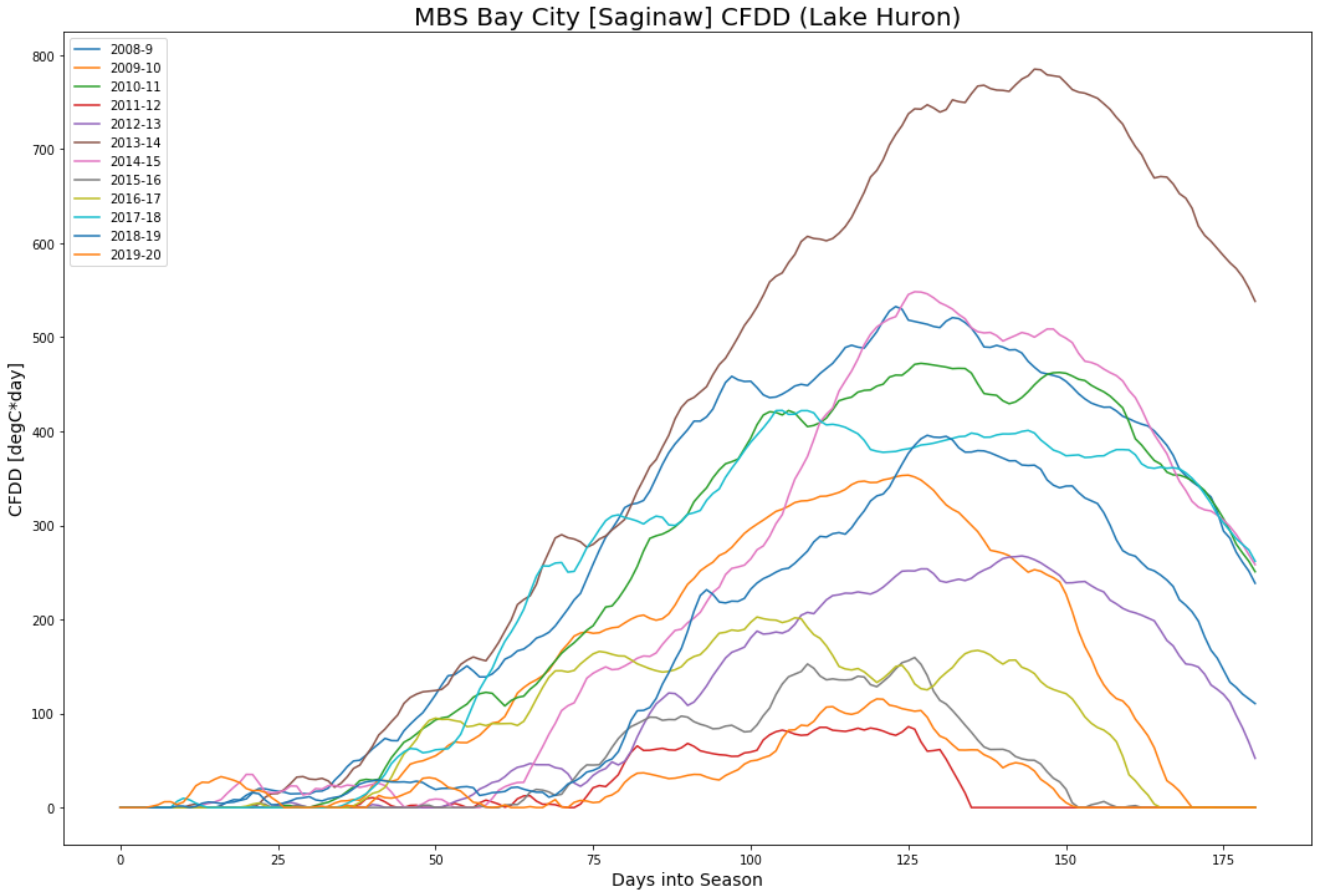
BCI Sensor Height: 203m

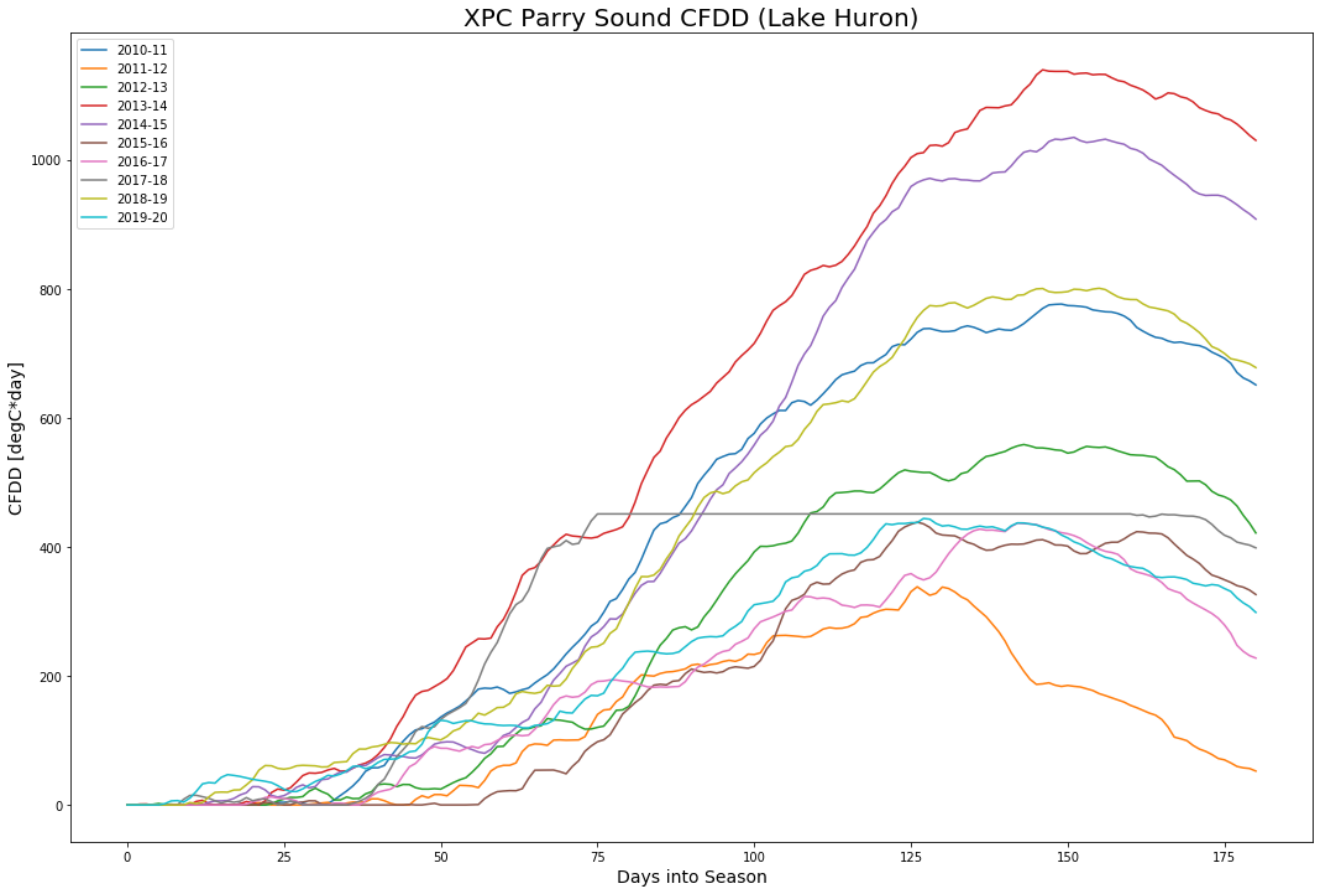
	Avg.Temp	Max.CFDD	Missing
2008-2009	-0.673132	532.684329	0.0
2009-2010	1.642872	353.394472	0.0
2010-2011	-0.692028	472.274707	0.0
2011-2012	3.614964	85.993530	0.0
2012-2013	0.622621	267.512939	0.0
2013-2014	-2.329347	785.073094	0.0
2014-2015	-0.981056	548.398406	0.0
2015-2016	2.947438	159.490032	0.0
2016-2017	2.336475	202.582486	0.0
2017-2018	-0.748682	422.214156	0.0
2018-2019	-0.336263	395.762013	0.0
2019-2020	1.242142	115.495877	0.0

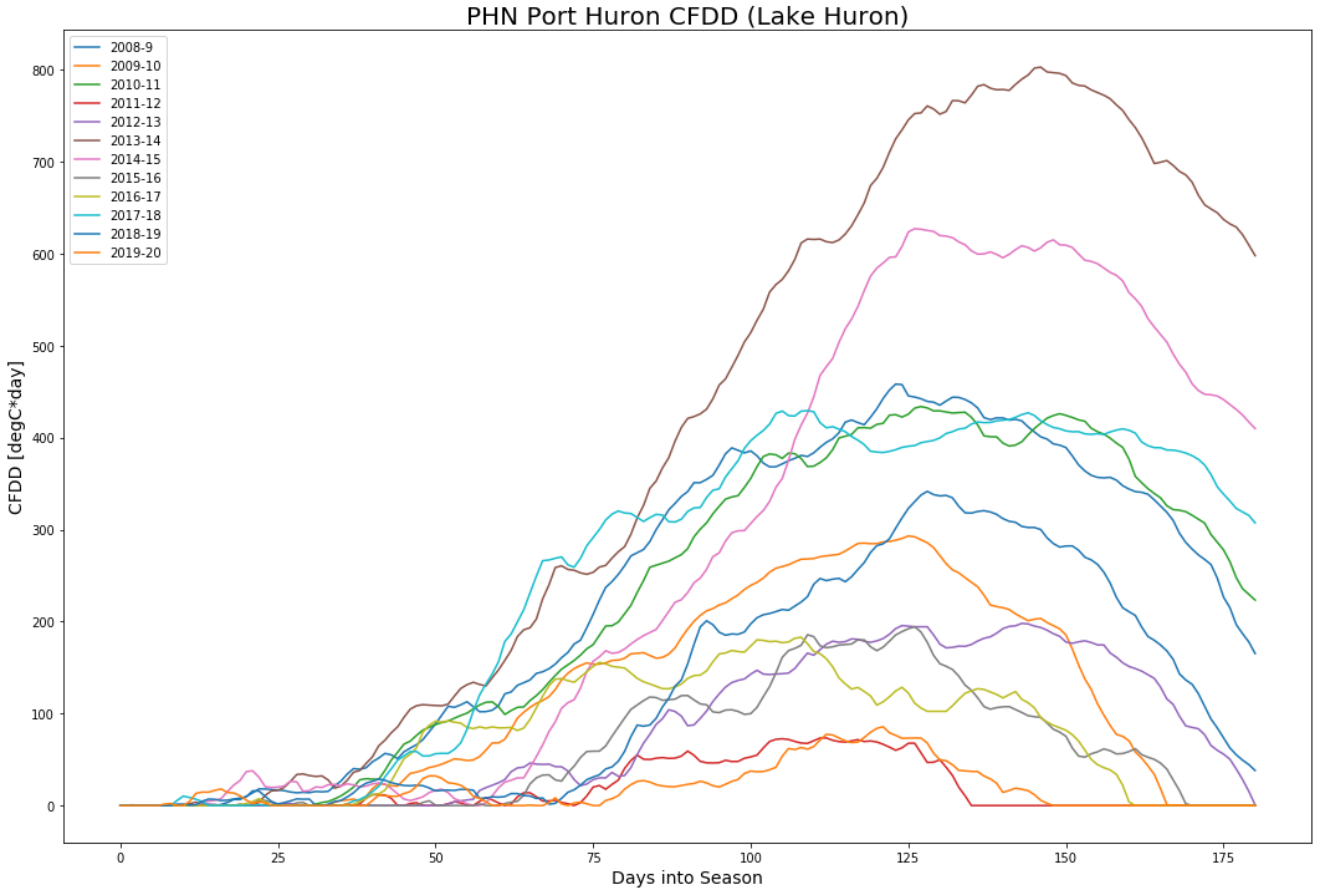
PAS Sensor Height: 178m

	Avg.Temp	Max.CFDD	Missing
2010-2011	-3.036775	776.409938	0.0
2011-2012	0.678618	338.172101	0.0
2012-2013	-1.909068	558.696551	0.0
2013-2014	-5.274204	1139.508329	0.0
2014-2015	-4.790116	1034.817147	0.0
2015-2016	-0.311530	438.335505	0.0
2016-2017	-0.361279	436.884552	0.0
2017-2018	-3.446631	451.081159	85.0
2018-2019	-3.498436	800.906592	0.0
2019-2020	-1.548411	444.064747	0.0


LPNM4 Alpena CFDD (Lake Huron)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


MBS Bay City [Saginaw] CFDD (Lake Huron)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


XPC Parry Sound CFDD (Lake Huron)
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


PHN Port Huron CFDD (Lake Huron)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season

D. **Lake Michigan**

ESC Sensor Height: 180m

	Avg.Temp	Max.CFDD	Missing
2008-2009	-3.653394	875.755078	0.0
2009-2010	-0.589042	521.823642	0.0
2010-2011	-2.878222	702.673543	0.0
2011-2012	0.423062	305.200299	0.0
2012-2013	-2.432408	589.377169	1.0
2013-2014	-6.285585	1249.575237	5.0
2014-2015	-4.402237	951.822581	1.0
2015-2016	-0.365223	374.123604	0.0
2016-2017	-0.909332	509.828816	0.0
2017-2018	-4.095598	833.084933	2.0
2018-2019	-4.030086	860.770036	0.0
2019-2020	-2.349841	510.858533	0.0

MIL Sensor Height: 210m

	Avg.Temp	Max.CFDD	Missing
2008-2009	-0.398916	504.002823	0.0
2009-2010	1.917087	330.742296	0.0
2010-2011	-0.166801	427.541949	0.0
2011-2012	3.576472	86.315776	0.0
2012-2013	0.386484	331.569756	0.0
2013-2014	-2.678159	822.834944	0.0
2014-2015	-0.814364	528.363240	0.0
2015-2016	2.382511	235.387504	0.0
2016-2017	2.872319	211.394622	0.0
2017-2018	-0.046723	371.636085	0.0
2018-2019	0.157113	356.539048	0.0
2019-2020	1.577616	98.083126	0.0

GRB Sensor Height: 211m

	Avg.Temp	Max.CFDD	Missing
2008-2009	-3.071147	879.585591	0.0
2009-2010	0.110275	556.615532	0.0
2010-2011	-2.467908	714.944735	0.0
2011-2012	2.173043	213.820931	0.0
2012-2013	-1.561527	547.317613	0.0
2013-2014	-5.196120	1193.792372	0.0
2014-2015	-2.490947	792.730732	0.0
2015-2016	0.690573	370.355993	0.0
2016-2017	1.111029	345.511392	0.0
2017-2018	-2.539218	629.959191	0.0
2018-2019	-1.636816	576.931423	8.0
2019-2020	-0.799930	366.328329	0.0

CHI Sensor Height: 176m

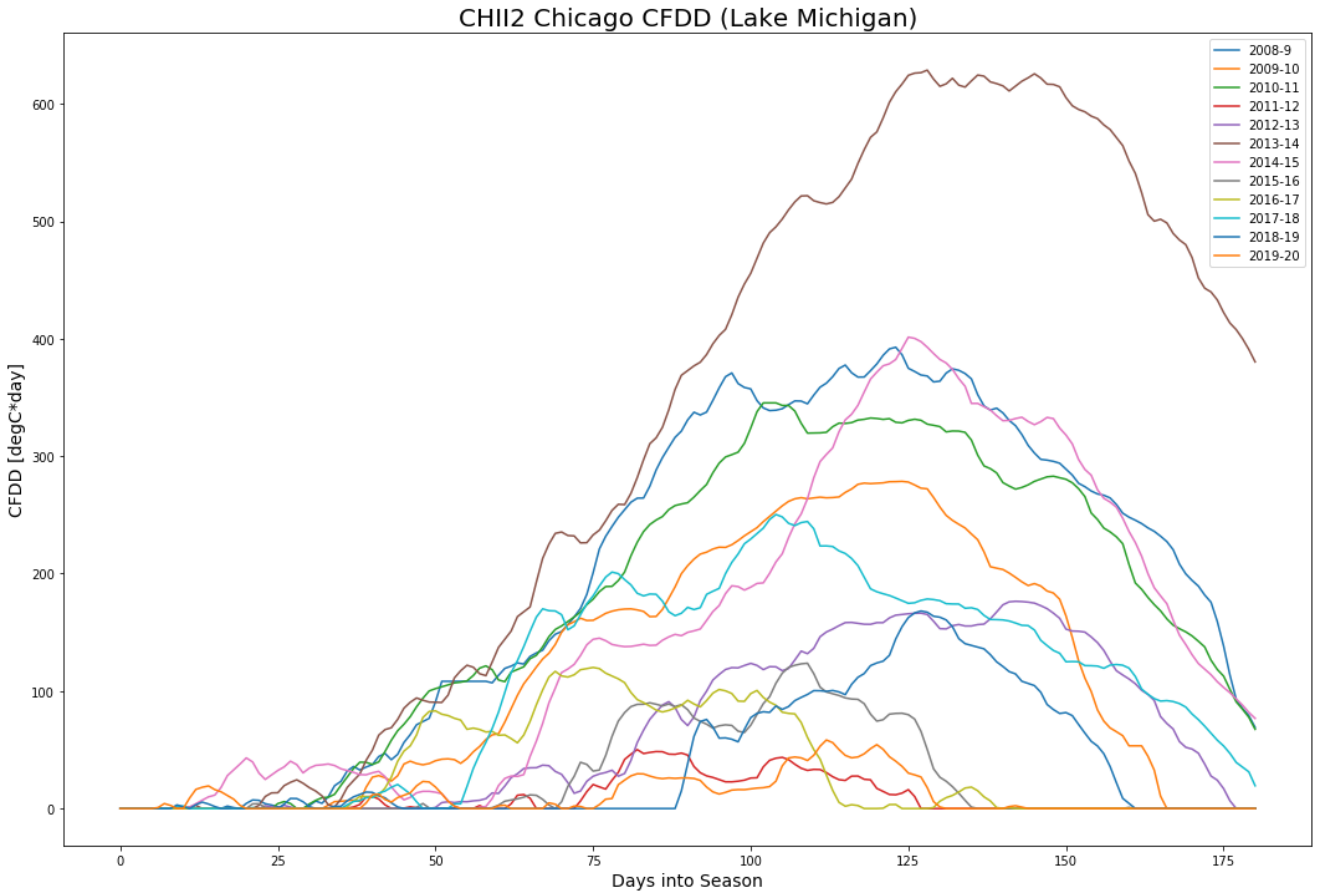
	Avg.Temp	Max.CFDD	Missing
2008-2009	0.458556	392.742180	7.0
2009-2010	2.234179	278.547283	2.0
2010-2011	0.720407	345.362839	2.0
2011-2012	4.727714	50.171617	0.0
2012-2013	1.751601	176.448668	1.0
2013-2014	-1.438356	628.674847	1.0
2014-2015	0.127757	401.355084	1.0
2015-2016	3.851940	123.706229	9.0
2016-2017	4.075517	120.100498	3.0
2017-2018	1.160279	250.358911	1.0
2018-2019	1.952608	168.194288	33.0
2019-2020	2.430911	58.353494	0.0

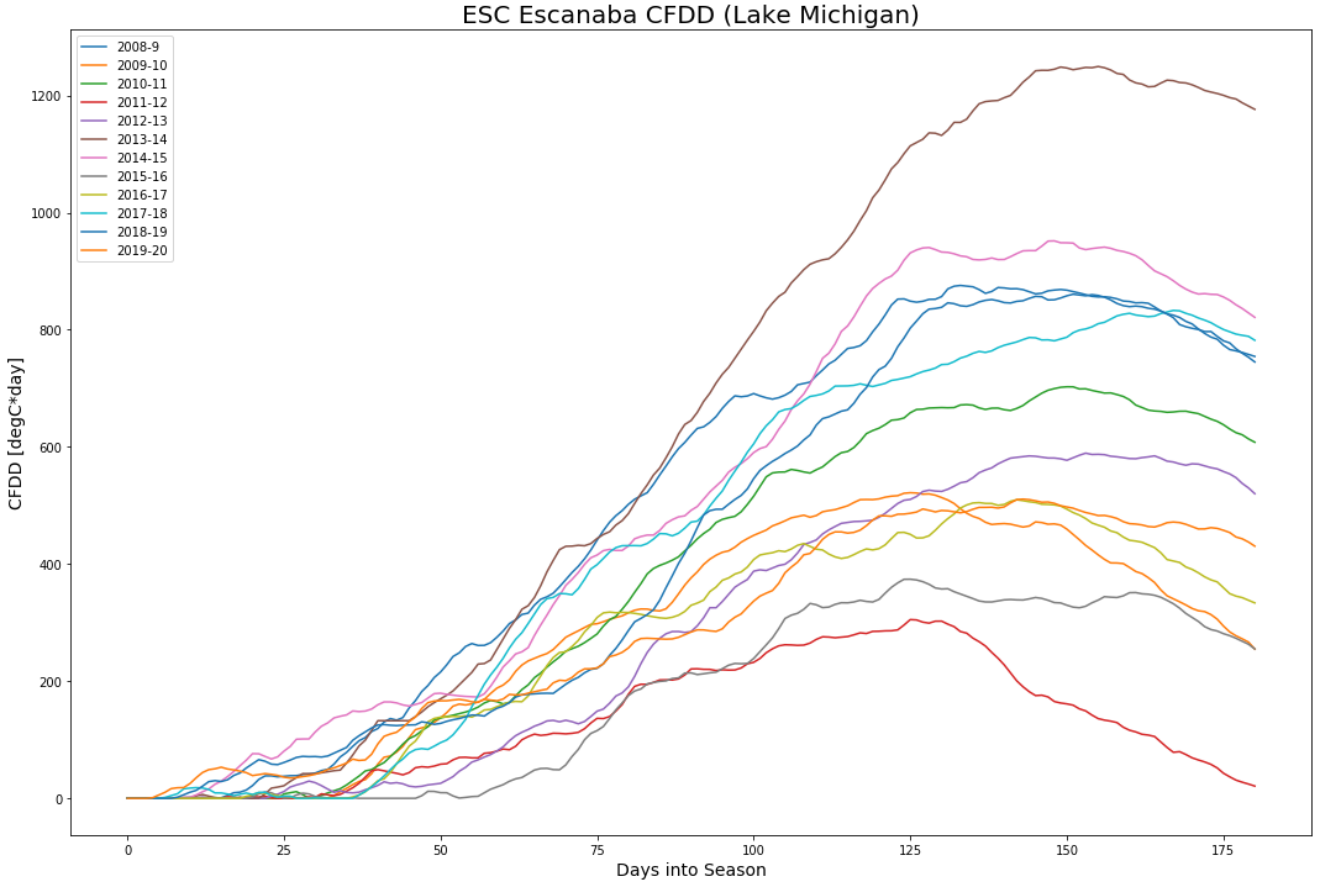
MSK Sensor Height: 179m

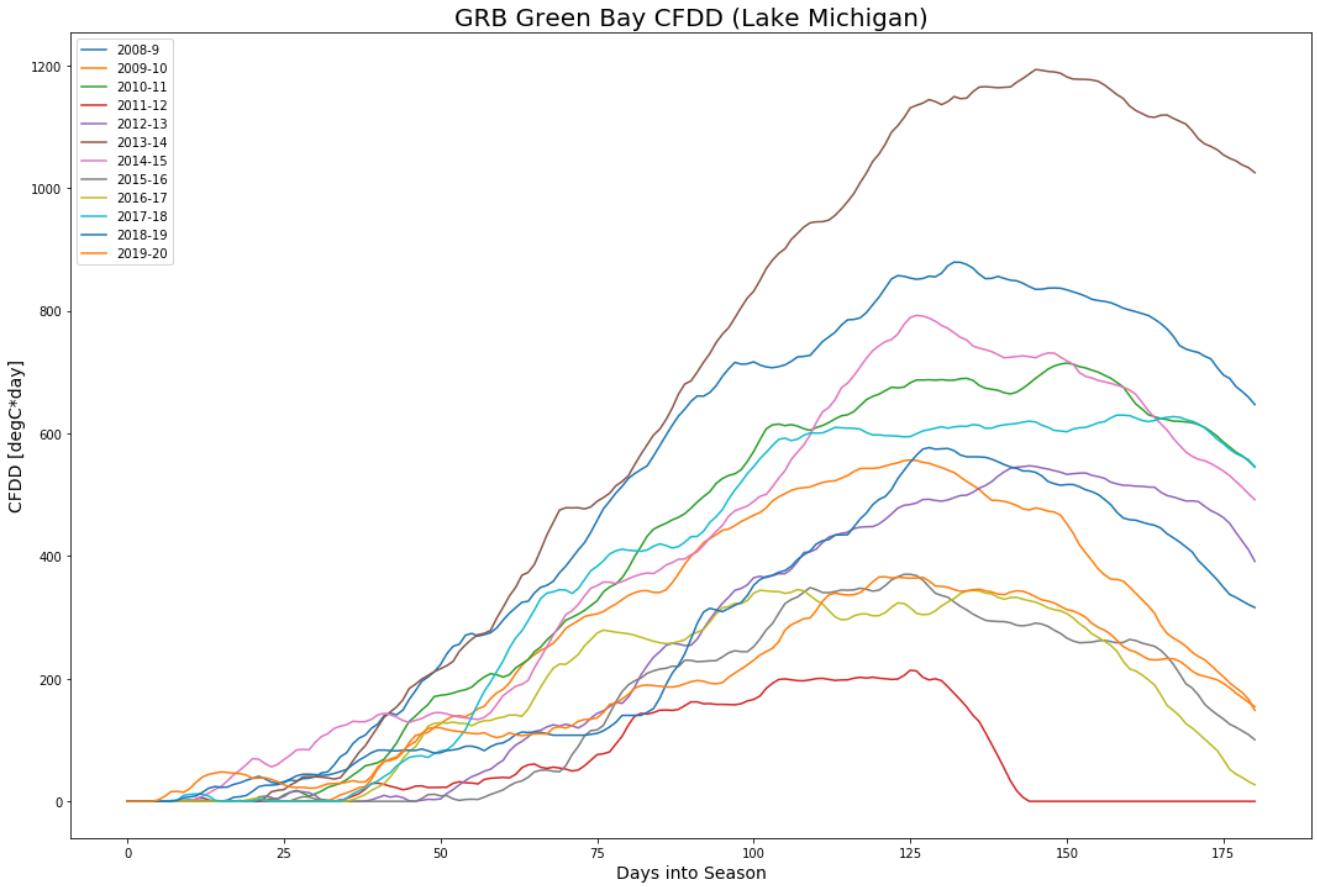
	Avg.Temp	Max.CFDD	Missing
2008-2009	-0.218984	408.725073	9.0
2009-2010	1.983092	231.779547	7.0
2010-2011	-0.041620	351.647524	2.0
2011-2012	4.098351	48.742464	11.0
2012-2013	1.028863	202.073998	3.0
2013-2014	-1.785113	594.984365	8.0
2014-2015	-0.784069	427.116641	3.0
2015-2016	2.976444	136.215465	0.0
2016-2017	2.625783	122.309999	1.0
2017-2018	-0.031959	300.703021	1.0
2018-2019	0.398898	209.300042	33.0
2019-2020	1.468954	72.615152	0.0

