Using Bootstrap Extreme Learning Machine and Least Squares Support Vector Machine Models For River Ice Thickness Estimation

By

Mahsa Ghasri

A thesis submitted to McGill University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Bioresource Engineering

MacDonald Campus of McGill University

Ste. Anne de Bellevue, Quebec, Canada

August 2017

Abstract

River ice plays a pivotal role in biophysical and socio-economic systems in Northern regions and significantly impacts climate variability and change in both regional and global scale. Collecting information and analyzing characteristics of river ice cover are important aspects to be considered in order to address engineering and environmental problems. However, poor accessibility and site-specific challenges of many river systems prevent up-to-date monitoring of ice cover regime. As a result, in this study, the potential of using easy to measure meteorological variables as predictors were investigated for estimation of river ice thickness with extreme learning machine (ELM), least squares support vector machine (LSSVM), and their bootstrap methodologies (BELM, BLSSVM, respectively). Based on the correlation analysis water level, accumulated freezing degree day, and mean temperature are employed to establish the best estimating model. Also, two imputation techniques, namely the Kendall–Theil Robust Line (KTRL) and the regularized expectation maximization (RegEm) were utilized to impute the missing values in the meteorological records. The estimation metrics defined as the correlation coefficient (R), Nash-Sutcliffe efficiency (E_{NS}), root-mean-squared error (RMSE), Bias, and mean absolute error (MAE) were computed to assess the models' accuracy. The results indicated that bootstrap ELM model outperformed ELM, LSSVM, and BLSSVM models in the testing phase across a number of statistical measures. Accordingly, R=0.90, RMSE=0.080 (m), E_{NS}=0.71, MAE=0.072, and Bias=0.991 was exhibited from BELM using KTRL imputation technique and R=0.93, RMSE=0.067, Nash=0.80, MAE=0.06, Bias=1.003 from BELM using RegEm. Based on the findings of this study, presented machine learning techniques using meteorological variables are promising tools for river ice thickness estimation.

Résumé

Les glaces fluviales jouent un rôle crucial dans les systèmes biophysiques et socioéconomiques des régions septentrionales, et ont une incidence considérable sur la variabilité et le changement climatique regional et global. De recueillir des informations sur le couvert de glace fluvial et d'en faire l'analyse des caractéristiques sont d'importants aspects à prendre en compte afin de s'adresser à la problématique technique et environnementale. Cependant, les difficultés d'acceder aux sites fluviaux et d'y travailler empêche une suivie au jour le jour du régime des glaces fluviales. En conséquence, la présente étude enquêta sur la possibilité d'employer des variables météorologiques faciles à mesurer comme variables explicatives de l'épaisseur des glaces fluviales avec des logiciels *extreme learning machine* (ELM) machine à vecteurs de support par moindres carrés (LSSVM), et leurs méthodes *bootstrap* (BELM, BLSSVM, respectivement). Suite à une analyse corrélative du niveau des eaux, des degrés-jours de gel cumulatifs et de la température moyenne, ces paramètres furent choisis pour établir le modèle d'estimation le plus performant. Deux techniques d'imputation statistique, soit la ligne Kendall-Theil (KT), et l'espérance-maximisation régularisée (EMR) servirent à attribuer des valeurs manquantes dans les dossiers météorologiques. Le coefficient de corrélation (R), le coefficient d'efficacité Nash-Sutcliffe (E_{NS}), l'erreur quadratique moyenne (RMSE), le biais (Bias), et l'erreur absolue moyenne (MAE) servirent d'indicateurs d'exactitude pour les modèles comparés. Selon plusieurs critères, l'exactitude du modèle BELM en phase de validation surpassa celle des modèles LSSVM et BLSSVM. Le modèle BELM avec imputation KT (R=0.90, RMSE = 0.080, E_{NS}=0.71, MAE=0.072, et Bias=0.991) fut surclassé par celui avec imputation EMR (R=0.93, RMSE=0.067, E_{NS}=0.80, MAE=0.06, Bias=1.003). Selon les résultats de cette étude, les méthodes *learning machine* offrent un outil prometteur pour l'estimation de l'épaisseur des glaces fluviales.

Acknowledgements

I would like to thank my supervisor, Dr. Zhiming Qi for the opportunity that he gave me to study under his supervision. I am very grateful for his support, guidance, and understanding over these last two years. He was always available for discussions, and he provided the most useful advice and comments on my research. Thank you for believing in me.

I would like to express my gratitude to Dr. Jan Adamowski who guide me on my research topic. He has provided valuable suggestions and feedback for my project, and I am very grateful for all the time that he spend in helping me and improving my work. Thank you for trusting me to carry on your research topic.

I would like to extend my gratitude to Dr. John Quilty, who has been very kind, helpful, and supportive of my research. He has encouraged me to try new techniques and inspired me with his enthusiasm for teaching me from very first day. I am very grateful for the time you spent in helping me whenever I run into problems and always being responsive to my emails.

I would also like to extend my gratitude to people who have assisted me during my studies and through some tough times. In particular, I would like to appreciate Dr. Meisam Vadiati for introducing me this field of research. Also, I wish to express my most heartfelt thanks to my close friends Amir Tabandehjooy, Faezeh Eslamian, Sahar Zamani, Hamed Vatankhah, Shima Keisandokht, and Maryam Kargar.

I wish to thank my parents, Morteza Ghasri and Marzieh Yavari, for supporting me unconditionally and making my dreams come true. I wish to thank my sister, Mina Ghasri, for showing me the light toward success and for her unconditional love. This thesis is for you.

Contribution of Authors

This thesis has been prepared for submission as manuscripts to Journal of Hydrology, which is a peer-reviewed journal.

The author of this thesis was responsible for determining the step-by-step procedures involved in the methodology for data analysis, performing the data analysis, and preparing the manuscript for journal submissions. Dr. Zhiming Qi is the supervisor of this thesis; He reviewed this thesis and is a co-author of the manuscripts of this thesis. Dr. Jan Adamowski also supervised this research; he provided the original data for this thesis, and provided guidance and advice regarding many different aspects covered by this thesis. He also reviewed this thesis and is a co-author of the manuscripts of this thesis. Dr. John Qulity of the Department of Bioresource Engineering at McGill University is also a co-author of the manuscripts. He provided statistical and technical guidance during the data analysis and helped review the thesis. He also assisted in compiling the MATLAB codes used to analyze the data.

Table of Contents

List of tables

List of figures

List of Acronyms

AFDD	Accumulative freezing degree day
ANFIS	Artificial neuro-fuzzy inference system
ANN	artificial neural network
ARIMA	Autoregressive integrated moving average
ASR	Accumulative solar redaction
AVHRR	Advanced Very High-Resolution Radiometer
BELM	bootstrap extreme learning machine
BIAS	bias
BLSSVM	bootstrap least squares support vector machine
CLIMO	Canadian Lake Ice Model
CUMS	Accumulative snow
ELM	extreme learning machine
GIS	Geographic information system
KTRL	Kendall-Theil robust line
LSSVM	least squares support vector machine
MAE	mean absolute error
MLP	multiple linear regression
MODIS	Moderate Resolution Imaging Spectrometer
MSC	Water Survey of Canada
NASH	Nash Criterion
R	correlation coefficient
RegEM	regularized expectation maximization

RICE	River ice
RMSE	Root-mean-square error
RSL	Revised Stefan's law
SAR	synthetic aperture radar
SL	Stefan's law
SVM	support vector machine

List of Mathematical Symbols

H= ice thickness

D_d= sum of the degree-days below the freezing point

A_0= an empirical constant

D_g= accumulation of freezing degree-days

C= adjustable parameter

(τ)= Kendall rank correlation coefficient

b_k= slope estimator

a_k= KTRL intercept

x_m = vector of missing values

μ_m= vector of means of the missing values

x_a = vector of available values

μ_a = vector means of the available values

B= matrix of regression coefficients

e= assumed residual

h= a positive number called the ridge parameter

$\widehat{D}$= the diagonal matrix

z= the model output

β= the output weights

H=nodes

G= hidden layer output matrix.

G^+= inverted hidden layer matrix,

β^*= estimated output weights from N data records

w=d-dimensional weight vectors

x_t = predictor

y_t= predictand,

$\emptyset$= the mapping function

b= the bias term.

γ= the margin parameter

e_t= the slack variable for x_t

E_0= The generalization error

T^s= bootstrap sample

$\hat{y}(x)$= the boostsrap model estimate

$x_{normalized}$=the normalized value

x_i= current input variable

x_{min}= minimum value within the historical dataset

x_{max}= maximum value within the historical dataset

y_t= observed variable

$\hat{y}_t$= estimated variable

$\bar{y}_t$= mean values of dependent variable

N= the sample size.

Chapter 1: Introduction

1.1 Introduction

Representing the frozen part of the terrestrial climate system, the cryosphere is comprised of several subsystems: ice sheets, ice shelves, ice caps, glaciers, sea ice, lake ice, river ice, ground ice, and snow. Ice sheets are large (exceeding 50,000 km^2) ice masses on land, whereas ice shelves consist of floating ice nourished by the inflow from an adjacent ice sheet, typically stabilized by large bays. In contrast, the smaller (under 50,000 km^2) land-based ice masses termed ice caps or glaciers, and are constrained by topographical features (*e.g.*, mountain valley). In contrast to an ice shelf, sea ice floats on the ocean and forms directly by freezing sea water. Similarly, lake ice and river ice form directly on the lake and river water, respectively. Ground ice occurs as permafrost: soil that stays in a frozen state year-round. Snow is precipitation of crystalline water ice, containing a multitude of snowflakes, which accumulate on the ground at a bulk density significantly less than that of ice (Greve and Blatter, 2009).

Freshwater lake and river ice is estimated to cover a total area of 1.7×10^6 km^2 over the Northern Hemisphere (estimated at peak thickness, north of the January 0°C isotherm, and excluding the Greenland ice sheet), and represents a volume of 1.6×10^3 km^3 (Brooks et al., 2013). The estimated fresh water ice is approximately equal to the Greenland ice sheet (the second largest ice body in the world), and its volume to that of snow on land (Duguay et al., 2015). Accordingly, it is considered a major component of the terrestrial landscape. Majoritarily located in Northern Hemisphere, lakes cover nearly 2% of the Earth's land surface. Regional climate and weather events are affected by the presence or absence of ice cover on lakes during the winter months (Brown and Duguay, 2010) . Consequently, monitoring of lake ice is key to forecasting high-latitude weather, climate, and river run off. Moreover, modeling the energy and water balance of high-latitude river basins, in order to improve numerical weather predictions in regions where lakes occupy a substantial fraction of the landscape makes the consideration of this phenomena even more important (Martynov et al., 2012; Zhao et al., 2012b). In the Northern Hemisphere, in addition to the lake ice cover, river ice also affects an extensive

portion of the global hydrologic system: significant ice cover develops on 29% and seasonal ice affects 58% of the total river length (Prowse et al., 2007). For large rivers in cold regions ice cover can persist over the entire river length for over half the year, rivers with more temperate headwaters experience the long-term ice on only some reaches (Prowse et al., 2011b). For such rivers, ice processes govern the timing and extent of extreme hydrological events such as low flows and floods (Prowse et al., 2007). The broad ecological and socio-economic significance of river ice raise scientific concerns regarding how future changes in climate might affect river ice processes (Strategy, 2007).

