Peter Annable <pannable@indiana.edu>
Indiana University Applied Data Science Final Project, May 2, 2018

Effect of Climate Change on Lake Matinenda Ice-Out Dates

Lake Matinenda in mid-February, 2006. (ice-out was on April 19 that year)

Introduction

My family has owned a cottage on Lake Matinenda in Blind River, Ontario since 1970. During the past 30 years, I've personally observed a number of unusually warm winters during trips to ski there, especially during the last 15 years. I wanted to understand if there truly has been a warming trend in local winters, and how they might correlate to regional and global changes. Sometimes this is known as "weather vs. climate". That is, local variation in daily or seasonal weather is incorrectly attributed to climate change. I would like to understand if my personal observation has a statistical correlation to broader climate trends.

The term "ice-out" refers to the date on which a lake becomes navigable again after winter ice has melted. The day on which ice-out occurs is usually a good indicator of how severe the winter has been that year in terms of temperatures and overall precipitation. This date tends to be highly variable from one year to the next. In the case of Lake Matinenda, ice-out typically occurs sometime between a 30-day period beginning April 10, and ending May 9.

There is a large body of data documenting the increase in the Earth's global surface temperatures during the last 137 years. Since 1932 there has been an overall increase in average surface temperature of 1.13° Celcius. If overall temperatures are increasing, it seems logical that ice-out dates would be occurring earlier on average. And in fact, this seems to be the case, as was analyzed by Sam Bledsoe. Thanks to record keeping by several cottage owners dating back to 1932, he found a decrease in the average ice-out date of about 3.15 days between 1932 and 2014.

In this paper, I will analyze the trend in ice-out date change over time, identify correlations with other local, regional and global trends, and attempt to project the future trends of ice-out dates.

Data

Sources

The following local, regional and global sources will be used for analysis. The table below describes each source and its purpose:

Peter Annable <pannable@indiana.edu>
Indiana University Applied Data Science Final Project, May 2, 2018

Data Type	Source	Measures and Purpose
Ice-Out Dates	Matinenda Cottage Owners Site	Ice-out date for past 82 years
Blind River Daily and Monthly Weather	Environment Canada Historical Climate Data	Max, Min and average local temperatures, and total precipitation for past 82 years.
Blind River monthly temperature averages	Farmzone.com	Average Max, Min and average monthly temperatures.
North America Snow Cover Extent	Rutgers University Global Snow Lab	Total square kilometers with snow cover by month. Gives regional data for total snow received each season.
Great Lakes Surface Temperatures	NOAA Great Lakes Coast Watch	Daily average surface temperatures of the Great Lakes. These bodies of water have a big influence on the weather on Lake Matinenda. They also act as something of a proxy to understand how warm a particular season has been.
Global Surface Temperatures	NASA Global Climate Change Center	Change in global surface temperature relative to 1951-1980 average temperatures
Global Co2 Concentrations	NOAA Earth System Research Labs	Monthly and annual Co2 parts per million in the atmosphere.

Data Preparation Steps

Each of the sources listed above provided data that was downloadable, but in widely different formats. For each data set, the data was acquired, harmonized and stored in separate excel worksheets. This step took a significant amount of time, the below sections highlight the key steps for each source:

Ice-Out Dates: these were provided from Sam Bledsoe's original analysis and used as-is. The Julian Data calculation was corrected to properly account for leap-years.

Blind River Historical Weather: To assemble a complete set of monthly data back to 1932, several weather station data sets were used from the Environment Canada site to get a complete set of data:

- Blind River
- Blind River Hydro
- Elliot Lake, and the Elliot Lake Stanleigh and Dennison locations
- Gore Bay
- Massy
- Mississagi
- Sault Saint Marie (only for 6 months in 2012)

These locations are all within 100 km of Lake Matinenda, except for Sault Saint Marie, and have similar weather. Sault Sainte Marie could have significantly different precipitation amounts, however only 6 records were used to fill in missing data for 2012.

North America Snow Cover Extent: This provides monthly data from 1967 to present on the total square kilometers in North America with snow cover. This was downloaded as is. The data was pivoted to provide and annual summary number for each year, providing an overall understanding of snow extent for each winter.

Great Lakes Surface Temperatures: NOAA has daily average surface temperatures for each of the Great Lakes from 1995 to present. These were downloaded for each lake, assembled into one file and pivoted to provide monthly and annual average temperatures for each lake.

Global Surface Temperatures: NASA provides a download file of actual and smoothed average temperature of the earth's surface by year, from 1880 to present. I was able to use this file as is.

Global Co2 Concentration: NOAA provides these by month from 1958 to present. These were loaded and pivoted to provide an annual average. The Interpolated Value was used for analysis, which corrected for errors in the raw data.

