PREDICTING RIVER ICE BREAKUP EVENTS USING MACHINE LEARNING AND HYBRID MODELLING

PREDICTING THE OCCURRENCE OF RIVER ICE BREAKUP EVENTS IN CANADA USING MACHINE LEARNING AND HYBRID MODELLING

By MICHAEL DE COSTE M.SC., B.SC.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2022) Hamilton, Ontario (Civil Engineering)

TITLE: Predicting the Occurrence of River Ice Breakup Events in Canada using Machine Learning and Hybrid Modelling AUTHOR: Michael De Coste, M.Sc. (University of Alberta), B.Sc. (University of Alberta) SUPERVISOR: Professor Z. Li NUMBER OF PAGES: xxii, 203

Lay Abstract

River ice breakup is a key event to the hydrology of rivers throughout Canada, playing a major role in their physical and ecological characteristics. The timing and mechanism of these events can, however, be unpredictable and volatile, with the effects of climate change only exacerbating these risks. This dissertation focuses on addressing these potential issues through the application of machine learning and hybrid modeling in the prediction of river ice breakup events. Advanced data driven techniques coupled with novel applications of other analytical methods are used to: i) predict the presence of ice jams through the application of stacking ensemble modelling; ii) predict the severity of mid-winter breakups through application of trend and variable analysis; iii) predict the occurrence and timing of mid-winter breakups using rare-event forecasting techniques; and iv) develop a novel hybrid modelling scheme coupling ontology-based semantic modelling and machine learning to predict spring breakup timing. Detailed case studies for each application are provided demonstrating the effectiveness of the discussed techniques.

Abstract

River ice breakup is a vital process to the morphology and hydrology of many rivers in Canada, often governing peak flows of the river. These events can occur through multiple mechanisms, with the potential for volatile or early breakup events that can have severe impacts to the river. Ice jam flooding can be a potentially devastating result of river ice breakup while early breakup of ice cover in a mid-winter breakup can be unpredictable and greatly alter the remaining ice season. These events are growing increasingly common as a result of climate change, and as a result there is a need to develop prediction tools for these events to aid in decision making support. Past investigations into developing such tools, especially from a data-driven modelling perspective, are challenged by the availability and complexity of the data related to these rare and dangerous to measure events. Therefore, the goal of this dissertation was to develop and apply methods to address the historical challenges and shortcomings in predicting these events through the use of data-driven modelling techniques. This includes: i) development of a stacking ensemble modelling framework for the prediction of ice jam presence during the spring breakup season of a river, utilising variable selection and rare-event forecasting techniques in combination with a comprehensive selection of machine-learning algorithms; ii) return period and trend analysis of mid-winter breakups in conjunction with comprehensive input analysis techniques to identify the key drivers of these events' severity and develop a means of classifying the flood risk based on hydroclimatic traits; iii) the development of a two-level modelling system for the prediction of the occurrence and timing of mid-winter breakups on a national scale utilising rare event forecasting techniques and imbalanced learning; and iv) development of a novel hybrid semantic and machine learning modelling system in which an ontology is used in conjunction with network analysis techniques to select variables for machine learning models, which is used on a national case study

of the prediction of spring breakup timing in Canada. The results of each study in application to their respective case studies demonstrate the effectiveness of the proposed techniques, which are shown to be easily adaptable to other regions or locations. These techniques can form the backbone of decision-making support for communities on rivers that are affected by the unpredictable and oftentimes volatile nature of river ice breakup.

Acknowledgements

The author would like to extend their sincere thanks to their supervisor, Dr. Zhong Li, for her support and guidance throughout the work presented in this thesis. Her expertise and advice were instrumental in the completion of this thesis. This research could also not have been completed without the guidance and support of the co-authors of these studies, Dr. Wei Sun, Dr. Yonas Dibike, and Dr. Ridha Khedri, who provided data, guidance, and suggestions. I would also extend my thanks to my supervisory committee, Dr. Paulin Coulibaly and Dr. Moataz Mohammed for their professional advice and comments that enhanced the work presented here. This thanks is also extended to the members of the Department of Civil Engineering at McMaster including Dr. Sarah Dickson, Dr. Robin Zhao, Dr. Peijun Guo, Dr. Lydell Wiebe, Sarah Sullivan, Amelia Brook, and Monica Han for their work in facilitating my studies and research. This thanks extends to NSERC and the MacData Foundation who provided vital financial support to this research. I would also thank the member of my research group, Dr. Qianqian Zhang, Dr. Maysara Ghaith, Dr. Wendy Huang, Xinyi Li, Pengxiao Zhou, Zehao Yan, David Kovacs, Fuwei Rao, and Jinyi Fang. I would finally like to thank Katarina Green for her support during this challenging and fulfilling process, and my parents, Frank and Shari De Coste, and brother, Derek De Coste, for their encouragement and guidance from the start to the finish of this journey.

Table of Contents

List of Figures

List of Tables

List of Symbols

$f(x)$	Predicted response of an algorithm
u	Logistic regression linear function
α	Logistic regression intercept
β_i	Logistic regression coefficients
w_j	Weight of an individual observation in nearest neighbours algorithms
d_j	Distance between neighbour points
y_k	Response variable in a support vector machine
ψ	Function for assigning constant values in support variable machine
M	Number of bootstrap samples
S	Time series Mann-Kendall trend statistic
z	Test statistic for Mann-Kendall test
R	Return period of water levels
$f(x, c)$	Probability density function of Weibull distribution with shape parameter *c*
$\rho_{X,Y}$	Pearson correlation between x and y
$cov(X, Y)$	Covariance between x and y
$\hat{\beta}$	Weight of a given variable in LASSO regression
λ	LASSO tuning parameter
ℓ	L1 regularization penalty for LASSO
a	Linear regression intercept
b_i	Linear regression coefficients
$f_T(x)$	AdaBoost final model prediction
c_t	Weight of each learner

$h_t(x)$	Ensemble of T hypotheses from vector x
$P_{j \leftarrow i}$	Probability that element labelled *i* gets labelled as *j*
w	Weight of switching rate
$p(n)$	Probability of resampling nth classifier
$m(n)$	Number of misclassifications
$p(y_i)$	Predicted probability of an event i
$r_k(z)$	Rank function in extreme gradient boosting
h	Second order gradient statistics
s_{ki}	Candidate split points in extreme gradient boosting
$C_D(v)$	Degree centrality of node v
$C_C(v)$	Closeness centrality of node v
$C_B(v)$	Betweenness centrality of node v
V	Total number of nodes in a network
$d_G(v,t)$	Distance between vertices v and t
$\sigma_{st}(v)$	Number of shortest paths through node v
tp	True Positives
tn	True Negatives
fp	False Positives
fn	False Negatives
AFDD	Accumulated freezing degree days
$\Delta AFDDi$	Change in accumulated freezing days over *i* days
Q	Daily flow rate
ΔQi	Change in daily flow rate

LASSO	Least Absolute Selection Shrinkage Operator
MWB	Mid-Winter Breakup
MSE	Mean Squared Error
MAE	Mean Absolute Error

Declaration of Academic Achievement

Following the sandwich thesis requirements set out by the McMaster University School of Graduate Studies, the thesis as presented consists of the following three published research articles and one article currently under preparation for publication in peer reviewed journals:

Paper I: Michael De Coste, Zhong Li, Darryl Pupek, and Wei Sun. (2021) A hybrid ensemble modelling framework for the prediction of breakup ice jams on Northern Canadian Rivers. *Cold Regions Science and Technology*, 103302.

Paper II: Michael De Coste, Zhong Li, and Yonas Dibike. (2022) Assessing and predicting the severity of mid-winter breakups based on Canada-wide river ice data. *Journal of Hydrology*, 127550.

Paper III: Michael De Coste, Zhong Li, and Yonas Dibike. (2022) Machine-learning approach for predicting the occurrence and timing of mid-winter breakups on Canadian Rivers. *Environmental Modelling and Software*, 105402.

Paper IV: Michael De Coste, Zhong Li, and Ridha Khedri. (2022) A hybrid ontology-based semantic and machine learning model for the prediction of spring breakup on Canadian rivers. Submitted for publication to the journal *Computer-Aided Civil and Infrastructure Engineering*.

Chapter 1: Introduction

The vast majority of rivers in Canada develop fully formed ice covers throughout the winter season. These ice covers generally form after an extended period of below freezing temperatures and continue to develop throughout the winter season. These ice covers are often accompanied by low water flow rates and water levels in the rivers and their thicknesses can vary from centimeters up to several meters depending on the river, precipitation, and temperatures throughout the season. With few exceptions, these ice covers will eventually breakup, typically in the spring, through one of two mechanisms. The first, thermal breakup, results in the ice cover melting in place as temperatures gradually increase and snow cover melts, increasing the amount of thermal energy absorbed by the ice. The second, dynamic breakup, occurs when the rise in water levels brought about by the increase in snowmelt runoff outpaces the melt of the ice cover. When this occurs, the ice is lifted out of place due to buoyancy and begins to flow with the river. Both breakup mechanisms are quite common throughout the country, though dynamic breakup has the potential to trigger more severe breakup events. This includes the potential for ice jam flooding, where the free flowing chunks of ice form a semi-permanent dam on the river, greatly restricting the flow capacity and triggering an upstream damming effect. These jams will eventually give way and release a wave of water and ice flowing downstream, similar to a dam release scenario. These events are also no longer limited to occurring in the spring, as the occurrence of Mid-Winter Breakups (MWBs), the breakup of ice cover outside of the typical spring season, is becoming much more common due to the increasing frequency of climate change induced mid-winter thaws. Because of the importance of these events to the flow regime of the rivers as well as their potential to trigger flooding, the ability to accurately predict these events is of vital importance to affected communities and has been the focus of several past studies. Though these needs are recognized,

past studies often focus on applications of process-based physical modelling or analysis of trends in drivers and occurrence, with applications of machine learning limited or non-existent depending on the event in question.

1.1 River Ice Events

1.1.1 Spring Breakup

As stated, the two primary breakup mechanisms are thermal and dynamic. Thermal breakups are often triggered by milder spring temperatures which result in low amounts of runoff entering the river. This is also dependent on the amount of rainfall, as increasing amounts of rainfall both increase precipitation runoff entering the river as well as increasing snowmelt and subsequent snowmelt runoff (Prowse et al., 2007). Conversely, warmer springs often coupled with larger amounts of rainfall are often the trigger of dynamic breakups. In both cases, a series of days with above freezing temperatures is required to trigger the melt required for a breakup to occur. The quantity and magnitude of short-wave radiation is a governing trigger for the type of breakup mechanism (Prowse and Bonsal, 2004). Other factors governing the timing and mechanism of spring breakups include the date and water level at freeze up, the speed of freeze up and subsequent ice roughness, the average temperatures throughout the winter, the amount of snowfall and snowpack depth, and the length of the ice season (Prowse et al., 2007). de Rham et al., (2008a) and de Rham et al., (2008b) investigated the factors and trends governing the spring breakups in a northern Canadian basin, finding that high water events in much of the basin were caused by ice cover breakups, and that breakup timing is trending towards earlier spring breakups. Ice thickness and speed of melt have been found to have a strong influence on the timing of subsequent breakup (Goulding et al., 2009).

1.1.2 Ice Jams

There are a complex number of interacting factors influencing the formation of ice jams including many hydroclimatic and physical factors. The location of a jam is largely governed by physical characteristics of a river, including the channel morphology (slope, width, depth, bed materials, presence of sharp curves, presence of changes in cross-sectional shape, presence of man-made structures) and watershed characteristics (size, retention time, infiltration, vegetation, urbanization) (Beltaos and Prose, 2002). The timing of both freeze up and breakup is a key factor in the occurrence of ice jams. A shorter ice season with mild temperatures and fast warming in the spring results in a less developed ice cover which is not conducive to the formation of ice jams (Beltaos and Burrell, 2003). Conversely early freezes with long ice seasons allow an ice cover to develop further, with early spring breakups often leading to ice jam formation as the cover is stronger at that time (Carr and Vuyovich, 2014). Though there have been many hydroclimatic variables identified as leading factors in the occurrence and severity of ice jams, the complexity of their interaction has led to the conclusion that no single variables can reliably predict these factors on their own, and that combinations must be considered in the development of machine learning models (Robichaud, 2003). The effects of ice jams on the location have been noted to be complex. Severe flooding is often a risk (Beltaos, 2003, De Coste et al., 2017), with communities located on or near the river often sustaining large financial costs and in some cases requiring the relocation of communities, as in the case of the town of Hay River. Conversely, ice jam flooding can also be a beneficial process, with standing basins in some norther regions of the country relying on their occurrence for replenishment (Beltaos et al., 2006).

1.1.3 Mid Winter Breakups (MWBs)

The trigger of these events and the factors governing their occurrence and severity are largely similar to those of spring breakup, with both thermal and dynamic breakup events being possible despite the earlier timing in the ice season (Prowse et al., 2002). These events have been increasing in frequency due to climate change, linked to an increasing trend towards earlier breakups and more common mid-winter warming periods throughout the country (Beltaos and Prowse, 2009). Mild winter days have been noted to be significantly increasing in frequency over the past 80 years, and with them the possibility of these events (Beltaos, 1999). These events, after the re-establishment of ice cover following the thaw, often result in milder spring breakups and reduced flows, and a significantly altered flow regime for the remainder of the season (Huntington et al., 2003).

1.2 Applications of Machine Learning

1.2.1 Spring Breakup

Shouyu and Honglan (2005) applied a combination of fuzzy optimization and ANN to predict both freeze up and spring breakup dates. Both events utilised a different combination of input variables and the input variables were fuzzified before application to the model. Though the model was noted to be successful, the results were noted to be based on limited data, particularly in the case of the breakup dates. Zhao et al., (2012) applied a three-layer feed forward ANN to predict the start of breakup in the Hay River with a specific target of the first crack in ice cover following the start of increasing water levels. The model was found to accurately predict the most recent breakup event, though results degraded over previous years in relation to the availability of data.

Sun and Trevor tested two fuzzy logic modelling systems for the prediction of breakup initiation on 39 years of data from the Athabasca River. Both of the models attained success in

different aspects, with one found to be more suitable for pre-screening of ice flooding risk and one more effective at predicting maximum breakup water levels than previously developed models. It was noted that limited data availability was a potential issue in the overall reliability of the models. Sun and Trevor (2018a) subsequently developed a two-level stacking ensemble system for predicting breakup dates, testing numerous combinations of model structures and configurations. The models used the same combinations of input variables with the simpler algorithms being found to outperform the more complicated and computationally intensive models, with results deemed promising but needing development. Sun (2018) applied the stacking ensemble system to predict breakup timing using a different combination of component models, with 24 models tested against 36 years of data using climatic ang hydrologic inputs. The simple averaging methods utilized were most successful, with the technique noted to have transferability to different locations. Sun and Trevor (2018b) applied multiple model combinations in predicting maximum breakup water levels including fuzzy logic, ANNS and multiple linear regressions as well as simple averaging methods. 36 years of data were used with input data including hydrologic and climatic variables, with varying but strong model performance obtained with the considered algorithms. ANN was found as the most successful combining model, but limitations in data were once again noted as a factor degrading results.

1.2.2 Breakup Ice Jams

Though extensive process-based modelling of ice jams has been conducted using varying physical and numerical based models (De Coste et al., 2018), the number of studies utilising machine learning to directly predict ice jam occurrence is limited. Madaeni et al., 2020 provides an review of the application of artificial intelligence techniques applied to ice jam forecasting, noting some of the earliest applications were conducted by Appelbaum (1982) and Barnes-Svarney

and Montz (1985), analysing ice jam frequency and predicting ice jam occurrence based on time series analysis respectively. White (1996) was one of the earliest applications of traditional data driven modelling to the forecasting of river ice jams. A logistic regression model was applied in the prediction of ice jams with a one-day lead time. The utilised data focussed on climatic values leading up to the jam occurrence, but the total dataset was limited to only 39 observations, with 28 being ice jams. Though the model was successful, its utility was left in question due to the limited analysis. Masie et al., (2002) followed this study with the application of Artificial Neural Networks (ANN) for same day prediction of ice jam occurrence. Though the developed model was found to be 93% accurate on predictions using variables combining both climatological and hydrological factors, only 17 ice jams were included leading to some problems in the selection of data and the potential model utility.

Mahabir et al., (2006) applied a hybrid neuro-fuzzy model similar to that of Shouyu and Honglan (2005), predicting the occurrence of ice jams. The model was based around a long lead time of 3-4 weeks and used corresponding variables, but the ensuing results were poor due to the combination of long lead times and limited data availability. Mahabir et al., (2007) subsequently tested the transferability of the previous model on a more remote location, with results being noted as particularly poor in quantitative assessments, indicating a need to reconstruct and reevaluate the model for the new location, though the transferability to low population and low data locations was shown to be promising. Guo et al., (2018) applied a backwards propagation ANN to predict ice jam occurrence against a historical record of 23 ice jam events. The model was found to be 85% accurate but the data limitations were again noted.

1.2.3 Mid-Winter Breakups

Direct applications of machine learning to the prediction of MWBs have been extremely limited, with no formal application of traditional algorithms. Prowse et al., (2002) proposed a threshold for the triggering of MWBs based on climatic data, finding it accurate for the limited region of its application. Likewise, Carr and Vuyovich (2014) developed a threshold including both temperature and precipitation as factors for the triggering of an MWB, again demonstrating its accuracy for the specific region it was developed for. Newton et al., (2017) developed a larger database of MWB events in the western portion of Canada and the United States where it was shown that the above proposed thresholds were quite inaccurate when translated to this region. Though this study featured further analysis of MWBs and their drivers, it included no formal application of machine learning techniques directly to this phenomenon.

1.3 Objectives and Organization

1.3.1 Research Objectives

Though the ice phenomena described through the literature review have been extensively researched from a physical and process-based perspective, the attempts at predicting these events using machine learning have been limited in scope and amount. Data availability is often the primary limiting factor, with studies generally focussing on single rivers or regions, and in the case of some events, relying on extremely simplified datasets. There is also no generally accepted methodology for modelling these events, with existing applications reliant on relatively simple machine learning applications and with minimal attempts at implementing non-traditional model structures. Rare event forecasting techniques are well suited to these event types, though these techniques have been applied minimally if at all. Further, these applications also face challenges

in presenting the data in a meaningful way for stakeholders, with decision making support tools often being unintuitive and difficult to transfer between locations.

With this in mind, this thesis aims to investigate the application of machine learning in 1) predicting the presence of ice jams on a river, 2) predicting the occurrence, timing, and severity of MWBs while implementing rare event forecasting techniques, and 3) the prediction of the timing of spring breakups in conjunction with ontology-based modelling. Case studies are implemented for each application, with large scale national applications achieved for all breakup investigations. The results obtained illustrate the effectiveness of the applied techniques while also demonstrating their viability as decision making support tools that can be easily implemented on regional or national scales.

1.3.2 Thesis Organization

To achieve the stated objectives, this dissertation is organized as follows:

In Chapter 2, a hybrid ensemble framework for the prediction of ice jam occurrence on the Saint John River in New Brunswick is proposed. This framework combines the results of multiple machine learning algorithms at two levels and after extensive testing was proved to outperform other modelling methods and single model algorithms for the same purpose.

In Chapter 3, an application of machine learning in assessing the severity of MWBs on a national scale in Canada is proposed. Key variables leading to the occurrence and affecting the severity of MWBs are identified, and a new threshold for the triggering of MWBs in Canada is described, outperforming the accuracy of previously proposed thresholds to a significant degree.

In Chapter 4, a method for predicting the occurrence and timing of MWBs in Canada using a two-level machine learning framework is proposed. Techniques implemented focus around rare-

event forecasting, with the final model configuration producing a high level of accuracy in the results.

In Chapter 5, a novel hybridization of machine learning and ontology-based modelling for the prediction of spring breakup timing is proposed. The ontology model allows the application of network analysis techniques to the data used in the machine learning model, allowing a refined selection of variables to be made. The success of the models is demonstrated on a national scale.

Chapter 6 presents the conclusions and contributions of this dissertation, with further recommendations for research provided.

Chapter 2: A Hybrid Ensemble Modelling Framework for the Prediction of Breakup Ice Jams on Northern Canadian Rivers

Prediction of ice jams using machine learning has been previously attempted but has encountered issues in the selection of both algorithms and organization of data. In this chapter, a stacking ensemble approach is presented wherein the results of a set of first level models are used as the input for a second level model in the prediction of ice jam presence on the Saint John River in New Brunswick. A variety of machine learning algorithms ranging from single models to complex ensembles were tested. Combinations of input variables and algorithms were extensively tested using grid searching with accuracy assessed based on common binary classification metrics. The final model achieved a high level of accuracy on the prediction of both ice jam and non-ice jam occurrence.

Michael De Coste was responsible for investigation, model development and testing, and results validation under the guidance of Dr. Zhong Li and Dr. Wei Sun. Data from the Saint John River was supplied by Darryl Pupek. The manuscript was drafted and prepared by Michael De Coste and revised by Michael De Coste, Dr. Zhong Li, and Dr. Wei Sun.

This chapter has been published: Michael De Coste, Zhong Li, Darryl Pupek, and Wei Sun. (2021) A hybrid ensemble modelling framework for the prediction of breakup ice jams on Northern Canadian Rivers. *Cold Regions Science and Technology*, 103302. (DOI: 10.1016/j.coldregions.2021.103302)

Abstract

The forecasting of river ice jams faces challenges relating to both the availability of data and the complexity of river ice dynamics, resulting in difficulties in model formulation. In this study, a hybrid ensemble modelling framework is developed to address the data scarcity issue and leverage advanced machine learning techniques for the prediction of ice jams with a one-day lead time. The proposed methodology utilises data easily monitored in advance of any ice jam events and maintains a realistic balance between ice jam and non-ice jam events. A combination of both single model algorithms, including classification trees, logistic regression, *k*-nearest neighbors, support vector machines, and artificial neural networks, and ensemble model algorithms, including random forest, adaptive boosting, gradient boosting, variance penalizing adaptive boosting, logistic boosting, class switching, and adaptive resampling and combining, are considered for both the member models of the first layer of the hybrid ensemble and for the ensemble combiner of the second layer. The final selection of both variables and member models for the hybrid ensemble is detailed, with a focus on the reduction of false negatives, the prediction of no ice jam on a day when one occurs. The proposed method is applied to the St. John River in New Brunswick, Canada, in a location particularly prone to ice jam flooding. Using the proposed methodology, a final model combining 6 different member models using a support vector machine as the ensemble combiner was produced, with a balanced prediction accuracy of 86%. This hybrid ensemble model outperformed the other tested ensemble models, as well as a series of generalized models produced using all available input variables and member models. The model also performed well against other ensemble techniques and against the individual member models. These results demonstrate the viability of the proposed methodology in constructing a hybrid ensemble model for the forecasting of ice jams on Northern Canadian Rivers. The techniques utilised can be adapted to

other locations to facilitate ice jam forecasting, requiring data that is easily available and monitored in advance of any potential flooding events.

Keywords: River ice; ice jams; flooding; forecasting; data-driven modelling; ensemble learning

2.1 Introduction

A problem faced by many northern Canadian rivers is vulnerability to ice jam flooding in the early spring during ice cover breakup. These floods typically occur following a dynamic ice cover breakup, a volatile event where the ice on the river is lifted out of place by increasing water levels before it melts. The ice will break first into discrete panels, and then smaller chunks, resulting in it flowing freely downstream. If the flow of the ice becomes restricted, either at a location where the channel geometry changes or where there is still intact ice cover, the ice chunks can bridge the channel and begin to accumulate and thicken. The increasing accumulation of ice is called an ice jam, and this often triggers a damming effect, reducing the available flow area in the channel (Massie et al., 2002). As daily temperatures continue to rise, increasing runoff and flows into the river while melting the ice further, the jam will weaken, eventually giving way and releasing a wave of ice and water that will flow downstream. In some rivers this can cause a cascading effect of ice jam formation and release, and as a result, communities both upstream and downstream of an ice jam are at risk for flooding. Because of their volatility, the development of accurate ice jam forecasting techniques is of great interest both for the purposes of early warning and mitigation.

Data driven modelling techniques including machine learning have been applied to both ice jam and ice breakup forecasting on rivers in past studies, with varying degrees of success. White (1996) applied logistic regression modelling to predict breakup ice jams on the Platte River in Nebraska, USA, with a dataset consisting of 28 historic ice jams and 11 non-ice jam dates. The model achieved a classification accuracy of 87%, though the dataset was heavily weighed towards ice jams. Massie et al., (2002) focussed on predicting the timing of ice jams using Artificial Neural Networks (ANN) in Oil City, Pennsylvania. Only 17 historic ice jams were available to be tested,

producing a final model with a 93% accuracy in predicting ice jams on the same day. Mahabir et al., (2006) utilized a hybrid neuro-fuzzy model for the prediction of ice jams. The system used four long lead time (3-4 week) input variables, but the obtained quantitative ice jam prediction results were poor and would require more years of calibration data and additional testing. Wang et al, (2010) aimed to forecast ice jam water levels and thickness in the Quyu reach of the Yellow River using a Back-Propagation Artificial Neural Network (BP-ANN) and a Support Vector Machine (SVM). The developed models outperformed existing regression models for both methods based on a tolerable relative error of 20%, however performance sharply declined as the tolerable relative errors were decreased, and the available data at the time of the study was extremely limited. Zhao et al, (2012) aimed to forecast the timing of spring breakup in Hay River using a three-layer feed forward ANN. The developed model showed potential, with accurate prediction of the 2011 breakup event; however, the performance in other years did not exceed that of existing models and limited data constrained the accuracy of results. Sun and Trevor (2015) aimed to forecast the breakup of ice cover on the Athabasca River using a Qualitative Fuzzy Logic Model (QFLM) and an Adaptive Neuro-Fuzzy Inference System (ANFIS). It was concluded that the QFLM model could act best as a seasonal risk assessment tool, with the ANFIS model as a forecasting tool, however improvement of the models with additional data was stated as a necessity as there were only 39 samples available in model development. Guo et al., (2018) used a BP-ANN to predict the occurrence of ice jams based on historical ice jam records. The developed model was 85% accurate in forecasting jam occurrence, with plans for future research being the addition of event observations to the data series as well as added climatic and hydrological variables as the base data only had 23 ice jam events.

Several limitations consistently arise in these and other river ice jam forecasting studies using data driven modelling. Many of the studies cite a lack of data resulting from the difficulties in obtaining both observations and direct field measurements, particularly in ice jam conditions. The configuration of ice jam dates in the modelling dataset is an additional issue that has arisen in these studies, with some opting to model seasonally, while others reduce the quantity of dates without ice jams through statistical algorithms to improve accuracy, but there is no well accepted procedure for this. Other techniques exist to address these data limitations such as surrogate data testing (Shaiknina and Khovanova, 2017), which builds additional data mimicking the statistical properties of the original data, or virtual sample generation and over-sampling (Li et al., 2018) which generates new data falling in the gaps between existing data, however these techniques rely on the use of generated data, not solely observational data. Further, new ensemble learning algorithms have been developed and applied to river ice related modelling tasks, but not yet to ice jam forecasting specifically. For example, Sun and Trevor (2018a) predicted ice breakup dates on rivers using a stacking ensemble model consisting of a two-level system of member and combining models. The combining models utilised Simple Averaging Methods (SAM) with varying combinations of component models tested. A SAM model using only ANN member models performed the best, with the results deemed promising for application to other river ice forecasting problems. Sun and Trevor (2018b) applied multiple model combination methods in the estimation of maximum water levels during the spring breakup of ice cover. The developed models utilise ANFIS, ANN, and multiple linear regression models as member models and as the combining method for the ensemble models for comparison, using data from 36 years. While performance for each of the member models varied in ability to predict maximum water levels optimal configurations of each were found. Sun (2018) focused on the use of a Stacking Ensemble Tree

Model (SETM) to predict breakup timing on the Athabasca River. Classification and regression trees and M5 (using regression tree with linear regressions at leaves) models were tested as both base and ensemble models. 12 of each model type were tested on 36 years of data. The SETM using simple averaging methods performed the best, and the flow and temperature in the freeze up and breakup months were the most important variables in prediction. The technique was noted as easily applicable to other river ice forecasting problems. Barzegar et al., (2019) developed an extreme learning machine model, a least squares support vector machine model, and their bootstrapping versions to estimate river ice thickness in the Mackenzie River Basin in the Northwest Territories, Canada. The results indicate that the bootstrapping extreme learning model approach using meteorological variables is a promising prediction tool for river ice thickness.

This study focuses on the development of a hybrid ensemble framework for forecasting the occurrence of ice jams. The model construction included variables to account for the trends in temperature and flow and utilised a stacking ensemble structure for predictions. The modelling dataset did not reduce the amount of non-ice jam dates included for model training, resulting in a more realistic proportion between these days and days with ice jams.

A case study of the ice jams on the Saint John River in New Brunswick, Canada, is used to demonstrate the applicability of the developed framework. The proposed hybrid ensemble structure can provide a more robust and flexible approach to account for the limited amount of ice jam data available for model training. It also leverages the advantages of various advanced machine learning techniques concurrently providing more diverse predictions than those produced from single model algorithm structures. To the best of our knowledge, this is the first application of these techniques in ice jam forecasting.

2.2 Methodology

2.2.1 Hybrid Ensemble Modelling

The structure of the hybrid ensemble model utilised in this study consists of two internal levels, illustrated in Figure 2.1. The first level of the ensemble consists of multiple member models, each of which is trained and tested on the same original input dataset and outputting an individual prediction. These predictions are combined into a second level input dataset, which is used to train and test the second level combiner model. The second level model only uses the outputs from the first level models as inputs in its predictions, outputting a final prediction for the overall model (Sun and Trevor, 2017). Major advantages of this model structure are the reduction of variance in forecasting errors and the correction of model biases that would not be possible with a single model structure. Multiple model types are considered for both the member models and combiner model and tested by various methods. Tested models for both the ensemble members and combiner included single model algorithms, including classification trees, logistic regression, *k*-nearest neighbors, support vector machines, and artificial neural networks, and ensemble model algorithms, including random forest, adaptive boosting, gradient boosting, variance penalizing adaptive boosting, logistic boosting, class switching, and adaptive resampling and combining.

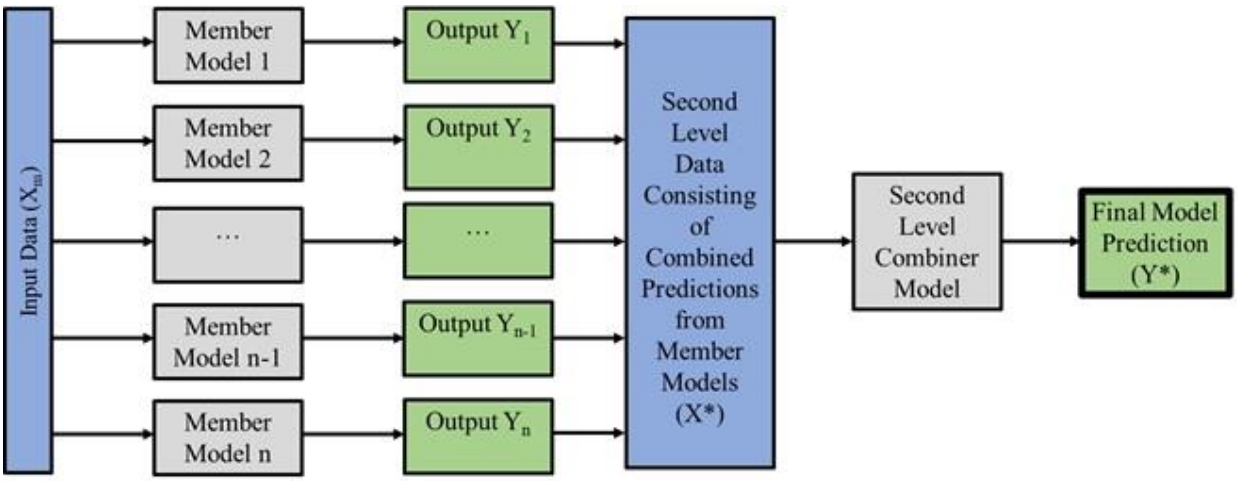


Figure 2.1: Schematic of hybrid ensemble model structure.

2.2.2 Utilised Model Algorithms

A comprehensive selection of binary machine learning algorithms were considered for both the member models and the combiner model to ensure the flexibility of the modelling framework. The selected algorithms consist of types that are commonly applied to binary prediction tasks and have not yet been applied in the field of river ice. As none of these models have yet been utilised in a hybrid ensemble framework for the predictions of river ice jams, the potential benefits each may have were not known prior to modelling. Thus, it was elected to treat each model equally without applying either ranking or differing weights to the model results. These algorithms are summarized below.

2.2.2.1 Single Model Algorithms

Classification Tree (CT) (Safavian and Landgrebe, 1991): A CT or decision tree model produces a single classification for a set of input variables based on decisions made at internal model splits. Each split is based on the value of a single input variable, with the number of branches within the model and the number of splits being tuned by the user. The model terminates in leaf nodes

representing a final classification, with the model structure resembling a tree. Singh et al. (2014) successfully applied CT to determine the relationship between catchment similarity and the performance of transferred parameters predicting runoff from rainfall.

Logistic Regression (LR) (White, 1996): An LR model produces a binary prediction outcome via a linear combination of variables in a similar fashion to linear regression. The logistic regression uses a logistic function (Equation 2.1) to model a binary dependent variable.

$$f(x) = \frac{e^u}{1 + e^u} \tag{2.1}$$

where *f(x)* is the probability of the binary response being 1, and *u* is a linear function of input variables $X_1, X_2, \ldots, X_i$ defined as below.

$$u = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i \tag{2.2}$$

where α is the regression intercept, and β_i is the regression coefficient for each variable X_i. Szelag et al., (2019) applied LR to predict the occurrence of storm overflow in a stormwater system based on total rainfall and duration.

***k*-Nearest Neighbors** (KNN) (Dudani, 1975): In a KNN model, unlabelled observations are classified based on the labels of the *k* nearest classified points, classifying the unlabelled point to the class most heavily represented amongst its *k* nearest neighbors. The value of *k* is the primary parameter for tuning the model, with each new observation being labelled based on a weighting factor comparing the distances to the *k* nearest neighbors from the labelled point. As described in Equation 2.3, w_j is the weight for an individual observation *j*, and the distance between neighbor points are given by d_j, j=1,…,*k*. The value of a new observation is assigned to the class with the highest sum of weights.

$$w_j = \begin{cases} \frac{d_k - d_j}{d_k - d_1}, d_k \neq d_1 \\ 1, \quad d_k = d_1 \end{cases} \tag{2.3}$$

Liu et al., (2020) used a KNN model in conjunction with a recurrent neural network to provide real time flood forecasting on three differing catchments, demonstrating its effectiveness for local decision-making support. Semenova et al., (2020) used KNN to forecast whether ice jams will occur in a given season at the confluence of the Sukhona and Yug Rivers in Russia, achieving an 82% accuracy.

Support Vector Machine (SVM) (Suykens and Vandewalle, 1999): An SVM maps data into a higher dimensional input space followed by constructing an optimal separating hyperplane for classification. This process entails the construction of a classifier from a set of *N* data points $\{x_k, y_k\}_{k=1}^{N}$ in the form of Equation 2.24, where α_k and b are positive real constants and $\psi(x, x_k)$ is described by Equation 2.5 in a polynomial SVM of degree *d*.

$$f(x) = \sum_{k=1}^{N} \alpha_k y_k \psi(x, x_k) + b \tag{2.4}$$

$$\psi(x, x_k) = (x_k^T x + 1)^d \tag{2.5}$$

Wang et al., (2019) used an SVM coupled with a Copula function to simulate runoff data for the Yangtze River in China, overcoming severe limitations in the available data which was deemed insufficient for traditional modeling methods.

Artificial Neural Network (ANN) (Agatonovic-Kustrin and Beresford, 1999): The feed-forward ANN structure utilised in this study maps the relationship between input and output variables by developing weights for the input variables, a layer of internal nodes, and the output nodes. A nonlinear activation function allows for nonlinear mapping between the variables and nodes, while

the number of nodes in the internal layer must be tuned alongside the training algorithm. Massie et al., (2002) applied ANNs in forecasting ice jam occurrence on rivers and Tsakiri et al., (2018) applied ANNs in forecasting seasonal river flooding.

2.2.2.2 Ensemble Model Algorithms

Random Forest (RF) (Liaw and Wiener, 2002): As an extension of the bootstrap aggregating (also called bagging) technique, RF is an ensemble of classification trees developed from bootstrap samples of the data set. The bootstrap samples are resampled from the overall dataset. Each bootstrap sample can be used to train an independent classification tree, and the prediction of the random forest, $\hat{f}(x)$, can be generated based on all bootstrap samples, using Equation 2.6.

$$\hat{f}\ (x) = \frac{1}{M}\sum_{m=1}^{M} \hat{f}^{m}\ (x) \tag{2.6}$$

where $\hat{f}^{m}(x)$ is the prediction trained on the m^{th} bootstrap sample, and M is the number of bootstrap ensembles. Each tree in the random forest is split based on a randomly selected subset of variables made available to each tree in its construction, providing robustness against overfitting. An RF model was used by Munoz et al., (2018) in the prediction of flash-flooding in the Ecuadorean Andes catchment.

Adaptive Boosting (AdaBoost) (Zhu et al., 2009): AdaBoost is an iterative ensemble procedure for two-class classification problems that builds a series of classification trees with associated class labels. The initial classifier is constructed first, producing class labels. If a training data point is incorrectly classified, the weight of that point is boosted with the Stagewise Additive Modelling using a Multi-class Exponential (SAMME) loss function, and a second classifier is constructed using the new weights. The process aims to minimize the misclassification rate through the

weighting of the weak learners' accuracy. AdaBoost was successfully applied to the prediction of flash flooding in the Lao Cai province of Vietnam by Bui et al., (2019).

Gradient Boosting (GBoost) (Friedman, 2002): GBoost alters the AdaBoost algorithm by solving the internal trees through fitting them by least squares to their current residual errors, followed by a single parameter optimization of the expansion coefficient based on the general loss criterion. Papacharalampous et al., (2019) applied gradient boosting in the prediction of hydrological model errors, successfully combining the results with those of other algorithms to reduce model errors.

Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016): XGBoost functions similarly to Gradient Boosting but uses the second order gradients of the loss function in the minimization of the overall model error. Additionally, the method utilises advanced regularization to improve model generalization. XGBoost was applied by Wu et al., (2019) in the prediction of mean evapotranspiration based on temperature data, achieving a higher level of accuracy than other non-tree-based methods applied for the same purpose.

Variance Penalizing Adaptive Boosting (VadaBoost) (Shivaswamy and Jebara, 2011): VadaBoost is a similar algorithm to the AdaBoost algorithm, with the weighting function combining both the empirical risk and empirical variance, creating a trade-off between the two quantities allowing a minimization of the upper-bound of true risk.

Logistic Boosting (LogitBoost) (Cai et al., 2006): Functioning similarly to the AdaBoost algorithm, LogitBoost utilizes the minimization of a binomial log-likelihood loss function changing linearly with classification error. This process recasts the AdaBoost model as a generalized additive model.

Class Switching (CS) (Narassiguin et al., 2016): CS is a variant of an output flipping ensemble approach where class labels are randomly flipped within the data set. The final classifier decision is given by majority vote over the base classifiers, with the parameter *p* representing the rate of class switching as a percentage of the overall dataset, specified by the user.

Adaptive Resampling and Combining-X4 (ArcX4) (Breiman, 1996): Belonging to the Arc algorithm family, this algorithm functions by sequentially training a series of classifiers similarly to the AdaBoost algorithm. At each step a resampled training set is constructed based on a set of probabilities which are then updated and weighted based on the number of misclassifications at the present step. Upon termination, the classifiers are then combined using unweighted voting, a difference from other Arc algorithms which use weighting for recombination, returning the majority predicted class.

2.2.3 Performance Metrics and Cross-Validation

The dataset was divided into training and testing sets using a randomized 70/30 split stratified between the classes, such that 70% of ice jam and non-ice jam days were in the training set and 30% were in the testing set. As the modelling task was based around a binary classification, model performance was assessed based around the quantity of correct predictions for both positive and negative results. There are four possible outcomes of predictions that may be obtained in binary classification, with true positive in this study denoting the correct prediction of the occurrence of an ice jam and true negative denoting the correct prediction of no ice jam occurrence. Conversely, false positive would denote the prediction of an ice jam when one does not occur, while false negative would denote the prediction of no ice jam when one does occur. Three performance metrics were considered in this study: Recall, specificity, and balanced accuracy, defined in Equations 2.7-2.9 (Brodersen et al., 2010).

$$Recall = \frac{tp}{tp + fn} \tag{2.7}$$

where *tp* is true positives and *fn* is false negatives.

$$Specificity = \frac{tn}{tn + fp} \tag{2.8}$$

where *tn* is true negatives and *fp* is false positives.

$$Balanced\ Accuracy = \frac{Specificity + Recall}{2} \tag{2.9}$$

As each of the considered models have degrees of freedom in the configuration of their hyperparameters, cross-validation was used to select the best performing model parameters. Cross-validation divides the training set into *K* equally sized parts, using *K-1* of the parts as a training set and 1 part as a validation set (Bengio and Grandvalet, 2004). This process is done *K* times, such that each individual part of the set is used as the validation set. This procedure is completed for each of the considered values of model configuration, producing an averaged error metric for each to facilitate parameter selection. A 5-fold grid search cross-validation was utilised in this study for the selection of model parameters. In ice jam forecasting it is of most importance to avoid false positives, where the model predicts that an ice jam does not occur when it does. As a measurement of the occurrence of true positives against combined false negatives and true positives, Recall was used as the primary assessment metric for the cross-validation conducted on all models.

2.2.4 Selection of Input Variables

Selection of input variables was conducted based around the variable importance values obtained from a trained RF model and the Recall values obtained from a trained CT model. The variable importance provides an indication of which variables provide the greatest contribution to lowering of Gini impurity of the model predictions. The Gini impurity is based around the overall

classification rate, assessing both the misclassification of ice jam and non-ice jam events. Because of this, the next step of in-depth variable testing used the construction of CT models considering all possible input variables followed by developing models with single variables removed and with the removal of pairs of variables, a process called Input Omission (IO) (Snieder et al., 2020). As the modelling target is a binary prediction, the changes in Recall values from that obtained using the full input variable set were assessed to determine the impact of the IO, allowing both important and redundant variables to be identified. Following this, redundant variables were removed, important variables retained, and remaining variables were further tested through an IO process, with a final reduced input variable set determined based on the best modeling performance.

2.2.5 Selection of Members and Combiner Models

Following the selection of the input variable set, each of the considered models' parameters were tuned using 5-fold cross-validation. Following this, the 70/30 data split was used to train and validate each of the tuned models. These outputs were then combined into a new input dataset for the second level model. Similar to the selection of input variables for the first level model, the selection of member models to include in the final hybrid ensemble began through assessing the Recall of a CT built with all of the member models, followed with IO of individual models and model pairs. Member models whose exclusion either had no impact on the final ensemble Recall or improved the results were noted, allowing a final selection of ensemble combiners to be made. Following this, a series of ensemble combiners were tuned using 5-fold cross-validation, with the tuned models then applied to the same 70/30 data split, providing final predictions. Recall, specificity, and balanced accuracy were used to make a final assessment of the most effective combiner. To further validate model selection, a generalized model was developed including all

candidate variables and model algorithms. This would allow further comparison and validation of the curated approach.

2.2.6 Algorithm Implementation

Analysis was completed using Python, version 3.7 (Van Rossum and Drake, 2009). Additional packages used in the implementation of the algorithms included *Keras* (Gulli and Pal, 2017), *scikit-learn* (Pedregosa et al., 2011), *Numpy* (Oliphant, 2006), and *Pandas* (McKinney, 2010). Additional code for the implementation of VadaBoost, LogitBoost, CS, and ArcX4 models was adapted from Narassiguin et al. (2016).

2.3 Case Study

2.3.1 Study Area

The location of the study area for this project is the middle reach of the Saint John River in New Brunswick, Canada. The middle reach, shown in Figure 2.2, begins just downstream of the town of Grand Falls and extends approximately 60 km to just upstream of Bath. Bed elevations in this region vary between 60-90 m above sea level, with water surface elevations typically varying between 72-96 m. Channel manning's roughness coefficient ranges from 0.012-0.024 while floodplain roughness ranges from 0.040-0.060. There are three major tributaries entering the middle reach, the Little, Tobique, and Aroostook Rivers. The reach has historically been highly vulnerable to ice jam flooding, with historical ice jam dates and locations shown in Figure 4. Historic ice jams have induced flows of up to 8000 m^3/s at the Beechwood Generating Station at the downstream end of the reach. Just upstream of the reach is the Grand Falls Generating Station, one of the five hydropower structures on the Saint John River, originally constructed in 1925. This structure is responsible for a significant portion of power generation in the region and is also a governing factor on flows in the middle reach. The middle reach is home to many small

communities, including the town of Perth-Andover, which has been particularly vulnerable to ice jam flooding, with a major flood occurring in 2012 resulting in a state of emergency being declared and several millions of dollars in damages (Knack and Shen, 2018).

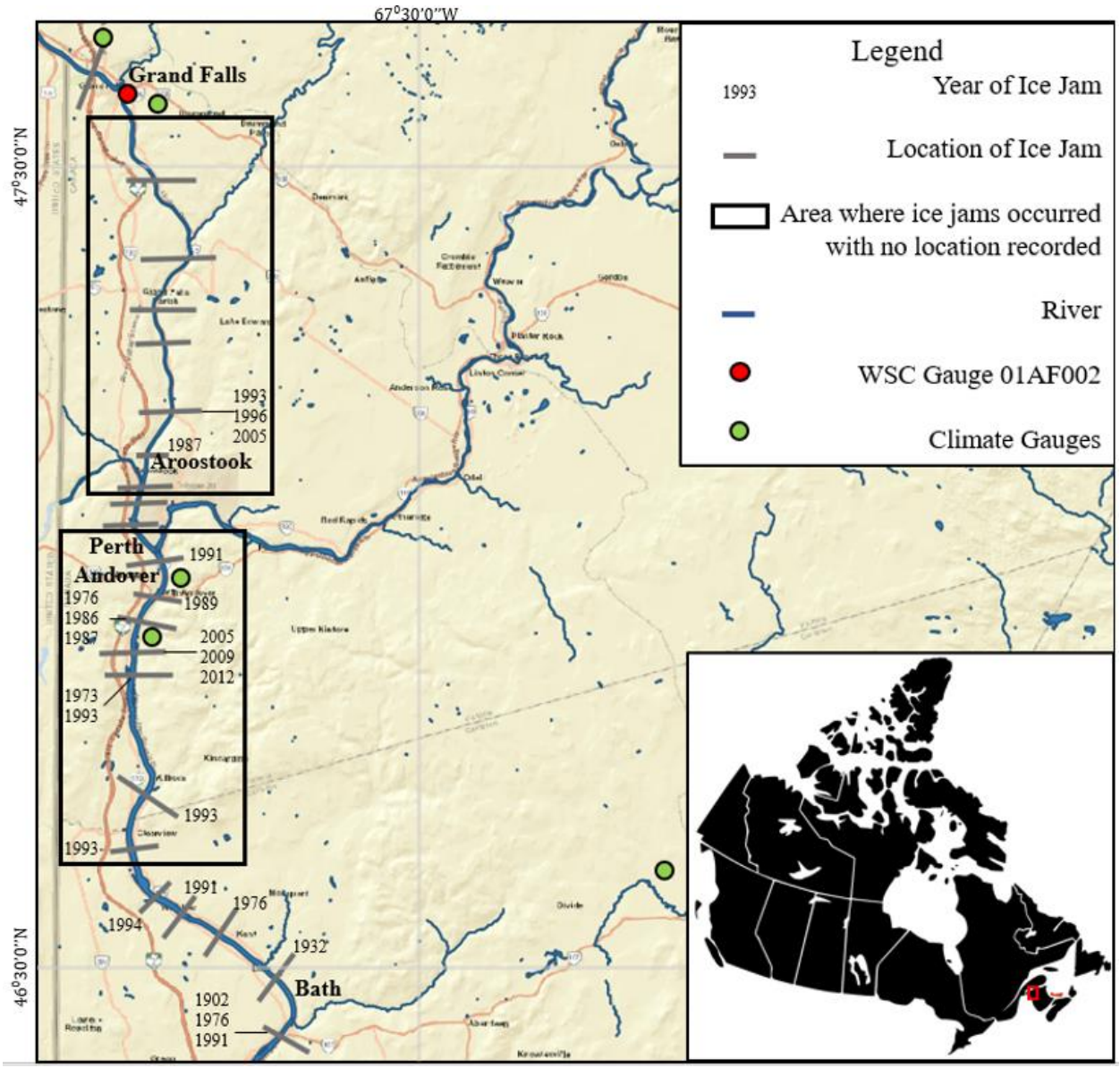


Figure 2.2: Map of the Middle Reach of the Saint John River showing locations and timing of historical ice jams (Adapted from Environment and Local Government, 2013).