Freshwater ice, a sensitive indicator of climate variability and change, has been significantly affected by air temperature changes. Long-term trends observable from terrestrial records reveal increasingly later freeze-up, earlier break-up dates, and decreased ice thickness, closely corresponding to increasing air temperature trends, but with greater sensitivity at the more temperate latitude(Brown and Duguay, 2010; Prowse et al., 2011a). Extensive spatial patterns in these trends are also associated with the principal atmospheric circulation patterns originating from the Pacific and Atlantic oceans: El Nino-La Nina/Southern Oscillation, the Pacific-North American pattern, the Pacific Decadal Oscillation, and the North Atlantic Oscillation/Arctic Oscillation (Bonsal et al., 2006; Prowse et al., 2011b).

Although the presence of freshwater ice can pose a significant ecological and socio-economic problems in northern regions, the absence of ice-induced hazards, such as ice jamming and flooding can decrease the flux of sediments and nutrients to wetlands. This process can result in increased soil salinity, along with shifts in flora and fauna which can eventually affect the traditional subsistence living of local and aboriginal communities along rivers and lakes (Jeffries et al., 2012; Prowse et al., 2011b; Prowse and Culp, 2003).

Despite the wide-ranging influence of freshwater ice, significant gaps exist in our knowledge and understanding of lake and ice river characteristics, properties, and processes. Additionally, poor accessibility and site-specific challenges of the collection of information on freshwater ice regimes add an extra burden to freshwater ice studies.

Consequently, different forecasting and simulation tools have been developed and applied for up-to-date monitoring of processes and patterns of ice cover. These tools have been used to map ice cover extent and characterize ice phenology and timing (*e.g*., timing of ice break up-freeze up, ice thickness, ice jam flooding over large freshwater surfaces, particularly large lakes. Since the formation and breakdown of ice cover on rivers (*vs.* lakes) occurs under much more dynamic conditions, river ice monitoring has proven to be more difficult (Chu and Lindenschmidt, 2016). In this regard, the reliability of developed forecasting tools must be enhanced to achieve the level of accuracy required by environmental modeling community (*e.g*., numerical weather forecasting, and hydrological forecasting) or public policy and decision-makers.

1.2 Thesis objective

The principal aim of this research was to investigate the accuracy of different machine learning techniques and their associated bootstrap methodologies for river ice thickness estimation in Alberta Canada. The particular focus of this thesis is as follow:

- To impute missing values of the presented dataset with two imputations methods namely The Kendall–Theil Robust Line (KTRL) and the regularized expectation maximization algorithm (RegEm). The performance accuracy of the applied methods will be compared based on performance metrics.

- To compare Artificial Neural Network (ANN) modelling with a relatively newer form of machine learning techniques namely Extreme Learning Machine (ELM) and Least Square Support Vector Machine (LSSVM) in terms of their performance accuracy for river ice thickness estimation.

- To apply the Bootstrap methodology of the Extreme Learning Machine (ELM) and Least Square Support Vector Machine (LSSVM) in river ice thickness estimation in order to investigate model precision.

Chapter 2: Literature review

Ice formation on lakes and rivers play a pivotal role in shaping water resource management and engineering in cold regions. The presence of freshwater ice has many implications: ice jams which lead to widespread flooding, reduced hydroelectric power generation, navigation and transportation disruption, as well as damage to man-made structures, the environment, and ecosystems (Duguay et al., 2015; Shen and Liu, 2003). Profoundly affected by river ice through a number of complex interactions with hydrological and meteorological conditions, the hydrodynamic, mechanical, and thermal processes of global hydrological system, particularly those portions situated in the Northern Hemisphere, become altered. The categorization of evolving river ice growth into essential phases of formation, evaluation, transport, accumulation, dissipation, and deterioration has led to significant advances in the understanding of the physics of river ice (Shen, 2010). However, gaps remain in our knowledge of river ice characteristics, properties, processes, and consequences that make it difficult to link global change and sustainable development with ice phenology (Jeffries et al., 2012).

Investigating different aspects of ice conditions requires extensive data on break-up date, freeze-up date, and ice thickness, amongst other parameters. Moreover, in northern regions, like Canada, the collection of data on freshwater ice is often challenging due to the site-specific nature of such processes (Chu and Lindenschmidt, 2016). Thus, different forecasting and simulation models are required.

Amongst these parameters, ice thickness is particularly useful in river-ice hydraulic models (Chokmani et al., 2007), flood routing models (She and Hicks, 2006), hydrodynamic modeling of freezing rivers (Shen, 2003) and flow control in regulated rivers (Tuthill, 1999). In addition, river ice cover influences biological ecosystems and aquatic fauna (Huusko et al., 2007).

Facing the lack or incompleteness of ice thickness measurements, different analytical, mathematical, and numerical models have been used to address this inadequacy of data.

While all numerical models developed for ice thickness simulation take into consideration energy budget components, they differ from each other by the level of detail used to compute the ice energy budget (Andres and Van der Vinne, 2001; Hicks et al., 1997; Ma and Fukushima, 2002). The one-dimensional numerical model for river ice processes, RICE, can simulate the growth and decay of the ice cover along with other associated parameters with a close agreement between observed and simulated data (Shen et al., 1991). The updated RICEN model provides substantial improvements in ice process simulation, including considering the effects of the wind, artificial icebreaking by icebreakers and flow resistance due to moving ice (Shen et al., 1995). However, these kinds of models require extensive input data (e.g., wind speed, vapor pressure, cloud cover), which are not always accessible. In addition, indices that are needed for estimating energy fluxes due to shortwave radiation, evaporation, and convection at the air-ice (or snow) interface cannot as yet be accurately determined (Shen and Yapa, 1985). Consequently, simplified versions of energy budget models can be developed using Stefan's law (SL) or the revised Stefan's Law (RSL) — detailed description of these equations can be found in the theoretical background section.

The date of the onset of the ice cover is a primary parameter in the Stefan's law equation as it determines the date of onset of the accumulation of freezing degree-days for any given winter. As in any one season this date is unknown for most sites, Stefan's law cannot be used; however, the RSL, based on the accumulation of freezing degree-days from the first day of below-freezing air temperatures can work reasonably well for ice thicknesses exceeding 0.10 m. However, for ice thicknesses inferior to 0.10 m, the RSL overestimates ice thickness (Ashton, 1989). Based on a theoretical analysis Shen et al. (1985) introduced an modified degree-day method capable of continuously simulating the variation of the thickness of the river ice cover from formation to break up. The method's application to the upper St. Lawrence River showed simulated ice thicknesses to compare well with field observations, but an empirical coefficient was still required to judge the ice thickness growth rate at different types of river sites.

Geographic Information System (GIS) and remote sensing techniques offer an alternative means of monitoring the processes and pattern of ice cover. There are different remote sensing system that can be used to map ice cover extent and ice phenology (Chu and Lindenschmidt, 2016). Optical sensors with low to medium spatial resolution and high temporal resolution, such as the Advanced Very High-Resolution Radiometer (AVHRR, 1 km to 4 km spatial resolution) and Moderate Resolution Imaging Spectrometer (MODIS, 250 m to 1 km spatial resolution) are two types of ice information collection methods used for large freshwater ice surfaces. Even though high temporal resolution data such as MODIS and AVHRR can provide timely information about river ice conditions, the low spatial resolution and dependence on atmospheric conditions for the obtention of these data can pose significant challenges in monitoring the details of river ice cover conditions. Given its higher spatial resolution, capacity to acquiring images under all weather and atmospheric conditions (*e.g*. clouds), as well as its sensitivity to water-ice surfaces, ice structure and ice thickness (Duguay et al., 2015; Nghiem and Leshkevich, 2007; Unterschultz et al., 2009), microwave remote sensing (*e.g.*, synthetic aperture radar — SAR) has proven to be another useful tool in monitoring and studying freshwater ice processes). Numerous studies have used SAR imagery to monitor river ice thickness, and have reported promising simulation results (Karl-Erich et al., 2010; Mermoz et al., 2014; Unterschultz et al., 2009). However, the main drawback of the numerical physically-based models is that they are complex and require some degree of user experience and expertise with the model.

This limitation of numerical models has encouraged researchers to develop analytical models which have an acceptable level of accuracy along with a simple operating procedure (Chokmani et al., 2007; Zaier et al., 2010). Parameters having an influence on ice thickness have been identified through single variable regression and factor analysis, and can then be used in combination in a multivariable linear regression analysis to address the complexity of ice growth and decay (Williams et al., 2004). However, while accumulated freezing degree days and mean temperature over a fixed period provided the best estimation of maximum ice thickness, freezing degree day is not

a viable predictor as it requires knowledge of the date of maximum ice thickness. Similarly, the effect of the wind, solar radiation, and snow cover were not considered. Accordingly, Dornan (2005) developed a multiple linear regression to determine the best model for river ice growth by considering the effect of climate variables such as snow cover, snow density, and solar radiation. For 11 hydrometric stations, 11 pairs of models were developed (one through Forward Stepwise Regression and one through Backwards Stepwise Regression for each pair) and compared to assess their overall ability to predict river ice growth. Six of the models performed adequately, and five inadequately. With regards to the correlation of climatic variables with ice thickness, cumulated solar radiation was deemed a non-significant parameter in the ice thickness growth models given its negative correlation with the most relevant model parameter, cumulative freezing degree days.