Final Data Assembly

To provide a unified set of features for analysis, the AnnualData.xlsx file was prepared, having one line per year with the following measures:

- Year
- Ice-Out Date
- Ice-Out Month
- Ice-Out Julian Day
- Ice-Out JTrend (the calculated Julian day based on trend line equation Julian Date = 192.21 - .0384*Year)
- Ice-Out Vs Trend ("Late" if actual Ice out was later then trend date, or "Early" if less)
- Global Temp Departure
- Global Temp Departure-Smoothed
- Global Co2 Conc
- Lake Superior Avg Temp
- Lake Huron Avg Temp
- Lake Michigan Avg Temp
- Lake Erie Avg Temp
- Lake Ontario Avg Temp
- Blind River Avg Low Temp
- Blind River Avg High Temp
- Blind River Total Snow (cm)
- Blind River Total Rain (mm)
- Blind River Total Precip (Rain + snow) (mm)
- ColdFactor: this was calculated as the sum of average local temperature for Jan-Mar. This gives a sense of the overall winter temperature and conditions for ice formation.
- ColdFactor Trend: The coldfactor trend number calculated using linear trend equation:
- ColdFactor Vs Trend: "Colder" or "Warmer" vs. Trend

Analysis Methods

Data Analysis was approached in three general phases, using a mix of tools for each:

1. Data Understanding: Tableau Visualizations, and R scripts for visualization and statistical understanding. In particular, data trends were identified and visualized

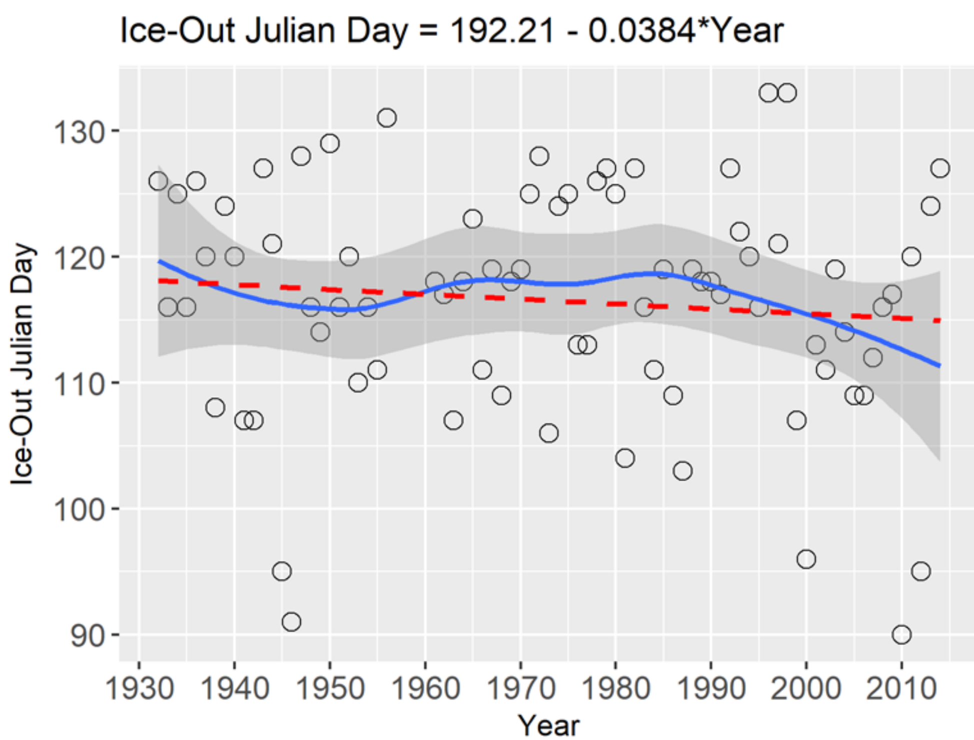
2. Correlation using Tableau.
3. Modeling and Future predictive algorithms using R.

In the results section, additional details will be provided about specific techniques used.