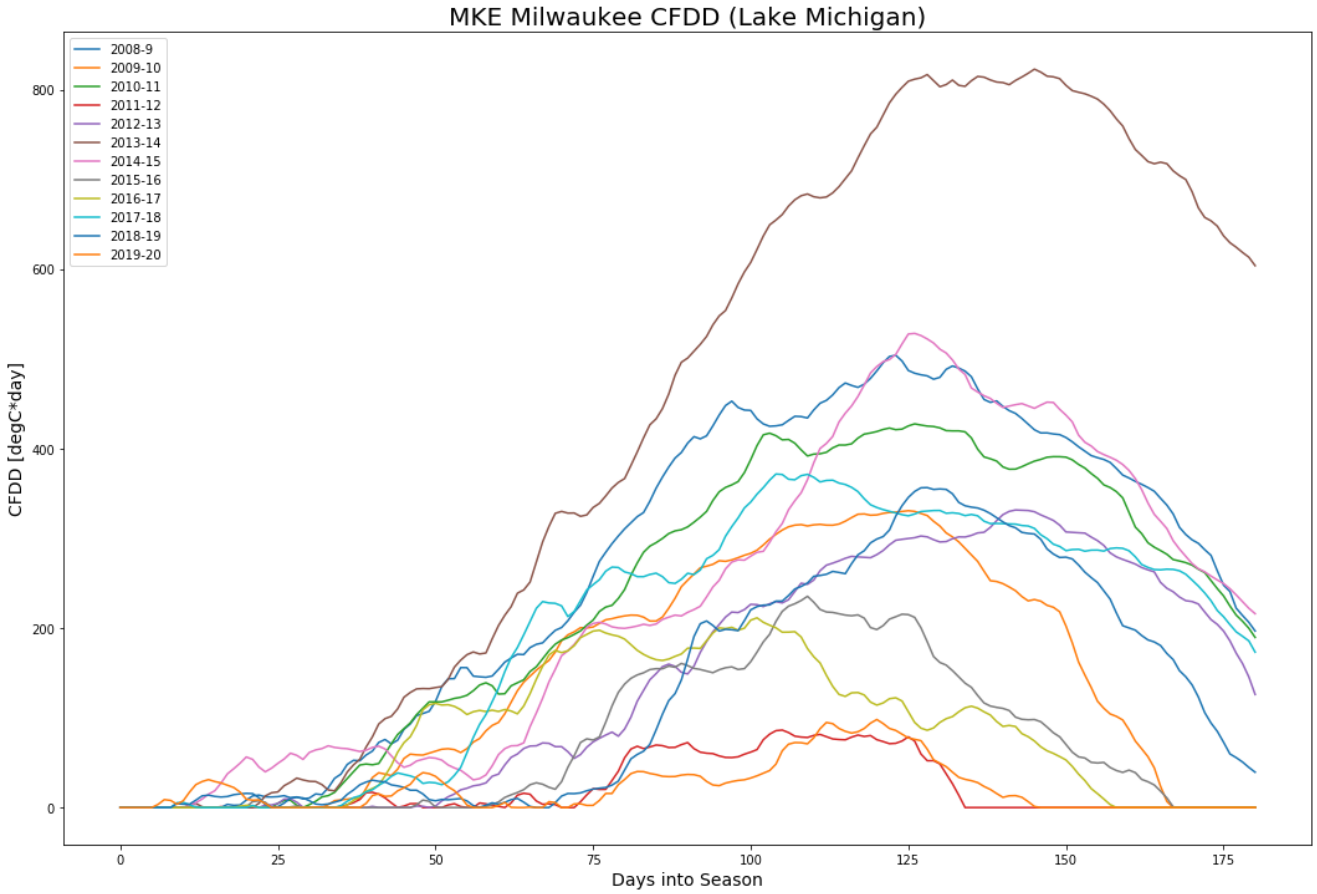
TVC Sensor Height: 190m

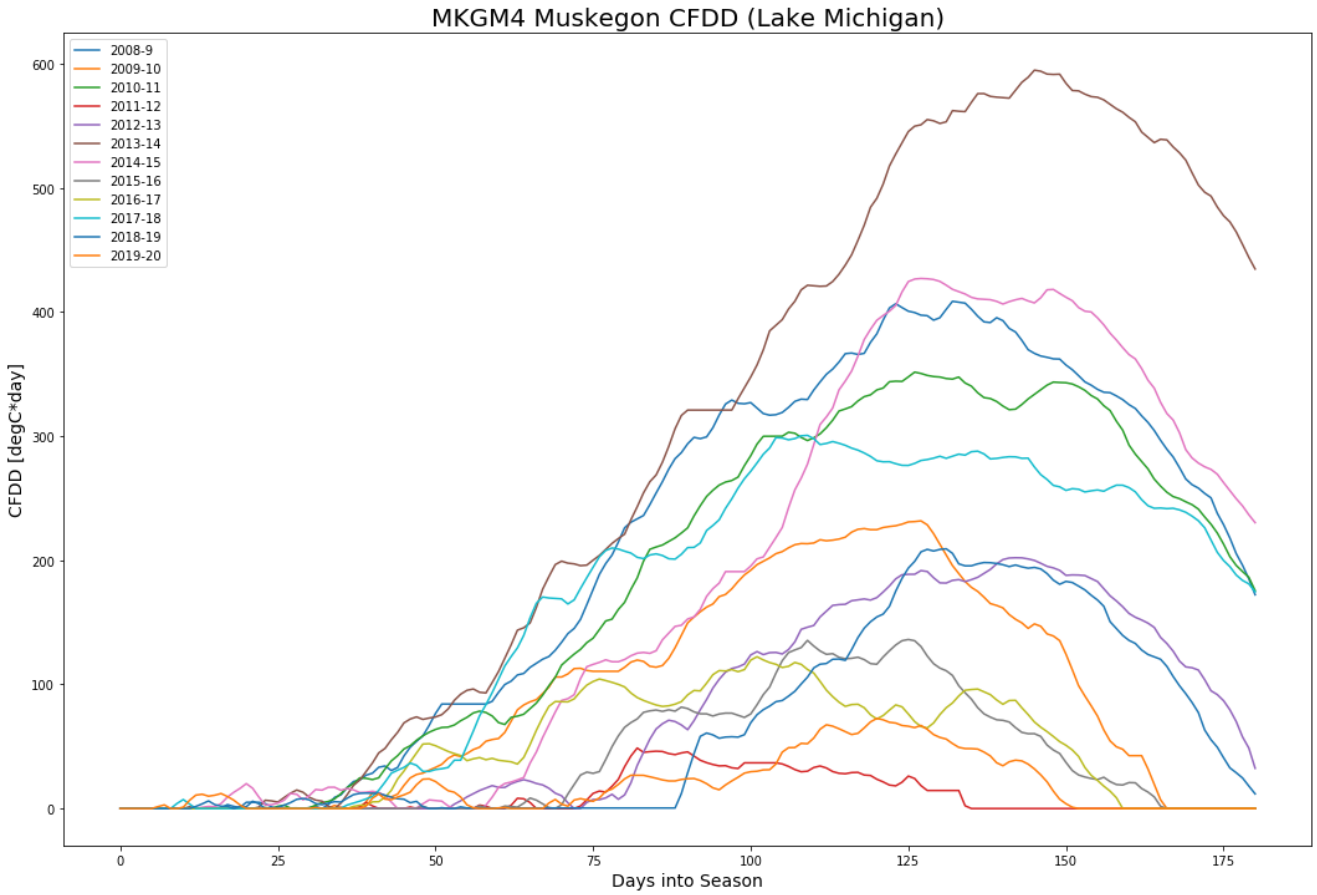
	Avg.Temp	Max.CFDD	Missing
2008-2009	-1.365757	585.992403	0.0
2009-2010	1.003509	388.928356	0.0
2010-2011	-0.836712	450.836658	0.0
2011-2012	2.917602	116.114243	0.0
2012-2013	0.054131	317.591437	0.0
2013-2014	-3.424531	893.282508	0.0
2014-2015	-2.199586	654.609588	0.0
2015-2016	2.047741	204.065264	0.0
2016-2017	1.916588	205.644607	0.0
2017-2018	-1.224261	408.686785	0.0
2018-2019	-1.056613	449.140878	0.0
2019-2020	0.624043	125.609192	0.0

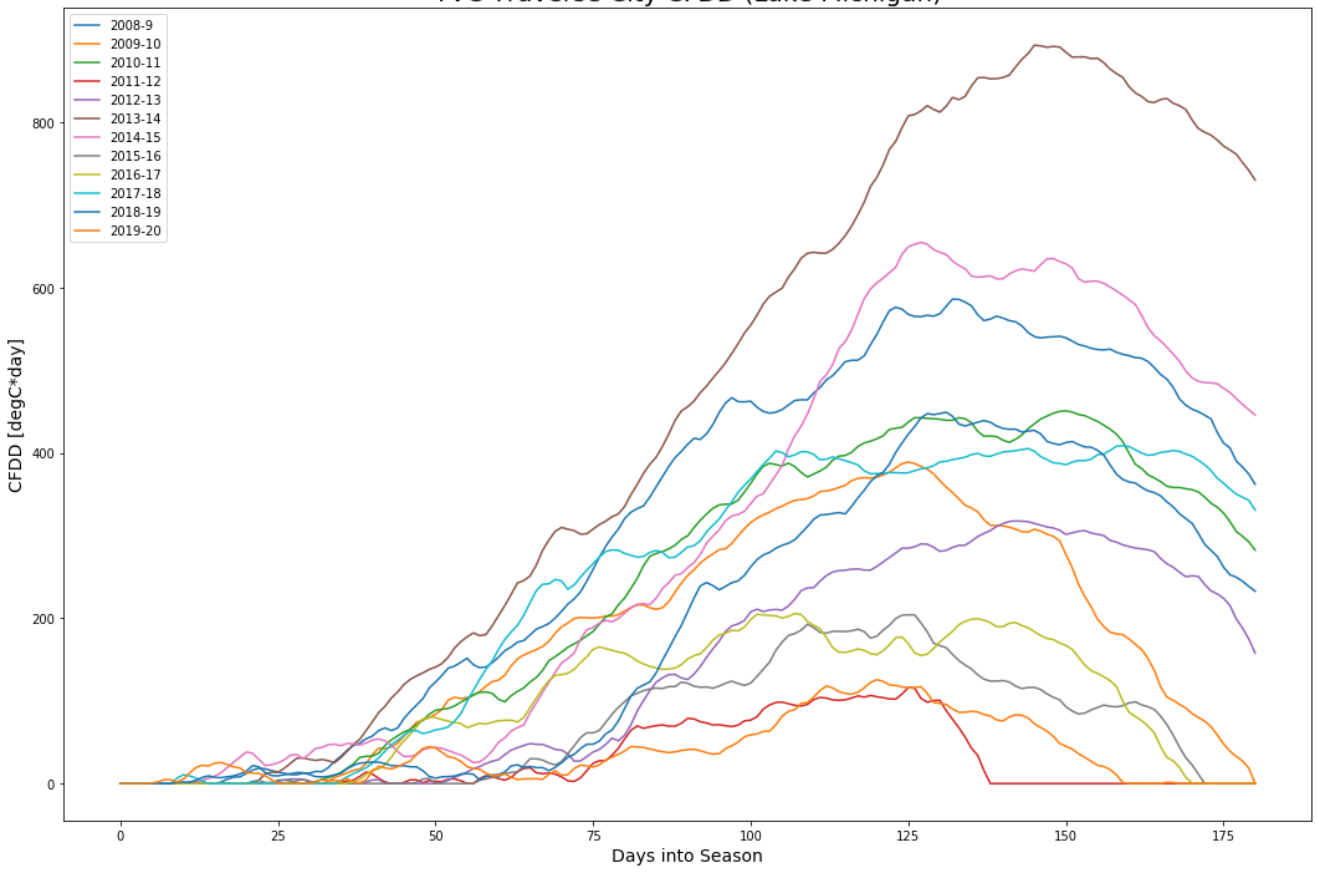

CHII2 Chicago CFDD (Lake Michigan)
CFDD [degC*day]
Days into Season
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20


ESC Escanaba CFDD (Lake Michigan)
CFDD [degC*day]
Days into Season
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20


GRB Green Bay CFDD (Lake Michigan)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


MKE Milwaukee CFDD (Lake Michigan)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


MKGM4 Muskegon CFDD (Lake Michigan)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season


TVC Traverse City CFDD (Lake Michigan)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season

E. **Lake Ontario**

ROC Sensor Height: 75m

	Avg.Temp	Max.CFDD	Missing
2008-2009	1.199219	253.956669	1.0
2009-2010	2.617460	210.400443	0.0
2010-2011	1.033728	273.567967	0.0
2011-2012	4.482820	39.329060	0.0
2012-2013	2.316934	124.368768	2.0
2013-2014	-0.450459	438.336191	1.0
2014-2015	-0.218212	468.597907	0.0
2015-2016	3.790960	84.389292	0.0
2016-2017	3.412682	47.350748	0.0
2017-2018	1.104128	209.305252	3.0
2018-2019	1.097995	229.480229	0.0
2019-2020	2.278341	64.343431	0.0

OSW Sensor Height: 78m

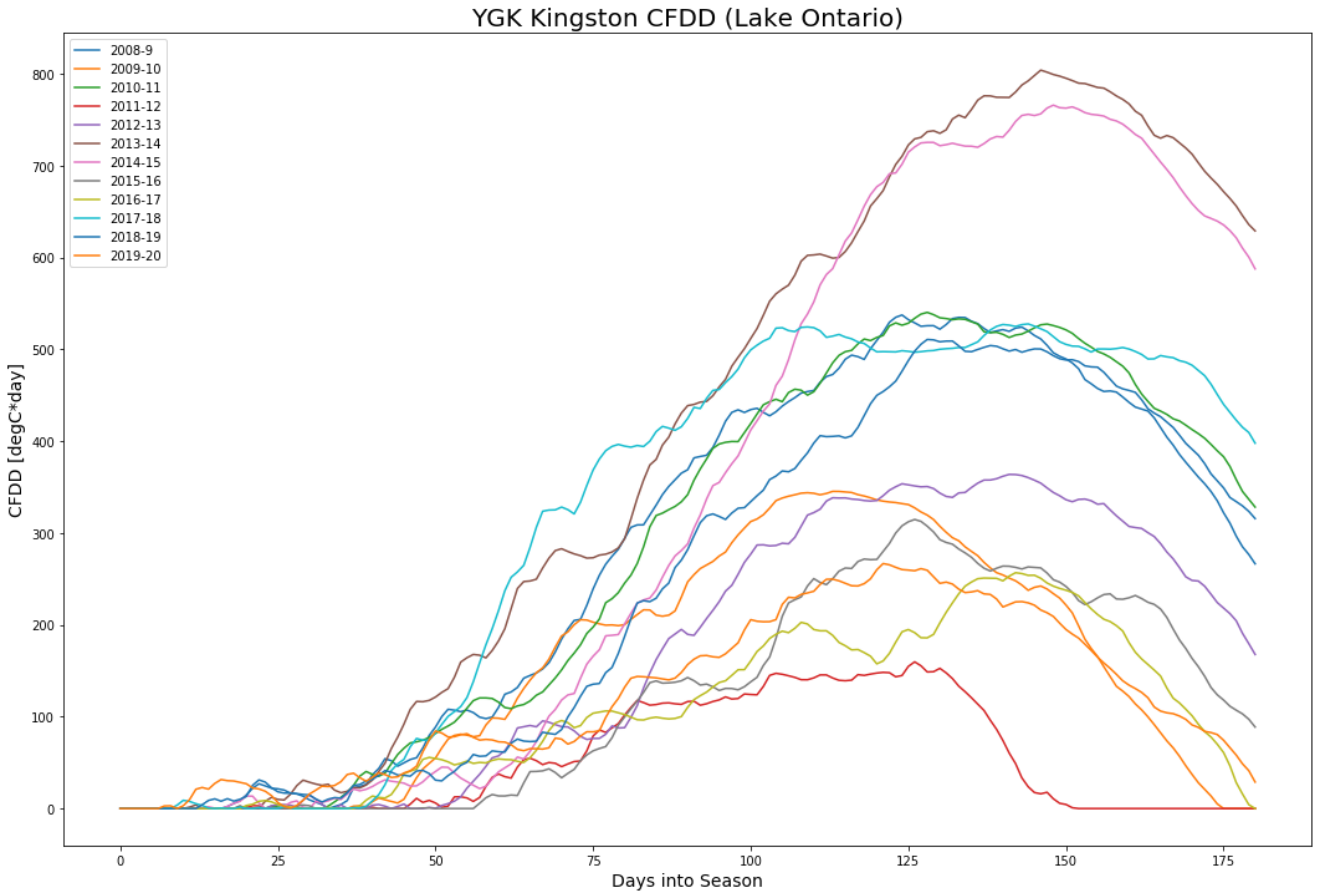
	Avg.Temp	Max.CFDD	Missing
2008-2009	0.452626	347.196624	0.0
2009-2010	2.266091	253.313829	0.0
2010-2011	0.409145	353.654845	0.0
2011-2012	3.783314	57.758185	0.0
2012-2013	1.569325	179.259091	2.0
2013-2014	-1.109361	521.503381	1.0
2014-2015	-1.134697	557.872589	1.0
2015-2016	2.944355	140.180581	2.0
2016-2017	2.652153	81.554385	1.0
2017-2018	0.260056	304.390279	4.0
2018-2019	0.205196	289.895074	4.0
2019-2020	1.844474	96.377280	71.0

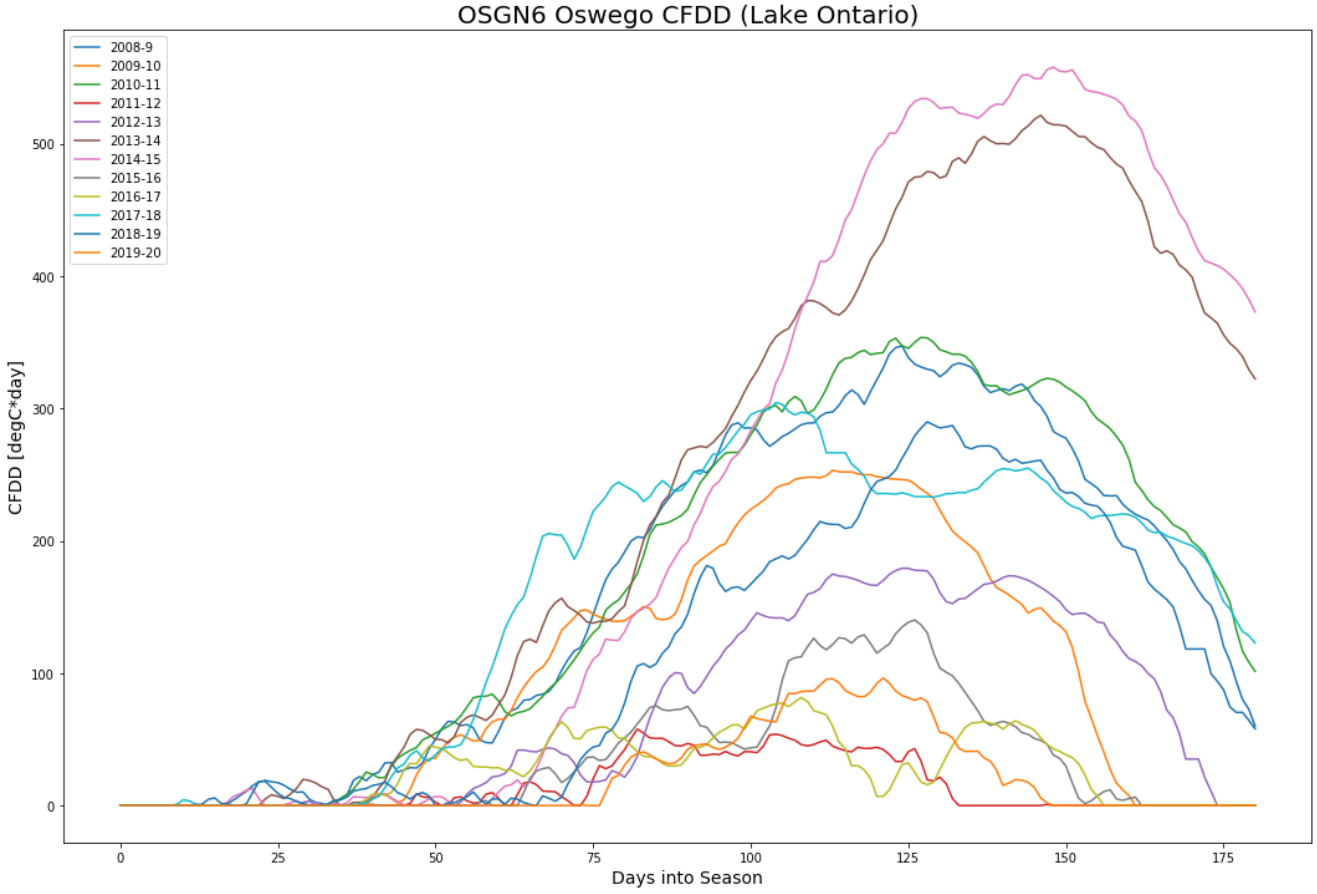
KIN Sensor Height: 93m

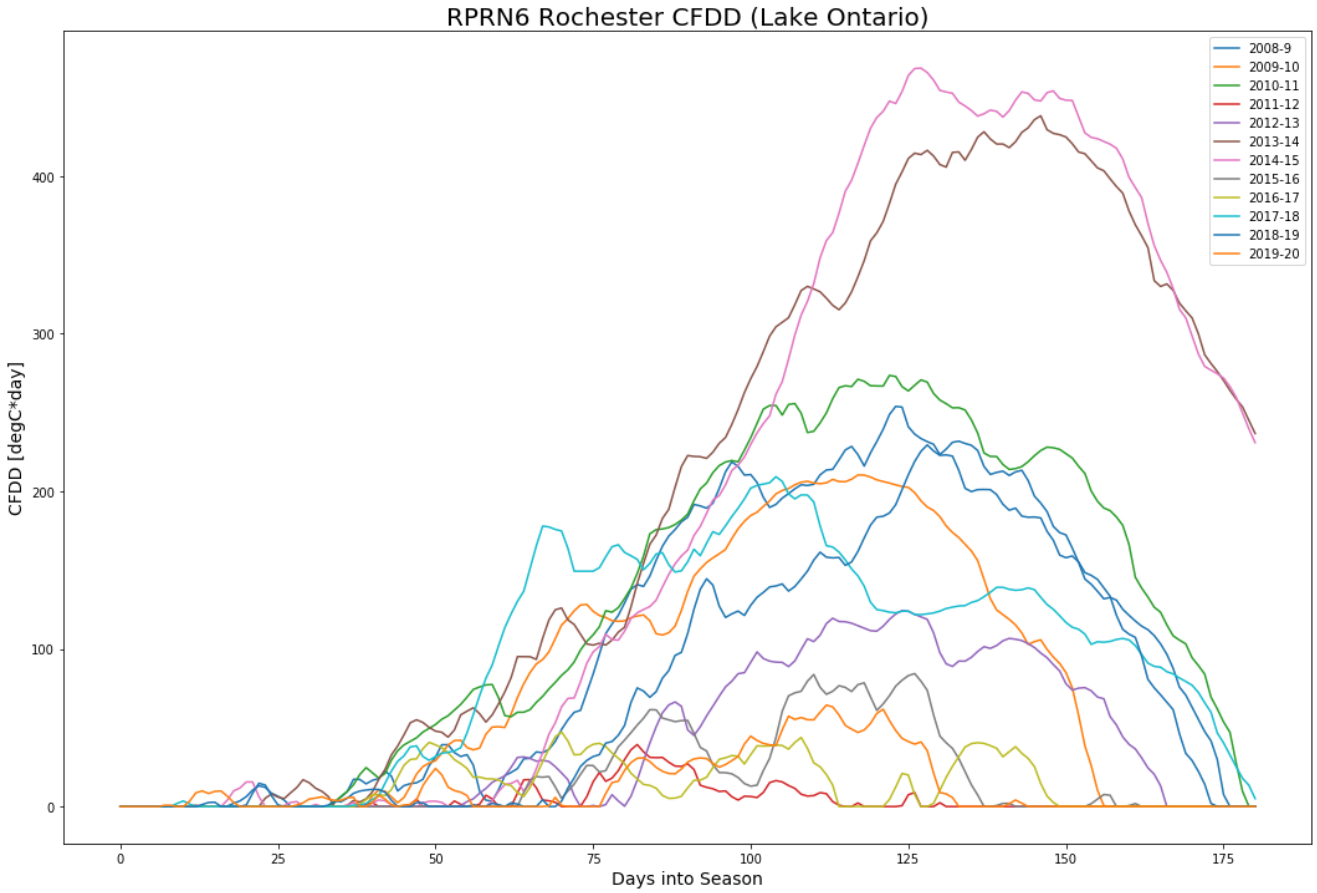
	Avg.Temp	Max.CFDD	Missing
2008-2009	-0.780910	537.393603	0.0
2009-2010	1.537626	345.342105	0.0
2010-2011	-1.028567	540.236133	0.0
2011-2012	2.544137	159.719642	0.0
2012-2013	-0.271684	363.808963	0.0
2013-2014	-2.850593	804.004147	0.0
2014-2015	-2.710024	765.904607	0.0
2015-2016	1.414072	314.736402	0.0
2016-2017	1.050093	256.785333	0.0
2017-2018	-1.412751	527.666669	0.0
2018-2019	-1.391479	510.753356	0.0
2019-2020	0.054017	266.745562	0.0

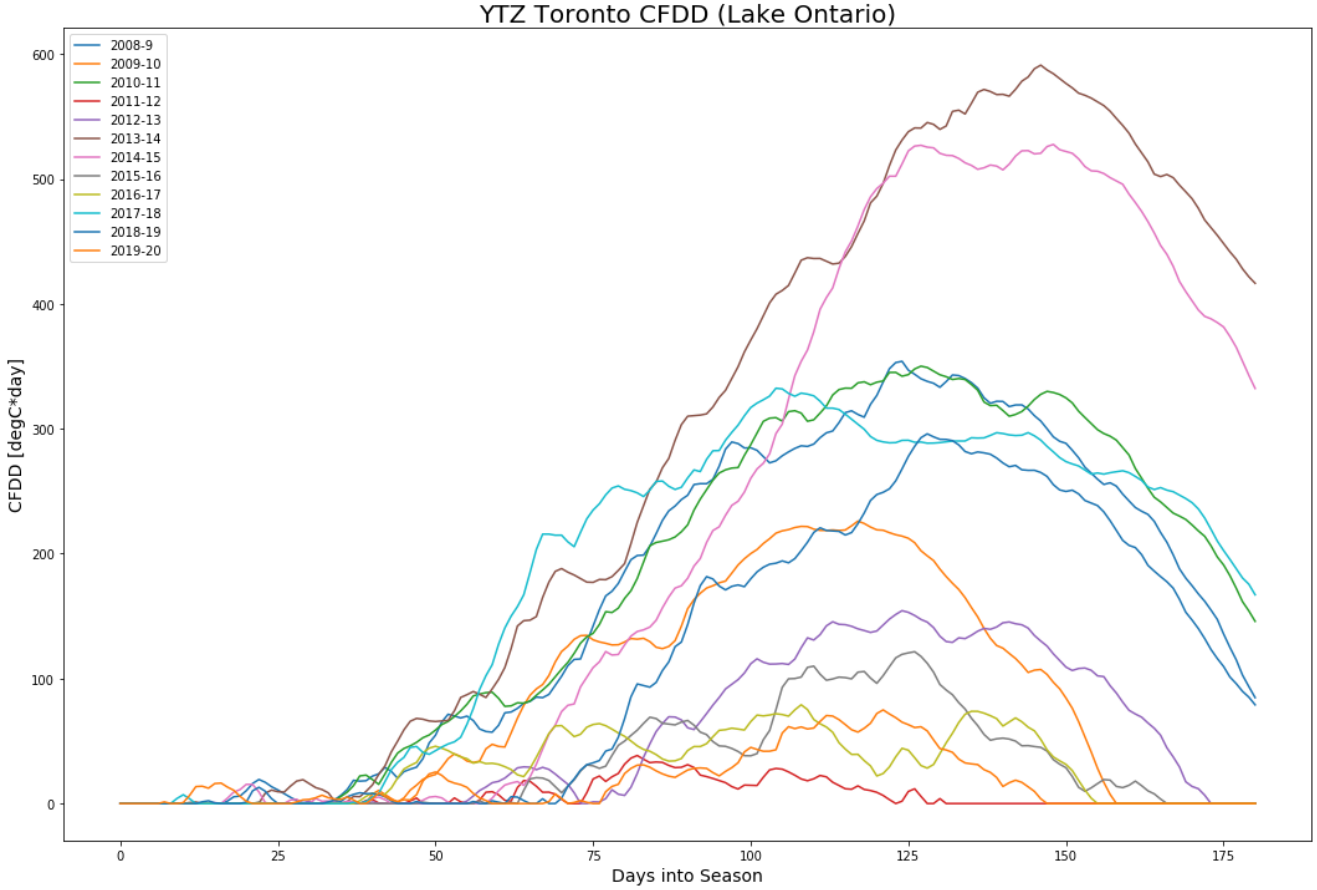
TOR Sensor Height: 77m

	Avg.Temp	Max.CFDD	Missing
2008-2009	0.242208	353.988457	0.0
2009-2010	2.503361	226.093890	0.0
2010-2011	0.123914	350.109148	0.0
2011-2012	3.629889	38.458333	0.0
2012-2013	1.529315	154.381055	0.0
2013-2014	-1.660257	591.069495	0.0
2014-2015	-1.098876	527.607615	0.0
2015-2016	2.671352	121.516934	0.0
2016-2017	2.531493	79.066863	0.0
2017-2018	0.006607	332.446786	0.0
2018-2019	0.265386	295.932570	0.0
2019-2020	1.622291	74.941707	0.0


YGK Kingston CFDD (Lake Ontario)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season

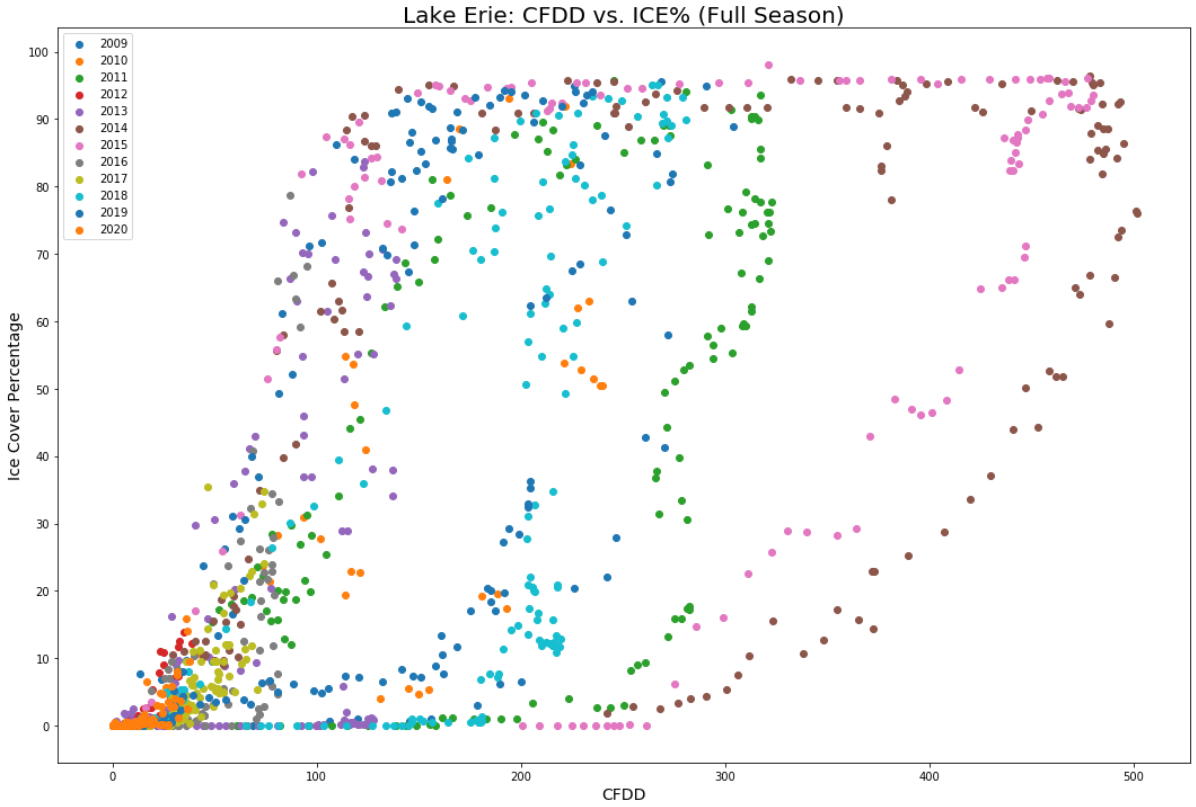

OSGN6 Oswego CFDD (Lake Ontario)
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
CFDD [degC*day]
Days into Season

