Abstract

Predicting the onset of river ice breakup has proven to be a difficult challenge due to the complex nature of the relationship between meteorological factors, streamflow hydraulics and river ice mechanics. This study aims to relate Alaska's climate variables to a unique dataset of yearly breakup dates for the Tanana River, by means of a machine learning regression model. First, a set of meteorological and environmental variables was established by investigating the ice breakup process, relevant previous research and various data from the site. Subsequently, a Random Forest (RF) regression model was conducted for different combinations of variables (configurations). The model configuration B (including temperature and solar heat input variables) performed best, and 5-fold cross validation reported for the prediction a RMSE and MAE of 2.80 and 2.26, respectively. Furthermore, the model performed worse on predicting outliers of the breakup date. In order to serve as a forecasting tool for the Nenana Ice Classic betting competition an attempt was made at conducting a forecasting model. Bounding the model input data to approximately one month before the breakup date resulted in significantly larger errors. While further research is recommended, the results from this study suggest Random Forest regression models have great potential in capturing the complex relationship between the river ice breakup date and environmental variables close to the breakup event itself.

Contents

Introduction

The Nenana Ice Classic betting competition began in 1917 when railroad engineers and gold miners, in search of entertainment, started betting money on guessing the exact date the ice in the Tanana River would break up (precise to the minute). Since then, Alaska residents each year have made an attempt at guessing the timing of the ice breakup correctly. The winning price was as high as 300000 dollars in recent years (Nenana Ice Classic, 2020). A tripod that is connected to a clock with a string is placed two feet into the river ice during freeze-up in November (Meier and Dewes, 2020). The following spring, when the ice starts to melt and move, the clock stops as the tripod pulls the string. This exact moment has been recorded for each year since 1917, yielding an extraordinary dataset.

Generally speaking, the ice breaks up in late April or early May, but a trend towards earlier breakup can be perceived. This dataset could be used for research on the process of ice breakup as well as climate change impacts for the Interior of Alaska. A wide range of factors likely influence the river ice breakup process. Examples are air temperature, ice thickness, snow cover and river discharge. Questions that arise are for example, how does river discharge influence the ice breakup date? Can the effects of climate change as mentioned in literature, especially the more frequent occurrence of extreme weather events like heat- and cold waves, be perceived from the data set? Subsequently, the question arises how the relationship between the variables and the breakup date can best be captured in a nonlinear numerical (Machine Learning) model in order to describe and predict the breakup process. What is the practical applicability of such a model, and how will it compare to its linear predecessor as conducted in previous research by Van Asselt, 2020, or other Machine Learning models like the Artificial Neural Network from Zhao et al., 2012?

The report is structured into six chapters, working its way up to the results in order to answer the research questions. Chapter 1 will establish a theoretical framework for the reader and will consist of an overview of previous research on the topic, as well as an overview of climate and environment variables that play a role in determining the breakup date. In Chapter 2, the ice breakup process will be examined in further detail and previous work on breakup date forecasting models will be highlighted. Next to that, the effect of heat waves and river discharge will be discussed based on data analysis, in preparation of the modelling. Subsequently, Chapter 3 will introduce the Machine Learning model concepts. Chapter 4 will describe the specific modelling approach and evaluation methods. Chapter 5 will present and discuss the results of the modelling approach. Finally, the study is concluded in Chapter 6.

Figure 1: Nenana Ice Classic watchtower at the Tanana River waterfront. *Bengel, 2021*

1

Context

This chapter introduces the context for this research, consisting of previous BSc Thesis research on the specific subject of the Nenana Ice Classic and the research gaps that followed. Next to that an overview is given of the climate, environment and other subjects relevant for the Nenana Ice Classic.

1.1. Previous Research

This research will be based on the work by Van Asselt (2020). The following section will provide the reader with a brief summary of the research approach, conclusions and knowledge gaps that might be of interest for further research.

1.1.1. Summary

The main objective of this research was to analyse the climate and ice breakup process, in order to describe the breakup process in a linear regression model with climate and environment related variables. In his work, he first analyzed the research area's climate and environment on different scales. Based on the Representative Concentration Pathways (RCP) model by the Intercontinental Panel On Climate Change (IPCC) for future climate scenarios based on CO2 concentrations, temperature projections for Nenana were established to provide input for an ice breakup prediction model later on. Next, the ice breakup process was discussed and a set of variables that might be linked to the Nenana Ice Classic dataset was introduced.

By performing linear ordinary least squares (OLS) regression analysis, the possible relationships of these variables to the breakup date were quantified. This led to two (linear) models. The first model was set up to make ice breakup date projections based on the aforementioned IPCC climate scenarios for the coming 80 years. All scenarios would result in an earlier avarge breakup date, with a significant spread between the scenarios. For the most severe IPCC climate scenario of global warming (RCP8.5), "a decrease of 15 days in average ice breakup dates is predicted by 2100." (van Asselt, 2020). The second model was time bounded by the betting deadline of April 1st and was set up to forecast the breakup date as a tool for the betting contest. This restriction on the use of environmental variables close to the breakup date resulted in a "less reliable model that showed a lower predictability". The conclusion followed that environmental variables close to the breakup event have a large influence on the breakup date. The average temperature during April and May for example, showed to be of great influence.

1.1.2. Research gaps

Several recommendations for further research were made by Van Asselt (2020). First of all it is suggested to look for a relationship between the betting behaviour of Nenana Ice Classic participants and their awareness of climate change. This however would require manual processing of the Books of Guesses at the Ice Classic office in Nenana, Alaska, which is currently not feasible. Furthermore, improvement of the model performance could be possible by performing a more detailed analysis, using different metrics of the data or shorter time intervals. A nonlinear or even machine learning approach to the modelling of the relationship is suggested. Work of Zhao et al., 2012 can be used as a reference here. This might lead to new findings on the relationship between the bounded environmental variables and the ice breakup dates (van Asselt, 2020).

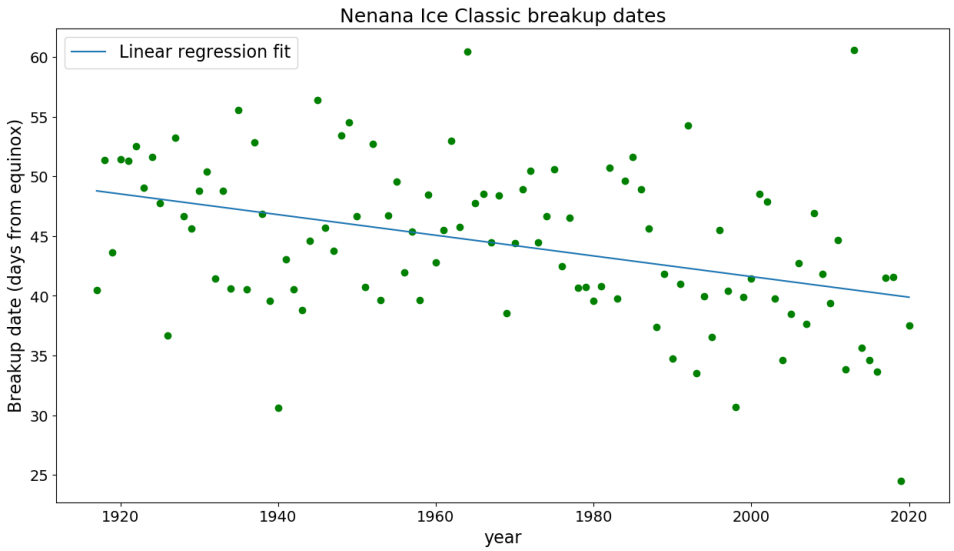


Figure 1.1: Nenana Ice Classic breakup dates with a linear regression analysis showing a downward trend, meaning earlier breakups are occurring more often in recent years. Research by Bieniek et al., 2011 into more river ice breakup measurements at different locations confirm this downward trend. *Based on a figure by Van Asselt, 2020*

Finally a more detailed investigation of local variables is recommended to improve the model, especially the river discharge.

It can be concluded further analysis of climate and river discharge data is necessary in order to better understand the breakup process. Using different metrics, e.g., the classification of heat waves from temperature data, might lead to new insights. Furthermore, an attempt will be made at describing the ice breakup process in a nonlinear Machine Learning model, like suggested by Van Asselt. The results and conclusions from the climate research and linear regression modelling by Van Asselt will be used to select promising input data for the nonlinear model.

1.2. Climate and Environment

The following section will provide an overview of the meteorological and environmental aspects relevant to the Nenana Ice Classic, working from the small scale of the Nenana city to larger scale of Alaska. While this section will mainly be based around the extensive analysis as carried out by Van Asselt, 2020, some new resources will be used to shape the analysis to the scope of this research.

1.2.1. Nenana

The city of Nenana is located in central Alaska, at the juncture of the Nenana River coming from the south, and the Tanana River from the east. About 40 km upstream the Tanana River, the large city of Fairbanks is located. Nenana has a small population of approximately 360 (“City of Nenana | Homepage”, 2017). Not only the existence of the city but also the Ice Classic betting contest have roots that go back to the gold mining industry around the year of 1920. The construction of the Alaska Railroad (ARR), connecting Fairbanks to the southern part of the state Alaska, played a major role in the development of the city of Nenana. The Mears Memorial Bridge, crossing the Tanana River, was completed in 1923 to facilitate this connection (Van Asselt, 2020). The city has experienced multiple floods in the last century, some caused by ’ice jams’. These obstructions in the river cause the river discharge to be blocked. In 1968, the Parks Highway Bridge was constructed (“City of Nenana | Homepage”, 2017). This highway is along with the ARR the main transport infrastructure leading south.