2.3.2 Data

Data used in this study included meteorological data obtained from both Environment Canada and the Government of New Brunswick, as well as ice jam observations from the New Brunswick Department of Environment and Local Government. Data sources and types are summarized in Table 2.1. The data spanned 1988-1990, 1995, and 2011-2017. For each year of available data, only the data in the typical ice jam season (March 20 – April 30) was included to reduce both the presence of redundant data and the effects of imbalanced classes in modelling. It was important to include all non-ice jam days present within this span to ensure that the developed model was structured around a realistic distribution of data, ensuring that it would also be able to function in years where no ice jams occur. Figure 2.3 shows the timeline of both ice jam and non-ice jam events during the modeling period. The historical ice jams all occurred within the middle reach of the St. John River however some location data was unable to be measured at the time, thus location data was excluded from modelling. A total of 29 input variables were considered, including daily flow rate (Q, m^3/s), average daily temperature (T, ℃), and daily precipitation (P, mm) up to day i, Accumulated Freezing Degree Days ($AFDD$, ℃), change in $AFDD$ over the past 1-15 days ($\Delta AFDDi$ or $DAFDDi$, ℃), and change in Q over the past 1-10 days (ΔQi or DQi, m^3/s). These variables are detailed in Table 2.2 and their selection was suggested by Mahabir et al., (2006), Massie et al., (2002), and White (1996) as potential indicators of ice jam occurrence. $AFDD$ measures the negative sum of temperatures from a set date, in this case October 1st, which marks the beginning of the water year for New Brunswick. The $\Delta AFDDi$ values account for the rate of temperature change during the spring breakup season, and ΔQi accounts for the changes in runoff entering the river. Each variable is available up to day i, with the prediction target of ice jam occurrence on day i+1 providing a one-day lead time.

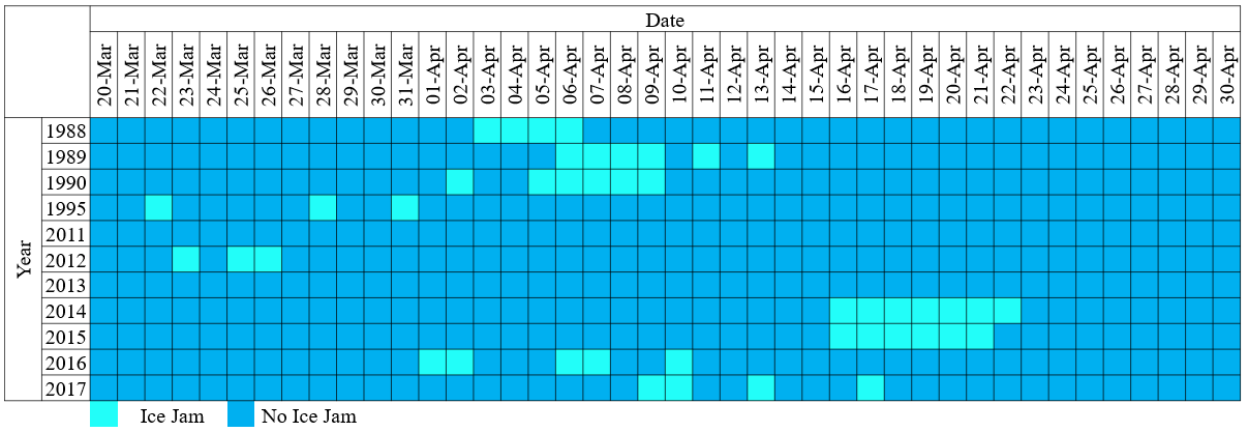


Figure 2.3: Timeline of ice events in modelled dataset.

Table 2.1: Summary of data sources and types.

DATA	TYPE	SOURCE
Flow	Continuous	Water Survey of Canada Gauge St. John River at Grand Falls
Temperature and Precipitation	Continuous	Environment Canada Gauge at Aroostook, missing data filled with Environment Canada gauges at Juniper and St. Leonard, and New Brunswick Climate Gauges at Andover and Grand Falls
Ice Jam Dates	Binary	NB Department of Environment and Local Government

Table 2.2: Summary of considered input variables.

Variable	Meaning	Unit	Minimum	Average	Maximum
Q	Flow	m^3/s	30.5	930.3	3880
T	Daily mean temperature	℃	-15.3	1.75	18.3
P	Precipitation	mm	0	2.04	33.2
AFDD	Accumulated Freezing Degree Days	℃	122.9	703.8	1110
DAFDD1	Change in AFDD in last 1 day	℃	-18.3	-1.76	15.3
DAFDD2	Change in AFDD in last 2 days	℃	-31.3	-3.24	29.1
DAFDD3	Change in AFDD in last 3 days	℃	-42.8	-4.51	43.1
DAFDD4	Change in AFDD in last 4 days	℃	-49.1	-5.56	51.6
DAFDD5	Change in AFDD in last 5 days	℃	-58.9	-6.45	59.4
DAFDD6	Change in AFDD in last 6 days	℃	-63.9	-7.12	66.4
DAFDD7	Change in AFDD in last 7 days	℃	-74.8	-7.59	73.7
DAFDD8	Change in AFDD in last 8 days	℃	-81.1	-7.85	76.5
DAFDD9	Change in AFDD in last 9 days	℃	-86.1	-7.91	88.1
DAFDD10	Change in AFDD in last 10 days	℃	-97.1	-7.77	102.1
DAFDD11	Change in AFDD in last 11 days	℃	-108.6	-7.36	109.1
DAFDD12	Change in AFDD in last 12 days	℃	-117.9	-6.70	116.4
DAFDD13	Change in AFDD in last 13 days	℃	-124.2	-5.79	123.2
DAFDD14	Change in AFDD in last 14 days	℃	-129.2	-4.57	133.2
DAFDD15	Change in AFDD in last 15 days	℃	-134.7	-2.98	146.6
DQ1	Change in Q in last 1 day	m^3/s	-980	43.6	2410
DQ2	Change in Q in last 2 days	m^3/s	-1490	86.7	3206
DQ3	Change in Q in last 3 days	m^3/s	-2070	127.7	3445
DQ4	Change in Q in last 4 days	m^3/s	-2380	166.1	3547
DQ5	Change in Q in last 5 days	m^3/s	-2610	203.9	3599
DQ6	Change in Q in last 6 days	m^3/s	-2760	242.4	3602
DQ7	Change in Q in last 7 days	m^3/s	-2891	280.2	3598
DQ8	Change in Q in last 8 days	m^3/s	-3000	316.4	3598
DQ9	Change in Q in last 9 days	m^3/s	-3077	351.6	3620
DQ10	Change in Q in last 10 days	m^3/s	-3130	384.6	3673

2.4 Results and Analysis

2.4.1 Variable Selection

Variable selection began with the assessment of variable importance obtained from an RF model. As the goal of the model was prediction of both ice jam and non ice jam events, the importance of variables in both predictions was used rather than just the importance in predicting

solely ice jam events. The model was first tuned using 5-fold grid search cross-validation including all 29 considered input variables. The tuned parameters were then used to train and validate a single model using the 70/30 data split. Figure 2.4 plots the obtained variable importance values from the trained model. *AFDD* had the highest importance, with several of the Δ*AFDDi* and Δ*Qi* variables having high importance relative to others. It is important to note that this variable importance metric considers the overall accuracy in predicting both possible outcomes, which weighs much more heavily towards the prediction of non-ice jam events given the configuration of the modelling dataset.

Following this, in depth testing of the effects of IO on Recall was conducted by both single variable omission and pairs variable omission in a CT model. Figure 2.5 shows the changes in model Recall that occurred from IO of a single variable and the IO of pairs of variables. In each case, the Recall obtained from the model built with all variables was subtracted from the Recall of a model built excluding the specified variables, with a negative value representing a decrease in model accuracy. The *AFDD* variable was again shown to have a higher degree of importance, with the majority of cases including its omission resulting in a reduction of Recall. Several combinations of the omission of delta variables resulted in either very little change or slight increases in Recall. It was indicated that some of these variables were redundant to the prediction capabilities of the data set.

With these results, final variable selection was completed through the omission of larger combinations of delta variables while maintaining enough to keep trends in the data. The total number of input variables was reduced from 29 to 11. The final model data set included *Q*, *T*, *P*, *AFDD*, *ΔAFDD5*, *ΔAFDD10*, *ΔAFDD15*, *ΔQ1*, *ΔQ4*, *ΔQ7*, and *ΔQ10*. The variable selection process eliminated several redundant variables and greatly simplified the input.

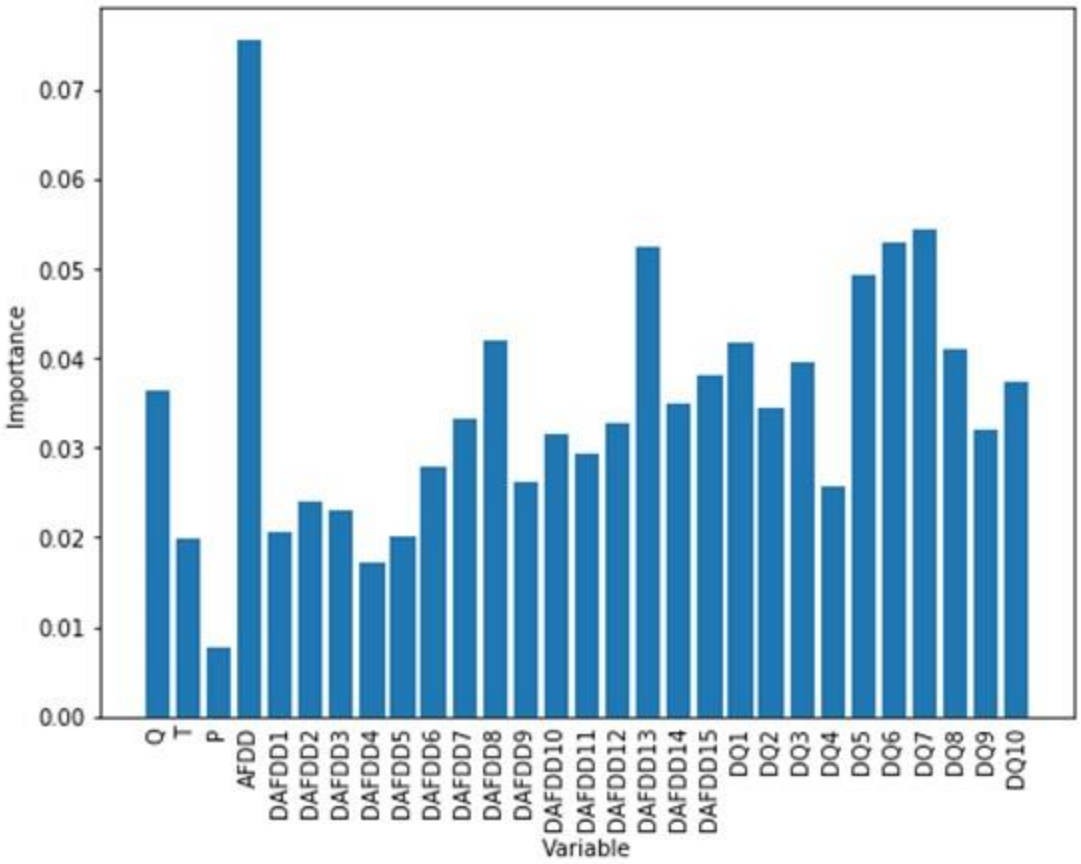


Figure 2.4: Random Forest variable importance plot using all considered input variables.

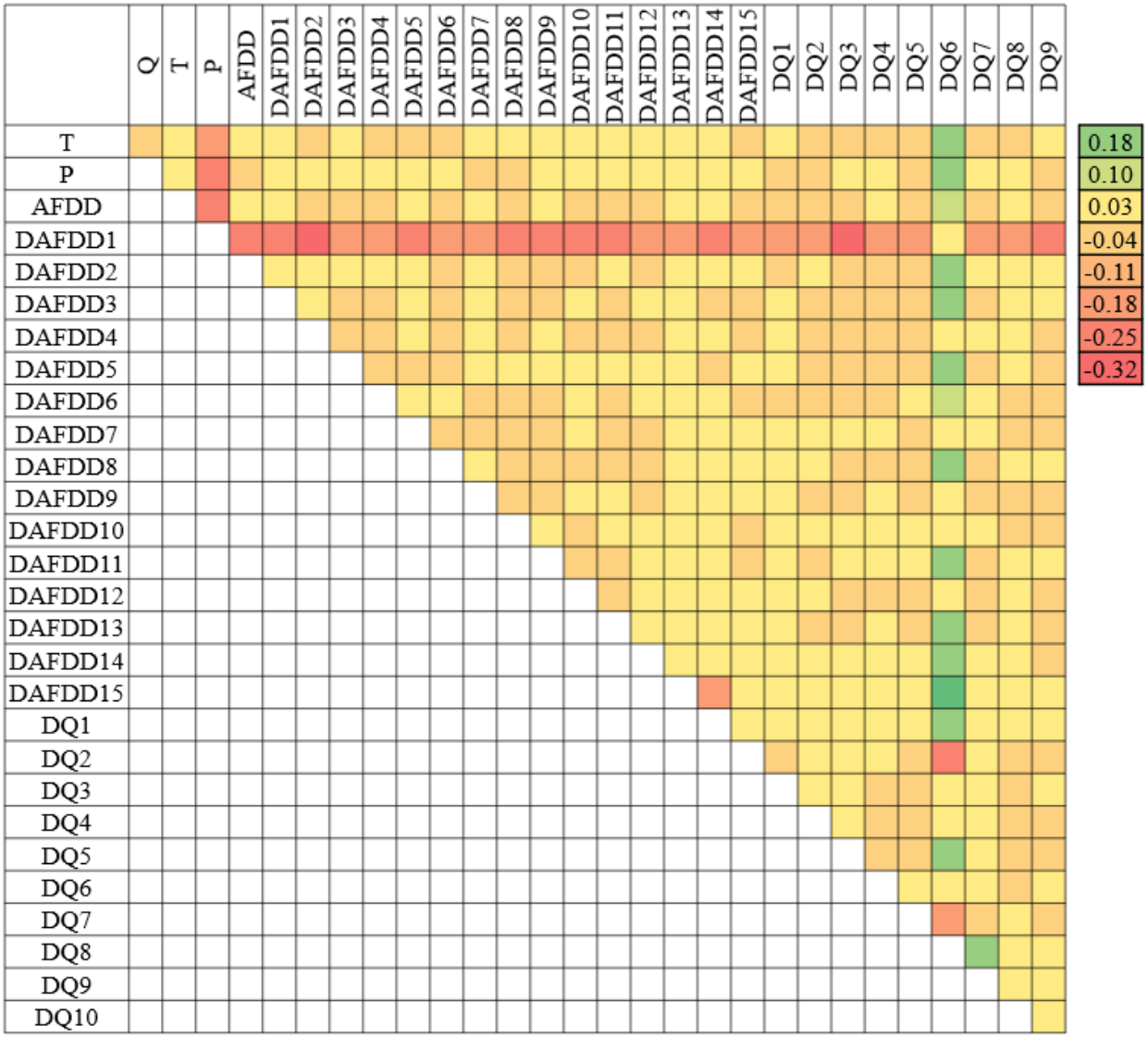


Figure 2.5: Changes in recall from the IO of single and pairs of variables.

2.4.2 Member Model Selection

With the finalized dataset, thirteen models were tuned using 5-fold grid search cross-validation, with the training metric being Recall. The tuned parameters were then used to train and validate the models using a 70/30 data split. Table 2.3 details the accuracy for each of the final models ranked by balanced accuracy. The obtained balanced accuracies ranged from 0.65 to 0.79, with an average of 0.73. The best performance was obtained from the CS model. LR, KNN, and

CS achieved the highest Recall values, with specificity over 0.87 amongst all of the models. CS and ArcX4 attained a specificity of 1.0, indicating perfect classification of non-ice jam events.

Table 2.3: Obtained member model classification rates and accuracies.

Model	TN	FP	FN	TP	Specificity	Recall	Balanced Accuracy
KNN	122	4	5	8	0.97	0.62	**0.79**
CS	126	0	6	7	1.00	0.54	**0.77**
ANN	125	1	6	7	0.99	0.54	**0.77**
SVM	123	3	6	7	0.98	0.54	**0.76**
CT	122	4	6	7	0.97	0.54	**0.75**
VadaBoost	119	7	6	7	0.94	0.54	**0.74**
LR	109	17	5	8	0.87	0.62	**0.74**
ArcX4	126	0	7	6	1.00	0.46	**0.73**
RF	125	1	8	5	0.99	0.38	**0.69**
XGBoost	125	1	8	5	0.99	0.38	**0.69**
LogitBoost	125	1	8	5	0.99	0.38	**0.69**
GBoost	124	2	8	5	0.98	0.38	**0.68**
AdaBoost	124	2	9	4	0.98	0.31	**0.65**

Note: inputs of these models are the variables *Q*, *T*, *P*, *AFDD*, *ΔAFDD5*, *ΔAFDD10*, *ΔAFDD15*, *ΔQ1*, *ΔQ4*, *ΔQ7*, and *ΔQ10*.

Member model selection continued with the assessment of variable importance using a tuned random forest model. Figure 2.6 shows the variable importance plot including each of the 13 considered member models. Several models (KNN, RF, AdaBoost, XGBoost, VadaBoost, CS, and ArcX4) ranked similarly highly, with importance exceeding 0.10, while two models (LR and LogitBoost) ranked extremely low. As these importance values are assessed using overall accuracy, assessment of changes in Recall caused by member omission from a CT was also conducted. Figure 2.7 shows the changes in Recall caused by single model omission and by pairs omission. All of the non-ensemble techniques resulted in decreases in Recall with their omission, while many ensemble models resulted in either minor changes or increases to Recall with their omission, with the exception of AdaBoost and XGBoost.

With the above results as the basis, the combination of six member models that produced the best model performance in terms of both Recall and balanced accuracy were selected for the first level of the hybrid ensemble, including four non-ensemble algorithms (CT, LR, KNN, and ANN) and two ensemble algorithms (VadaBoost and ArcX4). Using the predictions of these models as the second level input, the 13 combiner models were tuned using a 5-fold cross-validation and trained on the split data to produce the hybrid ensemble results listed in Table 2.4. The balanced accuracy ranged between 0.73 to 0.86, with an average of 0.79, which shows an overall improvement in prediction accuracy compared to those obtained from the single models. Specificity was very close to 1.0 amongst all models, while Recall ranged up to 0.77, which is an additional improvement from the single model results. Many of the boosting algorithms performed better as ensemble combiners than they did as single models. The single model algorithms of CT, KNN, and SVM performed well as both single models and as ensemble combiners. The hybrid ensembles were also compared to a voting ensemble technique, which counts the predictions from each model and assigns a value based on the quantities of "votes". The voting model was tuned to predict the occurrence of ice jams based on a number of votes for jams exceeding 2, and produced a balanced accuracy of 0.72, which was exceeded by all of the considered hybrid ensembles.

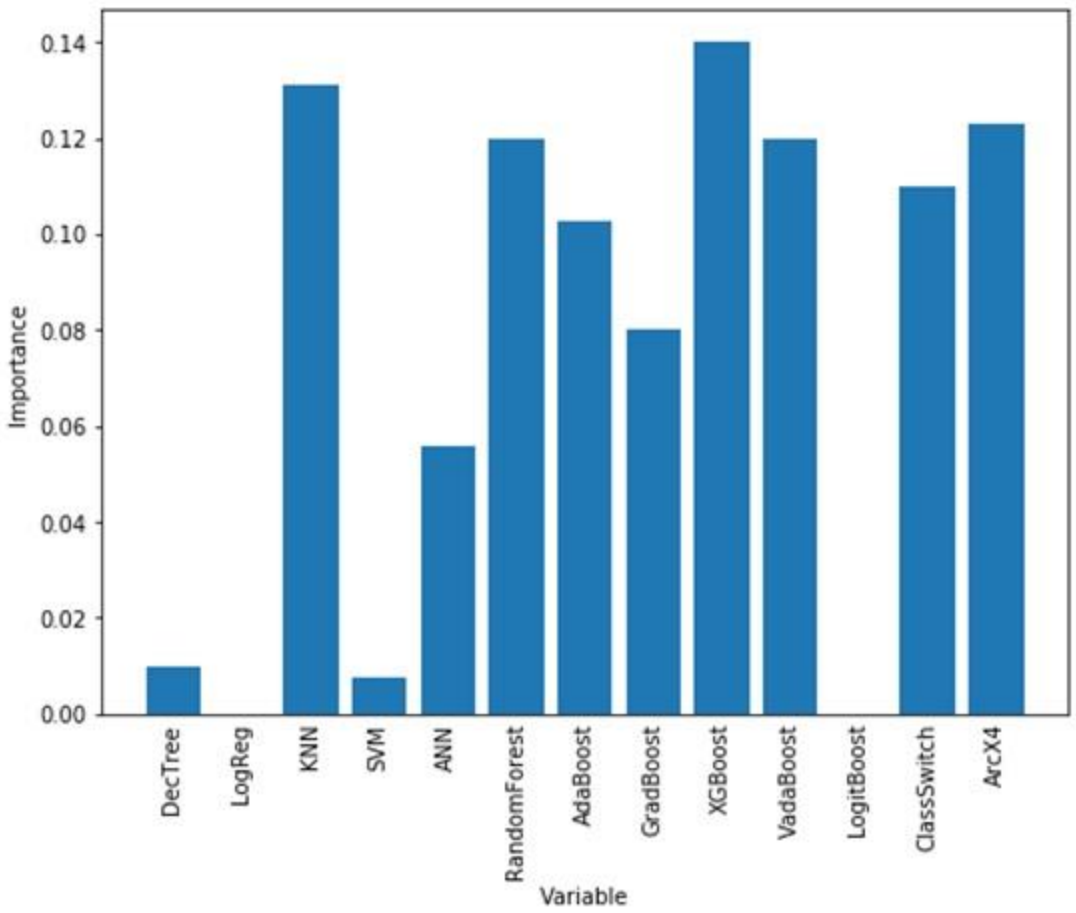


Figure 2.6: Random Forest variable importance plot using all considered member models in a hybrid ensemble.

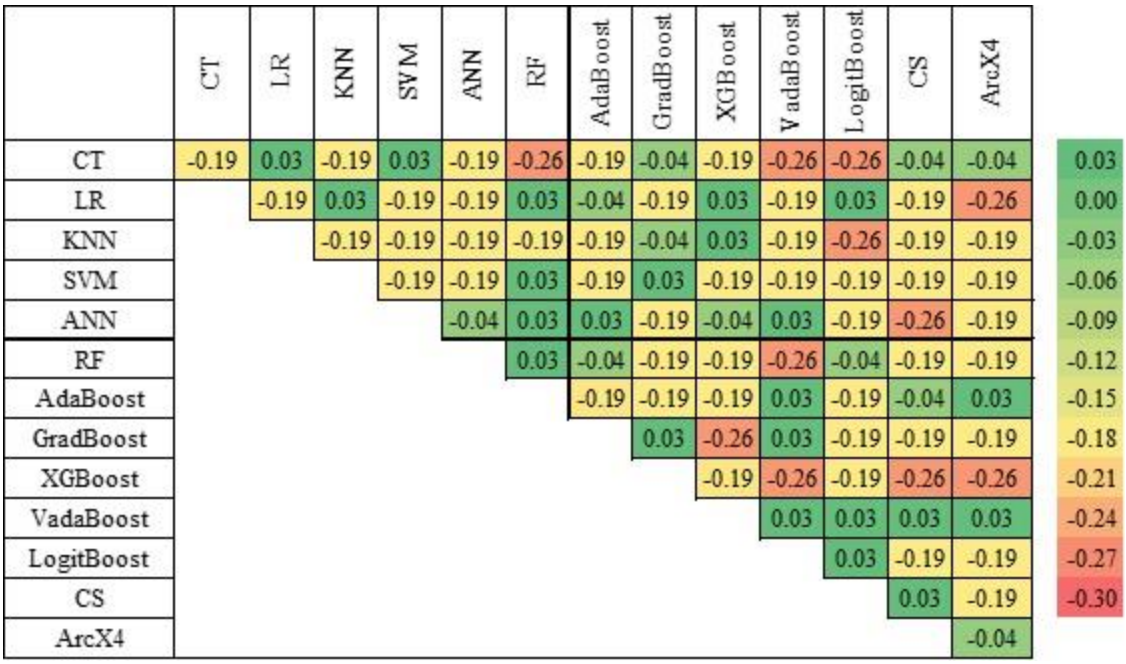


Figure 2.7: Changes in recall from the IO of single and pairs of member models.

Table 2.4: Obtained ensemble combiner model classification rates and accuracies.

Model	TN	FP	FN	TP	Specificity	Recall	Balanced Accuracy
SVM	121	5	3	10	0.96	0.77	**0.86**
KNN	126	0	5	8	1.00	0.62	**0.81**
LR	125	1	5	8	0.99	0.62	**0.80**
CT	122	4	5	8	0.97	0.62	**0.79**
RF	122	4	5	8	0.97	0.62	**0.79**
AdaBoost	122	4	5	8	0.97	0.62	**0.79**
GBoost	122	4	5	8	0.97	0.62	**0.79**
VadaBoost	122	4	5	8	0.97	0.62	**0.79**
LogitBoost	122	4	5	8	0.97	0.62	**0.79**
ArcX4	122	4	5	8	0.97	0.62	**0.79**
CS	122	4	5	8	0.97	0.62	**0.79**
ANN	125	1	6	7	0.99	0.54	**0.77**
XGBoost	125	1	7	6	0.99	0.46	**0.73**
Voting	124	2	7	6	0.98	0.46	**0.72**

Note: Input models of these combiners are CT, LR, KNN, and ANN, VadaBoost and ArcX4.

2.4.3 Best Hybrid Ensemble Model

The best prediction accuracy of 0.86 was obtained from SVM. SVM notably outperformed the next best models, which were grouped around a balanced accuracy of 0.8. The structure of the best model is detailed in Figure 2.8. Each of the model components were tuned using 5-fold grid search cross-validation for the selection of hyperparameters. The CT model used a maximum number of leaf nodes of 20, a minimum number of samples per leaf of 1, and a minimum number of samples per split of 2 in training. The LR model used a value of 0.001 for the inverse of regularization, the KNN a value of 1 for *k*, and the ANN 20 nodes in the hidden layer and a sigmoid activation function in model training. The VadaBoost model was trained using 750 trees with a lambda value of 0.1 and the ArcX4 model using 200 trees. The SVM combiner used a value of 0.1 for the inverse of regularization and a single degree sigmoid function in training.

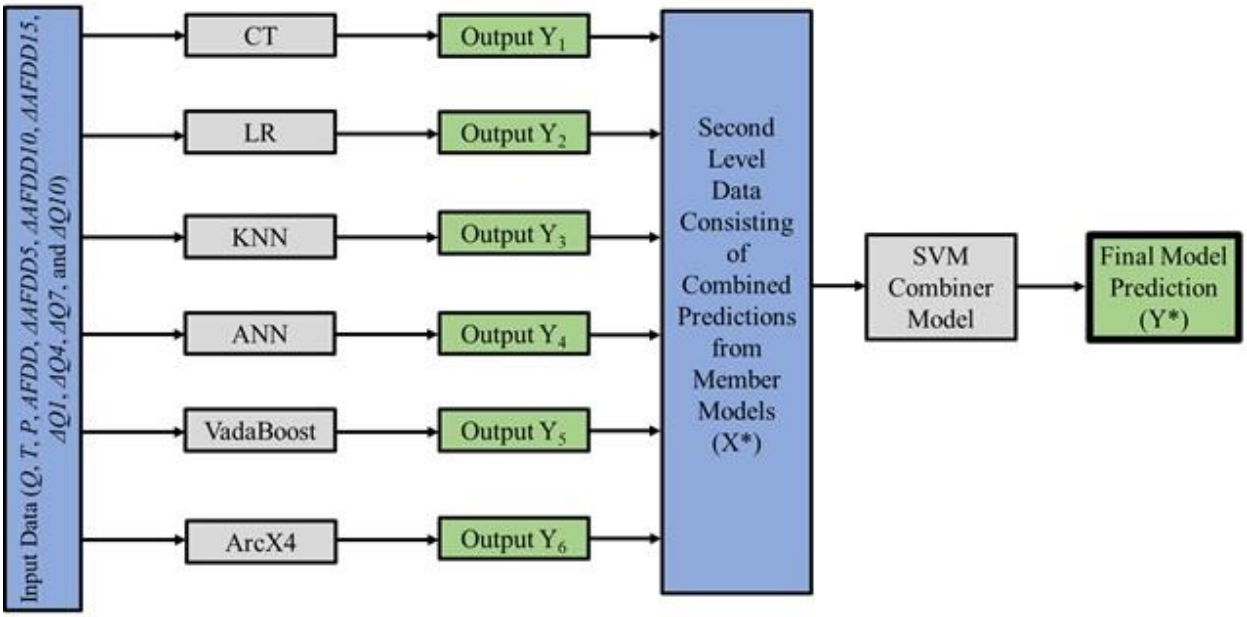


Figure 2.8: Structure of best hybrid ensemble model.

2.4.4 Discussion

To assess the individual prediction accuracy of the final ensemble models, the predictions the models obtained using 5-fold cross-validation using each of the models were plotted in a timeline. Figure 2.9 shows the timelines obtained from each of the tuned ensembles as well as the voting model. It was found that there is no single day where every model is incorrect, and that the incorrect predictions varied from model to model. False negatives were largely grouped at the start and end of ice jam events where jams were present for more than one day consecutively. Inaccuracies in the prediction of the end of these events are less severe, as a jam would already be in place and thus the risk would be known, but the inaccuracies in the start of events are a more severe issue. In this study the incorrect prediction of the start of events appears less common in the model results. False positives appear to be more randomly scattered throughout the timeline, with the only observable trend being their more common occurrence in the earlier years of the data rather than the more recent years. As many of these false positives occur after the presence of an ice jam in the timeline, this suggests that the conditions both climatically and hydraulically would be conducive to ice jamming in the river were the ice cover still present.

Figure 2.10 shows a timeline of the percentage of correct predictions from each of the considered models for every day in the dataset. The majority of poorly predicted days included the presence of ice jams, while many of the days without ice jams that had incorrect classifications still had high levels of correct predictions, typically close to a value of 1.0. Overall there were 20 days with incorrect predictions. The incorrect predictions appear to be grouped more heavily in the early years of record, with 2012 to 2017 having only 7 days that included incorrect classifications, two of which were less than 0.5.

Additional analysis was conducted through the creation of generalized models containing all of the considered variables and member models. The accuracies of the member models using all 29 input variables are listed in Table 2.5, ranked by balanced accuracy. The obtained balanced accuracies ranged from 0.65 to 0.76, with an average of 0.73, very similar to the values obtained from the simplified input set. Specificity was over 0.94 for the models with the exception of the LR model, which produced a value of 0.72. Recall ranged from 0.31 to 0.46, again with the exception of LR, which produced a value of 0.77. Using all of the models as member models, the resulting accuracies for the hybrid ensemble models are listed in Table 2.6, ranked by balanced accuracy. Balanced accuracies ranged from 0.68 to 0.77, with an average of 0.74. Specificities approaching 1.0 were obtained for all models and Recall values ranging from 0.38 to 0.62. The use of the hybrid ensemble approach resulted in improvements in the modelling results for several of the algorithms, similar to the improvement observed in the models developed with variable and member selection. A voting model, again tuned to more than 2 votes for ice jam classification, was also developed. This model obtained a final balanced accuracy of 0.82, outperforming the other generalized models, as well as several models using the curated variables and ensemble members. The overall results of the generalized model tests demonstrate that case-specific tuning is not always necessary if time and resources act as a constraint. A generalized approach to the selection of inputs and ensemble members produces results with accuracies that are not significantly lower due to the versatility of the hybrid ensemble technique.

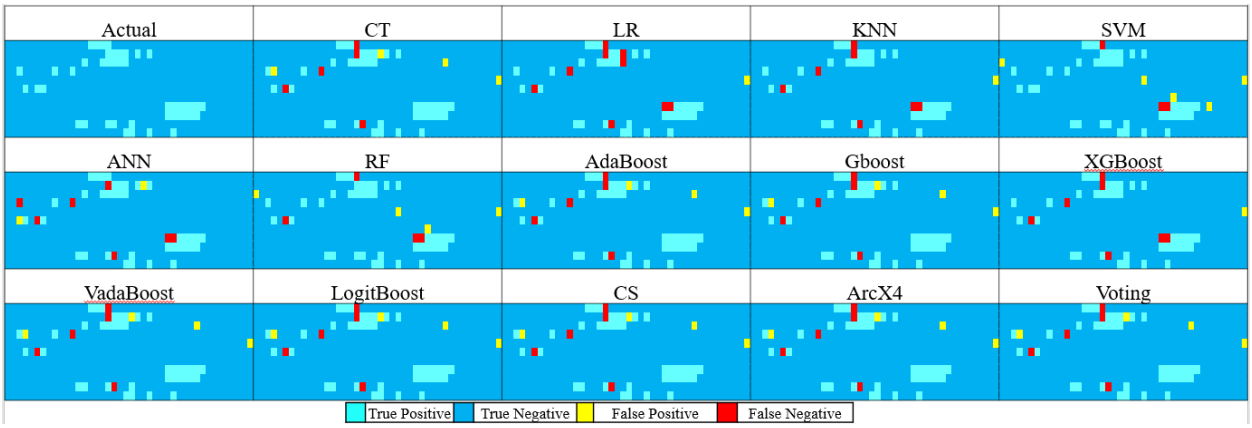


Figure 2.9: Timeline of ice event predictions from each ensemble combiner model.

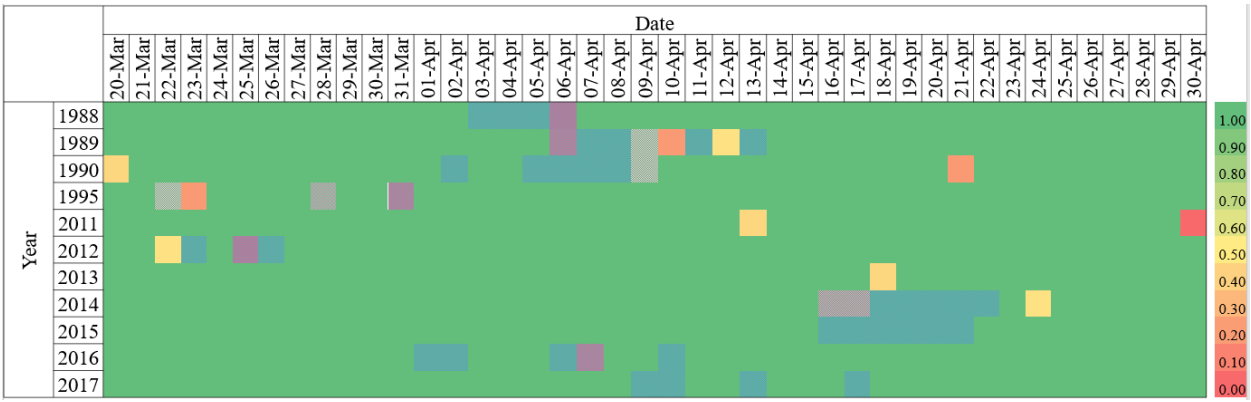


Figure 2.10: Timeline of proportion of correct model predictions from all ensemble combiners.

Dates where ice jams historically occurred shaded in blue.

Table 2.5: Obtained member model classification rates and accuracies using all 29 input variables.

Model	TN	FP	FN	TP	Specificity	Recall	Balanced Accuracy
CS	124	2	6	7	0.99	0.54	**0.76**
SVM	124	2	6	7	0.98	0.54	**0.76**
VadaBoost	124	2	6	7	0.98	0.54	**0.76**
GradBoost	123	3	6	7	0.98	0.54	**0.76**
CT	121	5	6	7	0.96	0.54	**0.75**
LR	91	35	3	10	0.72	0.77	**0.75**
ArcX4	125	1	7	6	0.99	0.46	**0.73**
LogitBoost	124	2	7	6	0.98	0.46	**0.72**
ANN	122	4	7	6	0.97	0.46	**0.71**
KNN	118	8	7	6	0.94	0.46	**0.70**
XGBoost	125	1	8	5	0.99	0.38	**0.69**
AdaBoost	124	2	8	5	0.98	0.38	**0.68**
RF	126	0	9	4	1.00	0.31	**0.65**

Table 2.6: Obtained ensemble combiner model classification rates and accuracies using all 13 member models.

Model	TN	FP	FN	TP	Specificity	Recall	Balanced Accuracy
VadaBoost	117	9	5	8	0.93	0.62	**0.77**
AdaBoost	117	9	5	8	0.93	0.62	**0.77**
ArcX4	126	0	6	7	1.00	0.54	**0.77**
CS	126	0	6	7	1.00	0.54	**0.77**
CT	125	1	6	7	0.99	0.54	**0.77**
LogitBoost	125	1	6	7	0.99	0.54	**0.77**
ANN	122	4	6	7	0.97	0.54	**0.75**
KNN	126	0	7	6	1.00	0.46	**0.73**
SVM	126	0	7	6	1.00	0.46	**0.73**
RF	126	0	7	6	1.00	0.46	**0.73**
GradBoost	126	0	7	6	1.00	0.46	**0.73**
LR	126	0	8	5	1.00	0.38	**0.69**
XGBoost	124	2	8	5	0.98	0.38	**0.68**
Voting (>2)	119	7	4	9	0.94	0.69	**0.82**

2.5 Conclusions

A hybrid ensemble framework that leverages a number of machine learning techniques was developed for the prediction of river ice jams with a one-day lead time, marking the first application of these techniques in the field of ice jam forecasting. An in-depth variable selection procedure was also introduced to produce a more accurate prediction. The input data was arranged such that there was a realistic proportion between the amount of ice jam and non-ice jam events used in model training and testing. Variables used in the predictions are accessible in advance for predictions and provide indications of the trends in the progression of warming and breakup within a river. The tested models provide a variety of algorithms, including both single model structures and ensemble structures, allowing for potential weaknesses in some algorithms to be offset by others. Models were each trained using cross-validation to select the optimal configuration. The variable selection methodology was also extended to the selection of ensemble members, allowing for the optimal combination of members to be determined for the final model. This hybrid ensemble model framework produces a robust model structure with high prediction accuracy that can both reduce forecasting error and provide correction for individual model biases. The data requirements are both flexible and simple, allowing for applicability to a variety of locations with varying data availability, which is a major limiting factor in ice jam forecasting.

This methodology was applied to the Saint John River in New Brunswick, Canada, for the prediction of ice jams in the typical breakup season over the course of 11 years of record. The most successful model combination was obtained using an SVM combined with 6 member models, obtaining a balanced accuracy in predictions of 86%. While this model was the most successful, several other hybrid ensemble models that were developed obtained results approaching the obtained values. This model also outperformed generalized models developed using all available

variables and member models, as well as voting ensemble models developed using both the curated and generalized datasets. In both the generalized and curated cases, the use of the hybrid ensemble technique improved balanced accuracies beyond what was obtained with single models. The hybrid ensemble technique is viable to improve modelling accuracy for both situations where time and resources allow for detailed variable and model selection and where that isn't possible.

These results demonstrated the viability of the hybrid ensemble model for the forecasting of ice jams in the spring breakup season. Many combinations of ensemble members and second level combiners produced similarly strong results, while the utilised modelling dataset provided a more accurate representation of the typical ice jam season without the need for reducing the number of non-ice jam observations to improve results. This model can work to provide decision making support for the St. John River in the ice jam season, and the applied methodology is adaptable to other rivers facing similar ice jam related problems. Within Canada, all data required to build models using the proposed framework is easily obtained from resources provided by organizations such as Environment Canada, such as local climate gauge stations, river flow gauges, and the Canadian River Ice Database. Similar data sources are available in many other northern countries facing the risk of ice jams. With these resources combined with observations of the timing of historical ice jams, this framework can be easily extended to many other rivers, not only in Canada but abroad.

Acknowledgements

This study was supported by Natural Sciences and Engineering Research Council of Canada (NSERC). Data is available from the New Brunswick Department of Environment and Local Government (https://www2.gnb.ca/content/gnb/en/departments/elg.html) and the Water Survey of Canada (https://wateroffice.ec.gc.ca/).

References

Agatonovic-Kustrin, S., and Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of Pharmaceutical and Biomedical Analysis, 22:717-727.

Barzegar, R., Ghasri, M., Qi, Z., Quilty, J., and Adamowski, J., 2019. Using bootstrap ELM and LSSVM models to estimate river ice thickness in the Mackenzie River Basin in the Northwest Territories, Canada. Journal of Hydrology, 577.

Bengio, Y., and Grandvalet, Y., 2004. No unbiased estimator of the variance of K-fold cross-validation. Journal of Machine Learning Research. 5:1089-1105.

Breiman, L., 1996. Bias, variance, and arcing classifiers. Technical Report, 460:1-25.

Brodersen, K., Ong, C., Stephan, K., and Buhmann, J., 2010. The balanced accuracy and its posterior distribution. 20th International Conference on Pattern Recognition.

Bui, D., Tsangaratos, P., Ngo, P., Pham, T., and Pham, B., 2019. Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. Science of the Total Environment 668: 1038-1054.

Cai, Y., Feng, K., Lu, W., and Chou, K., 2006. Using LogitBoost classifier to predict protein structural classes. Journal of Theoretical Biology, 238: 172-176.

Chen, T., and Guestrin, C., 2016. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.Dudani, S., 1975. The distance-weighted *k*-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, 6(4): 325-327.

Environment and Local Government, 2013. Historic ice jams in the Saint John River Basin.

Friedman, J., 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis, 38:367-378.

Gulli, A., and Pal, S., 2017. Deep learning with Keras. Packt Publishing Limited.

Guo, X., Wang, T., Fu, H., Guo, Y., and Li, J., 2018. Ice-jam forecasting during river breakup based on neural network theory. Journal of Cold Regions Engineering, 32(3): 04018010.

Knack, I., and Shen, H.T., 2018. A numerical model study on Saint John River ice breakup. Canadian Journal of Civil Engineering. 45:817-826.

Li, D., Chen, H., and Shi, Q., 2018. Learning from small datasets containing nominal attributes. Neurocomputing, 291: 226-236.

Liaw, A., and Wiener, M., 2002. Classification and regression by randomForest. R News, 2:18-22.

Liu, M., Huang, Y., Li, Z., Tong, B., Liu, Z., Sun, M., Jiang, F., and Zhang, H., 2020. The applicability of LSTM-KNN model for real-time flood forecasting in different climate zones in China. Water, 12:440.

Mahabir, C., Hicks, F., and Fayek, A.R., 2006. Neuro-fuzzy river ice breakup forecasting system. Cold Regions Science and Technology. 46:100-112.

Massie, D., White, K., and Daly, S., 2002. Application of neural networks to predict ice jam occurrence. Cold Regions Science and Technology. 35:115-122.

McKinney, W., and others, 2010. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 445: 51-56.

Munoz, P., Orellana-Alvear, J., Willems, P., and Celleri, R., 2018. Flash-flood forecasting in an Andean Mountain catchment – development of a step-wise methodology based on the random forest algorithm. Water, 10:1519.

Narassiguin, A., Bibimoune, M., Elghazel, H., and Aussem, A., 2016. An extensive comparison of ensemble learning methods for binary classification. Pattern Analysis Applications, 19:1093-1128.

Oliphant, T., 2006. A guide to NumPy (Vol. 1). Trelgol Publishing USA.

Papacharalampus, G., Tyralis, H., Langousis, A., Jayawardena, A., Sivakumar, B., Mamassis, N., Montanari, A., and Koutsoyiannis, D., 2019. Probabilistic hydrological post-processing at scale: why and how to apply machine-learning quantile regression analysis. Water, 11:2126.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and others, 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Safavian, S., and Landgrebe, D., 1991. A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 21(3): 660-674.

Semenova, N., Sazonov, A., Krylenko, I., and Frolova, N., 2020. Use of classification algorithms for the ice jams forecasting problem. E3S Web of Conferences, 163.

Shaikhina, T., and Khovanova, N., 2017. Handling limited datasets with neural networks in medical applications: A small-data approach. Artificial Intelligence in Medicine, 75: 51-63.

Shivaswamy, P., and Jebara, T., 2011. Variance penalizing AdaBoost. 24.

Singh, R., Archfield, S., and Wagener, T., 2014. Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments – a comparative hydrology approach. Journal of Hydrology, 517:985-996.

Snieder, E., Shakir, R., and Khan, U., 2020. A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. Journal of Hydrology, 583: 124299.

Sun, W., and Trevor, B., 2015. A comparison of fuzzy logic models for breakup forecasting of the Athabasca River. CRIPE 18th Workshop on the Hydraulics of Ice Covered Rivers, Quebec City, QC, Canada, August 18-20, 2015.

Sun, W., and Trevor, B., 2017. Combining k-nearest-neighbor models for annual peak breakup flow forecasting. Cold Regions Science and Technology, 143: 59-69.

Sun, W., 2018. River ice breakup timing prediction through stacking multi-type model trees. Science of the Total Environment, 644:1190-1200.

Sun, W., and Trevor, B., 2018a. A stacking ensemble learning framework for annual river ice breakup dates. Journal of Hydrology. 561:636-650.

Sun, W., and Trevor, B., 2018b. Multiple model combination methods for annual maximum water level prediction during river ice breakup. Hydrological Processes, 32:421-435.

Szelag, B., Suligowski R., Studzinski, J., and De Paola, F., 2020. Application of logistic regression to simulate the influence of rainfall genesis on storm overflow operations: a probabilistic approach. Hydrology and Earth Sciences, 24: 595-614.

Suykens, J., and Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Processing Letters, 9:293-300.