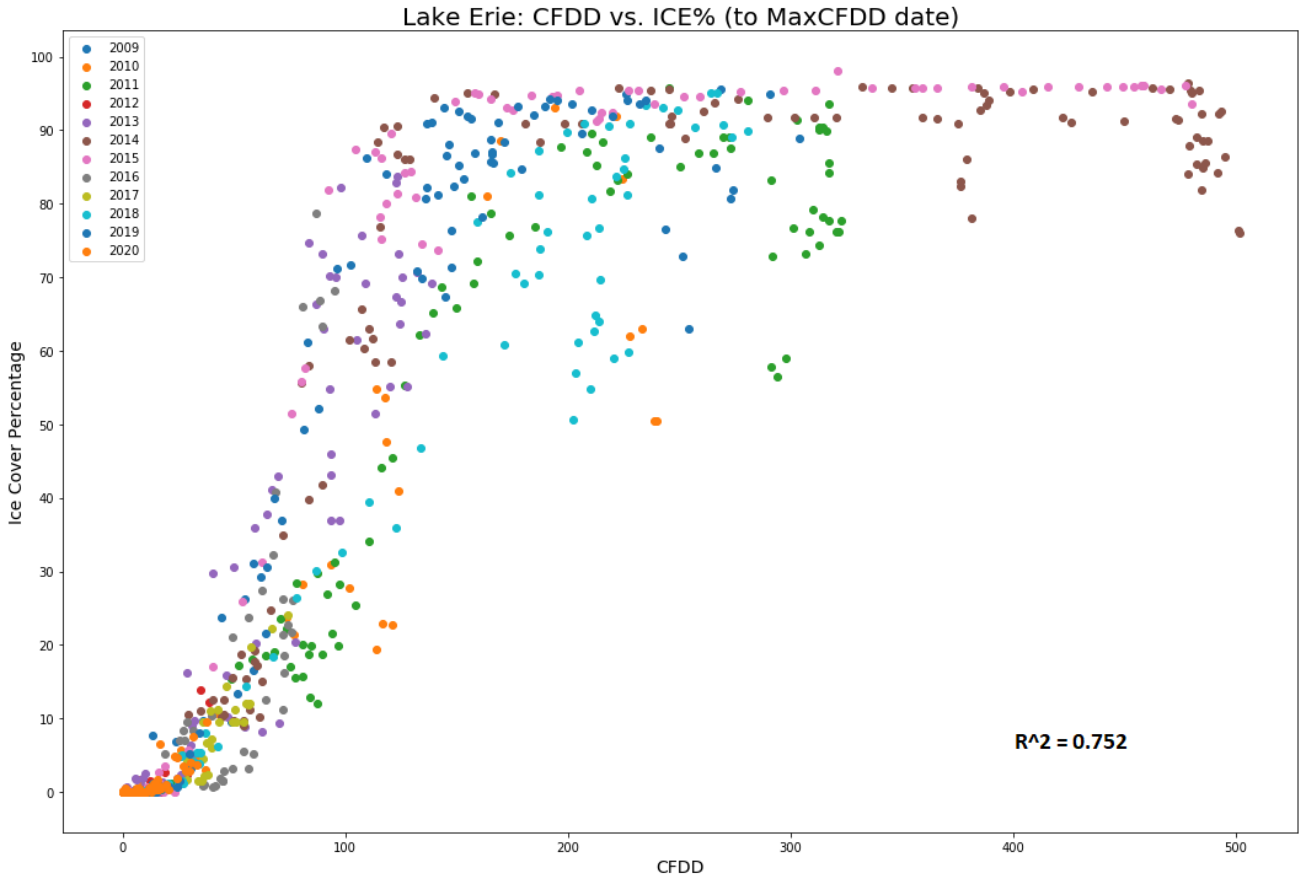

RPRN6 Rochester CFDD (Lake Ontario)
CFDD [degC*day]
Days into Season
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20

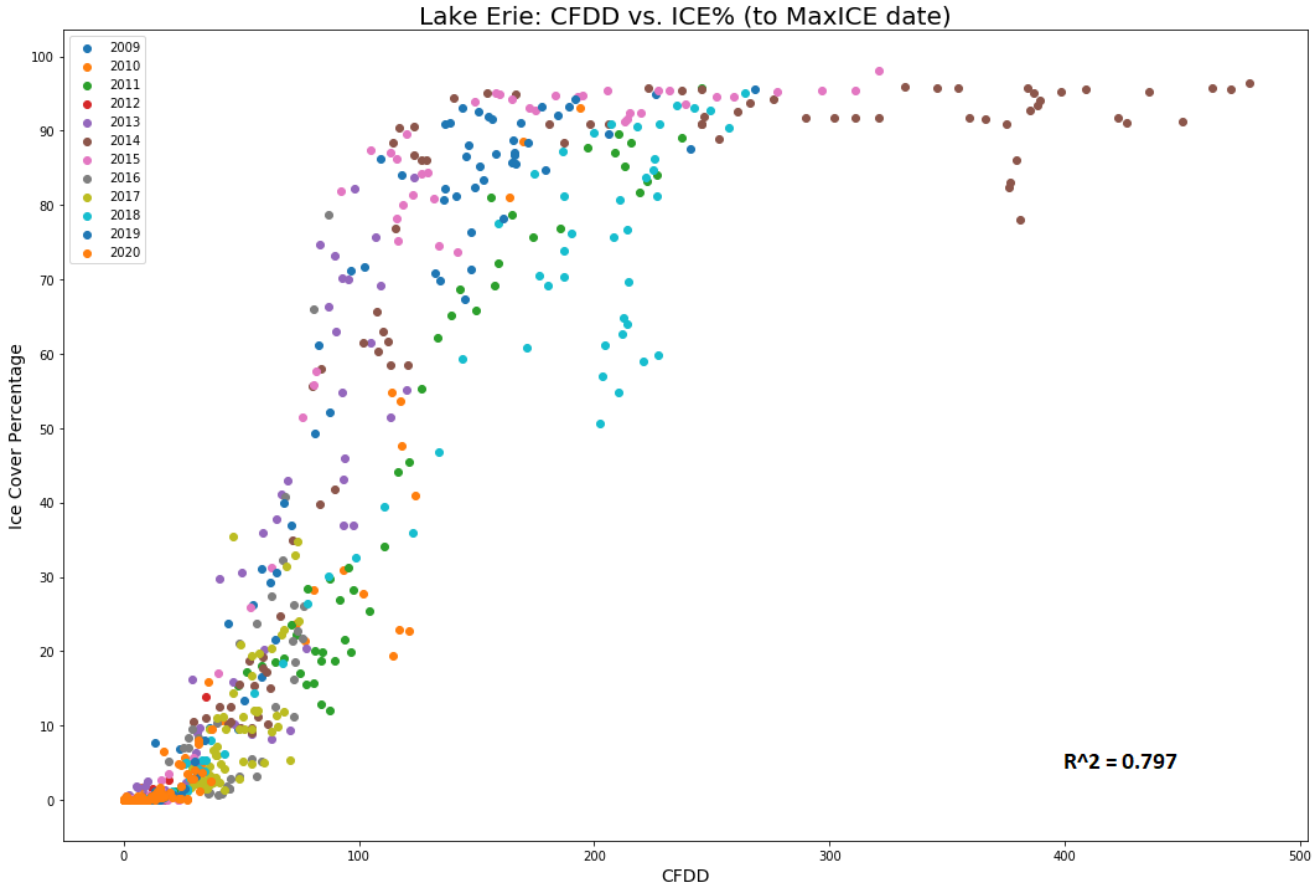

YTZ Toronto CFDD (Lake Ontario)
CFDD [degC*day]
Days into Season
2008-9
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20

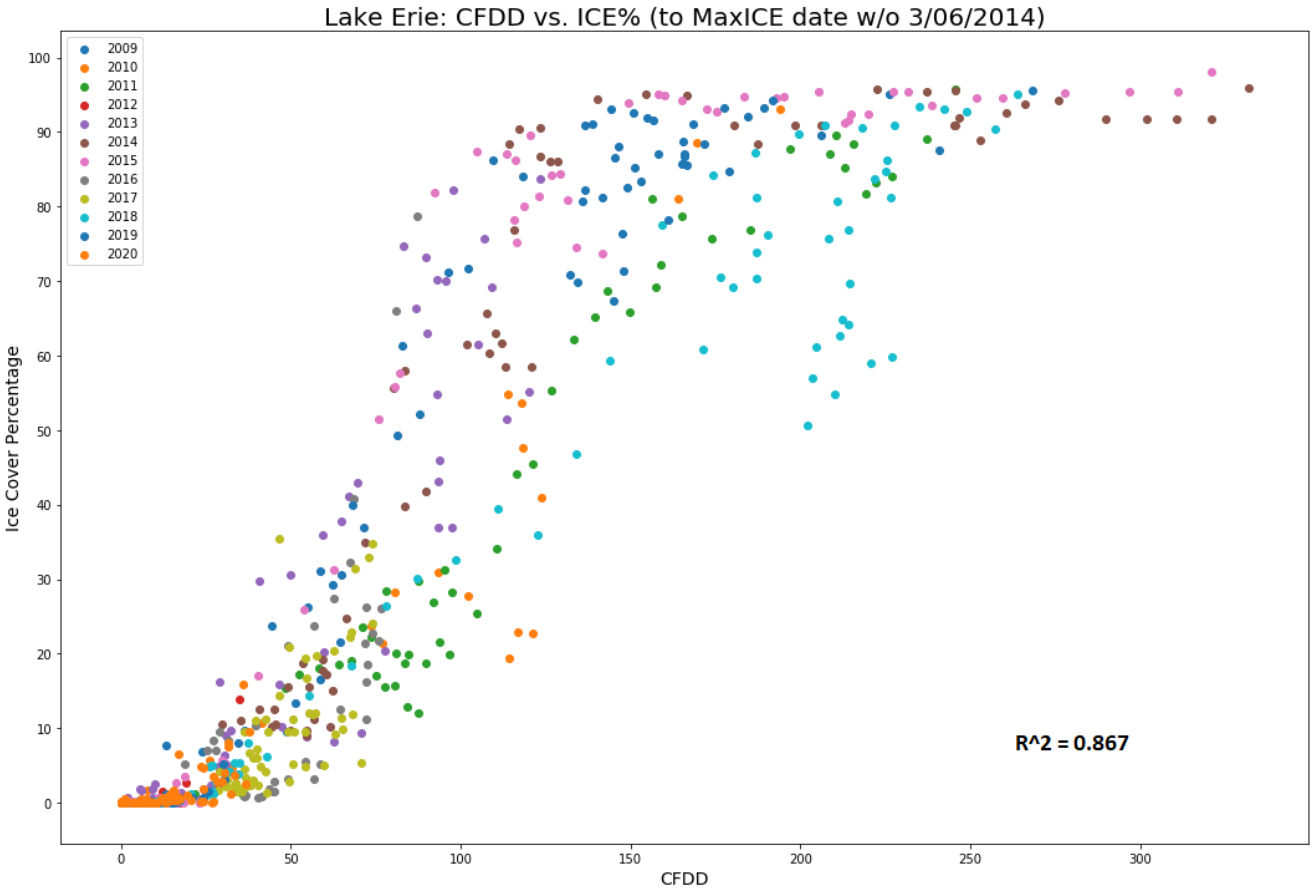
CFDD VS. ICE COVER SCATTER PLOTS

A. Lake Erie

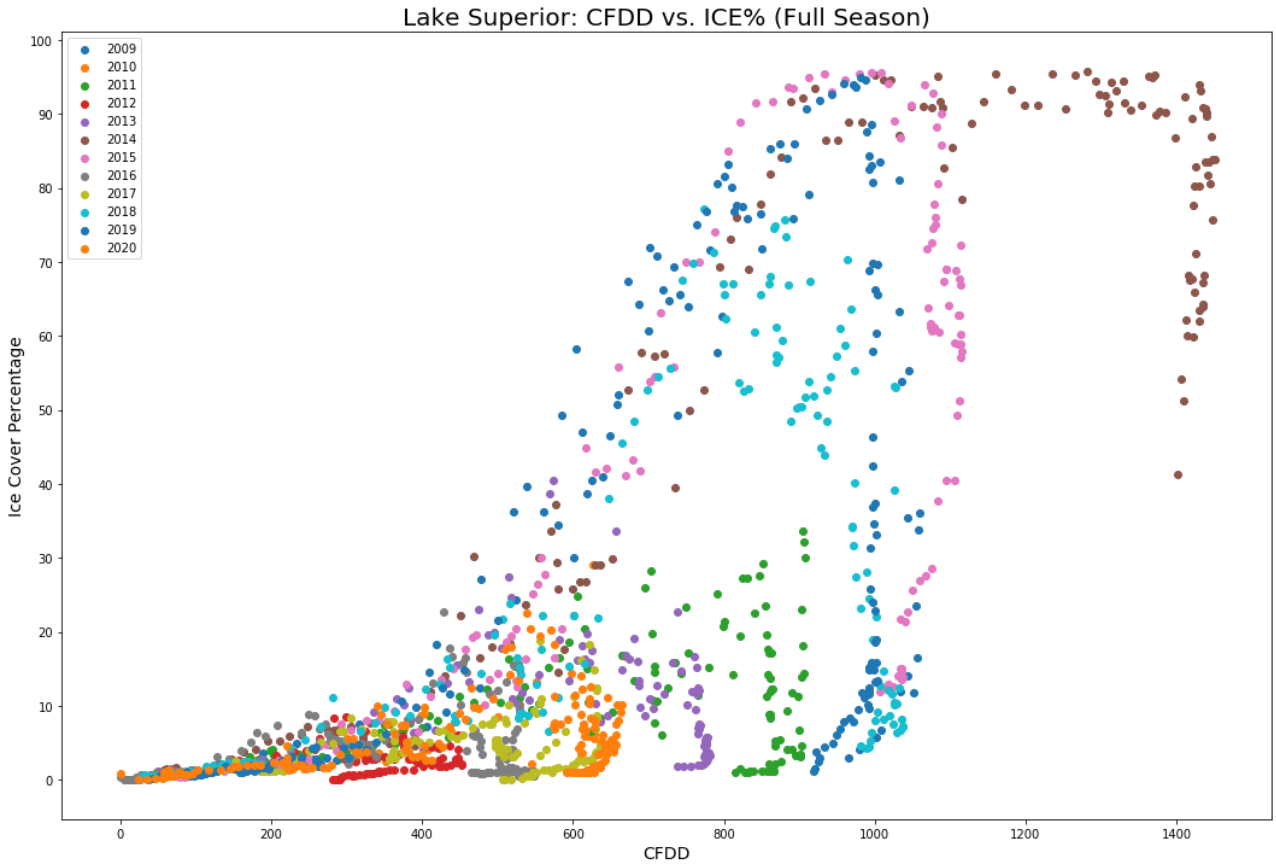


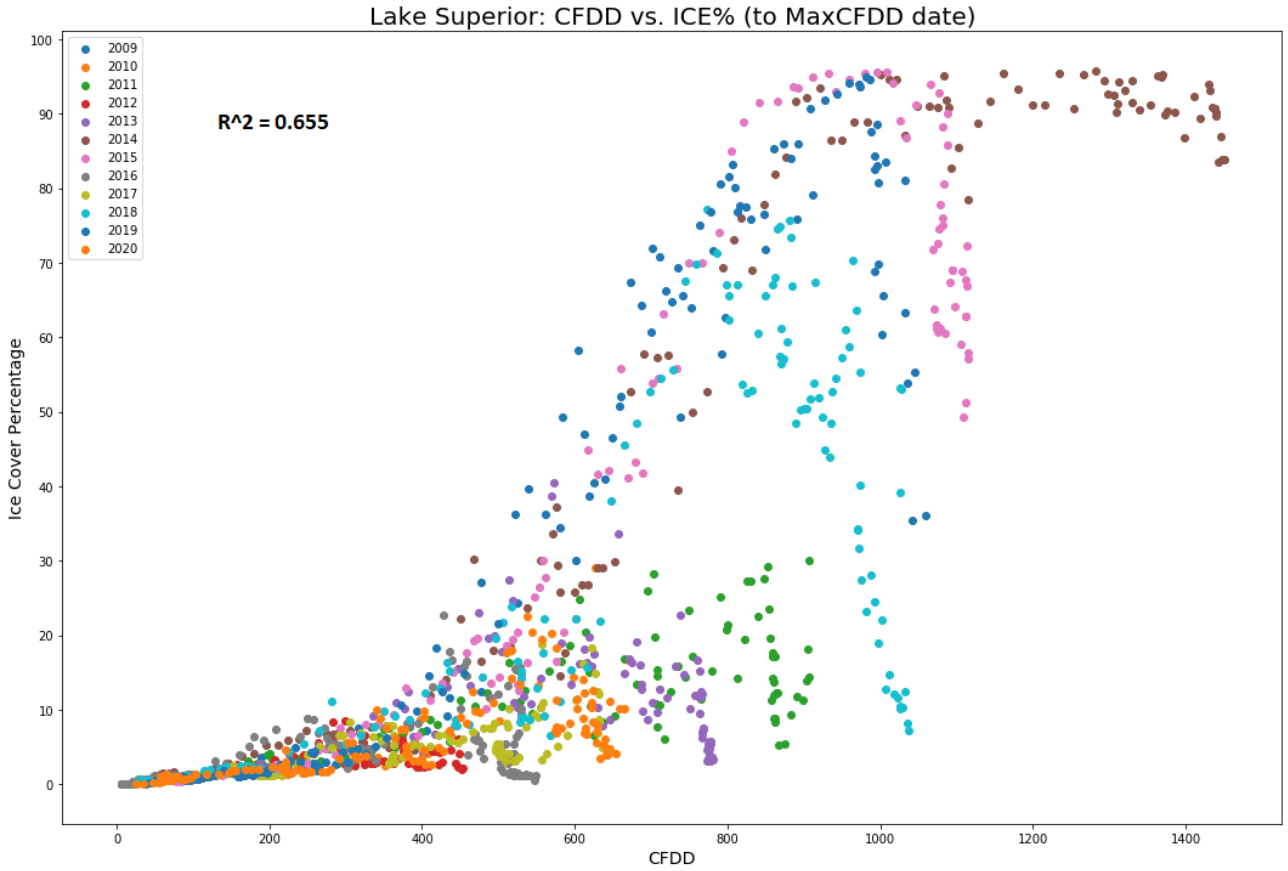

Lake Erie: CFDD vs. ICE% (to MaxCFDD date)
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Ice Cover Percentage
CFDD
R^2 = 0.752

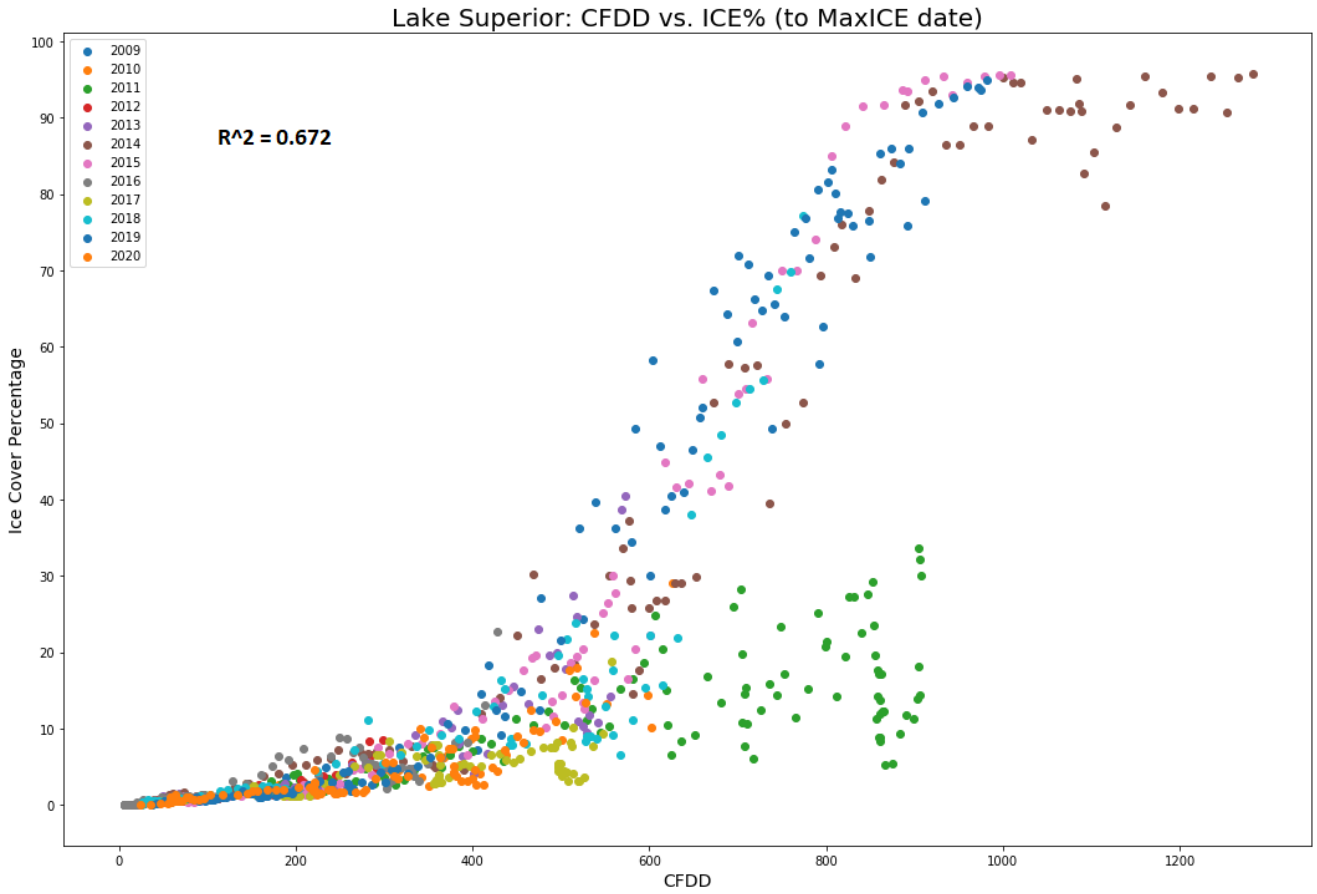

Lake Erie: CFDD vs. ICE% (to MaxICE date)
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Ice Cover Percentage
CFDD
R^2 = 0.797

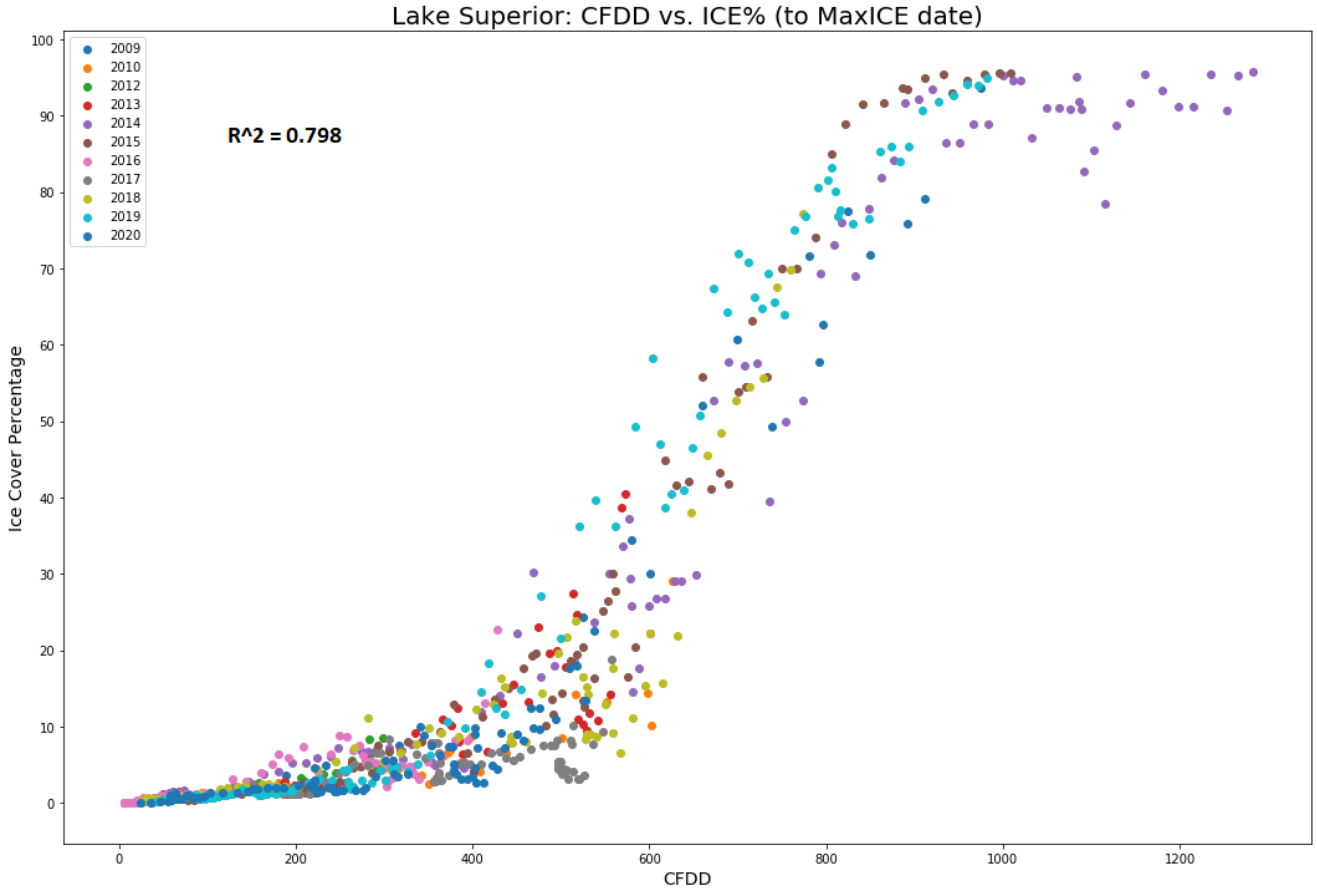

Lake Erie: CFDD vs. ICE% (to MaxICE date w/o 3/06/2014)
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Ice Cover Percentage
CFDD
R^2 = 0.867