The Nenana Ice Classic takes place on the Tanana River in between the Mears Memorial Bridge and the Parks Highway Bridge, just before colliding with the Nenana River (see figure 1.2). After analysis of video material it is estimated the tripod is installed 160 m from the riverside, which is done by hand and therefore includes an error (Van Asselt, 2020).

Figure 1.2: Approximate location of the Nenana Ice Classic tripod, taken from OpenStreetMap

1.2.2. Tanana River Basin

The Tanana River is located in east-central Alaska and the 'Tanana River Basin' covers around 115.500 km^2. The largest tributaries are the Alaska Range (South) and Yukon-Tanana Uplands (North), with both glacier based inflow from the Alaska Range and non-glacial inflow from the Yukon-Tanana uplands. The glacier inflow represents 85% of the Tanana River inflow (Collins, 1990).

The climate of the Tanana River Basin can be characterized as continental (Collins, 1990). Long winters and short summers are normal with annual average temperatures around -3.5 °C and extremes of 35°C and -52°C. Annual precipitation, mainly snow, differs between 250-560 mm/year of water, equivalent to 760-1500 mm of snow per year. Permafrost characterizes the Tanana River Basin soil profile (Yarie et al., 1998, Van Asselt, 2020).

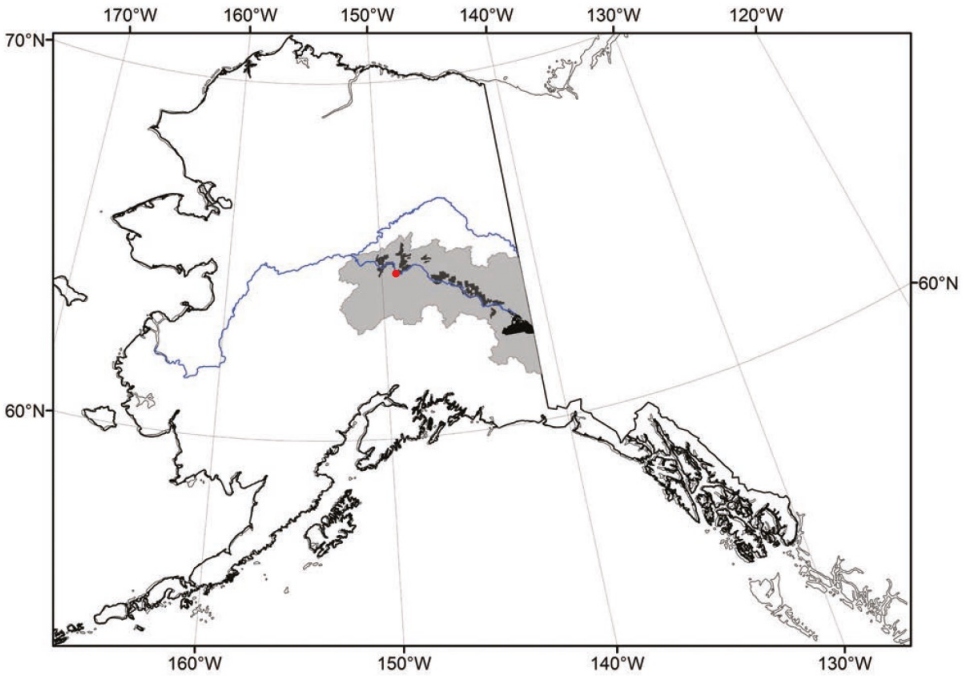


Figure 1.3: Map of Alaska, showing the Tanana River Basin in grey, the Yukon and Tanana Rivers in blue, and the City of Nenana as a red dot (edit from original figure by Andersen et al., 2018).

1.2.3. Alaska

Alaska, being the most northern state of the United States, has a subarctic climate (Kottek et al., 2006). The long term weather patterns are unsteady and significant changes in temperature and precipitation can occur in Alaksa as pointed out by Papineau, 2016. In his research, he linked these patterns to periodic, large scale *climate forcings*. The two forcings with the largest effect on the Alaskan climate are elaborated on below.

- Pacific Decedal Oscillation (PDO)

  PDO is the fluctuation of the sea surface temperature (SST) of the Central North Pacific Ocean with a phase of approximately 20 to 30 years, and a 2 to 5 year mini reversal occurrence. During a positive phase of the PDO, the SSTs in the Central North pacific are above normal, this results in warmer SSTs along the coast of Alaska.

- El Nino and Southern Oscillation (ENSO)

  The ENSO includes the El Nino and La Nina, which refers to oscillations in SST's in equatorial and sub-tropical zones of the Pacific and Indian Oceans. El Nino and La Nina have the opposite effect of eachother. El Nino is characterized by warming of the seawater, La Nina is linked to cooling of the seawater in these regions. Both events last for about 12-18 months and can be measured in terms of not only water temperature but also air pressure. During La Nina, polar jet air streams spend more time at lower latitudes, leading to cooler temperatures in Alaska. (Papineau, 2016, Van Asselt, 2020).

These large scale climate forcings were related to ice breakup by Bieniek et al., 2011. In his research, Van Asselt, 2020 based the PDO and ENSO variables for the linear regression model on Bieniek et al., as will be done here. Bieniek et al. concluded that river ice breakup is mainly influenced by local spring surface air temperatures and the river discharge. There also is a correlation between the two variables, since higher spring air temperatures cause more snow melt and thereby increase river discharge. Bieniek et al., 2011 also concluded that breakup is linked to ENSO-related climate anomalies, persisting from winter into spring. The large scale "climate-breakup" mechanism is summarized as follows: *"During El Ninõ in spring (AM), fewer storms occur in the Gulf of Alaska reducing cloudiness, warming air temperatures, and leading to earlier interior river ice breakup. (...) during La Ninã in spring (AM) more storms occur in the Gulf of Alaska increasing cloudiness, cooling temperatures, and resulting in later breakup."* Bieniek et al., 2011.

2

Ice Breakup

This chapter provides an overview of the river ice breakup process. The climate and environment related variables that influence this process are introduced, an overview on previous research on this relationship is given and relevant data is analyzed in preparation of conducting the numerical model. Finally, a set of input variables for the model is presented.

2.1. Process

2.1.1. Definition

River ice breakup is a yearly event taking place on rivers that have been frozen up during winter. Due to milder weather conditions during springtime, the ice that covers the river is broken, allowing flow of the water. Breakup ice jams frequently cause extreme flood events, which have major impacts on riverside communities and infrastructure, but also aquatic life and navigation. Therefore research output on this topic is significant (Beltaos, 2007). The ice breakup process is visualized as a succession of phases that can be described as *pre-breakup, onset, drive, wash* (Beltaos, 2003). During pre-breakup, the ice thickness and strength is thermally reduced. This means it is more susceptible to fracture. The onset of the breakup is then caused by increased flow discharge of the river, which is in turn caused by rain, warming weather and associated snow melt. The onset is followed by the drive, the movement of the now loose ice slabs by the current. Finally the wash is defined as the moment the ice slabs are transported downstream.

2.1.2. Ice Breakup Dynamics

Beltaos, 2003 and Jeffries et al., 2012 distinguish two breakup event extremes, the *thermal* or *overmature* breakup and the *premature* or *mechanical* breakup. The former type takes place when the mild springtime weather is accompanied by low river runoff, due to gradual snow melt and a lack of rain. This type of breakup can be seen as relatively slow and calm and frequently occurs at smaller rivers with low discharge magnitudes.

Premature or mechanical breakup is caused by high discharge magnitudes and results in mechanical fracture of the ice. The term premature is used because thermal processes have not yet affected the ice strength significantly. Rapid melt and heavy rain conditions upstream cause the high discharge. The hydrodynamical forces are sufficient to cause fractures and lifting of the ice, causing a phenomena characterized as 'ice push', where several breakups occur due to the impacts of ice from upstream (Beltaos, 1997). This type of breakup has potential for the forming of "ice jams", often resulting in flooding, with numerous socio-economic and ecological impacts (Beltaos and Kääb, 2014).

2.2. Forecasting models

The destructive character of river ice breakup and the resulting social relevance for understanding and predicting the phenomenon have resulted in several attempts to describe the breakup process in a forecasting model. Several models with different levels of detail were created up until now (Shulyakovskii, 1963, Beltaos, 1997, Hu et al., 2008, Kovachis et al., 2017). As concluded by Van Asselt, 2020, most researchers used the principle of driving forces and resisting forces in their modeling of the ice breakup. A variable can be seen as

either a driving force for earlier breakup (e.g. discharge, mild temperatures), or a resisting force causing later breakup (e.g. snowfall).

2.3. River Discharge

2.3.1. Relation to ice breakup

As mentioned earlier, the river discharge can be characterized as a driving force for earlier breakup and is associated with the mechanical breakup typology. The large scale meteorological effects causing this driving force is described by Bieniek et al., 2011. A high river discharge during April and May, due to high temperatures causing more rapid snow melt during this period, is mentioned as one of the key variables of influence. Similarly, Prowse et al., 2007 and Beltaos, 1997 mention this relationship in descriptions of the breakup process, and for example Zhao et al., 2012 and Wang et al., 2013 use the discharge variable as one of the key drivers in numerical models.

2.3.2. Discharge data

The discharge data that will be used for analysis is obtained from the USGS database USGS, 2021 at a discharge measurement station with ID 15515500 located about 200 m downstream of the location of the tripod. When the river is frozen, no discharge measurements can be made. As can be seen in figure 2.1, data is missing during the whole winter. The 'start' and 'ending' of the winter gaps in the data are however interesting to link to the freeze-up and breakup dates in further modelling of the ice breakup process. Of course, the yearly oscillation can be seen, along with the fact that temperature rise precedes discharge increase, which is in line with the conclusions from Bieniek et al., 2011 as highlighted in Section 1.2.3. A study that is noteworthy with regard to the missing data during winter is Beltaos and Kääb, 2014. Here it was shown the estimation of river discharge at ice breakup using satellite imagery is feasible.