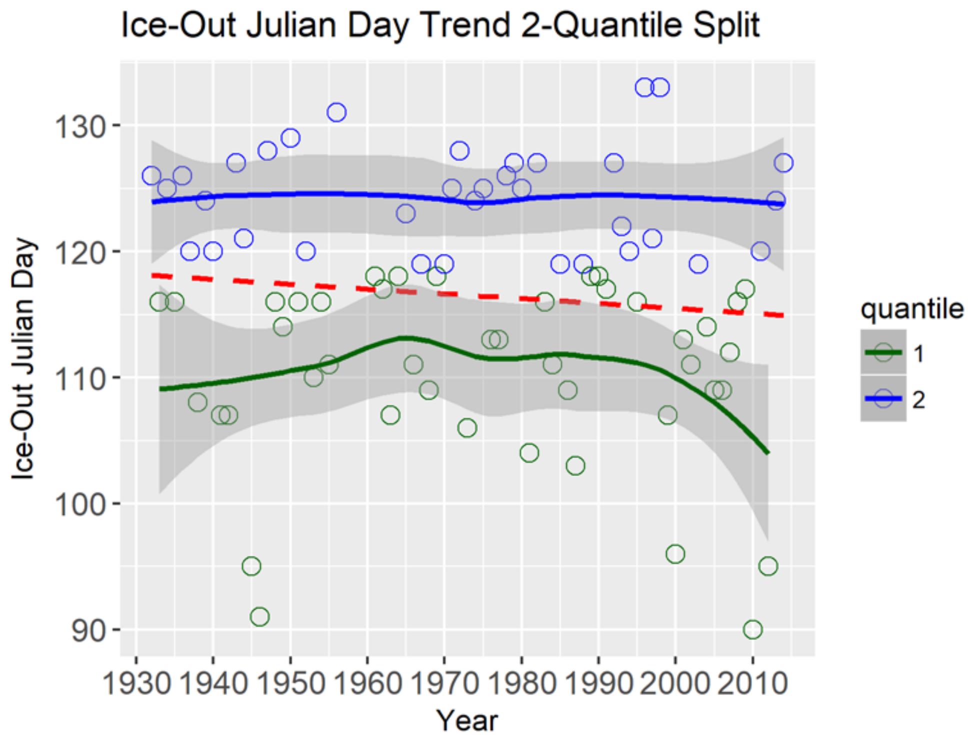
Discussion of Results

Ice-Out Trend Analysis

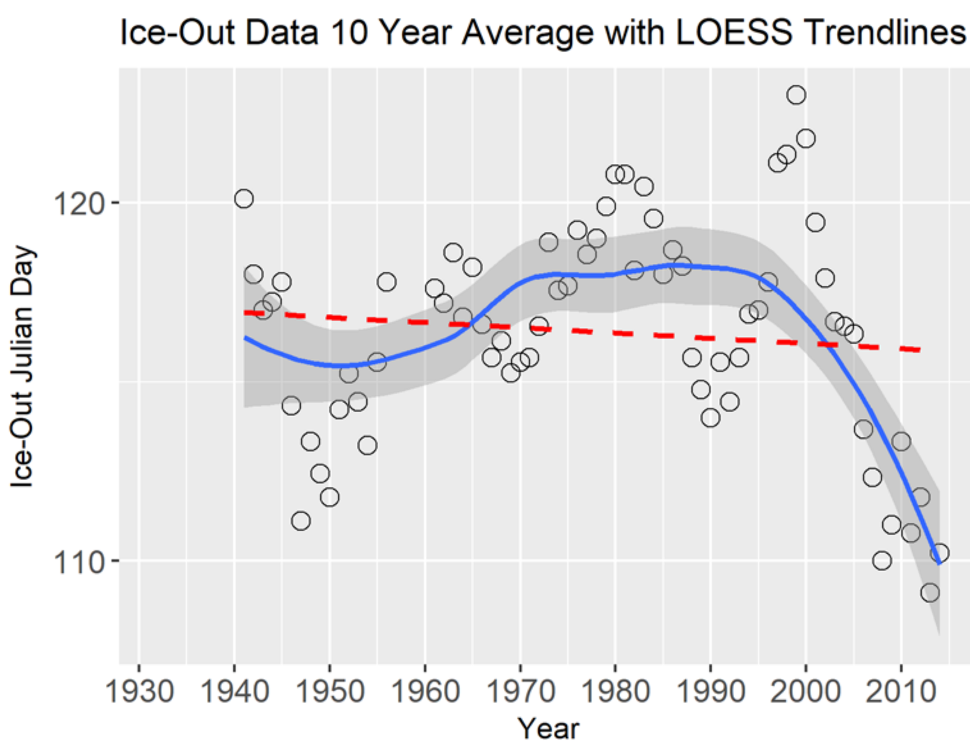
Using an R scatter plot, the Ice-Out Julian Day was plotted by Year. Both a linear model trend line and a LOESS (Locally Weighted Scatterplot Smoothing) line was added, with error bars shaded in gray. Here we can observe the overall trend downward in the Ice-Out date from 118 days in 1932, to 115 days in 2014. On average, Ice-Out now occurs three days earlier.



To further understand the trending, Ice-Out Julian dates were ranked in 2 quantiles, with a LOESS line for each. Here we can see that while Ice-Out dates above the median date are generally flat, the Ice-Out dates below the median are trending downward faster, especially since 1990. Overall this would indicate Ice-Outs will both be earlier on average, but also with higher variance.



Finally, as was done in Sam Bledsoe's original analysis, the ice-out Julian dates are plotted as a rolling 10-year average. This smooths the data and makes the recent trend in earlier ice-out timings more evident. Also notice the narrower confidence bands in light gray, indicating a higher R^2 correlation factor.

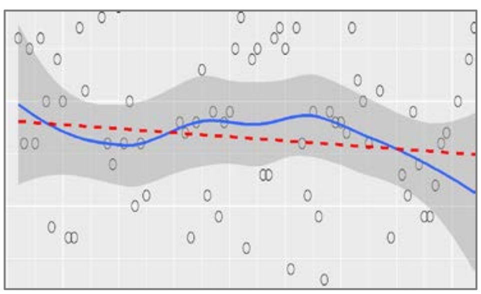


Peter Annable <pannable@indiana.edu>
Indiana University Applied Data Science Final Project, May 2, 2018

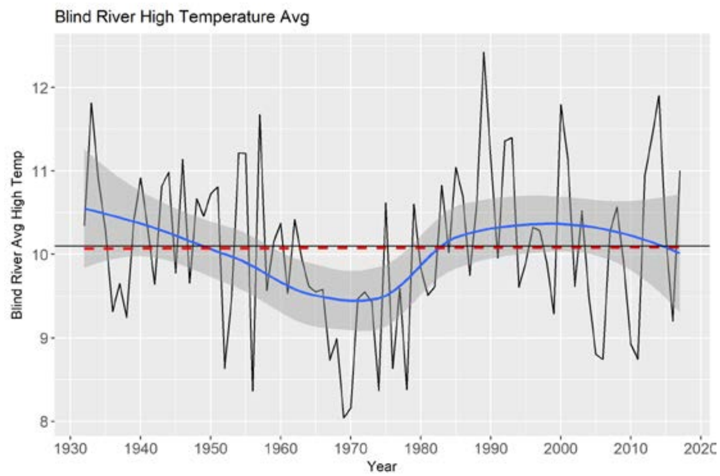
Blind River Local Temperature Trend Analysis