B. **Lake Superior**

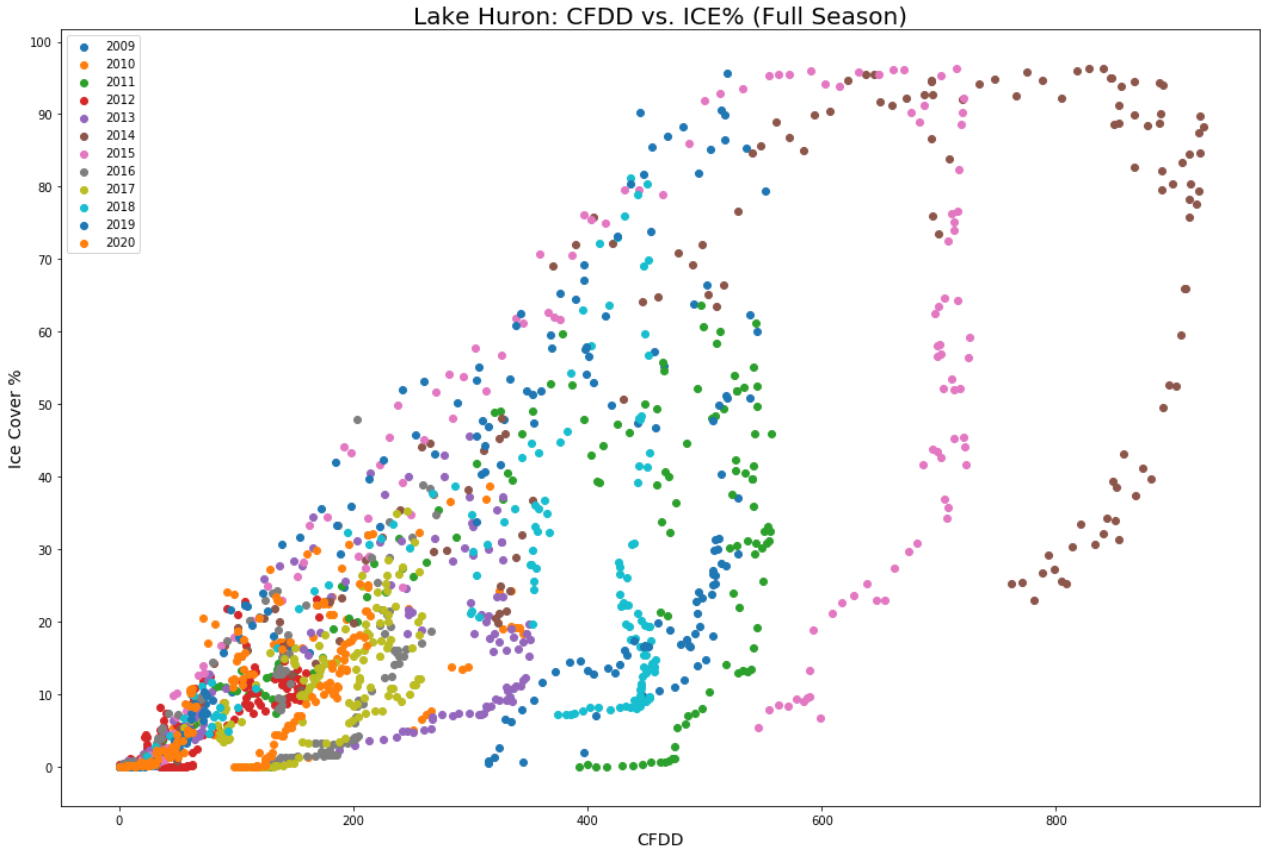


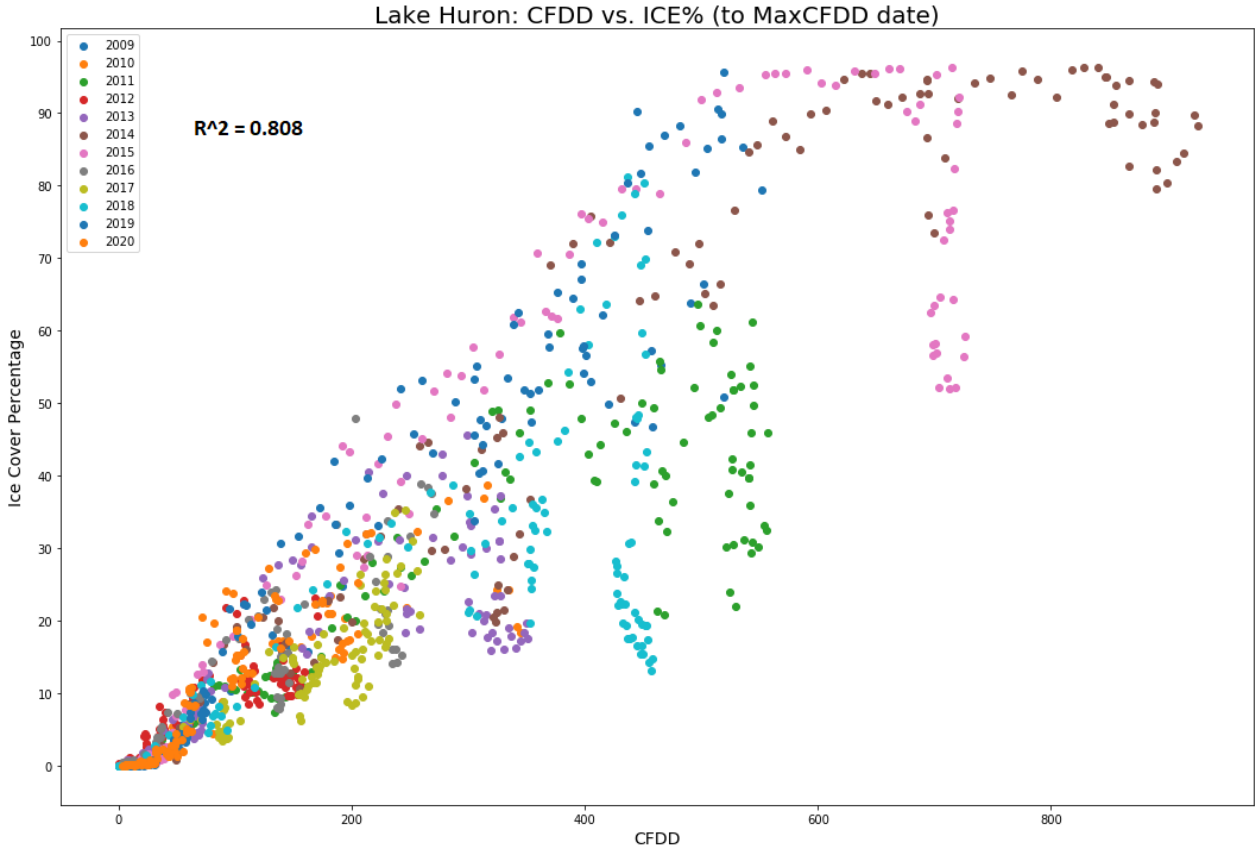


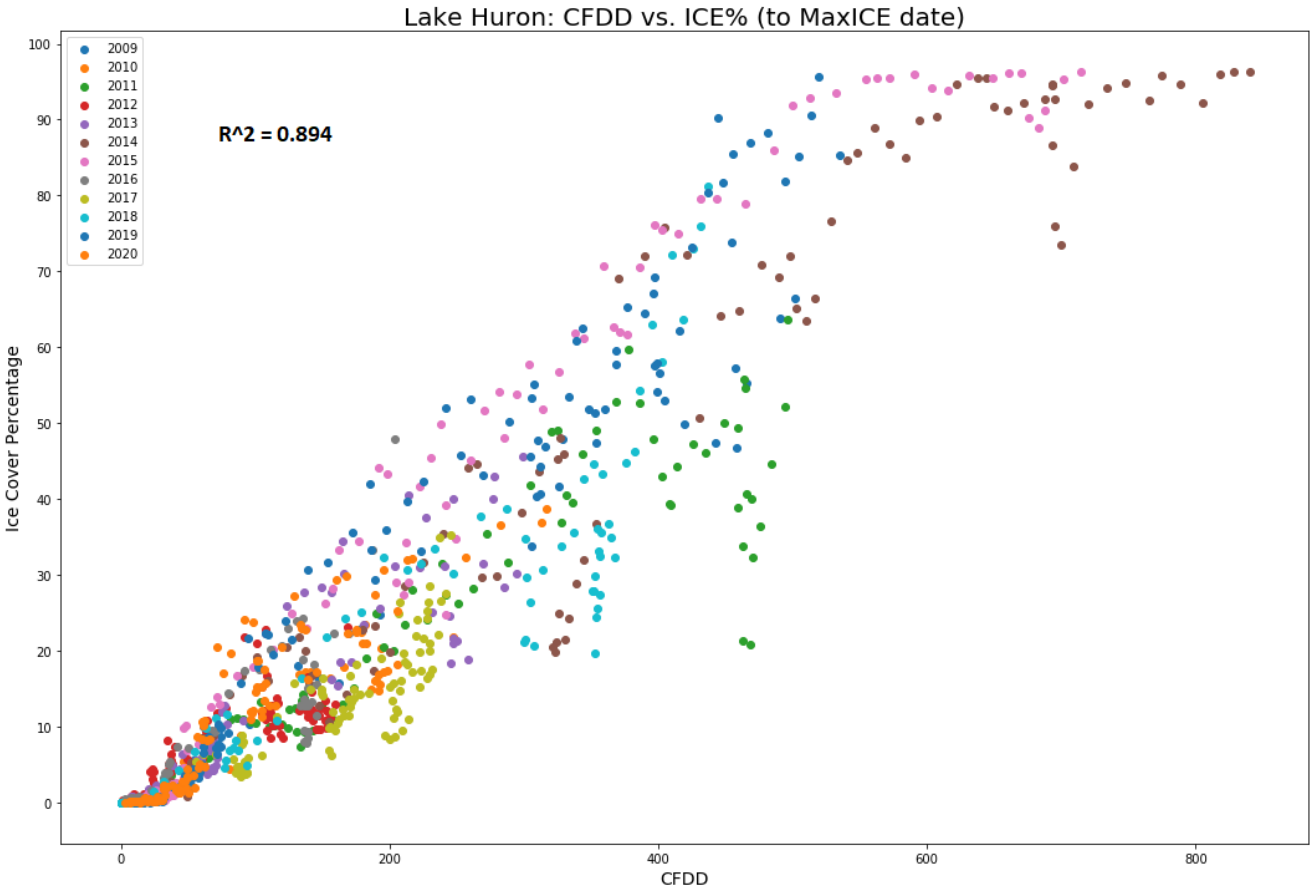

Lake Superior: CFDD vs. ICE% (to MaxICE date)
R^2 = 0.672
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Ice Cover Percentage
CFDD
0
10
20
30
40
50
60
70
80
90
100
0
200
400
600
800
1000
1200


Lake Superior: CFDD vs. ICE% (to MaxICE date)
R^2 = 0.798
2009
2010
2012
2013
2014
2015
2016
2017
2018
2019
2020
Ice Cover Percentage
CFDD
0
10
20
30
40
50
60
70
80
90
100
0
200
400
600
800
1000
1200

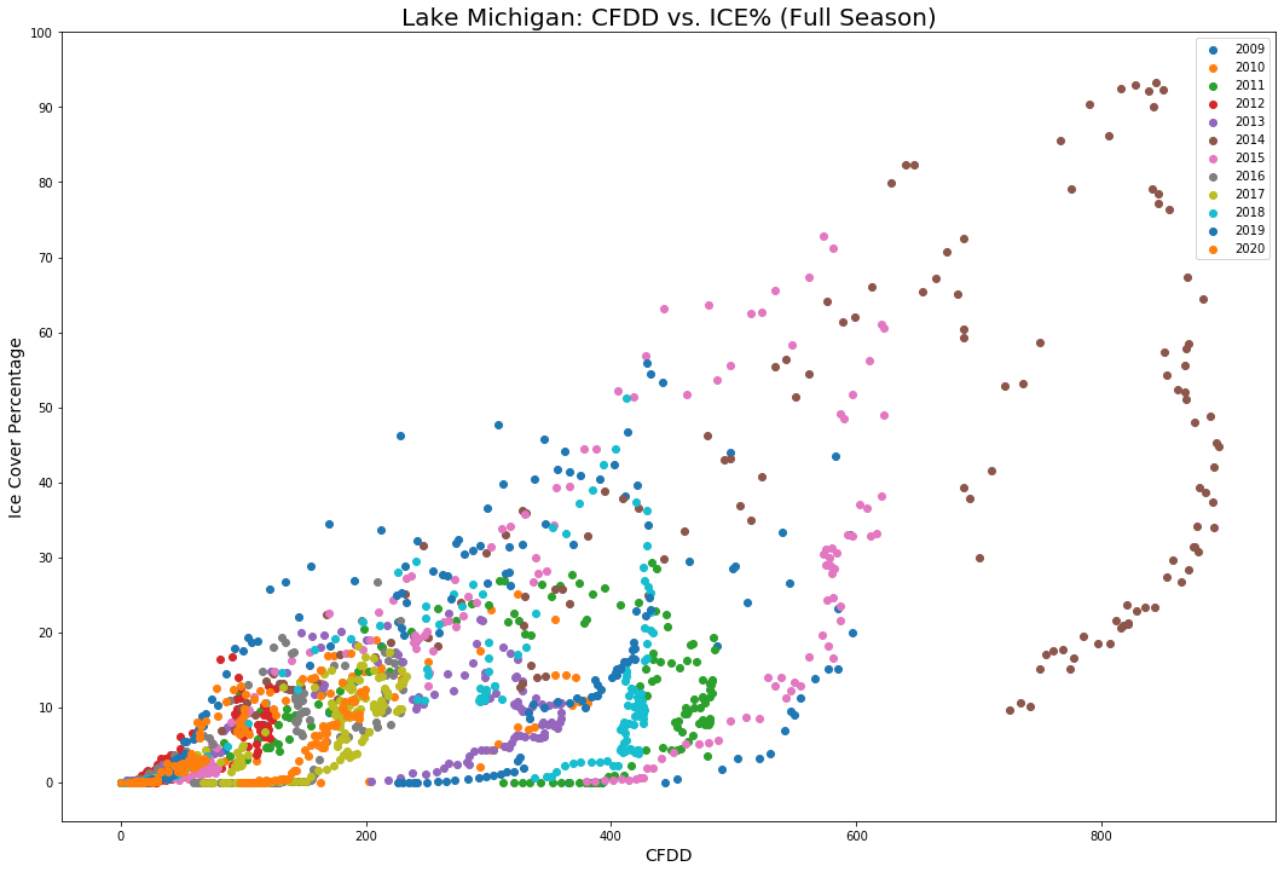
C. Lake Huron

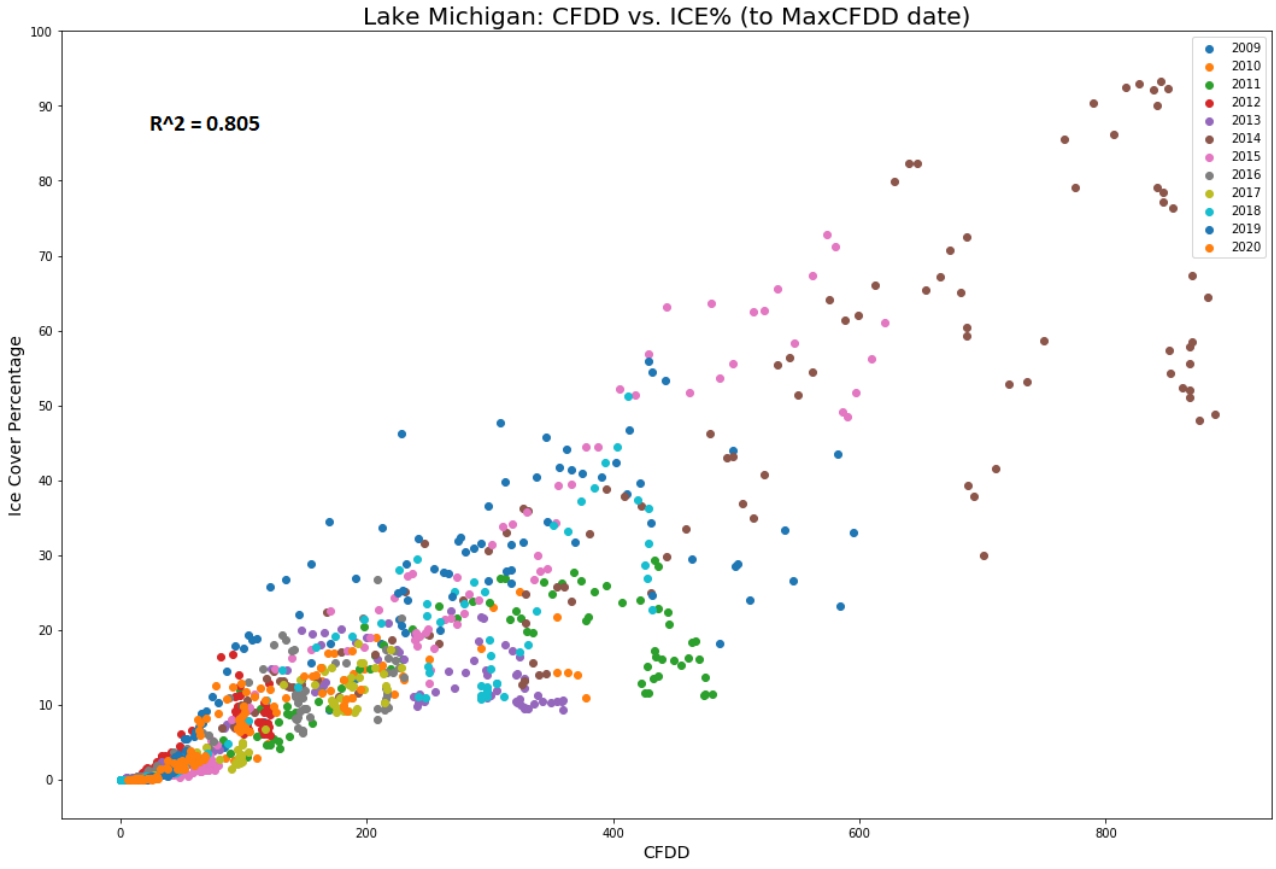

Lake Huron: CFDD vs. ICE% (Full Season)
Ice Cover %
CFDD
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020

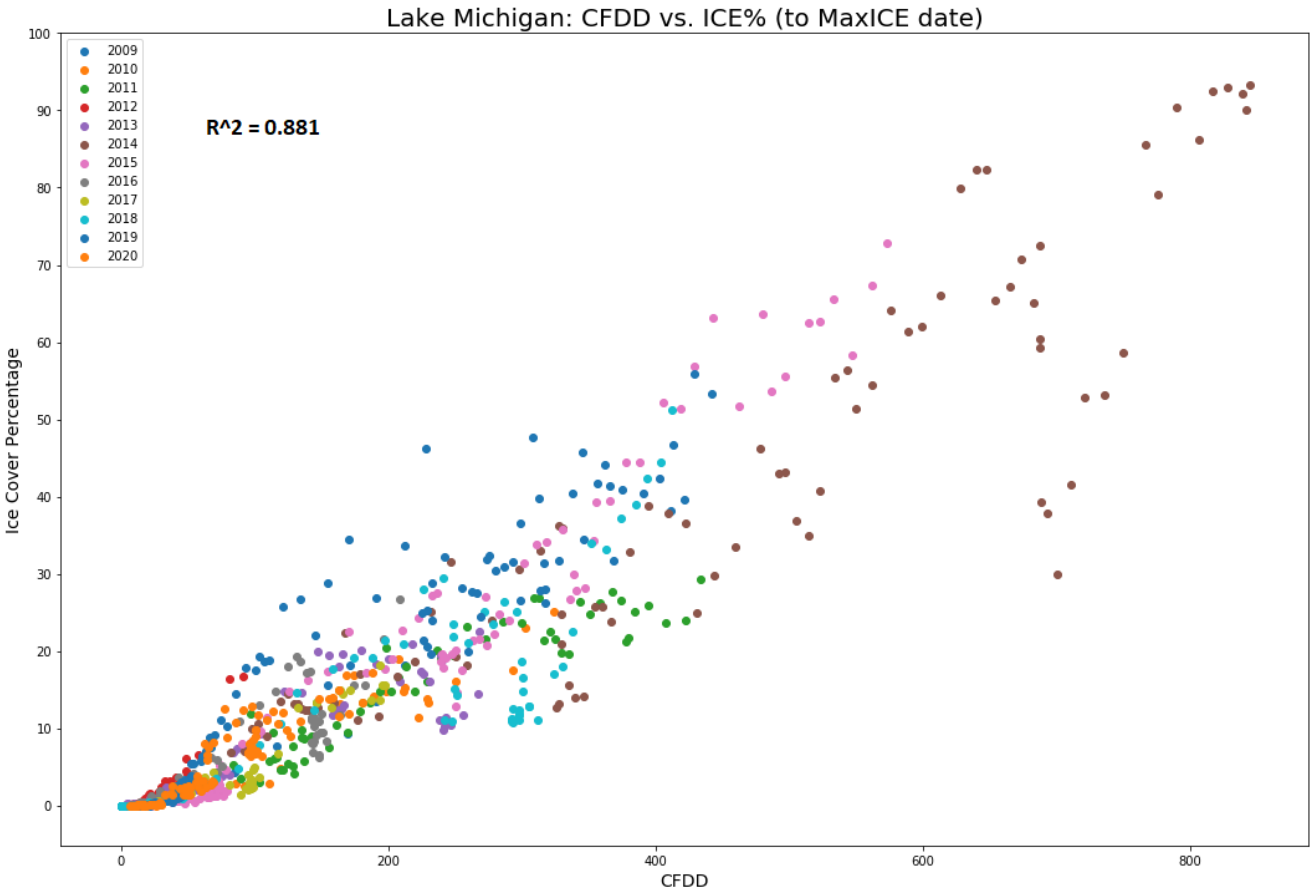

Lake Huron: CFDD vs. ICE% (to MaxCFDD date)
Ice Cover Percentage
CFDD
R^2 = 0.808
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020


Lake Huron: CFDD vs. ICE% (to MaxICE date)
R^2 = 0.894
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Ice Cover Percentage
CFDD
0
10
20
30
40
50
60
70
80
90
100
0
200
400
600
800

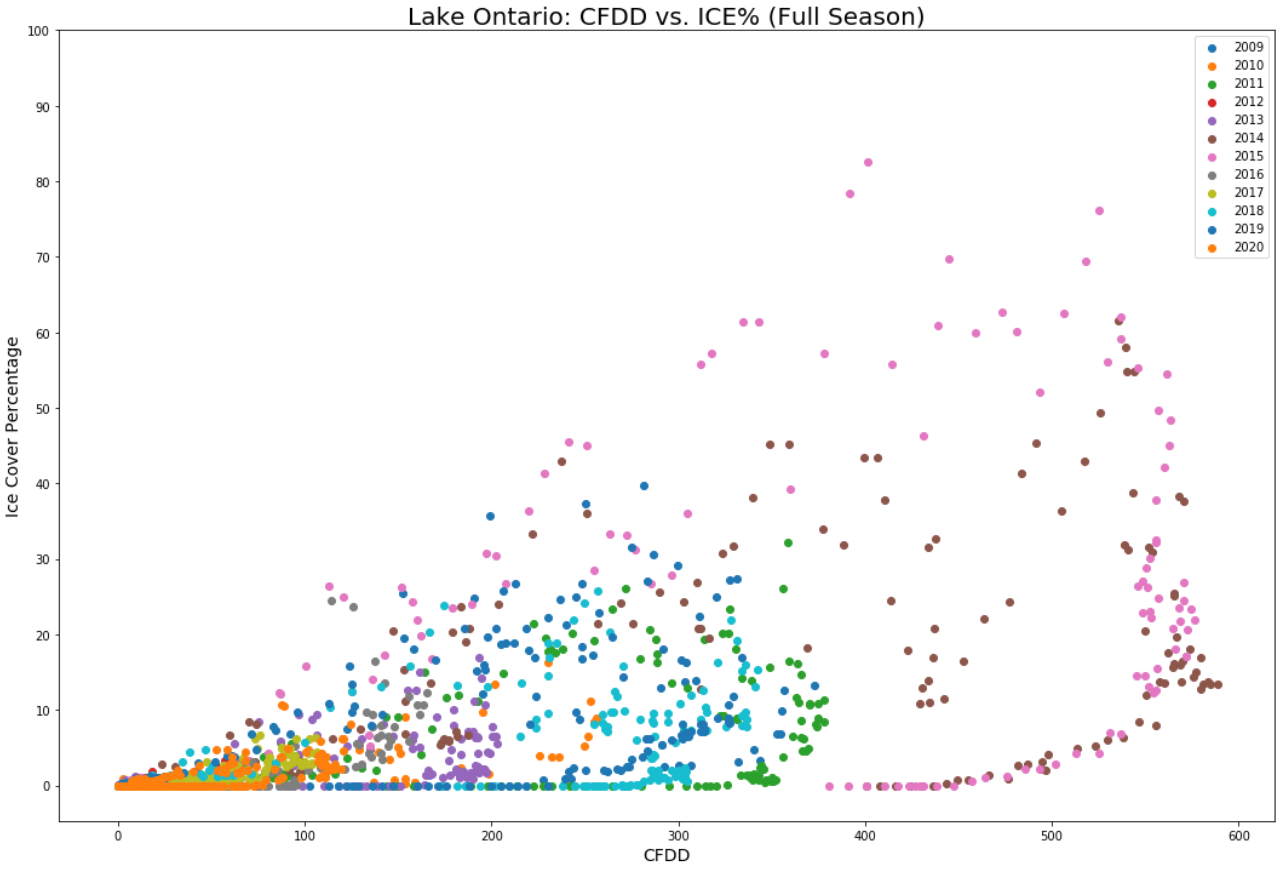
D. **Lake Michigan**

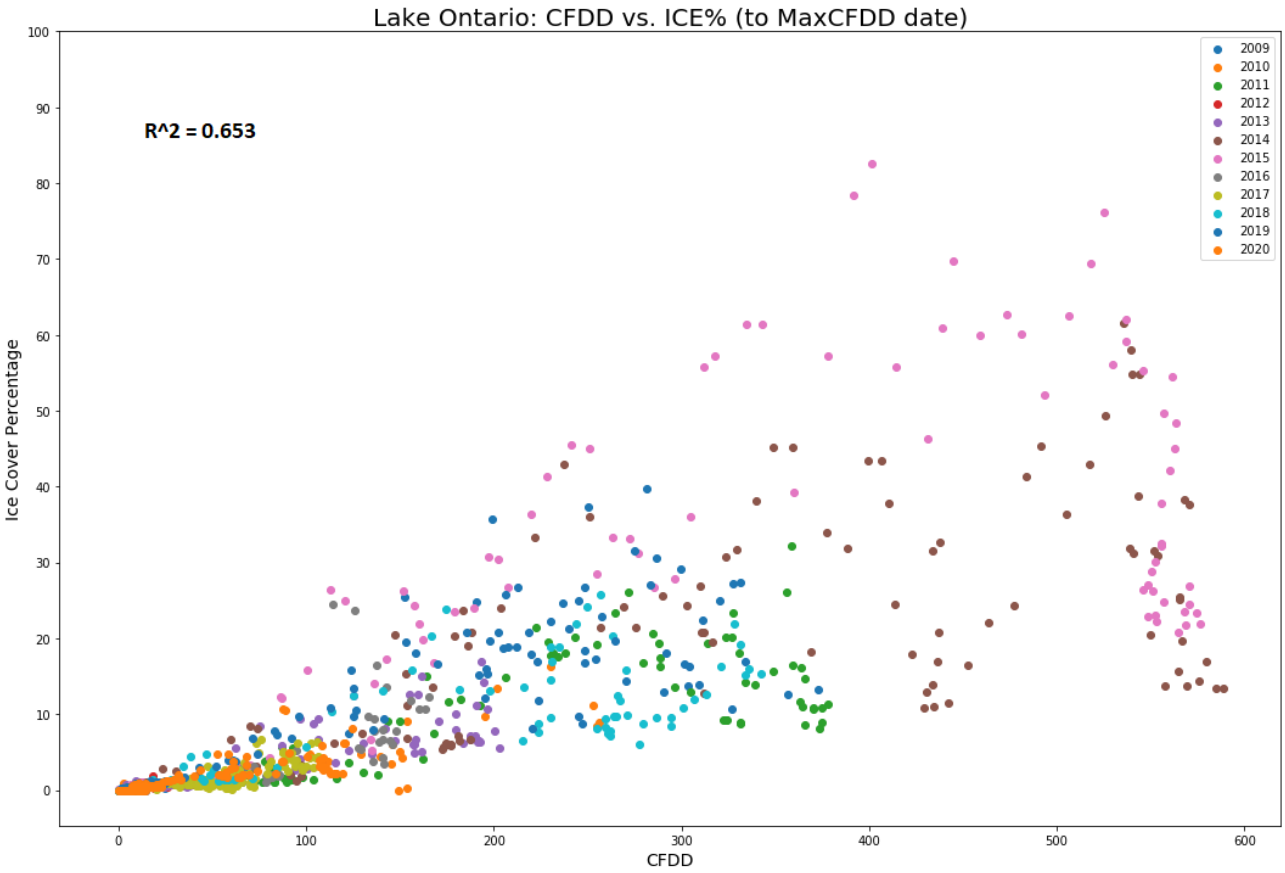


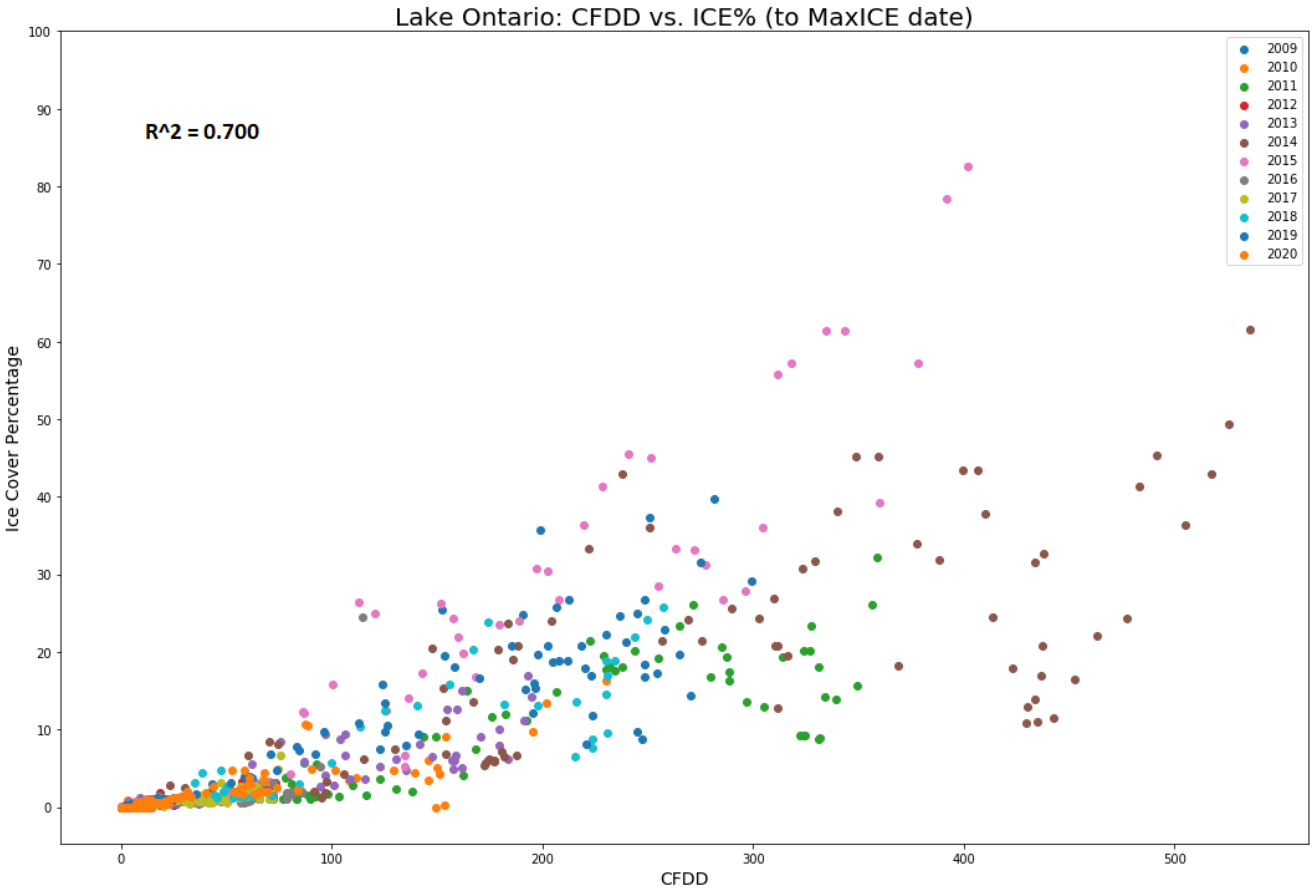



Lake Michigan: CFDD vs. ICE% (to MaxICE date)
R^2 = 0.881
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Ice Cover Percentage
CFDD
0
10
20
30
40
50
60
70
80
90
100
0
200
400
600
800