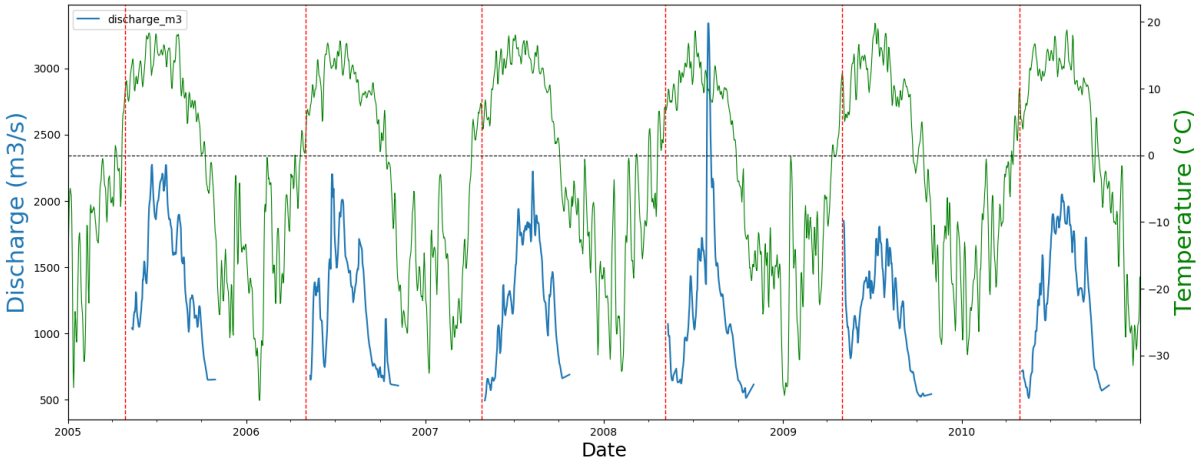


Figure 2.1: Visual analysis of temperature data (shown in green as 3 days moving average), discharge data (shown in blue as daily mean) and the breakup dates (shown in red). Seasonal repetition can clearly be seen. This plot was used to determine the metrics to work with in more detail. Data shown for 2005-2010 to increase readability. Data periods used for analysis however include for discharge 1989-2020 and for temperature 1917-2020.

2.4. Extreme temperature waves

2.4.1. Climate change

One of the observed trends associated with climate change is *"the increased frequency, intensity and duration of heat-related events, including heatwaves"* (IPCC, 2020). Especially in arctic regions like Alaska, the climate is changing rapidly due to a phenomena called *arctic amplification*. Which means that warming occurs at rates almost four times higher than the global means (You et al., 2021). This is especially evident during autumn and winter seasons in the form of higher average temperatures, but also more frequent extreme temperature events in the form of 'Artic winter warming events' (Serreze and Barry, 2011, Graham et al., 2017).

2.4.2. Definition

In literature, many different definitions of heat and cold waves exist. Most have a certain threshold normal temperature that has to be exceeded for a prescribed amount of days. An example of a definition would be from AghaKouchak et al., 2020: *"(...) a series of consecutive hot days that has significant impacts on human health, agricultural yield, water resources, energy consumption and the environment."* In order to quantify this definition specifically for this research, the daily temperature data for Nenana will be used (BerkeleyEarth, 2021). A heat wave is composed of three or more consecutive 'heat wave days'. The threshold for the mean daily temperature for which a day is considered a 'heat wave day' is determined using the 95th percentile of the temperature data for each month in the base period 1917-1947. An example is as follows; for the base period, the 95th percentile of the mean daily temperature on March 7th equals -2.1 °C. In 1963, the temperature on March 7th was equal to -0.3 °C, so this day is considered a 'heat wave day'. When the following two or more days are also above their respective thresholds, a heat wave event is noted.

2.4.3. Data analysis

Using the aforementioned temperature thresholds per month, the amount of heat waves in the months January until May is counted for each year. Extreme temperature effects during these months are assumed to have most effect on the breakup date. The results are shown in Figure 2.2. A trend towards the more frequent occurrence of heat waves can be seen in recent decades, when a linear regression is performed.

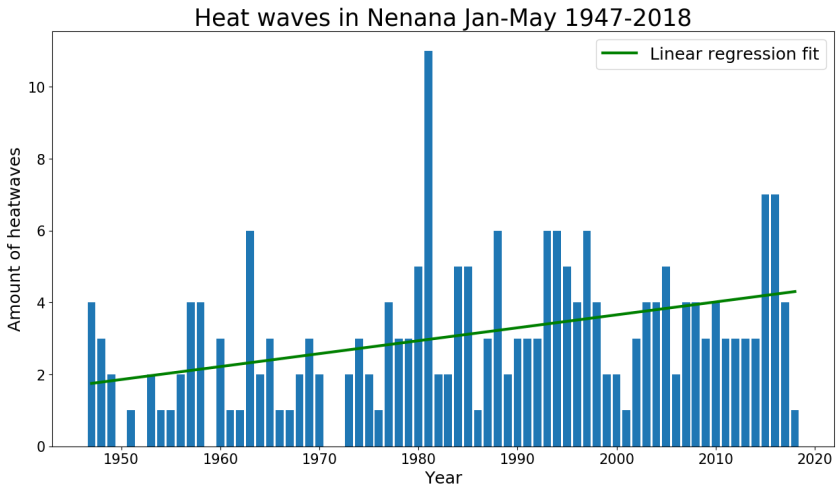


Figure 2.2: This figure shows the amount of heat waves in the period January - May for each year from 1947-2018. A heat wave is defined as three consecutive days in which the mean daily temperature is above the threshold based on the 95th percentile of the mean daily temperatures in the base period 1917-1947. The linear regression line shows an upward trend.

2.5. Variables

Based on previous work by Van Asselt, 2020, Bieniek et al., 2011, Beltaos, 2003, among others, a set of climate and environment related variables is chosen. These are shown in Table 2.1 below , along with the symbol that will be used further on in the report, the units, the period of observation and the source of the data. The Accumulated Degree-Days Thaw (ADDT) and Accumulated Degree-Days Frost (ADDF) are based on the reference point of -5°C. ADDF includes temperature data ranging from December up to March, while ADDT ranges from February up to April. Instead of using the amount of heat waves per month, the amount of heat wave days per month (HWd) is used as a variable, which means the data contains more information instead of just the amount of heat waves (three consecutive hot days would mean one 'heat wave', but nine consecutive days too). The ice thickness measurements are available yearly, for the following moments: February 20th, March 1st, March 15th, April 1st. Data sources are BerkeleyEarth, 2021, USGS, 2021, NOAA, 2020, CRU, 2021.

Ice breakup aspect	Variable	Symbol	Unit	Period	Source
Thermal ice deformation	Monthly average temperature	T_{month}	$°C$	1917-2019	BE
	Accumulated degree-days thaw (base - 5°C)	ADDT	$°C$	1917-2019	BE
	Heat wave days per month	HWd_{month}	d	1947-2019	BE
Solar radiation	Cloud coverage	CC_{month}	%	1917-2019	BE
	Average winter precipitation	P_{DJFM}	mm	1917-2019	CRU
River discharge	Monthly average discharge	Q_{month}	m^3/s	1963-2019	USGS
Ice thickness	Average winter temperature	T_{DJFM}	$°C$	1917-2019	BE
	Accumulated degree days frost (base -5°C)	ADDF	$°C$	1917-2019	BE
	Ice thickness	t	m	1989-2016	USGS
Large scale SST	ENSO effect in February - May	$ENSO_{FMA}$	$°C$	1951-2019	NOAA
	PDO effect in February - May	PDO_{FMA}	$°C$	1917-2019	NOAA

Table 2.1: Per ice breakup process aspect, the variables that will be used in the RF model are shown.

3

Random Forest model

This chapter introduces the concept of Machine Learning models and discusses previous application to the river ice breakup problem. Furthermore, the specific Random Forest Regression algorithm that will be used, is introduced and explained in detail.

3.1. Machine learning

3.1.1. Introduction to machine learning

Machine learning is a concept so abstract to most people, it often is described as 'magic', even by scholars and businesses for whom it might be most interesting. In fact, machine learning (ML) is in principle based on simple statistics and has been around for a while. ML can best be described as *"the science (and art) of programming computers so they can learn from data."* (Géron, 2017). An example of a ML application that first became mainstream in the 1990s, is the e-mail spam filter. The spam filter example is a ML program that can learn to distinguish spam emails from normal emails, using the emails the user flags as spam to learn to classify other spam emails (Géron, 2017).

The ML method that is most interesting for this thesis is regression. ML regression algorithms are used to represent the inherent relationship between input and output variables of a complex nonlinear problem of which the nature of the relationship is unknown a priori (Zhao et al., 2012). The machine learning regression algorithm 'learns' from the data and formulates the relationship between the input and output variables. These learning methods are dependent on the type of ML algorithm used.

3.1.2. Previous Research

As mentioned in Section 2.2, various researchers have described and modelled the ice breakup process, including Zhao et al., 2012, who used an Artificial Neural Network (ANN) system to describe and forecast the breakup of the Hay River ice in northern Canada. ANN is a ML regression method with a parallel and layered structure consisting of artificial neurons (nodes) that process the input signals to produce an output signal. In Zhao et al., 2012, the water level rise associated with the onset of ice breakup is the output of the ANN and it is used as the prediction of the breakup date.