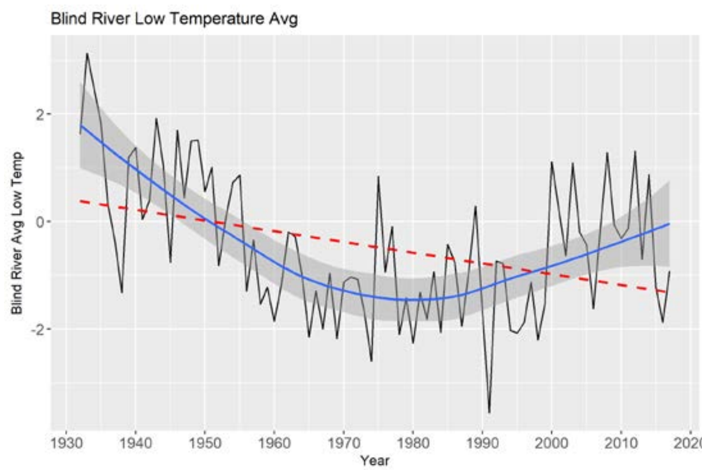
Using the annualized average high and low temperature, we can see overall temperatures trended downwards until the 1970s and have since then generally been rising. This is also true of the average low temperature for the Jan-Mar months each year.

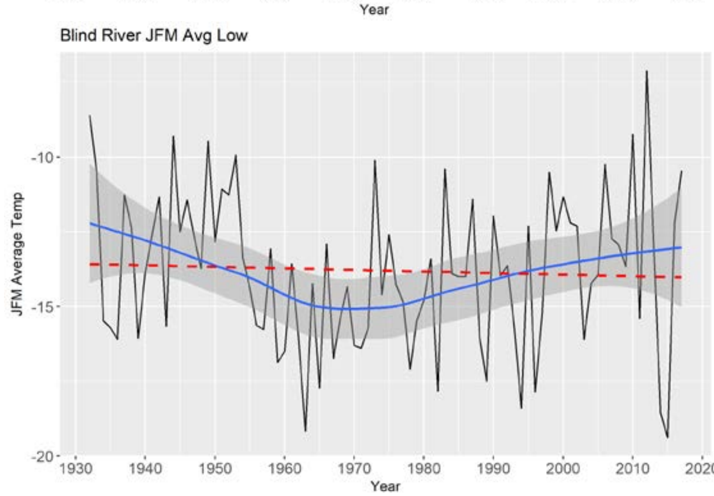
This appears somewhat consistent with the smoothed average Ice-Out dates in first chart, which exhibits the opposite shape, although not as strongly. This part of the graph is show below, slightly distorted vertically to show its shape. This will be further analyzed in the correlation analysis section



Ice-Out Dates Trend Shape







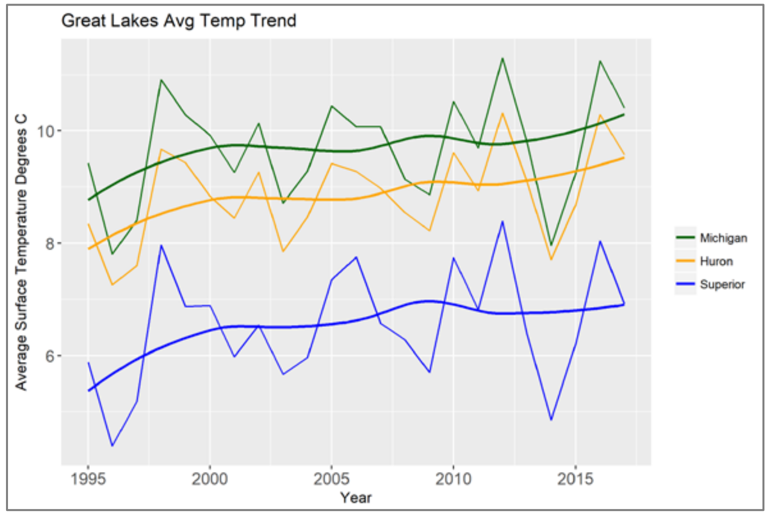
Peter Annable <pannable@indiana.edu>
Indiana University Applied Data Science Final Project, May 2, 2018

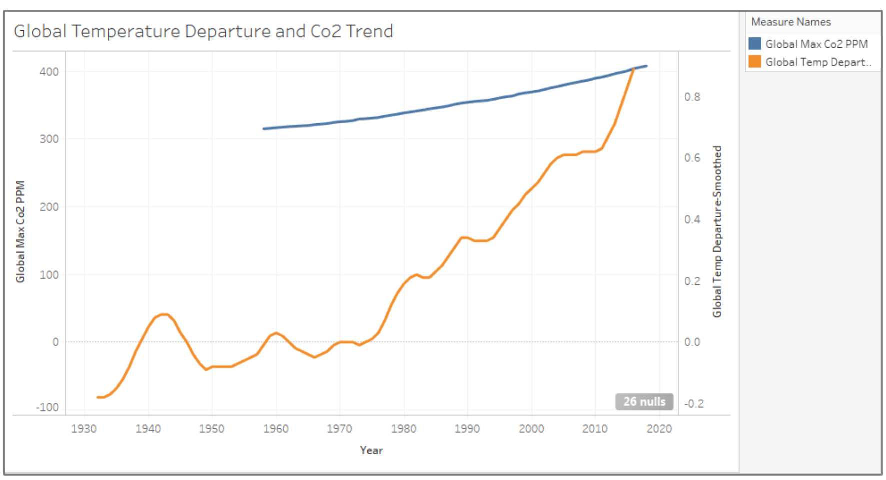
Regional and Global Trends

On a regional and global basis, the trends are very clear: the Great Lakes are consistently getting warmer since 1995, when data becomes available. Only Superior, Michigan and Lake Huron are shown since they are the closest to Blind River. The surface temperatures over time track very closely in shape.

On a global level, the average surface temperature is trending upward consistently, especially since 1970.