Another useful tool for river and lake ice growth modelling is the artificial neural network (ANN). Comparing the accuracy of lake ice growth simulation with an ANN model, multiple linear regression (MLR) model, revised Stefan's Law (RSL) model, or the Canadian Lake Ice Model (CLIMO), all drawing on data for daily mean air temperature, daily rainfall, daily total solar radiation, and daily snow depth, Seidou et al. (2006) reported that the ANN model to follow data variations more closely than the RSL. Simulating lake ice growth for a site in Fort Reliance, NWT, Canada with CLIMO for the years 1990-1997 resulted in an root-mean-square error (RMSE) of 180 mm, only slightly higher than that for a RSL model (171 mm) or an ANN model (168 mm). Thus, despite the CLIMO model's much greater complexity, its performance was comparable to that of ANN, MLR, and RSL models . Assessing ANN-based river ice thickness estimation methods drawing upon cumulative climate input variables such as freezing degree day (CFDD), solar radiation (CSR), and snow (CUMS), Chokmani et al. (2007) found that, for sites in Alberta, these models provided good estimates, with 90 mm < RMSE < 130 mm. However, as the models failed to estimate low and high values correctly, they were not considered optimum. In seeking to estimate lake ice thickness Zaier et al. (2010) employed an ANN ensemble technique, wherein the use of multiple models can improve the generalization

ability of a single model (Shu and Burn, 2004). Drawing on identical data from a number of Canadian lakes , Zaier et al. (2010) tested six ANN-based models, five employing ensemble modeling techniques, and one (control), a single ANN model developed by Seidou et al. (2006). Improvement in generalization ability was achieved by using stacking for combining member networks. Moreover, in most cases, boosting was deemed the best method for reducing the estimation error. Overall, with $0.74 \leq E_{NS} \leq 0.96$, the ANN ensemble provide satisfactory estimation for lake ice thickness at most stations.

Besides ANN and MLR, support vector machine (SVM) and fuzzy logic have been successfully used in the estimation of other ice characteristics, *e.g.*, ice affected stream flow (Chokmani et al., 2008), ice jam occurrence (Massie et al., 2002), water level and ice jam thickness (Wang et al., 2010), and date of ice break-up and freeze-up (Mahabir et al., 2005; Shouyu and Honglan, 2005; Tao et al., 2008; Zhao et al., 2012a).

Despite the extensive application of ANN in the modeling of hydrological processes (and freshwater ice), the forecasts generated still suffer from deficiencies such as over-fitting, slow learning speed, and local minima. To overcome these limitations, a relatively newer form of machine learning models, extreme learning machine (ELM) and least squares support vector machine (LSSVM) have been developed. ELM models' faster learning algorithms and improved generalization performance have led to their extensive use in different modeling processes (Deo and Şahin, 2015). In addition, advanced design features of ELM such as analytical output determination by a least squares problem and random generation of the parameters of hidden nodes without the need for tuning the algorithm makes ELM a proper substitution for traditional forms of machine learning models (Yaseen et al., 2016). Likewise, LSSVM models, modified versions of SVM using kernel methods, have shown the capacity to estimate output parameters more precisely than traditional techniques such as ARIMA, ANN, ANFIS, or neuro-fuzzy systems (Hong and Pai, 2006; Wang et al., 2009). The initial motivation for using kernel method for time series estimation arises from their greater capability in the modeling of non-linear, non-stationary, and not characterized phenomena (Sapankevych and Sankar, 2009).

In addition to the practice of newer forms of machine learning techniques, researchers have started to investigate particular procedures, such as bootstrapping, in their modeling processes. The bootstrap (Efron, 1979), an intensive resampling with replacement, is a computational technique that has been employed in specific applications to reduce uncertainty (Tiwari and Chatterjee, 2010b). This method has been successfully applied to hydrological modeling due to it yielding more accurate results, especially in the absence of adequate training data (Erdal and Karakurt, 2013).

To the best of our knowledge, no studies have been explored the potential of ELM and LSSVM coupled with the bootstrap technique in river ice thickness estimation. The performance of these models was compared with ANN to evaluate their relative estimation accuracy. Several quantitative performance indicators are employed for model comparison. In addition, the inherent accuracy of two imputation methods, KTRL and the regularized EM algorithm, were evaluated through the applied models' relative accuracy. The imputation of missing values has been considered in this study since historical river ice thickness records are minimal and ignoring any observations with missing variables would result in a significant loss of information.

Chapter 3: Methods

3.1 Study area

Arising from snow and glacial meltwater in the Columbia Icefields of the Rocky Mountains of southwestern Alberta (Figure 1), the Athabasca River is the second largest river in Alberta, and its largest unregulated river. The river initially passes through the rugged forested montane landscapes of Jasper National Park and onward through Brule and Jasper Lakes into rolling foothills. The Athabasca River travels northward through the boreal mixed wood forest and encounters sudden changes in its physical properties near Fort McMurray (Figure 1). These changes are responsible for the frequent formation of ice jam in this location. The Clearwater River joins the main stream immediately downstream of Fort McMurray and then drains into Lake Athabasca. The entire Athabasca River basin is approximately 1.59×10^5 km^2, which represents about 24% of Alberta's landmass (Peters et al., 2013). Socio-economic impacts of river ice have been felt in Fort McMurray over its entire history — from its early years as a Hudson Bay Company outpost to the modern era as an oil boom municipality (Alberta Environment, 1985). In response to these impacts and frequent ice jams, this area was chosen to test the feasibility of machine learning techniques in the estimation of ice properties.

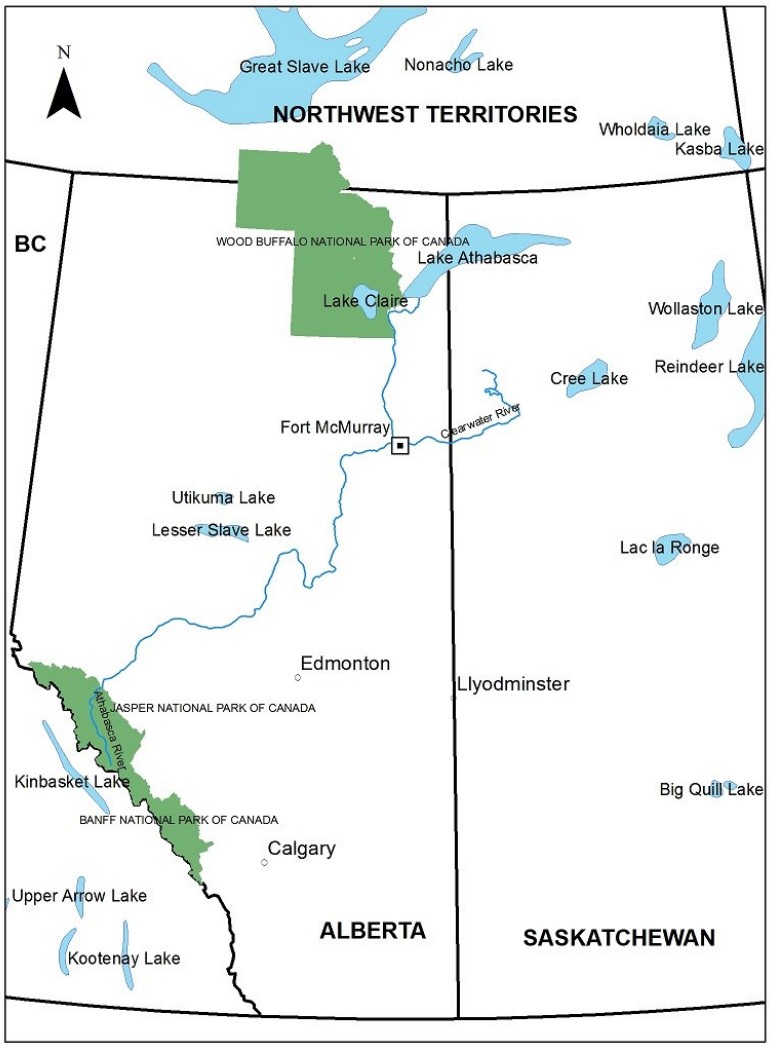


Figure 1. General overview of the study area

In the present study, direct measurements of ice thickness were collected from Water Survey of Canada (MSC) winter gaugings at the Clearwater River at Draper (07CD001) hydrometric station (Figure 2). The closest meteorological station to the hydrometric site, Fort McMurray A (Figure 2), was used for meteorological observation records in order to calculate the explanatory variables assumed representative to climate conditions at the hydrometric stations. The meteorological variables considered were maximum, minimum, and mean daily temperature, total precipitation, and snow on the

ground. These were recorded on a daily basis with some missing values. As river ice thickness is a function of temporal changes in meteorologic parameters, accumulated freezing degree days (AFDD) and accumulated solar radiation (ASR) were also calculated (Dornan, 2005). However, the absence of 'snow on the ground' data for the period of 2009-2012 led to its omission and that of cumulative snow (CUMS) as input variables. Measurement of ice thickness was limited to one to seven measurements per season. In total, 110 records of ice thickness for station 07CD001 were collected during the period from 1979 to 2012.

An important step in machine learning modeling is a selection of proper input variables. Irrelevant inputs can significantly influence model accuracy or add unnecessary complexity impacting model reliability (Hejazi and Cai, 2009). In this regard, correlation analysis was applied to all predictors to find the best set of input variables for river ice thickness estimation. Three predictors, water level (m), accumulated freezing degree days, and mean air temperature (°C), showed the greatest correlation with ice thickness, and were accordingly selected as inputs.

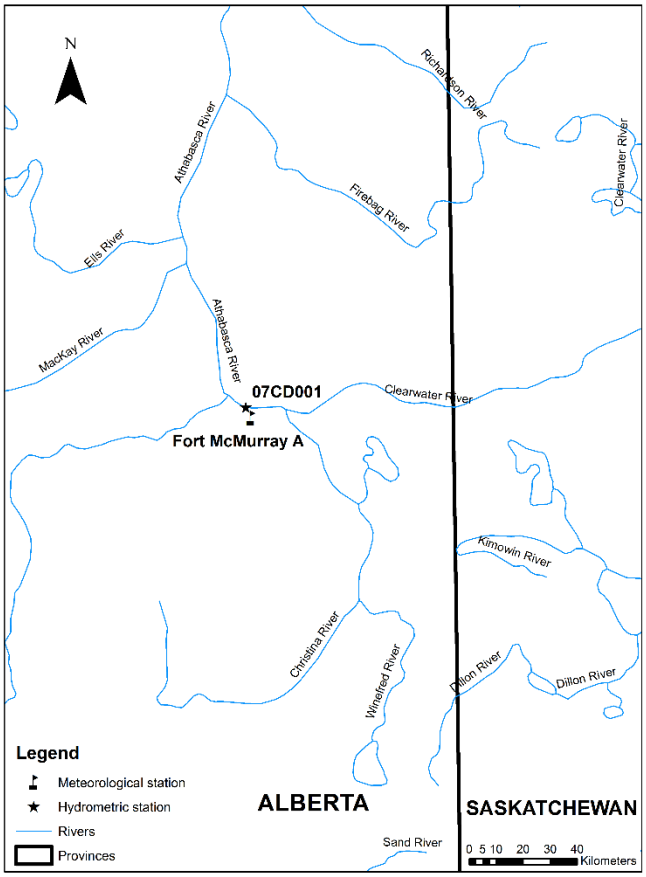


Figure.2 Meteorological and hydrometric stations' locations

3.2 Theoretical background

3.2.1 Stefan's law and revised Stefan's law

Stefan's law or degree-day method is the most widely used formula which it is driven by simplifying the equations obtained using energy balance (Lock, 1990):

$$H = A_0\sqrt{D_d} \tag{1}$$

where,

A_0 is an empirical constant that varies with the site,

D_d is the sum of the degree days below the freezing point, and

H is the ice thickness.