Tsakiri, K., Marsellos, A., and Kapetanakis, S., 2018. Artificial neural network and multiple linear regression for flood prediction in Mohawk River, New York. Water, 10:1158.

Wang, J., Sui, J., Guo, L., Karney, B.W., and Jupner, R., 2010. Forecast of water level and ice jam thickness using the back propagation neural network and support vector machine methods. International Journal of Environmental Sciences and Technologies, 7(2):215-224.

White, K., 1996. Predicting breakup ice jams using logistic regression. Journal of Cold Regions Engineering, 10(4):178-189.

Wu, L., Peng, Y., Fan, J., and Wang, Y., 2019. Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. Hydrology Research, 50.6:1730-1750.

Van Rossum, G., and Drake, F., 2009. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Zhao, L., 2012. River ice breakup forecasting using artificial neural networks and fuzzy logic systems. MSc thesis, Department of Civil Engineering, University of Alberta, Edmonton, AB.

Zhu, J., Zou, H., Rosset, S., and Hastie, T., 2009. Multi-class AdaBoost. Statistics and its Interface, 2:349-360.

Chapter 3: Assessing and Predicting the Severity of Mid-Winter Breakups Based on Canada-Wide River Ice Data

The prediction of MWB severity has been hampered by a lack of data in many studies. This chapter presents a new analysis of the severity of MWBs in Canada on a national scale, combining the Canadian River Ice Database (CRID) and the National Resources Canada (NRCan) gridded climate dataset. Multiple input selection techniques are utilised to identify the primary drivers of MWBs on a national scale, successfully selecting the key variables linked to the severity of the events. A random forest model developed using the identified inputs successfully classified the severity of the events, and a new threshold for the triggering of the events selected based on the identified variables was found to be significantly more accurate than previously proposed thresholds.

Michael De Coste was responsible for investigation, model development and testing, and results validation under the guidance of Dr. Zhong Li and Dr. Yonas Dibike, who also supplied the data used in the study. The manuscript was drafted and prepared by Michael De Coste and revised by Michael De Coste, Dr. Zhong Li, and Dr. Yonas Dibike.

This chapter has been published: Michael De Coste, Zhong Li, and Yonas Dibike. (2022) Assessing and predicting the severity of mid-winter breakups based on Canada-wide river ice data. *Journal of Hydrology,* 127550. (DOI: 10.1016/j.jhydrol.2022.127550)

Abstract

Mid-winter breakups (MWBs), consisting of the early breakup of the winter river ice cover before the typical spring breakup season, are becoming increasingly common events in cold region rivers. These events can lead to potentially severe flooding, while also altering the expected spring flow regime, yet data on these events is limited. In this study, a newly released Canadian River Ice Database (CRID), containing river ice data from 196 rivers across Canada obtained from time series analysis, was used to analyse these MWBs on a previously impossible national scale. The CRID data was combined with the Natural Resources Canada (NRCan) gridded daily climate dataset to identify a list of potential hydrologic and climatic drivers for MWB events. Techniques such as correlation analysis, Least Absolute Selection Shrinkage Operator (LASSO) regression, and input omission were combined to select 20 key drivers of the severity of MWB events. A random forest model that was trained with these drivers using data-driven modelling techniques successfully classified the MWBs as either low, medium, or high severity, achieving an overall accuracy of 80%. A new threshold for the prediction of MWB initiation based on climatic conditions was subsequently proposed through the use of optimization via an exhaustive grid search and its accuracy in identifying MWBs exceeded those proposed by previous studies. The new threshold used in conjunction with the random forest model provide valuable tools for both the prediction of MWBs and the assessment of their potential severity.

Keywords: River ice; breakup; flooding; prediction; data-driven modelling; time series analysis

3.1 Introduction

River ice formation and breakups are a yearly phenomena on many Canadian rivers and are often source of the highest annual flows. In recent years, there has been an increasing risk of Mid-Winter Breakups (MWBs), which are the early breakup of a river ice cover before winter ends (Beltaos et al., 2003). These typically result from unseasonably high temperatures and are followed by temperatures returning to typical seasonal values and the reformation of an ice cover. In addition to exposing the river to higher than usual flows more than once per season, they also carry the risk of producing ice jams, where broken river ice becomes restricted in the channel, leading to a damming effect on flows (Massie et al, 2002, Beltaos 2002, Rokaya et al., 2018). There is the potential for these jams to freeze in place when temperatures return to seasonal levels, creating a semi-permanent dam on the river (Beltaos 2003). Further, these events alter the subsequent spring breakup on the river, greatly affecting the expected flow regime by reducing expected flows which can have impacts on the hydrology and ecosystem of downstream basins (Beltaos et al., 2006). Many studies have noted the increasing frequency of these events throughout North America (Prowse et al., 2002, Huntington et al., 2003, Carr and Vuyovich, 2014). Because of this, it is of great interest to develop a better understanding of the key factors, both at the start of the ice season and before melt occurs, that can trigger a MWB and their effect on the severity of the resulting MWB.

Several past studies have applied large scale data analysis techniques to the spring breakup of river ice cover. For example, de Rham et al., (2008a) investigated annual high water events in the Mackenzie River Basin, investigating the spatial distribution of open water and breakup dominated high flows. They concluded that high water events in the southern portion of the basin were dominated by open water flows while those in the north were dominated by spring breakup flows.

This was further investigated through a spatiotemporal analysis of the last ice affected dates of 29 Water Survey of Canada (WSC) gauges in the basin (de Rham et al., 2008b). Through the extraction of the breakup and peak water levels and subsequent trend analysis it was found that the timing of breakup is trending earlier throughout the basin, with longer breakup periods expected. Goulding et al., (2009a) examined the drivers of breakup and ice jamming in the Mackenzie Delta through time series inspection. Trend analysis using Mann-Kendall and Sen's slope revealed that the peak stage in the Delta was most influenced by upstream discharge while timing of breakup was most influenced by ice conditions such as thickness and the speed of melt, and that breakups in the Delta were trending towards longer melts, lower peaks, and earlier timing. Further research incorporating spatial mapping and analysis of both ice driven and open water driven peak flows reinforced these conclusions (Goulding et al., 2009b). While these papers investigated the drivers and trends in breakup, they were focussed largely on single regions within the country and investigated only spring breakup and open water events.

Past studies on the phenomena of MWBs have looked at these events in a generalized sense, emphasizing their increasing frequency, but due to the lack of readily available data were unable to do any detailed modeling. Beltaos (1999) provided an overview of the effects of climate change on the Saint John Rivers ice breakup regime. It was noted that there was an increase in mild winter days over the past 80 years, with a trend of spring breakup occurring earlier that was tied to an increase of mid winter breakup events carrying a risk of ice jams freezing in place and flooding. Beltaos and Prowse (2009) investigated the impacts of climate change on the timing and severity of ice events, including breakup and ice jamming. Their findings indicated a trend towards shorter ice seasons and earlier breakups with the potential for mid winter ice jams increasing as a result. Further, it was found that the changes to breakups occurring both in mid winter and in spring would

be driven by changes in warming rates and snowpack depths, with potentially severe environmental impacts both to habitats and hydrology.

Research has also been conducted on MWB severity and drivers with varying levels of success. Prowse et al., (2002) investigated the drivers of MWBs, focussing on the effects of temperature in triggering these events. They defined a climatic threshold of 25 melting degree days (MDD), the sum of temperatures over 0°C over 7 days as a major trigger for breakups. Similarly, Carr and Vuyovich (2014) also defined a threshold for the trigger of an MWB. Their threshold considered temperature as well as precipitation, with values of 8 MDD or 2.8 mm of rain over a period of 5 days. Both of these thresholds were investigated by Newton et al., (2017) in their study of the hydro-climatic drivers of MWBs in western Canada. In this study, a database compiled from gauges on 90 unregulated rivers was used to study the influence of daily temperature and precipitation on MWBs in 5 different climate regions, with a larger "temperate" region defined as the area where solid, sufficiently thick ice cover is expected to form while still being susceptible to mid-winter warm spells. In total, 52 MWBs were identified from 21 rivers in the considered region. The climate conditions preceding these MWBs were then compared to the thresholds proposed by Prowse et al., (2002) and Carr and Vuyovich (2014), finding that the first threshold was inaccurate, only accurately predicting 11% of the MWBs, while the second was more accurate, predicting 60% of all MWBs. While the results of this research were promising, it was limited both by the quantity of data s and in the scale of the considered region, only including the western coast of the country.

Machine learning techniques have been previously applied to model and predict a variety of river ice phenomena. White (1996), Massie et al., (2002), and Mahabir et al., (2006) utilised logistic regression, artificial neural networks, and hybrid neuro-fuzzy systems for the prediction

of ice jam timing on rivers, however they encountered issues related to the availability of data. Spring breakup has also been modeled using techniques such as artificial neural networks (Zhao, 2012, Guo et al., 2018), adaptive neuro-fuzzy inference systems (Wang et al., 2012, Sun and Trevor, 2015, Sun and Trevor, 2018), stacking ensembles (Sun, 2018), and support vector machines (Wang et al., 2010, Barzegar et al., 2019). These studies often focus on predicting the water levels or ice thicknesses associated with these events, however some have focused on the specific timing of spring breakup, with data availability often being a limiting factor in the results.

This study focuses on the application of return period analysis, statistical and spatial analysis, and machine learning techniques to investigate the primary drivers of MWBs and predict the Severity of MWBs on a national scale. The study was facilitated through the use of a newly developed river ice dataset containing data on 458 MWBs occurring throughout the country. Through this investigation, **a generalised set of key indicators of MWB occurrence are identified and their linkages to MWB severity are examined**. A data driven model capable of accurate prediction of MWB severity and a new threshold providing highly accurate indication of MWB occurrence were both developed, outperforming results of past studies with similar goals. **This is the first attempt to assess and predict the severity of MWBs in Canada at a national scale. It allowed a greater understanding of drivers of MWBs, and a clearer relationship between the drivers and severity of the MWB events.**

3.2 Data

3.2.1 The Canadian River Ice Database (CRID)

One of the primary data sources for this study, the Canadian River Ice Database (CRID) contains an extensive amount of data on river ice throughout the country extracted from WSC gauge records by de Rham et al., (2020) using time series analysis. The extraction process involved

visual inspection of water level time series records according to a process laid out by Beltaos (1990), requiring detailed criteria for each river ice event to be identified before the timeline of each river ice season could be constructed. The definition of each event was primarily based on the time series of water levels, with peaks and lows occurring throughout the winter season acting as indicators of distinct events. The CRID contains data from 196 gauges in Canada, with 46 on regulated rivers and 150 on unregulated rivers. For each season, up to 15 river ice related events are identified and the corresponding water levels, discharges and event timings are recorded. One of the key events recorded is the occurrence of MWBs, identified as peaks in the flow record associated with higher-than-average winter temperatures at nearby climate stations, historically a difficult task (Beltaos 1990). The instantaneous MWB water level represents the onset of ice cover movement at a site during the winter season and is identified as a spike on the rising limb of the water level record (de Rham et al., 2020). For each MWB event, the event date and associated discharge are recorded along with the water level. Additional water level, flow, and ice thickness data were also recorded for each event if available. To demonstrate potential applications of the data, an initial trend analysis was conducted for the calculated interval representing ice cover duration, the span between the date of freeze-up and the date of break-up. Their primary findings through this analysis were that ice-duration was decreasing and breakups were occurring earlier throughout the country, with only some gauges in the southwest experiencing longer seasons (de Rham et al., 2019).

The 196 gauges of the CRID are shown in Figure 3.1, with the gauges containing historical MWBs highlighted with the number of MWB events. There are a total of 52 gauges with MWB events mainly concentrated in two major regions of the country: the southwest (in BC and Alberta) and the southeast (in Ontario, Quebec, New Brunswick, and Newfoundland and Labrador). The

locations of these gauges are largely outside of the colder northern portions of the country and are more frequently affected by mid winter thaws. Temporal coverage varies from gauge to gauge depending on period of operation, with the earliest MWB occurring in 1955 and 301 of the MWB station years occurring after 1990. 38 of the gauges are located on unregulated rivers, while 14 are on regulated rivers containing structures altering the flow regime in the rivers. Data on 452 MWB events are contained in the CRID amongst the 52 stations where they have been observed. Outliers in the time water level or discharge time series stemming from gauge repositioning or measurement error were identified and corrected after consultation with Environment and Climate Change Canada. Historical notes were used to identify the dates of gauge datum changes and the corresponding datum measurements, allowing corrections to be made. In total 7 gauges had portions of their record adjusted, including 28 station years with MWBs. The CRID data that was utilised in this study for the prediction of MWBs is listed in Table 3.1. This data consists of all ice cover related variables preceding the occurrence of MWB events.

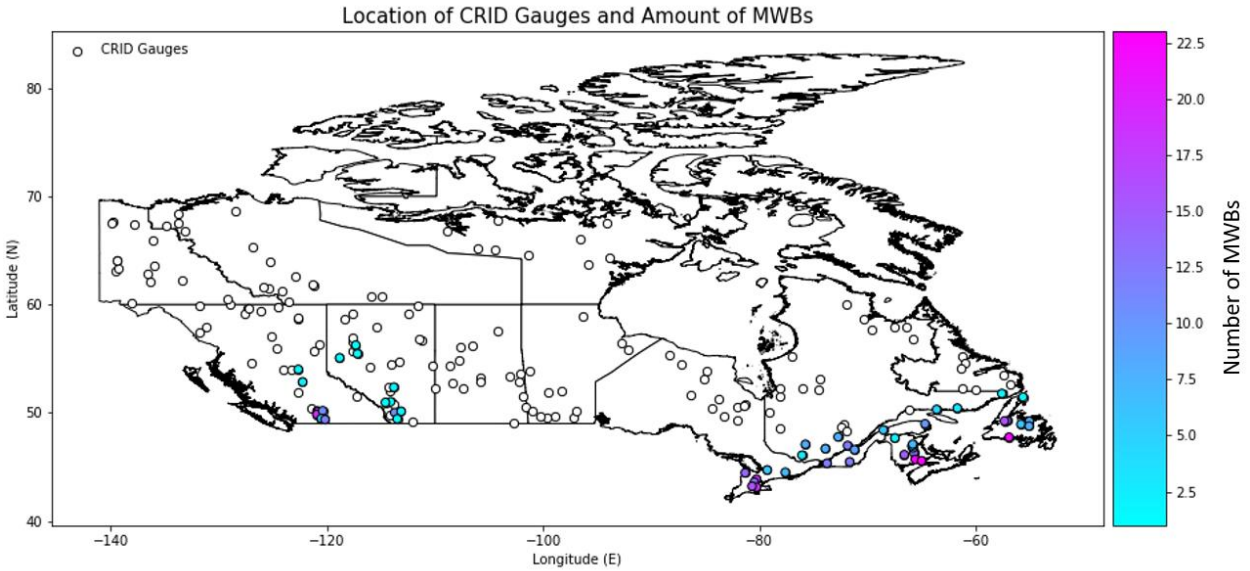


Figure 3.1: Locations of gauges included in the CRID. Number of MWBs at affected gauges is given by color scale.

Table 3.1: List of events and associated variables in the CRID used for prediction of MWBs.

Event	**Variables**
First Day with Ice Backwater	Date
Freeze-Over	Date, Time, Water Level, Flow, Water Levels of next 30 days
First Minimum Winter Water Level	Date, Water Level, Flow
First Minimum Winter Flow	Date, Water Level, Flow
MWB Initiation	Date, Time, Water Level, Flow
Maximum MWB Water Level	Date, Time, Water Level, Flow

3.2.2 NRCan Gridded Climate Data

The primary source of climate data for this research was the Natural Resources Canada (NRCan) Daily Gridded Climate set, providing 0.1-degree gridded data across the country spanning 1950-2010. Variables included in the initial data are daily minimum temperature, daily

maximum temperature, and daily precipitation. This data was derived from quality-controlled, but unadjusted (without homogenization or procedural changes), station data from the National Climate Data Archive (NCDA) of Environment and Climate Change Canada data (Hutchinson et al., 2009) and interpolated onto a high-resolution grid using thin plate splines. (McKenney et al., 2011, Hopkinson et al, 2011).

Geographic information system mapping was used to extract daily minimum and maximum air temperature and daily precipitation data from the nearest NRCan grid point at each of the gauges in the CRID where MWBs have occurred. The minimum and maximum temperature values were used to calculate a single daily mean temperature value for each day in the time series (McMaster and Wilhelm, 1997). These values were used to compile a series of climate variables for attribution analysis of MWBs. Temperature values were also used to calculate a daily value for Accumulated Freezing Degree Days (AFDD), the sum of temperatures below 0°C beginning from a reference date, in this case October 1st, the beginning of the Canadian water year (Boyd, D., 1979). Over the course of the winter these values would represent the magnitude of how cold the season had been. The daily total seasonal precipitation cumulatively summed from the beginning of freeze up was also calculated to account for the depth of snowpack over the winter.

3.3 Methodology

3.3.1 Trend Analysis

Trend analysis of MWBs was conducted using the Mann-Kendall (MK) trend test for time series data based on it's previous successes in applications related to river flooding (Du et al., 2015, Goulding et al., 2009a, Hamed and Rao, 1998). MK has been widely used for trend analysis in hydrological studies as it does not require the data it's applied to to be normally distributed

while also not being affected by missing data points. For a time series $X=x_1, x_2, \ldots, x_n$, a statistic S is calculated using Equation 3.1.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i) \tag{3.1}$$

where S is the MK trend statistic, the values of x_i are the data being analysed for trends, and n is the number of observations in the time series. Under the null hypothesis that no trend is present, the statistic S tends towards normality, with variance given by Equation 3.2.

$$var(S) = \frac{1}{18}[n(n-1)(2n+5)] \tag{3.2}$$

The test statistic z for the MK test is described by Equation 3.3.

$$z = \begin{cases} \frac{S-1}{se}, & S > 0 \\ 0, & S = 0 \\ \frac{S+1}{se}, & S < 0 \end{cases} \tag{3.3}$$

where *se* is the square root of *var*. The p-value derived from the area under a normal distribution curve at the value of z will be assessed against a 95% confidence interval for this MK test, where a value less than 0.05 for the p-value would indicate the presence of a trend, allowing the null-hypothesis of no trend to be rejected.

3.3.2 Return Period Analysis

The return period of each MWB was calculated at each of the gauges independently. At each gauge, ice affected flows and water levels were first normalised through the subtraction of the means. A Weibull distribution, traditionally used in the assessment of flood return periods on rivers (Apel et al., 2006, Vervuren et al., 2003), was then fitted to each of the water level and flow

curves, using maximum likelihood estimation to fit parameters (de Rham et al., 2008a). The probability density function (PDF) of the Weibull distribution is presented in Equation 3.4.

$$f(x, c) = c/2|x|^{c-1}exp(|x|^{c}) \tag{3.4}$$

where *c* is the shape parameter fitted for the distribution and is greater than 0. Equation 3.5 is used to calculate the plotting position of return periods of the historic MWBs.

$$R = (n + 1)/m \tag{3.5}$$

where *R* is the return period of the event in years, *n* is the number of years of water level records, and *m* is the magnitude ranking of the event. Stations with less than 30 years of ice season data, including both years with and without MWBS, were excluded from this analysis. In total, 50 of the 52 stations with MWBs had sufficient data to be included in this analysis.

3.3.3 Driver Identification

A series of potential drivers were identified for early indication of MWBs based on the NRCan climate data and the CRID data. To determine the drivers best suited to provide early indications of MWB occurrence, a multi-step variable selection process was employed consisting of correlation analysis, regression analysis, and input omission (Hall and Miller, 2009, Meinshausen and Buhlmann, 2006, Galelli et al., 2014). While there are various other techniques that can be employed including analysis of variable distributions or principal component analysis, this process ensures that each of the possible drivers are considered in the selection process to prevent the risk of excluding vital features while also ensuring that redundant or detrimental drivers are excluded. This combination of multiple selection methods considers each driver individually while also employing a combination of a typical forward selection approach and a modified backward elimination process employing shrinkage, avoiding biased regression coefficients typical of a standard backwards approach (Tibshirani, 1996). Potential issues

stemming from non-linear relationships between drivers are also addressed through the selection of a regression algorithm capable of functioning in non-linear problems in the second step.

The first step of this process involved correlation analysis of the candidate drivers against the three variables representing MWB severity: water level, discharge, and return period corresponding to the maximum MWB water level (H_{MWM}). Pearson's correlation was used as the primary metric for this analysis, described by Equation 3.6 (Benesty et al., 2009).

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{3.6}$$

where $\rho_{X,Y}$ is the Pearson correlation between variables X and Y, $cov(X,Y)$ is the covariance between variables X and Y, and σ_X and σ_Y are the standard deviations of variables X and Y. A Pearson's correlation value approaching 1 would represent an increasingly strong positive correlation, a value approaching -1 would represent an increasingly strong negative correlation, and a value of 0 would represent no correlation between the variables. By conducting this analysis for all drivers to the three MWB severity metrics, early indications of potentially redundant drivers could be identified while drivers with stronger correlations indicate stronger potential as predictors.

The second step of this process involved initial driver selection using Least Absolute Selection Shrinkage Operator (LASSO) regression, regressing the full input set against the three variables representing MWB severity. A LASSO model fits a regression with an L1 penalty on non-intercept coefficients in the regression equation (Tibshirani, 1996). This allows the regression to perform initial variable selection through using a shrinkage factor, detailed in Equation 3.7.

$$\hat{\beta} = argmax_{\beta}\left(\ell(\beta|y_{it}, A) - \lambda\sum_{q=1}^{p}|\beta_q|\right) \quad (3.7)$$

where β, the coefficient or weight for a particular variable, is estimated through maximizing the likelihood function $(\beta|y_{it}, A)$, with y_{it} representing the observed value of the predictand, A representing the pool of potential predictors, λ the tuning parameter, and ℓ the L1 regularization penalty which focuses on reducing the weights of each variable, often driving small weights to zero. The tuning parameter λ was selected through 5-fold cross validation on the data, subdividing the data into five portions and building a model based on four then testing the model against the fifth. Each of the five possible data combinations is tested with each of the possible λ values and the value producing the best performance is selected. The final equation developed resembles that of a multiple linear regression, shown in Equation 8, where the *b* parameters would be replaced by β.Through this, the drivers selected by the LASSO algorithm from the full input set (those with a non-zero β value) for regressing each of the three targets were able to be ranked based on the order in which the regression included them and the value of their final coefficient, providing an indication of their importance (Sermpinis et al., 2018). Drivers not selected for inclusion in the fitted LASSO regressions could be excluded from subsequent analysis.

Following this, the remaining ranked drivers selected from the LASSO analysis were tested using an Input Omission (IO) process against a Multiple Linear Regression (MLR) to the three targets, detailed in Equation 3.8.

$$y = a + b_1x_1 + b_2x_2 + \cdots b_nx_n \quad (3.8)$$

where *a* is the intercept and b_i are the coefficients for each of the predictor variables x_i, predicting the response *y* (Goswami and Brahma, 2019). An MLR equation was constructed using all drivers, and an additional equation was developed for each variable where it was omitted from the predictor

variables. Changes in mean squared error were calculated and drivers whose omission either decreased the value or resulted in no change were removed from the driver pool to develop a final driver selection (Snieder et al., 2020). This two-step process of LASSO and IO functions to reduce the initial candidate driver pool to those that showed the strongest link to the predictand variables. The final driver selection could conceptually be used to identify a new threshold for the initiation of MWBs by identifying the ideal climatic variables to base a threshold around. Identified climatic variables would be run through a grid search process to identify the values that maximize the prediction of MWBs from the full dataset.

3.3.4 Classification Algorithm

Following the final selection of drivers, a data-driven modelling algorithm, random forest, was used to test the effectiveness of the drivers in classifying the severity of MWB return periods. The amount of events in the severity classes would be uneven to ensure the model is capable of capturing the uneven distribution of rare events while also ensuring each severity class was distinct. A random forest model develops an ensemble of classification trees through bootstrap samples of the dataset, as shown in Figure 3.2 (Zhou et al., 2019a and 2019b). Each bootstrap sample set is used to develop an independent classification tree and a separate prediction, and the predictions of all bootstrap sample sets are used to output a final prediction (Liaw and Wiener, 2002). This process is represented in Equation 3.9.

$$\hat{f}\,(x) = \frac{1}{M}\sum_{m=1}^{M} \hat{f}^{m}\,(x) \tag{3.9}$$

where $\hat{f}(x)$ is the overall prediction of the random forest, M is the number of bootstrap ensembles, and $\hat{f}^{m}(x)$ is the prediction trained on the m^{th} bootstrap sample. Overfitting of random forest models is a concern particularly with smaller datasets, thus in some cases it is recommended to use

stopping criterion to limit tree growth to account for this problem. As this model was built for classification, which seldom encounters the overfitting effect as a result of the ensemble structure which also limits error increases due to bias (Segal, 2003, Wu and Zhang, 2016, and Javeed et al., 2019), a default stopping criterion of zero with no pruning was used producing fully grown trees, as recommended by Hastie et al., (2009). As a results, with no stopping criteria to meet, full trees will be grown for each member of the random forest according to the values selected for the model hyperparameters. These values would be selected using 5-fold cross-validation, with a grid search being used to try each possible combination of the considered hyperparameters one at a time (Probst et al., 2018, Albers et al., 2016). The hyperparameters included in model tuning were the number of estimators (number of trees included in the forest), maximum features (the number of features considered when determining the best split), maximum depth (maximum levels of the tree), minimum samples per split (number of samples needed to split a node), and minimum samples per leaf (number of samples needed at a leaf node). A minimized Gini impurity, the measure of an incorrect classification's likelihood, was used as the target value to split each node. A default learning rate of 0.1 and step-size of 10 (Hastie et al., 2009). The final model would be subsequently trained and tested on a stratified and randomized 90/10 data split with variables normalized. The variables of test data would be run through the trained model to obtain a set of predicted severity classes which would then be compared to the actual classes those observations belonged to.

Model accuracy can be assessed using metrics based around the class-by-class rate of true positives (tp_i, the number of correctly classified members of the class), true negatives (tn_i, the number of correctly classified non-members of the class), false positives (fp_i, the number of non-members of the class incorrectly classified into the class), and false negatives (fn_i, the number of

members of the class incorrectly classified as non-members). With these values, the metrics Overall Accuracy (average effectiveness of the classifier per class), average Precision (agreement of the classifier's labels with the true class labels per class), and average Recall (effectiveness of the classifier in identifying class labels) can be calculated using Equations 3.10-3.12. Additionally, the rate of true negatives, or average Specificity can be calculated using Equation 3.13.

$$Overall\ Accuracy = \frac{\sum_{i=1}^{k} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + tn_i}}{k} \tag{3.10}$$

$$Precision = \frac{\sum_{i=1}^{k} \frac{tp_i}{tp_i + fp_i}}{k} \tag{3.11}$$

$$Recall = \frac{\sum_{i=1}^{k} \frac{tp_i}{tp_i + fn_i}}{k} \tag{3.12}$$

$$Specificity = \frac{\sum_{i=1}^{k} \frac{tn_i}{tn_i + fp_i}}{k} \tag{3.13}.$$

where *k* is the total number of classes (Sokolova and Lapalme, 2009). These metrics are commonly used in the assessment of multi-class models (Shahri et al., 2020 and Shahri et al., 2021). By plotting the relationship between the Recall, or true positive rate, and 1 – Specificity, or the false positive rate, a Receiver Operating Characteristic (ROC) curve can be developed. This curve plots the ability of the model to classify each target class, with a curve produced for each of the considered classes by looking at the accuracy of their predictions. The Area Under the Curve (AUC) is a useful indication of model skill, producing a value between 0 and 1.0, with values of 0.5 representing no discriminatory ability, 0.7-0.8 being acceptable, 0.8-0.9, considered excellent, and anything above 0.9 demonstrating outstanding ability (Mandrekar, 2010).

The trained model can also provide an indication of variable selection accuracy through the output of variable importance values. The value of a variables importance in a random forest model indicates the level of contribution it provides in lowering the Gini impurity of the model predictions, based around overall classification rate. This allows an indication of whether any selected variables are further redundant and could potentially be excluded, or if each variable provides an even contribution to the model results. While non-stationarity can be a concern as a result of climate change in these modelling applications, the developed model will be built on a combination of hydrologic and climatic variables, which addresses this potential issue (Li et al., 2016, Ghaith et al., 2020). Random forest models have historically been successful in classifying flood hazard risks (Wang et al., 2005, Sadler et al., 2018).

This final model will be valuable in demonstrating the strength of the selected drivers in predicting the severity of a subsequent MWB in combination with the techniques applied in driver selection. The predictions of the model are based on the assumption that there are relationships present in the data, specifically between the variables and the severity, which are demonstrated through the variable selection process, though not necessarily described in detail. While random forest models have historically encountered issues related to computational time, especially in real-time predictions, these issues are not a concern in this application due to the size of the data. The rareness of MWB events limits the amount of data needed for initial model training, while the variable selection process reduces the amounts of data needed for real time monitoring, preventing issues stemming from computation time and model complexity.

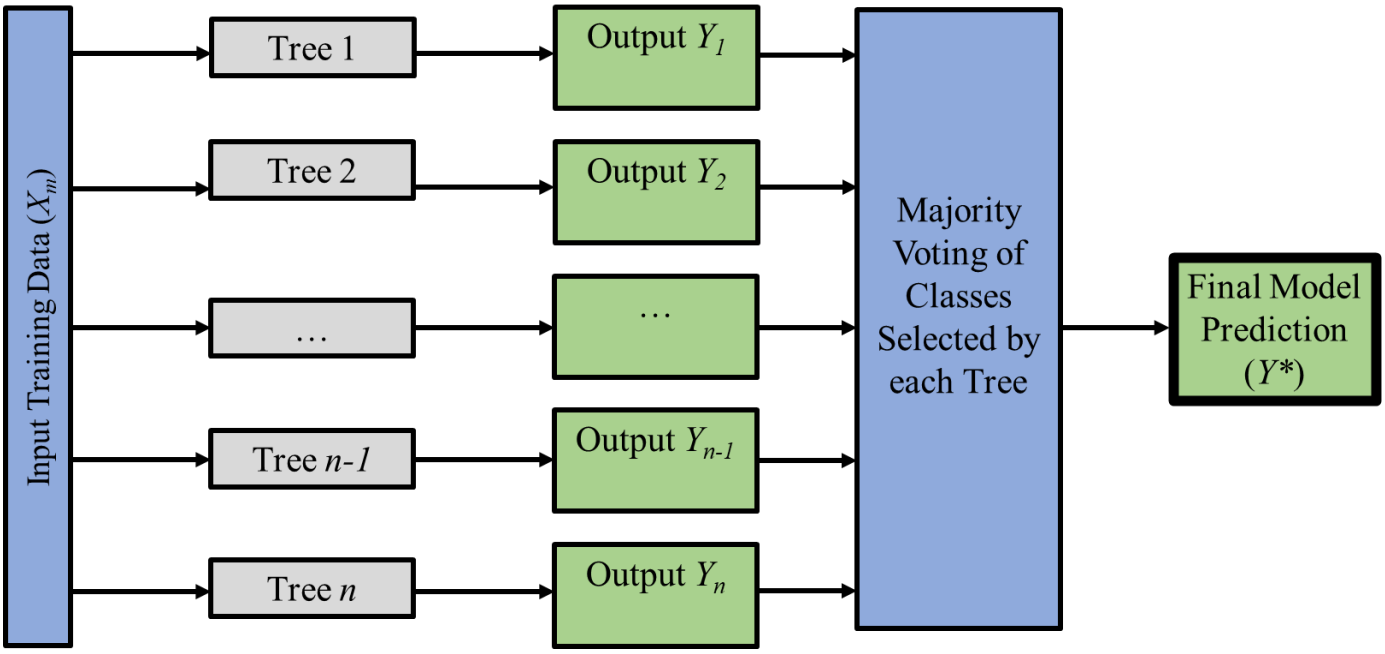


Figure 3.2: Random forest internal model structure.

3.3.5 Algorithm Implementation

Analysis in this study was completed using Python, version 3.7 (Van Rossum and Drake, 2009). Additional packages implemented in these applications included *Numpy* (Oliphant, 2006), *Pandas* (McKinney, 2010), *scikit-learn* (Pedregosa et al., 2011), *scipy* (Virtanen et al., 2020), and *seaborn* (Waskom, 2008).

3.4 Results and Discussion

3.4.1 MWB Trend Analysis

The annual number of MWBs throughout Canada was compiled from the CRID and plotted as a time series in Figure 3.3. This time series was used as the basis of the MK trend test on the annual number of MWBs in the country. This analysis produced an S value of 1061, which provides an indication of an increasing trend, an *se* value of 15.9, a z value of 6.63, and a p-value of 3.47×10^{-11}, which indicates a statistically significant trend at a 95% confidence interval. It can

be concluded based on both this analysis and the visible trend in Figure 3 that there is an increasing trend in the frequency of MWBs in Canada, emphasizing the importance of their analysis. A trend analysis of the MWBs at the individual gauges within the CRID was further conducted. MK testing of the water levels and timing of the events at individual gauges was less conclusive, displaying no statistically significant trend at all but one of the rivers, the exception of which had an increasing trend in water level. Thus, while the frequency of these events is clearly increasing, there is no clear increase or decrease in their severity on a regional scale.

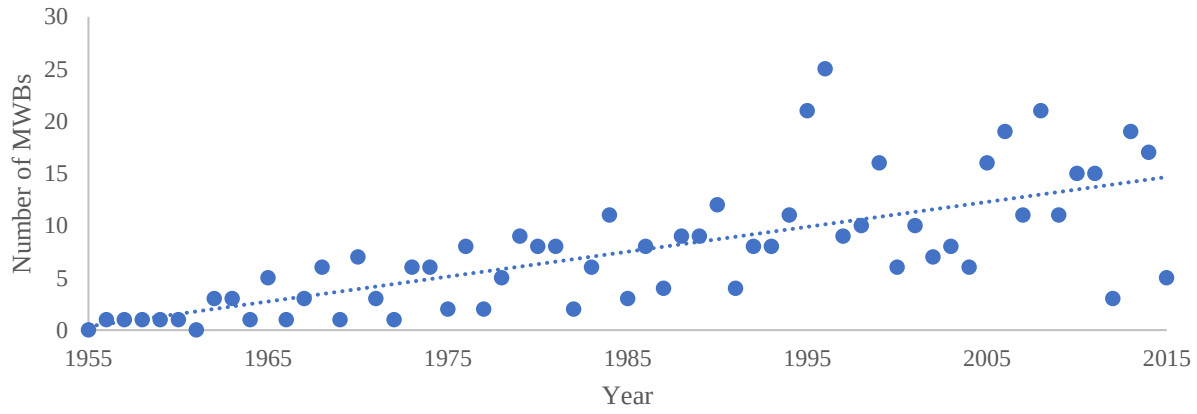


Figure 3.3: Number of MWBs occurring each year in the CRID from 1955-2015 including the trend line.

3.4.2 Return Period Analysis

At each of the gauges with MWB events and sufficient ice season data, a Weibull distribution was fitted to the maximum ice affected water level of each year. The fitted distribution was used to calculate the return period of each maximum ice affected water level for each ice season including the maximum MWB water levels. It was found that a significant portion of the high return period ice-affected water levels in the CRID were caused by MWBs. Figure 3.4A shows the number of MWB water levels with return periods over 5 years plotted at each gauge

location with the average return period being 3.47 years. It was found that more southern locations had a higher number of such events, and that every gauge where a MWB had occurred produced at least one with a return period exceeding 5 years. Figure 3.4B plots the percentage of water levels with a return period of over 5 years caused by MWBs. In the southern portions of BC and Alberta, as well as Southern Ontario and Quebec, the percentage is notably higher, approaching 85% at some gauges. While the MK tests on the water level and flows related to these MWBs indicated that the severity of these events has no significant increasing trend, on average these events are typically much more severe than the regular ice season flows. Because of this, the increasing trend in frequency that was found for these events indicates an increasing likelihood of high water level events related to MWBs on these rivers, which can be much more severe than typical spring breakups.

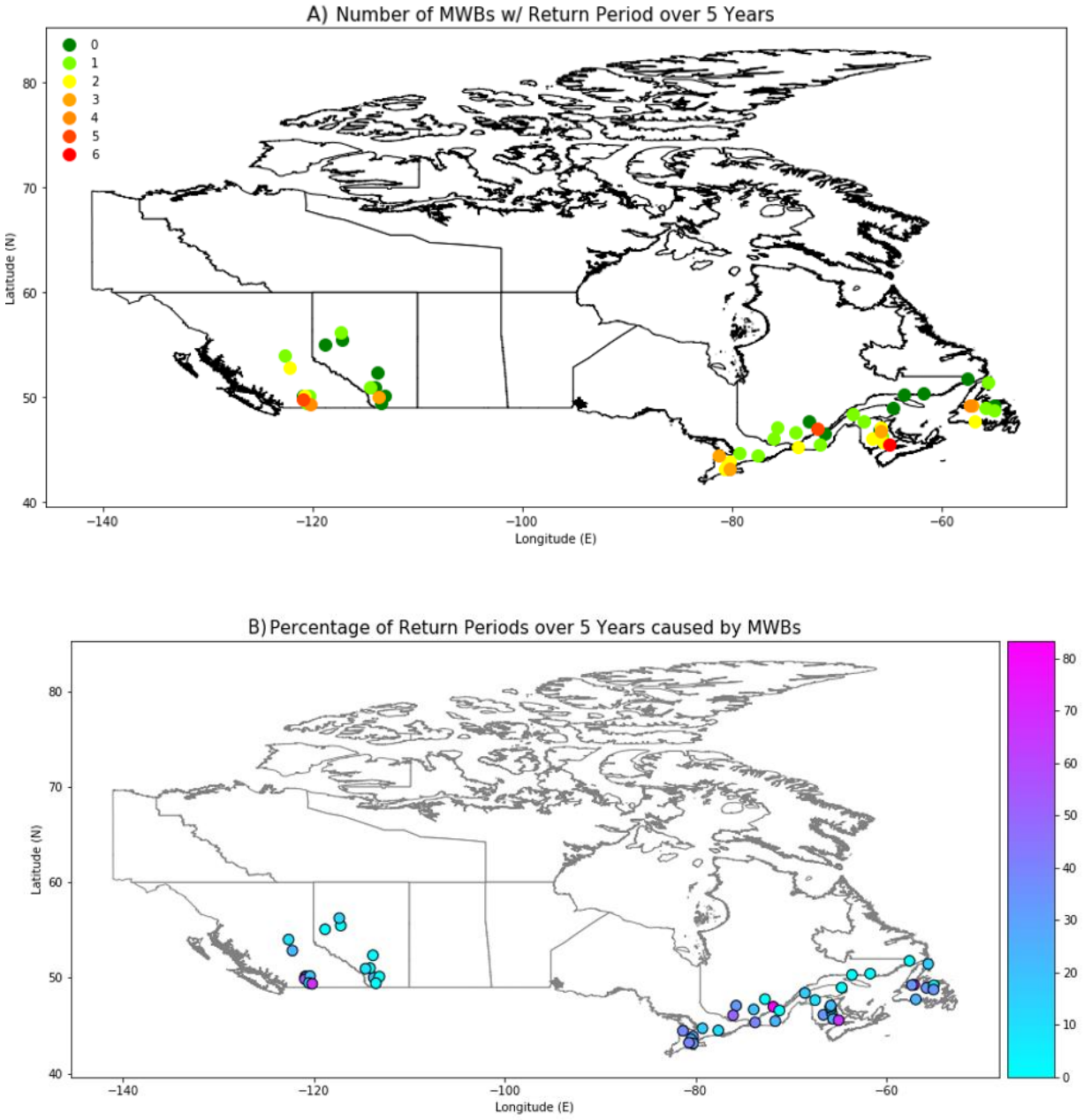


Figure 3.4: A) Number of MWBs where water levels with > 5 year return period and B) percentage of RPs over 5 years caused by MWBs at each gauge where they have historically occurred in the CRID.

3.4.3 Driver Identification

An initial selection of 130 potential indicator variables was assembled for each of the 452 MWBs in the CRID. Each MWB would be treated as a discrete event with the goal of using a selection of the 130 indicator variables to predict the level of MWB severity via classification. Table 3.2 details these variables, consisting of ice-related events preceding MWBs from the CRID and climatic indicators derived from the NRCan climate time series, all of which could provide early indication of ice, snow, or climatic conditions preceding MWBs occurrence. Each variable includes the name they were coded as in analysis. The events included the first ice affected day and the freeze up date, which describe the initial traits of the ice cover, and the dates of the first winter low flow and low water level, providing an indication of the development of both the ice cover and the flow regime in the river preceding an MWB. Dates used in the analysis were calculated as the number of days since the beginning of the reference period, October 1st. Water levels and flows were normalised at each station to prevent any inconsistencies being present in modelling. The AFDD was used as a measure of how cold the ice season had been preceding an MWB, with values included from each of the 30 days preceding the event. The change in AFDD between each of these 30 days and the value measured on the day an MWB occurred were also included, providing an indication of how quickly the warming occurred during the thaw preceding the MWB. Similarly, the total precipitation from the start of the winter season was calculated up to each of the 30 days preceding the MWBs, providing an early indication of how much snow was present before thaw occurred. The change in total precipitation between each of the 30 days preceding the MWB and the date of the MWB was also calculated, providing an indication of the amount of precipitation that fell directly preceding the MWB, potentially as liquid.

Table 3.2: Initial variables considered as potential MWB drivers.

Category	Initial Inputs	Coded Name	Number of Features
Date	First Ice Affected Date, Freeze Up Date, Winter Low Level Date, Winter Low Flow Date	FIRST_B_DATE, HF_DATE, HLW1_DATE, HLQ1_DATE	4
Water Level	Freeze Up Water Level, Winter Low Level Water Level, Winter Low Flow Water Level	HF_EXT_WL, HLW1_EXT_WL, HLQ1_EXT_WL	3
Flow	Freeze Up Flow, Winter Low Level Flow, Winter Low Flow Flow	HF_DISCHARGE, HLW1_DISCHARGE, HLQ1_DISCHARGE	3
Temperature	AFDD from 1-30 days, ΔAFDD from 1-30 days	AFDDi, DAFDDi	60
Precipitation	Total Seasonal Precipitation (P) from 1-30 days, P in last 1-30 days	TotP1, DPi	60

Initial analysis of Pearson correlation was conducted between each of the 130 candidate variables and the three severity variables. Figure 3.5A-E shows heatmap results of this analysis. Stronger correlations were found between ice event variables and precipitation variables to the three severity targets, while the variables related to temperature largely had a weaker correlation to these same targets, though these values varied.

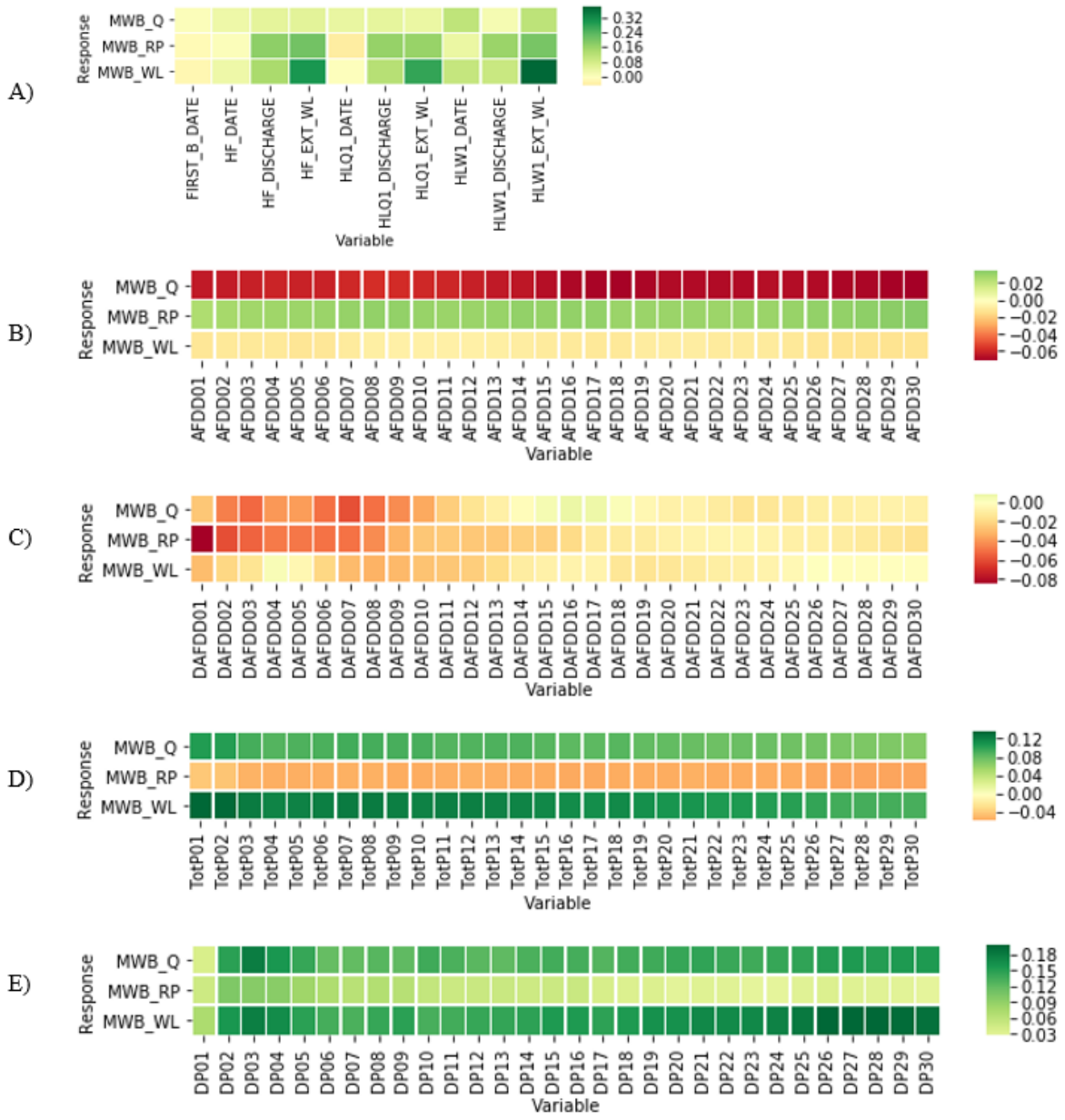


Figure 3.5: Heatmaps of the correlation between the Discharge (MWB_Q), Return Period (MWB_RP) and Water Level (MWB_WL) of each MWB and A) CRID ice event variables, B) AFDD up to day *i* before an MWB (AFDD*i*), C) change in AFDD up to day *i* before an MWB (DAFDD*i*), D) Total Precipitation up to day *i* before an MWB (TotP*i*), and E) change in Total Precipitation up to day *i* before an MWB (DP*i*).

LASSO models were constructed as the next step of variable selection from the 130 initial variables. Figure 3.6 shows an example of the variable implementation of the LASSO regressing a selection of variables to the return period. Models were developed regressing the entire driver pool to the water level, flow, and return period of each MWB. By removing those variables that were not included in the final regression equations (having a β of 0 in Equation 3.7), the initial variable selection was able to be reduced to 76 variables. These remaining variables were ranked based on the order of their implementation in the LASSO regressions and the value of their β coefficient in the final regression equation. Inspection of the rankings allowed an additional 35 variables to be rejected on the basis of their overall low contributions to the regression equations. The final 41 variables were then analysed through an IO process regressing them against the MWB return period using MLR equations. Each variable was removed from the equation in turn, allowing the changes in mean squared error with their absence to be calculated. Variables whose removal resulted in either no change or a decrease in mean squared error would be removed from consideration. Based on the change to mean squared error, a further 21 variables were removed, resulting in the final 20 variables listed in Table 3.3.