Average Co2 levels in the atmosphere are also increasing steadily during this time from about 320 ppm in 1958 to 410 ppm in 2018.



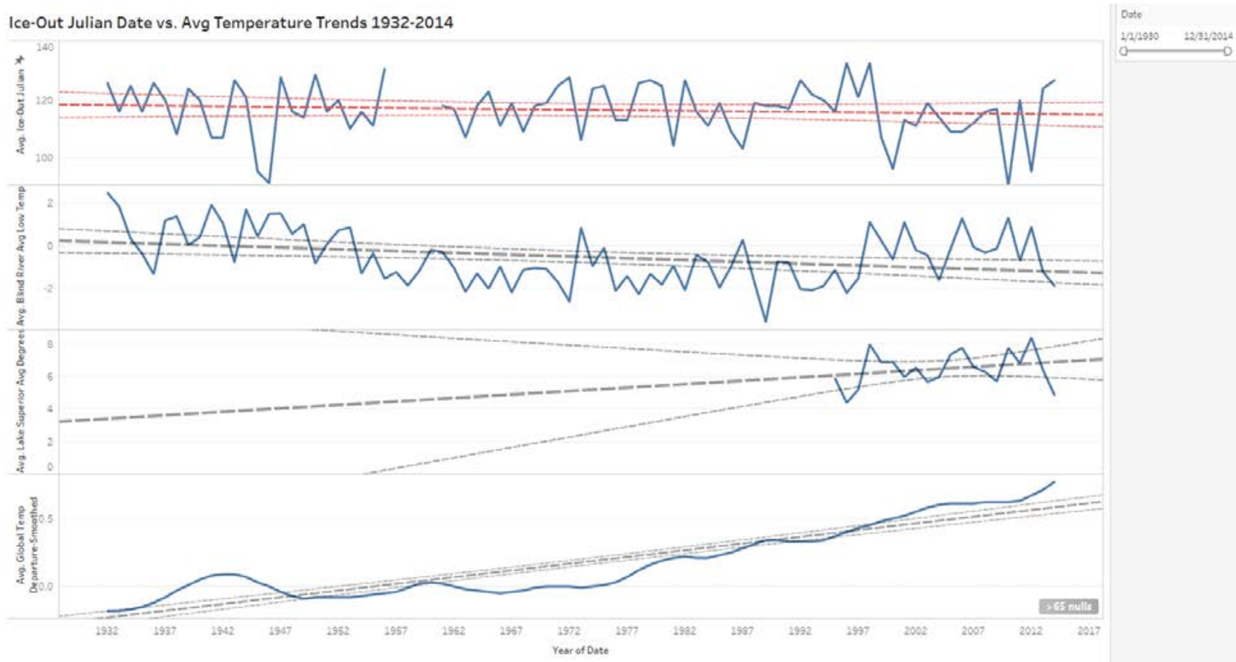


Correlation Analysis

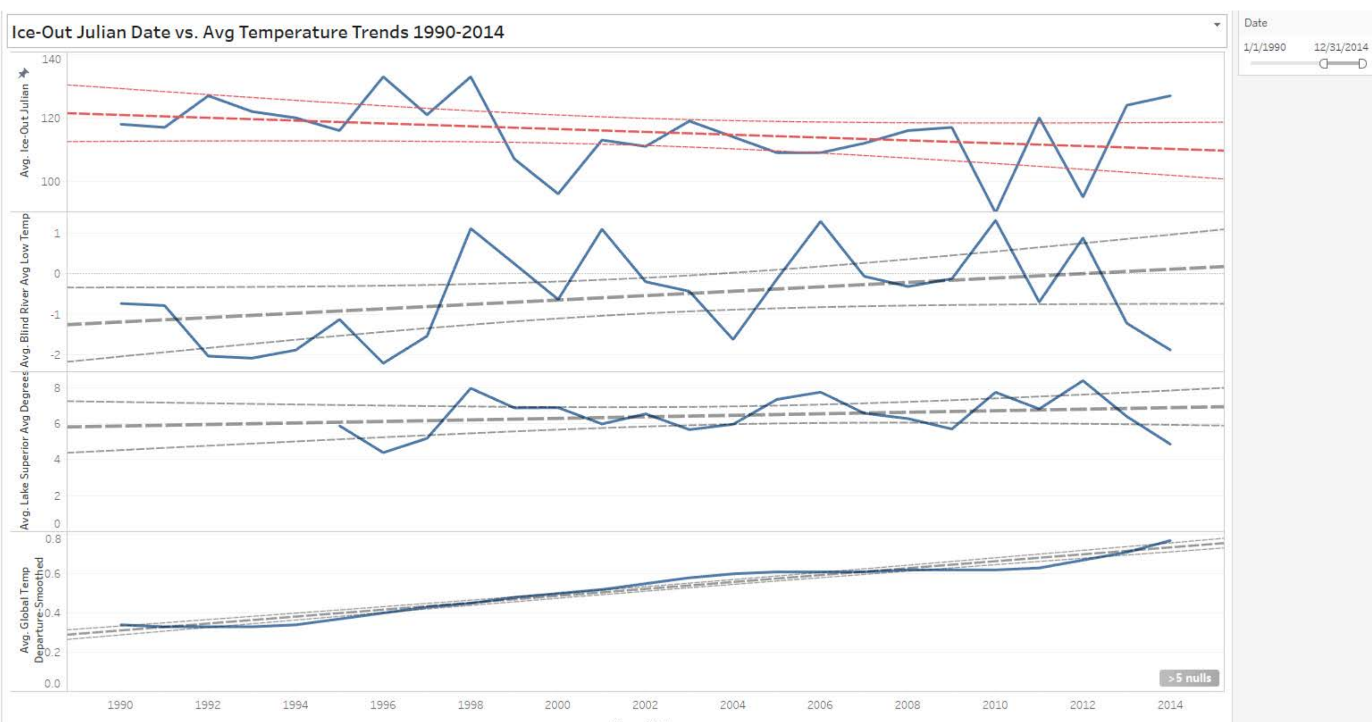
Based on the data trending and visual analysis, long term Ice-Out dates appear to be correlated both to the long term local temperature averages, and to the global trend towards increasing temperature, especially since 1970. That trend is also mirrored in the Great Lakes temperature data available since 1995.