In practice, $A_0$is used as an adjustable parameter with a value that is lower than the theoretical value to account for varying conditions of exposure and insulation. Michel (1971) gives a range of values adapted for a variety of lakes and rivers. As explained in the introduction, the date of onset of ice cover, a basic parameter in Stefan's equation, is not known for the majority of sites. The "Revised Stefan's law" (RSL) is a substitute equation which is based on D_g, the accumulation of freezing degree-days starting with the first day of below freezing air temperatures in any given season. The RSL presents one more adjustable parameter, C, the effective number of degree-days to be subtracted from $D_g$in order to obtain D_d (Seidou et al., 2006).

$$H = \begin{cases} A_0\sqrt{D_g - C} \;\; if \; D_g \geq C \\ 0 \; if \; D_g < C \end{cases} \tag{2}$$

Since the occurrence of the first day of the freezing daily mean air temperature usually arrives several days or weeks before the date of ice onset, the parameter C will normally take a positive value (Seidou et al., 2006).

3.2.2 Imputation of Missing Values

Missing or incomplete datasets pose a challenge to hydrologists attempting to model processes; however, as machine learning models rely solely on data to learn the underlying input-output relationships, they suffer more in comparison to physically-based models (Gill et al., 2007). Accordingly, applying imputation techniques instead of ignoring observations with any missing variables was implemented for the available data set. The total number of missing values for the 07CD001 station during ice thickness measurement dates is nine. The imputation techniques applied were the Kendall-Theil robust line (KTRL) and the regularized expectation maximization (RegEM). Although these two methods are both been suggested for imputation of scattered missing values, they substitute the missing records differently: the KTRL employs the correlation between one of the complete variables and the variable with missing values, whereas RegEM considers the

covariance between all variables to impute missing values in the dataset. As a result, both these methods were considered for substitution of missing records in this data set.

The KTRL was initially described by Theil (1950), and is based on the Kendall rank correlation coefficient (τ) (Theil, 1992). The KTRL robust slope estimator or Sen's slope is computed by comparing each pair of records to all others in a pair-wise manner. Considering a data set of (x, y) of size N will result in $N(N-1)/2$ pair-wise comparisons. For each set of comparison, a slope $\Delta y/\Delta x$ is computed. The slope estimate (b_k) is the median of all slopes computed (Theil, 1992):

$$b_k = median\frac{y_j - y_t}{x_j - x_t} \qquad \forall t < j \qquad t = 1,2, \dots N-1 \quad and \quad j = 2,3, \dots, N \tag{3}$$

The KTRL intercept (a_k) which is a function of its slope and the median values is defined as follow:

$$a_k = median(y_c) - [b_k * median(x_c)] \tag{4}$$

The KTRL technique can be considered as an analogue to ordinary least squares (OLS), with the main difference being that the KTRL regression line passes through the point representing the median values of the response and predictor (median x, median y), whereas for OLS it represents the mean values (mean x, mean y) (Helsel and Hirsch, 2002). Based on the KTRL method's robustness in the presents of outliers, this method for substitution of scattered missing values comes highly recommended (Khalil and Adamowski, 2014; Khalil et al., 2012) and was therefore implemented for the present study's data set.

The regularized expectation maximization (RegEm) is based on an iterative analysis of linear regression of missing and available values along with estimated regression coefficients obtained by ridge regression — a regularized regression method in which the filtering of the noise in the data set is controlled by a continuous regularization parameter (Schneider, 2001). The conventional EM algorithm and the RegEM are only applicable to datasets with scattered missing values, and share the advantage of being easy to program, having a low cost per iteration, and providing an

economy of storage (Li et al., 2005). The RegEm for a data set of size N is formulated as (Schneider, 2001):

$$x_m = \mu_m + (x_a - \mu_a)B + e \quad (5)$$

where,

e	is the assumed residual with unknown covariance matrix
B	is a matrix of regression coefficients
x_a	the vector of available values
x_m	is the vector of missing values,
μ_a	is the vector of means of the available values, and
μ_m	is the vector of means of the missing values.

The conditional maximum likelihood estimate of the regression coefficients can be written as:

$$\hat{B} = \hat{\sum}_{aa}^{-1}\hat{\sum}_{am} \quad (6)$$

where,

$\hat{\sum}_{aa}^{-1}$	is the estimated covarinace matrix of the available values, and
$\hat{\sum}_{am}$	is the estimated cross covariance of the available and missing values.

The regularization of the regression model in ridge regression is achieved by adjusting the inverse matrix $\hat{\sum}_{aa}^{-1}$ in Eq. (6) to be:

$$(\hat{\sum}_{aa} + h^2\hat{D})^{-1} \quad (7)$$

where,

$\hat{D}$	is the diagonal matrix consisting of the diagonal elements of the covariance matrix $\hat{\sum}_{aa}$, and
h	is a positive number termed the ridge parameter.

A detailed derivation and discussion of these equations can be found in Schneider (2001).

3.2.3 Extreme Learning Machine

The Extreme Learning Machine (ELM) was first proposed by Huang et al. (2006b) in order to overcome the limitation and shortcomings of its counterparts (*e.g.*, ANN and SVM)

(Huang et al., 2006a). The novel single layer feed forward network (SLFN) in ELM operates differently than the traditional feed forward backpropagation (FFBP) ANN. This confers on ELM a shorter modeling time, which is the main advantage of ELM algorithm. This arises because the input weights (and biases) are randomized, and the output weights have a unique least-squares solution solved by a Moore-Penrose generalized inverse function (Huang et al., 2006a; Huang et al., 2006b; Huang and Xiang, 2015). This process prevents iterative training techniques that tend to collapse to local, rather than global minima (Deo et al., 2017). Accordingly, three simple steps are employed in ELM algorithms:

(i) random generation of hidden layer weights and biases,

(ii) generation of the hidden layer output matrix by the inputs variables that passed through the hidden layer parameters, and

(iii) estimation of ELM output weights by inverting the hidden layer output matrix, as well as computation of its product with the response variable.

The number of hidden neuron nodes has typically been identified by trial and error processes using the validation data set. Randomisation of the hidden layer by a continuous probability distribution such as the uniform distribution, normal distribution, or triangular distribution must be generated to complete the process (Deo et al., 2016b).

Figure 3 illustrates the basic structure of an ELM model. Considering a data set, with x_t as the predictor and y_t as the predict, comprised of N training data records with d-dimentional vectors, where $t = 1, 2, \dots, N$ and $x_t \in R^d$ and $y_t \in R$, the SLFN with L hidden neurons is mathematically expressed as (Huang et al., 2006b):

$$\sum_{i=1}^{H} \beta_i G_i(a_i * x_t + b_i) = z_t \quad (8)$$

where,

a_i $\in R^d$

b_i $\in R$

β $\in R^H$, are the output weights between the hidden layers with H nodes and the model output,

$G_i(a, b, x)$ is the hidden layer activation function, and

z ($z_t \in R$) is the model output .

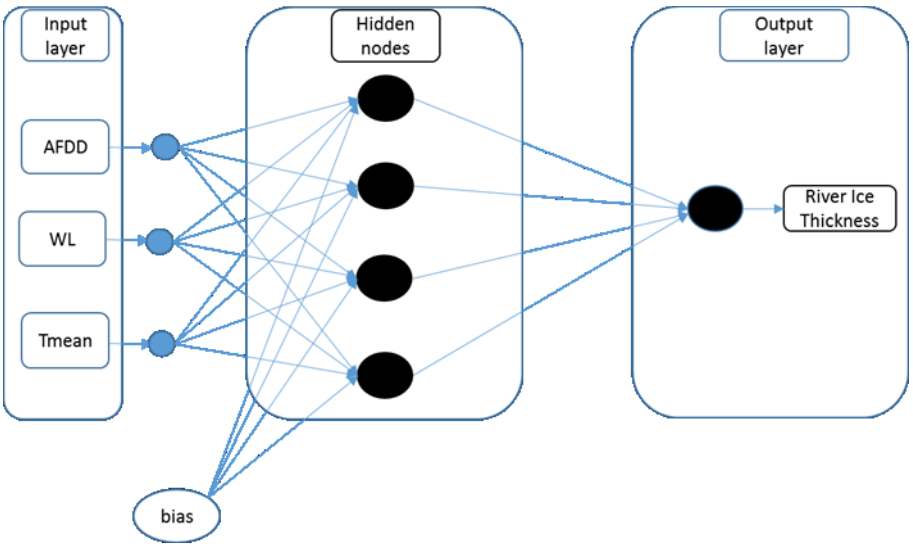


Figure 3. Basic structure of extreme learning machine model employed in this study.

There are different activation functions adopted in ELM models, *e.g.*, tangent sigmoid, logarithmic sigmoid, hard limit, triangular basis, and radial basis, etc.. Choosing a suitable activation function allows the ELM to achieve a more optimal generalization than traditional feedforward neural networks in which all parameters are learned (Liu et al., 2015). This results in ELM models outperforming traditional learning algorithms such as ANN and SVM (Deo et al., 2017). Among the activation functions, the logarithmic sigmoid is the one most commonly used in the field of hydrological forecasting (Adamowski et al., 2012; Kişi, 2008; Quilty et al., 2016), and is denoted as follow:

$$G(a_i,\ b_i,\ x) = \frac{1}{1+e^{(-ax+b)}} \tag{9}$$

According to Huang et al. (2006), Eq. 8 can approximate N training set samples with zero error as follows (Huang et al., 2006b):

$$\sum_{t=1}^{N} \|O_t - y_t\| = 0 \tag{10}$$

This equation illustrates the point that network parameters (a, b, β) can be obtained analytically for a given set of training records. In this context, β can be estimated directly

from the N input-output records as a linear system of equations denoted as (Huang et al., 2006b):

$$G\beta = Y \qquad \beta = \begin{bmatrix} \beta_1 \\ . \\ . \\ . \\ \beta_H \end{bmatrix}_{H\times 1} \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ . \\ . \\ . \\ y_N \end{bmatrix}_{N\times 1} \tag{11}$$

where,

G is the hidden layer output matrix.

In order to compute the output weights of the ELM network the hidden layer matrix is inverted using the Moore-Penrose generalized inverse function (+) (Huang et al., 2006b):

$$\beta^* = G^+ Y \tag{12}$$

where,

G^+ is the inverted hidden layer output matrix,

+ represents the Moore-Penrose generalized inverse function, and

β^* is the estimated output weights from N data records.