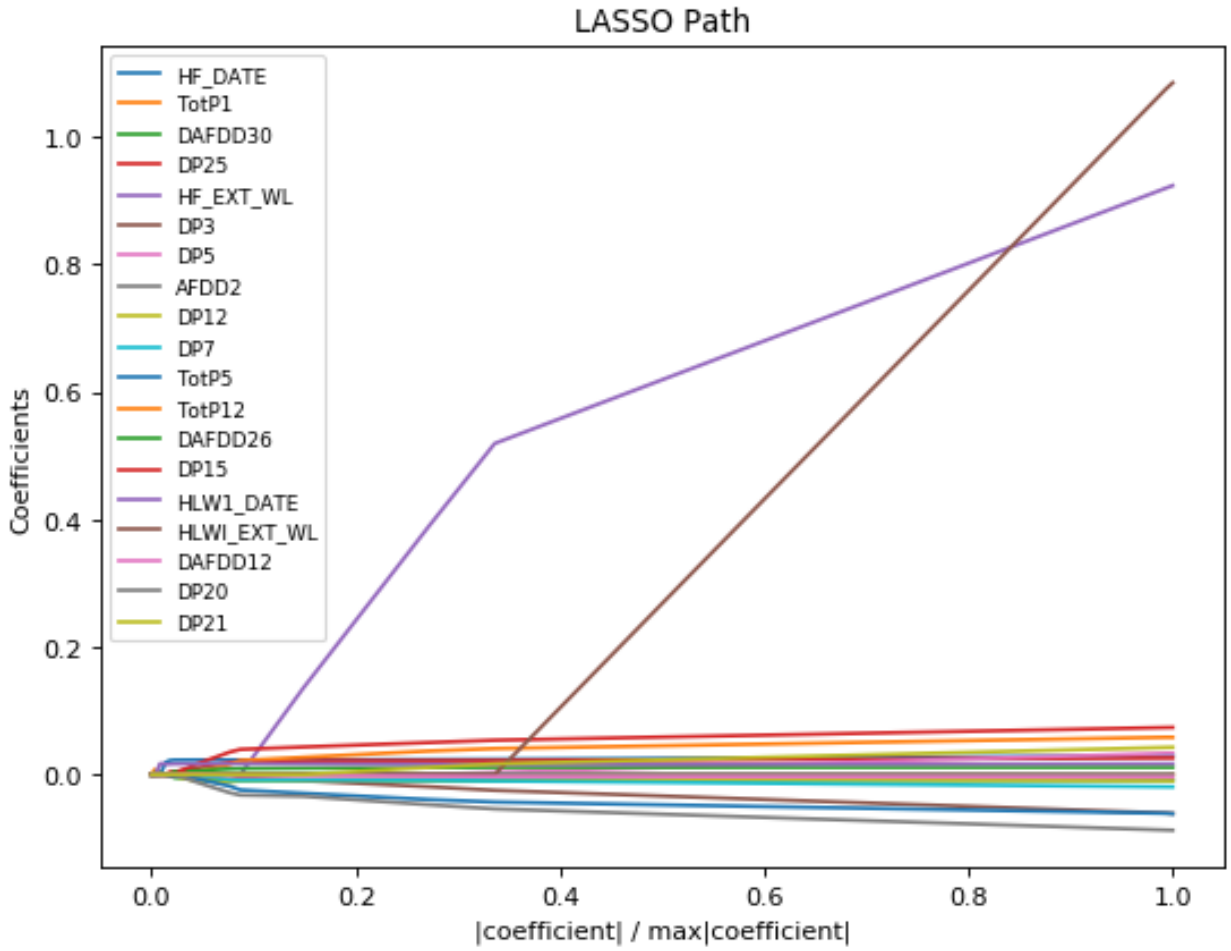


Figure 3.6: LASSO results example regressing a selection of variables against MWB return period, showing the changing values of the coefficients of each variable as more are introduced to the regression.

Table 3.3: Final selection of candidate driver variables after analysis.

Category	Initial Inputs	Number of Features
Date	Freeze Up Date, Winter Low Level Date	2
Water Level	Freeze Up Water Level, Winter Low Level Water Level, Winter Low Flow Water Level	3
Flow	None	0
Temperature	AFDD from 1 day before, ΔAFDD from 12, 26, and 30 days	4
Precipitation	Total P from 1, 5, and 12 days, P in last 3, 5, 7, 12, 15, 20, 21, and 25 days	11

3.4.4 Classification Algorithm

Using the final selection of 20 variables, a random forest model was next trained and tested to classify the severity of MWB event return periods. The return periods of the MWBs considered in the dataset were divided into three categories: low severity (return period less than 2 years), medium severity (return period between 2 and 15 years) and high severity (return period greater than 15 years). Figure 3.7 maps the dominant category at each of the gauges in the CRID included in MWB analysis. After completing a 5-fold cross validation to select model hyperparameters via a grid search (with a selected value of 100 estimators from a range of 100-1000, maximum features of 5, maximum depth of 30 from a range of 10 to 100, minimum samples per split of 2 from a range of 2 to 5, minimum samples per leaf of 3 from a range of 1 to 5, and bootstrap selection) producing the highest accuracy on the training set, the random forest was trained and tested using the 20 variables. The variables of a testing set with 40 events was classified by the model and compared to the actual classes of the data, with the resulting confusion matrix shown in Table 3.4.

The trained model produced an 80% overall accuracy, a precision of 87% and a recall of 71% on the testing data. The ROC curve for the testing data is shown in Figure 8. An AUC of 0.8 was obtained for low severity, 0.735 for medium severity, and 0.824 for high severity, indicating overall good ability of the model to capture the MWB severity. Many of the events were classified correctly, with the majority of incorrect classifications being the prediction of a low severity event as a medium severity event. Figure 3.9 shows the variable importance plot for this model. While the date variables obtained slightly higher values than the others, it can be seen that all of the selected variables provided a roughly equal contribution to the reduction of Gini impurity in the model classification. These results successfully demonstrate the strength of the selected variables in classifying MWBs and their use as early indicators for MWB severity. The classification accuracy of the trained model applied to the whole dataset is mapped by gauge in Figure 3.10. The overall accuracy was 96%, with a precision of 97% and a recall of 88%. An AUC of 0.99 was obtained for low severity, 0.99 for medium severity, and 0.936 for low severity. Nearly all gauges had high classification accuracies, with the exception of one gauge in northern Alberta, which had a 0% classification accuracy. Further analysis of this gauge shows that there was only one MWB in its record, which was considered a high severity event by the criteria described above. Based on the overall results, the high severity events are the hardest to classify due to the small number of them contained in the dataset.

To investigate the structural uncertainty of the final model and the driver selection process, additional models were developed for comparison. The first of these models was constructed using all 130 of the variables considered in Table 3.2. The constructed model was a random forest classifying the return period in the same manner as the curated model detailed above, with training and testing being conducted in the same manner. The resulting model had a 63% overall accuracy,

a recall of 46%, a precision of 67.5%, and an average AUC of 0.65, a decrease in comparison to the curated model. This indicates that the reduced variable set provides a much greater indication of MWB severity through the removal of redundant variables whose linkages to the severity are less significant, less relevant, and misleading (or potentially correlated non-linearly with other variables), while also increasing computational efficiency through the removal of unneeded data. A second model was developed using an alternative method of AFDD calculation from the method described in Section 3.3.1. In this method, AFDD was calculated from the first day of below 0 temperatures rather than the generalized start date of October 1st. As the only AFDD variable present in the final curated dataset was AFDD1, the AFDD up to the day of an MWB, this was the only factor that would change in the final model. DAFDDi variables would remain the same as they are the day-to-day difference in AFDD directly preceding the MWB event. By replacing the AFDD1 variable and retraining the model, an overall accuracy of 72%, recall of 65%, precision of 81%, and an average AUC of 0.72 were obtained, indicating that the generalized method of calculating the AFDD provides a better indicator of MWB severity.

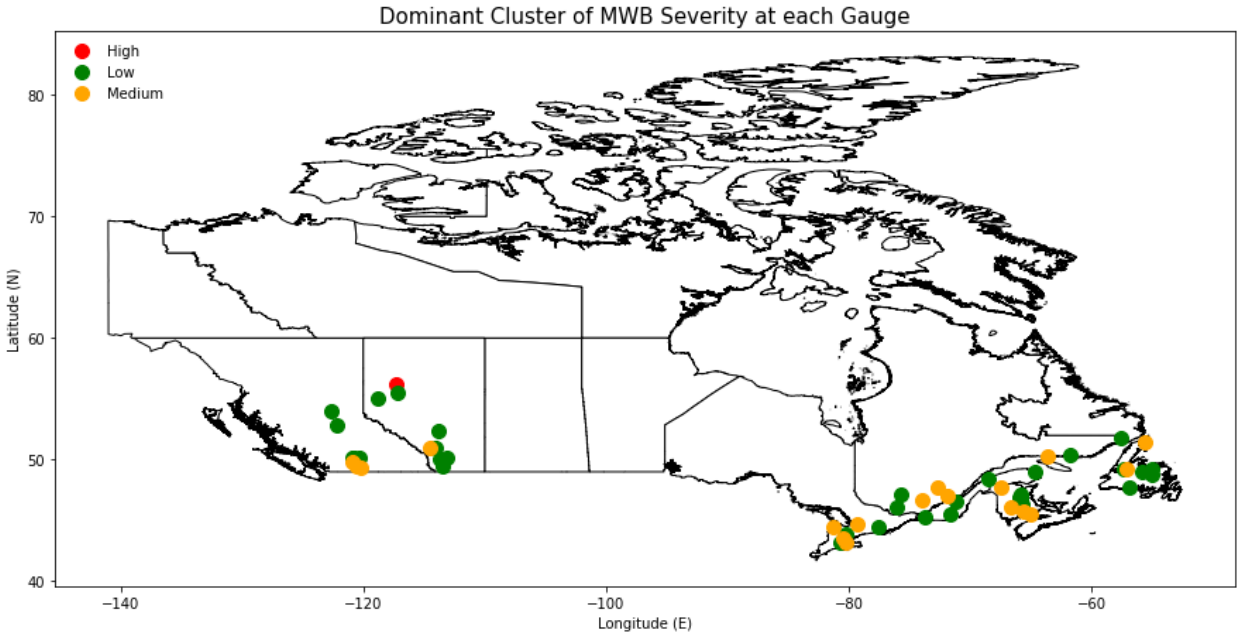


Figure 3.7: Map of the dominant cluster of MWB severity at each gauge in the CRID.

Table 3.4: Random forest model classification confusion matrix.

		Predicted		
		Low	Medium	High
Actual	Low	17	3	0
	Medium	4	14	0
	High	1	0	1

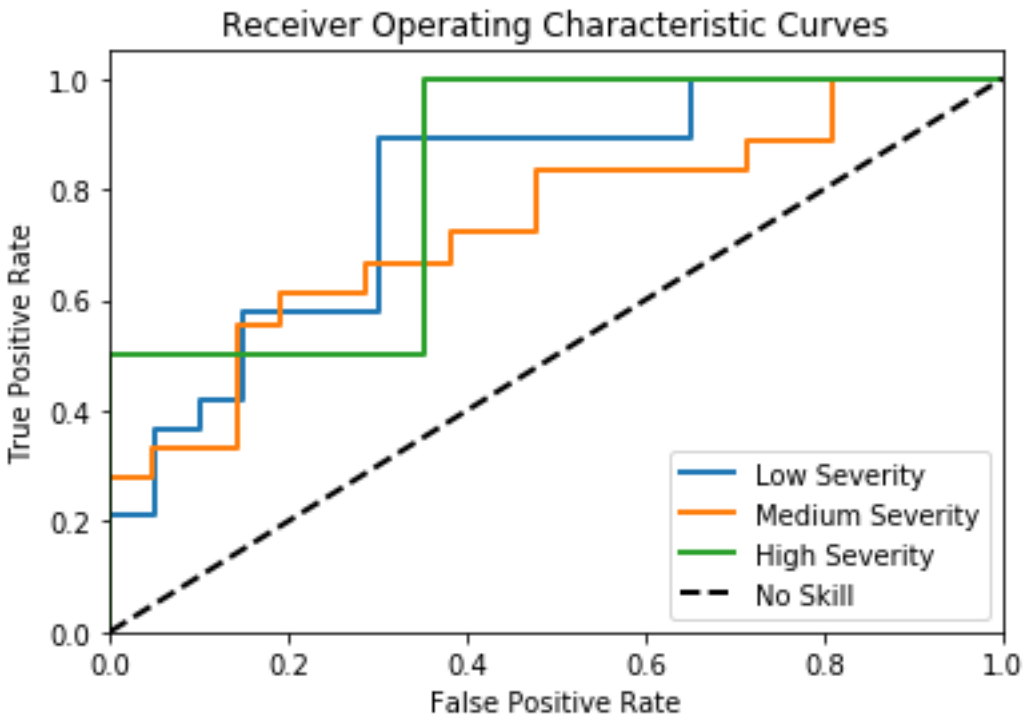


Figure 3.8: ROC curves from the trained random forest classification model for the test data.

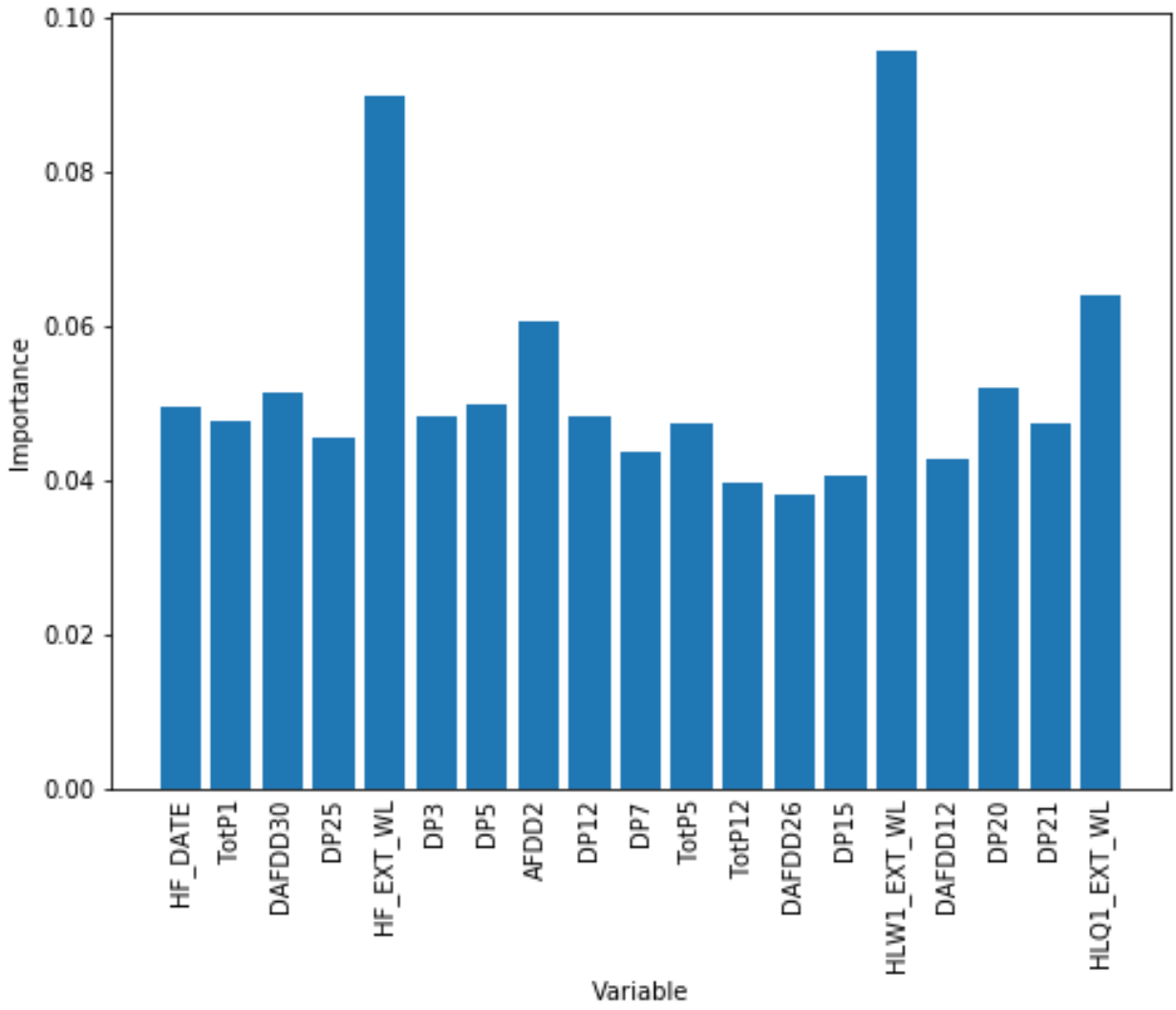


Figure 3.9: Variable importance plot from the trained random forest classification model.

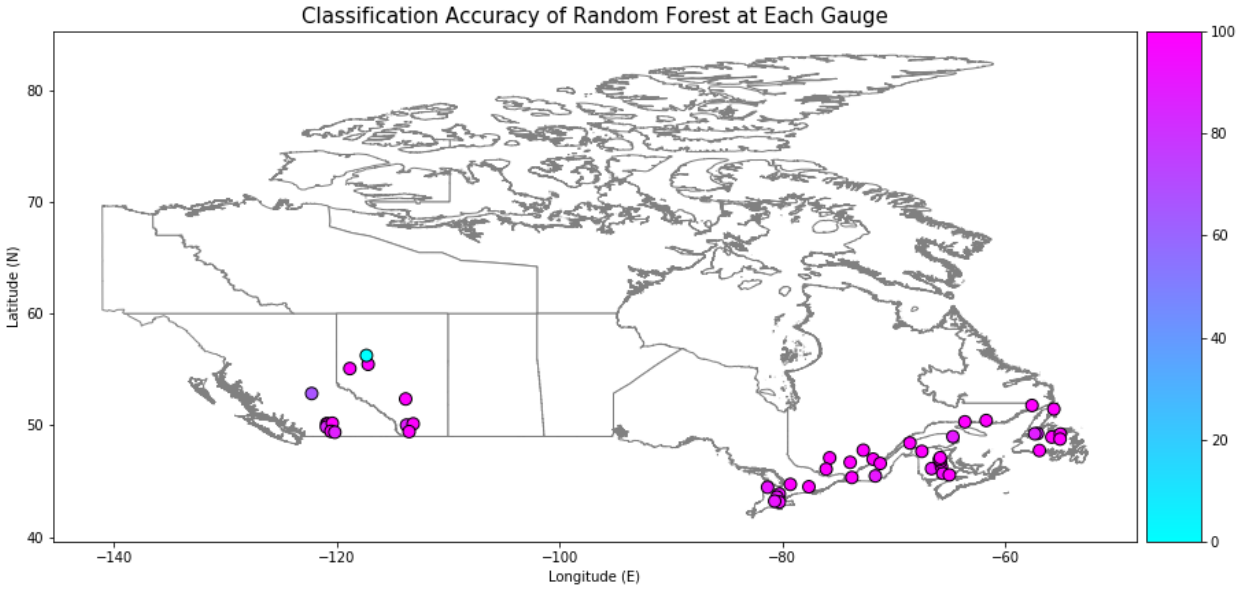
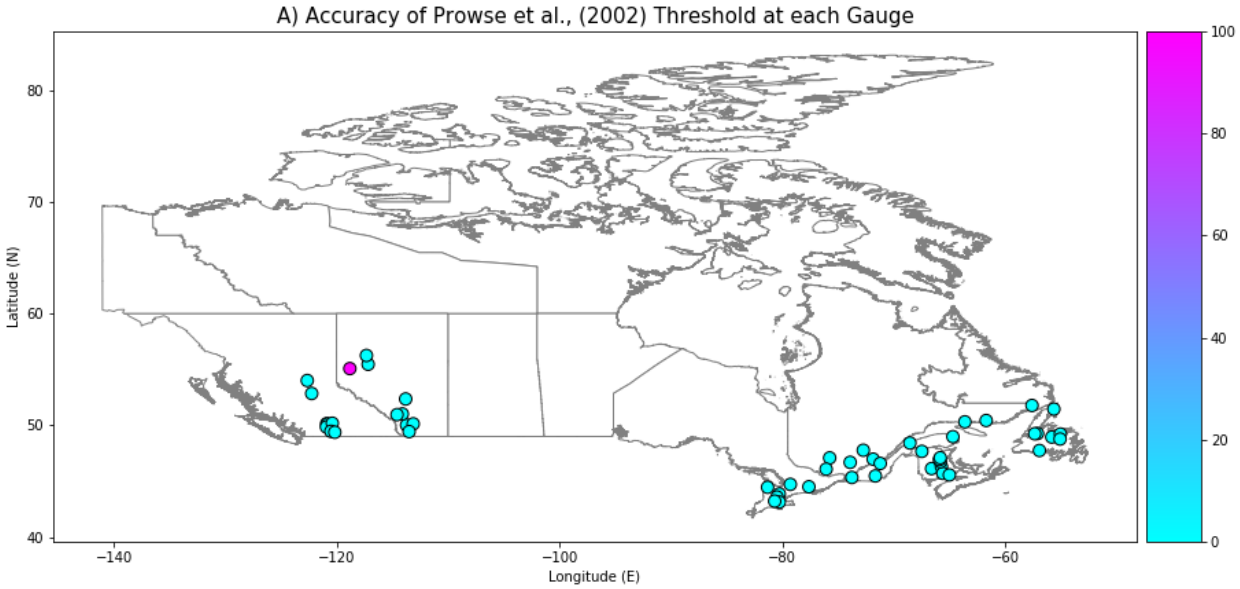


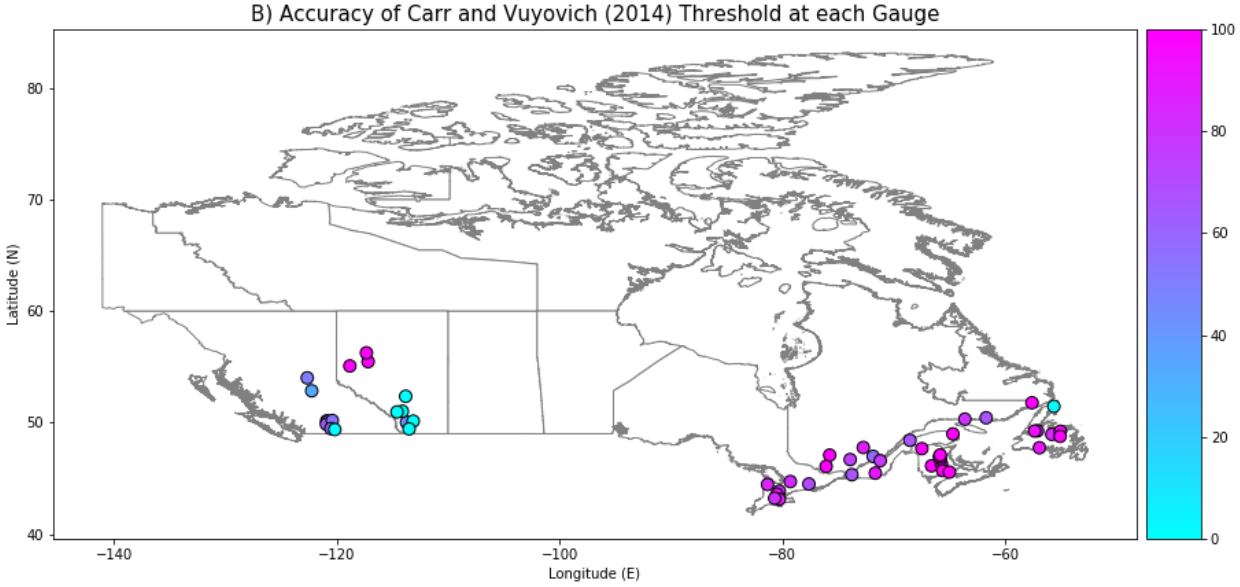
Figure 3.10: Classification accuracy of the random forest for the full dataset at each gauge in the CRID with MWBs.

3.4.5 Comparison with previous studies

Investigation of the use of thresholds to predict the occurrence of MWBs was also conducted. Previously, two thresholds for predicting the initiation of MWBs were proposed (Prowse et al., 2002;Carr and Vuyovich, 2014). The accuracy of each of these thresholds was evaluated against the data collected for this research, followed by the construction of a random forest model classifying MWB severity based solely on the threshold variables. The threshold proposed by Prowse et al., (2002) of a value of 25 MDD over the 7 days preceding an MWB, was accurate for only 0.22% of the MWBs in the CRID, as shown in Figure 3.11A. Using only the single variable of this threshold in a random forest classifying MWB severity, an overall accuracy of 26% was obtained. The threshold from Carr and Vuyovich (2014), 8 MDD or 2.8mm of precipitation over the 5 days preceding an MWB, was accurate for 12% (in the case of the MDD) and 80% (in the case of the precipitation) of the MWBs in the CRID, as shown in Figure 3.11B. A random forest model using these two variables obtained an overall accuracy of 48%.

With these results in mind, a new threshold was proposed through analysis of the values of MDD and precipitation for the 30 days preceding all MWBs contained in the CRID. An optimization process focused on maximization of the percentage of MWBs correctly predicted by threshold values was developed, with each possible combination of MDD ranging from 0 to 25 in increments of 1 and precipitation ranging from 0 mm to 10 mm in increments of 0.1 mm. A threshold of 4 MDD and 8 mm of precipitation in the 20 days preceding an MWB was selected, with an overall accuracy of 96%, shown in Figure 3.11C. The newly proposed threshold was also tested for use as classification variables, obtaining an overall accuracy for MWB severity of 64%. This new threshold exceeds both the accuracy in predicting MWBs and the classification accuracy of the previous thresholds when measured against the CRID data, however it does not outperform the final input set developed above in classifying MWB severity. This indicates that while these thresholds have some use in predicting MWB occurrence, predictions of the MWB severity are a much more complicated process requiring additional variables to be considered beyond the values presented by the thresholds. Additionally, while these thresholds may provide some early indication of the occurrence of MWBs, due to their simplicity they cannot be solely relied on for accurate prediction of MWB timing and additional research would be necessary to develop a more dependable method of predicting the timing of MWBs.


A) Accuracy of Prowse et al., (2002) Threshold at each Gauge
Latitude (N)
Longitude (E)
80
70
60
50
40
−140
−120
−100
−80
−60
100
80
60
40
20
0


B) Accuracy of Carr and Vuyovich (2014) Threshold at each Gauge
Latitude (N)
Longitude (E)
80
70
60
50
40
−140
−120
−100
−80
−60
100
80
60
40
20
0

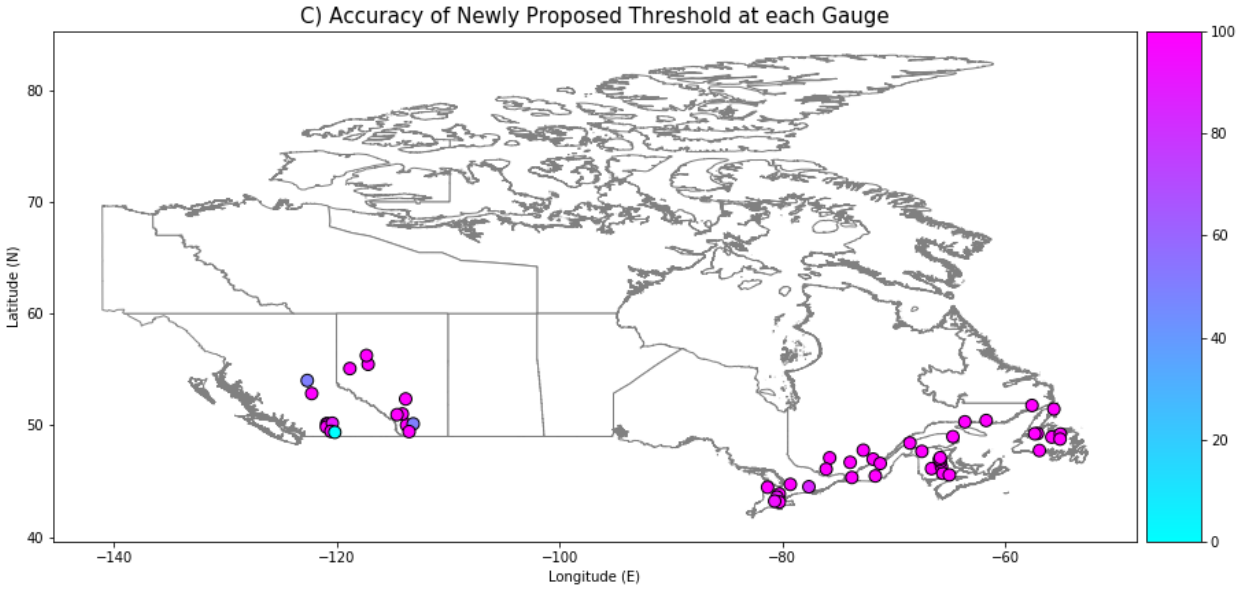


Figure 3.11: Maps of the accuracies of MWB thresholds at each CRID gauge for A) Prowse et al., 2002, B) Carr and Vuyovich, 2014, and C) the newly proposed threshold.

3.4.6 Discussion

In this study, a combination of LASSO, correlation analysis, and input omission was used to select input variables from the initial selection of 130 potential variables. Using this multi-step approach, some of the shortcomings of each method were overcome while ensuring that multiple facets of input relationships were considered. The variable selection methods have had past successes in previous research and were successful in this application in determining a set of drivers with a strong ability to predict MWB severity, but they still have shortcomings. For example, care must be taken when restricting variables to avoid excluding their potential effects in the classification model, and variable distribution should be considered when determining which variables to exclude (Chowdhury and Turin, 2020). Other methods for input selection exist with their own advantages and shortcomings including methods such as Bayesian variable selection, best subset regression, stepwise selection, and principal component analysis (George and

McCulloch, 1997, Abdi and Williams, 2010, and Zhang, 2016), although it is worth noting that there is no universally accepted method.

These concerns also extend to the model construction, where steps were taken in choosing hyperparameters, division of testing and training data, and model structure to avoid overfitting issues common in machine learning applications. *For example, although random forest is known for its ability to limit overfitting without substantially increasing error,* yet the risk remains for this model type. . *Also, though the 90/10 data split is commonly used in the literature for training and testing (Ahmad et al., 2018, You et al., 2020), there is no consensus on the optimal split and other ratios (e.g., 80/20 and 70/30) can be found in many other studies (Garcia and Muga, 2016, Wahid et al., 2017).* Future work to account for this can include the use of different proportions of training and testing data or different algorithms and training techniques.

Uncertainties remain present in the results both in terms of the used data and methods. Though the data has been quality controlled, uncertainties are still present stemming from the numerous sources and quantity of data, which increases risk of error. Some of these issues were encountered in this study and corrections were provided, but the risk remains present. Further uncertainties exist in both the application of variable selection and the random forest algorithm (Mentch and Hooker, 2016 and Rahmati et al., 2019). Detailed analysis of these inherent uncertainties may form the basis of future work.

Additionally, though the algorithm is successful for the utilised data, which includes all possible data for MWBs in CRID, the algorithm may have issues when encountering data which falls outside of the range of the data used (Hengl et al., 2018). It is recommended to continue refining both the model and thresholds as new data is acquired. This is especially important as climate change brings about more extreme and more frequent events, which may easily fall outside

of the considered ranges. Closely related to this concern are issues potentially stemming from spatial distribution, with a common consequence of models working accurately within the spatial regions for which they were developed but struggling when extended beyond the training region (Meyer et al., 2019). The current model includes data from many rivers across the country, however the climate regions they inhabit are quite similar, particularly in their capacity for MWBs. As climate change continues to alter the expected temperatures in winter, the distribution of these events can be expected to shift throughout the country, thus the current approach may no longer be applicable. With this in mind, revisions to the spatial distribution of data when training the model, or training separate regional models, may be necessary to account for spatial and climatic variability.

3.5 Conclusions

This study focussed on the analysis of the severity and causes of MWBs, which have become increasingly common in Canada as a result of climate change. This study utilised a newly released dataset, the CRID, to gain insight on MWB events occurring throughout the country. The extensive data contained in the CRID allowed this study to conduct MWBs analysis on a scale not possible before, providing the opportunity for a greater understanding of the main drivers of these events. A combination of trend analysis, return period analysis, attribution analysis, and machine learning techniques were employed to gain a better understanding of the early indicators of MWB severity. A clear increasing trend in the frequency of MWBs in Canada was identified, while return period analysis found that these events are consistently more severe than the average ice events on the affected rivers. Combining traits of the river ice cover in the early parts of the ice season and climate conditions directly preceding the MWB events, 130 initial candidate variables were identified. The initial variable set was reduced through the application of correlation analysis,

LASSO regression, and input omission. Through this analysis, 20 variables that have high link to the traits defining MWB severity were selected. These variables include the initial date of freeze up, the low flow water levels and flow rate, and several variables representing the warming rate and quantities of liquid precipitation occurring during the warming preceding and MWB. These variables can be easily monitored in advance of a potential MWB, providing the opportunity for detailed forecasting to be developed.

The strength of these selected variables' connection to MWB severity was demonstrated through the construction of a random forest model classifying the return period of the MWBs based on the identified variables. This model achieved a high level of accuracy, indicating a strong connection between the variables and the severity of the MWB events. Further, this model was compared to previously developed threshold values for the triggering of MWBs, as well as a new set of threshold values developed from the extensive data contained in the CRID. The new threshold, 4 MDD and 8 mm of precipitation in the 20 days preceding an MWB, consistently outperformed the previous thresholds and, coupled with the results of the random forest model, provides a new method through which both the occurrence and severity of MWBs throughout the country can be predicted. Through this, major indicators of MWB severity have been identified and their utility in the prediction of MWBs can continue to be researched. Future studies of these drivers may investigate alternative methods for selection of drivers or model variables such as global sensitivity analysis or principal component analysis (Ashegi et al., 2020, Ciriello et al., 2019). The study of multiple periodicities present in the time series data as well as data non-stationarity is also recommended.

Acknowledgements

This study was supported by Natural Sciences and Engineering Research Council of Canada (NSERC). Data used is available from Environment Canada (https://open.canada.ca/data/en/dataset/c5b58ccd-0011-4a80-8f24-034c86cbc14d) and Natural Resources Canada (https://cfs.nrcan.gc.ca/projects/3/4).

References

Abdi, H., and Williams, L., 2010. Principal component analysis. Computational Statistics, 2: 1-47.

Ahmad, I., Basheri, M., Iqbal, M., and Rahim, A., 2018. Performance comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection. IEEE Access, 6: 33789-33795.

Albers, S., Dery, S., and Petticrew, E., 2016. Flooding in the Nechako River Basin of Canada: A random forest modeling approach to flood analysis in a regulated reservoir system. Canadian Water Resources Journal, 41:250-260.

Apel, H., Thieken, A., Merz, B., and Bloschl, G., 2006. A probabilistic modelling system for assessing flood risks. Natural Hazards, 38: 79-100

Ashegi, R., Hosseini, S., Mojtaba, S., and Shahri, A., 2020. Updating the neural network sediment load models using different sensitivity analysis methods: a regional application. Journal of Hydroinformatics, 22.3: 562-577.

Barzegar, R., Ghasri, M., Qi, Z., Quilty, J., and Adamowski, J., 2019. Using bootstrap ELM and LSSVM models to estimate river ice thickness in the Mackenzie River Basin in the Northwest Territories, Canada. Journal of Hydrology, 577.

Beltaos, S., 1990. Guidelines for extraction of ice break-up data from hydrometric station records. Working Group on River Ice Jams: Field Studies and Research Needs. NHRI Science Report No. 2, National Hydrology Research Institute, Environment Canada, Saskatoon, SK, Canada, pp. 37–70.

Beltaos, S., 1999. Climatic effects on the changing ice-breakup regime of the Saint John River. Proceedings of the 10th Workshop on the Hydraulics of Ice Covered Rivers, Guidelines for Extraction of Ice Break-up Data from Hydrometric Station Records. Winnipeg, Canada, 251-264.

Beltaos, S., 2002. Effects of climate on mid-winter ice jams. Hydrological Processes, 16: 789–804.

Beltaos, S., 2003. Threshold between mechanical and thermal breakup of river ice cover. Cold Regions Science and Technology, 37: 1–13.

Beltaos, S., Ismail, S., and Burrell, B., 2003. Midwinter breakup and jamming on the upper Saint John River: a case study. Canadian Journal of Civil Engineering, 30: 77–88.

Beltaos, S., Prowse, T., Bonsal, B., MacKay, R., Romolo, L., Pietroniro, A., and Toth, B., 2006. Climatic effects on ice-jam flooding of the Peace-Athabasca Delta. Hydrological Processes, 20: 4031-4050.

Beltaos, S., and Prowse, T., 2009. River-ice in a shrinking cryosphere. Hydrological Processes, 23: 122-144.

Benesty, J., Chen, J., Huang, Y., and Cohen, I., 2009. Pearson Correlation Coefficient. Noise Reduction in Speech Processing. Springer Topics in Signal Processing, 2: 1-4.

Boyd, d., 1979. Degree days: The different types. Building research note, National Research Council of Canada from the Atmospheric Environment Service Department of Fisheries and Environment.

Chowdhury, M., and Turin, T., 2020. Variable selection strategies and its importance in clinical prediction modelling. Family Medicine and Community Health, 8:1-7.

Carr, M., and Vuyovich, C., 2014. Investigating the effects of long-term hydro-climatic trends on Midwest ice jam events. Cold Regions Science and Technology, 106-107: 66-81.

Ciriello, V., Lauriola, I., and Tartakovsky, D., 2019. Distribution-based global sensitivity analysis in hydrology. Water Resources Research, 55: 8708-8720.

de Rham, L., Prowse, T., Beltaos, S., and Lacroix, M., 2008a. Assessment of annual high-water events for the Mackenzie river Basin, Canada. Hydrological Processes. 22: 3864-3880.

de Rham, L., Prowse, T., and Bonsal, B., 2008b. Temporal variations in river-ice break-up over the Mackenzie River Basin, Canada. Journal of Hydrology, 349: 441-454.

de Rham, L., Dibike, Y., Beltaos, S., Peters, D., Bonsal, B., and Prowse, T., 2020. A Canadian river ice database from national hydrometric program archives. Earth System Science Data, Open Access Discussions.

Du, S., Gu, H., Wen, J., Chen, K., and Van Rompaey, A., 2015. Detecting flood variations in Shanghai over 1959-2009 with Mann-Kendall tests and newspaper-based database. Water, 7:1808-1824.

Galelli, S., Humphrey, G., Maier, H., Casteletti, A., Dandy, G., and Gibbs, M., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environmental Modelling and Software, 62: 33-51.

Garcia, F., and Muga, F., 2016. Random forest for malware classification. Cryptography and Security, *arXiv preprint arXiv:1609.07770.*

George, E., and McCulloch, R., 1997. Approaches for Bayesian variable selection. Statistica Sinica, 7:339-373.

Ghaith, M., Siam, A., Li, Z., and El-Dakhakhni, W., 2020. Hybrid hydrological data-driven approach for daily streamflow forecasting. Journal of Hydrological Engineering, 25(2): 04019063.

Goswami, P., and Braahma, H., 2019. River water level prediction modelling using artificial neural network and multiple linear regression. International Journal of Engineering and Management Research, 9:23-31.

Goulding, H., Prowse, T., and Bonsal, B., 2009a. Hydroclimatic controls on the occurrence of break-up and ice-jam flooding in the Mackenzie Delta, NWT, Canada. Journal of Hydrology, 379: 251-267.

Goulding, H., Prowse, T., and Beltaos, S., 2009b. Spatial and temporal patterns of break-up and ice-jam flooding in the Mackenzie Delta, NWT. Hydrological Processes, 23: 2654-2670.

Guo, X., Wang, T., Fu, H., Guo, Y., and Li, J., 2018. Ice-jam forecasting during river breakup based on neural network theory. Journal of Cold Regions Engineering, 32(3): 04018010.

Hall, P., and Miller, H., 2009. Using generalized correlation to effect variable selection in very high dimensional problems. Journal of Computational and Graphical Statistics, 18: 533-550.

Hastie, T., Tibshirani, R., and Friedman, J., 2009. The elements of statistical learning: Data mining, inference and prediction. Springer-Verlag, 2009.

Hamed, K., and Rao, A., 1998. A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology, 204: 182-196.

Hengl, T., Nussbaum, M., Wright, M., Heuvelink, G., and Graler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6:e5518

Hopkinson, R., McKenney, D., Milewska, E., Hutchinson, M., Papadopol, P., and Vincent, L., 2011. Impact of aligning climatological day on gridding daily maximum-minimum temperature and precipitation over Canada. Journal of Applied Meteorology and Climatology, 50: 1654-1665.

Huntington, T., Hodgkins, G., and Dudley, R. 2003. Historical trend in river ice thickness and coherence in hydroclimatological trends in Maine. Climate Change 61: 217–236.

Hutchinson, M., McKenney, D., Lawrence, K., Pedlar, J., Hopkinson, R., Milewska, E., and Papadopol, P., 2009. Development and testing of Canada-wide interpolated spatial models of daily minimum-maximum temperature and precipitation for 1961-2003. American Meteorological Society, 48: 725-741.

Javeed, A., Zhou, S., Yongjian, L., Qasim, I., Noor, A., and Nour, R., 2019. An intelligent learning system based on random forest search algorithm and optimized random forest model for improved heart disease detection. IEEE Access, 7:180235-180243.

Li, Z., Huang, G., Wang, X., Han, J., and Fan, Y., 2016. Impacts of future climate change on river discharge based on hydrological inference: A case study of the Grand River Watershed in Ontario, Canada. Science of the Total Environment, 548-549:198-210.

Liaw, A., and Wiener, M., 2002. Classification and regression by randomForest. R News, 2:18-22.

Mahabir, C., Hicks, F., and Fayek, A.R., 2006. Neuro-fuzzy river ice breakup forecasting system. Cold Regions Science and Technology. 46:100-112.

Mandrekar, J., 2010. Receiver operating characteristic curve in diagnostic test assessment. Biostatistics for Clinicians, 5:1315-1316.

Massie, D., White, K., and Daly, S., 2002. Application of neural networks to predict ice jam occurrence. Cold Regions Science and Technology. 35:115-122.

McKenney, D., Hutchinson, M., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., Milewska, E., Hopkinson, R., Price, D., and Owen, T. 2011. Customized spatial climate models for North America. Bulletin of American Meteorological Society-BAMS December: 1612-1622.

McKinney, W., and others, 2010. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 445: 51-56.

McMaster, G., and Wilhelm, W., 1997. Growing degree-days: one equation, two interpretations. Agricultural and Forest Meteorology, 87: 291-300.

Meinshausen, N., and Buhlmann, N., 2006. High-dimensional graphs and variable selection with the LASSO. The Annals of Statistics, 34: 1436-1462.Mentch, L., and Hooker, G., 2016. Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. Journal of Machine Learning Research, 17:1-41.

Meyer, H., Reudenbach, C., Wollauer, S., and Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecological Modelling, 411: 108815.

Newton, B., Prowse, T., and de Rham, L., 2017. Hydro-climatic drivers of mid-winter break-up of river ice in western Canada and Alaska. Hydrology Research 48.4: 945-956.

Oliphant, T., 2006. A guide to NumPy (Vol. 1). Trelgol Publishing USA.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and others, 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Probst, P., Wright, M., and Boulesteix, A., 2018. Hyperparameters and tuning strategies for random forest. arXiv, arxiv:1804.03515.

Prowse, T., Bonsal, B., Lacroix, M., and Beltaos, S., 2002. Trends in river-ice breakup and related temperature controls. Proceedings of the 16th IAHR International Symposium on Ice. Trends in River-Ice Breakup and Related Temperature Controls. Dunedin, New Zealand, 64-71.

Rahmati, O., Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., Mollaefar, E., Tiefenbacher, J., Cipullo, S., Ahmad, B., and Bui, D., 2019. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. Science of the Total Environment, 688: 855-866.

Rokaya, P., Budhathoki, S., and Lindenschmidt, K., 2018. Trends in the timing and magnitude of ice-jam flood in Canada. Scientific Reports 8, 1:1-9.

Sadler, J., Goodall, J., Morsy, M., and Spencer, K., 2018. Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and random forest. Journal of Hydrology, 559: 43-55.

Segal, M., 2003. Machine learning benchmarks and random forest regression. UCSF: Center for Bioinformatics and Molecular Biostatistics. Retrieved from https://escholarship.org/uc/item/35x3v9t4

Sermpinis, G., Tsoukas, S., and Zhang, P., 2018. Modelling market implied ratings using LASSO variable selection techniques. Journal of Empirical Finance 48: 19-35.

Shahri, A., Larsson, S., and Renkel, C., 2020. Artificial intelligence models to generate visualized bedrock level: a case study in Sweden. Modeling Earth Systems and Environment, 6:1509-1528.

Shahri, A., Pashamohammadi, F., Ashegi, R., and Shahri, H., 2021. Automated intelligent hybrid computing schemes to predict blasting induced ground vibration. Engineering with Computers, https://doi.org/10.1007/s00366-021-01444-1

Snieder, E., Shakir, R., and Khan, U., 2020. A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. Journal of Hydrology, 583: 124299.

Sokolova, M., and Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information Processing and Management, 45: 427-437.

Sun, W., and Trevor, B., 2015. A comparison of fuzzy logic models for breakup forecasting of the Athabasca River. CRIPE 18th Workshop on the Hydraulics of Ice Covered Rivers, Quebec City, QC, Canada, August 18-20, 2015.

Sun, W., 2018. River ice breakup timing prediction through stacking multi-type model trees. Science of the Total Environment, 644:1190-1200.

Sun, W., and Trevor, B., 2018. A stacking ensemble learning framework for annual river ice breakup dates. Journal of Hydrology. 561:636-650.

Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society, 58: 267-288.

Van Rossum, G., and Drake, F., 2009. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Vervuren, P., Blom, C., and De Kroon, H., 2003. Extreme flooding events on the Rhine and the survival and distribution of riparian plant species. Journal of Ecology, 91: 135-146.

Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S., Brett, M., Wilson, J., Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, I., Feng, Y., Moore, E., VanderPlas, J., Laxalde, D., Pertkold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17: 261-272.

Wahid, F., Ghazali, R., Shah, A., and Fayaz, M., 2017. Prediction of energy consumption in the buildings using multi-layer perceptron and random forest. International Journal of Advanced Science and Technology, 101:13-22.

Wang, J., Sui, J., Guo, L., Karney, B., and Jüpner, R., 2010. Forecast of water level and ice jam thickness using the back propagation neural network and support vector machine methods. International Journal of Environmental Science & Technology, 7(2): 215‑224.

Wang, T., Yang, K.L., Guo, X.L., and Fu, H., 2012. Application of adaptive network based fuzzy inference system to ice condition forecast. Journal of Hydraulic Engineering, 1:18.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., and Bai, X., 2015. Flood hazard risk assessment model based on random forest. Journal of Hydrology, 527: 1130-1141.