It should be noted that the R^2 value for the Ice-Out data itself is fairly weak at .0113. This is due the high variability in the data. When smoothed using the 10 year averaging, R^2 jumps to .19. And using the higher order trend line, R2 jumps to over 0.6

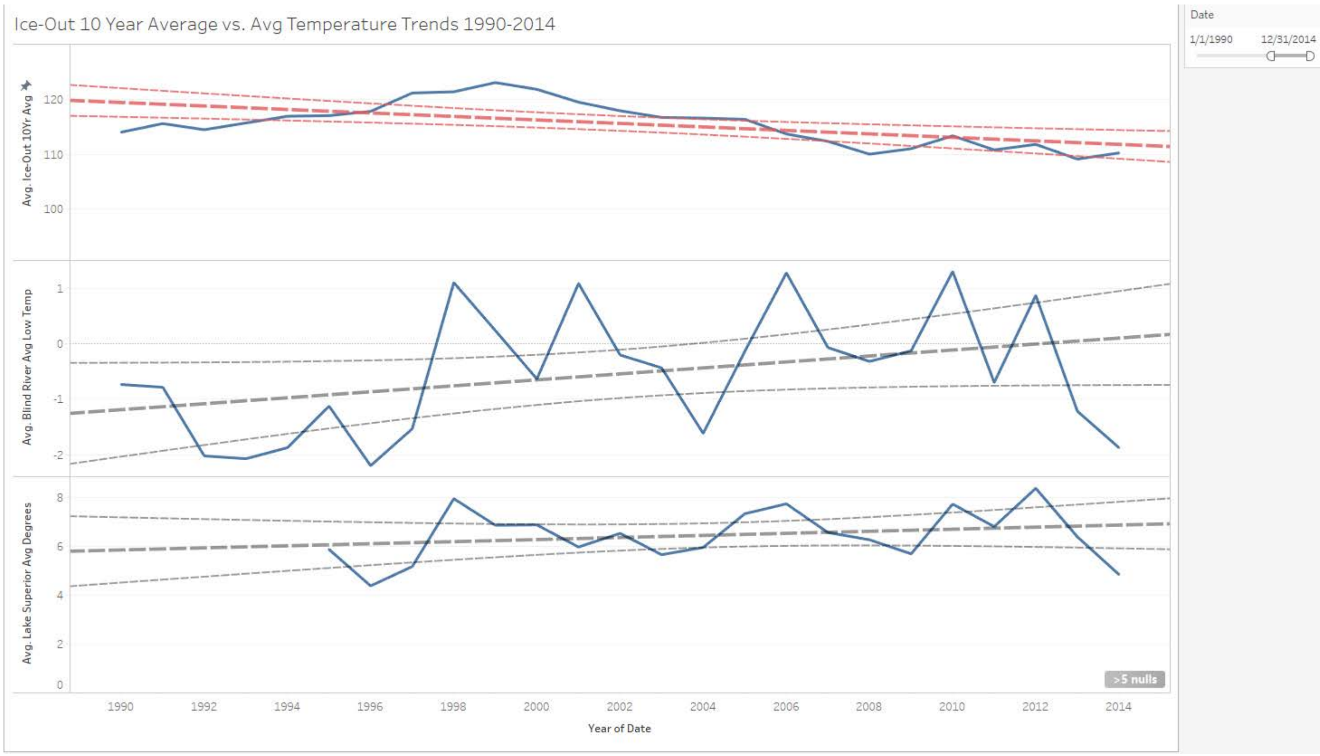
To see the trends visually, comparing Global Temperature Change, Average Surface Water temperature of Lake Superior, Average Low Temperatures in Blind River, and average Ice-Out date, the following chart was rendered using Tableau:



Here we can see the (subtle) downward slope on ice-out dates, increasing global and regional temperatures, but an overall downward trend on the local Bind River Temperatures. However, if we zoom in to data since 1990, the picture changes. Now we can see all temperatures are rising on average, while ice-out times are dropping.