In the end, the forecasted values can be generated by a new input vector (testing set) (Akusok et al., 2015):

$$\hat{y} = \sum_{i=1}^{H} \hat{\beta}_i G_i(a_i x_{new} + b_i) \tag{13}$$

3.2.4 Least Squares Support Vector Machine

Originating from support vector machines (SVM), and first proposed by Suykens and Vandewalle (1999), the least squares support vector machine's (LSSVM) compelling aspect is its ability to solve problems of non-linear classification and function estimation (Kumar and Kar, 2009). The LSSVM formulation computes a linear system in dual space under a least squares cost function, whereas SVM uses a quadratic optimization problem approach. The LSSVM has been employed in a variety of subjects (*e.g.*, pattern recognition, signal processing, and non-linear regression estimation) due to its more

limited computation compared to conventional models such as back propagation neural networks (BPNN), and partial least square regression (PLS) (Kisi and Parmar, 2016).

Figure 4 illustrates the basic structure of an LSSVM model. A data set comprising of, x_t as the predictor, y_t as the predictand, and w as d-dimentional weight vectors, where $x_t \in R^d$ and $y_t \in R$, is considered. The LSSVM non-linear function is written as:

$$f(X) = w^T \emptyset(X) + b \tag{14}$$

where,

b is the bias term, and

$\emptyset$ is the mapping function that maps X into d-dimensional feature vector.

The regression problem can be expressed according to the structural minimization principle , considering the complexity of function and fitting error as:

$$minJ(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{t=1}^{d} e_t^2 \tag{15}$$

that has the following constraints:

$$y_t = w^T \emptyset(x_t) + b + e_t \qquad (t = 1, 2, \dots, d) \tag{16}$$

where,

e_t is the slack variable for x_t. and

γ is the margin parameter.

Introducing the Lagrange multipliers α_t and changing the constraint problem into an unconstrained one, the objective function can be gained in order to solve the optimization problems in Eq. 15 as:

$$L(w, b, e, \alpha \propto) = J(w, e) - \sum_{t=1}^{d} \propto_t \{w^T \emptyset(x_t) + b + e_t - y_t\} \tag{17}$$

By taking the partial derivatives of Eq. 17, the optimal condition can be obtained according to the Karush-Kuhn-Tucker (KKT), as follows:

$$\begin{cases} w = \sum_{t=1}^{d} \propto_t \emptyset(x_t) \\ \sum_{t=1}^{d} \propto_t \\ \propto_t = \gamma e_t \\ w^T \emptyset(x_t) + b + e_t - y_t = 0 \end{cases} \quad (18)$$

Thus, the linear equations are generated as:

$$\begin{bmatrix} 0 & -y^T \\ y & ZZ^T + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ \propto \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (19)$$

where,

t = $1, \dots, d$

y = $y_1, \dots, y_d$

Z = $\emptyset(x_1)^T y_t, \dots, \emptyset(x_d)^T y_d$, and

$\propto$ = $[\propto_1, \dots, \propto_i]$.

By defining the kernel function $k(x, x_t) = \emptyset(x)^T \emptyset(x_i),\ i = 1, \dots, d$ the LSSVM regression for new points becomes:

$$\hat{f}(x) = \sum_{t=1}^{d} \propto_t k(x_{new}, x_i) + b \quad (20)$$

Many kernel function such as linear, polynomial, radial basis, and sigmoidal have been proposed for LSSVM. The radial basis function (RBF) kernel is the most commonly used function in regression problem (Kisi, 2012), and is denoted as:

$$k(x, x_t) = \exp(-\|x - x_t\|^2 / 2\sigma^2) \quad (21)$$

where,

σ^2 is the bandwidth of the radial basis kernel function.

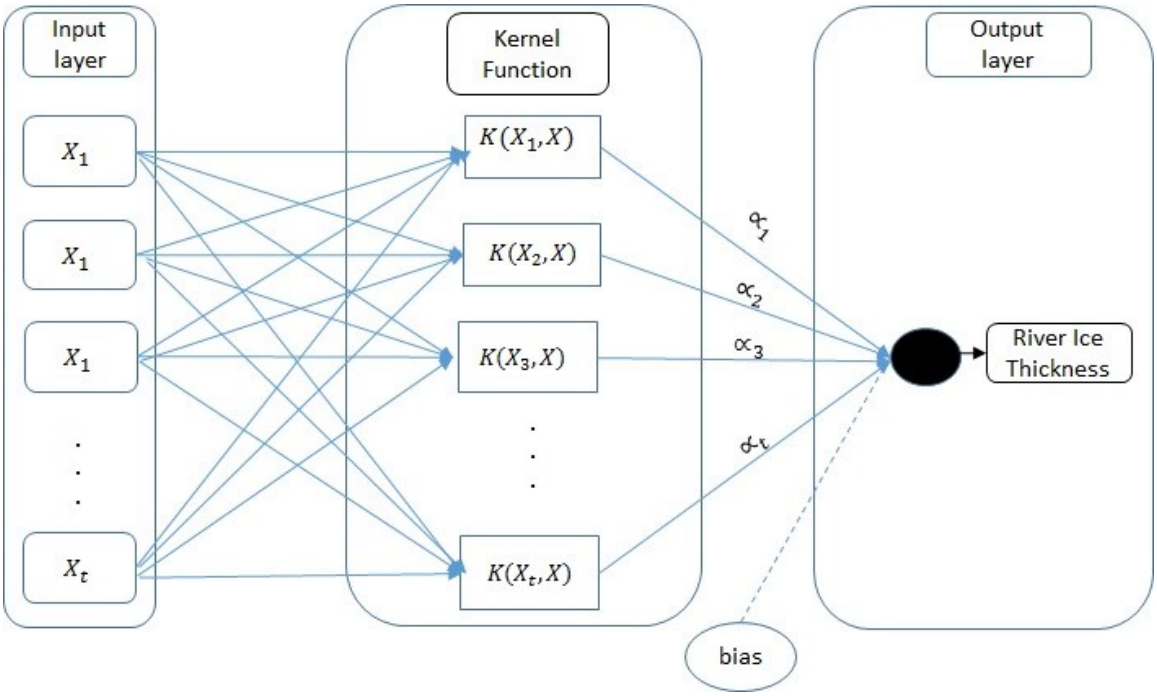


Figure 4. the basic struct of least square support vector machine in this study.

3.2.4 Bootstrap Technique

For each set of bootstrap samples,T^s, ELM and LSSVM models were developed and trained. The model output were then evaluated using the set of A^s observation pairs, $t_t = (x_t, y_t)$, that were not a part of T^s. The generalization error for the models then represents the performances of the models with the validation sets which were subsequently averaged. The generalization error, E_0, can then be estimated as (Tiwari and Chatterjee, 2010a):

$$\hat{E}_0 = \frac{\sum_{s=1}^{S} \sum_{i \in A_s} (yi - f_{model}\left(x_t, \frac{w_s}{T^s}\right))^2}{\sum_{s}^{S} (\#A_s)} \quad (22)$$

where,

$f_{model}\left(x_t, \frac{w_s}{T^s}\right)$ is the output of the model (ANN and LSSVM) developed from bootstrap sample T^s,

in which,

w_s is the weight vector, and

x_t is a particular input vector.

The bootstrap model estimate $\hat{y}(x)$ of all developed models is given by the mean of the S bootstrap estimates (Tiwari and Chatterjee, 2010a):

$$\hat{y}(x) = \frac{1}{S}\sum_{s=1}^{S} f_{model}(x, w_s) \tag{23}$$

and the variance is computed as:

$$\hat{\sigma}^2(x) = \frac{\sum_{s=1}^{S}\sum_{i=A_S}(yi - f_{model}\left(x_t, \frac{w_s}{T^s}\right))^2}{S-1} \tag{24}$$

Chapter 4: model development

4.1 Model Development

The river ice thickness data spanned a period of 35 years (1978-2012). For all the applied models, the data set was split into training (82%), validation (10%), and test (8%) subsets. It is important to note that there is no thumb rule for data division for a predictive dataset and it varies with the problem of interest (Deo et al., 2016a). Table 1 shows the univariate statistics for each data set. As explained in study area section, the proper input variables were chosen based on correlation analysis between river ice thickness and climate variables. Cumulated freezing degree days, water level (m), and mean temperature ($°C$) were the input variables chosen. In order to prevent parameters with large numerical ranges dominating over of those with smaller numerical ranges in the models, the input variables were scaled prior to the modelling processes, as follow:

$$x_{norm} = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (25)$$

where,

x_i is the current input variable that is to be normalized,
x_{min} is the minimum value within the historical dataset,
x_{max} is the maximum value within the historical dataset, and
x_{norm} is the normalized value (between 0 and 1) of the input variable,

Table 1. Descriptive statistics for river ice thickness.

Partition	No.records	River ice thickness (m)				
		mean	St.Dev.	median	minimum	maximum
Training	90	0.5836	0.2001	0.5831	0.2511	1.8440
Validation	11	0.5580	0.1495	0.6266	0.3121	0.7396
Test	9	0.5460	0.1596	0.5362	0.2884	0.7308

In this study, a Back-propagation Levenberg-Marquardt (LM) algorithm was used in training the single layer ANN. In function approximation problems, when training moderate size networks with fewer than a few hundred weights, the LM algorithm proves to be the fastest and most accurate, generating lower mean square errors than any other

training algorithms tested (Demuth and Beale, 2009). The number of hidden neurons in the hidden layer was obtained through careful trial and error. The optimum number of hidden neurons, based on the lowest RMSE in the validation phase, was found to be 9. The maximum number of training epochs was set to 100.

ANN models are prone to overfitting, that is the model can fit the training data set precisely, but cannot necessarily perform accurately with the test set. The ELM model was developed to circumvent this flaw. To create an appropriate ELM structure for river ice thickness estimation, the optimum number of hidden neurons was investigated by trial and error. One to fifty hidden neurons were tested, and the optimal model structure was chosen based on the model performance with the validation set. The minimum value of RMSE was generated for 39 hidden neurons. For bootstrap ELM, besides the number of hidden neurons, the bootstrap ensemble size must be determined. Ensemble sizes of 10, 25, 50, 75, 100, 150, 200, 250, and 500 were tested and the best performance achieved with an ensemble size of 10. The sigmoid activation function was applied for ELM and BELM model development.

In LSSVM, the Gaussian radial basis function was used as the selected kernel function. In order to define the nonlinear function in LSSVM two parameters, γ and σ^2, were employed, where γ is a regulization constant and σ^2 is the bandwidth of the radial basis kernel function (RBF). Based on the model performance with the validation set, $\gamma = 1.5$ and $\sigma^2 = 0.5$ were the optimal values (trial and error). For the bootstrap LSSVM, a trial and error procedure was undertaken to determine the best ensemble size, γ value and σ^2 value — values of 100, 5.5, and 0.5, respectively were determined. All the predictive models' algorithms were developed in the MATLAB programming environment.