E. **Lake Ontario**



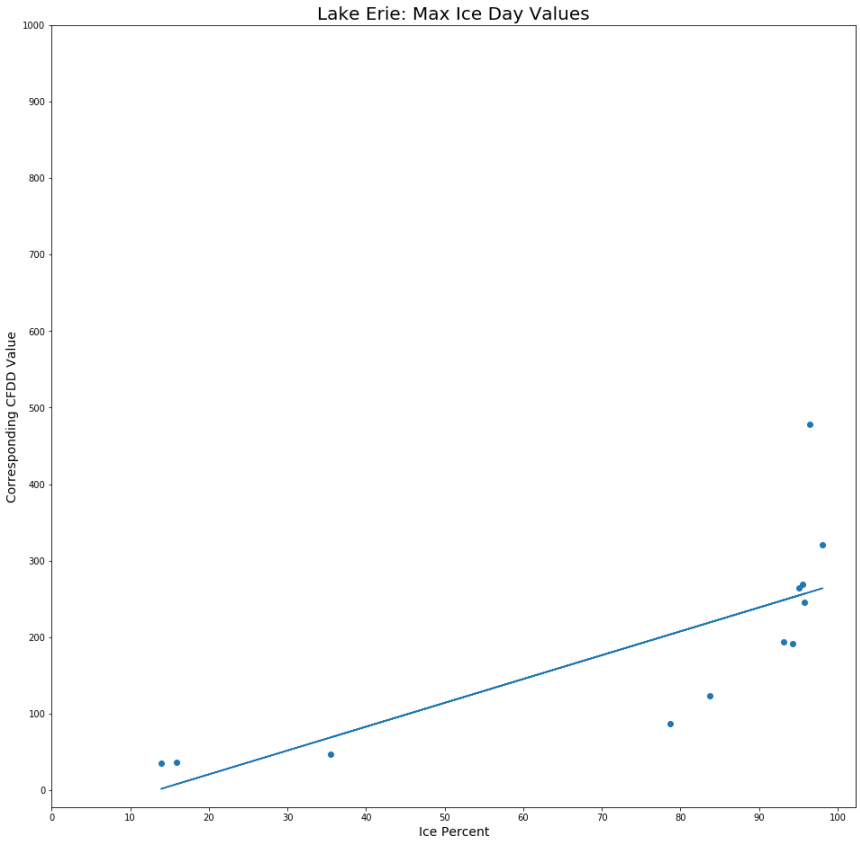


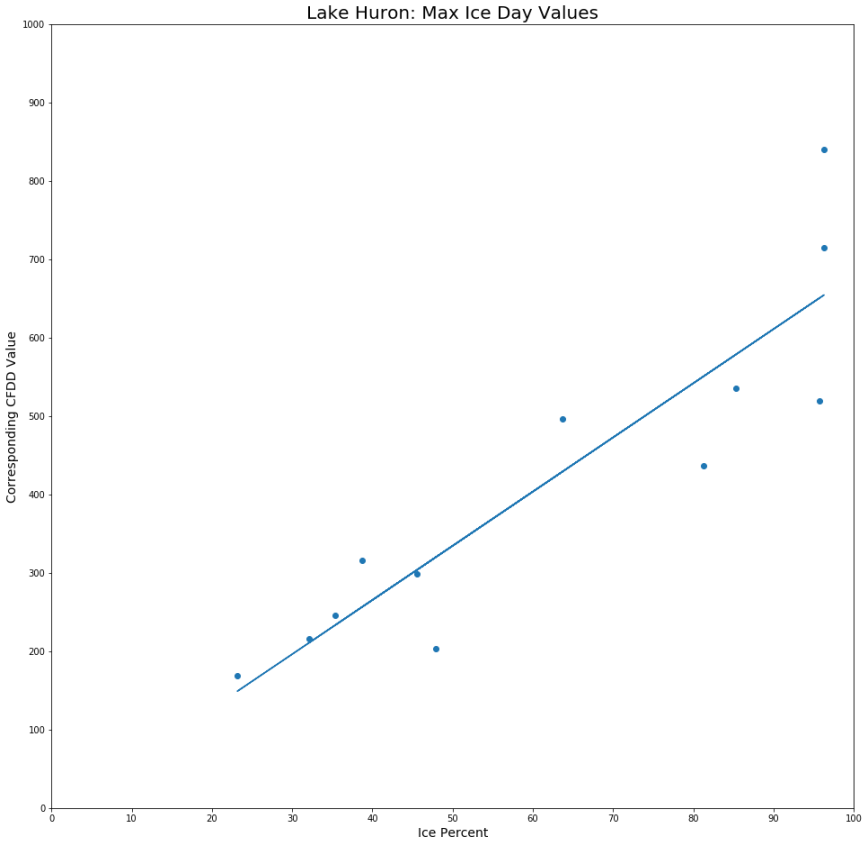

Lake Ontario: CFDD vs. ICE% (to MaxICE date)
R^2 = 0.700
Ice Cover Percentage
CFDD
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020

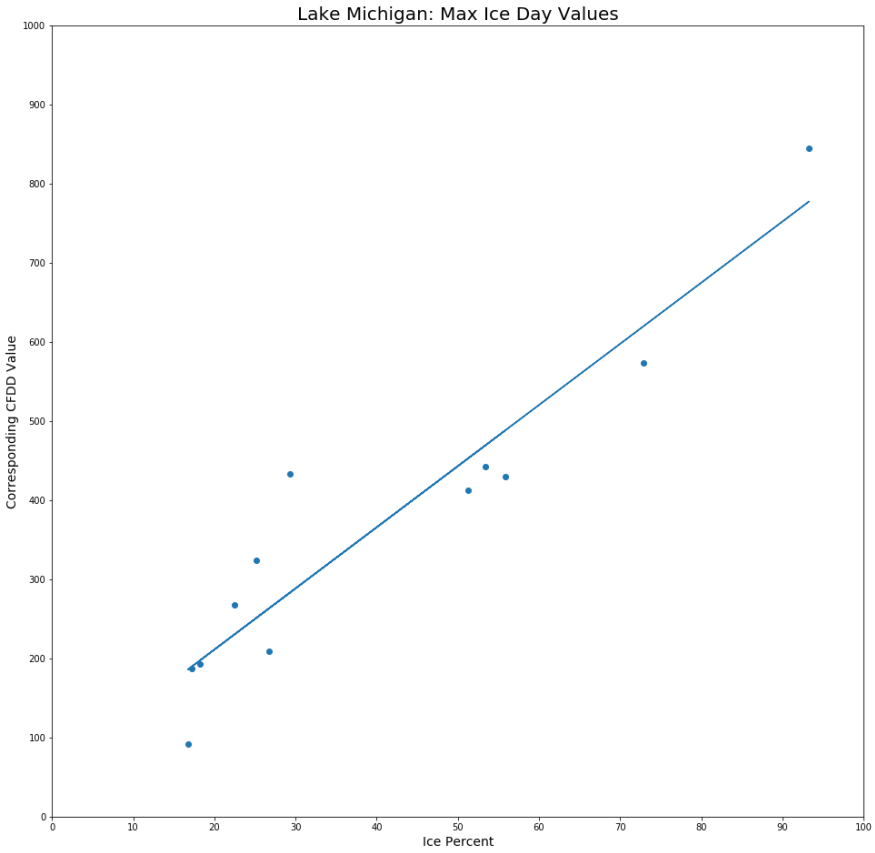
A. Ice Cover Percentage vs. CFDD for Maximum Ice Cover Dates for each year

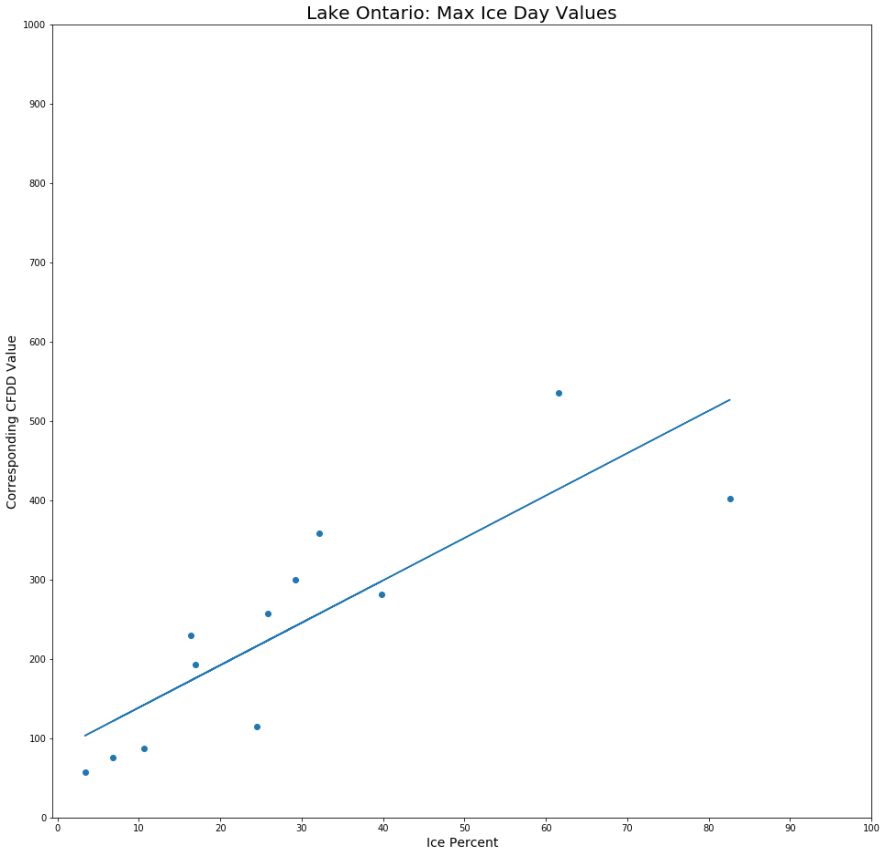
R Squared Values:

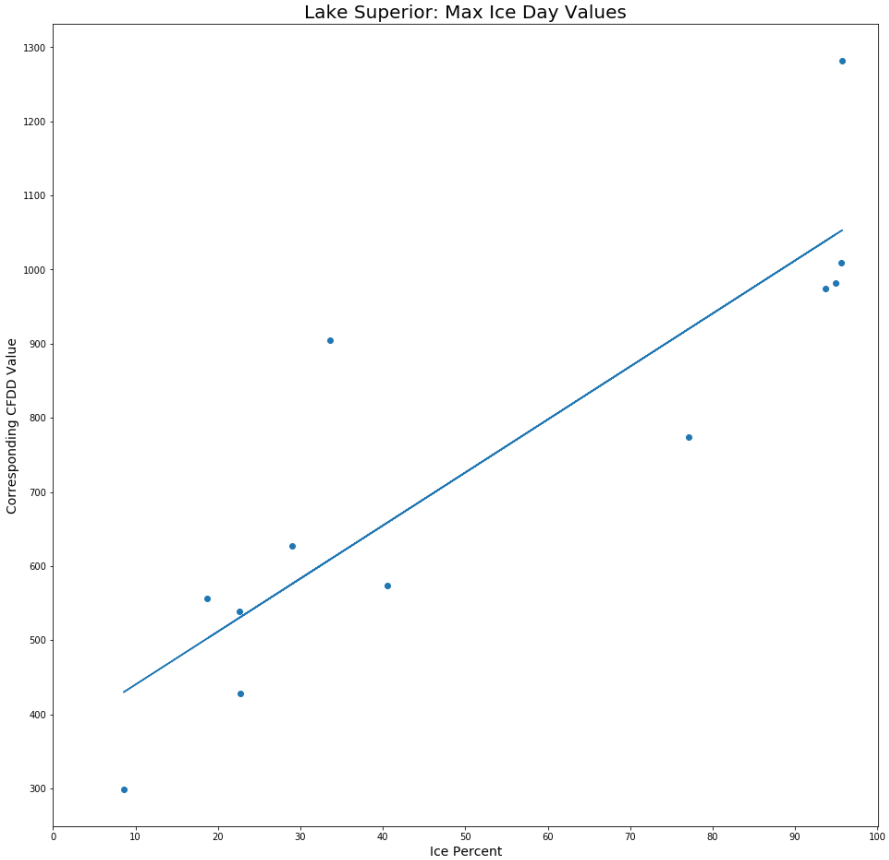
ONT	0.688
MICH	0.872
HUR	0.796
SUP	0.745
ERI	0.536



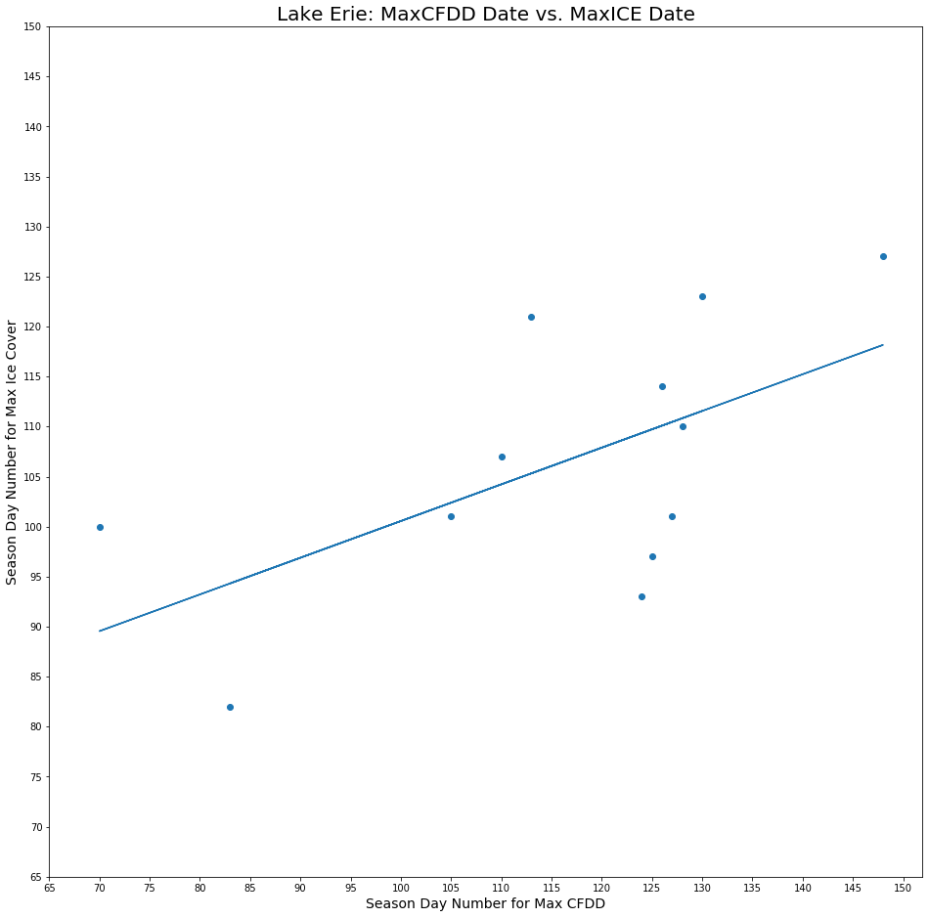

Lake Huron: Max Ice Day Values
Corresponding CFDD Value
Ice Percent
0
100
200
300
400
500
600
700
800
900
1000
0
10
20
30
40
50
60
70
80
90
100


Lake Michigan: Max Ice Day Values
Corresponding CFDD Value
Ice Percent
0
100
200
300
400
500
600
700
800
900
1000
0
10
20
30
40
50
60
70
80
90
100


Lake Ontario: Max Ice Day Values
Corresponding CFDD Value
Ice Percent
0
100
200
300
400
500
600
700
800
900
1000
0
10
20
30
40
50
60
70
80
90
100

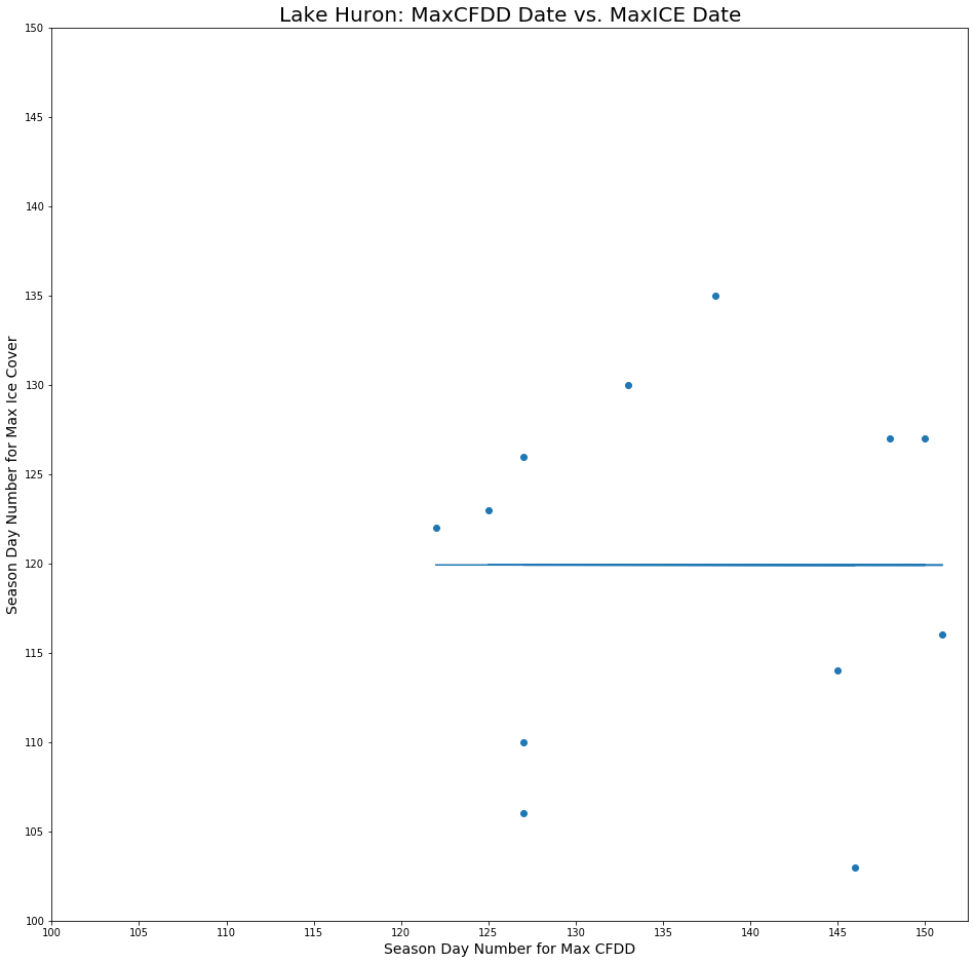

Lake Superior: Max Ice Day Values
Corresponding CFDD Value
Ice Percent
300
400
500
600
700
800
900
1000
1100
1200
1300
0
10
20
30
40
50
60
70
80
90
100

B. **Maximum Ice Cover Percentage Date vs. Maximum CFDD Date for each year**



	ERIE		
	Season Day of Max:		
	MaxCFDD	MaxICE	Difference
YEAR			
2009	125	97	28
2010	127	101	26
2011	124	93	31
2012	83	82	1
2013	126	114	12
2014	148	127	21
2015	128	110	18
2016	110	107	3
2017	70	100	-30
2018	105	101	4
2019	130	123	7
2020	113	121	-8

Average Difference: 9.416666666666666
R^2 Value = 0.288



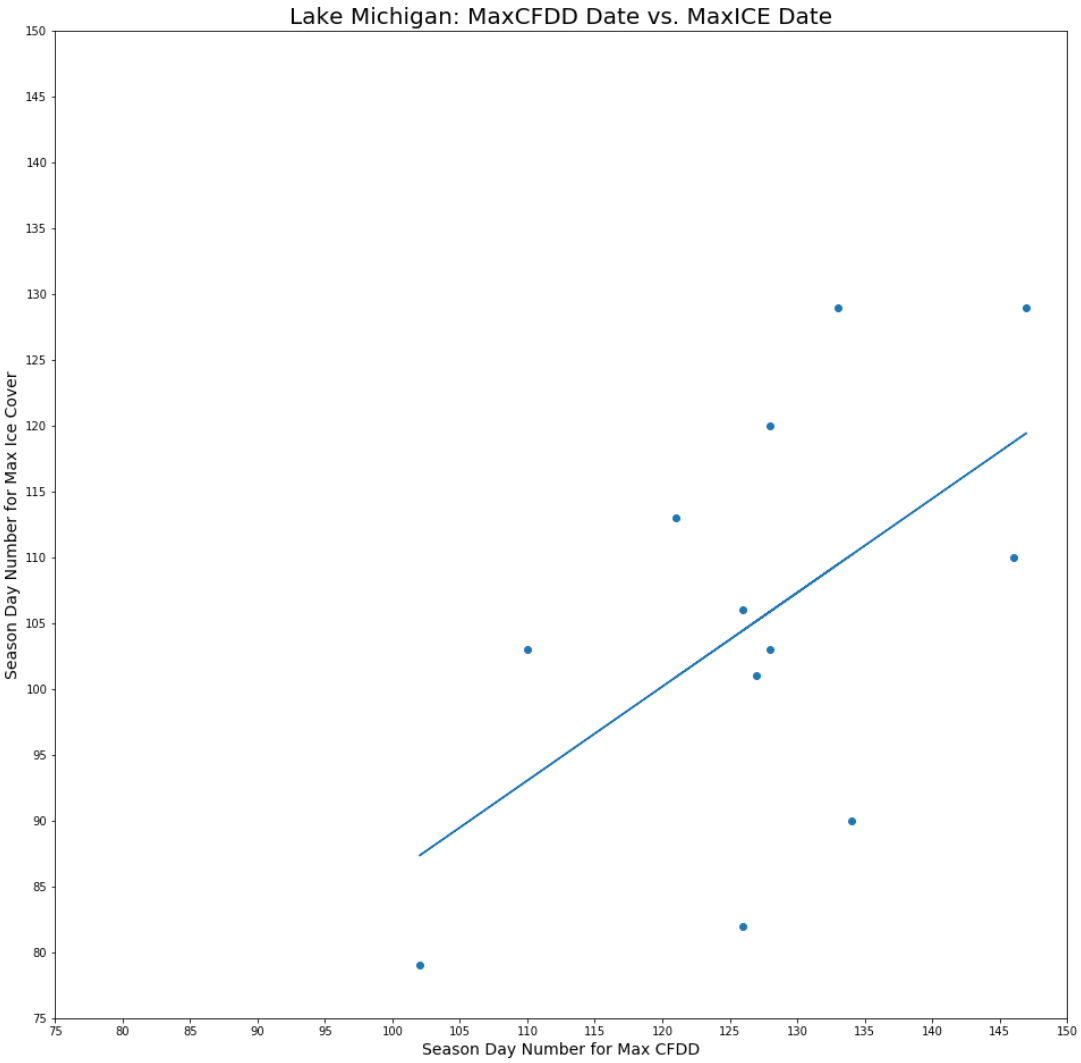
HURON

Season Day of Max:

YEAR	MaxCFDD	MaxICE	Difference
2009	125	123	2
2010	127	110	17
2011	151	116	35
2012	127	126	1
2013	145	114	31
2014	148	127	21
2015	150	127	23
2016	127	106	21
2017	138	135	3
2018	146	103	43
2019	133	130	3
2020	122	122	0

Average Difference: 16.666666666666668

R^2 Value = -0.100



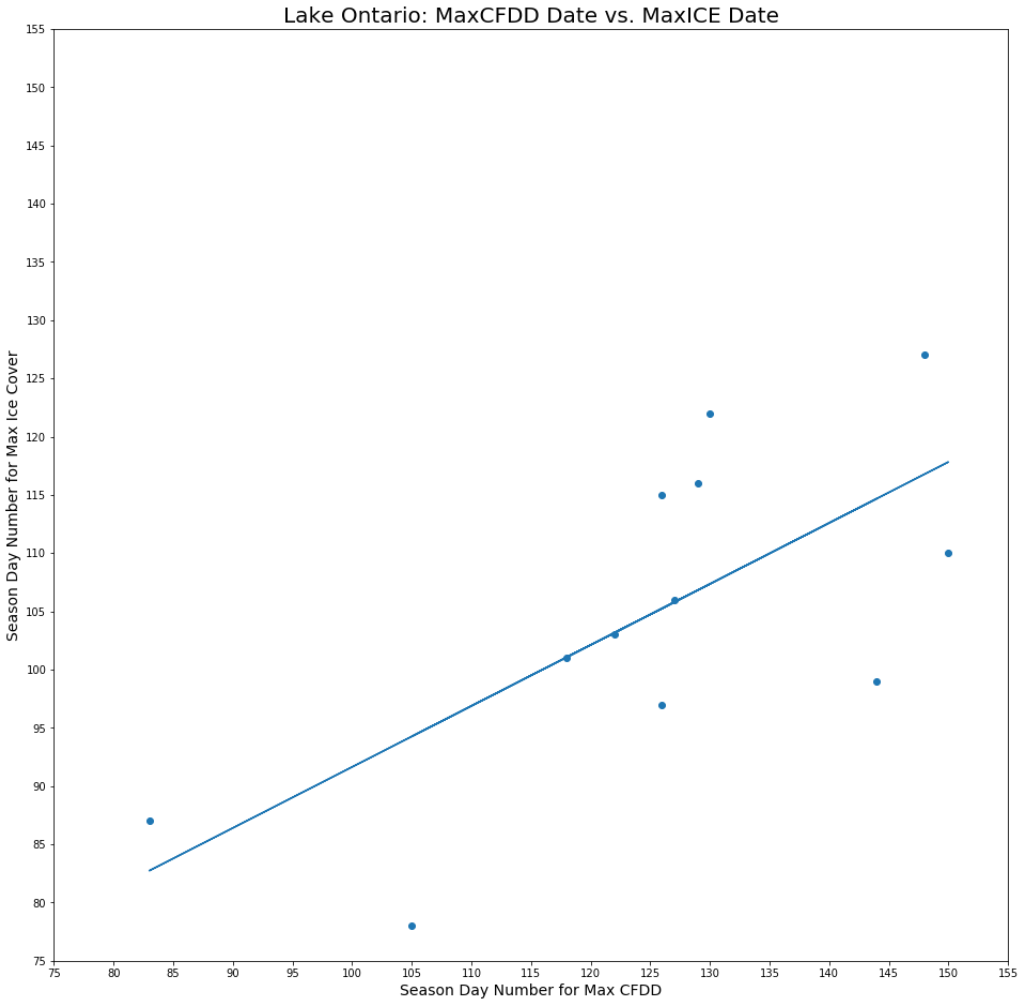
MICHIGAN

Season Day of Max:

YEAR	MaxCFDD	MaxICE	Difference
2009	134	90	44
2010	127	101	26
2011	128	103	25
2012	126	82	44
2013	146	110	36
2014	147	129	18
2015	128	120	8
2016	126	106	20
2017	102	79	23
2018	110	103	7
2019	133	129	4
2020	121	113	8

Average Difference: 21.916666666666668

R^2 Value = 0.245



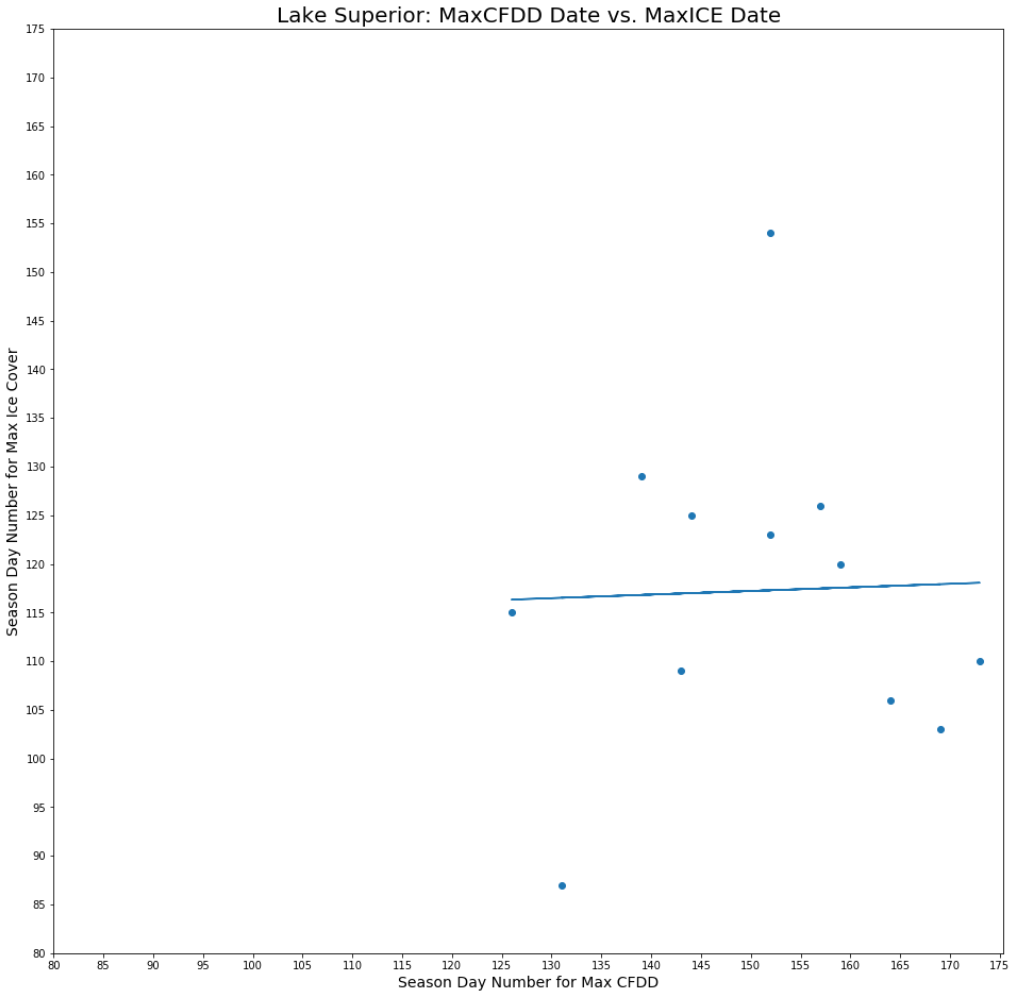
ONTARIO

Season Day of Max:

YEAR	MaxCFDD	MaxICE	Difference
2009	126	97	29
2010	118	101	17
2011	129	116	13
2012	83	87	-4
2013	126	115	11
2014	148	127	21
2015	150	110	40
2016	127	106	21
2017	144	99	45
2018	105	78	27
2019	130	122	8
2020	122	103	19

Average Difference: 20.583333333333332

R^2 Value = 0.420



SUPERIOR

	Season Day of Max:		
YEAR	MaxCFDD	MaxICE	Difference
2009	152	123	29
2010	126	115	11
2011	152	154	-2
2012	131	87	44
2013	173	110	63
2014	157	126	31
2015	159	120	39
2016	164	106	58
2017	144	125	19
2018	169	103	66
2019	139	129	10
2020	143	109	34

Average Difference: 33.5
R^2 Value = -0.099

A. Monthly ABNA Index from 1980 - 2020

Years	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov
1980	2.15899	-0.1091	-1.5754	2.22332	-0.0869	-2.4445	-0.064	0.58388	-0.9634	-0.8417	-1.074	-0.5538
1981	0.34711	-0.4649	1.28398	0.40358	-1.6352	1.02684	0.93333	1.18407	-0.5304	-0.3454	-1.4422	-1.3538
1982	0.78721	-0.9186	-1.6005	0.82301	-0.4336	-1.1309	-1.6797	-0.4091	0.28205	-0.3431	-0.3722	-0.2386
1983	1.09128	1.55862	1.50947	0.43133	0.76488	0.71733	-0.1385	0.1048	0.39691	0.13281	-0.4119	0.34308
1984	-2.3316	1.0251	1.66504	-2.0817	0.40236	1.21671	1.06282	0.91359	0.37065	-2.0075	-2.192	-0.6785
1985	-1.0072	-2.8614	1.43699	0.01564	-3.5916	1.67187	-0.0642	-1.3248	0.1906	-2.0252	-1.1593	-3.4793
1986	-2.8044	0.83584	-0.5732	-3.4658	0.24448	-0.8766	-0.3343	-0.9951	1.15902	-1.7427	-2.3547	-2.974
1987	0.41238	0.48978	-0.5017	-0.411	-0.1705	-0.7838	0.83627	-2.3418	-2.3024	-0.5884	0.09212	0.00174
1988	-0.3825	-0.8611	0.21624	-0.9221	-1.2192	-0.4781	1.06033	1.78786	1.56926	-0.0683	0.05062	-0.3719
1989	-0.4723	2.20445	-2.7957	-0.745	2.43088	-1.9687	0.7564	0.2586	-1.318	-0.1775	-0.1679	1.84588
1990	-1.7775	-0.0932	2.53828	-2.2612	0.07013	3.23205	0.67865	-0.6139	-0.5531	0.60924	2.47328	2.09399
1991	1.49507	0.04463	0.56259	2.21928	-0.3989	-0.0641	-1.6903	-2.1994	-0.3813	-0.9956	0.27276	0.73677
1992	1.47325	1.20595	-0.702	1.05226	0.37929	-0.8868	0.06124	-3.5217	-0.4909	0.41577	1.29385	-1.35
1993	0.13386	1.16968	-0.1662	0.78318	1.39808	-0.5191	-1.3553	-1.7209	-0.3657	-1.0738	-1.3186	-0.6649
1994	1.17302	-2.1078	-1.3075	0.67915	-1.9215	-0.8925	0.18183	-0.1155	-1.3108	0.87004	0.83131	0.8923
1995	0.8808	0.35245	-0.1209	0.48946	-0.0661	-0.5684	-0.6988	-1.3529	-1.5698	-1.4769	1.27669	-0.3483
1996	-0.2891	-0.3014	1.02335	-0.7494	-0.3008	1.21202	-0.1215	0.66122	1.17276	-0.722	-0.1035	-1.9503
1997	1.01782	0.87276	-0.6763	1.66695	0.44589	-0.9965	-0.2633	-0.9465	-1.7022	-1.2167	-0.4996	0.69628
1998	1.28795	0.42764	2.34823	0.4415	-0.0991	1.81323	0.56314	1.12688	0.81661	0.35728	0.52111	-2.0931
1999	-1.4803	0.8984	1.94717	-1.4184	0.78266	1.94487	0.47106	-0.1557	0.3662	1.11502	0.5831	0.95984
2000	1.10691	2.77806	-0.575	0.94704	3.09035	-1.0816	-0.7101	0.12979	1.09418	0.46451	-0.4408	-0.5533
2001	-2.0084	0.03683	-2.3025	-2.6097	-0.7697	-1.9999	-2.1526	-0.2471	0.84892	0.0172	1.44907	0.19537
2002	0.38619	-0.101	2.02333	0.17634	-0.0515	1.80355	-0.4718	0.91861	0.39198	0.12655	-2.7108	-2.9571
2003	-0.6573	-0.6019	-0.7658	-1.5287	-1.3463	-1.0853	-0.2958	1.56609	-0.3434	1.03817	0.70917	0.29324
2004	1.4969	-1.3715	0.50748	0.97405	-1.6358	-0.0116	-2.573	-0.5693	-1.7801	1.15477	1.43563	1.73267
2005	0.89752	-0.0573	1.35342	0.73749	-0.0774	1.24498	-1.2055	1.84096	-0.3339	1.19035	1.47778	1.26709
2006	-0.4838	2.01213	0.03793	-1.113	1.65419	0.09701	1.91109	0.78929	0.51792	-0.1794	-0.6177	1.5807
2007	1.9288	1.33593	-1.0715	0.70393	0.80222	-1.0758	0.43608	-0.9307	-0.5864	0.89431	0.73763	-0.2715
2008	-0.4894	0.21436	0.58836	-0.587	0.49112	0.21727	0.72788	2.4408	0.95339	0.43465	2.01367	1.50151
2009	0.19402	0.93105	-0.6899	0.92871	0.43701	-0.0163	0.28115	-1.4164	-0.4815	0.03258	-2.3073	2.54946
2010	-1.0121	0.81068	0.16394	-1.2541	-0.0069	-0.1798	1.18194	1.99814	0.86513	-0.3643	1.39164	0.87325
2011	1.77977	-0.4072	-0.4901	2.63034	-0.973	0.29972	-0.2693	1.23702	1.38293	0.57831	1.55931	0.88849
2012	0.8353	2.55655	-0.6302	0.68612	2.0716	-0.9415	0.84547	0.48212	1.97012	3.00114	-0.1437	0.47321
2013	0.26434	-1.3953	0.65003	0.83601	-1.559	0.41866	-0.1838	0.5776	-0.8433	1.19558	-0.5068	-0.0591
2014	-1.6112	-0.8614	0.49449	-1.0373	-1.3364	0.84104	1.46437	1.4901	0.6737	-0.145	1.43977	-0.9489
2015	0.52425	0.86284	1.70105	0.22944	0.4831	1.23619	-0.9518	-0.7138	1.31546	1.07724	-0.5111	2.05037
2016	1.96746	-0.2264	1.24658	1.35107	-1.4428	0.34325	-0.0057	1.45774	-0.5326	0.53993	-2.2335	1.36918
2017	-0.8421	1.25019	0.64529	-0.6455	1.06995	0.5321	-0.9172	0.44586	0.47529	-0.7429	-0.3359	-1.0962
2018	-1.0804	0.34564	-1.2728	-1.4607	0.13818	-0.5164	0.03936	2.46886	0.70862	-0.3149	-0.8613	-0.7132
2019	0.41753	0.61605	-3.1479	-0.1046	0.23405	-2.3527	-0.9273	-0.0809	0.03167	0.33018	1.08041	-1.2831
2020	0.75694	1.44614	1.53039	0.64252	1.50796	1.33305	2.1425	1.14761	0.45217	2.21655	0.45151	1.11705

B. **Monthly New Index from 1980 - 2020**

Years	D	J	F	M	A	M	J	J	A	S	O	N
1980	0.76347	-0.2073	-0.9862	0.79613	-0.3508	-1.0939	-0.4154	0.45474	-0.5306	-0.1516	-1.0976	-0.7614
1981	-0.345	-0.3637	1.71989	-0.4207	-0.3941	1.53708	0.98604	0.34236	0.25424	0.37916	-0.2477	0.30183
1982	0.57548	-0.5636	-1.4368	0.63888	-0.4136	-1.2898	-0.8672	-0.3008	-0.8263	-0.3617	-0.5266	0.0494
1983	0.63922	1.254	1.8984	0.33933	0.99114	1.69639	0.77807	0.82474	0.3235	0.30042	1.10475	-0.1364
1984	-1.543	0.51629	2.30036	-1.5714	0.33439	2.13071	0.25254	0.90239	0.80849	-0.9439	-1.1918	0.42829
1985	0.26072	-2.6006	0.73144	0.44686	-2.5551	0.80814	0.62536	-0.3335	-0.2785	-0.6616	-0.2743	-2.123
1986	-2.3089	1.25849	-0.3571	-2.3898	1.1077	-0.4471	-0.3286	-0.5976	0.99374	-1.572	-1.4981	-1.6024
1987	-0.0928	0.78105	-0.3129	-0.2982	0.51424	-0.2705	0.25014	-0.7623	-1.6691	-0.6845	0.20405	-0.0222
1988	-0.5258	-0.4257	0.61733	-0.6567	-0.5955	0.43389	0.3949	0.5856	0.06255	-0.4528	0.62632	0.45749
1989	-0.679	1.33786	-2.6528	-0.6007	1.3386	-2.3309	0.38438	-0.7336	-1.8007	-0.295	-0.1978	0.74895
1990	-2.8859	0.75178	0.09814	-2.896	0.70584	0.16373	0.5656	-1.1635	0.08263	0.76791	1.22378	0.51202
1991	0.87659	-1.3323	-0.0684	1.06498	-1.3905	-0.1115	-1.1982	-0.9365	0.08118	-0.4027	0.22982	0.99046
1992	0.9516	0.63474	-0.6498	0.78517	0.4613	-0.7254	-0.0267	-2.0859	-0.3746	1.61356	1.08525	-1.1376
1993	0.12113	0.29488	-0.094	0.12517	0.24289	-0.308	-0.954	-1.3321	0.30804	-0.4019	-0.706	-0.5997
1994	0.71515	-1.2923	-0.993	0.60208	-1.1893	-0.9978	0.50941	-0.1657	-1.5254	0.3728	0.53883	0.02905
1995	1.28764	-0.2119	-1.2139	1.16252	-0.253	-1.2486	-0.4549	-0.4351	0.01721	-0.753	0.75127	0.0651
1996	-0.4068	0.31967	1.59738	-0.4672	0.23107	1.60375	0.2774	0.49619	1.51082	-0.4145	0.47003	-0.7891
1997	1.19017	0.087	-0.7954	1.33417	-0.0251	-0.8775	-0.0624	-0.4322	-0.6369	-0.4112	0.19836	-1.0951
1998	0.15526	0.69253	1.50289	-0.0483	0.46925	1.51811	-0.5716	-0.0075	0.65078	-0.3504	-0.5996	-0.8294
1999	-0.9321	1.2484	3.009	-0.9826	1.17144	2.87852	0.95318	-0.8652	-0.3564	1.1545	-0.0646	0.91711
2000	0.92365	1.63365	0.25793	0.81713	1.71028	0.16648	-0.3325	0.51764	0.57938	-0.0729	0.50167	0.27893
2001	-1.3043	0.02789	-0.2232	-1.4331	-0.0606	-0.1902	-0.4727	0.04963	0.30321	0.47611	1.0377	0.75723
2002	1.18809	-0.9562	0.76005	1.18965	-0.7794	0.75405	0.35812	0.94851	0.95533	0.13655	-2.186	-1.9109
2003	-0.2903	-1.1	-1.4438	-0.443	-1.1762	-1.2886	-0.4783	0.94405	0.58605	1.31429	0.45379	-0.1115
2004	1.64627	-1.9083	0.44991	1.50291	-1.9523	0.48801	-2.8439	0.10862	-0.6051	0.59008	1.43421	0.74585
2005	0.71537	-0.166	1.74558	0.66975	-0.1941	1.60476	-0.9647	1.68423	-0.7624	0.48628	0.7855	1.1035
2006	0.11098	1.82001	-0.3245	0.05368	1.65914	-0.1809	1.47893	0.83621	-0.4712	-1.2131	-0.8704	0.67433
2007	0.74887	0.24682	-1.743	0.51789	0.09681	-1.7037	0.14364	-0.2857	-0.733	0.21893	0.60865	0.59502
2008	-0.5675	1.21326	0.6824	-0.5809	1.28474	0.48075	0.62536	1.6481	0.87297	0.24056	1.46386	1.35659
2009	-0.3756	-0.9954	0.18677	-0.1498	-1.01	0.37581	-0.1263	-0.995	-0.2384	0.05154	-1.7993	1.93686
2010	-1.3957	0.16574	-0.4813	-1.521	-0.0354	-0.4239	-0.5474	0.71411	-0.1229	-0.3229	0.15818	0.15886
2011	1.03159	-0.1816	-1.2598	1.14776	-0.1363	-1.1049	0.48784	0.48724	1.02362	0.30885	1.05313	0.67937
2012	1.1284	2.0726	0.3209	1.097	1.91296	0.3024	0.42816	0.05259	0.38065	1.15578	-0.1997	0.18613
2013	0.90508	-1.4281	0.93681	0.94493	-1.3021	0.89709	-1.162	-0.4224	-0.1816	0.64044	-1.6191	-0.3875
2014	-1.7289	-0.624	-1.123	-1.5533	-0.6555	-1.0157	0.38945	1.30799	0.21552	-0.0703	1.15158	-1.8203
2015	0.98047	-0.7206	-0.8494	0.9783	-0.6562	-0.7889	-0.1511	-0.5386	0.97969	0.63549	-0.335	0.33961
2016	0.85606	-0.7308	0.59942	0.66707	-0.9592	0.53928	0.24876	-0.179	-1.7245	0.10513	-0.3827	1.82637
2017	-0.7674	0.61752	0.21752	-0.7549	0.6011	0.2164	-0.2319	-0.3221	0.62585	-0.1254	0.27732	-0.156
2018	-1.1211	-0.0897	-0.3929	-1.03	-0.0135	-0.149	-0.1037	1.25485	1.44084	-0.6376	-1.5824	-1.9221
2019	0.15228	0.39517	-2.0728	0.04949	0.41359	-1.8764	-0.1148	-0.6716	0.04974	-0.2427	-0.1589	-0.8428
2020	0.17907	0.38171	0.45879	0.30045	0.45895	0.40703	1.98308	0.5983	0.05502	0.45166	-0.0418	0.94014

The correlation coefficients among our calculated teleconnection indices, ice cover proxies, and recorded AMIC from 1951 and 1983. The bold proxy is the better model. All values are significant at 5%.

	Superior			Michigan			Huron			Erie			Ontario		
	NMDD	CFDD	Recorded	**NMDD**	CFDD	Recorded	NMDD	**CFDD**	Recorded	NMDD	**CFDD**	Recorded	**NMDD**	CFDD	Recorded
Calculated ABNA	0.58	-0.53	-0.52	0.62	-0.62	-0.53	0.65	-0.65	-0.52	0.6	-0.6	-0.51	0.54	-0.59	-0.4
New Index	0.55	-0.51	-0.51	0.65	-0.65	-0.5	0.68	-0.69	-0.55	0.67	-0.67	-0.46	0.58	-0.68	-0.46

The correlation coefficients among the Great Lakes AMIC and various teleconnection indices over the DJFs of 1980 to 202. Numbers in bold represent a correlation significant at 5%.

	Superior	Michigan	Huron	Erie	Ontario	All Lakes
Calculated ABNA	**-0.52**	**-0.50**	**-0.54**	**-0.56**	**-0.35**	**-0.56**
New Index	**-0.58**	**-0.51**	**-0.60**	**-0.48**	**-0.44**	**-0.62**
AO	0.02	-0.14	-0.01	-0.19	-0.13	-0.04
NAO	0.02	0.04	-0.02	-0.14	0.08	0.03
PNA	-0.29	-0.18	-0.27	-0.12	-0.11	-0.26

*bold = statistically significant

C. Sample Python Code

Full code can be found at https://github.com/InigoP/NOAA/blob/master/index_calculation.ipynb

1. Sample 1: NCEP Reanalysis II Data Parsing

```
def select_data(data_set, climate_var,months):
    '''
    parse the geopotential height NCEP data that is found at https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.pressure.ht
    temperature data is found at https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.gaussian.html

    Parameters
    ---
    xarray dataset: string
    climate_var: string
        parse either hgt or temperature
    time_var: string
        parse time.month
    months: list
        list of months eg.[1,2,3]

    Returns
    ---
    xarray dataset
    '''
    result = data_set[climate_var].sel(time=np.in1d(data_set['time.month'],[months])).squeeze()
    if climate_var == 'hgt':
        if months == [12, 1, 2]:
            hgt = result.sel(time = slice('1979-12-01','2020-3-01'),level = slice(500,500), lon = slice (100, 360), lat = slice(
            return hgt
        elif months == [3,4,5]:
            hgt = result.sel(time = slice('1980-3-01','2020-6-01'),level = slice(500,500), lon = slice (100, 360), lat = slice(-
            return hgt
        elif months == [6,7,8]:
            hgt = result.sel(time = slice('1980-6-01','2020-9-01'),level = slice(500,500), lon = slice (100, 360), lat = slice(-
            return hgt
        elif months == [9,10,11]:
            hgt = result.sel(time = slice('1980-9-01','2020-12-01'),level = slice(500,500), lon = slice (100, 360), lat = slice(
            return hgt
        elif months == [10, 11,12]:
            hgt = result.sel(time = slice('1980-10-01','2021-1-01'),level = slice(500,500), lon = slice (100, 360), lat = slice(
            return hgt
        elif months == [11,12]:
            hgt = result.sel(time = slice('1980-11-01','2021-1-01'),lon = slice (190, 300), lat = slice(13)).squeeze()
            return hgt

    elif climate_var == 'air':
        if months == [12, 1, 2]:
            temp = result.sel(time = slice('1979-12-01','2020-3-01'),lon = slice (190, 300), lat = slice(13)).squeeze()
            return temp
        elif months == [3,4,5]:
            temp = result.sel(time = slice('1980-3-01','2020-6-01'),lon = slice (190, 300), lat = slice(13)).squeeze()
            return temp
        elif months == [6,7,8]:
            temp = result.sel(time = slice('1980-6-01','2020-9-01'),lon = slice (190, 300), lat = slice(13)).squeeze()
            return temp
        elif months == [9,10,11]:
            temp = result.sel(time = slice('1980-9-01','2020-12-01'),lon = slice (190, 300), lat = slice(13)).squeeze()
            return temp
        elif months == [10,11,12]:
            temp = result.sel(time = slice('1980-10-01','2021-1-01'),lon = slice (190, 300), lat = slice(13)).squeeze()
            return temp
        elif months == [11,12]:
            temp = result.sel(time = slice('1980-11-01','2021-1-01'),lon = slice (190, 300), lat = slice(13)).squeeze()
            return temp
```

2. Sample 2: Maximum Covariance Analysis

```
def svd(temp, hgt):
    '''
    calculate maximum covariance of hgt and temperature using singular value decomposition
    MCA can isolate important coupled modes of variability between two fields

    Parameters
    ---
    temp: numpy array
    hgt: numpy array

    Returns
    ---
    tuples
        contains first and second temperature and hgt MCA pattern
        and its corresponding expansion coefficients
    '''
    #Preprocess - convert 3D temp and hgt data to 2D arrays
    temp_time, temp_row, temp_col = temp.shape
    temp_2d = np.reshape(temp, (temp_time, temp_row*temp_col), order = 'F')

    ntime_anom, hgt_row, hgt_col = hgt.shape
    hgt_2d = np.reshape(hgt, (ntime_anom, hgt_row*hgt_col), order = 'F')

    #Carry out Maximum Covariance Analysis
    #Create covariance matrix of temp and hgt by finding the dot product
    Cxy = np.dot(temp_2d.T, hgt_2d)/(temp_time -1.0)

    #Apply Singular Value Decomposition to decompose the covariance matrix
    #s is the diagnoal matrix of eigenvalues
    #U is the left singular vectors of Cxy
    #V is the right singular vectors of Cxy
    U, s, V = np.linalg.svd(Cxy, full_matrices=False)
    scf = s**2./np.sum(s**2.0) #calculate cumulative fraction of squared covariance fraction (SCF, %) explained
    V = V.T
    imode = 0
    print(scf[imode])
    print(scf[1])

    #Extract the leading (mode0) Temperature MCA Pattern and project onto the original data
    U1 = np.reshape(U[:,0, None], (temp_row, temp_col), order='F')
    #Expansion coefficient 1 (EC1) of temperature pattern
    a1 = np.dot(temp_2d, U[:,0, np.newaxis])
    #Normalize the leading mode by standardizing with EC1 so patterns correspond to 1-std variation in EC1
    U1_norm = U1*np.std(a1)
    a1_norm = a1/np.std(a1)

    #Exract and project the second leading (mode1) Temperature MCA and EC2
    U2 = np.reshape(U[:,1, None], (temp_row, temp_col), order='F')
    a2 = np.dot(temp_2d, U[:,1, np.newaxis])
    U2_norm = U2*np.std(a2)
    a2_norm = a2/np.std(a2)

    #Extract the leading (mode0) HGT MCA pattern and project onto the original data
    V1 = np.reshape(V[:,0, None], (hgt_row, hgt_col), order='F')
    #Expansion coefficient 1 (EC1) of HGT pattern
    b1 = np.dot(hgt_2d, V[:,0, np.newaxis])
    #Normalize by standardizing with HGT EC1
    V1_norm = V1*np.std(b1)
    b1_norm = b1/np.std(b1)

    #Extract and project the second leading (mode1) HGT MCA pattern and EC2
    V2 = np.reshape(V[:,1, None], (hgt_row, hgt_col), order='F')
    b2 = np.dot(hgt_2d, V[:,1, np.newaxis])
    V2_norm = V2*np.std(b2)
    b2_norm = b2/np.std(b2)
    return U1_norm, a1_norm, U2_norm, a2_norm, V1_norm, b1_norm, V2_norm, b2_norm, scf
```

GREAT LAKES ICE COVER : Enriching Database and Improving Forecast

Master's Project #P19

Part B : PREDICTION USING MACHINE LEARNING MODELS - LSTM AND XGBOOST

A project submitted
in partial fulfillment of the requirements
for the degree of
Master of Science
(Environment and Sustainability:Geospatial Data Sciences)
In the University of Michigan

21 April 2021

University of Michigan
School for Environment and Sustainability

Client: Dr. Philip Chu, NOAA Great Lakes Environmental Research Laboratory
Faculty Advisor:
Dr.Ayumi Fujisaki-Manome, Assistant Research Scientist

Abstract

The Laurentian Great Lakes (hereafter the Great Lakes), cover more than 94,000 square miles in the United States and Canada and are interconnected by a series of rivers, straits, and connecting channels, forming the world's largest freshwater system. The Great Lakes waterway is a system of natural channels and artificial canals which enable navigation between the Great Lakes[1]. Among these waterways, St. Marys River is a key waterway that extends from Brush Point in the southeast corner of Lake Superior to the northwest section of Lake Huron. The massive Soo Locks and dredged channels, constructed in the St. Marys River, support navigation activities including commercial shipping in the Great Lakes. This navigational lock system is closed annually from late January to late March due to the development of ice cover over the river. However, a notable year-to-year variability in ice condition exists in the transition periods, namely when ice cover starts to form in early winter and melt in spring. This poses a challenge to safe and effective planning of shipping and icebreaking operations around the region. Consequently, it is significant to find a way to predict the ice coverage in order to help the shipping community plan their schedule in advance.

While the ice prediction for Great Lakes has been done in the past several years, most of them applied statistical and numerical modeling methods, such as Regression analysis. However, because of the focused geographical area and complex physics in the river system, the St. Marys River area is not covered by these traditional models, including NOAA's Great Lakes Operational Forecast System.

Machine learning is a rising technique that is well developed and has been massively applied in many scientific fields, like in medicine, finance, geophysics and climate research. Previous studies have shown that compared with normal statistical methods, Machine learning is more likely to detect the internal mechanisms among data, contributing to a higher prediction accuracy. In this study, we applied two supervised Machine learning methods namely Long Short-memory (LSTM) and Extreme Gradient Boost (XGBoost) and compared their predictive abilities on the Great Lakes' ice prediction. We trained these models by using the four weather stations data around the St. Marys River from the Coastal Marine Automated Network and the satellite-based ice coverage data from the NOAA Coastwatch Great Lakes node. Apart from these machine learning algorithms, we use various packages implemented in Python, like Datetime, Pandas, Sklearn for data processing and Matplotlib for data visualization.

After the respective models were built and prediction was conducted for the next 7 days : Both the models accurately forecasted ice cover during stable phase. Based on the metrics of mean absolute error and root mean square error, it was found that the model skill tended to be worse in early winter and spring months compared with the mid-winter period because of highly dynamic conditions in these periods. The differences between the original and predicted ice-on/off date are within 3-5 days for both models. LSTM has a higher prediction accuracy than XGBoost based on the result. XGBoost and LSTM have the potential to be used as a good reference for the shipping community to help them plan safe and effective operations.

Acknowledgements

Our client, National Oceanic Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL)

NOAA GLERL and its partners conduct innovative research on the dynamic environments and ecosystems of the Great Lakes and coastal regions to provide information for resource use and management decisions that lead to safe and sustainable ecosystems, ecosystem services, and human communities.(https://www.glerl.noaa.gov/)

This research was carried out with support of the NOAA GLERL, awarded to CIGLR through the NOAA Cooperative Agreement with the University of Michigan (NA12OAR4320071)

Our advisor and mentor, Dr.Ayumi Fujisaki Manome, Assistant Research Scientist, Cooperative Institute of Great Lakes Research

Dr. Manome's extensive knowledge and overall experience in the field of oceanography were foundational in the development of this project. She was very supportive along the way and provided us the necessary resources, suggestions and opportunities to improve our learning on the subject.

Our staff advisor, Haoguo Hu, Ice Modeler, Cooperative Institute of Great Lakes Research

Haoguo's knowledge in Ice modeling using Machine Learning has been invaluable in providing us the technical help and help with the modeling process. His critical questions have been useful to improve our models.

Our client advisor, Dr.Philip Chu, Supervisory Physical and Branch Chief,NOAA GLERL

We would like to thank him for his support and contribution.

List of Figures

List of Tables

TABLE OF CONTENTS

1.Introduction

1.1. Overview of the Great Lakes and St.Marys River

Lake ice always plays an important role in the shipping industry in the Great Lakes (Fig.1). In the Great Lakes, lake ice starts to form in late November and early December, and causes severe problems in navigation from mid-December until early March (Figs. 1 and 2). Usually, federal and commercial icebreakers help keep the shipping routes open in early winter and spring. a In St. Marys river (Fig. 3), a key waterway in the Great Lakes, the navigational lock is closed from mid January to late March. In the transition periods (i.e. when lake ice starts to form and melts), the capability of ice forecasting with sufficient quality and lead time is of considerable concern in lock operations, the shipping industry, and icebreaking operations

In any given year, the formation, movement, and timing of ice cover on the Great Lakes is temperamental and changes substantially with shifts in weather and climate patterns. Extremely cold air across the Great Lakes has been the major contributor to ice formation on the Great Lakes. Air temperature and its yearly variability is a major factor in determining when and how much ice cover develops. Other factors such as El Nino in the Pacific Ocean, Lake effect snow etc also affect ice cover in the Great Lakes.
In 2018, NOAA noted the downward trend in the Ice Cover on the Great Lakes since the 1970s.

Fig 1. Image of the Great Lakes taken on Feb. 14, 2020 shows significantly less ice cover compared to the average..
Taken by NOAA-NASA Suomi NPP/NASA Earth Observatory.
Source: https://www.ibtimes.com/noaa-nasa-satellite-image-shows-lower-average-great-lakes-ice-cover-2925431

Fig 2. Great Lakes, MODIS, March 25, 2019.
Source:https://earthobservatory.nasa.gov/images/144747/a-clear-spring-view-of-the-great-lakes

Fig 3. Clipped area of St.Marys River from Fig. 2.

1.2. Background on previous Ice Cover and analysis methods

Ice cover prediction has always been a topic of great concern in the Great Lakes and other cold waters. Here is the background and some previously done research on ice prediction.
Chi and Kim[2] used a total of 446 months of monthly Arctic sea ice concentration data, acquired from November 1978 to December 2015, for sea ice prediction in the Arctic using machine learning modeling. The data acquired from November 1978 to December 2014 was used as the training data and the data acquired from January 2015 to December 2015 was used as the test data. The Machine learning models used in this research were Long Short-term Memory (LSTM) and Extreme Gradient Boost (XGBoost), which generated good prediction results. The special aspect of this research is that both the input and output data are remote sensing images. However, this research only includes one feature (previous ice value) in the model, without considering other environmental features, such as air temperature. Geophysical Fluid Dynamic Laboratory[3] developed a quasi-operational prediction system that is run every month and produces seasonal forecasts of the climate system, including sea-ice extent.
In the Great Lakes, GLERL has conducted research on ice cover forecasting on two different time scales, short term (1-5) days[4] and seasonal[5,6]. It has successfully forecasted the annual maximum Great Lakes ice cover and the long term average annual maximum ice cover for the whole Great Lakes as well as for each lake. And the main predictor of the forecast model is the latest surface air temperature. However, these models for ice forecasting have not covered the Great Lakes river systems or waterways mostly because of the computational challenges to capture the detail, complex physics at these focused geographic scales at the same time as they cover the lake-wide scale phenomena. As a result, the ice forecasting capability both at the short-term and seasonal time scales for the key river systems and waterways has been a gap in the Great Lakes.