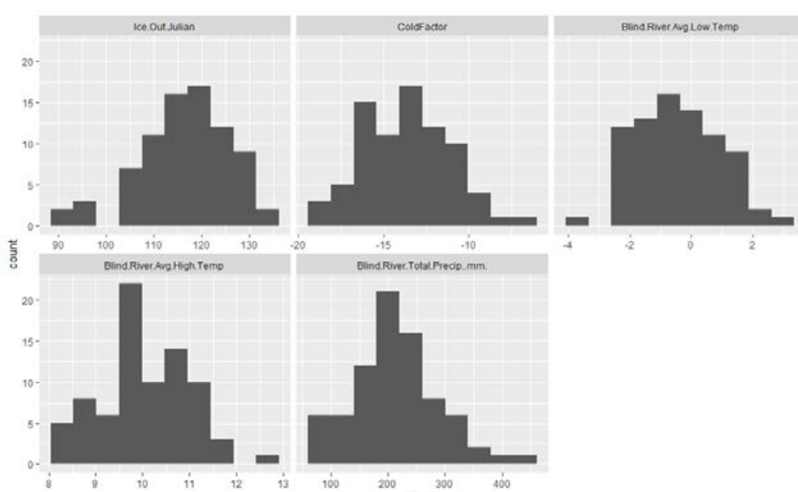
Lastly, if we use the 10-year Ice-Out Average, and zoom in a bit by only looking at the average Blind River Temperatures and Average Surface Temperature of Lake Superior, the trend looks more dramatic:



Since 1990, the rolling 10-year average ice-out date has changed from day 120 to 112 while the average JFM temperature has increased from -1.2 to just above 0 degrees Celsius.

Prediction Models

So far, we've seen that ice-out times are highly variable by year, and many of the data elements used so far have a linear, non-Gaussian distribution. However, there are a few variables that will be useful for prediction algorithms, with histograms shown below.



These are the ColdFactor, Average Temps, and total precipitation per year. Given that these are variables we only know after the fact, the best we will be able to do is predict an Ice-Out date with minimal advance timing. Possibly in March we can predict the Ice-Out time for April to May.

To determine if this was the case, two models were selected after some trial and error with Linear Models, Logistic Regression, Artificial Neural Networks, and Naïve Bayes.

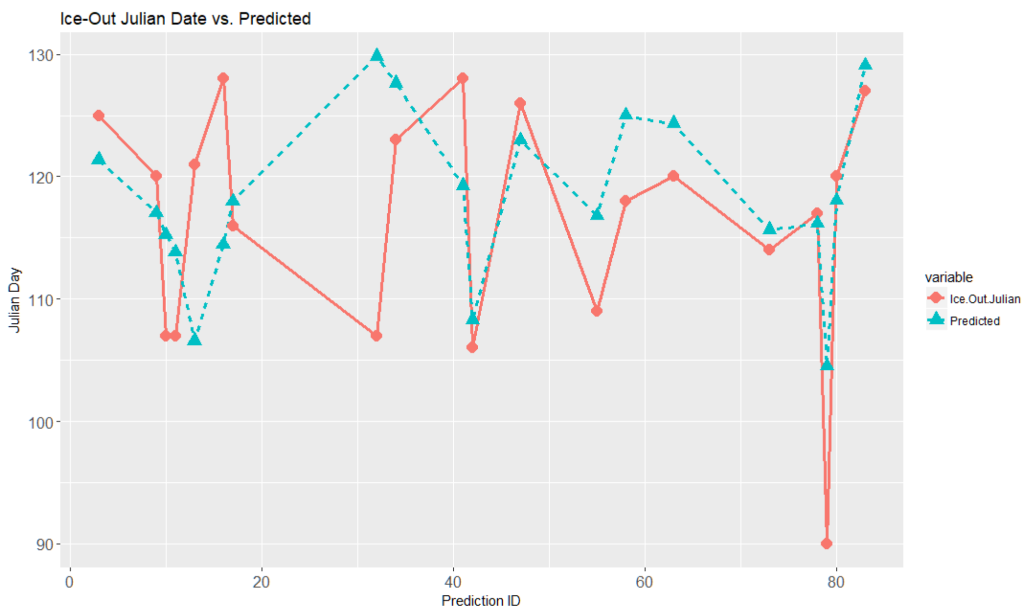
Peter Annable <pannable@indiana.edu>
Indiana University Applied Data Science Final Project, May 2, 2018

Linear Model to Predict Ice-Out Julian Date

The available data was randomly split into a training set using 75% of the data, and 25% for test. A linear model was fitted to the training data and used for prediction using the following R code:

```
lm <- lm(Ice.Out.Julian ~ Global.Temp.Departure+ColdFactor+Blind.River.Total.Rain..mm.+
            Blind.River.Total.Snow..cm., data=tr)
ps <- predict(lm, ts)
```

Average absolute error for the predictions was 6.7 days. This is best seen in visual form, with a couple exceptions the predictions follow the actual reasonably closely.



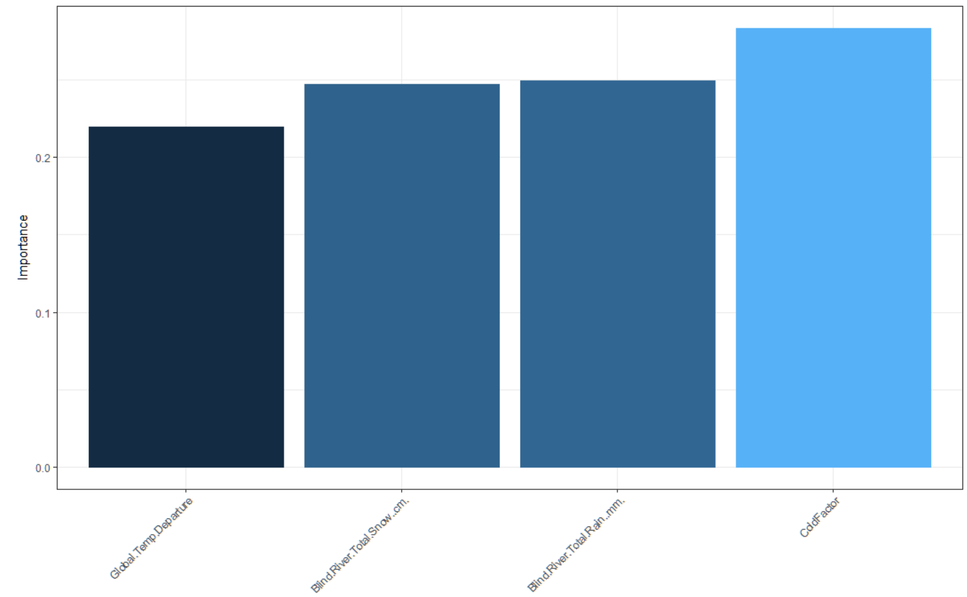
ANN to predict Early or Late Ice-Out vs. trend