4.2 Performance Assessment

In order to assess the proposed models, several quantitative performance indicators were employed. A combination of different metrics is often required for an extensive evaluation of a model's performance (Chai and Draxler, 2014). The Root-mean-square error (RMSE), Nash-Sutcliffe Efficiency (E_{NS}), correlation coefficient (r), mean absolute error (MAE), and bias (BIAS) were considered in assessing the accuracy of river ice thickness (RIT). These defined as follow:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{t=N}(RIT_t^{obs} - RIT_t^{pred})^2} \tag{26}$$

$$r = \frac{\sum_{t=1}^{t=N}\left[(RIT_t^{obs} - \overline{RIT^{obs}})(RIT_t^{pred} - \overline{RIT^{pred}})\right]}{\sqrt{\sum_{t=1}^{t=N}(RIT_t^{obs} - \overline{RIT^{obs}})^2} \cdot \sqrt{\sum_{t=1}^{t=N}(RIT_t^{pred} - \overline{RIT^{pred}})^2}}, -1.0 \le r \le 1.0 \tag{27}$$

$$E_{NS} = 1 - \frac{\sum_{t=1}^{t=N}\left(RIT_t^{obs} - RIT_t^{pred}\right)^2}{\sum_{t=1}^{t=N}\left(RIT_t^{obs} - \overline{RIT^{obs}}\right)^2}, -\infty \le E_{NS} \le 1.0, \tag{28}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{t=N}\left|RIT_t^{obs} - RIT_t^{pred}\right| \tag{29}$$

$$BIAS = \frac{1}{N}\sum_{t=1}^{t=N}\left(RIT_t^{obs} - RIT_t^{pred}\right) \tag{30}$$

where,

N	is the sample size,
RIT_t^{obs}	is the t^{th} observed *RIT* value,
$\overline{RIT^{obs}}$	is the mean observed *RIT* value,
RIT_t^{pred}	is the t^{th} predicted *RIT* value, and
$\overline{RIT^{pred}}$	is the mean predicted *RIT* value.

Chapter 5: Results and Discussion

The goals of this study were to investigate the feasibility of using ELM and LSSVM, and their bootstrap enhanced equivalents in the river ice thickness estimation in comparison to an ANN model benchmark. Also, the performance of applied imputation techniques were compared by various performance metrics which were computed to assess the models' accuracy. A direct comparison of the models is presented in Table 2 and 3 for each partitioning method.

Table 2. Performance indicators for the ANN, ELM, BELM, LSSVM, and BLSSVM models evaluated for each modelling phase — KTRL imputation technique.

	Modelling phase														
	TRAINING					VALIDATION					TEST				
Model	r	*RMSE* (m)	E_{NS}	*MAE* (m)	*BIAS* (m)	r	*RMSE* (m)	E_{NS}	*MAE* (m)	*BIAS* (m)	r	*RMSE* (m)	E_{NS}	*MAE* (m)	*BIAS* (m)
ANN	0.76	0.096	0.56	0.071	1.002	0.67	0.103	0.35	0.082	0.953	0.60	0.147	0.32	0.126	1.02
ELM	0.81	0.116	0.65	0.093	1.000	0.81	0.089	0.60	0.068	0.944	0.80	0.091	0.63	0.079	0.993
BELM	0.46	0.177	0.21	0.107	0.993	0.70	0.101	0.50	0.084	1.012	**0.90**	**0.080**	**0.71**	**0.072**	**0.991**
LSSVM	0.85	0.116	0.66	0.077	1.000	0.69	0.086	0.63	0.070	0.983	0.72	0.103	0.52	0.083	0.990
BLSSVM	0.85	0.117	0.67	0.071	1.003	0.80	0.088	0.61	0.071	0.970	0.71	0.105	0.51	0.081	0.997

Table 3. Performance indicators for the ANN, ELM, BELM, LSSVM, and BLSSVM models evaluated for each modelling phase — RegEm imputation technique.

	Modelling phase														
	TRAINING					VALIDATION					TEST				
Model	r	*RMSE* (m)	E_{NS}	*MAE* (m)	*BIAS* (m)	r	*RMSE* (m)	E_{NS}	*MAE* (m)	*BIAS* (m)	r	*RMSE* (m)	E_{NS}	*MAE* (m)	*BIAS* (m)
ANN	0.73	0.103	0.52	0.083	0.980	0.75	0.110	0.49	0.077	1.056	0.46	0.091	0.30	0.086	1.002
ELM	0.88	0.068	0.78	0.047	1.000	0.76	0.090	0.57	0.068	0.9742	0.73	0.1143	0.43	0.082	0.988
BELM	0.66	0.1140	0.40	0.090	0.970	0.54	0.117	0.27	0.098	0.978	**0.93**	**0.067**	**0.80**	**0.060**	**1.003**

LSSVM	0.86	0.0797	0.71	0.065	1.000	0.68	0.103	0.43	0.085	0.957	0.80	0.090	0.64	0.074	1.025
BLSSVM	0.87	0.075	0.73	0.060	1.004	0.68	0.104	0.42	0.087	0.946	0.78	0.095	0.64	0.074	0.074

In general, the correlation coefficient, *r*, represents the strength of the linear regression between observed and predicted river ice thickness and compares them directly. However, for the observed *vs.* predicted linear relationship, an 'ideal' *r* = 1.0 can occur even if the slope and ordinate intercept differ from 1.0 and zero, respectively, so scatter plots and other metrics must be consulted. The E_{NS} measures the model's overall prediction skill, and it is sensitive to differences in the observed and estimated means and variances. An $E_{NS} > 0.5$ is considered good; $E_{NS} = 0$, represents predicted values no better than using $\overline{RIT^{obs}}$ for all predicted values, and $E_{NS} < 0$ to $-\infty$ represents increasingly poorer predictions. The E_{NS} therefore represents a better assessment of model accuracy than simply *r* (Govindaraju, 2000). The MAE gives the same weight to all errors, whilst the RMSE squares errors, thereby giving greater weights to errors with larger absolute values.

Based on Table 2, the ANN model performed acceptably in the training phase (*r*=0.76, RMSE=0.096 m, E_{NS}=0.56, MAE=0.071 m, and BIAS=1.002 m), but performed poorly in the validation and test phases (*r*=0.60, RMSE=0.147 m, E_{NS}=0.32, MAE=0.126 m, and BIAS=1.02 m for test phase). In contrast, the performance of the LSSVM and BLSSVM models only differed slightly (0.01 to 0.02 m difference in RMSE) between training, validation, and test phases (for each partitioning), indicating that these two models did not privilege any particular set of river ice thickness values and both performed satisfactorily. On the other hand, the ELM and BELM models' performances in different modelling phases behaved differently (Table 2), with the ELM model producing acceptable results in training, validation, and test phases (E_{NS}= 0.65, 0.60, 0.63, respectively), whereas the BELM model only generated satisfactory results for the validation and test phases (E_{NS}= 0.46, 0.50, 0.71 for training, validation, and test phases, respectively).

It is therefore evident that the bootstrap ELM model can generate a satisfactory performance even though the model is not sufficiently trained. A similar conclusion can

be drawn based on the RMSE, *i.e*., the lowest RMSE value ($\approx 0.080\ m$) was obtained for the BELM model in the testing phase, although, in the training phase, the ELM and LSSVM models performed better in terms of RMSE. Considering model performance based on the evaluation metrics oresented for the test phase, the BELM outperformed the ELM, LSSVM, and BLSSVM in river ice thickness estimation, indicating that the BELM model had a slightly better ability to estimate river ice thickness for out-of-sample records.

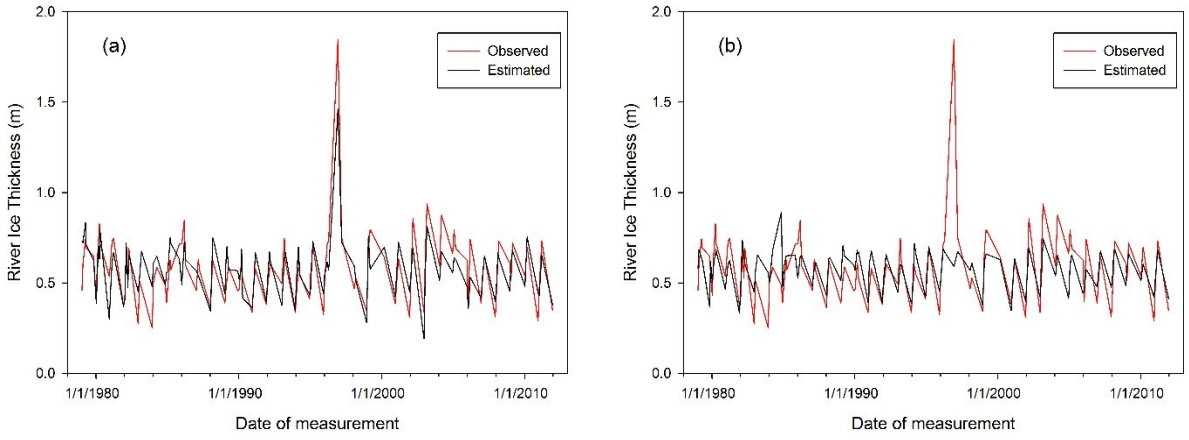
The better performance of the BELM model over the LSSVM and BLSSVM can be attributed to the fact that, for this study:

(i) the parameterization of BELM better fits the given dataset and,

(ii) the trained network in BELM is more robust to a training dataset that contains input patterns that are drastically different than of other input sets.

Comparing the KTRL and RegEm imputation techniques (Table 2 *vs*. Table 3) it is evident that with RegEm method, and similarly with the KTRL method, the BELM model performs best in the test phase (R=0.93, RMSE= 0.067 m, E_{NS}=0.80, MAE=0.060 m, and BIAS=1.003 m. However, the overall performance of KTRL in different modelling phases outperformed the RegEm method, since for the latter at least one of the subsets generated a performance metric outside the acceptable range.

Time series graphs provided a closer examination of model performance with respect to their ability to capture minimum and maximum values of river ice thickness.