1.3. Overview of Machine Learning models

Machine Learning techniques have been massively applied in many scientific fields, like in medicine, finance, geophysics and climate research[7,8,9]. Machine Learning approaches are being increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data which assist in the identification of useful connections in the climate system. They are used to train statistical models which mimic the behavior of climate models and also to identify and leverage relationships between climate variables. These trained statistical models allow us to quantify non-linear relationships between the climate variables we input to train the models .

However, current approaches may not be optimal when system behaviour is dominated by spatial or temporal context. Contextual cues should be used as part of Machine learning (an approach that is able to extract spatio-temporal features automatically) to gain further understanding of climate science, thus improving the predictive ability of seasonal forecasting and modelling of long-range spatial connections across multiple timescales.

In the Great Lakes, applications of machine learning models to date are limited, except for a few pioneering works that focused on waves.[10,11] Based on the historical research, we

know that machine learning models, such as LSTM and XGBoost are attractive modeling approaches to examine in ice forecasting in lieu of numerical modeling. An obvious advantage with use of a machine learning model against a numerical geophysical model is reduced computational cost. This is particularly true for a small but complex system like Great Lakes waterways. Thus, the potential of machine learning modeling in Great Lakes ice forecasting warrants pilot research: This includes finding the most appropriate machine learning model, identifying the suitable features, and adjusting the best parameters for our prediction problems. Such work is critical in order to support future products that can support decision making by lock operators, vessel managers, ship captains, and Coast Guards around St Marys river water system in a way that they can maximize their shipping time and avoid unnecessary cost.

1.4. Project Objectives

The general objective of our project is to address how we can apply Machine learning to predict the ice cover in the St. Marys River system in the Great Lakes accurately. The ultimate goal is to build a pilot modeling framework that can support decision making of the stakeholders around the St. Marys River system, one of the key waterways in the Great Lakes.

To achieve this objective, we address the following tasks:

- Prediction of ice cover on St Marys River using Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) models. Developing these models will help us understand the relationships among Great Lakes ice cover, surface meteorology, and climate indices inorder to inform better prediction.
- Compare the models' prediction skills. Each of the models have their own learning methods and capabilities in prediction. Identifying the conditions where each model performs well is important to evaluate which model to use at what stage.
- Provide the prediction result for local shipping community

2.Methods

2.1. Model Selection

The main neural network models selected for this project are Long Short-Term Memory (LSTM) and Extreme Gradient Boost (XGBoost). LSTM is a kind of neural network that is widely used for predicting the time series data. It has been widely used in climate science and got great outcomes[12,13].XGBoost is a widely used Machine learning method that uses Gradient tree boosting technique.

2.2. Data Preparation

2.2.1. Data Collection of Weather data

Weather data was collected from the National Oceanic and Atmospheric Administration's National Data Buoy Center(NDBC) website. Station ID search was used to search for the respective Station's data. Data is found under 'Historical data', 'Standard meteorological data'

for each year. The historical data was downloaded as a text file for each of the weather stations - SWPM4, LTRM4, WNEM4, RCKM4 and DTLM4 for the years 2007 to 2020 as shown below.

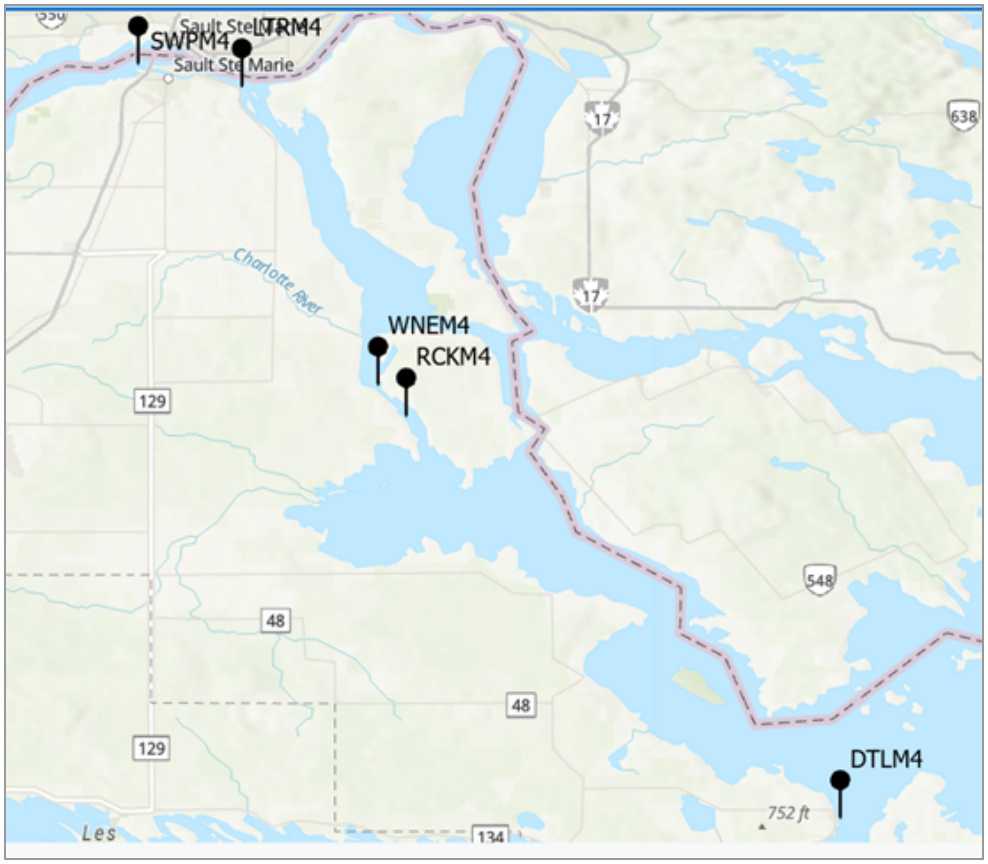


Fig 4. Weather stations' (SWPM4, LTRM4, WNEM4, RCKM4 and DTLM4) locations on St.Marys River.
Refer to Appendix C for coordinates of the stations.

Though the station RCKM4 was present on St.Marys river, it was not considered in our study as it had no data on Air Temperature(ATMP) that was needed specifically for our study.
The stations recorded data on several weather parameters as shown in Table 2 of Appendix C. However, parameters relevant to our study such as Wind Direction (WDIR), Wind Speed (WSPD), Gust speed (GST), Sea Level Pressure (PRES) and Air Temperature (ATMP) were considered for the analysis.

2.2.2. Data Collection of Ice data

Ice data was collected from the National Oceanic and Atmospheric Administration's CoastWatch Great Lakes Node website. This data consists of Great Lakes ice concentration data obtained from the US National Ice Center. The gridded ice analysis products are produced from available data sources including Radarsat-2, Envisat, AVHRR, Geostationary Operational and Environmental Satellites (GOES), and Moderate Resolution Imaging Spectroradiometer (MODIS). Spatial resolution of the ice concentration data is 2.55 km in 2005, and 1.8 km from 2006–2017. The resulting NIC data set defines ice concentration values from 0 to 100% on 10% The format of the ice data is netcdf, the package netCDF4 is used to read ice data.

Trim the research area to the area of St Marys River by its latitude and longitude (46.05° N - 46.59° N, 84.65° W - 83.80° W)
We extracted the ice concentration value at the location of the above weather stations. The average ice concentration over the St Marys River was also calculated. An average ice concentration spatial map of St Marys River is shown below.

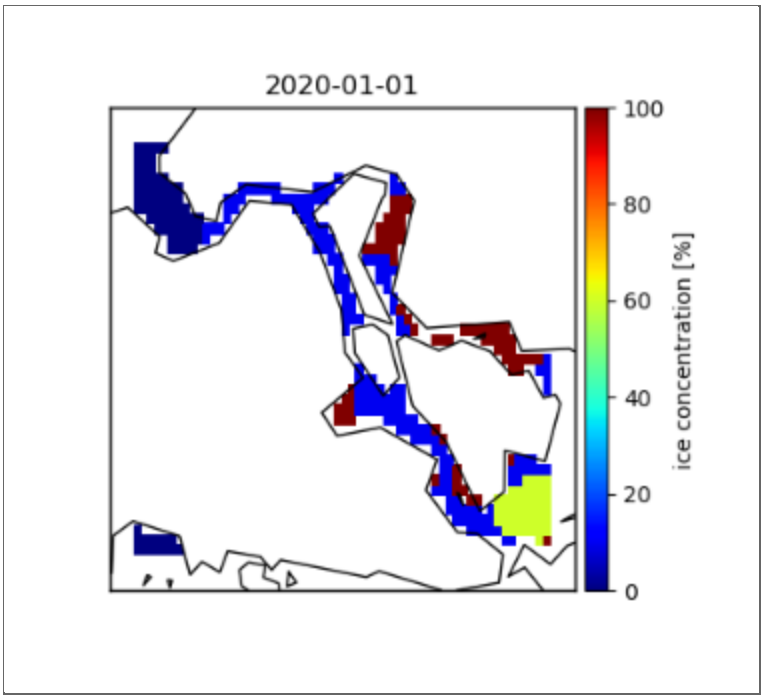


Fig 5. The spatial distribution of ice concentration in St Marys River on January 1st, 2020

2.3. Data Pre-processing(Data Wrangling) and Exploratory Data Analysis

Python programming and Jupyter Notebook IDE was used for the analysis of data and Model development. Pandas was further used to convert the data into Machine readable format. Relevant libraries(given below) were imported for the analyses.

- Scientific Computing libraries (Pandas, NumPy)
- Visualisation library(Matplotlib)
- Algorithmic library(Statsmodels)

The screenshot of relative code can be found in Appendix A.

2.3.1 Data Preprocessing of Weather data

For each station, the downloaded data was read into dataframes for that station and column names were assigned. Any missing information was replaced with Nan's and interpolated. Descriptive Statistics(summary functions) was used to understand the attributes of the data.The features that had more than 30% of missing data could not be interpolated and were dropped. Exploratory data analysis(EDA) was performed on each station. "DateTime" package in Python was used to create a "DateTime" feature that was used as a JoinKey to merge the different data frames. It was later used as an Index to represent the parameters or features in the data set. Since the weather data was collected at a 6 minute interval on a daily basis, a data frame

was created for each day of the year using Groupby operation. “corr() method” was used to create a correlation matrix to observe the feature across all stations.Air temperature and Ice Cover were considered as important features.

2.3.2 Data Preprocessing of Ice data

The average ice concentration value was calculated across St Marys river from 2007 to 2020. A dataframe for ice data was built that had the mean ice concentration (%) value for the whole river as well as for the individual weather station for each day. Summary functions were used to detect and remove anomalies in the data set. Ice concentration value was normalized as the Machine Learning models are more sensitive to the normalized data. The dataframe was exported into an Excel format. Weather data and Ice cover data for all the weather stations were merged to be used for analysis.

2.4. Model Development

In both LSTM and XGBoost models, we used the surface weather data (see section 2.3.1) as input data and ice concentration data (see section 2.3.2) as the target variable. Considering that these are time series data, in which serial correlation exists between successive observations, the usual approach of random assignment to the three partitions is not followed. Instead, continuous periods of the time series are assigned to each partition. The data was divided into the training set (2007-2015), the validation set (2016-2017), and the test set (2018-2019). With the training set, reliable estimates of the trainable model parameters are achieved. The validation set is used in the selection of model hyperparameters such as tree depth and to check for overfitting on the training data. Finally, the test set is used to test how well the ML models generalize to unseen conditions. The model specific configurations are described in the sections 2.4.1 and 2.4.2.
Moreover, the data from Nov.01 to May.10 are further divided into Freezing Phase (Nov .01-Jan. 14), Stable Phase (Jan. 15–Mar. 25) and Melting Phase (Mar. 26–May 10). Apart from predicting for the whole entire, we also conducted predictions based on these three phases and observed the performance of our models in different phases.

2.4.1. LSTM

For LSTM, the number of features and time steps were two significant parameters that need to be determined[14]. According to the data downloaded from the weather stations, there were five relevant features.They are air temperature(ATMP), wind direction(WDIR), wind speed(WSPD), gust speed(GST) and atmospheric pressure(PRES). Descriptions of the features are provided in Appendix C. Along with previous ice value, these features and ‘day of year’ were considered. We calculated how these features are related with the target value (ice value) using regression (f_regression package in python) and found that previous ice value, air temperature and ‘day of year’ have the most significant effect on our prediction, while atmospheric pressure, wind speed,

wind direction and gust speed don't have an obvious influence on the future ice value. Consequently, we chose air temperature, previous ice value and 'day of year' as three features for the LSTM input data. The graph that shows how accuracy changes with different combinations of features can be found in Appendix B, figure 13.
Regarding the time steps (how many previous days of information are used for the input data), we tried running the model from 1 day to 30 days time. The results showed that when time step equals five days, which means that when we input 5 days of data into the model, the prediction accuracy is highest. The graph that shows how accuracy changes with time steps can be found in Appendix B, figure 15.
As for other hyper-parameters, we tried different functions and model structure (different hidden layers and hidden units) but the accuracy did not change a lot. So we use the most widely used ones. Relu was used as activation function, 'mean squared error' was used as the loss function and RMSprop is used as the optimizer. Our model has one hidden layer with ten hidden units in it. And the batch size is set as 32, the specific code used for constructing the LSTM model can be found in Appendix A, figure 10.

2.4.2. XGBoost

Packages for XGBoost were imported. Sklearn(or Scikit-learn) library was used for the model selection. Daily lags (for 7 days) and moving averages or Rolling Means (for 3,4,5 and 6 days respectively) for Air Temperature and Ice Cover were created as additional features in the dataset (in addition to existing features WDIR, WSPD, GST, PRES, ATMP and Ice Cover). This is normal given that XGBoost is not set up for Time series data (like an LSTM is). So providing lags and rolling means of ICE is essential to achieve decent models. Feature engineering and feature importance was determined as explained in Appendix B figure 14.
The input variables to be predicted (Ice Cover) and the variables to be used for training (the remaining features in the dataset) were identified. The model was trained using training data and then it was tested on the testing data. "Time Series split cross validation" method was used in which no future observations can be used in constructing the forecast. "GridSearchCV" function was used to create multiple splits in training data across different time periods with the training data expanding in each field. segments.The forecast accuracy is computed by averaging over the tests.
XGBoost Regressor model was run to predict the ice cover based on the weather data. XGBoost regressor was run with a wide range of hyper parameters(Learning_rate, max_depth, subsample, colsample_bytree, n_estimators) and 5 cross validation time series splits. The specific code used for XGBoost can be found in Appendix A, figure 11. Mean absolute error was used to find the difference between the prediction and actual values (test data). In order to forecast the ice cover in the future days using the predicted value on any day, the latter was used as the "ground truth" for forecasting for the former. In our case the predicted ice cover value on 01 Jan 2019 was used to forecast the ice cover for the next 14 days or 2 weeks from 2-15 Jan 2019. XGBoost models were built for individual stations and the outputs were used to create the forecast for all the stations merged together.

2.5 Evaluation methods

The following evaluation methods were used to evaluate the accuracy and predictive power of the Models. Both evaluation metrics are calculated for LSTM, XGBoost and our baseline.

1. Root Mean Squared Error (RMSE) and Mean Absolute Error(MAE) on the test set.
2. Differences between the original and predicted ice-on/off date.

Definition of baseline: In order to verify our predictions do perform better than simply taking the average of the ice concentrations in the previous years, we use the average of the ice-concentrations in the past 9 years as our baseline. This baseline provides the approximate information of ice coverage at a given time of a season in the 'normal' year. If the error for our model is lower than the baseline, it indicates that our models do have predictive ability better than using the normal-year information as forecast.

Definition of ice-on/off date: If consecutive 3 days that have ice cover more than 10% and less than 10%, it will be chosen as Ice on and Ice off date respectively. Figure 12 in Appendix B can be used for understanding ice-on/ice-off dates.

3.Results

3.1 Time series of ice cover on the test set

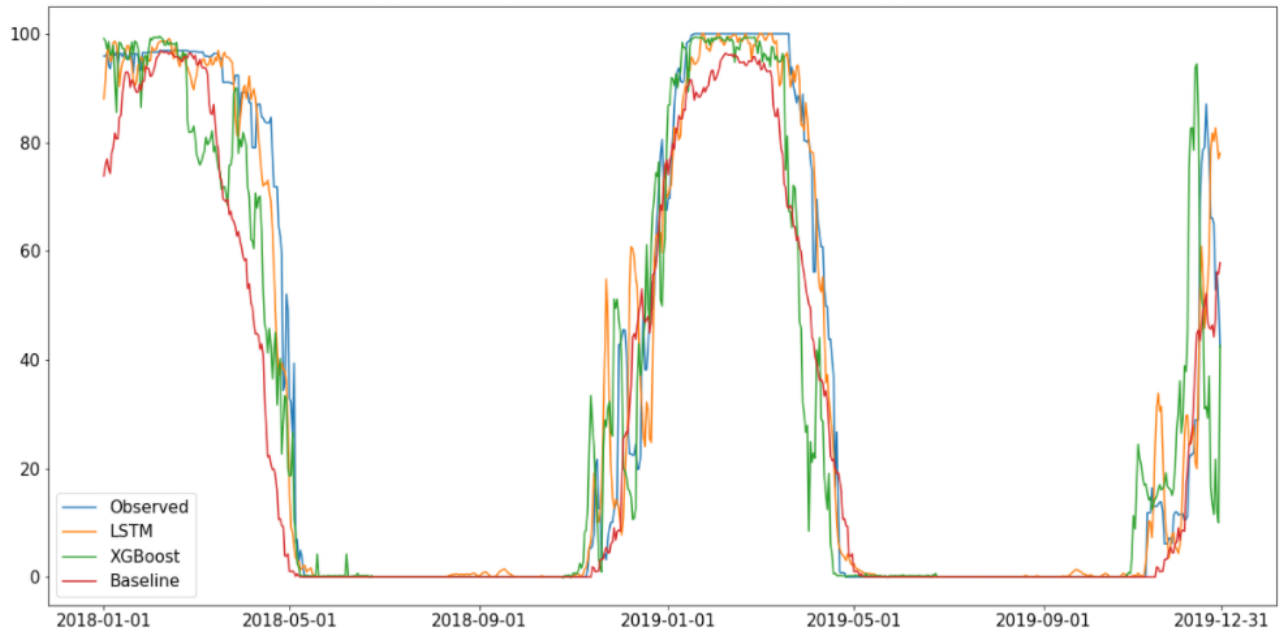


Fig 6. Time series of ice cover on the test set

This time series plot compares the relationship among the original ice value, the prediction result of XGBoost, the prediction result of LSTM and our baseline from 2018 - 2019 (test set). The blue, orange,green and red curves represent the original ice value, LSTM predicted ice

value , XGboost predicted ice value and the baseline, respectively. According to the plot, it is clear that compared with the red curve (baseline), the variation trend of the orange curve (LSTM) and green curve (XGBoost) are both much more similar to the blue curve, which indicates that the predictions by our models are much more accurate than simply taking the average of the past several years.

Apart from that, we can also locate where the prediction error comes from. According to this time series plot, the orange curve and the green curve overlap with the blue curve for most of the time, except for December and April for each year.

During these two periods, we can see that the blue curve (original) contains some small fluctuations, which indicates that the ice value varies dramatically in a short period of time. During such a period, we can see the orange curve and the green curve fall apart with the blue curve (original), which means such periods should be the main source of our prediction. After our analysis, this period may be caused by a drastic change in the local air temperature, which is hard to predict by using the previous weather data, especially when the predicted interval is large.

As for the comparison between LSTM and XGBoost, it is hard to judge which one performs better only based on this plot as they are very similar and close to each other. Consequently, we calculated the MAE and RMSE of these two models so that we can compare them quantitatively. The comparison section will be explained in more detail in the next section

3.2 Prediction error (RMSE and MAE) on test set

Table 1: Prediction error (RMSE and MAE) on test set with 7 predict interval

	MAE			RMSE		
Metric	**XGBoost**	**LSTM**	**Baseline**	**XGBoost**	**LSTM**	**Baseline**
Freezing Phase	5.99%	7.50%	9.54%	9.82%	15.17%	17.21%
Stable Phase	5.19%	2.89%`	10.27%	7.73%	3.38%	10.53%
Melting Phase	4.50%	7.11%	9.33%	7.82%	12.44%	16.22%
The Whole Year	4.57%	2.90%	6.73%	7.78%	8.71%	13.38%

This table compares the MAE and RMSE among the XGBoost, LSTM and the Baseline when the prediction is conducted for the next 7 days. According to the table, the prediction result of both Machine learning models are better than the baseline. This indicates that our prediction models do have better accuracy than simply taking the average of the ice cover value in the previous several years.

For the prediction in different phases, we can observe that both models accurately forecast the ice cover during the stable phase, while the accuracy in freezing and melting phase are relatively low. After our analysis, the reason for the low prediction accuracy for these two phases might be the ice cover changes frequently and drastically, which makes the prediction to be difficult.
For the comparison between LSTM and XGBoost, We can see that for the stable phase, the RMSE and MAE of LSTM are smaller than those of XGBoost, which indicates that LSTM performs better during the stable phase. However, for freezing phase and melting phase, the RMSE and MAE of the XGBoost is smaller. This means XGBoost is better at predicting when the ice changes more frequently.

3.3 Ice season duration with 7 Predict Interval

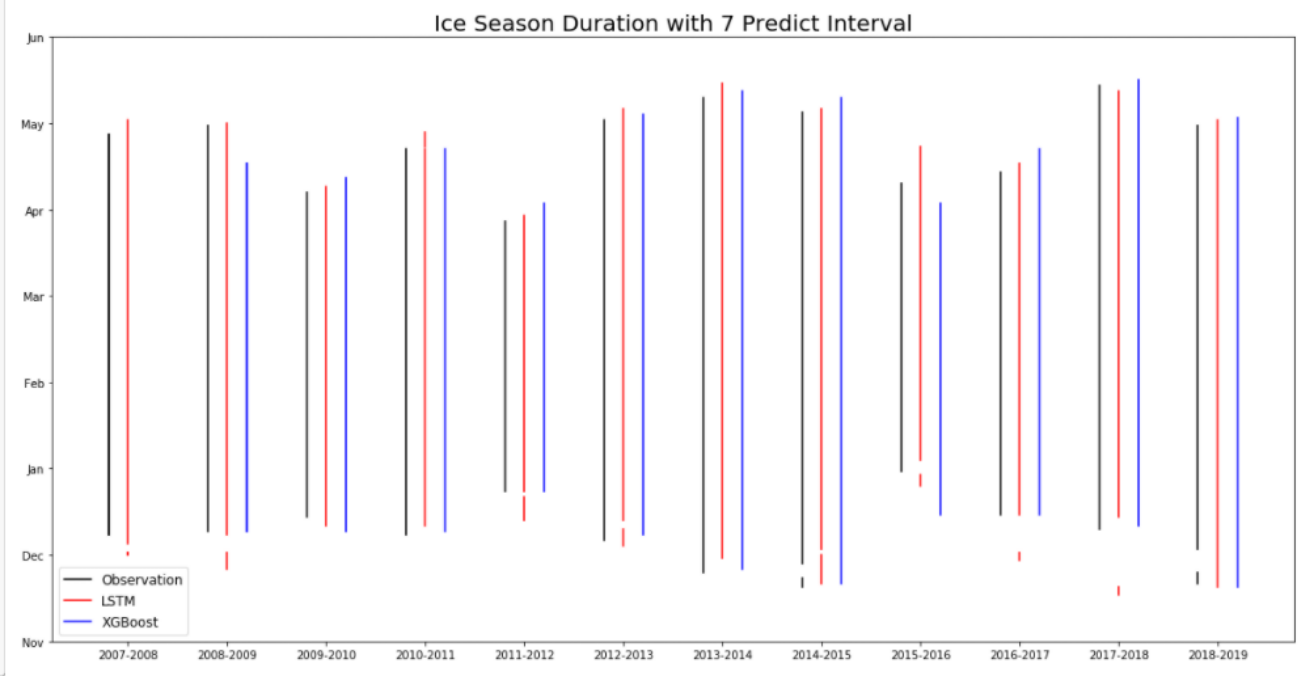


Fig 7. Ice season duration with 7 predict interval

Figures 7 show the relationship between the original ice-on/ice-off dates and the predicted ice-on/ice-off dates when the prediction is conducted for the next 7 days. The black, red and blue line represents original ice-on/ice-off dates and ice-on/ice-off dates of LSTM and ice-on/ice-off dates of XGBoost, respectively. To be noted, because the weather station only contains data from 2008, the ice-on/ice-off dates for XGBoost in 2007-2008 are missing. Considering the fact that the local shipping community may be more concerned about when the ice will freeze and melt, this evaluation method might have more reference value in this project. As this plot shows, the differences between the original and predicted ice-on/off dates for both

models are small, mostly within 3-5 days, even when predicting the ice cover after 1 week. However, for the small ice periods (very short black lines in the plot), both models can't predict them accurately, which is what we need to improve way forward.
In general, both models can control the error of ice-on/ice-off almost within 5 days when the predicted interval equals 1 week, which can absolutely provide helpful information for the local shipping communities.

3.4 Predictions accuracy with predicted interval

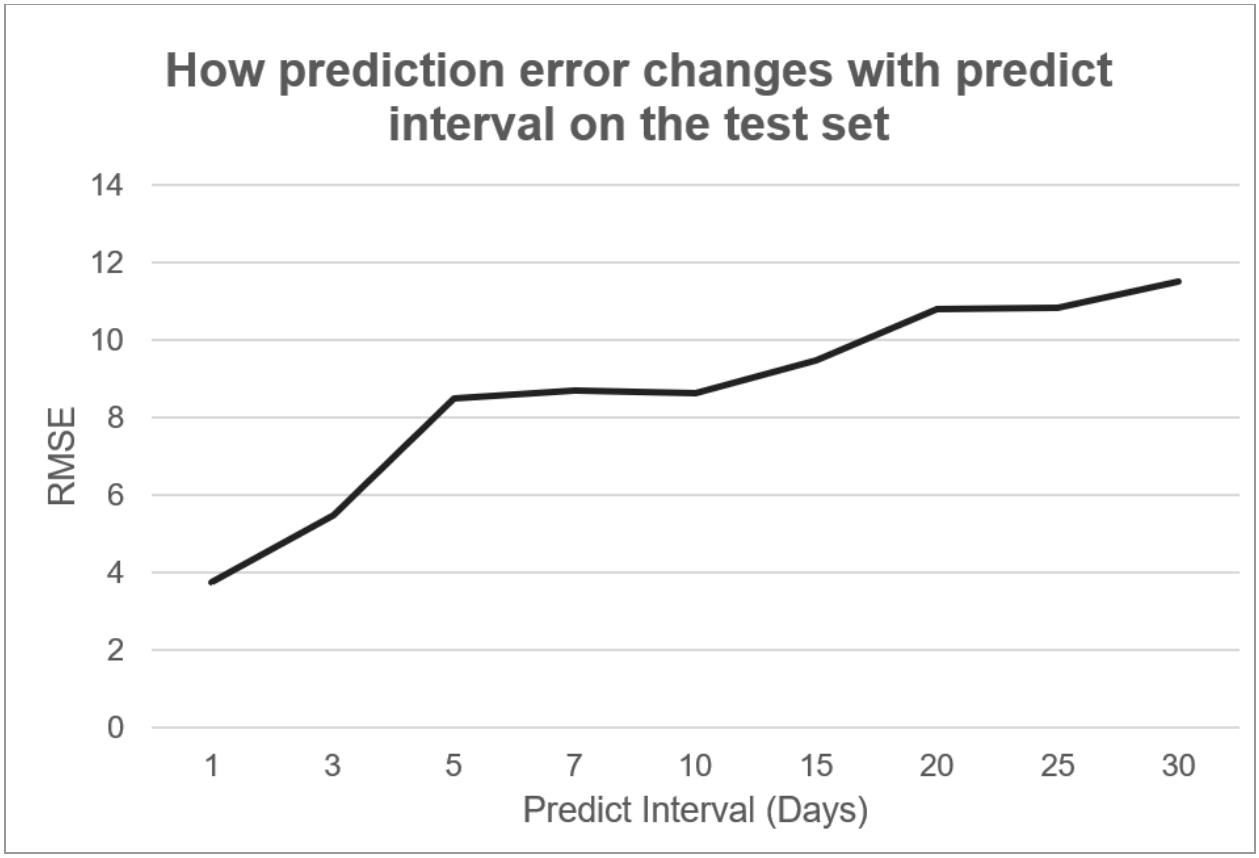


Fig 8. How prediction error changes with predict interval on the test set using LSTM

For prediction problems, the final concern is how long our predict interval could be. And the graph shows how the prediction error (RMSE) changes with the predict interval on the test set. As the graph in Fig. 9 shows, when the predict interval increases from 1 to 5, the RMSE of our model increases quickly. However, as the predict interval continues to increase (from 5 days to 30 days), RMSE increases quite slowly. According to the graph, we can conclude that as long as the predict interval is less than 15 days (around 2 weeks), our prediction result will not cause much error. When the predicted interval is less than 3 days, the result will be very accurate.

4.Achievement & Outreach

Due to the unexpected pandemic situation, we faced a few challenges in executing the planned outreach activities. First, because of the travel restrictions placed in 2020, we had to cancel the in-person visit to the Soo Lock and a meeting with lock operators of the U.S. Army Corps of Engineers (USACE). Second, the summer internship at GLERL that was originally anticipated, but it had to be converted to a virtual setting due to the restrictions in building access. However, we managed to adjust our activities in order to achieve meaningful outreach from our work. First, we routinely communicated with Dr. Philip Chu, the client contact at GLERL and held a virtual seminar to showcase our findings on April 22, 2021. Second, the faculty advisor Fujisaki-Manome continues to communicate with professionals at USACE and NOAA on the project progress and seek potential continuation of the work. Third, most importantly, we presented our study at the 101st AMS(American Meteorological Society) Annual Meeting and hosted virtually during 10-14 January 2021. We also presented our study at the AMS 20th Annual Student Conference in the form of a poster (snapshot shown in Fig. 10) and recorded hosted virtually during 9-10 January 2021.