Zhao et al., 2012 shows that the ANN model performed significantly better in describing the ice breakup process when compared to Multiple Linear Regression models using the same input variables. A figure representing the ANN model structure can be seen in Appendix A. The downside of ANN models is that due to the complexity of the large amount of (hidden) nodes in the network (see Figure A.1), the relationship between input and output is difficult to interpret. A ML regression model that does allow for more insight in this relationship, is Random Forest. This algorithm will be applied to the problem in this thesis and will be elaborated on in the next section.

3.2. Random Forest Regression

3.2.1. Principles

Random Forest is an ensemble machine learning technique introduced by Breiman, 2001. The Random Forest Regression (RFR) algorithm is based on the law of large numbers. The algorithm uses a large amount of

'tree predictors'. A tree predictor is a decision trees that gives an estimate for the output variable of the model. An example of a tree predictor can be seen in Figure 4.1. Taking the mean of all the predictions gives a better result than if a single model for prediction was set up. The name for the use of multiple predictors with different inputs in parallel is called *ensemble methods*, and is widely used in ML problems (Dietterich, 2000).

Furthermore, the Random Forest algorithm is based on another principle. The principle of bootstrap aggregating, i.e., bagging. The bagging principle generates multiple versions of a predictor. The different versions are formed by making 'bootstrap' samples (picked at random from the total dataset) of the training set, and using these to train different predictors, the 'trees'. These different predictors then form an ensemble, the 'forest', and the individual prediction results are averaged to increase the output accuracy (Breiman, 1996).

3.2.2. Training and Testing

In RF regression, the 'tree' predictors are functions relating the input variables to the output variable (Breiman, 2001). This relationship is formed by 'training' the RF model. This is done by splitting the input variables in two sub samples, the 'training set' and 'testing set'. An example would be to use 75% for training the model and the remaining 25% to test or validate the model (metrics used to validate the model are discussed in Chapter 4). The model is trained by learning from the input and corresponding output data. For the test set, only the input data is given to the model. Then the model makes predictions for the output data of the test set. The accuracy of these predictions (in comparison to the true values) says something about the quality or robustness of the model. A graphical representation of the testing procedure can be seen in Figure 3.1.

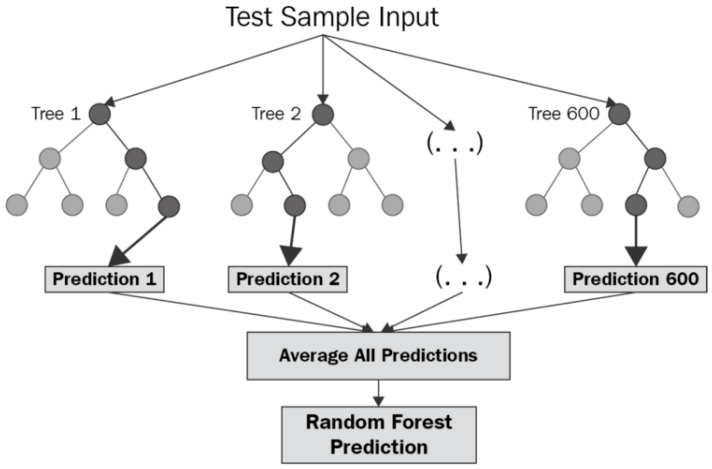


Figure 3.1: This figure shows a visual representation of the testing procedure of a RF model. The predictors, of which there are 600 in this case, have already been formed by training the model with the training data. The test procedure consists of handing the testing data to the model, without the corresponding output values. Now the model makes predictions for these output values, and the model accuracy is defined as the difference between these predictions and observed values for the output. (Chakure, 2019)

3.3. Python Package and Hyperparameters

The Python machine learning package *scikit-learn* (Pedregosa et al., 2011) will be used in implementing and validating the RF model. A RF regression module is readily available, where several *hyperparameters*, parameters that influence the module's workings, can be tweaked. An example is the amount of tree predictors or 'trees in the forest'. The larger the forest the better, but this will take more computing time. In addition, the results will not get significantly better after a certain number of predictors (Pedregosa et al., 2011). More hyperparameters are adjustable, but elaboration on these is emitted here. These hyperparameters will be left constant.

4

Methodology

In this chapter the RF regression model is applied on the specific river ice breakup prediction problem. The model structure, error metrics and validation procedures are discussed.

4.1. Model Setup

The application of the RF algorithm is based around the assumption of a relationship between the climate and environmental variables and the output variable: the river ice breakup date. This assumption is based on the physics of river ice breakup, also investigated in research by Bieniek et al., 2011, Beltaos, 1997, Beltaos, 2007, Zhao et al., 2012, Van Asselt, 2020 and Hu et al., 2008, amongst others, as outlined in Chapters 1 and 2. This relationship is assumed to be of the following form:

$$y = f(\mathbf{X}) \tag{4.1}$$

y being the output variable (date of ice breakup).
f being some arbitrary function describing the (possible) relationship.
X being a matrix containing the input variables (per year), e.g., the mean temperature in April for the year 1999.

To formulate the relationship function f, the RF algorithm will be used. Unlike linear regression, the algorithm will not be able to explicitly express this relationship in a mathematical formula like is the case with linear regression.

A general method for setting up a RF model for regression as presented by Hastie et al., 2009 is outlined below. Here, a total of B tree estimators (T_b with minimum node depth n_{min}) is created, using a training set or 'bootstrap sample' (X). The training set consists of numerical values for the total of m input variables, but to individually train each tree predictor only a random amount of $p < m$ variables is used.

1. For $b = 1$ up to B tree predictors:
   (a) Draw a bootstrap sample (X) of size N from training data (75% of the total, in this example).
   (b) Grow a random-forest tree T_b to the bootstrapped data by recursively repeating the following steps for each terminal node of the tree, until the minimum node depth n_{min} is reached.
      i. Select p variables at random from the total of m variables.
      ii. Pick the best variable/split-point among the m.
      iii. Split the node into two daughter nodes.
2. Finally: output the ensemble of tree predictors $\{T_b\}_1^B$. To make a prediction at a new point x, average the results of the tree predictors:

$$f_{rf}^B = \frac{1}{B} \sum_{B}^{b=1} T_b(x) \tag{4.2}$$

The above process will not be programmed by hand, but is already implemented in the scikit-learn Python package discussed earlier. One decision tree is constructed with a root node and multiple middle leaf nodes, which together form a function which can take on different output values. An example of such a tree can be seen in Figure 4.1.

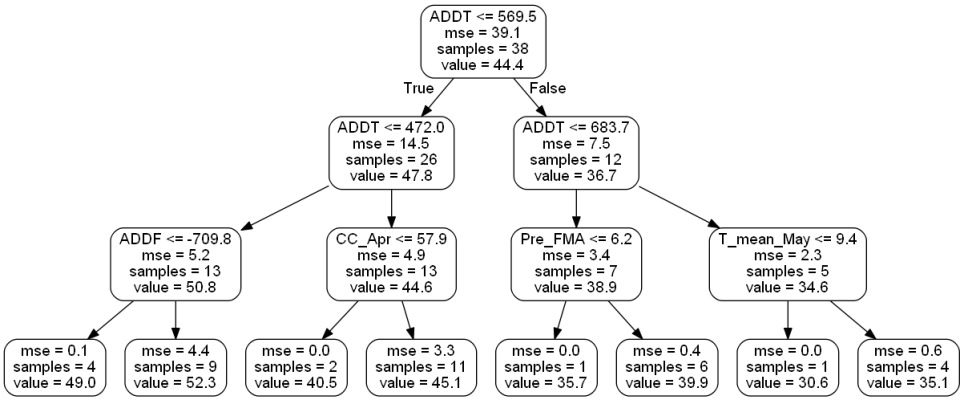


Figure 4.1: Example of a single tree predictor of a RF regression model. It can be interpreted as follows: using for example data point 2018, where ADDT = 486.3 and CC_Apr = 58.2 (other values are not important for the example). The first node is True (move to the left), since ADDT is smaller than 569.5. The second node is False, since ADDT is larger than 472.0. The third node is also False (move to the right), since CC_Apr is larger than 57.9, meaning the predictor arrives at the 'leaf' that concludes the estimation for the breakup date is equal to 45.1 days from equinox. This is one of the many predictions that are averaged to formulate the final output of the RF model. This three predictor has depth 3, since three steps are taken before the result is obtained. The actual model used has trees with a larger and variable depth, meaning more variables are considered in the estimation procedure.

4.2. Model Evaluation

4.2.1. Over-fitting

First, the question arises whether or not such a complex model tends to over-fit the data, overestimating the accuracy, because RF models tend to have an amount of tree predictors in the order of magnitude range of approximately $10^2 - 10^3$. Generally speaking, RF models rarely over-fit due to the random training data selection and the averaging of a large amount of predictions (Breiman, 2001). RF models do however perform poorly when the fraction of relevant variables is small in relation to the total (Hastie et al., 2009). This is something to keep in mind when evaluating the model performance and choosing input variables.

4.2.2. Variable Importance

The *variable importance evaluation* module of the RF regression package will be used to quantify the relative importance of the input variables used by the RF model. This value indicates that a permutation of a relatively influential variable will have a large consequence for the model prediction. The variable importance can also be interpreted as how often a decision is made using that variable (as shown in Figure 4.1). It will also be possible to identify variables of insignificant importance, only causing noise in the model prediction. Subsequently, it is possible to reevaluate the selection of variables, in order to comply with the requirement of multiple relevant variables stated earlier in this paragraph.

4.2.3. Metrics

To quantify the model performance, several metrics will be used, outlined below. The model generates predictions (fitted values) f_i and is tested with n samples of observed values y_i.