Waskom, M., and the seaborn development team, 2020. Seaborn. Zenodo, https://doi.org/10.5281/zenodo.592845

White, K., 1996. Predicting breakup ice jams using logistic regression. Journal of Cold Regions Engineering, 10(4):178-189.

Wu, X., and Zhang, P., 2016. A two-stage random forest method for short-term load forecasting. 2016 IEEE Power and Energy Society General Meeting (PESGM), 1-5.

You, J., van der Klein, S., Lou, E., and Zuidhof, M., 2020. Application of random forest classification to predict daily oviposition events in broiler breeders fed by precision feeding system. Computers and Electronics in Agriculture, 175: 105526.

Zhang, Z., 2016. Variable selection with stepwise and best subset approaches. Annals of Translational Medicine, 4(7):136.

Zhao, L., Hicks, F., and Fayek, A., 2012. Applicability of multilayer feed-forward neural networks to model the onset of river breakup. Cold Regions Science and Technology, 70:32-42.

Zhou, P., Li, Z., Snowling, S., Baetz, B., Na, D., and Boyd, G. (2019). A random forest model for inflow prediction at wastewater treatment plants. Stochastic Environmental Research and Risk Assessment, 33(10), 1781-1792.

Zhou, P., Li, Z., Snowling, S., Goel, R., and Zhang, Q. (2019). Short-term wastewater influent prediction based on random forests and multi-layer perceptron. Journal of Environmental Informatics Letters, 1(2), 87-93.

Chapter 4: Machine-Learning Approach for Predicting the Occurrence and Timing of Mid-Winter Ice Breakups on Canadian Rivers

Forecasting of extreme river ice events is often challenged by the rarity of said events relative to the larger time scale. This chapter focusses on the application of novel rare-event forecasting techniques in the prediction of MWBs in Canada, developing a unique two-level model structure to predict the occurrence of these events. Both model levels utilised rare-event forecasting techniques in the configuration of data, the selection of model algorithms, and the assessment of model accuracy. The final model configuration achieved a high level of accuracy on the tested events at both model levels.

Michael De Coste was responsible for investigation, model development and testing, and results validation under the guidance of Dr. Zhong Li and Dr. Yonas Dibike, who also supplied the data used in the study. The manuscript was drafted and prepared by Michael De Coste and revised by Michael De Coste, Dr. Zhong Li, and Dr. Yonas Dibike.

This chapter has been published: Michael De Coste, Zhong Li, and Yonas Dibike. (2022) Machine-learning approach for predicting the occurrence and timing of mid-winter breakups on Canadian Rivers. *Environmental Modelling and Software*, 105402. (DOI: 10.1016/j.envsoft.2022.105402)

Abstract

The increasingly common occurrence of Mid-Winter Breakups (MWBs) in Canadian rivers, consisting of the early breakup of ice cover outside of the typical spring season, is a cause for concern. This study applied various data-driven modelling techniques to predict MWBs occurrence and timing with sufficient lead times on a national scale using a new Canadian River Ice Database (CRID) coupled with National Resources Canada gridded climate data. A two-level machine learning model structure was developed, with the first level model predicting MWB occurrence within a given period and the second level model predicting the timing of an MWB occurrence within that period. Machine learning techniques that can handle class imbalance were employed to address many of the issues inherent in rare event forecasting, including the implementation of data preprocessing, the selection of algorithms and performance metrics suited to rare events. Multiple configurations of both model levels, including variations on time series arrangement and input variables, were tested to select the optimal model structure. The best performing configuration, focussing on a biweekly time period, attained overall accuracies of 80.1% and 77.6% for the first and second level models respectively on the 452 MWBs in the CRID. In addition, probabilistic prediction results were analyzed to evaluate model uncertainty and robustness. This new modelling framework provides the first tool capable of predicting MWBs on a national scale, with easily extendable methodology to locations that have not yet experienced MWBs and can form the basis of vital decision-making support to affected communities.

Keywords: River ice; mid-winter breakup; rare event forecasting; imbalanced learning; prediction; data-driven modelling

Software and Data Availability

Analysis in this study was completed using Python v3.7 (https://www.python.org/downloads/release/python-370/). Data used in this study is available through Environment and Climate Change Canada (https://open.canada.ca/data/en/dataset/c5b58ccd-0011-4a80-8f24-034c86cbc14d) and Natural Resources Canada (https://cfs.nrcan.gc.ca/projects/3/4).

4.1 Introduction

The spring breakup of river ice cover is a yearly occurrence on rivers throughout cold regions of the world, with many environmental processes depending on both the timing and mechanism of the breakup (Beltaos et al., 2003). The combination of runoff from snowmelt and melting of ice cover is important to ecosystems, especially those that depend on the higher than usual water levels generated by these types of events (Beltaos et al., 2006). However, in recent years it has been noted that many rivers in Canada have become vulnerable to Mid-Winter Breakups (MWBs), which are the early breakup of river ice cover caused by unseasonably high temperatures and the occurrence of liquid precipitation (Prowse et al., 2002, Carr and Vuyovich, 2014). While the ice cover typically re-establishes as temperatures return to seasonal values, the flow regime for the remainder of the season is significantly altered (Huntington et al., 2003). Spring breakups following these events are often milder, with lower flows and water levels, which can have significant environmental impacts to those processes that rely on spring breakups to drive higher water levels (Beltaos et al., 2003). However, MWBs also carry the risk of river ice jam formation where the ice rubble can freeze in place, creating a semi-permanent dam on the river that can trigger severe flooding (Beltaos, 2003). To reduce such risk of severe flooding and their long-term impacts on the river system, effective means of predicting MWB events are of great interest to affected communities. Though past studies have attempted to forecast these events, they have been limited both by the effectiveness of the methodology and the availability of data (Newton et al., 2017).

Analysis of the spring breakup processes on rivers, typically driven by more gradual temperature increases in combination with snowmelt runoff (Beltaos et al., 2006), has been conducted through application of large-scale data analysis techniques at regional scales. For

example, de Rham et al. (2008a) studied the high water events driven by the changing spatial distribution of breakups in the Mackenzie River Basin in Canada. Through analysis of Water Survey of Canada gauges within the basin, it was concluded that spring breakups are trending towards earlier occurrence and longer affected periods (de Rham et al., 2008b). Trend analysis of these events identified factors that influenced the timing and magnitude of breakups while also reinforcing the trend towards earlier and longer breakups (Goulding et al., 2009), with additional spatial mapping and analysis of peak flows reinforcing these findings. The breakup trends in the Peace Athabasca Delta in Canada has also been well studied. The delta has seen a notable decreasing trend in ice jam occurrence, leading to perched basins which rely on the flooding triggered by these events to be replenished beginning to reduce (Timoney, 2002). Regulation on the Peace River upstream has been generally accepted as the cause, and subsequent research has focussed on investigating changes in ice jam frequency and severity through statistical and modelling based means in the complex delta (Beltaos, 2017, Timoney et al., 2018, Jasek, 2019a, and Jasek, 2019b). Statistical data analysis has also identified factors that are best indicators of ice jam occurrence in the delta, such as winter temperatures and precipitation (Lamontagne et al., 2021). Though these studies have investigated factors driving breakup and breakup related events on a regional scale, their focus was limited to the more easily monitored spring breakup events and limited to singular regions.

Research into MWBs has historically encountered issues related to data availability and spatial coverage, with trends being successfully modelled while forecasting is often difficult. A trend towards earlier spring breakups on the Saint John River triggered by an increase in MWBs was noted by Beltaos (1999), with a notable increasing trend in the occurrence of mild winter days. Beltaos and Prowse (2009) subsequently noted an increasing risk of spring and mid winter ice jams

on the river as a results of climate change effects. Prowse et al., (2002) and Carr and Vuyovich (2014) conducted research into the drivers of MWBs, producing predictive thresholds based on trends in temperature and precipitation. Though these thresholds were entirely accurate for the specific location for which they were developed, they were found to not translate with a meaningful accuracy to other locations through a regional investigation of MWBs in Western Canada and the United States (Newton et al., 2017). Their research identified 52 MWBs in the considered region, with the aforementioned thresholds being accurate for only 11% and 60% of the events respectively as those studies were limited by the available data. The MWB database developed by Newton et al., (2017), which covers the western portion of Canada and the United States, was one of the largest collections of data on these types of events. Though these studies developed some methods of simple predictions of MWBs, there has been no attempt at the creation of a comprehensive forecasting tool for these events.

In comparison to MWBs, the prediction of spring breakup events timing through machine learning has been much more thoroughly researched, though these projects often focussed on a single river or river network and are based around events with a more reliable timing. Simple single-model algorithms such as artificial neural networks (Zhao, 2012, Guo et al., 2018) and support vector machines (Wang et al., 2010, Barzegar et al., 2019), as well as more complicated algorithmic structures such as adaptive neuro-fuzzy inference systems (Wang et al., 2012, Sun and Trevor, 2015, Sun and Trevor, 2018) and stacking ensembles (Sun, 2018), have had ranges of success in spring breakup prediction. One such example is the prediction of spring breakup ice jams, which have been predicted with logistic regression (White, 1996), artificial neural networks (Massie et al., 2002), hybrid neuro-fuzzy systems (Mahabir et al., 2006), and stacking ensembles (De Coste et al., 2021). Like prediction of breakup timing, these studies were often limited to

specific rivers, with an added problem of accounting for rare event forecasting, a challenging data issue in machine learning that can lead to biased results from models. The rare event forecasting issue would extend to MWB forecasting attempts as well and has never been addressed so far.

This paper focusses on the development and application of a two-level modelling framework for the prediction of MWB occurrence and timing on a national scale, applying machine learning and probabilistic analysis techniques while avoiding common issues that challenge rare event forecasting. The study combines data of 452 MWBs from affected 52 rivers throughout Canada with a national scale climate dataset to facilitate development of four machine learning algorithms (Adaptive Boosting, Class Switching, K-Nearest Neighbors, and Adaptive Resampling and Combining X4). This new approach allows for the prediction of an MWBs occurrence within a given time period using the first level model and the subsequent prediction of the MWBs timing within the period using the second level model. The developed models are tested and validated from both a classification and probabilistic standpoint, allowing the best model configuration to be selected. **This is the first successful application of these modelling techniques in the prediction of MWBs on a national scale. It provides an easily configured methodology for the development of models predicting these events that is easily transferable to other locations.**

4.2 Data

4.2.1 CRID

The primary source of historical ice data for this study was the Canadian River Ice Database (CRID), an extensive collection of data taken from a select set of Water Survey of Canada gauges located across the country (de Rham et al., 2020). This dataset was compiled through analysis and extraction of water level time series, utilising visual inspection techniques first proposed by

Beltaos (1990) to identify the key river ice events of each season based on the observed changes of water levels. Data from 196 river gauges in Canada are included in the CRID, with 46 located on regulated rivers and 150 on unregulated rivers. Fifteen river ice related events (e.g., freeze up, winter low flows and levels, MWBs, secondary low flows and levels, breakups, etc.) are included with corresponding water levels, flows, and dates recorded if available. Included in these events are MWBs at gauges where they have been historically observed, identified as peaks in the flow record outside of typical spring breakup periods for the gauge and associated with higher-than-average air temperatures recorded at nearby climate gauges. Each MWB includes the date, flow, and water level at both the initiation of the event and the peak flow if available.

Figure 4.1 shows the 196 gauges of the CRID with the 52 gauges that have had historical MWB events highlighted. These gauges were concentrated in the southwest (British Columbia and Alberta) and southeast (Ontario, Quebec, New Brunswick, and Newfoundland and Labrador) portions of the country. These regions fall within the relatively warmer portions of the country that are more vulnerable to mid-winter thaws that trigger MWBs. In total 452 MWBs are included in the CRID amongst the 52 identified stations. The time span covered by the CRID varies between gauges, with the earliest MWB on record occurring in 1955, but most (301 out of 452) MWBs occurring after 1990. Water levels at seven gauges with MWBs (28 station years) were corrected for outliers using historical notes relating to gauge datum changes and in consultation with Environment and Climate Change Canada.

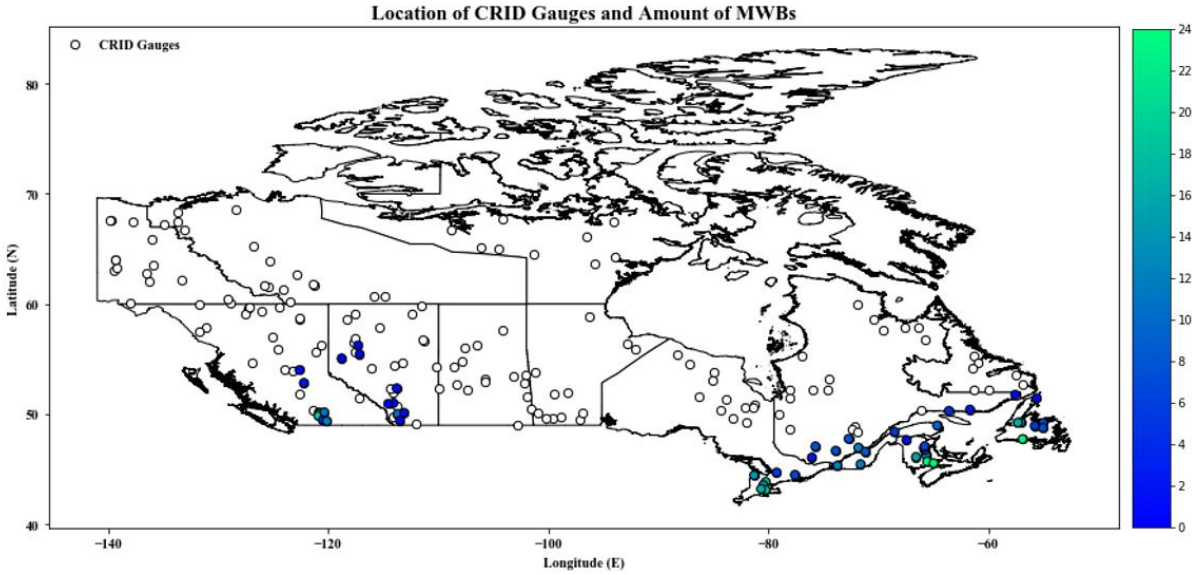


Figure 4.1: Locations of CRID gauges with locations and quantities of MWBs in color and gauges without MWBs as open circles.

4.2.2 NRCan Gridded Climate Data

Climate data corresponding to MWB stations were extracted from the National Resources Canada (NRCan) Daily Gridded Climate Dataset (https://cfs.nrcan.gc.ca/projects/3/4). The NRCan dataset provides daily maximum and minimum air temperatures and total precipitation data with a resolution of 0.1˚ at a national scale. The data spans 1950-2015 and is derived from quality controlled, non-homogenized station data taken from the National Climate Data Archive (NCDA) of Environment and Climate Change Canada data (Hutchinson et al., 2009). This station data has been interpolated onto a high-resolution grid (i.e., 0.1˚ × 0.1˚) through thin plate splines (McKenney et al., 2011, Hopkinson et al, 2011). In this study, climate data at the closest grid point to each CRID gauge was extracted using geographic information system mapping techniques, providing a complete time series of the three extracted variables for all CRID gauges. Maximum and minimum air temperatures were averaged to obtain daily mean air temperatures. Then, the Accumulated Freezing Degree Days (AFDD) was calculated as the sum of temperatures below

0°C beginning from a reference date of October 1st, which is the beginning of the Canadian water year (Boyd, 1979). Total precipitation cumulatively summed from the freeze-up date of river ice cover was also calculated. These two values represent how cold a particular season has been and how much snowpack has fallen, respectively.

4.3 Methods

4.3.1 Model Construction with Imbalanced Data

Data imbalance is a common issue in rare event forecasting, especially in the case of binary classification where there is a significant difference in the number of observations in the majority class and the minority class (Maalouf and Trafalis, 2011, Li et al., 2016). In this study there is a massive imbalance present in the data between the majority class (periods or days with no MWB) and minority class (periods or days with an MWB). The majority class makes up 92-99% of the dataset depending on the configuration. Without any modifications to the traditional algorithmic construction process, a common result of these imbalances is the classification of all cases as belonging to the majority class. While this would represent an accuracy of 92-99%, the results would be useless in practice, as none of the rare events were properly classified. To address these issues, changes to the algorithm construction process can be implemented to data preprocessing, the algorithm selection, and the performance metric selection (Haixiang et al., 2017).

4.3.1.1 Data Preprocessing and Resampling

Resampling is the most common and versatile data preprocessing technique for imbalanced datasets due to its independence from the selected classifiers (Lopez et al., 2013). Two common methods of resampling are over-sampling, which works by duplicating instances of the minority class (Chawla et al., 2002), and under-sampling, which works by removing instances of the majority class (Tahir et al., 2009). Hybrid methods combining both over-sampling and under-

sampling exist, but their utilisation is rare. Under-sampling was selected in this study due to its versatility and effectiveness in computation time when there are hundreds of samples in the minority class, as there are in the MWB datasets (Napierala and Stefanowski, 2015, Loyola-Gonzalez et al., 2016). An even balance between the classes was used due to its past successes in application to classification without affecting generalization of the model results (Koziarski, 2020). The under-sampling and even balance techniques have had previous success in rare event forecasting of class imbalanced data (Li et al., 2013, Kumar et al., 2014, Sun et al., 2015).

4.3.1.2 Algorithm Selection

Ensemble techniques which train multiple classifiers and some traditional single model classifiers are recommended for imbalanced learning, including cases where data preprocessing has been applied (Galar et al., 2012). In this study, four model algorithms with previous success in addressing class imbalances in combination with preprocessing and performance metric selection (Haixiang et al., 2017), were tested and are detailed below.

Adaptive Boosting (AdaBoost): The AdaBoost algorithm develops an ensemble of iterative classification trees with associated class labels (Zhu et al., 2009). An initial classifier is constructed and used to classify observations until an incorrect classification occurs. A subsequent tree is then developed with the weight of that point "boosted" using the Stagewise Additive Modelling using a Multi-class Exponential (SAMME) loss function. This process continues until all training data has been run through the iterative ensemble. The overall goal of the process is the minimization of misclassification rates through the boosting of the weight of a weak learners' accuracy. The final model label output by the ensemble is calculated using Equations 4.1 and 4.2.

$$f_T(x) := \sum_{t=1}^{T} c_t h_t(x) \tag{4.1}$$

$$\bar{f}_T(x) := sign(f_T(x)) \tag{4.2}$$

where $h_t(x)$ is the ensemble of *T* hypotheses (tuned by the user) from the input vector *x* (ranging from 0 to 1), c_t is the weight of each learner (with c_t satisfying the conditions $c_t \geq 0$ and $\sum_{t=1}^{T} c_t = 1$), $f_T(x)$ is the weighted hypothesis, and $\bar{f}_T(x)$ is the final model hypothesis (Ratsch et al., 2001). These models have been applied in the forecasting of rainfall-runoff and rare dust storms (Liu et al., 2014 and Zhang et al., 2014).

k-Nearest Neighbors (KNN): KNN produces single models which function by classifying unlabelled samples based on the applied labels of the *k* nearest neighbors, with *k* tuned by the user (Dudani, 1976). Each new observation classified to the model is labelled based on a weighting factor, w_j, described by Equation 4.3.

$$w_j = \sum_{j=1}^{k} \begin{cases} \dfrac{d_k - d_j}{d_k - d_1}, & d_k \neq d_1 \\ 1, & d_k = d_1 \end{cases} \tag{4.3}$$

where the distance between neighboring points is described by d_j, j=1,…,k, and the weight for an individual point is w_j (Guo et al., 2003). These models have been successfully applied to the rare event forecasting of ice jam occurrence (Semenova et al., 2020) and real time flood forecasting (Liu et al., 2020).

Class Switching: Class Switching is an ensemble approach where class labels are randomly flipped during model training, allowing a more diverse classifier to be developed (Breiman, 2000). The ensemble is composed of classification trees, similar to the AdaBoost algorithm, with each

individual classifier generated based on the original training data with class labels switched at a rate defined by Equations 4.4, 4.5, and 4.6.

$$P_{j \leftarrow i} = wP_j \; for \; i \neq j \tag{4.4}$$

$$P_{i \leftarrow i} = 1 - w(1 - P_i) \tag{4.45}$$

$$w = \frac{p}{1 - \sum_j P_j^2} \tag{4.6}$$

where $P_{j \leftarrow i}$ is the probability that an element with the label *i* gets labelled as *j*, $P_{i \leftarrow i}$ the probability that an element with the label *i* remains labelled *i*, P_i the proportion of elements in the training set labelled *i*, P_j the proportion of elements in the training set labelled *j*, and *w* is proportional to the switching rate *p*, a factor tuned by the user. Martinex-Munoz and Suarez (2005) applied the algorithm successfully to a variety of binary classification tasks, including rare event forecasting.

Adaptive Resampling and Combining X4 (ArcX4): ArcX4 is an algorithm belonging to the Adaptive Resampling and Combining (ARC) family of iterative ensemble classifiers (Breiman, 1996). At each stage of model construction, the model resamples a portion of the training set to generate a classifier C_n (where *n* ranges from 1 to *k*) based on a probability described by Equation 4.7.

$$p(n) = \frac{1 + m(n)^4}{\sum(1 + m(n)^4)} \tag{4.7}$$

where $p(n)$ is the probability for resampling of the *n*th classifier, and $m(n)$ is the number of misclassifications of the *n*-1th classifier. The results of each classifier are combined through unweighted voting, with the number of classifiers *k* tuned by the user. When applied to imbalanced classification, the algorithm was found to outrank several other ensemble techniques (Narassiguin et al. 2016).

Each of the above-described models requires the tuning of hyperparameters to select the best model configuration for the given task. For this application, models were trained using five-fold cross-validation. The data is subdivided into five portions, with a training and testing process being conducted five times such that each partition is used as the testing data. The possible considered values of the parameter(s) of each model are organized into grids, and an exhaustive grid search was used to test every possible combination of parameters for each partition of data (Refaeilzadeh et al., 2009). The ability of each configuration to correctly classify MWBs was averaged to give a performance assessment, and the configuration with the best performance was used to select the optimum model topology. Finally, a stratified 80/20 training/testing split was used to develop the classification models.

4.3.1.3 Performance metrics

Performance metrics for the assessment of model tuning and final model accuracy were selected to focus on the correct classification of the minority case or rare event, the occurrence of an MWB. These metrics are based around a comparison of the actual class and the predicted class of an observation, assessing it with one of four outcomes. True positive (*tp*) denotes the correct prediction of an MWB, true negative (*tn*) denotes the correct prediction of no MWB, false positive (*fp*) denotes the prediction of an MWB when one does not occur, and false negative (*fn*) denotes the prediction of no MWB when one does occur. The primary metric for assessment is recall, which assesses the correct classification of the minority class, with additional assessment metrics of specificity (correct classification of the majority case) and balanced accuracy also calculated (Broderson et al., 2010). For each, values will range between 0 and 1, with 1 representing perfect classification, and are calculated using Equations 4.8 – 4.10.

$$Recall = \frac{tp}{tp + fn} \tag{4.8}$$

$$Specificity = \frac{tn}{tn + fp} \tag{4.9}$$

$$Balanced\ Accuracy = \frac{Specificity + Recall}{2} \tag{4.10}$$

Each of the four considered algorithms in Section 4.3.1.2 internally computes binary classifications by first producing a probabilistic prediction that an observation falls into a given class before conversion into a final binary value. In the case of a single model algorithm a single probability is produced, while ensemble algorithms use a mean of the values produced by each ensemble member (Pedregosa et al., 2011). These initial probabilistic predictions were also extracted and assessed using the metrics of Brier Loss and Log Loss (Roulston, 2007). Both loss functions compare the predicted probability of a sample being in the minority class to the actual binary value according to Equations 4.11 and 4.12.

$$Brier\ Loss = \frac{1}{N}\sum_{i=1}^{N}(p(y_i) - y_i)^2 \tag{4.11}$$

$$Log\ Loss = \frac{1}{N}\sum_{i=1}^{N} y_i \log\big(p(y_i)\big) + (1 - y_i) * \log\,(1 - p(y_i)) \tag{4.12}$$

where y_i is the binary value of the event and $p(y_i)$ is the predicted probability of the event for N events. For each of the above metrics, a value of 0 represents perfect probabilistic prediction, while values above 0.25 for Brier loss and 0.675 for Log loss represent unacceptable accuracy no better than chance (Roulston, 2007 and Vovk, 2015). Calibration curves, plotting the mean of predicted probabilities against the mean of the target variable for a set of observations were also used to evaluate the model performance. A perfectly calibrated probabilistic classifier would output a straight line from 0 to 1 (Niculescu-Mizil and Caruana, 2005).

4.3.2 Two-Level Model Framework

A two-level modeling framework was developed to allow for versatility in monitoring practice. The first level model focuses on an individual period of the year, producing a binary prediction of whether or not an MWB would occur within that period. If the first level model predicts an MWBs occurrence, the second level model would then be used to predict the timing of the event, producing a three day ahead binary prediction of an MWB. Figure 4.2 shows a flowchart of the model structure. Both models would use unique datasets featuring data easily monitored well in advance of an MWB. Though the first level model is used to trigger the use of the second level model, there is no direct feed of data or results from the first level model into the second level.

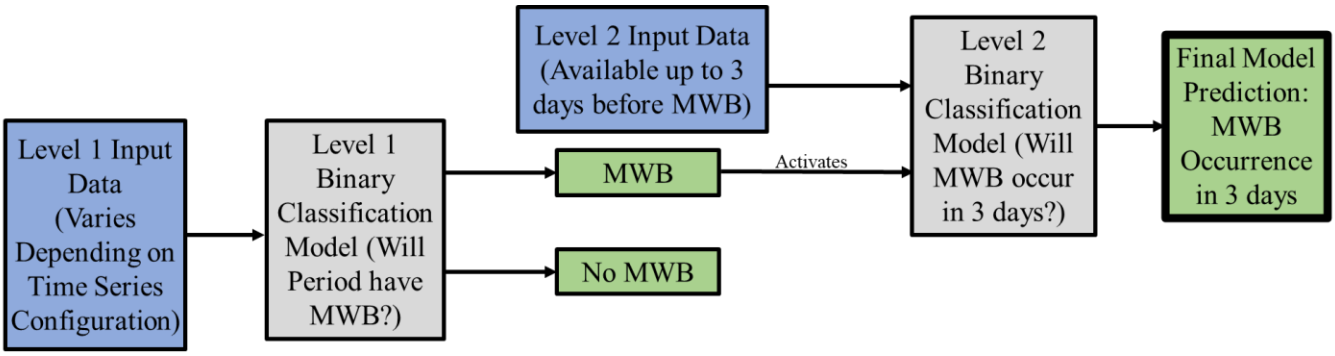


Figure 4.2: Two-Level model structure flow chart for the prediction of MWB occurrence and timing.

4.3.2.1 First Level Model

The primary goal of the first level model was to identify whether an MWB would occur in a given time period between December 15 to March 15. The duration of the target time period was the first aspect of model development requiring testing. Four different time series configurations

were considered: weekly, biweekly, triweekly, and monthly, as illustrated in Figure 4.3. In this figure, three hypothetical example MWB timings are provided to illustrate how data would be included, with all time periods preceding an MWB included and periods after the MWB has occurred excluded. In years where an MWB does not occur, all periods would be included. The different input variables for each shown time series configuration are listed in Table 1. These variables include values related to the key river ice conditions preceding an MWB and the climatic conditions at the start of freeze up and directly preceding the period of interest, suggested by Sun (2018) and De Coste et al., (2021). Once the optimum time series configuration was identified, it would be used to train and test the four algorithms to assess the accuracy of the chosen methods.

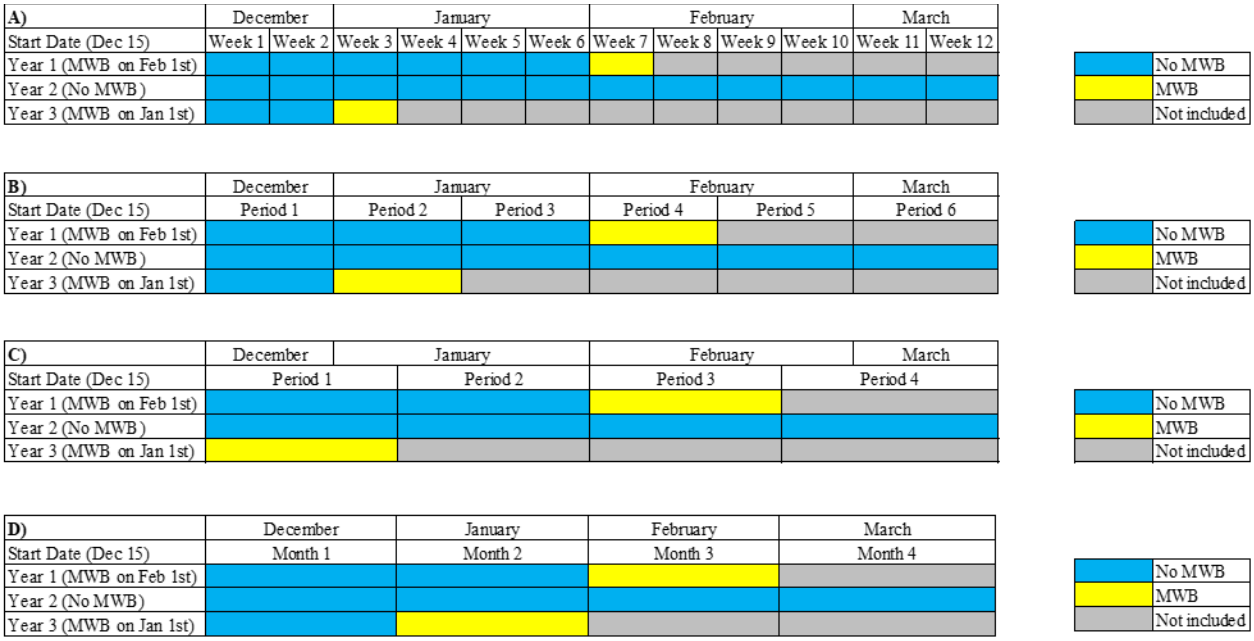


Figure 4.3: Examples of tested modelling time series configurations for A) weekly, B) biweekly, C) triweekly, and D) monthly periods. For the example station, two MWBs occurred during Years 1-3.

Table 4.1: Input variables used for each time series configuration of first level model.

Category	Variables	Time Period	Number of Features
Date	Freeze Up Date, Winter Low Level Date, Winter Low Flow Date	n/a	3
Water Level	Freeze Up Water Level, Winter Low Level Water Level, Winter Low Flow Water Level	n/a	3
Discharge	Freeze Up Discharge, Winter Low Level Discharge, Winter Low Discharge	n/a	3
Temperature	Maximum, Minimum, Mean, AFDD	October, November, December, and configured period preceding period of interest	16
Precipitation	Total Precipitation	October, November, December, and configured period preceding period of interest	4

4.3.2.2 Second Level Model

The second level model focusses on three-day ahead prediction of MWB occurrence within an identified period of interest in the first level model, providing sufficient medium range notice of potential alterations to flow (Le et al., 2019, Jurlina et al., 2019). This model would look at each possible day within the period and classify the MWB occurrence based on data available up to three days ahead of the day of interest. Variables would be limited to climatic variables, as factors such as warming and precipitation preceding the event have been found to be the primary drivers of MWB occurrence (Carr and Vuyovich, 2014). Ice-event related variables (such as those associated with freeze, low winter flow, and low winter level) used as inputs for the first level model would be too far from this event on a time scale and would provide no useful information at this prediction resolution. The initial pool of variables would include daily temperatures, AFDD and changes in AFDD, and daily precipitation and total seasonal precipitation. An input omission process would be used to select the optimum variable set for this prediction. A generalized AdaBoost model would be constructed using all variables to be sure that the effects of each was considered, followed by the exclusion of each variable in turn to see how the recall and balanced accuracy change with the removal. If the values increased or did not change, it would indicate that the variable in question is potentially redundant and can be removed from the model without negative impact. This would be further verified through the testing of the omission of larger groups of potentially redundant variables to reduce the input set to the most relevant values (Snieder et al., 2020). The final input set would then be used to test and train the four algorithms detailed above for the prediction of MWB timing.

4.3.3 Model Implementation

This study utilised Python, version 3.7, for all analysis (Van Rossum and Drake, 2009). Additional analysis was conducted using the packages *Numpy* (Oliphant, 2006), *seaborn* (Waskom, 2008), *Pandas* (McKinney, 2010), *scikit-learn* (Pedregosa et al., 2011), *ensemble* (Narassiguin et al., 2016), *imblearn* (Lemaitre et al., 2017) and *scipy* (Virtanen et al., 2020).

4.4 Results and Analysis

4.4.1 First Level Model

4.4.1.1 Time Series Configuration

The first level model aimed to predict the occurrence of an MWB in a given time period between December 15 to March 15, with the four time periods considered being weekly, biweekly, triweekly, and monthly. The initial number of observations was 4,098 for the monthly configuration, 6,965 for the triweekly configuration, 8,316 for the biweekly configuration, and 10,456 for the weekly configuration. The variables for each configuration were first normalised through division by means, and then balanced between positive and negative cases using under-sampling. An AdaBoost model was trained for each configuration using 5-fold cross-validation with an 80/20 training/testing split. The performance of each configuration using two classification metrics, i.e., balanced accuracy and recall, is presented in Table 4.2. The biweekly model was selected as the most successful of the tested configurations, greatly outperforming the monthly and triweekly models and producing comparable performance to the weekly model, however with a simpler initial dataset.

Table 4.2: Performance of AdaBoost Model for each tested time series configuration.

Period	Balanced Accuracy	Recall
Weekly	0.779	0.771
Biweekly	0.785	0.809
Triweekly	0.677	0.718
Monthly	0.651	0.687

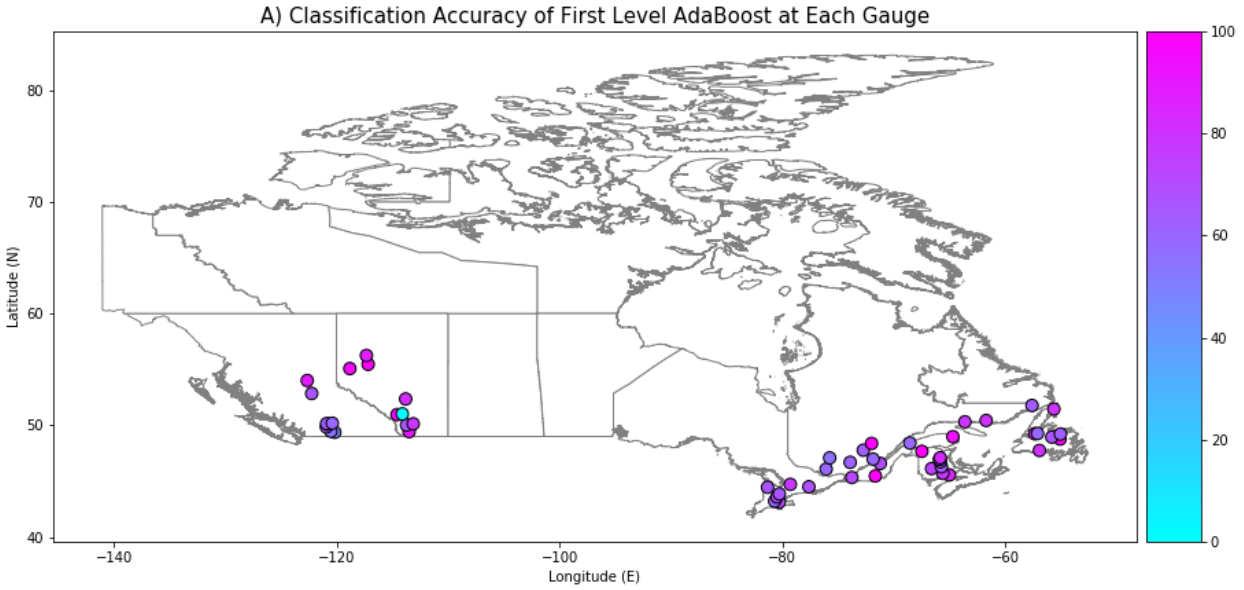
4.4.1.2 First Level Model Results and Analysis

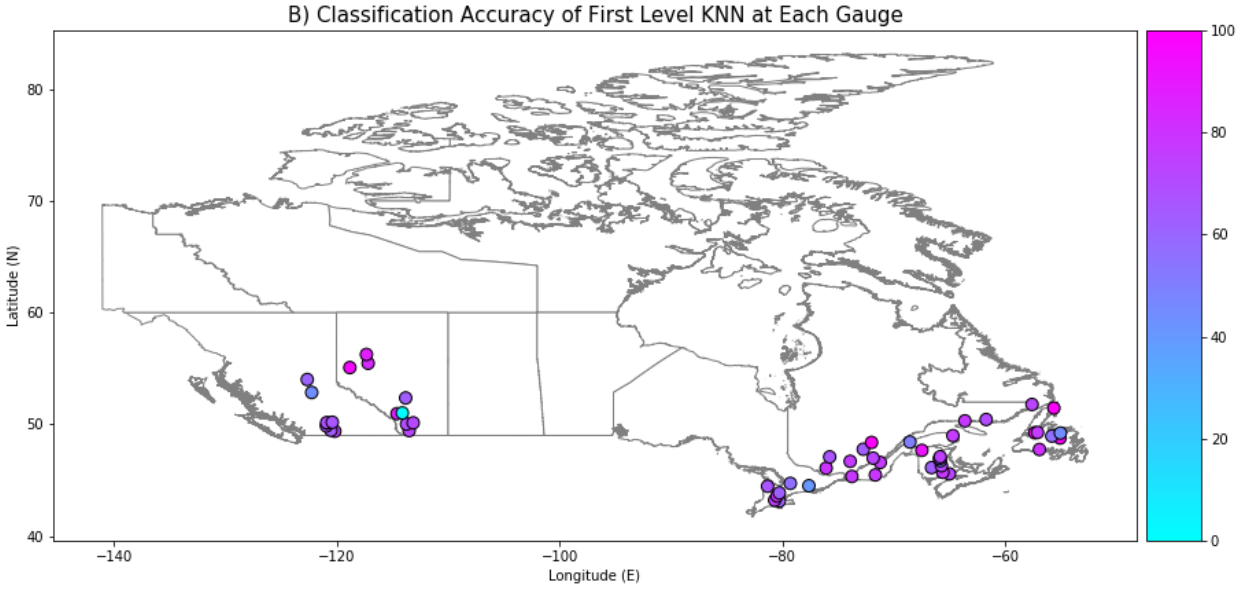
Additional KNN, Class Switching, and ArcX4 models were trained on the same data from the biweekly time series configuration as the AdaBoost model, with the results of each for both classification and probabilistic predictions shown in Table 3. Deterministic classification metrics indicate that the ArcX4 model performed the strongest at binary classification, with the highest obtained balanced accuracy and recall. The ArcX4 also performed well with probabilistic predictions; however, it was slightly outperformed by the Class Switching model in all considered metrics. While the AdaBoost and KNN models performed comparably well to the ArcX4 and Class Switching models in deterministic classification, they were found to have a more significant reduction in accuracy in probabilistic predictions. While their performance was reduced in comparison, each model did achieve an acceptable level of accuracy for each of the probabilistic prediction metrics for a binary classification. The success of each of the models is further demonstrative of the effectiveness of the selected biweekly model configuration, as a change to triweekly or monthly would degrade the obtained results based on the performance of the AdaBoost model, while selection of the weekly configuration would greatly increase data requirements.

Figure 4.4 shows the binary classification accuracy of each model at each gauge. Some of the gauges, particularly those with relatively few MWBs in their flow history like the Elbow River below Glenmore Dam in Alberta, were challenging for some of the developed models but no gauges were completely unclassifiable by all of the developed algorithms, demonstrating both the strength of the selected variables and the trained algorithms in modelling all considered MWBs. Figure 4.5 shows calibration curves for each model, a representation of the calibration error listed in Table 4.3. The smoothest curves were those of the ArcX4 and Class Switching models, which also had the widest ranges of probabilistic prediction values. The AdaBoost model concentrated its probabilistic predictions between 0.4 to 0.6. This is further emphasized in Figure 4.6, which shows the distribution of probabilistic predictions produced by each model, with ArcX4 and Class Switching again having the widest range of predictions followed by KNN, with the AdaBoost model greatly concentrating its predicted probabilities around 50%. This stems from the AdaBoost algorithms tendency to fit predicted probabilities to a transformation of the true probabilities, rather than the actual probabilities (Niculescu-Mizil and Caruana, 2012). Figure 4.7 shows the probabilities predicted by each model against the corresponding correct classification. KNN, ArcX4, and CS performed very well on MWB predictions with the majority of their incorrect predictions falling very close to the 50% cut-off. Conversely, the predictions of the AdaBoost model are so concentrated around 50% that all predictions, whether they are for an MWB or are not, are very close to each other, thus the practical use of these predictions can be called into question.

Table 4.3: Final performance of trained first level algorithms for deterministic and probabilistic predictions.

Model	Deterministic Prediction		Probabilistic Prediction		
	Balanced Accuracy	Recall	Log Loss	Briar Loss	Calibration Error
AdaBoost	0.785	0.809	0.609	0.208	0.128
KNN	0.721	0.838	0.621	0.211	0.098
Class Switching	0.794	0.824	0.514	0.174	0.023
ArcX4	0.801	0.853	0.519	0.176	0.045


A) Classification Accuracy of First Level AdaBoost at Each Gauge
Latitude (N)
Longitude (E)
80
70
60
50
40
−140
−120
−100
−80
−60
100
80
60
40
20
0


B) Classification Accuracy of First Level KNN at Each Gauge
Latitude (N)
Longitude (E)
80
70
60
50
40
−140
−120
−100
−80
−60
100
80
60
40
20
0

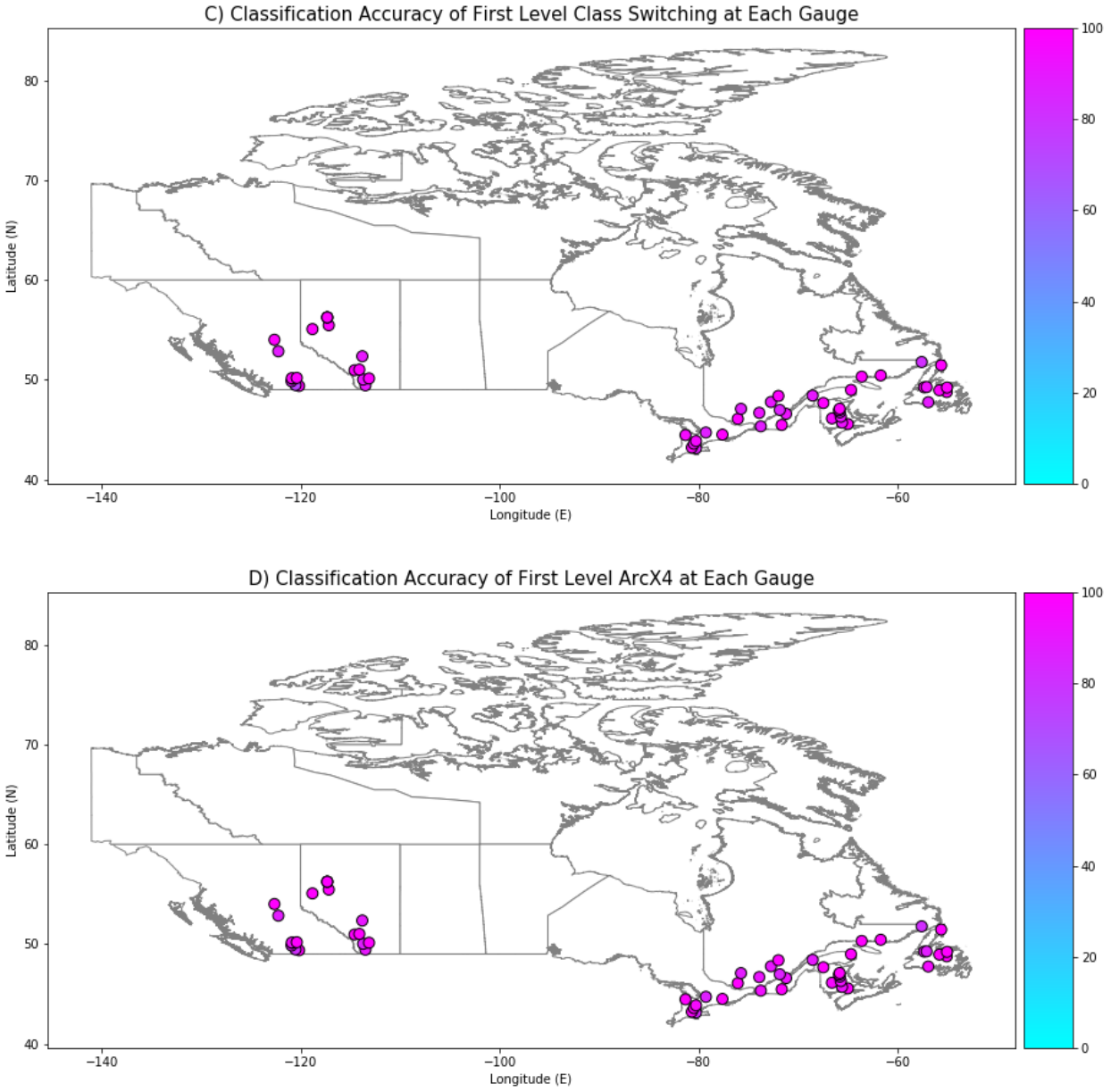


Figure 4.4: Total accuracies of first level models at each considered gauge for A) AdaBoost, B) KNN, C) Class Switching, and D) ArcX4.

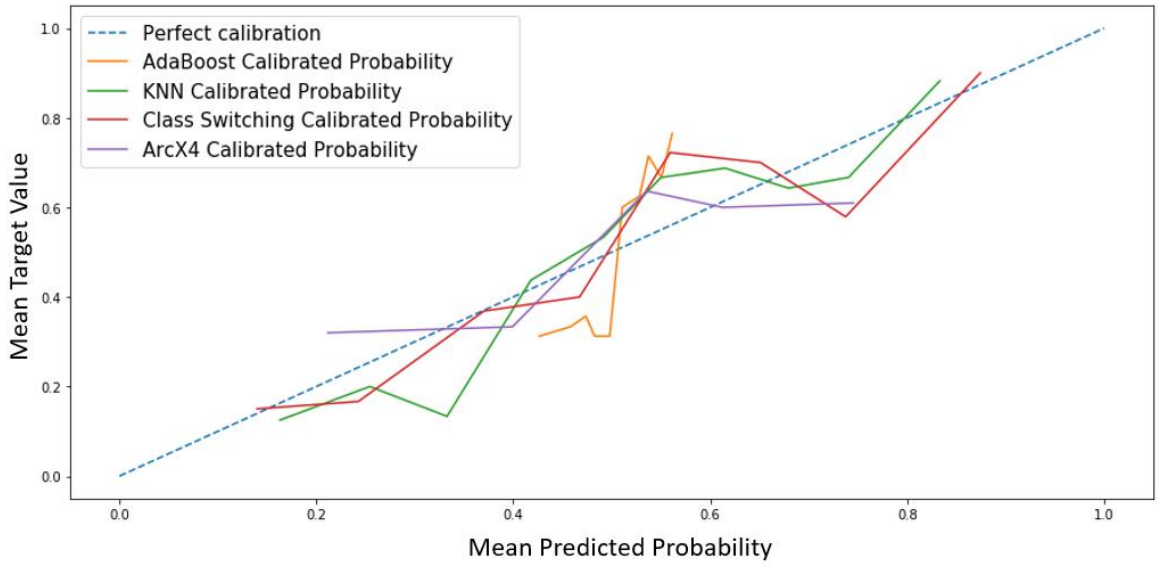


Figure 4.5: Calibration curves for probabilistic predictions for each of the four first level models (Naeini et al., 2015).