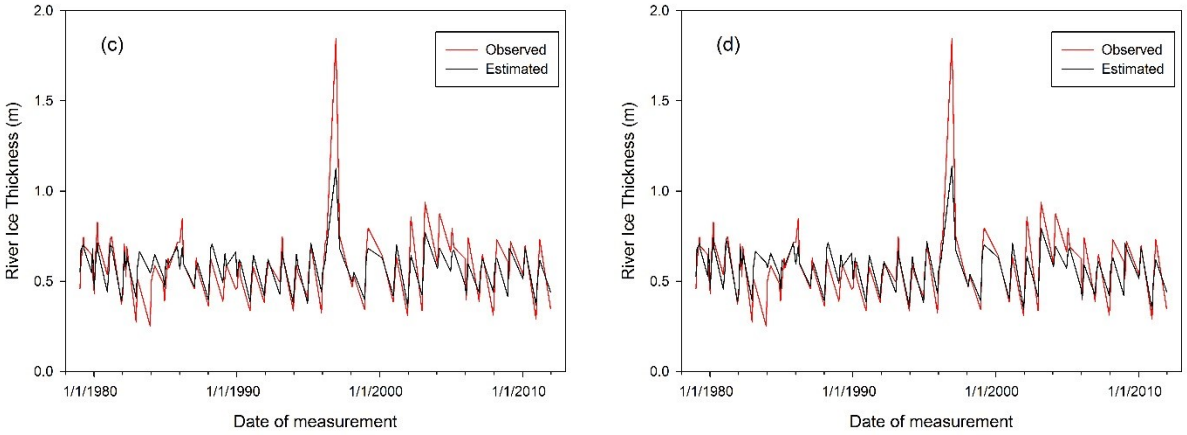


Figure 5. Performance of various machine learning models developed with KTRL imputation technique for river ice thickness estimation. (a) ELM, (b) BELM, (c) LSSVM, (d) BLSSVM.

The ELM model clearly showed its greater capacity to capture minimum and maximum values of river ice thickness. Confirming model performance metrics, the LSSVM and BLSSVM performed equally well in catching peak values. In contrast, the BELM model underestimated the maximum value of the dataset, while mid-range values (0.3-0.8 m) were more precisely estimated with this technique.

With the RegEm imputation technique, the ELM model could estimate the peak value with better accuracy than the ELM model with the KTRL imputation technique. Also, although the performance of BALM in catching the peak value was still very poor with both imputation techniques, the BELM with RegEm technique generated better results. Likewise, with the KTRL imputation technique, LSSVM, and BLSSVM generated similar results in capturing the maximum and minimum values.

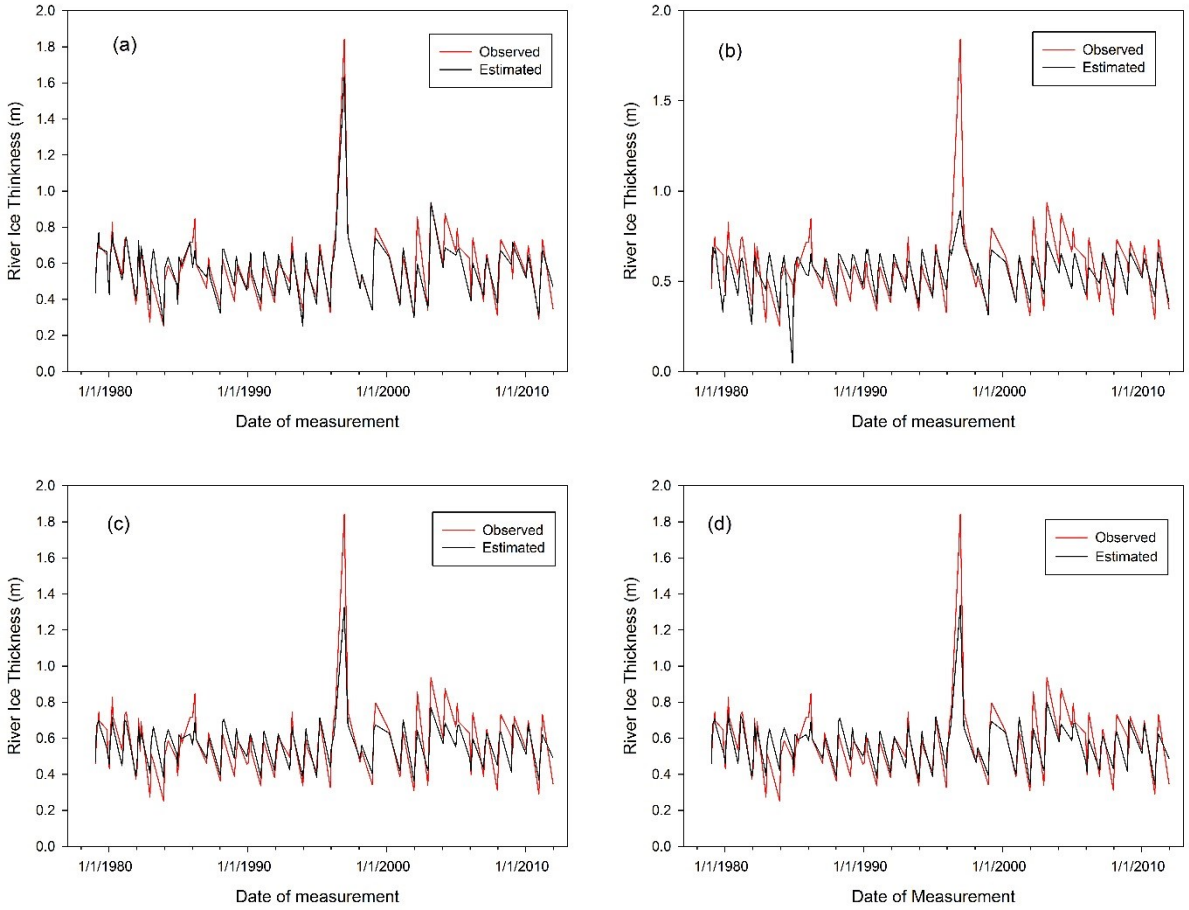


Figure 6. Performance of various machine learning models developed with RegEm imputation technique for river ice thickness estimation. (a) ELM, (b) BELM, (c) LSSVM, (d) BLSSVM.

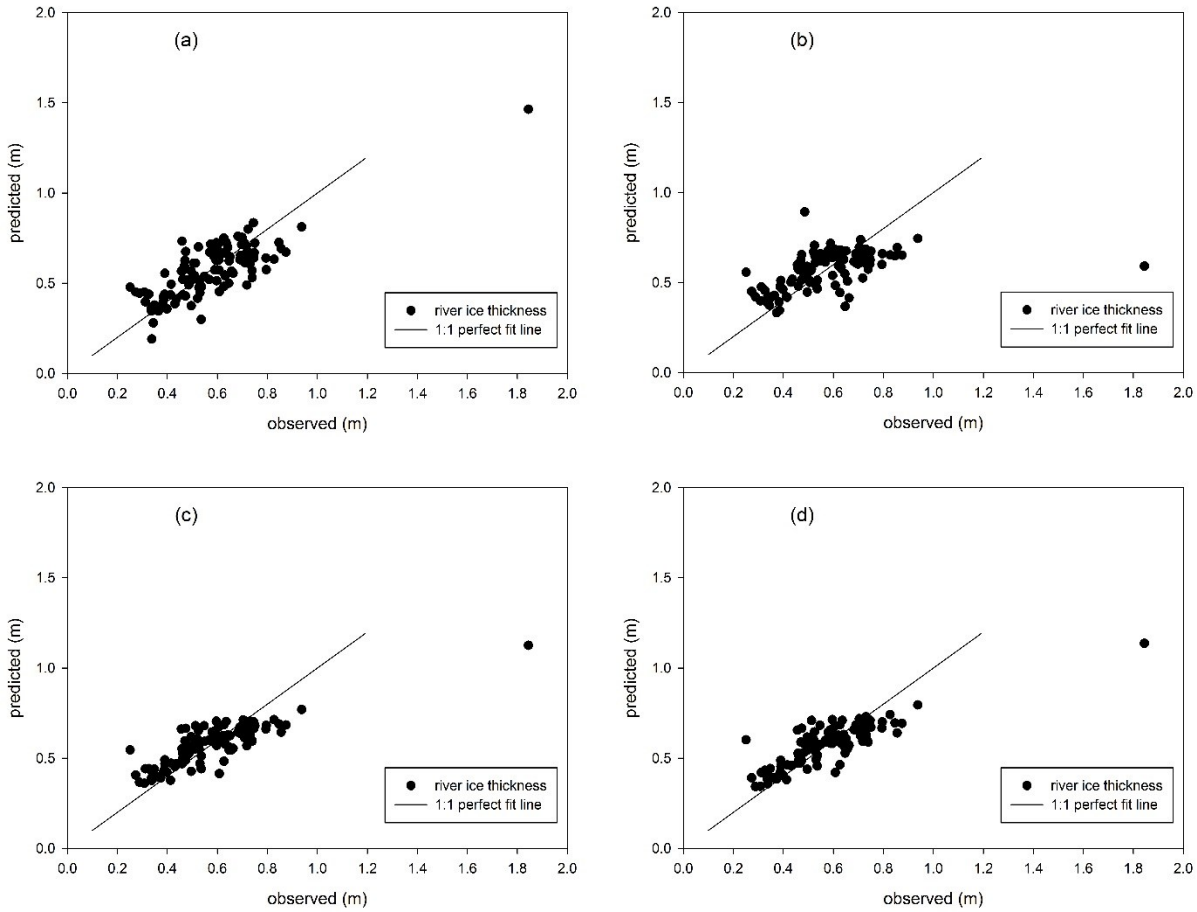


Figure 7. Performance of various machine learning models developed with KTRL imputation technique for river ice thickness estimation. (a) ELM, (b) BELM, (c) LSSVM, (d) BLSSVM.

For further examination of model performances, the scatter plots of ELM, BELM, LSSVM, and BLSSVM are shown in Figures 7 and 8. Although none of the applied models could estimate the peak value of river ice thickness, there appears to be a good agreement between the estimated river ice thickness and observed values. Despite some degree of scattering, LSSVM and BLSSVM generated the closest estimations to a 1:1 perfect fit line (Figure 7). This concurred with the results of the correlation coefficient for the overall performance of the LSSVM and BLSSVM models (R=0.84 and 0.85, respectively).

For the RegEm imputation technique, compared to LSSVM and BLSSVM, the ELM model provided similar overall estimations in regards to 1:1 perfect fit line (R=0.86, 0.84, and 0.85, respectively).