- Root Mean Squared Error (RMSE), a common metric for evaluating numerical predictions. RMSE is a measure of accuracy when comparing prediction errors of different model configurations for a particular (set of) variable(s) and not between variables, as it is scale dependent (Neill and Hashemi, 2018). The RMSE will be used to evaluate model performance in the process of tweaking the hyperparameters and as an error metric for validating the finalized model. From Glantz and Slinker, 2000:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \tag{4.3}$$

- Mean Absolute Error (MAE). The MAE is an error metric that is easily to interpret as it is just the mean value of the (absolute) errors. This also means it is in the unit of the model output variable; days. The MAE is useful for the comparison of different model configurations with different input variables. The MAE will be used to interpret the final model performance when an optimal model is chosen in terms of what input variables are used and how the RF model hyperparameters are tweaked. From Hastie et al., 2009:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |f_i - y_i|^2 \tag{4.4}$$

4.2.4. Cross Validation

In order to obtain the metrics above to asses the model performance, the *K-fold cross validation* method will be applied. Ideally we would have sufficient data split the data into a training dataset and a set for the validation of the prediction model performance (the example of 75% for training and 25% for testing was used in Section 4.1). This is however regularly not possible, since data are often scarce, as is the case here. K-fold cross validation splits the data into $k = 1, 2, ..., K$ roughly equal-sized parts or *folds*. Then the validation procedure is as follows; for example for $k = 1$, the model is trained using the other $K - 1$ parts of the data, and the prediction error is calculated (the testing of the model) with the $k = 1$ part. This is done cycling through $k = 1, 2, ..., K$, and the K estimates of the error are averaged. Typical choices of K are 5 or 10 (Hastie et al., 2009). Here, 5-fold will be used. 10-fold cross validation would infer a high variance for the estimation of RMSE and MAE, since the sample size for a single fold for a model with small sample size (the smallest is $N = 28$) would be too small (Hastie et al., 2009). In addition, the K-fold validation will serve as a sensitivity analysis for the splitting of the data in different random subsets. When the model performs consistently across different data splits or folds, the robustness of the model is hereby verified.

5

Results and discussion

This chapter describes the findings of and the results of the Random Forest regression model discussed in Chapter 4. First, the results for a model using all explanatory variables are presented. Upon further selection of different model configurations, the results of these different models are presented. Finally, a forecasting model is conducted, where the Nenana Ice Classic betting deadline bounds the input data and these forecasting results are presented.

5.1. Model configuration

5.1.1. Variable Importance

First, a RF model was constructed using all variables from Table 2.1. A box plot for all (normalized) data can be seen in Appendix B. The total sample size was small, at $N = 28$, considering the variable *ice thickness* only has data ranging from 1989-2016 and the RF algorithm cannot work with missing values (Pedregosa et al., 2011). The model was evaluated using 5-fold cross validation. Despite the small sample size, it is assumed the results can be used to evaluate the relative importance of the variables, in order to select a subset of variables to be used for further configuration of the RF model. The results are shown in Figure 5.1 below. An example of a tree predictor of this model can be seen in Appendix C.

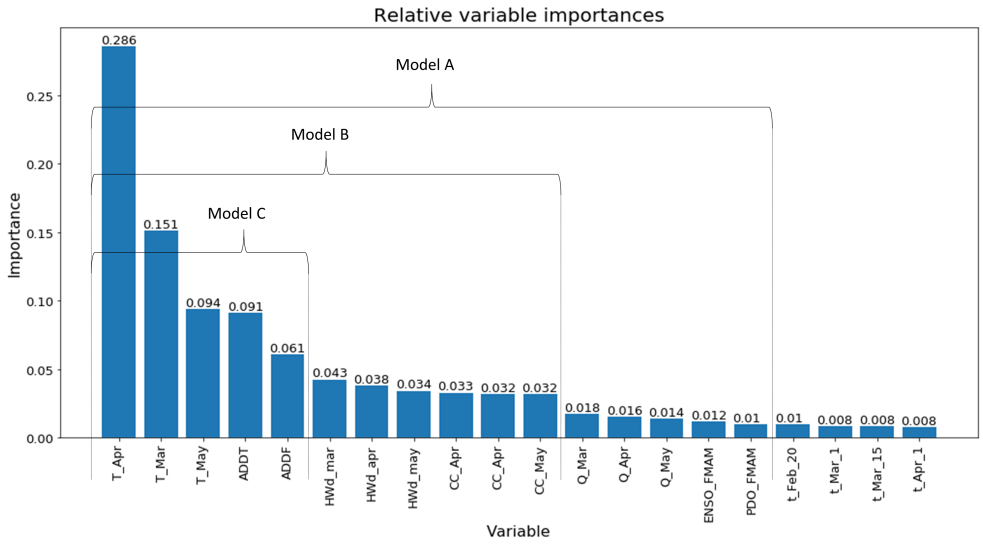
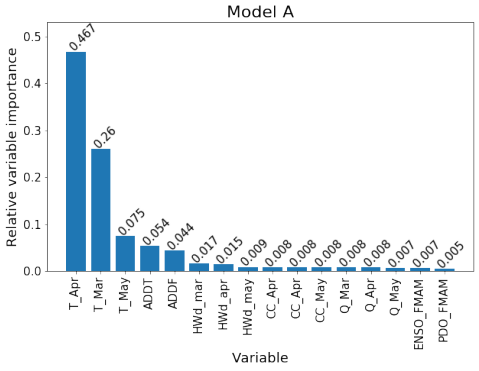
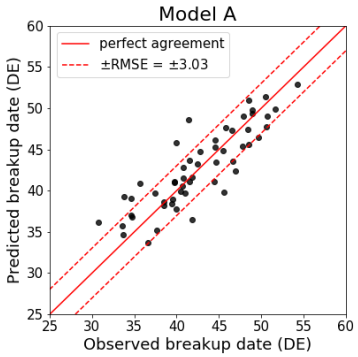


Figure 5.1: This graph shows the relative variable importances for an RF model with 100 tree predictors, based on all available input variables for the relatively short period of 1989-2016, evaluated using 5-fold cross validation. Based on this graph, the ice thickness variables are immediately left out from the model configuration, since it is assumed these only cause noise in the model output. The combinations of variables for models A, B and C are shown using the horizontal brackets. The cross validation reported an average RMSE of 4.53 and MAE of 3.82 for this model.

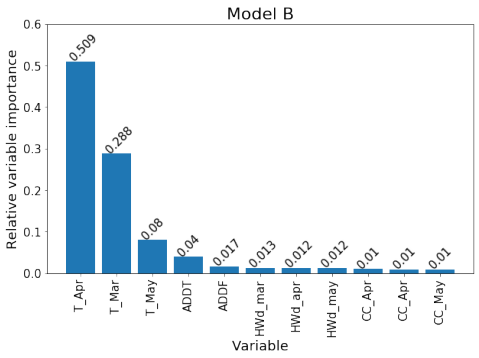
Based on Figure 5.1, it can be concluded the temperature data variables have a significant influence on the model output when compared to the other variables shown. This is in line with the results from Bieniek et al., 2011 and Van Asselt, 2020, since his linear model showed best results using the average temperature in April and May, T_{AM}, and $ADDT$. For further modelling, the ice thickness variables are left out of the configuration since it is assumed they only cause noise and significantly decrease the sample size, thereby decreasing model performance. Subsequently, the models A, B and C are chosen for further investigation and are made up of the variables as shown in the figure. Model A is based on all remaining variables and has sample size $N = 56$, since the discharge variable Q only has data ranging from 1963-2018. Model B includes the cloud coverage and heat waves variables, and ranges from 1947-2018 (the heat wave threshold is determined using 1917-1947 data and is therefore not used in this dataset). Model C has the least amount of variables with only mean monthly temperatures and Accumulated Degree Days Thaw and Frost. Model C has the largest possible sample size of $N = 102$ since the data ranges from 1917-2018.



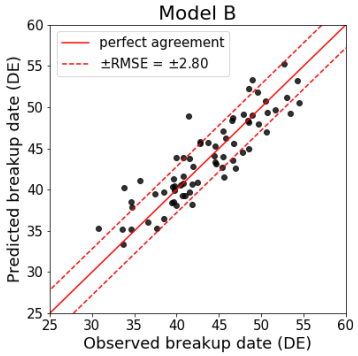
(a)



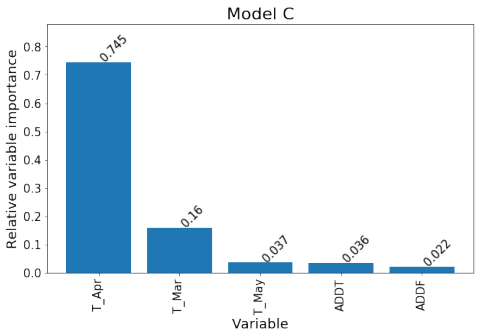
(b)



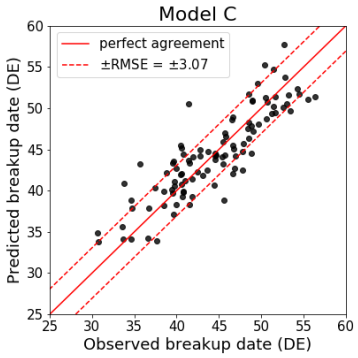
(c)



(d)



(e)



(f)

Figure 5.2: The relative variable importances ((a), (c), (e)) and observed versus predicted breakup dates ((b), (d), (f)) are shown for the three RF model configurations validated using 5-fold cross validation. Model B performed best, with a RMSE and MAE of 2.80 and 2.26, respectively. The variable importance plots show a high reliance of all models on temperature data, especially from April.