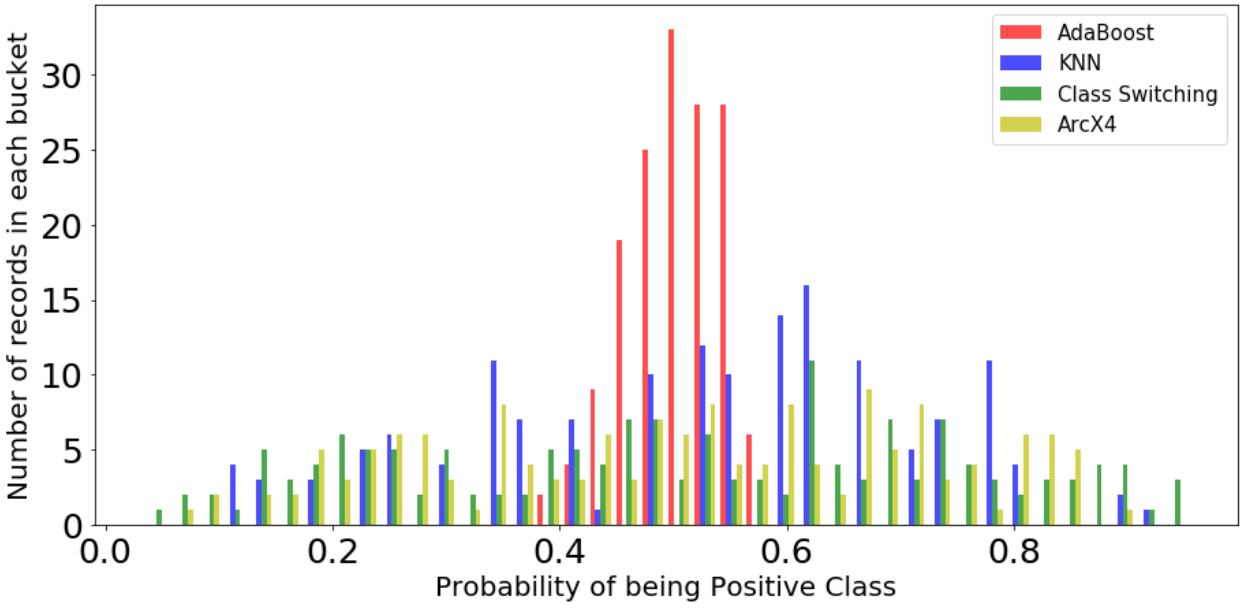
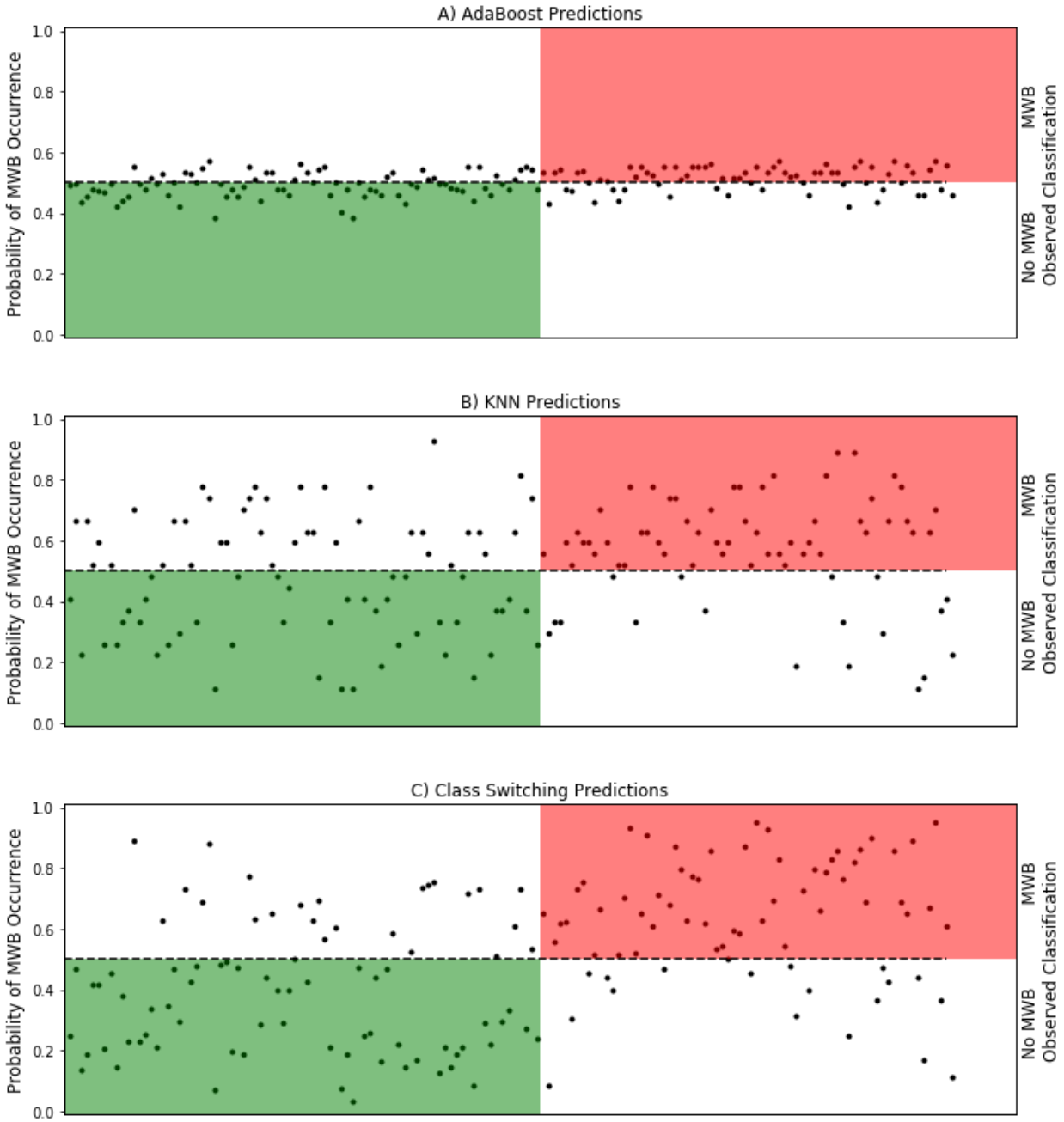


Figure 4.6: Distribution of probabilistic predictions of MWB occurrence for each of the four first level models.


A) AdaBoost Predictions
B) KNN Predictions
C) Class Switching Predictions
Probability of MWB Occurrence
1.0
0.8
0.6
0.4
0.2
0.0
MWB
No MWB
Observed Classification

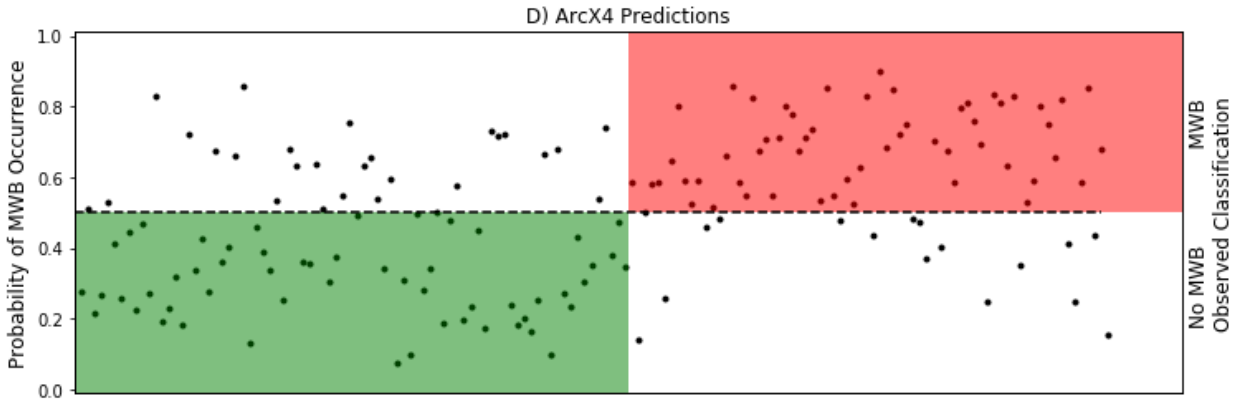


Figure 4.7: Probabilistic predictions from the first level models against the actual classification of each value in the testing set for A) AdaBoost, B) KNN, C) Class Switching, and D) ArcX4. Dots in the green quadrant represent the model properly predicting no MWB, while dots in the red quadrant represent the model properly predicting MWB occurrence.

4.4.2 Second Level Model

4.4.2.1 Variable Selection

The goal of the second level model was to identify if an MWB would occur in 3 days from a day of interest in the MWB period identified by the First Level Model. Input omission was used to test the removal of variables from an initial variable pool listed in Table 4.4. These variables consist of climatic variables representing air temperature and precipitation in the days immediately preceding the day of interest. The input omission process first started with the removal of each variable from the pool one at a time, with the reduced pool then used to train and test an AdaBoost model in a similar fashion to the first level model. The accuracies obtained from the removal of each individual variable were similar, thus no immediate conclusions could be drawn. A subsequent input omission procedure removing combinations of variables from each of the five groups identified in Table 4.4 starting from 15 days before the day of interest and moving closer was then conducted, with the resulting model accuracies shown in Figure 4.8. With these larger

combinations of variable removals tested, it was identified that the majority of considered variables were redundant, and the dataset was reduced to the variables shown in Table 4.5, which were the main variables whose removal began to show larger reductions in accuracy and recall in Figure 4.8. The selected variables were those illustrating short-term changes in the days immediately preceding an MWB, with AFDD and daily precipitation excluded entirely. These short-term changes in trends can be inferred to have a stronger impact on MWB occurrence than those over a longer timespan.

Table 4.4: Initial considered variables for the second level model.

Category	Variables	Number of Features
Daily Temperature	T from 1-15 days before	15
AFDD	AFDD from 1-15 days before	15
Change in AFDD	DAFDD from 1-15 days before	15
Precipitation	P from 1-15 days before	15
Total Precipitation	TOTP from 1-15 days before	15

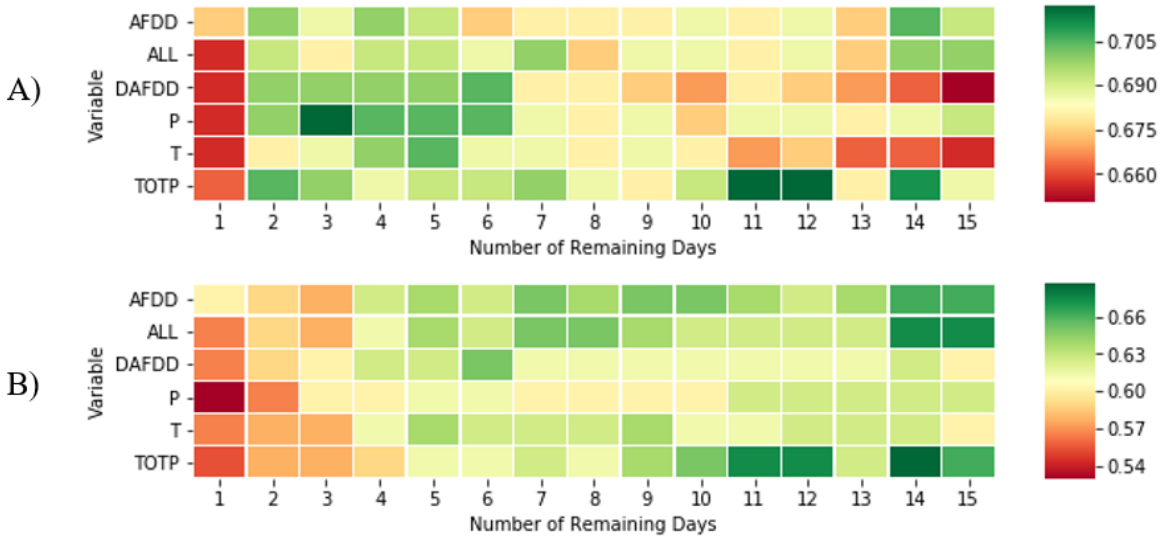


Figure 4.8: Heatmap of variable removal accuracies where A) is balanced accuracy and B) is recall.

Table 4.5: Final selected inputs for second level models.

Category	Variables	Number of Features
Daily Temperature	T from 1-3 days before	3
Change in AFDD	DAFDD from 1-3 days before	3
Total Precipitation	TOTP from 1-3 days before	3

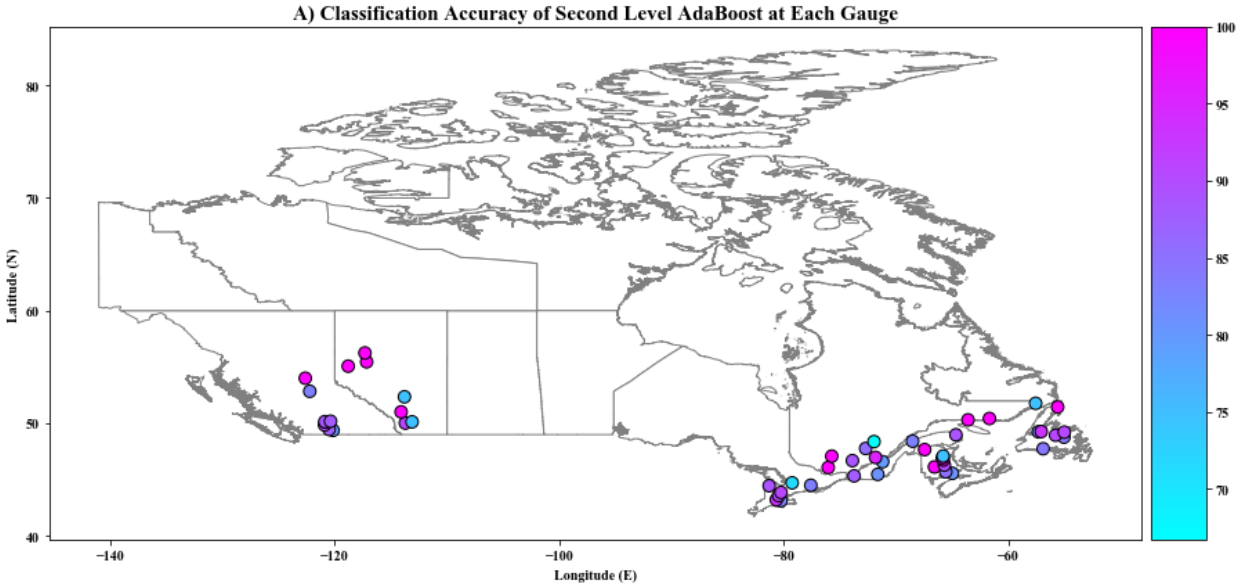
4.4.2.2 Second Level Model Performance

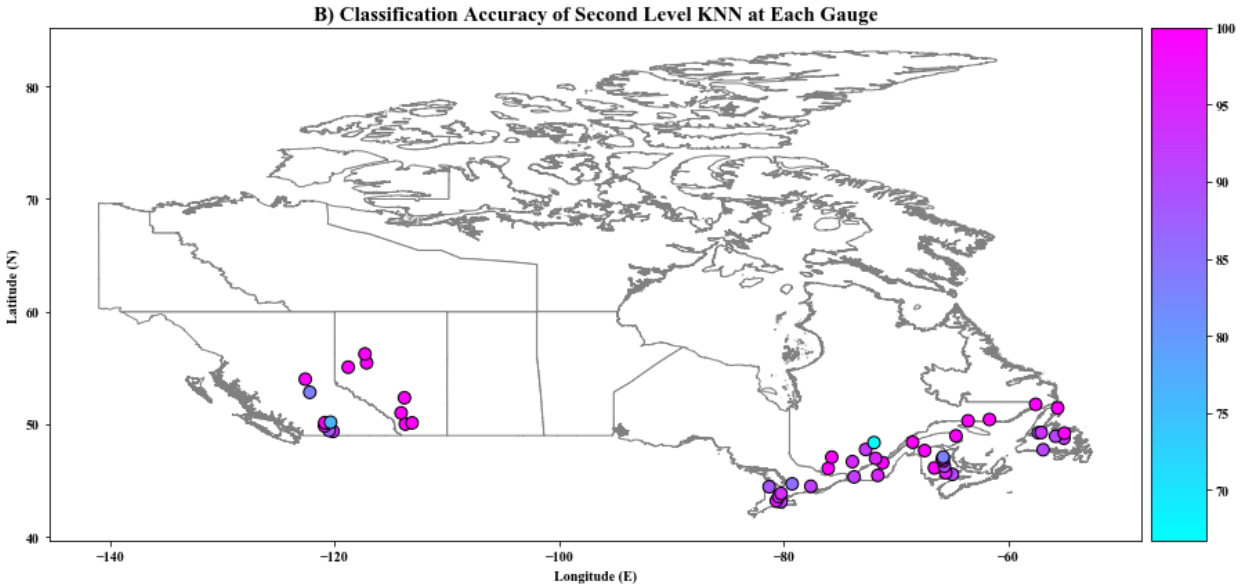
Data used for the second level models was first normalised and balanced using under-sampling in the same fashion as the first level model, with an 80/20 training/testing split. An AdaBoost, KNN, Class Switching, and ArcX4 model was trained separately using the prepared data, with the results of each model for both deterministic classification and probabilistic prediction detailed in Table 4.6. The obtained values for the considered performance metrics

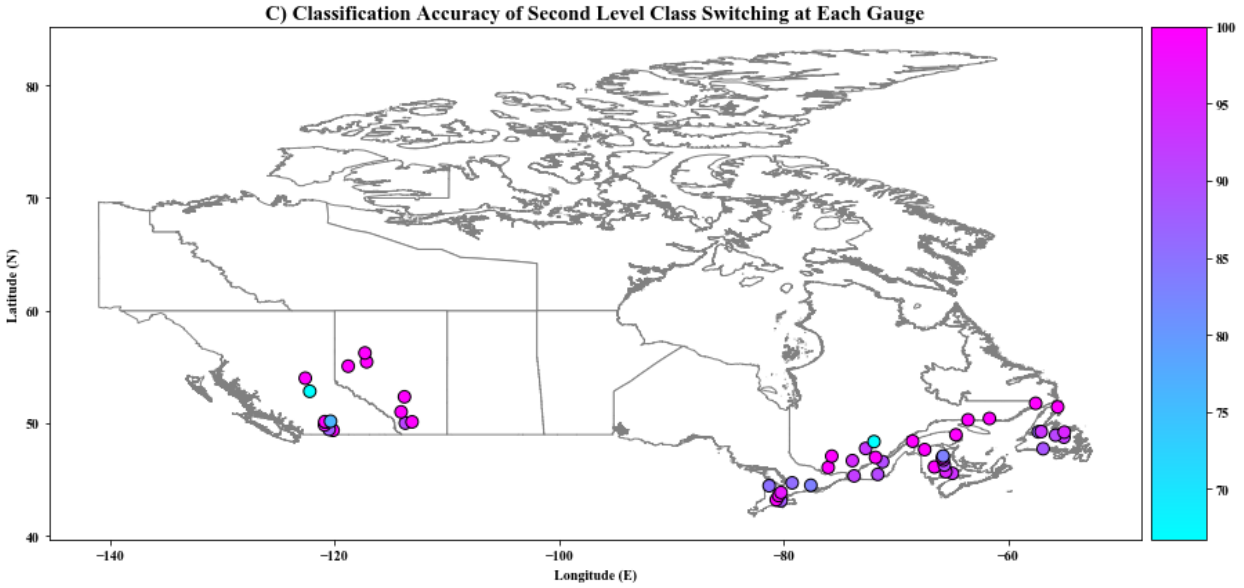
indicate that the KNN model had the highest accuracy for binary classification with the other three models performing quite similarly. The Class Switching and ArcX4 models achieved the highest performance in probabilistic predictions, with KNN and AdaBoost performing slightly worse in both Briar loss and calibration error. Notably, both the KNN and AdaBoost fall outside of the acceptable range of 0.675 for log loss score for a binary calibration. Similar to the first level model, Figure 4.9 shows the classification accuracies of each model at each gauge. Much like the first level model, the majority of the gauges were properly classified by at least one model, demonstrating the models strength and the importance of the selected variables. The calibration curves of each model for probabilistic predictions are shown in Figure 4.10. The smoothest curves were those of the ArcX4 and Class Switching models, which also had the widest ranges of probabilistic prediction values, while the KNN model had a smaller range. The values of the AdaBoost model were heavily concentrated at 0.5, resulting in a curve very far off from a perfect calibration. This shortcoming in predictions is further emphasized in Figure 4.11, which shows the distribution of probabilistic predictions of each model, with all values from the AdaBoost model being concentrated at 0.5. Figure 4.12 shows the probabilistic predictions against the correct classification. KNN, ArcX4, and Class Switching again performed well in predicting both classes, with most incorrect predictions very close to the 50% cut-off. The AdaBoost model produced predictions so concentrated around 50% that there is again very little distinction between what the model is classifying between the two classes.

Table 4.6: Final performance of trained second level algorithms for deterministic and probabilistic predictions.

Model	Deterministic Prediction		Probabilistic Prediction		
	Balanced Accuracy	Recall	Log Loss	Briar Loss	Calibration Error
AdaBoost	0.767	0.773	1.461	0.206	0.116
KNN	0.773	0.8	1.596	0.204	0.092
Class Switching	0.746	0.76	0.597	0.184	0.056
ArcX4	0.76	0.773	0.621	0.175	0.072


A) Classification Accuracy of Second Level AdaBoost at Each Gauge
Latitude (N)
Longitude (E)
80
70
60
50
40
−140
−120
−100
−80
−60
100
95
90
85
80
75
70


B) Classification Accuracy of Second Level KNN at Each Gauge
Latitude (N)
Longitude (E)
80
70
60
50
40
−140
−120
−100
−80
−60
100
95
90
85
80
75
70



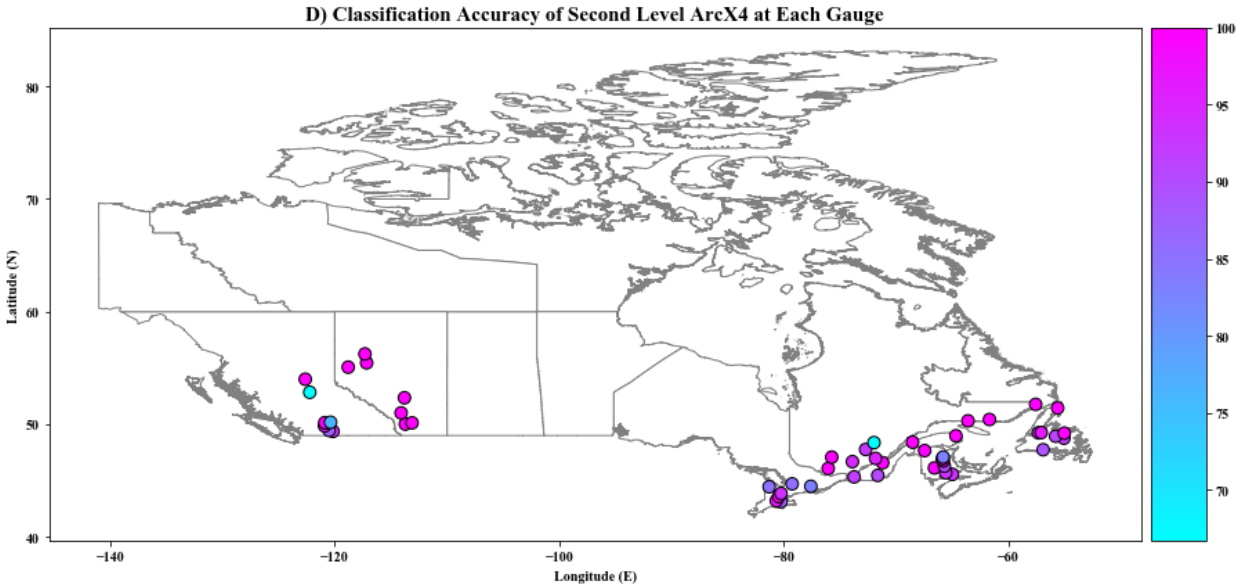


Figure 4.9: Total accuracies of second level models at each considered gauge for A) AdaBoost, B) KNN, C) Class Switching, and D) ArcX4.

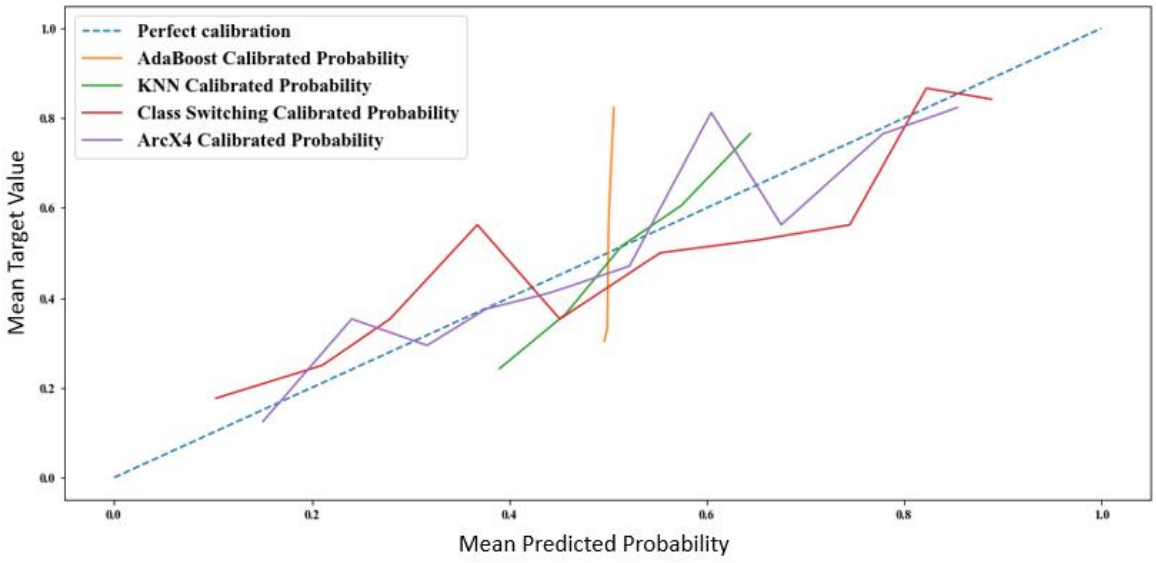


Figure 4.10: Calibration curves for probabilistic predictions for each of the four second level models (Naeini et al., 2015).

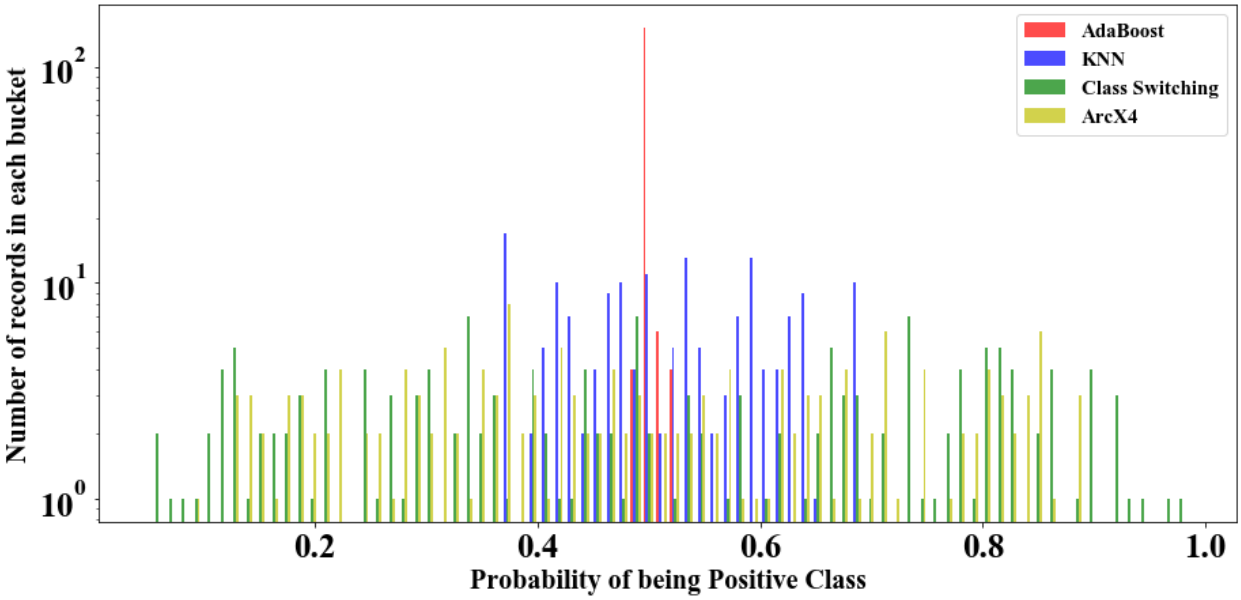


Figure 4.11: Distribution of probabilistic predictions of MWB occurrence for each of the four first level models.

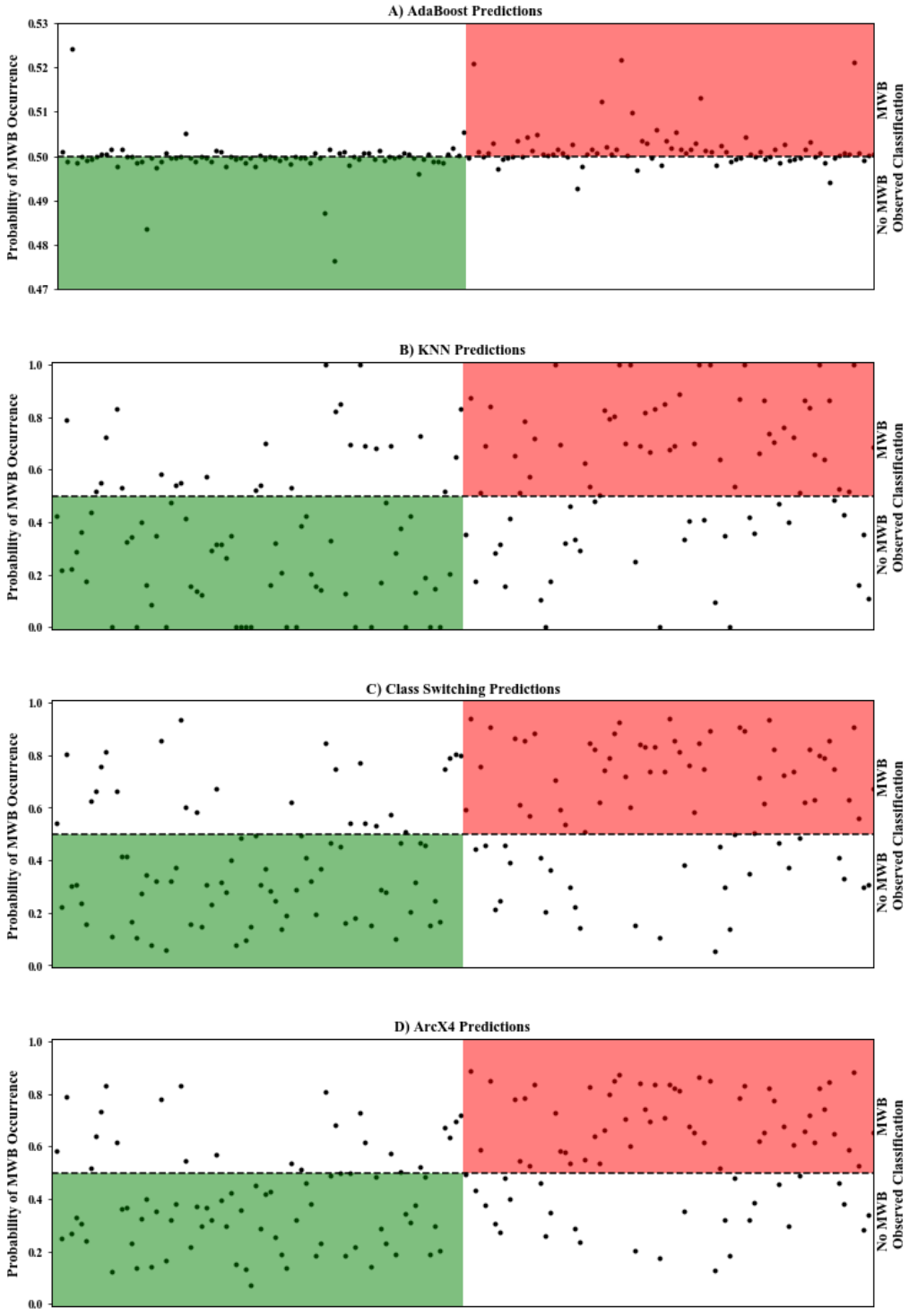


Figure 4.12: Probabilistic predictions from the second level models against the actual classification of each value in the testing set for A) AdaBoost, B) KNN, C) Class Switching, and

D) ArcX4. Dots in the green quadrant represent the model properly predicting no MWB, while dots in the red quadrant represent the model properly predicting MWB occurrence.

4.4.3 Discussion

At both of the developed model levels, the ArcX4 and Class Switching models achieved the best performance. The other two models obtained comparable performance in deterministic classification but showed a deterioration when probabilistic results were considered. Though the values of the performance metrics for the AdaBoost model were comparable to the KNN, the AdaBoost models had significant issues in the distribution of predicted probabilities, with their predictions so concentrated around 0.5. Because of this, shift in the classification threshold, which could be possible for other models, would ruin its results, especially in the second level, unless additional calibration of the results was performed, which is not necessary for the other models. Considering both the deterministic and probabilistic predictions can provide a clearer picture of the accuracy of the developed models because of factors such as this. On a national scale, the ArcX4 performs well, with no clear examples of a gauge being difficult to classify on both levels, while also obtaining some of the highest performance of the considered models for both prediction methods. Some spatial variability was present in the results of the second level model, with select gauges in British Columbia and Maritime provinces having slightly reduced accuracy, stemming from a combination of variations in the amounts of data available for these gauges and the milder climates of these regions. However, despite the lowered performance at some of these locations, the minimum balanced classification accuracy for this model was 68%, indicating an overall reliable performance.

The utilised two-level structure also provides versatility in predictions. Predicting the date of an MWB from the start of the season would be far less accurate based on the degrading performance of the models used to test the time series configuration in the first level as the time step increased in size. Conversely, predicting MWB occurrence using solely the second level model would be very data intensive as it would have to be run for every possible day in the considered period of MWB vulnerability. The two-level structure reduces the data complexity while increasing the model accuracy. Though the model accuracy of each level was assessed independently due to the differing timescales, objectives, and input variables at each model level, the results obtained are very promising, especially given the large spatial scale of the data used in the analysis. Due to the ease of configuration, both binary classification and probabilistic predictions are easily output from the models, with errors and shortcomings of both prediction methods well known. This can provide strong decision-making support for the affected rivers, producing predictions of occurrence with associated probabilities rather than just a binary prediction, allowing the strength of the prediction to be considered alongside the value of the prediction. Additionally, the methods used were successfully tested on seasons without any MWBs, thus the models could be applied to rivers that have not yet experienced MWBs but may encounter them in the future.

4.5.0 Conclusions

This study focussed on the development of a modelling methodology for the prediction of MWBs on a national scale, an event becoming increasingly common due to climate change and having an affect on both ecosystems and communities. An extensive database on river ice events in Canada, the CRID, along with gridded climate data were used to build data-driven models predicting both the occurrence and timing of these events on 52 rivers with historical MWBs across

Canada. Due to the rarity of these events and the subsequent issues introduced by the resulting imbalanced dataset, modifications were implemented throughout model construction to ensure the best results. These changes affected the data preprocessing, algorithm selection, and performance metric selection to strengthen the results against imbalanced data issues. The developed framework utilised a two-level structure, with multiple dataset configurations tested for the first level model, with detailed input selection performed for the second level model. The finalized model included a first level predicting the occurrence of MWBs within a given two-week period, and a second level predicting the timing within that period with a three-day lead time. Several machine-learning algorithms with both deterministic and probabilistic outcomes were considered with the objective of achieving best accuracy in the final models.

The final recommended model, utilising an ArcX4 algorithm at each model level, produced a high level of accuracy given the timescale of the predictions. This model was easily modified to output probabilistic and classification results, where the model performance was similar in both cases. This was not the case for the other considered models, which were found to have unacceptable performances from a probabilistic standpoint. The two-level structure was demonstrated to be a versatile method of predicting MWBs, avoiding the issues of degrading performance with increasing timescale that would be introduced using solely the first level model, while providing a far less data intensive method than would be required using solely the second level. The models themselves were tested both on years with MWBs as well as years where they did not occur, demonstrating the capacity for the model to be applied to rivers that have not yet experienced MWBs. The models are dependent solely on climatic, hydrometric and river ice data that can be monitored in advance of events, allowing for ease of use in forecasting. The model was constructed at a national scale not previously possible because of data scarcity, with the possibility

of easy reconfiguration for other northern regions vulnerable to MWBs. Future work may include detailed investigation of the transferability of this method, requiring data from other vulnerable regions. Further testing of additional machine learning algorithms and using additional climate data, such as snowpack depths and historical weather forecasts that can further increase the dependability of predictions, will also be considered.

Acknowledgements

This study was supported by Natural Sciences and Engineering Research Council of Canada (NSERC). Data used is available from Environment and Climate Change Canada (https://open.canada.ca/data/en/dataset/c5b58ccd-0011-4a80-8f24-034c86cbc14d) and Natural Resources Canada (https://cfs.nrcan.gc.ca/projects/3/4).

References

Barzegar, R., Ghasri, M., Qi, Z., Quilty, J., and Adamowski, J., 2019. Using bootstrap ELM and LSSVM models to estimate river ice thickness in the Mackenzie River Basin in the Northwest Territories, Canada. Journal of Hydrology, 577.

Beltaos, S., 1990. Guidelines for extraction of ice break-up data from hydrometric station records. Working Group on River Ice Jams: Field Studies and Research Needs. NHRI Science Report No. 2, National Hydrology Research Institute, Environment Canada, Saskatoon, SK, Canada, pp. 37–70.

Beltaos, S., 1999. Climatic effects on the changing ice-breakup regime of the Saint John River. Proceedings of the 10th Workshop on the Hydraulics of Ice Covered Rivers, Guidelines for Extraction of Ice Break-up Data from Hydrometric Station Records. Winnipeg, Canada, 251-264.

Beltaos, S., 2003. Threshold between mechanical and thermal breakup of river ice cover. Cold Regions Science and Technology, 37: 1–13.

Beltaos, S., Ismail, S., and Burrell, B., 2003. Midwinter breakup and jamming on the upper Saint John River: a case study. Canadian Journal of Civil Engineering, 30: 77–88.

Beltaos, S., Prowse, T., Bonsal, B., MacKay, R., Romolo, L., Pietroniro, A., and Toth, B., 2006. Climatic effects on ice-jam flooding of the Peace-Athabasca Delta. Hydrological Processes, 20: 4031-4050.

Beltaos, S., and Prowse, T., 2009. River-ice in a shrinking cryosphere. Hydrological Processes, 23: 122-144.

Beltaos, S., 2017. Frequency of ice jam flooding of Peace Athabasca Delta. Canadian Journal of Civil Engineering, 45: 71-75.

Boyd, d., 1979. Degree days: The different types. Building research note, National Research Council of Canada from the Atmospheric Environment Service Department of Fisheries and Environment.

Breiman, L., 1996. Bias, variance, and arcing classifiers. Technical Report, 460:1-25.

Breiman, L., 2000. Randomizing outputs to increase prediction accuracy. Machine Learning, 40: 299-242.

Brodersen, K., Ong, C., Stephan, K., and Buhmann, J., 2010. The balanced accuracy and its posterior distribution. 20th International Conference on Pattern Recognition.

Carr, M., and Vuyovich, C., 2014. Investigating the effects of long-term hydro-climatic trends on Midwest ice jam events. Cold Regions Science and Technology, 106-107: 66-81.

Chawla, N., Bowyer, K., Hall, L., and Kegelmeyer, W., 2002. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 321-357.

De Coste, M., Li, Z., Pupek, D., and Sun, W., 2021. A hybrid ensemble modelling framework for the prediction of breakup ice jams on Northern Canadian Rivers. Cold Regions Science and Technology, 189: 103302.

de Rham, L., Prowse, T., Beltaos, S., and Lacroix, M., 2008a. Assessment of annual high-water events for the Mackenzie river Basin, Canada. Hydrological Processes. 22: 3864-3880.

de Rham, L., Prowse, T., and Bonsal, B., 2008b. Temporal variations in river-ice break-up over the Mackenzie River Basin, Canada. Journal of Hydrology, 349: 441-454.

de Rham, L., Dibike, Y., Beltaos, S., Peters, D., Bonsal, B., and Prowse, T., 2020. A Canadian river ice database from national hydrometric program archives. Earth System Science Data, Open Access Discussions.

Dudani, S., 1976. The distance-weighted k-Nearest-Neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, 6: 325-327.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., and Herrera, F., 2012. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 42*:* 463–484.

Goulding, H., Prowse, T., and Bonsal, B., 2009. Hydroclimatic controls on the occurrence of break-up and ice-jam flooding in the Mackenzie Delta, NWT, Canada. Journal of Hydrology, 379: 251-267.

Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K., 2003. KNN Model-based approach in classification. Lecture notes in Computer Science, 2888: 986-996.

Guo, X., Wang, T., Fu, H., Guo, Y., and Li, J., 2018. Ice-jam forecasting during river breakup based on neural network theory. Journal of Cold Regions Engineering, 32(3): 04018010.

Haixiang, G., Yijing, L, Shang, J., Mingyun, G., Yuanyue, H., and Bing, G., 2017. Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications, 73: 220-239.

Hopkinson, R., McKenney, D., Milewska, E., Hutchinson, M., Papadopol, P., and Vincent, L., 2011. Impact of aligning climatological day on gridding daily maximum-minimum temperature and precipitation over Canada. Journal of Applied Meteorology and Climatology, 50: 1654-1665.

Huntington, T., Hodgkins, G., and Dudley, R. 2003. Historical trend in river ice thickness and coherence in hydroclimatological trends in Maine. Climate Change 61: 217–236.

Hutchinson, M., McKenney, D., Lawrence, K., Pedlar, J., Hopkinson, R., Milewska, E., and Papadopol, P., 2009. Development and testing of Canada-wide interpolated spatial models of daily minimum-maximum temperature and precipitation for 1961-2003. American Meteorological Society, 48: 725-741.

Jasek, M., 2019a. An emerging picture of Peace River break-up types that influence ice jam flooding of the Peace-Athabasca Delta, part 1: the 2018 Peace River breakup. In: Proceedings of the 20th Workshop on the Hydraulics of Ice Covered Rivers, Ottawa, ON, Canada.

Jasek, M., 2019b. An emerging picture of Peace River break-up types that influence ice jam flooding of the Peace-Athabasca Delta part 2: Insights from the comparison of the 2014 and 2018 break-ups. In: Proceedings of the 20th Workshop on the Hydraulics of Ice Covered Rivers, Ottawa, ON, Canada.

Jurlina, T., Baugh, C., Pappenberger, F., and Prudhomme, C., 2019. Flood hazard risk forecasting index (FHRFI) for urban areas: The Hurricane Harvey case study. Meteorological Applications, 27: e1845.

Koziarski, M., 2020. Radial-Based undersampling for imbalanced data classification. Pattern Recognition, 102: 107262.

Kumar, N., Rao, K., Govardhan, A., Reddy, K., and Mahmood, A., 2014. Undersampled K-means approach for handling imbalanced distributed data. Progress in Artificial Intelligence, 3: 29–38.

Lamontagne, J., Jasek, M., and Smith, J., 2021. Coupling physical understanding and statistical modeling to estimate ice jam flood frequency in the norther Peace-Athabasca Delta under climate change. Cold Regions Science and Technology, 192: 103383.

Le, X., Ho, H., Lee, G., and Jung, S., 2019. Application of Long Short-Term Memory (LSTM) neural network for flood forecasting. Water, 11:1387.

Lemaitre, G., Noguiera, F., and Aridas, C., 2017. Imbalanced-Learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. Journal of Machine Learning Research, 18: 1-5.

Li, Q., Yang, B., Li, Y., Deng, N., and Jing, L., 2013. Constructing support vector machine ensemble with segmentation for imbalanced datasets. Neural Computing and Applications, 22: 249–256.

Li, J., Liu, L., Fong, S., Wong, R., Mohammed, S., and Fiaidhi, J., 2016. Adaptive swarm balancing algorithm for rare-event prediction in imbalances healthcare data. Computerized Medical Imaging and Graphics, 12(7).

Liu, S., xu, J., Zhao, J., Xie, X., and Zhang, W., 2014. Efficiency enhancement of a process0based rainfall-runoff model using a new modified AdaBoost.RT technique. Applied Soft Computing, 23: 521-529.

Liu, M., Huang, Y., Li, Z., Tong, B., Liu, Z., Sun, M., Jiang, F., and Zhang, H., 2020. The applicability of LSTM-KNN model for real-time flood forecasting in different climate zones in China. Water, 12:440.

Lopez, F., Garcia, S., Palade, V., and Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences, 250: 113-141.

Loyola-González, O., Martínez-Trinidad, J., Carrasco-Ochoa, J., and García-Borroto, M., 2016. Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases. Neurocomputing, 175: 935–947.

Maalouf, M., and Trafalis, T., 2011. Robust weighted kernel logistic regression in imbalanced and rare events data. Computational Statistics and Data Analysis, 55: 168-183.

Mahabir, C., Hicks, F., and Fayek, A.R., 2006. Neuro-fuzzy river ice breakup forecasting system. Cold Regions Science and Technology. 46:100-112.

Martinez-Munoz, G., and Suarez, A., 2005. Switching class labels to generate classification ensembles. Pattern Recognition, 38: 1483-1494.

Massie, D., White, K., and Daly, S., 2002. Application of neural networks to predict ice jam occurrence. Cold Regions Science and Technology. 35:115-122.

McKenney, D., Hutchinson, M., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., Milewska, E., Hopkinson, R., Price, D., and Owen, T. 2011. Customized spatial climate models for North America. Bulletin of American Meteorological Society-BAMS December: 1612-1622.

McKinney, W., and others, 2010. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 445: 51-56.

Naeini, M., Cooper, G., and Hauskrecht, M., 2015. Obtaining well calibrated probabilities using Bayesian Binning. Proceedings of the AAAI Conference on Artificial Intelligence, 2015: 2901-2907.

Napierala, K., and Stefanowski, J., 2015. Types of minority class examples and their influence on learning classifiers from imbalanced data. Journal of Intelligent Information Systems, 1–35.

Narassiguin, A., Bibimoune, M., Elghazel, H., and Aussem, A., 2016. An extensive comparison of ensemble learning methods for binary classification. Pattern Analysis Applications, 19:1093-1128.