A potentially more reliable prediction is to predict if Ice-Out will be Early or Late vs the historical average. This was done by adding an Early/Late classification column to the data, and the fitting an artificial neural network with four neurons to the training data with same features as the linear model. As before, 75% of the data was used for training, 25% for test.

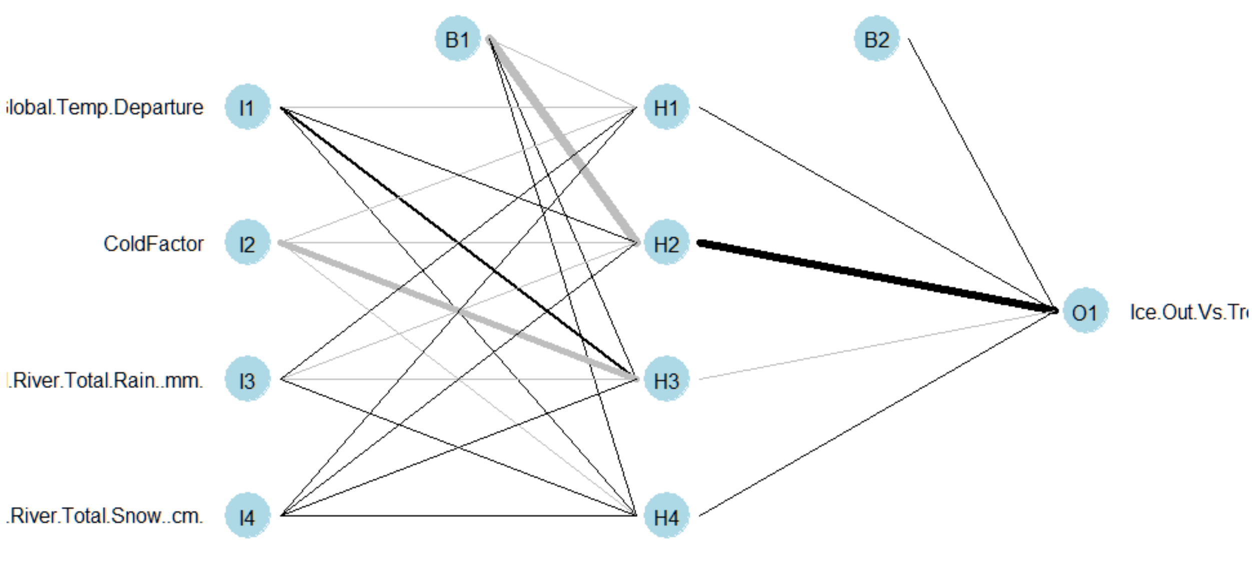
```
n <- nnet(Ice.Out.Vs.Trend ~ Global.Temp.Departure+ColdFactor+Blind.River.Total.Rain..mm.+
            Blind.River.Total.Snow..cm., data=tr, size=4 ,trace=FALSE, maxit=10000, decay=.001) # train
an ANN over ice out data
ps <- predict(n, ts, type="class")  #  classify test data

Confusion Matrix:
ps       Early Late
  Early     5    1
  Late      3   11
> 100*(1-sum(diag(cm))/sum(cm))  # determine error rate
[1] 20
```

Again the results were pretty encouraging, only a 20% error rate in predicting Early vs. Late.

The importance of the ANN features are plotted below. As we might expect, the ColdFactor, or sum of average temperature for the January-March period was the most important feature.



The resulting Neural Net is shown below:



Peter Annable <pannable@indiana.edu>
Indiana University Applied Data Science Final Project, May 2, 2018

Conclusion

It seems clear that the average ice-out date is decreasing over time, especially since 1990. However, the high variability makes further assumptions about change fairly uncertain. It seems likely that the average ice-out times will decrease, but total variability is likely to increase. In fact, as this analysis is being written, the Northern Ontario region experienced an unusually cold March and April. Ice-Out has not occurred yet and is likely going to be on the late end of its range, possibly as late as May 10.

Average temperatures are also rising globally, regionally and locally. This correlates well with the ice-out trends. There is strong evidence that the rising global temperatures are driving these other changes, but this analysis was not robust enough to conclude that definitively. Further data is needed and the next 10-20 years will be important to understand if trends will go beyond their current 100-year range. A deeper analysis using daily local weather data may provide some additional understanding and ability correlate data with more confidence.

Predicting the specific ice-out date is only possible within a month or two of the actual date. It may be possible to find indicators before the beginning of winter, but at best they are likely to only show if the ice-out will be early or later than normal.

As a final analysis, the ice-out data since 1990 was trended and projected to 2030. If trends continue as they have for the last 28 years, <u>ice-outs could be as early as March 26 in 2030</u>. At the same time, ice-outs in early May could also continue as seasonal variations become more extreme.

So was my personal observations of the unusual winter weather on Lake Matinenda indicative of a broader trend? The answer is yes, at least based on the last 80 years of data we have.