5.1.2. Three Configurations

For the three model variants, using again 5-fold cross validation, the results are shown in Figure 5.2. Model B showed the lowest RMSE and MAE values, at 2.80 and 2.26, respectively. This can probably be attributed to the fact a compromise is made between the sample size and the amount of explanatory variables, considering model B is in between A and C, based on these matters. As outlined in Section 4.2, the large amount of irrelevant variables for model A might very well decrease model performance.

All three models seem to perform better at predicting a breakup date within the 40-50 days from equinox range, and worse at predicting outliers. This hypothesis can be confirmed using the predictions and observed values from 1998-2018, shown in Figure 5.3. Next to that it can be seen all three models agree relatively well on the prediction, except for year 2001, where model C over-predicted the breakup date, while A and B made a prediction that was too early. Upon further inspection of the data for that year, it could be concluded the data for all explanatory variables was very close to the mean of the total data. This could be a reason the models went in different directions regarding the prediction. When the exhibited performance is compared to results from Zhao et al., 2012, a similar error (MAE) is found in predicting the breakup date, at 2-3 days. Overall, the relatively good performance implies great potential for RF models to generalize the complex relationship between the explanatory climate variables and the breakup date.

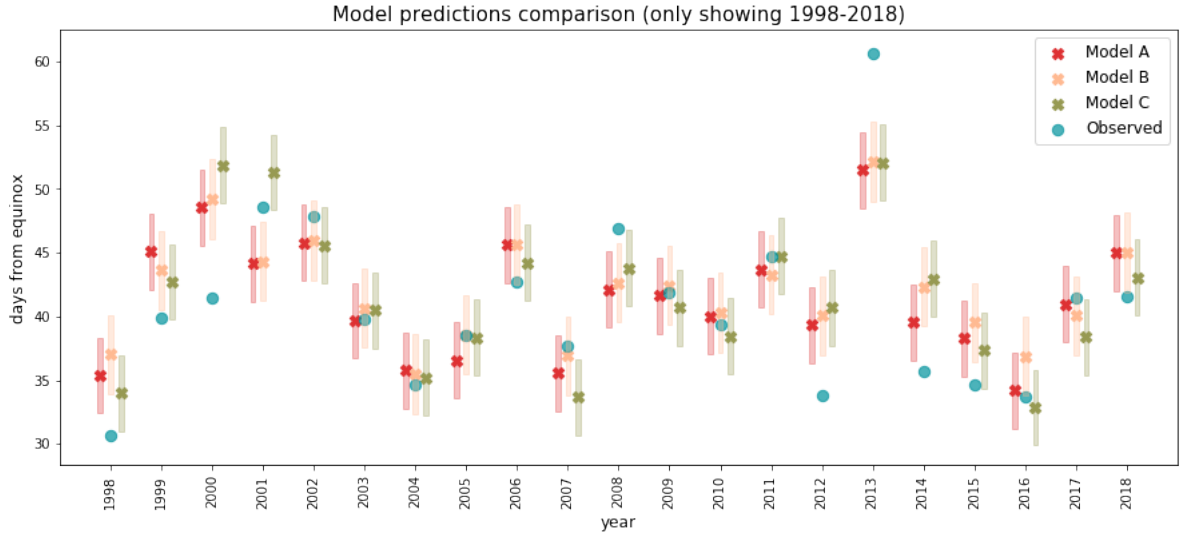


Figure 5.3: This figure shows a comparion between all three model predictions and the observed breakup date, for the years 1981-2018. The coloured bars around the prediction are the RMSE's for each model. Based on this figure it can be said all three models seem to have more difficulty predicting extreme breakup date values.

5.1.3. Hyperparameter tuning

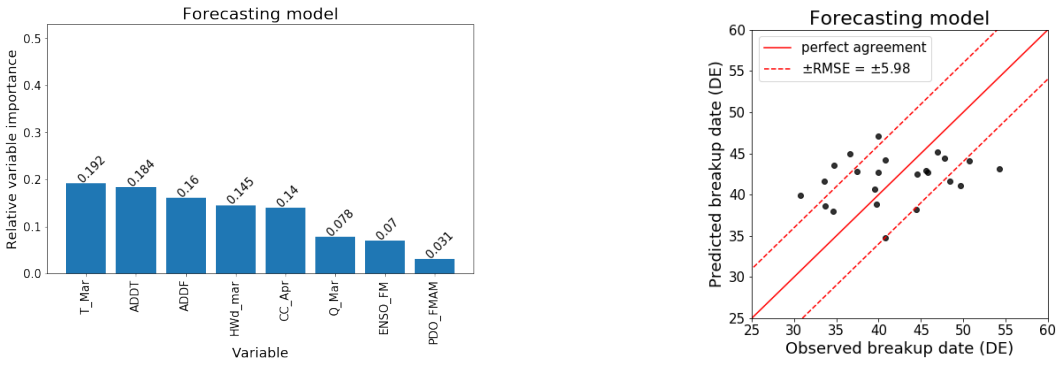
An attempt was made at tuning a single hyperparameter, as literature shows these parameters directly control the behaviors of algorithms and have significant potential effect on the performance (Wu et al., 2019). The amount of predictor trees (n_estimators in the scikit-learn RF regression package) was varied from 1 to 250, and the RMSE was calculated for all cases using model B. The result can be seen in Appendix C. The result however is not used to draw conclusions and adjust the model accordingly, since all possible hyperparameters would have to be tuned simultaneously in order to prevent undesirable changes due to the permutation of the effects of one parameter on the other, when changed.

5.2. Forecasting model

For the forecasting model, the data was bounded by the Nenana Ice Classic betting deadline of April 1st. This means either variables are completely removed from the input dataset, or are cut to only include data up to April 1st, e.g., ADDF. When compared to for example model A, the amount of explanatory variables is reduced from sixteen to eight.

Forecasting model performance, as shown in Figure 5.4b, is far worse at an RMSE of 5.98 and MAE of 5.07. The variable importances plot shows a more even spread across the explanatory variables. Besides the vast

reduction in data to train the model on, the reduce in performance can most probably be attributed to the fact RF models are not well suit for extrapolation problems (Hastie et al., 2009).



(a) (b)

Figure 5.4: (a) shows the relative variable importances for the forecasting model, bounded by the betting deadline April 1st. (b) shows the predicted vs observed breakup date. In comparison to model B, the best performing model, results are far worse. A significant amount of data close to the breakup date is left out of the training dataset, reducing model performance and robustness.

6

Conclusion

The aim of this study was to build a Random Forest (RF) regression model in order to describe the river ice breakup process and accurately predict the breakup date, using the Nenana Ice Classic dataset. Climate data related to the ice breakup was investigated and processed based on previous studies and the underlying physical and mechanical processes. An example is the increase of the frequency of heat waves in Nenana over the last decades. These variables were then used to develop a Random Forest regression model. Because the amount of data was relatively small, this model was validated using 5-fold cross validation. The descriptive RF model, not bounded by the betting deadline, showed promising results when tested. Different model configurations were made, based on the relative importance of the input variables in the different model configurations, as summarized in Figure 5.1. Model configuration B, with various temperature metrics and cloud coverage as input variables, showed best performance with a RMSE of 2.91 and MAE of 2.26. This outcome reflects the high correlation between air temperature and heat input, i.e., cloud coverage, and the breakup date. Similar relationships were found by Van Asselt, 2020. When compared to the ANN Machine Learning model as conducted by Zhao et al., 2012, similar mean error values are found, at 2-3 days. The predictions of the different RF model configurations were shown in Figure 5.3. From this it can be concluded that the model performed worse when the observed breakup date was very early or late. Besides this problem with outliers however, performance was promising. This result illustrates the potential for Machine Learning models to generalize the complex relationship between the explanatory climate variables and the river ice breakup date.

Concerning forecasting, plenty of improvements can still be made. The RF forecasting model, conducted with data bound by the Nenana Ice Classic betting deadline of April 1st, showed relatively poor performance. RMSE and MAE values were 5.98 and 5.07, respectively. This decline in performance upon bounding the data is in line with forecasting results shown by the linear model from Van Asselt, 2020. The poor performance could be attributed to the fact that a Machine Learning model not only relies heavily on input data (which was drastically reduced by the betting deadline). The specific Random Forest model type is also not well suited for extrapolation cases, as shown in literature. The ANN model from Zhao et al., 2012 performed well in forecasting, while the amount of data was very small. For further research on this topic, it might be interesting to apply a similar model to the Nenana Ice Classic data.

In conclusion, this research has provided a new perspective on the prediction of ice breakup dates and the effect of environmental variables on this very complex process. Considering the Random Forest regression model, significant improvements can still be made in gathering more input data for the explanatory variables, since this will have the largest effect on performance. Furthermore, a hyperparameter sensitivity analysis could be performed, in order to tune the model in a most optimal way. More generally, it would be interesting to explore the relationship between the betting behaviour of Nenana Ice Classic participants and their perception of the climate, also proposed by Van Asselt. Specifically, the perception of climate change in the Arctic region, e.g., the more frequent occurrence of heat waves as shown in this research, might very well influence this betting behaviour. Inversely, the trend towards earlier breakup dates of the Nenana Ice Classic, as visible in the data, could be used to raise more awareness for the devastating effects of human induced climate change on Arctic ecosystems.