Newton, B., Prowse, T., and de Rham, L., 2017. Hydro-climatic drivers of mid-winter break-up of river ice in western Canada and Alaska. Hydrology Research 48.4: 945-956.

Niculescu-Mizil, A., and Caruana, R., 2005. Predicting good probabilities with supervised learning. In Proceedings of the 22nd International Conference on Machine Learning.

Niculescu-Mizil, A., and Caruana, R., 2012. Obtaining calibrated probabilities from boosting. In Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence.

Oliphant, T., 2006. A guide to NumPy (Vol. 1). Trelgol Publishing USA.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and others, 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Prowse, T., Bonsal, B., Lacroix, M., and Beltaos, S., 2002. Trends in river-ice breakup and related temperature controls. Proceedings of the 16th IAHR International Symposium on Ice. Trends in River-Ice Breakup and Related Temperature Controls. Dunedin, New Zealand, 64-71.

Ratsch, G., Onoda, T., and Muller, K., 2001. Soft margins for AdaBoost. Machine Learning, 42: 287-320.

Refaeilzadeh, Payam, Lei Tang, and Huan Liu. "Cross-validation." Encyclopedia of database systems 5 (2009): 532-538.

Roulston, M., 2007. Performance targets and the Brier Score. Meteorological Applications, 14: 185-194.

Semenova, N., Sazonov, A., Krylenko, I., and Frolova, N., 2020. Use of classification algorithms for the ice jams forecasting problem. E3S Web of Conferences, 163.

Snieder, E., Shakir, R., and Khan, U., 2020. A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. Journal of Hydrology, 583: 124299.

Sun, W., and Trevor, B., 2015. A comparison of fuzzy logic models for breakup forecasting of the Athabasca River. CRIPE 18th Workshop on the Hydraulics of Ice Covered Rivers, Quebec City, QC, Canada, August 18-20, 2015.

Sun, Z., Song, Q., Zhu, X., Sun, H., Xu, B., and Zhou, Y., 2015. A novel ensemble method for classifying imbalanced data. Pattern Recognition, 48: 1623–1637.

Sun, W., 2018. River ice breakup timing prediction through stacking multi-type model trees. Science of the Total Environment, 644:1190-1200.

Sun, W., and Trevor, B., 2018. A stacking ensemble learning framework for annual river ice breakup dates. Journal of Hydrology. 561:636-650.

Tahir, M., Kittler, J., Mikolajczyk, K., and Yan, F., 2009. A multiple expert approach to the class imbalance problem using inverse random under sampling. Multiple Classifier Systems, 82-91.

Timoney, K., 2002. A dying delta? A case study of a wetland paradigm. Wetlands, 22: 282.

Timoney, K., Smith, J., Lamontagne, J., and Jasek, M., 2018. Discussion of "Frequency of ice-jam flooding of Peace-Athabasca Delta". Canadian Journal of Civil Engineering, 46 (3).

Van Rossum, G., and Drake, F., 2009. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S., Brett, M., Wilson, J., Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, I., Feng, Y.,

Moore, E., VanderPlas, J., Laxalde, D., Pertkold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17: 261-272.

Vovk, V., 2015. The fundamental nature of the log loss function. Fields of Logic and Computation II, 307-318.

Wang, J., Sui, J., Guo, L., Karney, B., and Jüpner, R., 2010. Forecast of water level and ice jam thickness using the back propagation neural network and support vector machine methods. International Journal of Environmental Science & Technology, 7(2): 215 – 224.

Wang, T., Yang, K.L., Guo, X.L., and Fu, H., 2012. Application of adaptive network based fuzzy inference system to ice condition forecast. Journal of Hydraulic Engineering, 1:18.

Waskom, M., and the seaborn development team, 2020. Seaborn. Zenodo, https://doi.org/10.5281/zenodo.592845

White, K., 1996. Predicting breakup ice jams using logistic regression. Journal of Cold Regions Engineering, 10(4):178-189.

Zhao, L., Hicks, F., and Fayek, A., 2012. Applicability of multilayer feed-forward neural networks to model the onset of river breakup. Cold Regions Science and Technology, 70:32-42.

Zhang, Z., Ma, C., Xu, J., Huang, J., and Li, L., 2014. A novel combinational forecasting model of dust storms based on rare classes classification algorithm. Communications in Computer and Information Science, 482: 520-537.

Zhu, J., Zou, H., Rosset, S., and Hastie, T., 2009. Multi-class AdaBoost. Statistics and its Interface, 2:349-360.

Chapter 5: A Hybrid Ontology-Based and Machine Learning Model for the Prediction of Spring Breakup on Canadian Rivers

The management of complex data in the prediction of river ice breakups is a challenge in many studies, with the presentation of results in a user-friendly way often hampered by the resulting difficulties in organization and delivery. In this study, an ontology-based semantic model of the domain of an ice season is proposed, providing a way to organize data and describe relationships between the key events present in the formation and development of river ice cover. This ontology is used in a novel modelling system in conjunction with machine learning models to predict the timing of spring breakup on these rivers. The ontology allows the application of network analysis techniques in the selection of variables in the machine learning algorithms, producing predicting results that exceed the accuracy of similar models developed without application of the hybrid system. This is the first application of these new techniques and provides an avenue for further analysis of both river ice phenomena and other domain systems in general.

Michael De Coste was responsible for the investigation, development, and implementation of the models, and results validation under the supervision of Dr. Zhong Li and Dr. Ridha Khedri, who provided additional guidance on the development of the hybrid modelling system. The presented manuscript was drafted and prepared by Michael De Coste and revised by Michael De Coste, Dr. Zhong Li, and Dr. Ridha Khedri.

This chapter has been submitted for publication in the journal *Computer-Aided Civil and Infrastructure Engineering*.

Abstract

River ice breakup is a yearly event on the majority of rivers in Canada. These events carry the potential for high flows and flooding and are therefore of great interest to accurately predict. A challenge in developing models for the prediction of these events, especially on a national scale, is the management of the massive amounts of data associated with a typical ice season. Physically based models require case-by-case calibration, which is data and time intensive, and there is no generalized approach for the prediction of these events. This study couples the unique capabilities of semantic-based and machine-learning-based models in developing a new hybrid modelling framework capable of predicting spring breakup on a national scale. The semantic model uses ontology web language to describe the relationships between the key events and measurements taken throughout the ice season, producing an Ice Season Ontology that sorts the data and allows for a user-friendly means of analyzing any ice season of interest. The structure of the ontology is used to further analyse the relationships between the ice season features (i.e., variables), providing an insight on which of the datapoints are most central to the typical ice season and conversely which are the least. With this, a refined variable selection is able to be made for the machine learning model, for which several algorithms were tested. The most successful developed model, a random forest, was able to predict the timing of spring breakups with a mean absolute error of 10.85 days and an R^2 of 0.884, a high level of accuracy for the long and varying lead time of the produced predictions. This new modelling framework combines the first ever ontology developed for river ice with data-driven modelling techniques in a novel way, allowing for accurate and easily understood prediction of spring breakup throughout Canada, providing a means for decision-making support for river bound communities.

Keywords: River Ice; breakup; semantic modelling; ontology; machine learning; data-driven modelling; ontology-based analytics

5.1 Introduction

The spring breakup of river ice cover is an expected event on many rivers throughout northern regions, often occurring after the initial spring thaw begins. As temperatures rise, snow melt increases, resulting in an increase in runoff into rivers found in the same basin. As this occurs, one of two breakup mechanisms can result: thermal or dynamic (Beltaos, 1997). Thermal breakup is the gradual melt of an ice cover in place and is more often linked to slow increases in air temperature and gradual increases in snow melt runoff and water level increases. Conversely, dynamic breakup is brought about by faster increases in air temperature and runoff flows, leading water levels rising faster than the ice can melt and lifting out of place (Beltaos, 2003). In the case of dynamic breakup, ice jamming can result, where the free flowing ice cover can become trapped at locations where channel geometry changes, potentially resulting in severe flooding (Lindenschmidt et al., 2016). While dynamic breakups can have potentially more severe consequences, both mechanisms play important parts in the flow regime, morphology, and environmental characteristics of the river in questions, and also impact any communities that are present along the river (Beltaos and Prowse, 2009). With this in mind, it is of great interest to be able to accurately predict the timing of these events with sufficient lead times to provide decision making support. Challenges in the prediction of these events, especially on larger regional or national scales, are often linked to data availability and management.

The processes leading to spring breakup are generally well researched and understood. The primary drivers are the changes in air temperature and flow occurring during the spring thaw (Beltaos et al., 2006), with trends in the spatial distributions and timing of breakups moving towards earlier and longer breakup periods (de Rham et al., 2008a and de Rham et al., 2008b). Geomorphological factors also play a role, with the width, slope, and shape of channels playing a

factor in the breakup mechanism that can be expected (Turcotte and Morse, 2013). Globally, the extent and duration of the river ice season is projected to decrease between the present and 2100, with season lengths decreasing up to 75 days and spring breakups occurring much earlier (Yang et al., 2020). Breakup trends also have a complex relationship to the presence of regulation on rivers, with operational dams affecting both the timing and mechanisms of breakup that can be expected (Timoney, 2002, Beltaos, 2017, and Jasek, 2019). The impacts of preceding winter temperatures and precipitation and the related trends in breakup occurrence have also been noted, with the factors potentially leading to ice jams or early breakups also well understood (Goulding et al., 2009, Lamontagne et al., 2021). This knowledge has formed the basis of past attempts at data-driven modelling and forecasting of spring breakups. Algorithms including support vector machines (Wang et al., 2010, Barzegar et al., 2019), artificial neural networks (Zhao, 2012, Guo et al., 2018), k-nearest neighbors (Sun et al., 2020), stacking ensembles using regression trees (Sun, 2018), and adaptive neuro-fuzzy inference systems (Sun and Trevor, 2015, Sun and Trevor, 2018) have been applied with varying levels of accuracies to spring breakup prediction. These predictions are often limited to single rivers or regions and have encountered challenges in managing and selecting data. Similar problems have been encountered in the prediction of mid-winter breakups (the early breakup of ice cover brought about by mid-winter thaws) using data-based thresholds (Prowse et al., 2002, Carr and Vuyovich, 2014, Newton et al., 2017) and machine learning analysis (De Coste et al., 2022a, De Coste et al., 2022b), or in the prediction of breakup ice jams using artificial neural networks (Massie et al., 2002), neuro-fuzzy systems (Mahabir et al., 2006), and stacking ensembles (De Coste et al., 2021). These techniques have also been extended to generalized spring flooding using boosting and random forest (Kulin et al., 2021), ensembles of regression and snowmelt models (Sarafanaov et al., 2021), and recurrent neural

networks (Cai and Yu, 2022). The management of data in these hydrological forecasting studies, especially when attempting modelling on regional or national scales, is often a challenge, as is delivering a user-friendly means of delivery. Challenges are also present in the selection of variables and the delivery of final decision-making tools for end users and stakeholders.

Knowledge and data management through the use of ontology-based semantic modelling presents an opportunity to address many of the issues present in the management and prediction of spring breakups. Ontologies are representations of a domain in terms of concepts and their relationships. They allow the structure of a system domain to be conceptualized and formalized through the establishment of the relevant concepts and entities and relationships present within the system (Guarino et al., 2009). Ontologies can be easily designed to match a specific criterion or purpose and multiple ontologies can be constructed to conceptualize the same domain knowledge body, while providing an easily understood and evaluated means of illustrating a network (Brank et al., 2005). Further, developed ontologies can also be combined with other ontologies to further elaborate on contained concepts or account for multiple related concepts, allowing more comprehensive models to be developed (Noy and Musen, 2004). Ontologies have been developed to describe flood forecasting systems (Agresta et al., 2014), flood risk assessment systems (Scheur et al., 2013), riverine systems (Mughal and Shaikh, 2017), remote sensing (Potnis et al., 2018), and many flood management systems (Mughal et al., 2021, Agresta et al., 2002, Roller et al., 2015, Kollarits et al., 2009, Yi and Sun, 2013). Though these systems have been well established in related fields, there has not been any application of these techniques to river ice related systems, and very limited applications in river-based hydrology.

This paper focuses on the prediction of spring breakup timing using a novel hybrid ontology and machine learning modelling framework. The Ice Season Ontology organizes the data

and connections between variables (i.e. concepts/entities) to build a semantic model of the relevant data, allowing network analysis techniques to assess the centrality of relevant variables. The machine learning model is then constructed and refined based on the output of the ontology model with the variable selection process enhance by the results of the ontology. **This new framework leverages the advantages of both semantic and data-driven modelling techniques to produce more versatile predictions, providing and easy to use and understand model for stake holders. This is the first attempt at combining these separate modelling techniques in this fashion as well as the first application of semantic modelling to river ice hydrology.** This framework is applied to a national case study of river ice breakups in Canada, focussing on rivers throughout the country with a variety of climate regions, to demonstrate the effectiveness of the proposed framework.

5.2 Data

The primary data source in this study is the Canadian River Ice Database (CRID) (de Rham et al., 2020). The CRID includes data from 196 rivers in Canada spread between 10 of the 11 climate regions within the country. Forty six of these rivers are regulated (meaning flows are affected by man-made structures such as dams or reservoirs) and 150 are non-regulated or natural. These river stations are mapped in Figure 5.1. The CRID contains data on 15 river ice events occurring within the typical river ice season. These events were extracted using careful analysis of the water level time series using a technique developed by Beltaos (1990). The first of the included events is freeze up, denoted as First B Date and H_f. First B Date indicates the first day that ice backwater effects were detected in the channel and thus the initial presence of ice forming in the river, however it does not indicate that a full ice cover has developed. The H_f date, which marks the point where a full ice cover has developed and freeze up has occurred, was extracted based on

observation of a spike in water level caused by the increase in frictional resistance caused by the presence of the full ice cover. The water level and flow measured on the H_f date as well as 30 subsequent days of freeze up water levels are also included for several of the rivers if the records were available. The next key events were the first winter low flow (HLQ1) and low water level (HLW1), both occurring after a decrease in flows caused by drawdown of water contributions to the river. These are determined through observation of the respective time series for each measure and may not occur on the same day as the open water stage-discharge relationship is no longer applicable under ice cover conditions (United States Geological Survey, 1977). For both of these events, the date, water level, and flow rate are also included at the occurrence of the event. Many of the ice seasons in the CRID also include measurements of ice thickness, taken based on direct measurement occurring during site visits to the gauges in question, with corresponding dates included. Though these ice thicknesses can be valuable in analysis, it is important to note that there is no fixed date or timing in the ice season for these measurements. Additional ice season events included in the CRID prior to ice cover breakup are the occurrence of mid-winter breakups (HMWB and HMWM) and the subsequent freeze up, maximum water level and minimum water level and flows measured for these events (HF2, HF2MAX, HLW2, and HLQ2 respectively), with each measured in the same manner as the initial events. The final event of the typical ice season is the breakup, starting with the onset of breakup (HB) determined as a spike in the rising limb of the water level hydrograph triggered by a decrease in flow resistance. The subsequent peak breakup flow (HM) is determined through analysis of the flow hydrograph following the onset of breakup, with the last ice affected date (Last B Date) determined as the final day with ice backwater effect. For each of these events, flows and water levels are also included if available. Table 5.1 summarises the events included in the CRID.

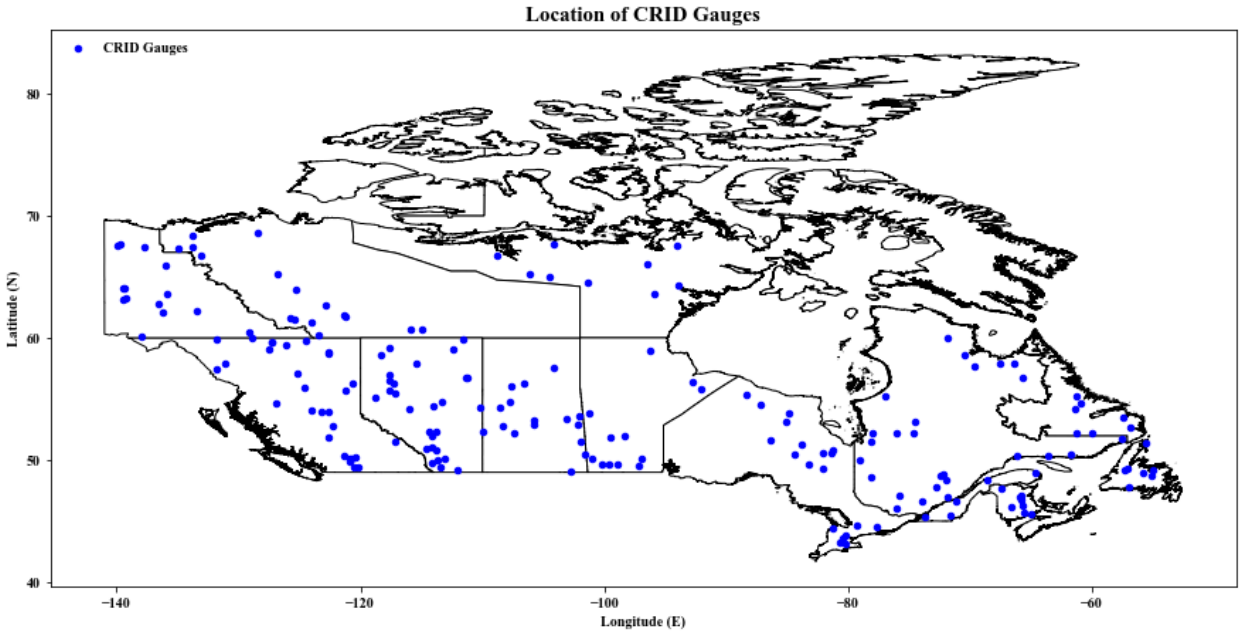


Figure 5.1: National map of CRID gauges included in this study.

Table 5.1: Key Ice Season Events included in the CRID and the number of measurements of each.

Event	Variables	Number of Values
Freeze Up	First B Date	8351
	Freeze Date	5933
	Water Level	5934
	Flow	5896
First Winter Low Flow	Date	7554
	Water Level	4841
	Flow	7554
First Winter Low Level	Date	4327
	Water Level	4329
	Flow	4310
Ice Measurement	Date	5893
	Thickness	5893
Breakup	Date	5367
	Water Level	5366
	Flow	5335
	Last B Date	8715

5.3 Methodology

5.3.1 Semantic Model Development

The intent of this study was the development of an Ice Season Ontology, which sorts the data and entities, and connections between the relevant concepts of an "ice season". The focus of this ontology needed to be clearly defined before construction began (Mughal et al., 2021). In this case, the ontology's key concept describes an ice season, with the associated events and measures from those events being the internal entities within the ontology. Each of the events has associated dates, flows, and levels, and the values of those falls under the umbrella of both the event and the

data type, thus the importance of the subdivisions detailed below. Within an ontology's structure, axioms define the relationships between model nodes, relationships are the links between nodes with a relevant descriptor, and individuals are the specific values present for a node (Guarino et al., 2009). The key individuals in this study are the flows, levels, and dates associated with the events of an ice season. Each node is defined with a datatype (float, integer, date, etc.,) to better represent any individuals mapped to the node and allow for further calculation if necessary.

The construction process of this ontology followed the METHONTOLOGY framework, consisting of 4 stages: *Specification, Conceptualization, Formalization and Implementation,* and *Evaluation* (Lopez et al., 1999). The *Specification* stage consists of defining the scope of the ontology and formulating the usage case and competency questions for assessment of the ontology accuracy. The *Conceptualization* stage is where the structure of the ontology is defined, with key entities identified and model concepts, properties, relationships, and axioms mapped. *Formalization and Implementation* consists of the construction of the ontology using OWL. The *Evaluation* stage consists of the testing of the ontology against the constructed competency questions and any other testing criteria that are developed, at which point the testing criteria can be refined, or the scope and mapping of the model can be revisited to reformulate problematic sections. If the model is found to be satisfactory at this stage, construction has concluded.

5.3.2 Machine Learning Model Development

5.3.2.1 Considered Algorithms

Five machine learning algorithms were tested in this study based on their previous applications in hydrology related modelling studies as described below. The models range in complexity from single equations to complex ensemble algorithms and are detailed below.

Multiple Linear Regression (MLR): MLR models develop a single equation to describe relationship between a set of *i* independent variables (x_i) and a dependent target variable (y) as shown in Equation 5.1 (Uyanik and Guler, 2013). This model is fit using the technique of least squares, through which the Mean Squared Error (MSE) between predicted values and actual values is minimized, as in Equation 5.2 below.

$$y = a + b_1 x_1 + b_2 x_2 + \cdots b_n x_n \tag{5.1}$$

$$MSE = \frac{1}{N}\sum_{i}^{n}(Y_i - y_i)^2 \tag{5.2}$$

where a is the intercept for the MLR equation, b_i the coefficient for each variable x_i, N the total number of observations, and Y_i the actual value for the predicted value y_i. MLR has been successfully applied to the prediction of stream flow on rivers in both open water and ice affected conditions (Seidou and Ouarda, 2007 and Chokmani et al., 2007).

k-Nearest Neighbours (KNN): This model type assigns values to observations based on feature similarity to its "k-nearest neighbours", with k being a user specified function (Song et al., 2017). A distance function is used to assign the value of each new observation to the model point based on the selected value of k neighbours, with the factor, w_j, described in Equation 5.3.

$$w_j = \sum_{j=1}^{k} \begin{cases} \dfrac{d_k - d_j}{d_k - d_1}, & d_k \neq d_1 \\ 1, & d_k = d_1 \end{cases} \tag{5.3}$$

where w_j is the weight for the individual point and the distance between the k neighboring points is described by d_j, j=1,…,k . KNN regression has been successfully applied to open water streamflow prediction and river ice breakup timing (Poul et al., 2019 and Sun et al., 2020).

Regression Tree (RT): Part of the Classification and Regression Tree (CART) model family, an RT model assesses the appropriate value for the predictand by using a series of decision nodes that

assess the values of individual variables (Lewis, 2000). Based on this value, the current observation will pass to the next appropriate decision node until it arrives at a final "leaf" node representing the predicted value for the observation. The total number of splits, leafs, and total tree depth are user selected parameters, along with the metric for which the regression accuracy is assessed, typically a form of MSE similar to the value used by the MLR in Equation 2. RT models have commonly been applied to river hydrology, finding success in modelling flood susceptibility and breakup timing (Janizadeh et al., 2021 and Sun, 2018).

Random Forest (RF): The RF algorithm is an extension of the RT model type, in which an ensemble, or "forest", of decision trees is constructed using bootstrap samples of data taken from the full dataset (Segal, 2003). Each of the trees is built from a unique bootstrap resampling of the overall dataset, leading to a final prediction of the model based on the ensemble of the bootstrap samples, developed from Equation 5.4 (Liaw and Wiener, 2002).

$$\hat{f}(x) = \frac{1}{M}\sum_{m=1}^{M} \hat{f}^{m}(x) \tag{5.4}$$

where $\hat{f}(x)$ is random forest final prediction for the given observation, M is the total number of bootstrap samples, and $\hat{f}^{m}(x)$ is the prediction from tree developed from the m^{th} bootstrap sample. RF has been successfully applied to the assessment of mid-winter breakups in Canada (De Coste et al., 2022).

Extreme Gradient Boosting (XGBoost): XGBoost models function in a similar fashion to other boosting algorithms, constructing an iterative ensemble of classification trees. Each tree is built sequentially, with points incorrectly classified used to refine subsequent trees. The XGBoost algorithms unique feature is the use of second order gradients of the loss function in order to minimize overall model error, while also employing regularization. The kth feature values and

second order gradient statistics of the training instances are represented by $D_k = \{(x_{1k}, h_1), (x_{2k}, h_2), \ldots (x_{nk}, h_n)\}$ with candidate split points represented by $\{s_{k1}, s_{k2}, \ldots s_{kl}\}$, thus satisfying Equations 5.5 and 5.6 (Chen and Guestrin, 2016).

$$r_k(z) = \frac{1}{\sum_{(x,h)\epsilon D_k} h} \sum_{(x,h)\epsilon D_k, x<z} h \tag{5.5}$$

$$\left|r_k(s_{k.j}) - r_k(s_{k,j+1})\right| < \epsilon, s_{k1} = \min(x_{ik}), s_{kl} = \max(x_{ik}) \tag{5.6}$$

where the rank function is $r_k: R \rightarrow [0, +\infty)$, representing the proportion of instances with a feature value *k* smaller than *z*, hi is the weight of each point, and ϵ is an approximation factor indicating that there is 1/ ϵ candidate points. XGBoost has been successfully applied to flood mapping on rivers and prediction of ice phenomena (Abedi et al., 2021 and Graf et al., 2022).

5.3.2.2 Model Assessment

Model accuracy for each of the considered algorithms was assessed using the regression metrics Mean Absolute Error (MAE) and R^2. These are calculated using Equations 5.7 and 5.8 below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{5.7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5.8}$$

where *n* is the number of observations, e_i the error for observation *i*, y_i is the *i*th response value, $\hat{y}_i$ is the *i*th fitted response, and $\bar{y}$ is the mean of the responses (Chai and Draxler, 2014, and Nakagawa and Schielzeth, 2013). Each of the models required the use of five-fold cross-validation to select the optimal configuration of hyperparameters. In this process, the entire data pool is subdivided into five equal segments, with each segment being used as a testing set against the remaining data used as training data. The considered values for the parameter(s) of each model are sorted into grids, and an exhaustive grid search is used to test each possible combination of parameters against the five data partitions, allowing the

most successful configuration to be selected (Refaeilzadeh et al., 2009). The models were subsequently trained and tested on a 75/25 training/testing split of the overall data.

5.3.3 Hybrid Model Structure

A hybrid model structure was developed for this study which integrates knowledge gained from the semantic model of the ice season to refine the abilities of the machine learning models to predict the timing of spring breakup. Analysis of the structure of the developed ontology using network analysis techniques provides a means to assess the critical and redundant nodes in the ontological structure, allowing variable selection in the machine learning models to be reassessed.

5.3.3.1 Data Preparation

Prior to analysis within either type of the models, the data was first cleaned and organized. Observations with missing values were removed from the modelling pool as, due to the large time spans that can be present between measurements coupled with the inconsistencies in when the measurements were taken, interpolation or other replacement methods were not possible without comprehensive reanalysis of the original data (Dastorani et al., 2009). Due to the large variance in both flows and water levels between the gauges considered in this study, the values for variables related to these measures were normalized through division by means to ensure consistency in the modelling results (Dery et al., 2009). Dates were also altered to be better used by the machine learning model, converted from calendar dates to durations from a reference date. In this study the selected reference date was October 1st, the start of the Canadian water year (Boyd, 1979). One group of variables that was excluded from analysis were those relating to mid-winter breakups, thus limiting the scope of the domain to only consider seasons without these events. These events are so rare and result in such a differing flow regime and progression for the remainder of the ice season that they would result in potential inaccuracies were they included in this study (Huntington

et al., 2003). Further, a differing ontology describing these events would need to be developed, separating them from the typical progression of the ice season.

5.3.3.2 Hybridization and Variable Selection

The ontology was structurally analysed using metrics common to network analysis, producing values describing the importance or dependency of the network on a given node or variable. The three metrics selected were unweighted measures of centrality or criticality. The first metric, degree centrality, describes the amount of links a particular node has in comparison to the total links in the network, giving and indication of the centrality of that node against the total network connectivity(Zhang and Luo, 2017). The second metric, closeness centrality, describes how close a particular node is relative to all other nodes present in the network, providing an indication of the overall connectivity of the node in question to the network (Okamoto et al., 2008). The final metric, betweenness centrality, describes the ratio of the amount of shortest paths between nodes in the network passing through the node of interest against the total number of shortest paths in the network, indicating the level of connectivity of that node to the other nodes of the network. These metrics are described in Equations 5.9-5.11 below.

$$C_D(v) = \frac{\sum_{i=1}^{V} Xi}{(V-1)(V-2)} \tag{5.9}$$

$$C_C(v) = \frac{1}{\sum_{t \in V} d_G(v,t)} \tag{5.10}$$

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{5.11}$$

where C_D is the degree centrality, v is the node of interest, V is the total number of nodes in the network, Xi is the amount of links of a particular node, C_C is the closeness centrality, $d_G(v,t)$ is the distance between vertices v and t in the network, C_B is the betweenness centrality, $\sigma_{st}(v)$ is the number of shortest paths passing through node v, and σ_{st} the total number of shortest paths in the

network. For each of the considered centralities, the values can vary between 0 and 1, with smaller values indicating decreasing centrality of the node in the network. Through assessing the values of centralities for each of the nodes representing a relevant variable in the machine learning model, the least critical values can be identified and removed from the model, producing a refined variable selection. This process is described in Figure 5.2.

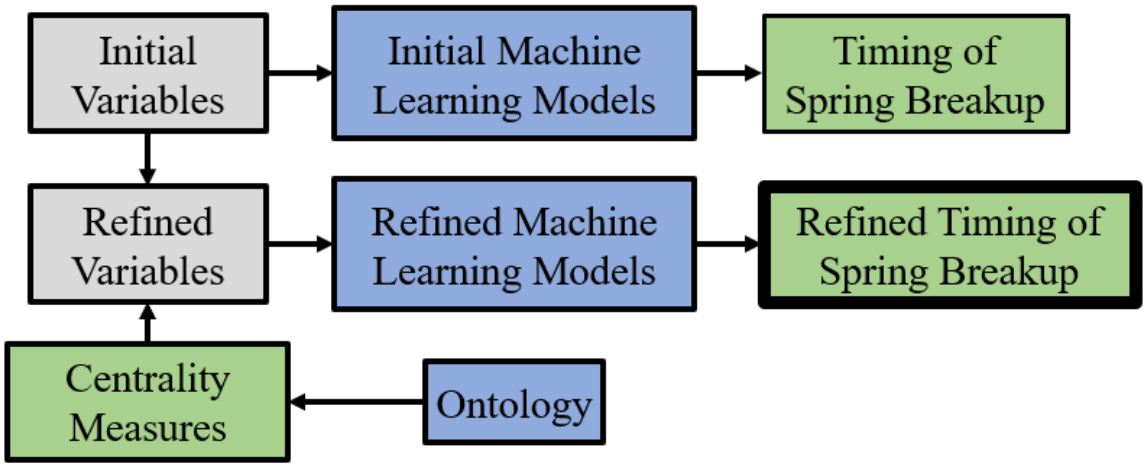


Figure 5.2: Flow chart of hybrid modelling framework combining the results of the Ontology and machine learning models in producing refined spring breakup timing predictions.

5.3.4 Model Implementation

This study utilised Protégé, version 5.5.0 for the development of the Ice Season Ontology (Musen, 2015). Additional analysis was conducted in Python, version 3.7, using the packages *ddot* (Yu et al., 2019), *geopandas* (Jordahl, 2014), *Pandas* (McKinney, 2010), *Numpy* (Oliphant, 2006), *scikit-learn* (Pedregosa et al., 2011), *scipy* (Virtanen et al., 2020), *seaborn* (Waskom, 2008), and *xgboost* (Chen and Guestrin, 2016).

5.4 Results and Analysis

5.4.1 Ice Season Ontology

As describes in Section 5.3.2, the Ice Season Ontology was constructed according to the METHONTOLOGY framework. In the *Specification* stage, the ontology was constructed to describe an "ice season", including all relevant events and traits described in the CRID, hence the title Ice Season Ontology. All variables listed within Table 5.1 were included in the model in addition to the years, station, and coordinates of each ice season in the record. This produced a more complete model and allowed for more robust predictions to be produced by the hybrid model framework. At this stage competency questions assessing the intended function of the finished model were developed. These included 1) "Which Ice Seasons include Breakup Flow?", 2) "Which Ice Seasons include the First Winter Low Flow Level?", 3) "What are the End Years and Station Names for each Breakup Level?", and 4) "What are the Dates of all Freeze Ups?". The next step of model construction was the *Conceptualization* stage, where the primary relationships between the intended data were identified and model structure was organized. Each individual variable value fell under the description of both the relevant event (i.e. freeze up, first winter low flow, first winter low level, breakup, ice measurement) and the type of value (i.e. date, flow, water level, thickness). Additional ice season characteristics were included separately from the key ice season events. The model structure was implemented in the *Formalization and Implementation* stage in the Protege software, with an OWL model constructed describing an Ice Season. At this stage, the data from the CRID was loaded in using the CELLFIE plugin, allowing the appropriate relationships between each ice season component to be assigned to each individual data point loaded. The final ontology structure is shown in Figure 5.3, including all events and data types. In Figures A1 – A8 of Appendix A, detailed views of the components of the ontology focussed on the event type and detailed views divided by the data type are provided to give a complete view of

the structure. The overall summary statistics of the model are detailed in Table 5.2, giving a breakdown of the model complexity and the amount of relationships, classes, components, and datatypes included in the final model with all CRID data imported.

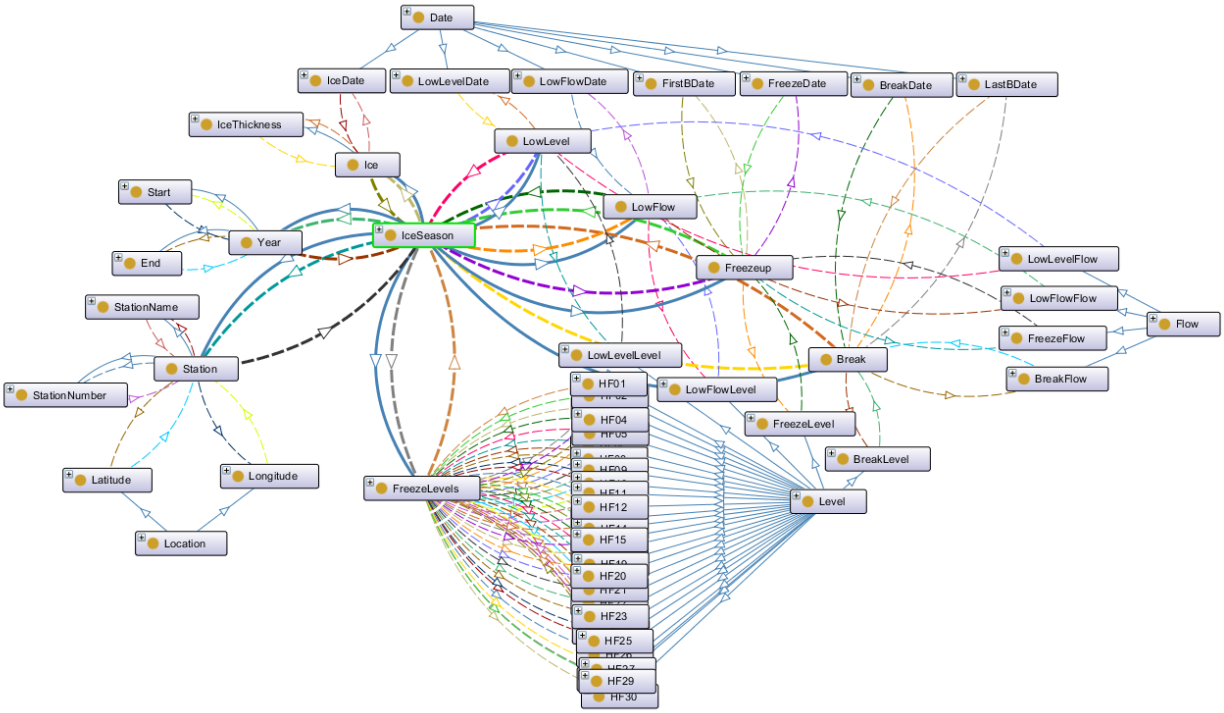


Figure 5.3: Full structural overview of the developed Ice Season Ontology.

Table 5.2: Summary statistics of Ice Season Ontology describing contained logical and data connections.

Metric	Count
Axiom	983624
Logical Axiom	891545
Declaration Axiom	92079
Class	65
Class Assertions	235040
Object Property	120
Object Property Assertions	654354
Data Property	7
Individual	91215

The model was tested against the competency questions listed above in the *Evaluation* phase and was found to successfully provide the intended data. In response to competency question 1, the response shown in Figure 5.4 was obtained, demonstrating the expected ability to produce the relevant breakup flows and associated ice seasons. Figure 5.5 shows the response to question 2, producing a similar list of the expected seasons and associated first winter low flow levels. Figure 5.6 shows the ice seasons, breakup water levels, end years, and station names expected in response to question 3, demonstrating the ability to handle more complex queries and link together separate entities from the ontology. The response to question 4 is shown in Figure 5.7, providing a list of ice seasons and freeze up dates, demonstrating the ability to call the last of the relevant variable datatypes.

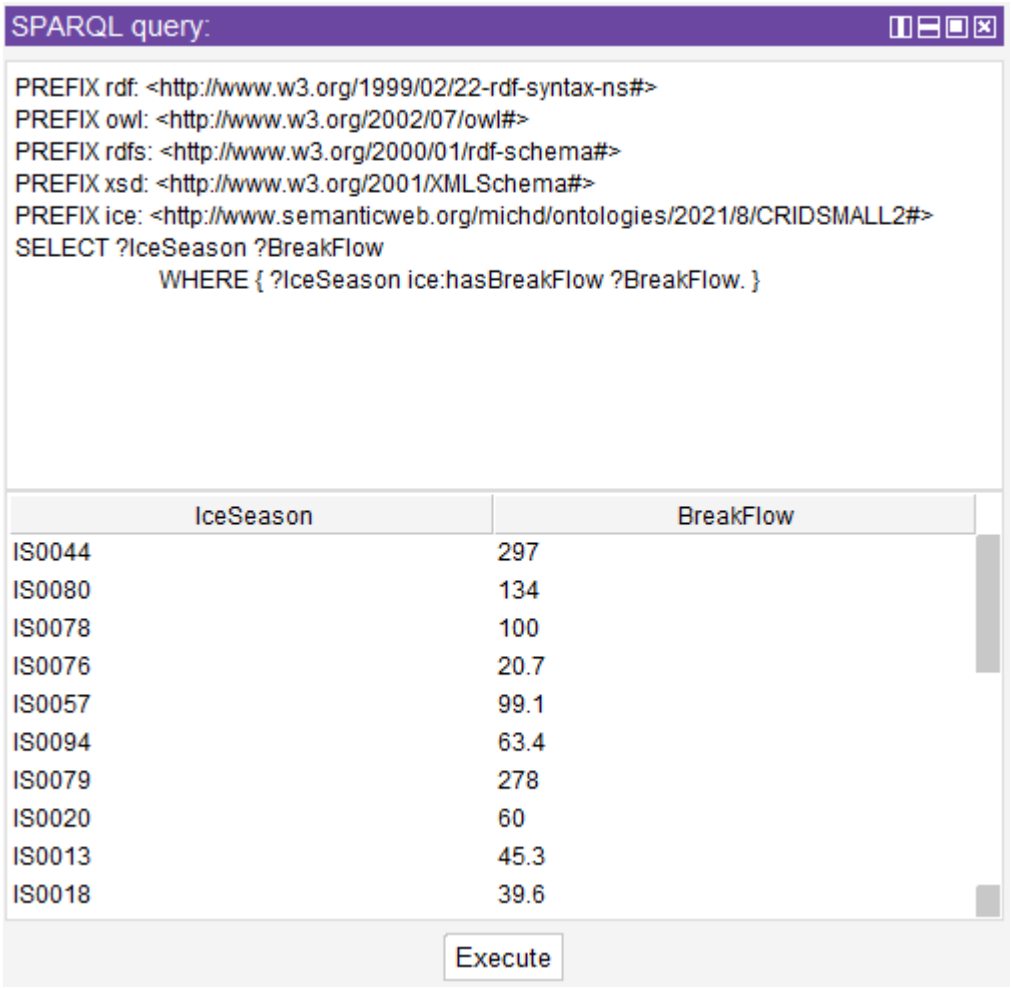


Figure 5.4: Sample of the Ice Season Ontology response to competency question 1) Which ice seasons include breakup flows?

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ice: <http://www.semanticweb.org/michd/ontologies/2021/8/CRIDSMALL2#>
SELECT ?IceSeason ?LowFlowLevel
                WHERE { ?IceSeason ice:hasLowFlowLevel ?LowFlowLevel. }
```

IceSeason	LowFlowLevel
IS0002	2.515
IS0048	1.284
IS0037	1.062
IS0007	1.695
IS0046	1.114
IS0087	10.404
IS0086	10.634
IS0088	10.699
IS0067	10.56
IS0057	10.903

Execute

Figure 5.5: Sample of the Ice Season Ontology response to competency question 2) Which ice seasons include first winter low flow level?

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ice: <http://www.semanticweb.org/michd/ontologies/2021/8/CRIDSMALL2#>
SELECT ?IceSeason ?BreakLevel ?End ?StationName
                WHERE { ?IceSeason ice:hasBreakLevel ?BreakLevel.
?IceSeason ice:hasEnd ?End.
?IceSeason ice:hasStationName ?StationName}
```

IceSeason	BreakLevel	End	StationName
IS0020	1.658	1981	NASHWAAKRIVERATDURHAMBRIDGE
IS0070	12.213	1989	SALMONRIVERATCASTAWAY
IS0006	1.646	1967	NASHWAAKRIVERATDURHAMBRIDGE
IS0024	1.37	1985	NASHWAAKRIVERATDURHAMBRIDGE
IS0084	13.986	2003	SALMONRIVERATCASTAWAY
IS0093	13.183	2012	SALMONRIVERATCASTAWAY
IS0019	2.017	1980	NASHWAAKRIVERATDURHAMBRIDGE
IS0029	1.765	1990	NASHWAAKRIVERATDURHAMBRIDGE
IS0038	1.931	1999	NASHWAAKRIVERATDURHAMBRIDGE
IS0078	12.416	1997	SALMONRIVERATCASTAWAY

Execute

Figure 5.6: Sample of the Ice Season Ontology response to competency question 1) What are the end years and station names for each breakup level?