The AMS is a global community committed to advancing weather, water, and climate science and service. Atmospheric scientists, oceanographers, hydrologists, earth system scientists, students, practitioners from Science and Engineering, academia etc. are members of the AMS community. Conferences and events are facilitated to meet, share ideas and collaborate on research and implementation.

Our presentation received positive reviews and some suggestions for improvement. We made modifications to our analyses and incorporated the reviews accordingly.

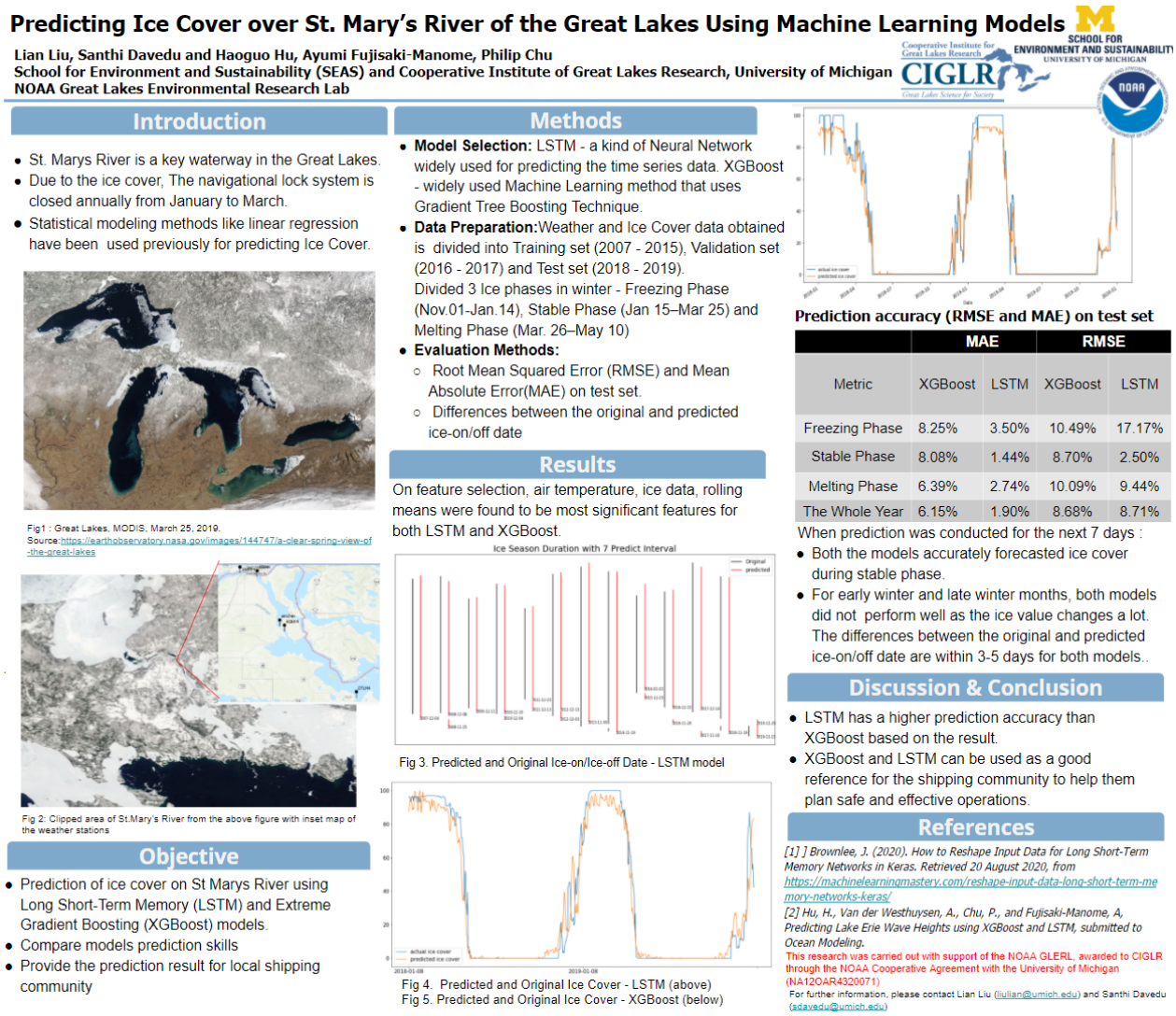


Fig 9. Poster presented at 101st AMS Annual Meeting and 20th Student Conference

5.Discussion & Conclusions

Based on our analyses, it can be observed that in general, LSTM has a higher prediction accuracy than XGBoost in stable phase. The three significant features in the LSTM model are temperature, previous ice cover and the day of year. And the input data contain the information in five consecutive days. The performance of LSTM is better when predicting for the mid ice phase, while the performance for the freezing and melting phase can still be improved. And the error of ice-on and ice-off dates can be controlled within 3-5 days even when we predict 1 week later.

For XGBoost, After executing the forecasting model using XGBoost using best feature combination and predicting for next 14 days it can be concluded that, the XGBoost model

performs well and accurately during the mid winter season(Feb and March) and when the ice cover is constant i.e either always low or always high. Thus, the XGBoost model can be used for 60-70 days during the mid-winter to reliably predict the Ice Cover. For early winter months and late winter months, the model does not perform as well as it does in the mid winter season. This is because we are using one day's ground truth to create features for making predictions. Rolling means which is an important feature/variable always has lag1 data which affects the predictability.
Both models outperformed the basic forecast using the normal-year condition. In conclusion, XGBoost and LSTM can be used as a good reference for the shipping community to help themplan safe and effective operations.

The limitations of our project are identified as follows. Limitation of Machine learning modelling in general as compared to Physics based learning is that these models cannot predict any unprecedented event since the model forecasts based on past data. For example, the XGBoost models are largely unable to extrapolate target values beyond the limits of the training data when making predictions. Another limitation of our project is that the ice data and some part of weather data are both contained from the satellite image, which is pixel based. If the spatial resolution of the satellite image is not high enough (in other words, one small pixel represents a large area), the data we obtain will then be highly inaccurate.
Our model only performs well when predicting for the whole St Marys river. In terms of predicting for a small area, such as the location of an individual weather station, the performance of the model decreases because the weather and ice value change more often in a small area.

In terms of the way forward, our project has future scope to be worked upon and developed. Some potential areas are :

- User engagement and co-design the study along with the end users and customise the project as per their need. For example, making changes to our analyses based on requirements such as:
    - Need for analysis on a specific point in the region instead of the whole St. Marys river
    - Need for analysis for a broader region
    - Users need the forecast of ice thickness instead of percentage of ice cover
- Inclusion of Spatial analysis and predictions in the study
- Hyper parameter tuning that was done in this project can be worked upon further more. For example in XGBoost model hyper-parameters include things like the maximum depth of the tree, the number of trees to grow, the number of variables to consider when building each tree, the minimum number of samples on a leaf, the fraction of observations used to build a tree etc. For LSTM the list includes the number of hidden layers, the size (and shape) of each layer, the choice of activation function, the drop-out rate and the L1/L2 regularization constants etc.

6.Bibliography

[1] Assel, R., Drobot, S., & Croley, T. (2004). Improving 30-Day Great Lakes Ice Cover Outlooks*. *Journal Of Hydrometeorology*, *5*(4), 713-717. doi: 10.1175/1525-7541(2004)005<0713:idglic>2.0.co;2

[2] Chi, J., and Kim, H., (2017), Prediction of Arctic Sea Ice Concentration Using a Fully Data Driven Deep Neural Network.. *Remote Sensing*, *9*(12), 1305. doi: 10.3390/rs9121305

[3] Geophysical Fluid Dynamics Laboratory, "Arctic Sea Ice Predictions" https://www.gfdl.noaa.gov/arctic-sea-ice-predictions/, last accessed on April 27, 2021.

[4] Anderson, E. J., Fujisaki-Manome, A., Kessler, J., Lang, G. A., Chu, P. Y., Kelley, J. G. W., et al. (2018). Ice Forecasting in the Next-Generation Great Lakes Operational Forecast System (GLOFS). Journal of Marine Science and Engineering, 6(123), 17 pages. https://doi.org/10.3390/jmse6040123

[5] Assel, R., Drobot, S., and Croley, T.E. (2004), Improving 30-Day Great Lakes Ice Cover Outlooks, Journal of Hydrometeorology, 5, 713-717.

[6] Wang, J., J. Kessler, X. Bai, A.H. Clites, B.M. Lofgren, A. Assuncao, J.F. Bratton, P. Chu, and G.A. Leshkevich (2018), Decadal variability of Great Lakes ice cover in response to AMO and PDO, 1963-2017. Journal of Climate 31(18):7249-7268, DOI:10.1175/JCLI-D-17-0283.1.

[7] Stewart, J., Sprivulis, P. and Dwivedi, G., 2018. Artificial intelligence and machine learning in emergency medicine. Emergency Medicine Australasia, 30(6), pp.870-874.

[8] Kolchinsky, E., 2018. Machine Learning for Structured Finance. The Journal of Structured Finance, 24(3), pp.7-25.

[9] Koc, M. and Acar, A., 2021. Investigation of urban climates and built environment relations by using machine learning. Urban Climate, 37, p.100820.

[10] Hu, H., Van der Westhuysen, A., Chu, P., Fujisaki-Manome, A., Lake Erie Wave Heights using XGBoost and LSTM, Ocean Modeling, in revision.

[11] Feng, X., Ma, G., Su, S., Huang, C., Boswell, M., Xue, P. (2020). A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan. Ocean Engineering, 211, 107526

[12] Tinaqi Chen, & Carlos Guestrin (2016). XGBoost: A scalable Tree Boosting System. Retrieved 21 Aug 2020, from https://dl.acm.org/doi/pdf/10.1145/2939672.2939785

[13] Choi, M., De Silva, L., & Yamaguchi, H. (2019). Artificial Neural Network for the Short-Term Prediction of Arctic Sea Ice Concentration. Remote Sensing, 11(9), 1071. doi: 10.3390/rs11091071

[14] Brownlee, J. (2020). How to Reshape Input Data for Long Short-Term Memory Networks in Keras. Retrieved 20 August 2020, from https://machinelearningmastery.com/reshape-input-data-long-short-term-memory-networks-keras/

7.Appendices

Appendix A - Screenshots of Code

```
# Importing packages for XGBoost and K-fold Gridsearch

import xgboost as xgb
from sklearn.model_selection import TimeSeriesSplit, GridSearchCV, KFold
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error
```

```
# Importing  necssary libraries

import pandas as pd
from pylab import *
import numpy as np
from matplotlib import dates
import matplotlib.pyplot as plt
from datetime import datetime, timedelta
import pylab
import sys
import matplotlib.dates as mdates
from glob import glob
```

In [13]:
```
# Dividing training set and test set

# Dividing data from 2007 to 2015 as training set
# Dividing data from 2016 to 2019 as test set
for i in range(0, len(train_df)):
    if (train_df[i][0] == datetime(2015,12,31,0,0,0)):
        num = i
    if (train_df[i][0] == datetime(2017,12,31,0,0,0)):
        num_validation = i

entire_trainX = train_df[0:num+1,1:]
entire_trainY = test_df[0:num+1,[1]]
entire_validX = train_df[num+1:num_validation+1,1:]
entire_validY = test_df[num+1:num_validation+1, [1]]
entire_testX = train_df[num_validation+1:train_df.shape[0],1:]
entire_testY = test_df[num_validation+1:train_df.shape[0],[1]]
```

In [15]:
```
# Reshape the data
# (Samples、Timesteps、Features)
entire_trainX = np.reshape(entire_trainX, (entire_trainX.shape[0], look_back, num_features))
entire_validX = np.reshape(entire_validX, (entire_validX.shape[0], look_back, num_features))
entire_testX = np.reshape(entire_testX, (entire_testX.shape[0], look_back, num_features))
```

In [16]:
```
# Build the LSTM model
def LSTM_model(look_back, num_features,
               trainX, validX, testX,
               trainY, validY, testY,
               scaler_test):

    batch_size = 32
    model = Sequential()
    model.add(LSTM(10,activation='relu' ,input_shape=(look_back,num_features)))
    model.add(Dense(1))
    model.compile(loss='mean_squared_error', optimizer ='RMSProp', metrics=['mse'])
    model.fit(trainX, trainY, epochs=50, validation_data=(validX,validY), batch_size=batch_size, verbose=0)

    trainPredict = model.predict(trainX)
    validPredict = model.predict(validX)
    testPredict = model.predict(testX)
```

Fig 10. Python code for building LSTM neural network

```
### XGBoost model to forecast the ice for the next 14 days. Here we select 2019-03-01 as the
### test data set.

split_date = '2019-03-01'

df_stable_1 = df_stable[['DTLM', 'Ice_lag_1', 'Ice_RM_3', 'Ice_RM_4', 'Ice_RM_5',
                         'ATMP_RM_4', 'ATMP_RM_3', 'ATMP_RM_5', 'Ice_lag_2', 'ATMP_lag_4',
                         'ATMP', 'Ice_lag_3','Ice_lag_4', 'Ice_lag_5']]

df_stable_train = df_stable_1.loc[df_stable_1.index < split_date].copy()
df_stable_test = df_stable_1.loc[df_stable_1.index == split_date].copy()

X_train = df_stable_train.drop(['DTLM'], axis = 1)
X_test = df_stable_test.drop(['DTLM'], axis = 1)
y_train = df_stable_train['DTLM']
y_test = df_stable_test['DTLM']

print(X_train.shape, y_train.shape)
print(X_test.shape, y_test.shape)

xgb1 = xgb.XGBRegressor()
parameters = {'objective':['reg:squarederror'],
              'learning_rate': [0.001, 0.01, 0.05, 0.1],
              'max_depth': [3, 4],
              'subsample': [0.7, 0.8],
              'colsample_bytree': [1.0, 0.6],
              'n_estimators': [100, 200]}

tscv = TimeSeriesSplit(n_splits=5)

xgb_grid = GridSearchCV(xgb1,
                        parameters,
                        n_jobs = -1,
                        cv = tscv,
                        verbose=True)

xgb_grid.fit(X_train, y_train)

print(xgb_grid.best_score_)
print(xgb_grid.best_params_)
```

In [108]:
```
######### Using the predicted value on any date as ground truth and forecasting for next 14 days #################

for i in range(1, 15):
    X_test['Ice_lag_5'] = X_test['Ice_lag_4']
    X_test['Ice_lag_4'] = X_test['Ice_lag_3']
    X_test['Ice_lag_3'] = X_test['Ice_lag_2']
    X_test['Ice_lag_2'] = X_test['Ice_lag_1']
    X_test['Ice_lag_1'] = y_test_pred

    X_test['Ice_RM_3'] = np.mean([X_test['Ice_lag_1'], X_test['Ice_lag_2'], X_test['Ice_lag_3']])
    X_test['Ice_RM_4'] = np.mean([X_test['Ice_lag_1'], X_test['Ice_lag_2'], X_test['Ice_lag_3'], X_test['Ice_lag_4']])
    X_test['Ice_RM_5'] = np.mean([X_test['Ice_lag_1'], X_test['Ice_lag_2'], X_test['Ice_lag_3'],
                                  X_test['Ice_lag_4'], X_test['Ice_lag_5']])

    res = (datetime.strptime(split_date, '%Y-%m-%d') + timedelta(days=i)).strftime('%Y-%m-%d')

    df_stable_test = df_stable_1.loc[df_stable_1.index == res].copy()

    y_test = df_stable_test['DTLM']

    y_test_pred = xgb_grid.predict(X_test)

    print(res, y_test, y_test_pred)
```

In [112]:
```
# Running the XGBoost regressor with a wide range of hyper parameters and 5 CV timeseries splits

t0 = time.time()

xgb1 = xgb.XGBRegressor()
parameters = {'objective':['reg:squarederror'],
              'learning_rate': [0.001, 0.01, 0.05, 0.1],
              'max_depth': [3, 4],
              'subsample': [0.7, 0.8],
              'colsample_bytree': [1.0, 0.8],
              'n_estimators': [100, 200]}

tscv = TimeSeriesSplit(n_splits=5)

xgb_grid = GridSearchCV(xgb1,
                        parameters,
                        n_jobs = -1,
                        cv = tscv,
                        verbose=True)

xgb_grid.fit(X_train, y_train)

tF = time.time()

print(xgb_grid.best_score_)
print(xgb_grid.best_params_)
print('Time to train = %.2f seconds' % (tF - t0))
```

```
Fitting 5 folds for each of 64 candidates, totalling 320 fits
```

```
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done  52 tasks      | elapsed:    3.0s
```

```
0.19795293152105495
{'colsample_bytree': 0.8, 'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 100, 'objective': 'reg:squarederror', 'subsamp
le': 0.7}
Time to train = 18.99 seconds
```

```
[Parallel(n_jobs=-1)]: Done 320 out of 320 | elapsed:   18.8s finished
```

Fig 11. Python code for building XGBoost model

Appendix B - QA/AC

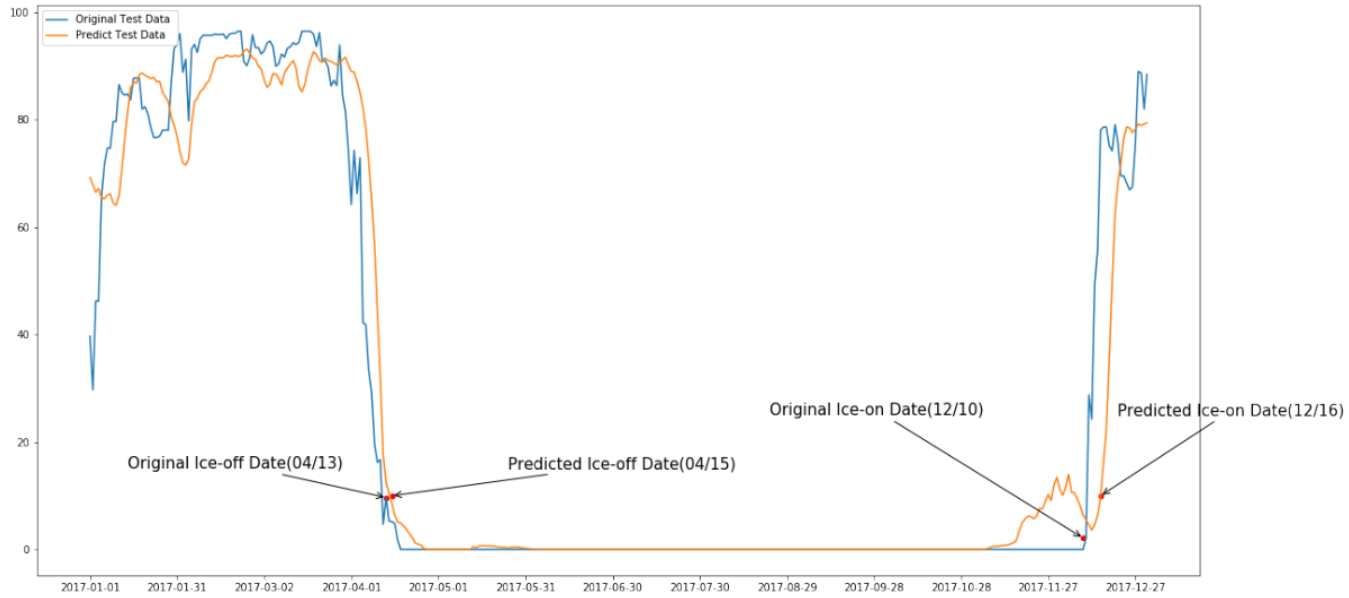


Fig 12. Ice-on/Ice off dates in time series plot

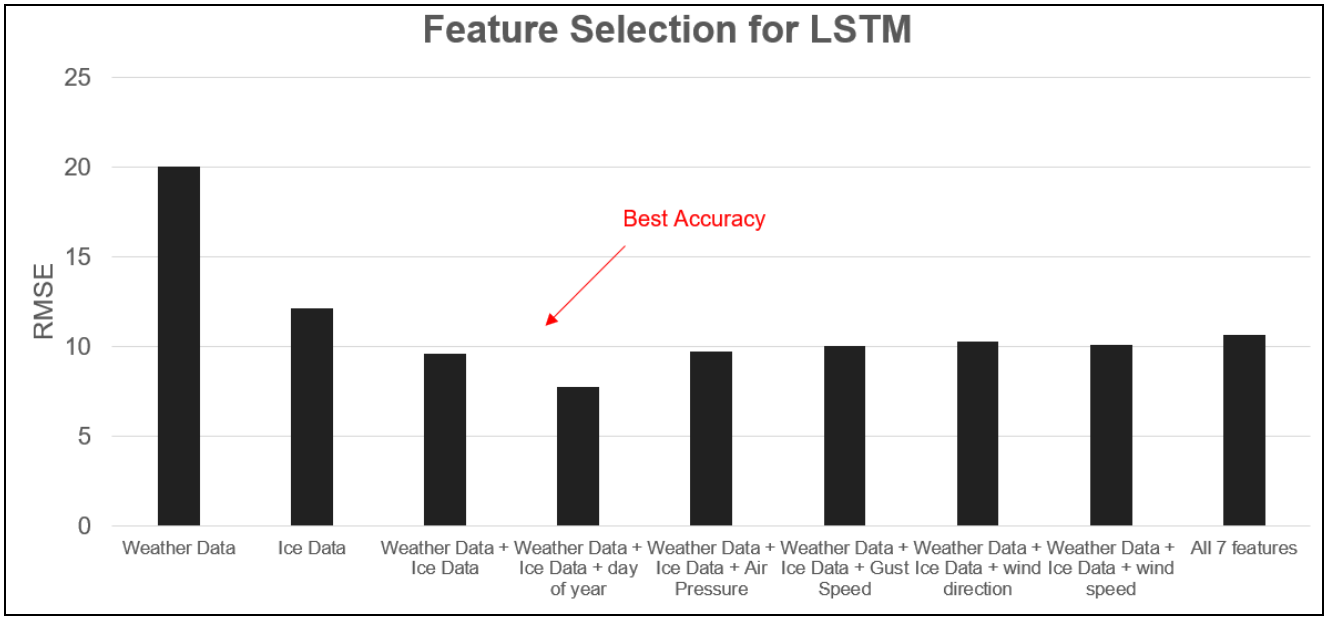


Fig 13. Feature selection for LSTM

Feature importance indicates how useful each feature was in the construction of decision trees within the model. The more the feature is used to make a decision, the higher is its relative importance.The trained XGBoost model calculated the feature importance on the predictive modeling problem. The XGBoost model was run again using the important features identified along with Ice Cover data and showed better accuracy.For XGBoost model:

- The Air temperature (ATMP) feature showed seasonality across the years of our study 2007-2019. The features Gust Speed (GST) and Wind Speed(WSPD) showed strong correlation and indicated that they be used as supporting data to predict ATMP. No correlation was observed between other variables. In general, ATMP correlations between the stations DTLM4, SWPM4, WNEM4 and LTRM4 were positive and were greater than 0.97. This shows us that we can explain the variation in ATMP at one station by understanding the variation in ATMP at another station.Highest correlation was between stations WNEM4 and LTRM4. Lowest was between SWPM4 and DTLM4.
- Feature engineering indicates creating or deriving additional predictive features . For XGBoost, additional features were created that could potentially aid in improving the RMSE of the XGBoost model. The following features were engineered to run the model:
    - The initial models were built by calculating the mean of features for the day. Built models with median values as well.
    - From the Wind Direction (WDIR), which is in degrees from North, exact direction of wind was created. That is, if the wind direction is 90, it is considered East. 180 is South, 270 is West etc.
    - For PRES feature, z scores were generated as a feature.

  It was observed that these features did not contribute significantly to improve the XGBoost model. The other features do not have predictive power. Ice data is more significant than ATMP for prediction. When the XGBoost model was run using only the significant variables it performed well. The error showed similar variance in test and train

data sets. There was also no overfitting of the model. Using features too far in the past does not add much value to increase the predictive power of the XGBoost model.

feature	importance
Ice_lag_1	0.607106
Ice_RM_3	0.151491
Ice_RM_4	0.043230
ATMP_RM_5	0.019847
Ice_lag_2	0.017084
Ice_lag_3	0.015140
Ice_RM_5	0.013205
ATMP_RM_4	0.012299
Ice_lag_5	0.011627
ATMP_lag_2	0.011431
ATMP_RM_3	0.010763
ATMP_lag_3	0.010305
ATMP_lag_1	0.009690
ATMP_lag_4	0.009444
ATMP_lag_5	0.008982
Ice_lag_4	0.008949
ATMP	0.008409
WSPD	0.006848
GST	0.006797
WDIR_New	0.006444
PRES	0.006387
Wind_North-West	0.004524
Wind_South-West	0.000000
Wind_South-East	0.000000
PRES_zscore	0.000000

Fig 14. Feature importance for XGBoost

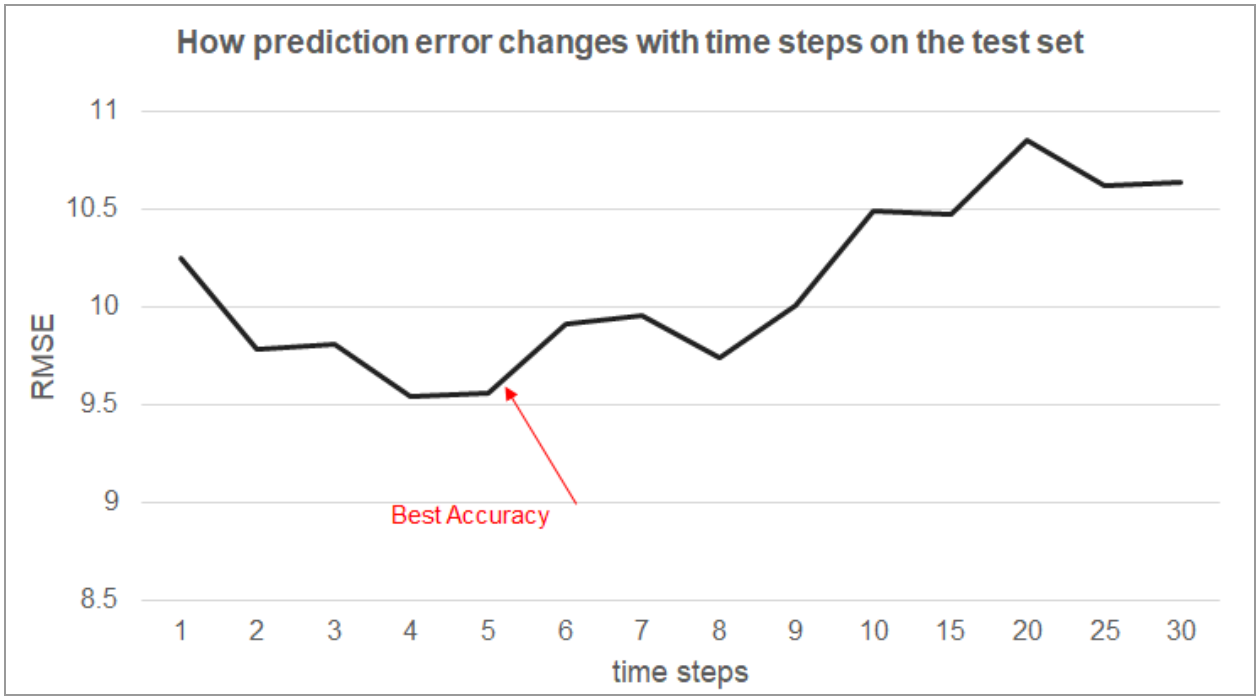


Fig 15. How accuracy changes with time steps on the test set for LSTM

Appendix C - Additional Information

1. Weather Stations' on St.Marys River coordinates:
    - SWPM4 : S.W. Pier, MI, 46°30'5" N 84°22'20" W
    - LTRM4: Little Rapids, MI, 46°29'9" N 84°18'6" W
    - WNEM4 : West Neebish Island, MI, 46°17'5" N 84°12'35" W
    - RCKM4 : Rock Cut, MI, 46°15'51" N 84°11'28" W
    - DTLM4 : De Tour Village, MI, 45°59'33" N 83°53'54" W

2. Feature description:

Table 2 : Weather Station features and their description.

WDIR	Wind direction (the direction the wind is coming from in degrees clockwise from true N) during the same period used for WSPD.
WSPD	Wind speed (m/s) averaged over an eight-minute period for buoys and a two-minute period for land stations. Reported Hourly.
GST	Peak 5 or 8 second gust speed (m/s) measured during the eight-minute or two-minute period. The 5 or 8 second period can be determined by payload.
WVHT	Significant wave height (meters) is calculated as the average of the highest one-third of all of the wave heights during the 20-minute sampling period.
DPD	Dominant wave period (seconds) is the period with the maximum wave energy.
APD	Average wave period (seconds) of all waves during the 20-minute period.
MWD	The direction from which the waves at the dominant period (DPD) are coming. The units are degrees from true North, increasing clockwise, with North as 0 (zero) degrees and East as 90 degrees.
PRES	Sea level pressure (hPa). For C-MAN sites and Great Lakes buoys, the recorded pressure is reduced to sea level using the method described in NWS Technical Procedures Bulletin 291 (11/14/80). (labeled BAR in Historical files)
ATMP	Air temperature (Celsius). For sensor heights on buoys, see Hull Descriptions. For sensor heights at C-MAN stations, see C-MAN Sensor Locations
WTMP	Sea surface temperature (Celsius). For buoys the depth is referenced to the hull's waterline. For fixed platforms it varies with tide, but is referenced to, or near Mean Lower Low Water (MLLW).

DEWP	Dewpoint temperature taken at the same height as the air temperature measurement.
VIS	Station visibility (nautical miles). Note that buoy stations are limited to reports from 0 to 1.6 nmi.
PTDY	Pressure Tendency is the direction (plus or minus) and the amount of pressure change (hPa)for a three hour period ending at the time of observation. (not in Historical files)
TIDE	The water level in feet above or below Mean Lower Low Water (MLLW).