A

Zhao et al. ANN model

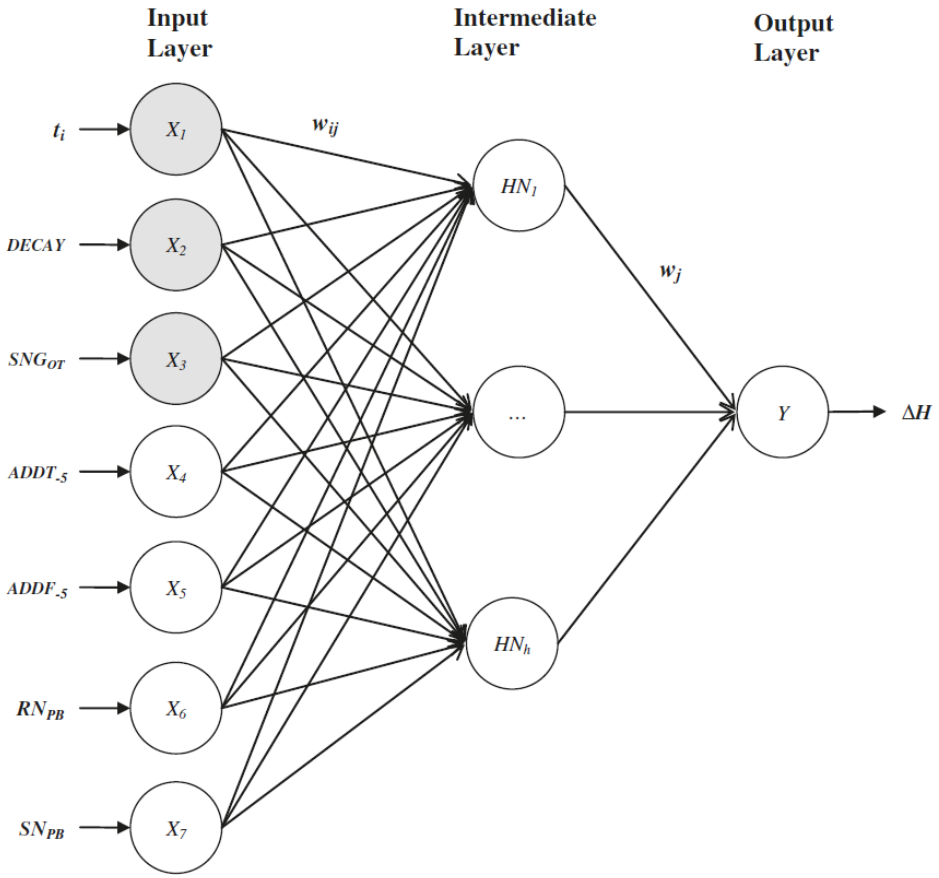


Figure A.1: A graphical representation of the Artificial Neural Network model as used by Zhao et al., 2012 in describing the ice breakup process. The model has a 'three-layer feed-forward' structure. The first layer, the input layer, is comprised of all the input variables, for example the Accumulated Degree Days Thaw (ADDT) also used in this model. The last output layer is the output variable. The intermediate (or hidden) layer is composed of a specified number of nodes, which process the input layer data. The nodes of adjacent layers are connected by weighing factors (w_{ij}). This is an example of a 'black box' machine learning model, because the nodes of the intermediate layer are hidden to the user.

B

Data boxplots

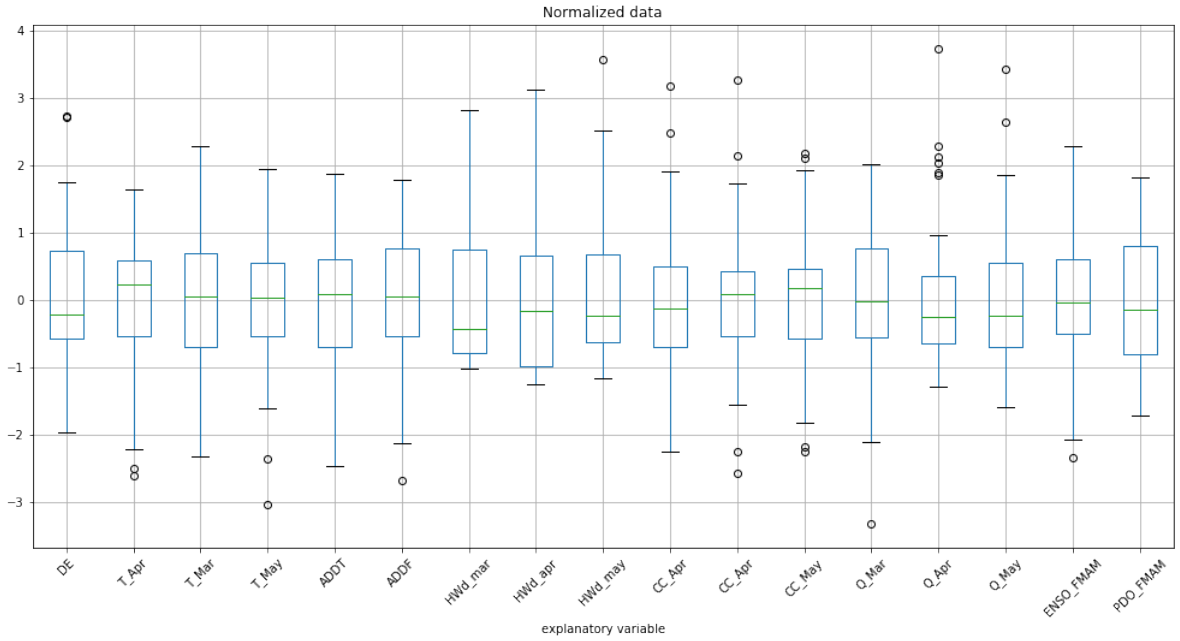


Figure B.1: The explanatory variables used in the RF models, normalized to be shown in this boxplot. A significant amount of high outliers can be identified in the discharge data for April, but there are no real other significant anomalies.

C

Tree predictor

The figure is shown on the next page.

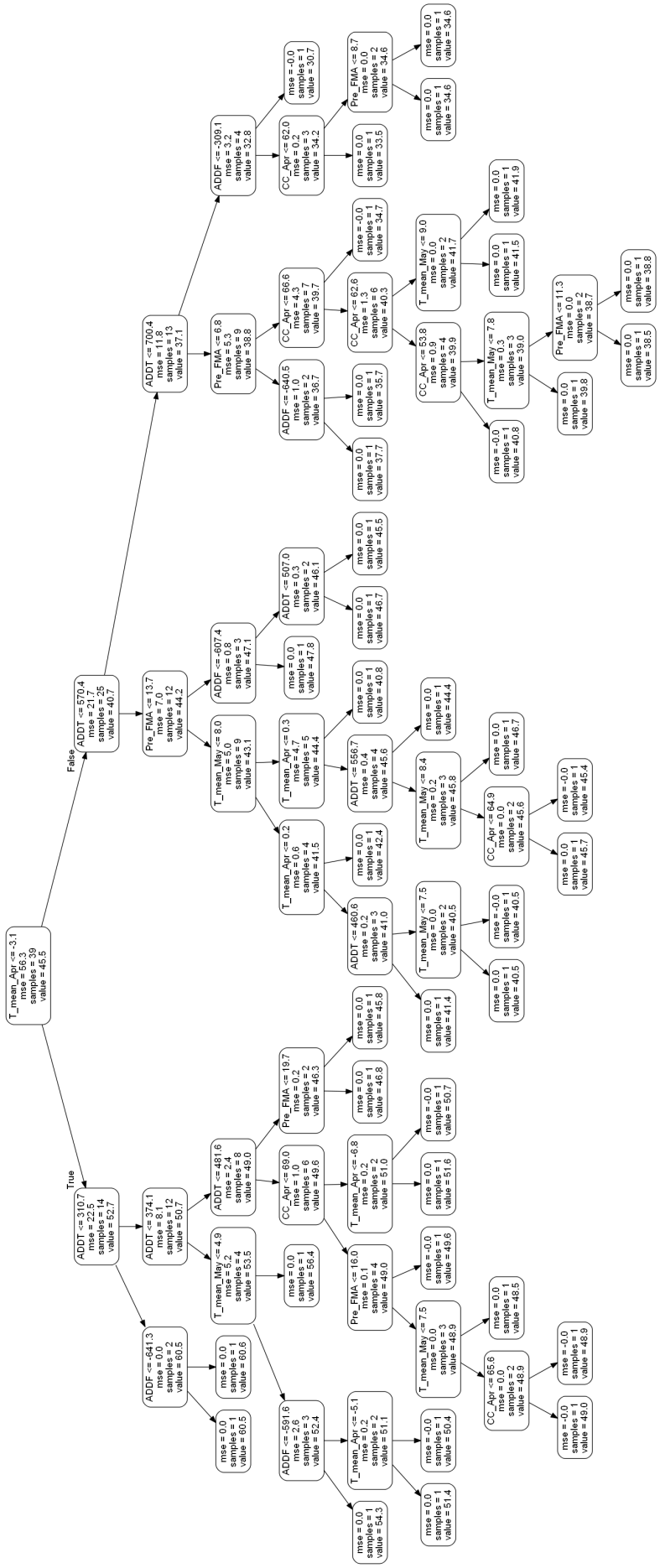


Figure C.1: This figure shows an example of a single tree predictor in the RF model. Since the different 'leafs' might be too small to read, this figure is added to illustrate the following: The variable 'depth' of the predictor (the steps taken to the output result) can clearly be seen, along with the overall complexity of such a predictor. For an explanation of the workings, the reader is referred to Figure 4.1.