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ice: <http://www.semanticweb.org/michd/ontologies/2021/8/CRIDSMALL2#>
SELECT ?IceSeason ?FreezeDate
        WHERE { ?IceSeason ice:hasFreezeDate ?FreezeDate.}
```

IceSeason	FreezeDate
IS0068	28/11/1986
IS0056	29/12/1974
IS0021	28/12/1981
IS0077	05/12/1995
IS0031	15/12/1991
IS0047	09/12/2007
IS0072	01/01/1991
IS0040	18/12/2000
IS0046	14/01/2007
IS0009	01/01/1970

Execute

Figure 5.7: Sample of the Ice Season Ontology response to competency question 1) What are the dates of all freeze ups?

5.4.2 Hybrid Model Results

Five machine learning models were constructed using all potential variables in the CRID for initial accuracy. An MLR, KNN, DT, RF, and XGBoost model were constructed to predict the length of the ice season from the beginning of freeze up to the breakup date using the key events of the season in between. Each of the models was trained and tested on a 75/25 split, and the model hyperparameters were selected using 5-fold cross-validation. The models included a total of 10 variables for the prediction of the season length. The obtained accuracies of each of the five models

on the testing set is shown in Table 5.3. The R-Squared values and MAEs of each of the models varied, though performance was generally high. The best performing models were the RF and XGBoost, with near comparable values in accuracy, both algorithms being ensemble models that produce more versatile predictions.

Table 5.3: Model accuracies of each machine learning algorithm prior to hybridization.

Model	R-Squared	Mean Absolute Error (days)
MLR	0.774	13.82
KNN	0.795	12.82
DT	0.777	14.66
RF	0.864	11.86
XGBoost	0.854	11.25

Following this, the Ice Season Ontology structure was used to calculate the values for degree, closeness, and betweenness centrality for all of the variables included in the initial models, as well as the larger categories that the variables belong to (event type and data type). The obtained values are listed in Table 5.4. Based on the structure of the ontology itself, it was found that each of the 10 considered variables in the initial machine learning models had the same values for each of the three centralities considered, thus further analysis of the larger categories was necessary. For both degree and betweenness centrality, the values for the event categories of First Low Flow (LowFlow in Table 5.4) and First Low Level (LowLevel in Table 5.4) were lower than those of freeze up and breakup. The degree centrality for both the level and flow categories was also lower

than the value for date. Through these values it was concluded to remove four variables from the initial selection: First Winter Low Flow Level, First Winter Low Flow Flow, First Winter Low Level Level, and First Winter Low Level Flow. The refined variable selection now consisted of six variables for the prediction of breakup timing.

Table 5.4: Centrality metrics from Ice Season Ontology for each relevant variable in hybrid modelling framework.

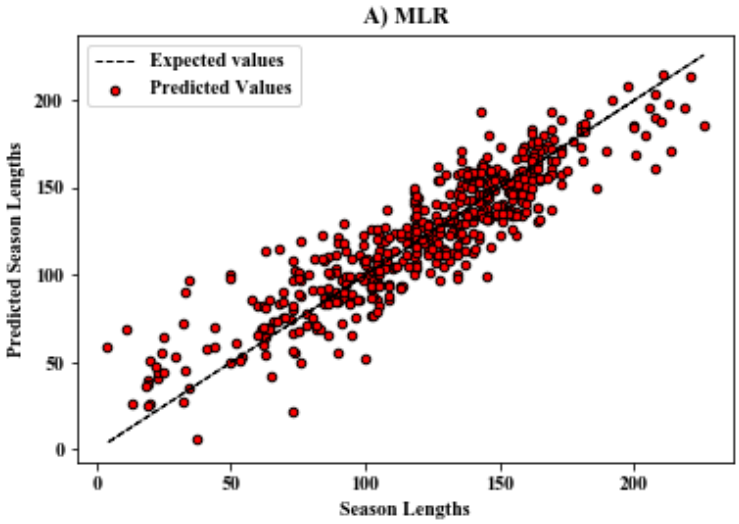
Variable	Degree Centrality	Closeness Centrality	Betweenness Centrality
IceSeason	0.212	0.000	0.000
Date	0.212	0.000	0.000
Flow	0.121	0.000	0.000
Level	0.121	0.000	0.000
Freezeup	0.152	0.030	0.004
FirstBDate	0.061	0.068	0.000
FreezeDate	0.061	0.068	0.000
FreezeFlow	0.061	0.068	0.000
FreezeLevel	0.061	0.068	0.000
LowFlow	0.121	0.030	0.003
LowFlowDate	0.061	0.068	0.000
LowFlowFlow	0.061	0.068	0.000
LowFlowLevel	0.061	0.068	0.000
LowLevel	0.121	0.030	0.003
LowLevelDate	0.061	0.068	0.000
LowLevelFlow	0.061	0.068	0.000
LowLevelLevel	0.061	0.068	0.000
Break	0.152	0.030	0.004
BreakDate	0.061	0.068	0.000
BreakFlow	0.061	0.068	0.000
BreakLevel	0.061	0.068	0.000
LastBDate	0.061	0.068	0.000

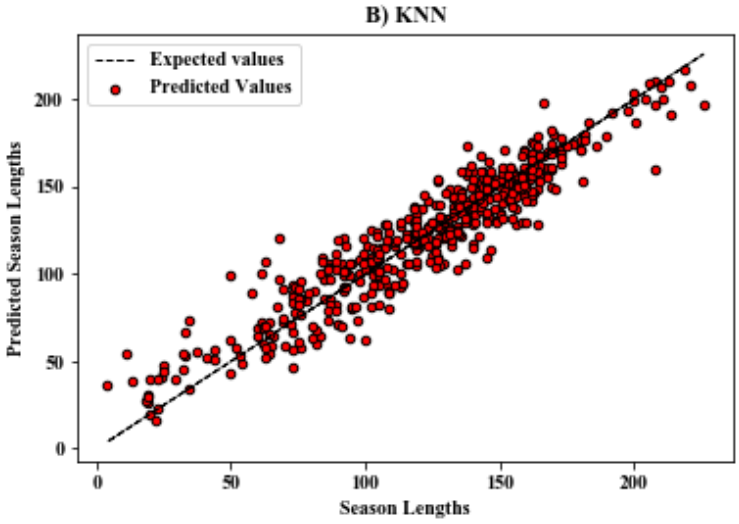
With the refined variable selection, the five machine learning models were reconstructed using the same methodology previously discussed. Each model was trained and tested on a 75/25

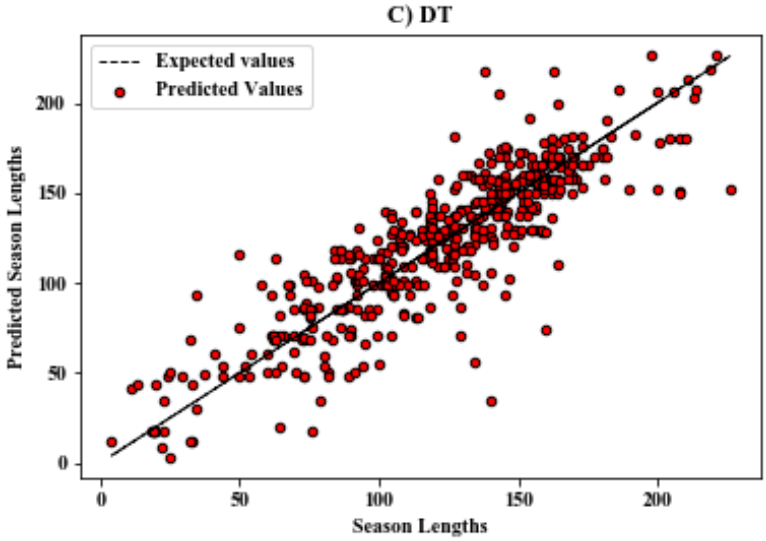
split, with hyperparameters again selected through five-fold cross-validation. The new performances achieved by the models are listed in Table 5.5. The performance of each of the models uniformly increased with the refined variable selection, with the RF and XGBoost models again being the best performing of the five considered algorithms, though with the RF more consistently outperforming the XGBoost model. Figure 5.8 graphs the predictions of each of the models against the actual values for each of the seasons in the testing set. Though all five models achieved reasonable R-Squared values, the closeness of the predictions of the RF is best illustrated here. Further, the MAE of 10.85 days can be seen against the maximum time span for the considered seasons of 224 days.

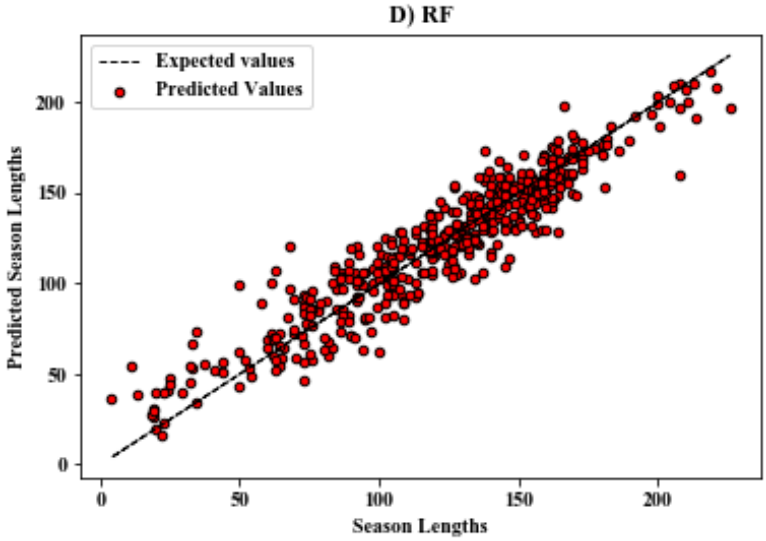
Table 5.5: Accuracies of considered machine learning algorithms post-hybridization using refined variable selection.

Model	R-Squared	Mean Absolute Error (days)
MLR	0.779	13.61
KNN	0.797	12.77
DT	0.781	14.12
RF	0.884	10.85
XGBoost	0.856	11.18


A) MLR
Expected values
Predicted Values
Predicted Season Lengths
Season Lengths
0
50
100
150
200


B) KNN
Expected values
Predicted Values
Predicted Season Lengths
Season Lengths
0
50
100
150
200


C) DT
Expected values
Predicted Values
Predicted Season Lengths
Season Lengths
0
50
100
150
200


D) RF
Expected values
Predicted Values
Predicted Season Lengths
Season Lengths
0
50
100
150
200

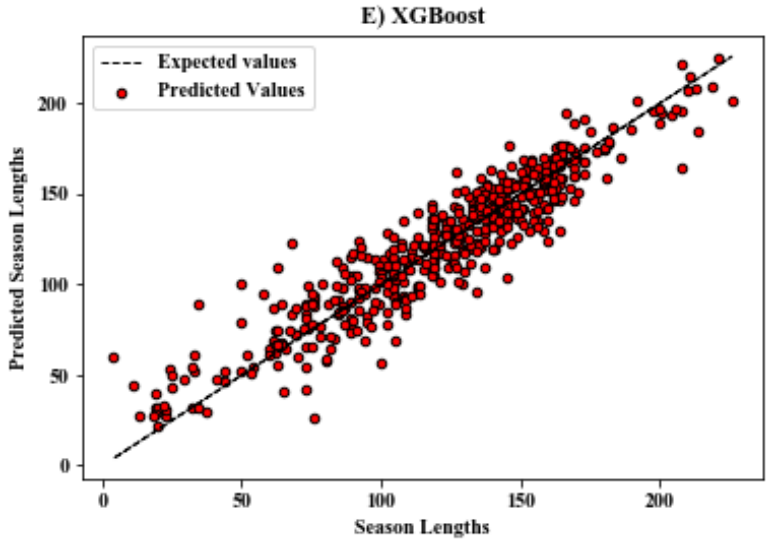


Figure 5.8: Graphs of the accuracies of predicted season lengths from the A) MLR, B) KNN, C) DT, D) RF, and E) XGBoost models.

5.4.3 Discussion

Overall prediction performance for the machine learning models was acceptable both pre and post hybridization, though an improvement in model accuracy was noted after a refined variable selection was made. In particular, a decrease in MAE of 1.01 days was achieved for the RF model. While the decrease in error was not as significant for other models, with the XGBoost having only a decrease of 0.08 days, each of the models was seen to increase in accuracy. As the model is intended for decision making support for river-bound communities, increases in accuracy of this level are vital for providing sufficient notice of breakup initiation. The accuracy of the model was also assessed on a national scale by running all included CRID data through the successful RF and mapping it by gauge, shown in Figure 5.9. The MAE achieved at each gauge was found to be quite consistent and in line with the overall model, though there were some exceptions. Two gauges in particular (Nicola River near Merritt in British Columbia and Whitemud River near Dixonville in Alberta) had notably high values of MAE (24.79 days and

23.37 days respectively). The Nicola River gauge was located on a heavily regulated river, while the Whitemud River gauge had a significant amount of missing data, impacting the amount of data available for modelling. Though this study investigated both regulated and unregulated rivers simultaneously, future research may investigate the separation of these river types and the impact this has on accuracy. Additional research into the effects of the physical characteristics of the rivers could also include investigation of the hydrological factors of the river, such as basin size, slope, or flow regime, and their inclusion in both the Ice Season Ontology and the machine learning models.

Though other variable selection methods exist, including importance assessment (De Coste et al., 2021) and input omission (Snieder et al., 2020), this technique provides a refined selection in combination with the development of a usable ontology. The ontology itself provides an easily understood tool for organizing and viewing data and allows assessments at specific locations to be extracted and modelled with relative ease for an inexperienced user. This also allows for easy communication of modelling results with strong visuals for end-users, and a means of sorting and providing machine learning predictions.

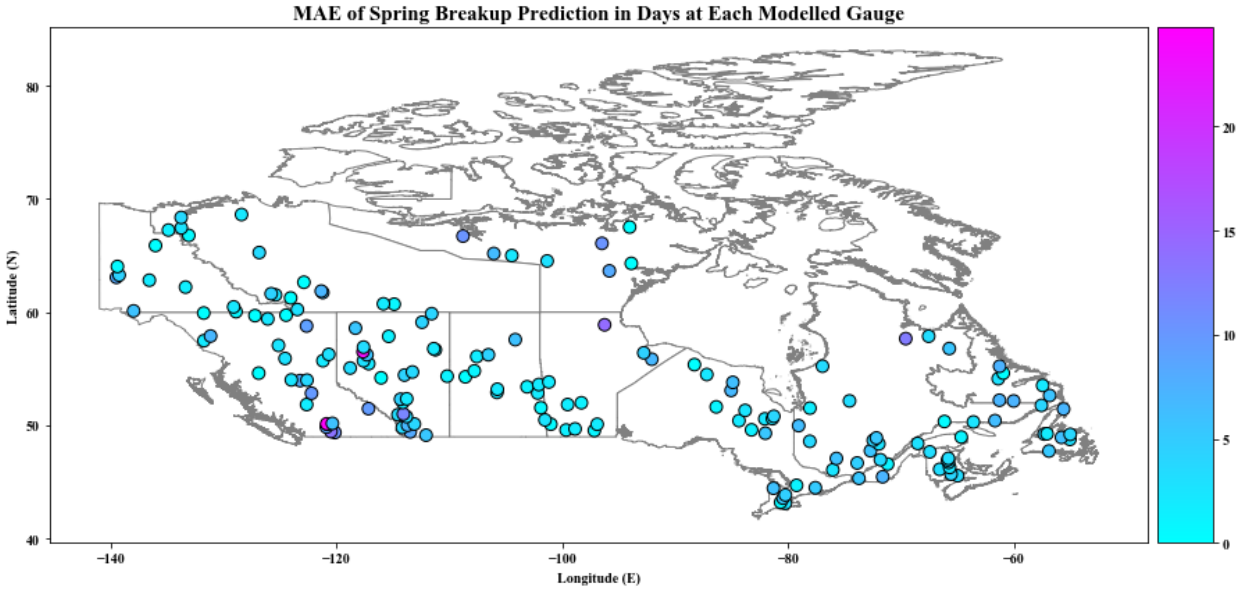


Figure 5.9: Map of the MAE of the hybrid RF model at each considered gauge using the full dataset.

There are several avenues for future development of both the Ice Season Ontology itself as well as the hybrid modelling framework. Currently, the ontology excludes the occurrence of mid-winter breakups in its structure, as these are relatively rare events, and they result in significant alteration to the remainder of the ice season and the type of events that occur. These events have been noted to be significantly increasing in frequency however (De Coste, 2022a), and their inclusion would be of great benefit to the model. Similarly, data from outside of the ice season, including maximum spring flows and climatic statistics from the year preceding freeze up would also provide additional avenues for analysis, as well as a more robust database to draw from for alternative modelling approaches. Currently, the ontology and machine learning aspects of the framework consist of separate models, though research could focus on full hybridization of the two, providing a single means for organization and analysis of the data and an easier method to add and assess new data and observations.

5.5 Conclusions

This study focussed on the development of a new hybrid modelling framework combining the advantages of ontology-based semantic modelling and data-driven modelling to describe and predict spring breakups on Canadian rivers. The CRID, an expansive dataset containing large amounts of data on the key events of the river ice season, was used to form the basis of an Ice Season Ontology. This ontology describes the key events and associated values of the typical Canadian Ice Season, and application of competency questions demonstrated the effectiveness of the model structure, which was designed according to typical semantic modelling standards. This model was then analysed using network modelling techniques to assess the importance of each event in the structure, allowing a refined selection of variables for a series of machine learning models. The refined selection lead to a smaller number of critical variables and was demonstrated to increase accuracy for all tested models, with the best performing model gaining a significant increase in accuracy. The model was found to be successful for the majority of gauges across the country, demonstrating the transferability of the proposed framework.

This study was the first application of semantic modelling to river hydrology and river ice hydrology. The obtained results demonstrate the value of these techniques both as an analytical tool, providing new means of assessing data and improving models, as well as acting as decision-making support. The intuitive structure of the semantic model coupled with it's relative ease-of-use compared to other data-driven modelling techniques allows it to be easily adopted by affected communities with a minimal learning curve. The hybrid framework acts as a novel combination of the results of semantic and data-driven modelling, with a successful demonstration of the value of these techniques used in parallel. These techniques are easily extended to similar locations and

applications in river ice studies, as well as other forecasting applications where network structures exist and ontology-based models have not yet been applied.

Acknowledgements

This study was supported by Natural Sciences and Engineering Research Council of Canada (NSERC). Data used is available from Environment and Climate Change Canada (https://open.canada.ca/data/en/dataset/c5b58ccd-0011-4a80-8f24-034c86cbc14d).

References

Abedi, R., Costache, R., Shafizadeh-Moghadam, H., and Pham, Q., 2021. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto International.

Agresta, A., Fattoruso, G., Pollino, M., Pasanisi, C., Tebano, S., Vito, S., and Francia, G., 2002. Multiagent decision support for flood emergency management. Proceedings of the 7th Asia-Pacific Decision Science Institute Conference, 1-10.

Agresta, A., Fattoruso, G., Pollino, M., Pasanisi, C., Tebano, S., Vito, S., and Francia, G., 2014. An ontology framework for flooding forecasting. Proceedings of the International Conference on Computer Science Applications, 417-428.

Barzegar, R., Ghasri, M., Qi, Z., Quilty, J., and Adamowski, J., 2019. Using bootstrap ELM and LSSVM models to estimate river ice thickness in the Mackenzie River Basin in the Northwest Territories, Canada. Journal of Hydrology, 577.

Beltaos, S., 1990. Guidelines for extraction of ice break-up data from hydrometric station records. Working Group on River Ice Jams: Field Studies and Research Needs. NHRI Science Report No. 2, National Hydrology Research Institute, Environment Canada, Saskatoon, SK, Canada, pp. 37–70.

Beltaos, S., 1997. Onset of river ice breakup. Cold Regions Science and Technology, 25: 183-196.

Beltaos, S., 2003. Threshold between mechanical and thermal breakup of river ice cover. Cold Regions Science and Technology, 37.1: 1-13.

Beltaos, S., and Prowse, T., 2009. River-ice hydrology in a shrinking cryosphere. Hydrological Processes, 23.1: 122-144.

Beltaos, S., Prowse, T., Bonsal, B., MacKay, R., Romolo, L., Pietroniro, A., and Toth, B., 2006. Climatic effects on ice-jam flooding of the Peace-Athabasca Delta. Hydrological Processes, 20: 4031-4050.

Beltaos, S., 2017. Frequency of ice jam flooding of Peace Athabasca Delta. Canadian Journal of Civil Engineering, 45: 71-75.

Brank, J., Grobelnik, M., Mladenic, D., 2005. A survey of ontology evaluation techniques. Proceedings o the Conference on Data Mining and Data Warehouses, 166-170.

Boyd, D., 1979. Degree days: The different types. Building research note, National Research Council of Canada from the Atmospheric Environment Service Department of Fisheries and Environment.

Cai, B., and Yu, Y., 2022. Flood forecasting in urban reservoir using hybrid recurrent neural network. Urban Climate, 42: 101086.

Carr, M., and Vuyovich, C., 2014. Investigating the effects of long-term hydro-climatic trends on Midwest ice jam events. Cold Regions Science and Technology, 106-107: 66-81.

Chai, T., and Draxler, R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7: 1247-1250.

Chen, T., and Guestrin, C., 2016. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

Chokmani, K., Ouarda, T., Hamilton, S., Ghedira, M., and Gingras, H., 2007. Comparison of ice-affected streamflow estimates computed using artificial neural networks and multiple regression techniques. Journal of Hydrology, 349(3-4): 383-396.

Dastorani, M., Moghadamnia, A., Piri, J., and Rico-Ramirez, M., 2009. Application of ANN and ANFIS models for reconstructing missing flow data. Environmental Monitoring and Assessment, 166: 421-434.

De Coste, M., Li, Z., Pupek, D., and Sun, W., 2021. A hybrid ensemble modelling framework for the prediction of breakup ice jams on Northern Canadian Rivers. Cold Regions Science and Technology, 189: 103302.

De Coste, M., Li, Z., and Dibike, Y., 2022a. Assessing and predicting the severity of mid-winter breakups based on Canada-wide river ice data. Journal of Hydrology, 607:127550.

De Coste, M., Li, Z., and Dibike, Y., 2022b. Machine-learning approach for predicting the occurrence and timing of mid-winter ice breakups on Canadian rivers. Environmental Modeling and Software, 152: 105402.

de Rham, L., Prowse, T., Beltaos, S., and Lacroix, M., 2008a. Assessment of annual high-water events for the Mackenzie river Basin, Canada. Hydrological Processes. 22: 3864-3880.

de Rham, L., Prowse, T., and Bonsal, B., 2008b. Temporal variations in river-ice break-up over the Mackenzie River Basin, Canada. Journal of Hydrology, 349: 441-454.

de Rham, L., Dibike, Y., Beltaos, S., Peters, D., Bonsal, B., and Prowse, T., 2020. A Canadian river ice database from national hydrometric program archives. Earth System Science Data, Open Access Discussions.

Dery, S., Stahl, K., Moore, R., Whitfield, P., Menounos, B., and Burford, J., 2009. Detection of runoff timing changes in pluvial, nival, and glacial rivers of western Canada. Water Resources Research, 45(4).

Goulding, H., Prowse, T., and Bonsal, B., 2009. Hydroclimatic controls on the occurrence of break-up and ice-jam flooding in the Mackenzie Delta, NWT, Canada. Journal of Hydrology, 379: 251-267.

Graf, R., Kolerski, T., and Zhu, S., 2022. Predicting ice phenomena in a river using the artificial neural network and extreme gradient boosting. Resources, 11, 12.

Guarino, N., Oberle, D., and Staab, S., 2009. "What is an Ontology?" Handbook on Ontologies, Springer, Berling, Heidelberg, 1-17.

Guo, X., Wang, T., Fu, H., Guo, Y., and Li, J., 2018. Ice-jam forecasting during river breakup based on neural network theory. Journal of Cold Regions Engineering, 32(3): 04018010.

Huntington, T., Hodgkins, G., and Dudley, R. 2003. Historical trend in river ice thickness and coherence in hydroclimatological trends in Maine. Climate Change 61: 217–236.

Jasek, M., 2019. An emerging picture of Peace River break-up types that influence ice jam flooding of the Peace-Athabasca Delta, part 1: the 2018 Peace River breakup. In: Proceedings of the 20th Workshop on the Hydraulics of Ice Covered Rivers, Ottawa, ON, Canada.

Janizadeh, S., Vafakhah, M., Kapelan, Z., and Dinan, N., 2021. Novel Bayesian additive regression tree methodology for flood susceptibility modeling. Water Resources Management, 35: 4621-4646.

Jordahl, K., 2014. GeoPandas: A Python tool for geographic data. URL: https://github.com/geopandas/geopandas.

Kollarits, S., Wergles, N., Siegel, H., Liehr, C., Kreuzer, S., Torsoni, D., Sulzenbacher, U., Papez, J., Mayer, R., Plank, C., and Maurer, L., 2009. MONITOR – An ontological basis for risk management. Technical Report, Monitor.

Kulin, N., Kozlov, E., and Zhuk, Y., 2021. Forecasting the spring flood of rivers with machine learning methods. Scientific and Technical Journal of Information Technologies, Mechanics, and Optics, 21: 135-142.

Lamontagne, J., Jasek, M., and Smith, J., 2021. Coupling physical understanding and statistical modeling to estimate ice jam flood frequency in the norther Peace-Athabasca Delta under climate change. Cold Regions Science and Technology, 192: 103383.

Lewis, R., 2000. An introduction to Classification and Regression Trees (CART). Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California.

Liaw, A., and Wiener, M., 2002. Classification and regression by randomForest. R News, 2:18-22.

Lindenschmidt, K., Das, A., Rokaya, P., and Chu, T., 2016. Ice-jam flood risk assessment and mapping. Hydrological Processes, 30.21:3754-3769.

Lopez, M., Gomez-Perez, A., Sierra, J., and Sierra, A., 1999. Building a chemical ontology using methontology and the ontology design environment. IEEE Intelligent Systems, 14: 37-46.

Mahabir, C., Hicks, F., and Fayek, A.R., 2006. Neuro-fuzzy river ice breakup forecasting system. Cold Regions Science and Technology. 46:100-112.

Massie, D., White, K., and Daly, S., 2002. Application of neural networks to predict ice jam occurrence. Cold Regions Science and Technology. 35:115-122.

McKinney, W., and others, 2010. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 445: 51-56.

Mughal, M., and Shaikh, Z., 2017. WaterOnto: Ontology of context-aware grid-based riverine water management system. Journal of Computer Math and Science, 1: 1-12.

Mughal, M., Shaikh, Z., Wagan, A., Khand, Z., and Hassan, S., 2021. ORFFM: An ontology-based semantic model of river flow and flood mitigation. IEEE Access 9: 44003-44031.

Musen, M., 2015. The Protégé Project: A look back and a look forward. AI Matters, Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4): 4-12.

Nakagawa, S., and Schielzeth, H., 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods in Ecology and Evolution, 4:133-142.

Newton, B., Prowse, T., and de Rham, L., 2017. Hydro-climatic drivers of mid-winter break-up of river ice in western Canada and Alaska. Hydrology Research 48.4: 945-956.

Noy, N., and Musen, M., 2004. Ontology versioning in an ontology management framework. IEEE Intelligent Systems, 19.4:6-13.

Oliphant, T., 2006. A guide to NumPy (Vol. 1). Trelgol Publishing USA.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and others, 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Potnis, A., Durbham S., and Kurte, K., 2018. A geospatial ontological model for remote sensing scene semantic knowledge mining for the flood disaster. IEEE International Geoscience and Remote Sensing Symposium, 5274-5277.

Poul, A., Shourian, M., and Ebrahimi, H., 2019. A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. Water Resources Management, 33: 2907-2923.

Prowse, T., Bonsal, B., Lacroix, M., and Beltaos, S., 2002. Trends in river-ice breakup and related temperature controls. Proceedings of the 16th IAHR International Symposium on Ice. Trends in River-Ice Breakup and Related Temperature Controls. Dunedin, New Zealand, 64-71.

Refaeilzadeh, Payam, Lei Tang, and Huan Liu. "Cross-validation." Encyclopedia of database systems 5 (2009): 532-538.

Roller, R., Roes, J., and Verbree, E., 2015. Benefits of linked data for interoperability during crisis management. International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences, 40.3: 211.

Sarafanov, M., Borisova, Y., Maslyaev, M., Revin, I., Maximov, G., and Nikitin, N., 2021. Short-term river flood forecasting using composite models and automated machine learning: The case study of Lena River. Water, 13: 3482.

Scheur, S., Haase, D., and Meyer, V., 2013. Towards a flood risk assessment ontology-knowledge integration into a multi-criteria risk assessment approach. Computers, Environment and Urban Systems, 37: 82-94.

Seidou, O., and Ouarda, T., 2007. Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. Water Resources Research, 43(7).

Segal, M., 2003. Machine learning benchmarks and random forest regression. Technical Report, Center for Bioinformatics and Molecular Biostatistics, University of California, San Francisco.

Shanzen, Y., and Sun, Y., 2013. Upper level ontology and integration assessment modeling in digital watershed. 21st International Conference on Geoinformatics, 1-6.

Snieder, E., Shakir, R., and Khan, U., 2020. A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. Journal of Hydrology, 583: 124299.

Song, Y., Liang, J., Lu, J., and Zhao, X., 2017. An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing 251: 26-34.

Sun, W., 2018. River ice breakup timing prediction through stacking multi-type model trees. Science of the Total Environment, 644: 1190-1200.

Sun, W., and Trevor, B., 2015. A comparison of fuzzy logic models for breakup forecasting of the Athabasca River. CRIPE 18th Workshop on the Hydraulics of Ice Covered Rivers, Quebec City, QC, Canada, August 18-20, 2015.

Sun, W., and Trevor, B., 2018. A stacking ensemble learning framework for annual river ice breakup dates. Journal of Hydrology. 561:636-650.

Sun, W., Lv, Y., Li, G., and Chen, Y., 2020. Modeling river ice breakup dates by k-nearest neighbor ensemble. Water, 12, 220.

Timoney, K., 2002. A dying delta? A case study of a wetland paradigm. Wetlands, 22: 282.

Turcotte, B., and Morse, B., 2013. A global river ice classification model. Journal of Hydrology, 507: 134-148.

Van Rossum, G., and Drake, F., 2009. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S., Brett, M., Wilson, J., Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, I., Feng, Y., Moore, E., VanderPlas, J., Laxalde, D., Pertkold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17: 261-272.

Wang, J., Sui, J., Guo, L., Karney, B., and Jüpner, R., 2010. Forecast of water level and ice jam thickness using the back propagation neural network and support vector machine methods. International Journal of Environmental Science & Technology, 7(2): 215 – 224.

Waskom, M., and the seaborn development team, 2020. Seaborn. Zenodo, https://doi.org/10.5281/zenodo.592845

United States Geological Survey, 1977. National Handbook of Recommended Methods for Water Data Acquisition. Office of Water Data Coordination, Reston, VA.

Uyanik, G., and Guler, N., 2013. A study on multiple linear regression analysis. Procedia – Social and Behavioural Sciences, 106: 234-240.

Yang, X., Pavelsky, T., and Allen, G., 2020. The past and future of global river ice. Nature, 577: 69-73.

Yu, M., Ma, J., Ono, K., Zheng, F., Fong, S., Gary, A., Chen, J., Demchak, B., Pratt, D., and Ideker, T., 2019. DDOT: A Swiss army knife for investigating data-driven biological ontologies. Cell Systems, 27: 267-273.

Zhang, J., and Luo, Y., 2017. Degree centrality, betweenness centrality, and closeness centrality in social network. Advances in Intelligent Systems Research, 132: 300-303.

Zhao, L., Hicks, F., and Fayek, A., 2012. Applicability of multilayer feed-forward neural networks to model the onset of river breakup. Cold Regions Science and Technology, 70:32-42.

Chapter 6: Conclusions

6.1 Conclusions and Contributions

This thesis presents a series of studies into the prediction of rare and volatile river ice breakup events. Machine learning techniques are coupled with input selection, imbalanced learning, novel modelling framework, and ontology-based techniques to achieve high levels of accuracy. These novel techniques present means of conducting these analyses that have not been achieved before, with levels of accuracy greatly exceeding traditional methods. Case studies include single rivers up to a national scale of Canada, demonstrating the success of the proposed techniques through extensive testing and analysis. The success of these studies demonstrates the value of non-traditional modelling frameworks in achieving high accuracy in cases where modelling is limited by small data sets and rare events. The developed techniques and models are easily transferable and can form the basis of decision-making support for affected communities throughout the country.

Chapter 2 developed a stacking ensemble framework for the prediction of river ice jam presence on the Saint John River in New Brunswick. Data requirements for the implementation of these techniques are outlined in depth. The techniques utilised in the selection of both variables and models as components of the ensemble are explained, with detailed overviews of each considered algorithm provided. The basis for model selection and means for assessing the best model combinations, in addition to other potential ensemble techniques such as voting are explored, with the potential for other ensemble combination techniques also present in adapting this methodology to other locations. The finished model delivered a combination of high accuracy with low computational demand after extensive testing of model combinations, with this study as

a whole providing a framework for the development of similar ensemble models for other rivers affected by ice jams.

Chapter 3 outlines the analysis of MWBs in Canada using a newly available dataset. The trends in the occurrence and severity of these events are first analysed using well established techniques, providing valuable results for subsequent modelling. These results are combined with input selection techniques to comprehensively analyse the drivers of MWBs in Canada on a national scale. The results of this analysis identified the key drivers of both MWB occurrence and severity on a national scale, and comparisons to past studies in combination with new machine learning analysis demonstrated both the accuracy and utility of these identified drivers. Through this, a new threshold for the occurrence of MWBs on a national scale is proposed whose accuracy greatly exceeds previous efforts and provides a valuable early warning tool to determine potential flood levels of these events.

Chapter 4 describes the development of a two-level modelling framework for the prediction of occurrence and timing of MWBs on a national scale in Canada using insights gained from the results of Chapter 3. This study features the first application of several rare-event forecasting techniques to river ice related phenomena, providing a basis for their implementation in future studies. The effectiveness of the developed two-level structure both as a means for the prediction of MWBs in Canada and as a framework for application to rare events forecasting in general is also demonstrated through a detailed and extensive case study. This novel framework can easily be transferred to other river phenomena such as ice jams, or in a more general sense to other rare events with similar data availability challenges.

Chapter 5 describes the development of a novel hybrid modelling system combining the advantages of semantic modelling and machine learning in the first study of this kind. The new

methodology developed within is used to produce an Ice Season Ontology, the first application of these techniques to river ice. This ontology is used in conjunction with network analysis techniques in the first hybridization of ontology-based modelling with machine learning to produce models capable of predicting spring breakup timing with increased accuracy over the non-hybrid models. This new methodology is easily transferred to other domains in both hydrology and civil engineering in general and represents a promising first combination of separate model techniques.

6.2 Recommendations for Future Work

(1) Though a variety of machine learning algorithms were tested for predicting each event, there remain a variety of algorithms and modelling techniques that are still untapped in these fields. Differing combinations of algorithms, deep learning, and applications of fuzzy logic and uncertainty analysis in these predictions has not been attempted, particularly in the case of MWBs, for which the work presented in this dissertation was the first application of machine learning directly to these event types. Further investigation of other model algorithms and techniques could yield better results and new insights into the drivers and severity of these events. Data used in these studies was inclusive of all that was available for the researchers at the time the models were developed, but increasing the data pool through acquisition of new data beyond the timespans of the existing data sets, or supplementing the data with additional variables that were either not available or obtainable through different means from those applied in these studies could also lead to improved results or new conclusions.

(2) The application of the ontology-based modelling in Chapter 5 represented the first use of these techniques in describing these river ice events, as well as the first integration of these techniques directly with machine learning. As presented, the ontology and machine learning models currently operate in separate software with different programming languages used in their

construction. Direct integration of these models would be an optimal next step in their application, allowing for results obtained from the machine learning model to be fed directly back into the ontology, and new observations added to the ontology to be directly assessed by the machine learning model. Through this, a single, user-friendly tool for the forecasting of breakup events could be developed and delivered to stakeholders. This would allow an intuitive means of decision-making support for affected communities while also allowing additional means of analysis to be conducted on the integrated data.

Appendix A: Detailed view of Ice Season Ontology divided by Event and Data Types

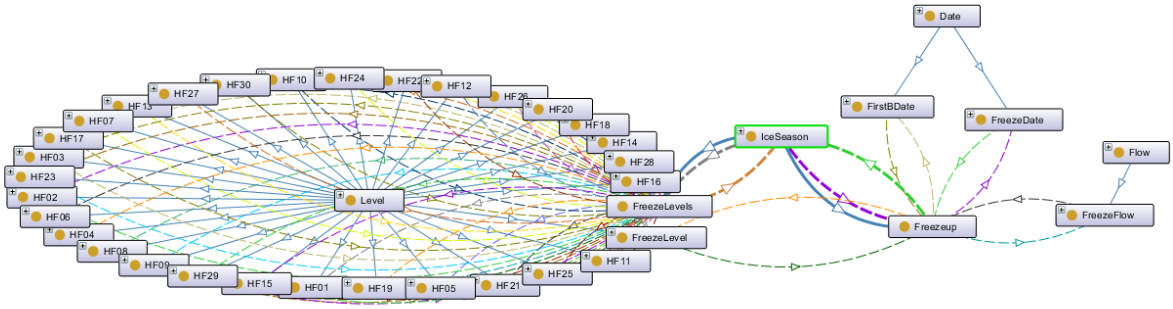


Figure A1: Subsection of Ice Season Ontology divided by event detailing Freeze-up events.

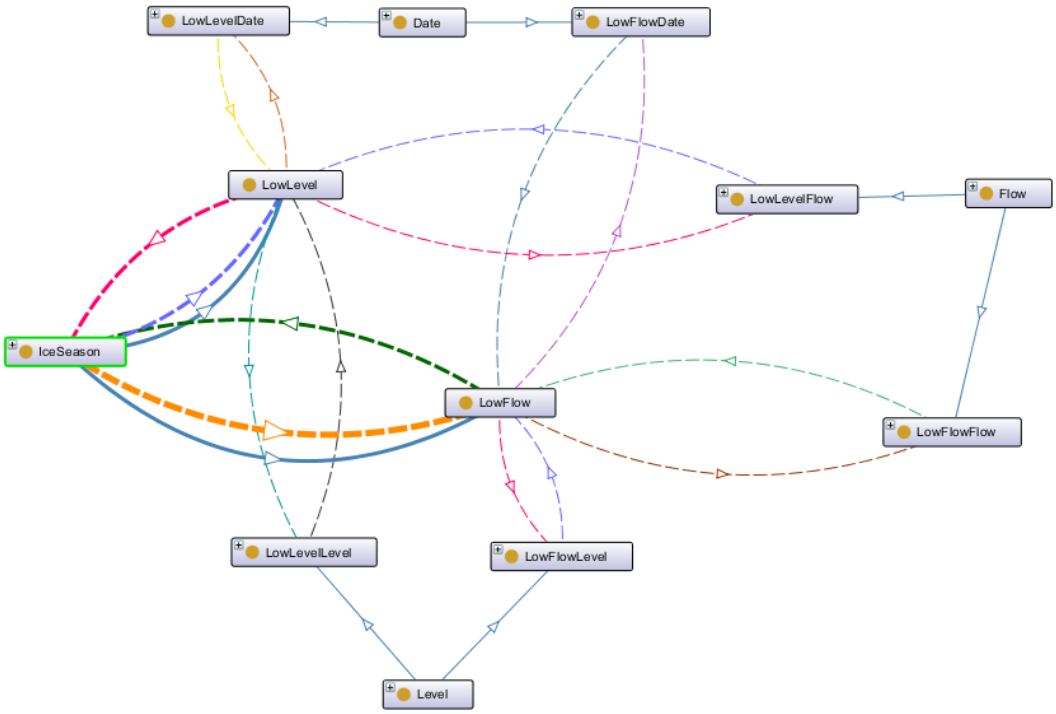


Figure A2: Subsection of Ice Season Ontology divided by event detailing Winter Low events.

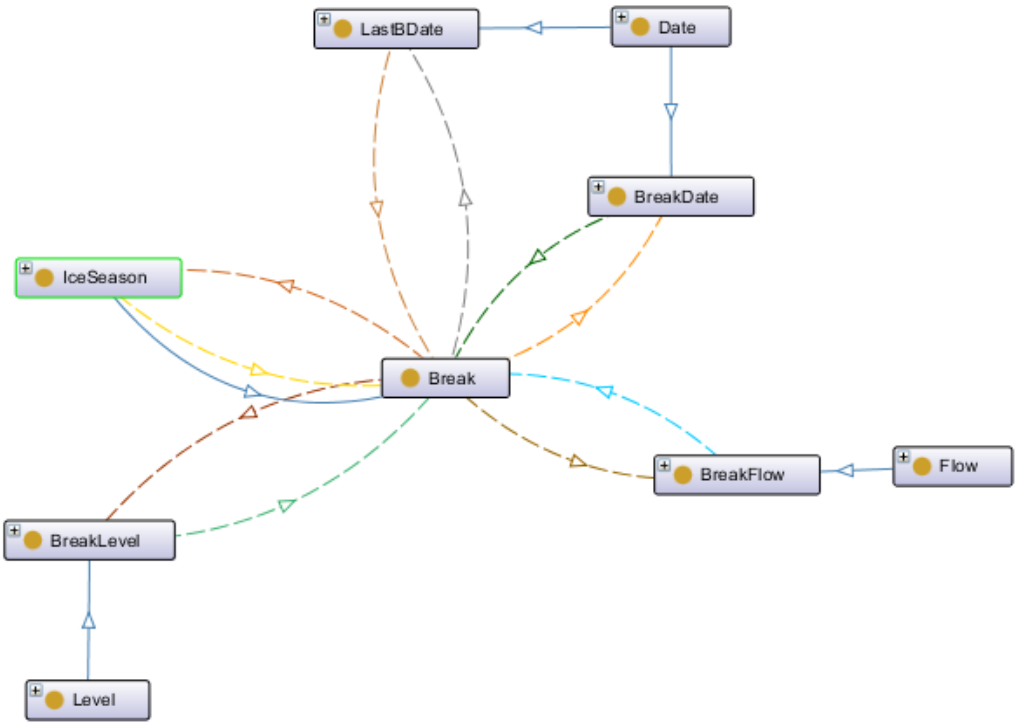


Figure A3: Subsection of Ice Season Ontology divided by event detailing Breakup events.

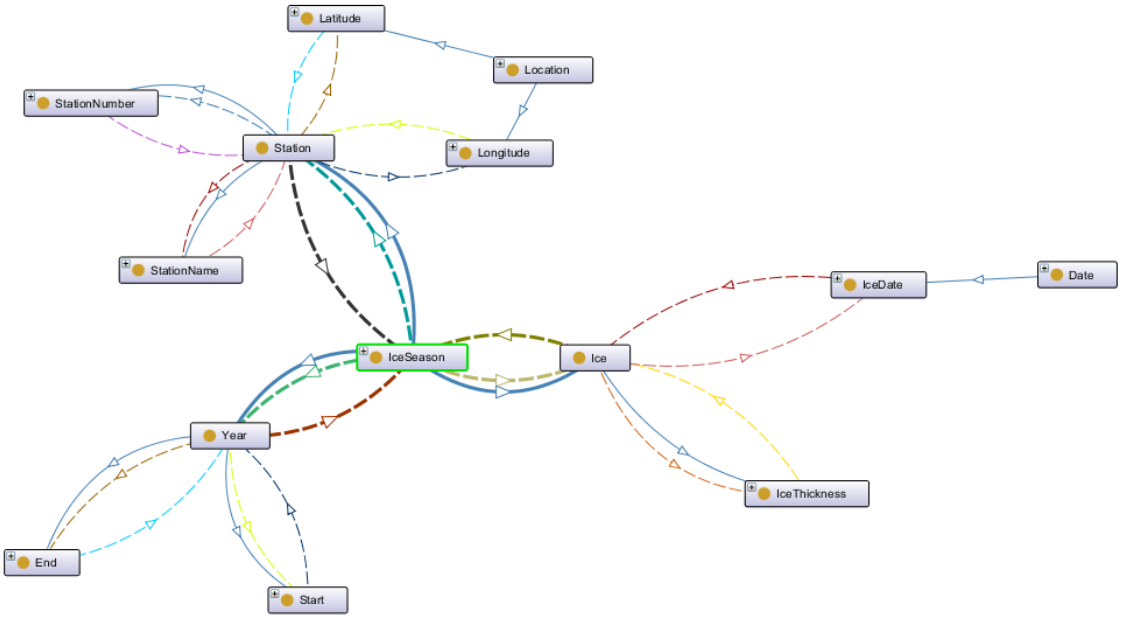


Figure A4: Subsection of Ice Season Ontology divided by event detailing Miscellaneous Season Values.

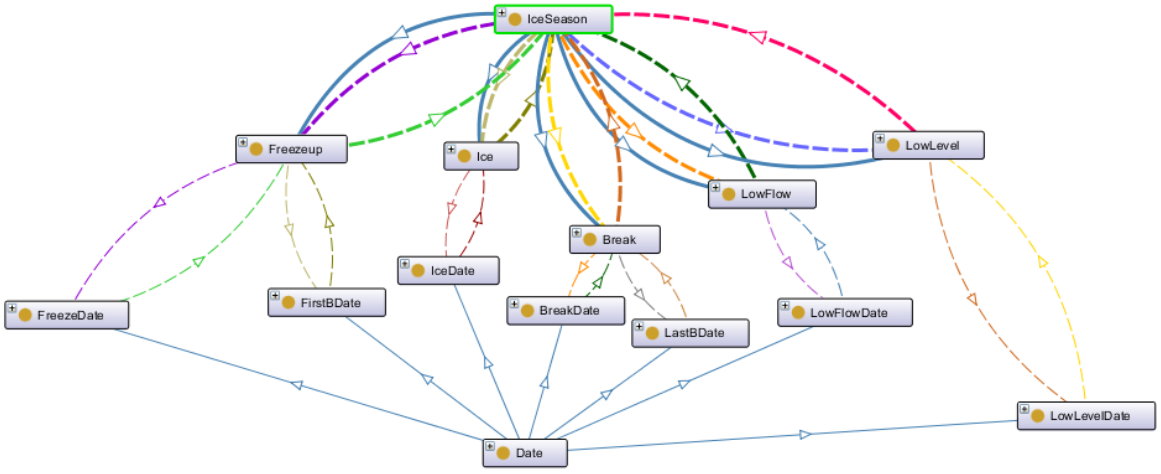


Figure A5: Subsections of Ice Season Ontology divided by datatype detailing Dates.

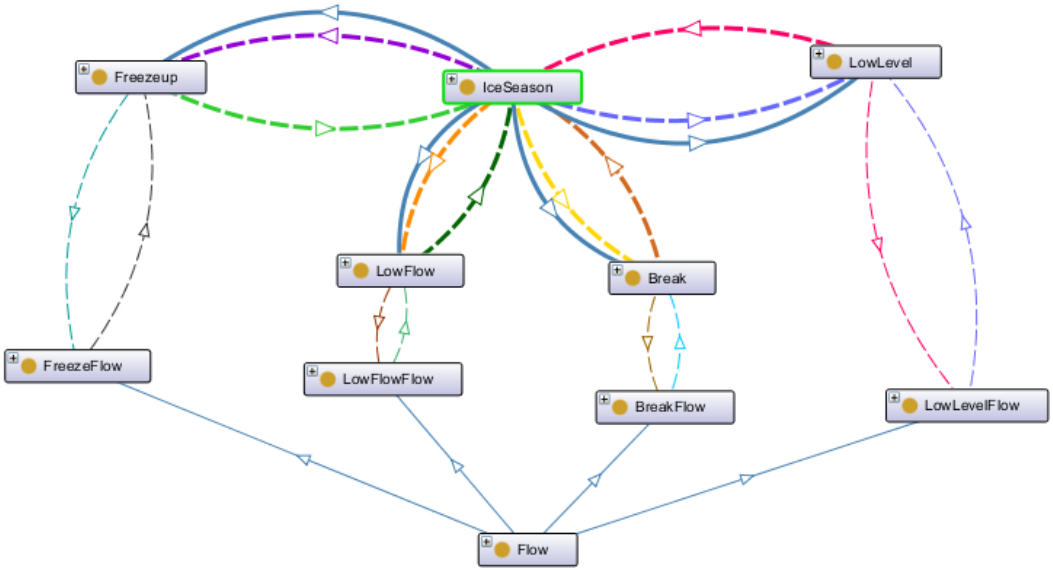


Figure A6: Subsections of Ice Season Ontology divided by datatype detailing Flows.

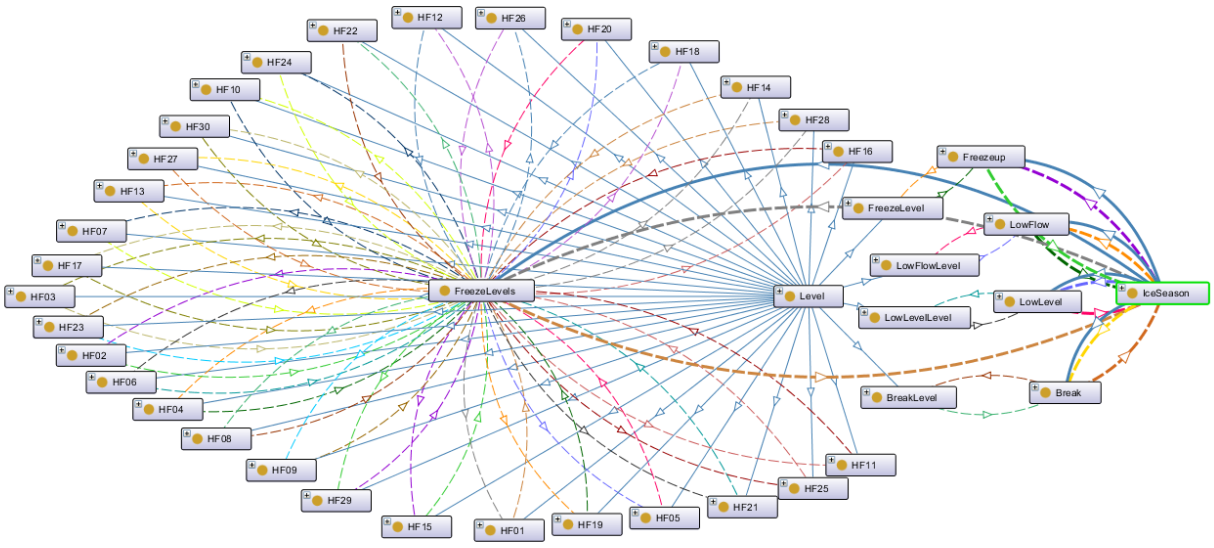


Figure A7: Subsections of Ice Season Ontology divided by datatype detailing Water Levels.

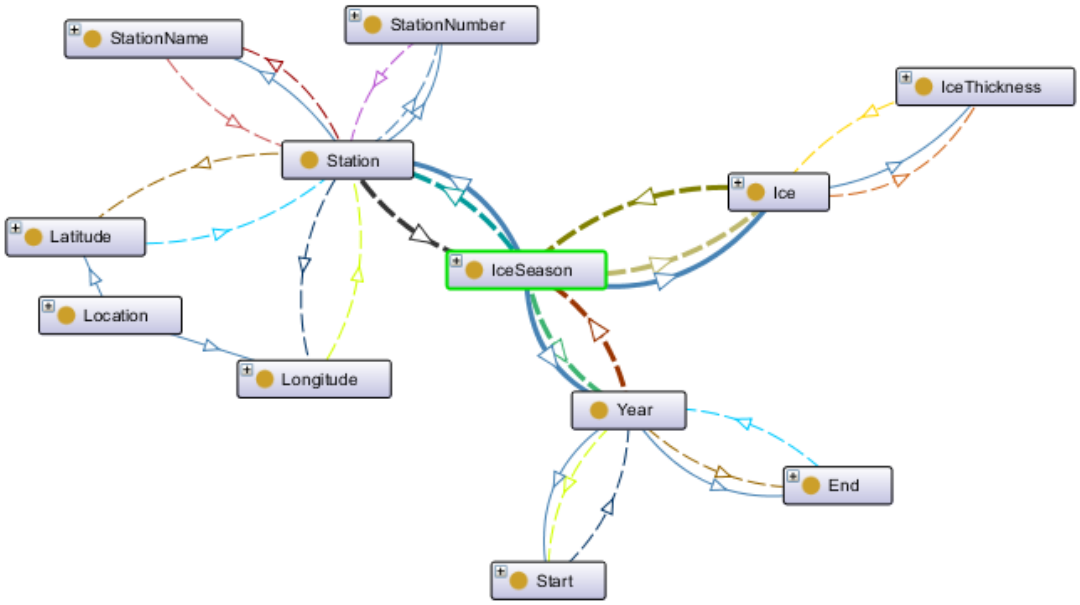


Figure A8: Subsections of Ice Season Ontology divided by datatype detailing Miscellaneous Season Values.