D

Hyperparameter tuning

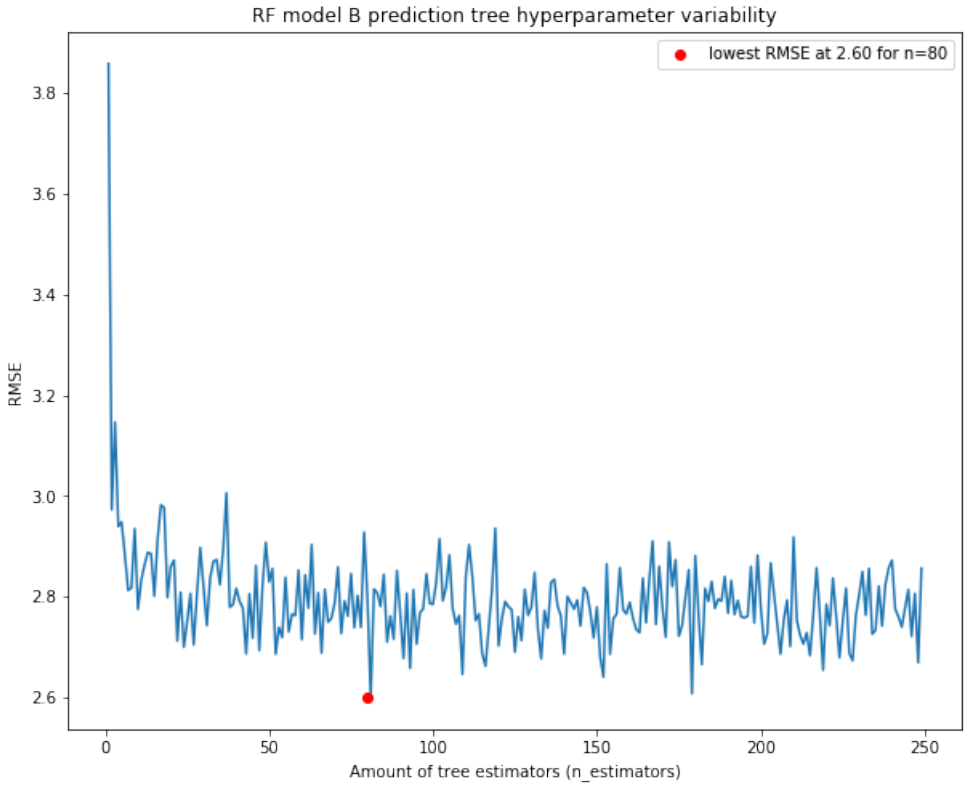


Figure D.1: This figure shows the RMSE of model B for different amount of tree estimators, ranging from 1 to 250. A steep decline in RMSE can be seen for the first 30, while further on no significant trends are visible. The lowest RMSE is marked with a red dot, at 2.6 for n=80.

Bibliography

AghaKouchak, A., Chiang, F., Huning, L., Love, C., Mallakpour, I., Mazdiyasni, O., Moftakhari Rostamkhani, H., Papalexiou, S. M., Ragno, E., & Sadegh, M. (2020). Climate extremes and compound hazards in a warming world. *Annual Review of Earth and Planetary Sciences*, *48*. https://doi.org/10.1146/annurev-earth-071719-055228

Andersen, Pattison, R., Gray, A., Schulz, B., smith, & Jovan, S. (2018). *Forests of the tanana valley state forest and tetlin national wildlife refuge, alaska: Results of the 2014 pilot inventory*. https://doi.org/https://doi.org/10.2737/PNW-GTR-967

Beltaos, S. (1997). Onset of river ice breakup. *Cold Regions Science and Technology*, *25*(3), 183–196. https://doi.org/10.1016/S0165-232X(96)00011-0

Beltaos, S. (2003). Threshold between mechanical and thermal breakup of river ice cover. *Cold Regions Science and Technology*, *37*, 1–13. https://doi.org/10.1016/S0165-232X(03)00010-7

Beltaos, S. (2007). River ice breakup processes: Recent advances and future directions. *Canadian Journal of Civil Engineering*, *34*(6), 703–716. https://doi.org/10.1139/l06-021

Beltaos, S., & Kääb, A. (2014). Estimating river discharge during ice breakup from near-simultaneous satellite imagery. *Cold Regions Science and Technology*, *98*, 35–46. https://doi.org/10.1016/j.coldregions.2013.10.010

Bengel, A. (2021). Nenana ice classic sees steady sales after ticket slump in 2020. https://www.webcenterfairbanks.com/2021/03/11/nenana-ice-classic-sees-steady-sales-after-ticket-slump-in-2020/

BerkeleyEarth. (2021). *Berkeley temperature data*. http://berkeleyearth.org/data/

Bieniek, P. A., Bhatt, U. S., Rundquist, L. A., Lindsey, S. D., Zhang, X., & Thoman, R. L. (2011). Large-scale climate controls of interior Alaska river ice breakup. *Journal of Climate*, *24*(1), 286–297. https://doi.org/10.1175/2010JCLI3809.1

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Chakure, A. (2019). *Random forest structure*. https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f

City of nenana | homepage. (2017). https://www.cityofnenana.com/

Collins, C. M. (1990). *Morphometric Analyses of Recent Channel Changes on the Tanana River in the Vicinity of Fairbanks, Alaska* (tech. rep.). U.S. Army Corps of Engineers. https://apps.dtic.mil/dtic/tr/fulltext/u2/a229511.pdf

CRU. (2021). *Cru data*. http://www.cru.uea.ac.uk/data/

Dietterich, T. G. (2000). Ensemble methods in machine learning. *Proceedings of the First International Workshop on Multiple Classifier Systems*, 1–15.

Géron, A. (2017). *Hands-on machine learning with scikit-learn and tensorflow : Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Glantz, S., & Slinker, B. (2000). *Primer of applied regression & analysis of variance*. McGraw-Hill Education. https://books.google.nl/books?id=fzV2QgAACAAJ

Graham, R. M., Cohen, L., Petty, A. A., Boisvert, L. N., Rinke, A., Hudson, S. R., Nicolaus, M., & Granskog, M. A. (2017). Increasing frequency and duration of arctic winter warming events. *Geophysical Research Letters*, *44*(13), 6974–6983. https://doi.org/https://doi.org/10.1002/2017GL073395

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning, second edition*. Springer New York Inc. https://doi.org/10.1007/b94608_15

Hu, J., Liu, L., Huang, Z., You, Y., & Rao, S. (2008). Ice breakup date forecast with hybrid artificial neural networks. *Proceedings - 4th International Conference on Natural Computation, ICNC 2008*, *2*, 414–418. https://doi.org/10.1109/ICNC.2008.169

IPCC. (2020). *An IPCC Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems Climate Change and Land Summary for Policymakers WG I WG II WG III* (tech. rep.).

Jeffries, M., Morris, K., & Duguay, C. (2012). *Floating ice: Lake ice and river ice*.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rube, F. (2006). Köppen-Geiger climate classification. http://koeppen-geiger.vu-wien.ac.at/present.htm

Kovachis, N., Burrell, B. C., Huokuna, M., Beltaos, S., Turcotte, B., & Jasek, M. (2017). Ice-jam flood delineation: Challenges and research needs. *Canadian Water Resources Journal*, *42*(3), 258–268. https://doi.org/10.1080/07011784.2017.1294998

Meier, W., & Dewes, C. (2020). Nenana ice classic: Tanana river ice annual breakup dates, version 2. https://doi.org/https://doi.org/10.5067/CAQ58H42LQY2 (accessed: 13.05.2021)

Neill, S. P., & Hashemi, M. R. (2018). Chapter 8 - ocean modelling for resource characterization. In S. P. Neill & M. R. Hashemi (Eds.), *Fundamentals of ocean renewable energy* (pp. 193–235). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-810448-4.00008-2

Nenana Ice Classic. (2020). Nenana Ice Classic. https://www.nenanaakiceclassic.com/

NOAA, N. W. S. (2020). Flooding in Alaska.

Papineau, J. (2016). Understanding Alaska's Climate Variation, 1–14.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Prowse, T., Bonsal, B. R., Duguay, C. R., & Lacroix, M. P. (2007). River-ice break-up/freeze-up: A review of climatic drivers, historical trends and future predictions. *Annals of Glaciology*, *46*, 443–451. https://doi.org/10.3189/172756407782871431

Serreze, M. C., & Barry, R. G. (2011). Processes and impacts of arctic amplification: A research synthesis. *Global and Planetary Change*, *77*, 85–96. https://doi.org/10.1016/j.gloplacha.2011.03.004

USGS. (2021). *Tanana river at nenana: Discharge data*. https://waterdata.usgs.gov/monitoring-location/15515500/#parameterCode=00060&period=P365D&compare=true

Van Asselt, P. A. K. (2020). *Nenana ice classic: Gambling with river ice* (BSc thesis). TU Delft.

Wang, J., He, L., Chen, P. P., & Sui, J. (2013). Numerical simulation of mechanical breakup of river ice-cover. *Journal of Hydrodynamics*, *25*, 415–421. https://doi.org/10.1016/S1001-6058(11)60380-7

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on bayesian optimizationb. *Journal of Electronic Science and Technology*, *17*(1), 26–40. https://doi.org/https://doi.org/10.11989/JEST.1674-862X.80904120

Yarie, J., Viereck, L., Van Cleve, K., & Adams, P. (1998). *Flooding and Ecosystem Dynamics along the* (tech. rep. No. 9).

You, Q., Cai, Z., Pepin, N., Chen, D., Ahrens, B., Jiang, Z., Wu, F., Kang, S., Zhang, R., Wu, T., Wang, P., Li, M., Zuo, Z., Gao, Y., Zhai, P., & Zhang, Y. (2021). Warming amplification over the arctic pole and third pole: Trends, mechanisms and consequences. *Earth-Science Reviews*, *217*, 103625. https://doi.org/https://doi.org/10.1016/j.earscirev.2021.103625

Zhao, L., Hicks, F. E., & Fayek, A. R. (2012). Applicability of multilayer feed-forward neural networks to model the onset of river breakup. *Cold Regions Science and Technology*, *70*. https://doi.org/10.1016/j.coldregions.2011.08.011