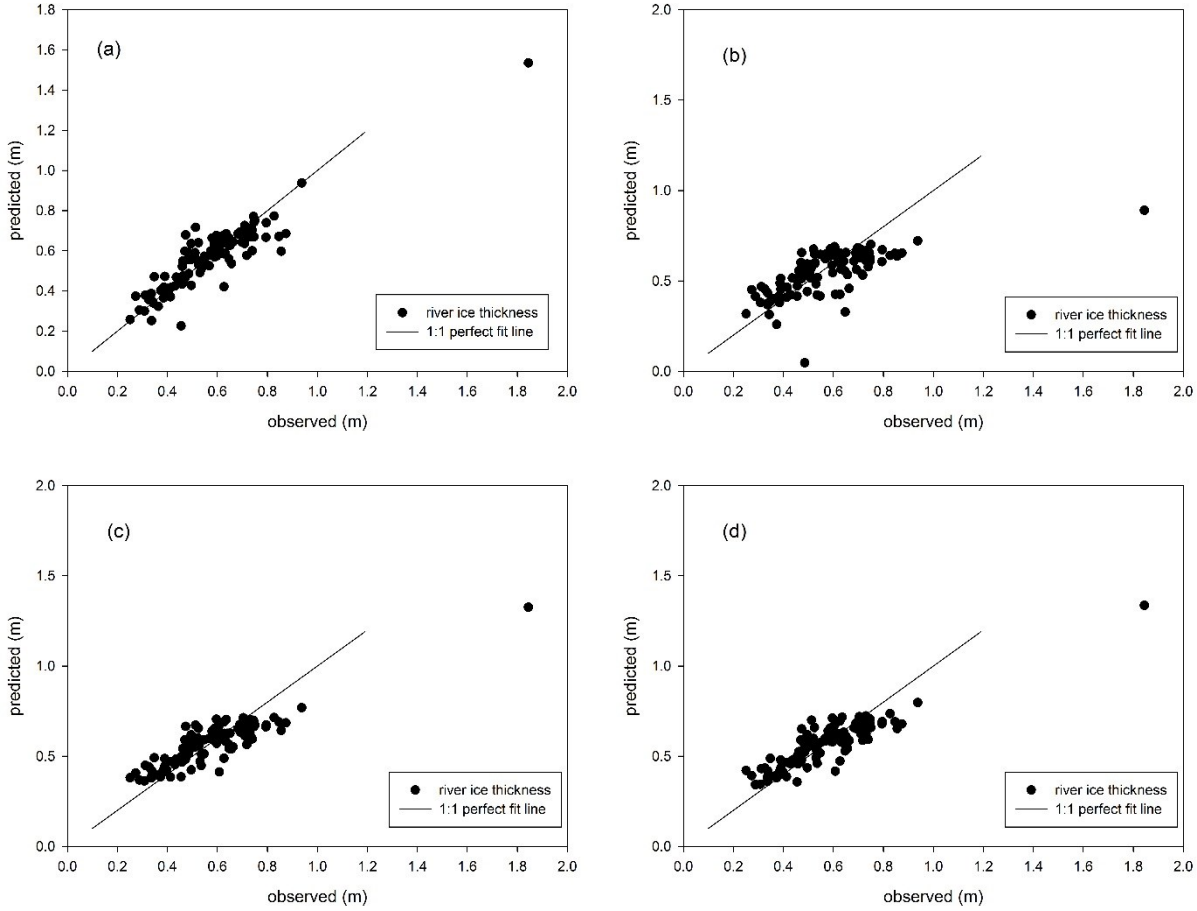


Figure 8. Performance of various machine learning models developed with RegEm imputation technique for river ice thickness estimation. (a) ELM, (b) BELM, (c) LSSVM, (d) BLSSVM.

Chapter 6: summary and conclusion

Accurate Freshwater ice estimation is increasingly important in the light of growing demand for water resource management and engineering in cold regions. Given the high importance of the Fort McMurray community on the effect of various river ice processes, the importance of accurate estimation is highlighted. This study attempts to compare different machine learning techniques and their bootstrap methodologies to estimate river ice thickness in Alberta, Canada. Extreme learning machine and least squares support vector machine were used for the estimation of river ice thickness. To date, these methods have not been explored for river ice thickness estimation until the present study.

The Athabasca River is the second largest river in Alberta, and its largest unregulated river which travels northward through the boreal mixed wood forest and encounters sudden changes in its physical properties near the Fort McMurray. These changes are responsible for the frequent formation of an ice jam in this location. The Clearwater River joins the main stream immediately downstream of Fort McMurray and then drains into Lake Athabasca. Direct measurements of ice thickness were collected from Water Survey of Canada (MSC) winter gaugings at the Clearwater River at Draper (07CD001) hydrometric station. The closest meteorological station to the hydrometric site, Fort McMurray A, was used for meteorological observation records in order to calculate the explanatory variables assumed representative to climate conditions at the hydrometric stations. The meteorological variables considered were maximum, minimum, and mean daily temperature, total precipitation, and snow on the ground were. These were recorded on a daily basis with some missing values. As river ice thickness is a function of temporal changes in meteorologic parameters, accumulated freezing degree days (AFDD) and accumulated solar radiation (ASR) were also calculated. However, the absence of 'snow on the ground' data for the period of 2009-2012 led to its omission and that of cumulative snow (CUMS) as input variables. Measurement of ice thickness was limited to one to seven measurements per season. In total, 110 records of ice thickness for station 07CD001 were collected during the period from 1979 to 2012.

Correlation analysis was applied to all predictors to find the best set of input variables for river ice thickness estimation. Accordingly, three predictors, water level (m), accumulated freezing degree day, and mean temperature (°C), which has the maximum correlation with ice thickness, were selected as inputs.

Since machine learning models solely rely on data to learn the underlying input-output relationships, imputation of missing values in the available data set instead of ignoring the observations with any missing values has been considered. The Kendall-Theil robust line (KTRL) and the regularized expectation maximization (RegEM) methods were applied for data imputation. Although these two methods are both been suggested for imputation of scattered missing values, they substitute the missing records differently: the KTRL employs the correlation between one of the complete variables and the variable with missing values, whereas RegEM considers the covariance between all variables to impute missing values in the dataset.

Extreme learning machine and least square support vector machine were utilized as the newer form of machine learning models because they have shown improved generalization performance when compared to the traditional form of such techniques (ANN and MLR). In addition, bootstrapping were employed in order to improve the result of single model output. In order to prevent the data patterns and attributes with large numerical ranges dominating the role of the smaller numerical ranges the scaling of input variables, prior to the modeling processes, was taken into consideration. For ANN model, Back-propagation Levenberg-Marquardt algorithm trained a single layer network with 9 hidden layer and 100 epochs size. The sigmoid activation function was applied for ELM and BELM model development, and the number of hidden neurons and ensemble size were acquired by careful trial and error. In LSSVM and BLSSVM, the Gaussian radial basis function was used as the selected kernel function. Besides the gamma and sigma values for kernel bandwidth, the number of ensemble BLSSVM were obtained via trial and error and best on the best performance in the validation set.

The primary objective of this study was to demonstrate the use of a machine learning algorithms (Extreme Learning Machine and Least Square Support Vector Machine), and their bootstrap versions, for the estimation of Athabasca River (Alberta, Canada) river ice thickness, using “easy to measure” meteorological variables. The performance of the applied models was compared to that of an artificial neural network. The secondary object was to evaluate the performance accuracy of two imputation methods (KTRL and RegEm) implemented in machine learning models, when applied to the same scenario. Gathered over a period of 35 years, the predictive variables considered were accumulated freezing degree days, water level (m), and mean temperature (°C) for. The dataset were divided into 82% (training), 11% (validation), and 8% (testing) subsets. While the ANN model was not sufficiently accurate in estimating river ice thickness in this study area, the newer form of machine learning techniques (ELM, LSSVM, and their bootstrap versions) were able to perform acceptably. The testing phase indicated that the BELM model could potentially improve the predictive accuracy of the modeling process compared to the ELM, LSSVM, and BLSSVM models. According to the evaluation metrics, RMSE and MAE were decreased by about 12%-24% and 8%-15%, respectively when the BELM model with KTRL imputation technique was evaluated in comparison with the ELM, LSSVM, and BLSSVM models. Moreover, for the BELM model employing the RegEm imputation technique, RMSE and MAE were reduced by about 25%-41% and 19%-27% compared to the other applied models in this study. Finally, the machine learning models developed in this study appear to be promising techniques for the estimation of river ice thickness. However, the potential of machine learning models in river ice thickness estimation requires further investigation in regions where sufficient historical climate data are available, and there is more frequent ice thickness measurements across each year.

Chapter 7: contribution to knowledge and future work

The use of Extreme Learning Machine, Least Square Learning Machine, and their bootstrap equivalent contributes significantly to the existing literature on the rive ice properties estimation. With regards to the rive ice thickness estimation, to date, the presented study is the only study that explored the aforementioned methodolgy to estimate river ice thickness. The results of this study confirms that bootstrap methodolgy help improve the results of ELM and LSSVM for river ice thickness estimation. The results also show that ELM and LSSVM and their bootstrap methodologies are effective estimation tools in the Athabasca river in Alberta, and should be explored in other areas.

While this study contribute new research to the field of river ice properties estimation there are still areas that need to be expanded upon. For instance, this study estimate the river ice thickness with limited number of data records. Future studies could attempt to investigate river ice thickness estimation in regions inwhere sufficient historical data and ice thickness measurement exists.

With respect to modeling techniques, this study apply two machine learning models and their bootstrap equivalents. Future studies could experiment different ensemble techniques to see wheter they are effective in improving the estimation accuracy in the region. In addition, other machine learning techniques could be tested.

Chapter 8: References

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga‐Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour. Res., 48(1).

Akusok, A., Björk, K.-M., Miche, Y., Lendasse, A., 2015. High-performance extreme learning machines: a complete toolbox for big data applications. IEEE Access, 3: 1011-1025.

Andres, D.D., Van der Vinne, P., 2001. Calibration of ice growth models for bare and snow covered conditions: a summary of experimental data from a small Prairie pond, 11th workshop on river ice: River ice processes within a changing environment. CGU HS Committe on River Ice Proccesses and the Environment (CRIPE). Citeseer.

Ashton, G.D., 1989. Thin ice growth. Water Resour. Res., 25(3): 564-566.

Bonsal, B.R., Prowse, T.D., Duguay, C.R., Lacroix, M.P., 2006. Impacts of large-scale teleconnections on freshwater-ice break/freeze-up dates over Canada. Journal of Hydrology, 330(1): 340-353. DOI:http://dx.doi.org/10.1016/j.jhydrol.2006.03.022

Brooks, R.N., Prowse, T.D., O'Connell, I.J., 2013. Quantifying northern hemisphere freshwater ice. Geophysical Research Letters, 40(6): 1128-1131.

Brown, L.C., Duguay, C.R., 2010. The response and role of ice cover in lake-climate interactions. Progress in Physical Geography, 34(5): 671-704.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3): 1247-1250.

Chokmani, K., Khalil, B., Ouarda, T., Bourdages, R., 2007. Estimation of river ice thickness using artificial neural networks, Proceedings of the 14th Workshop on the Hydraulics of Ice Covered Rivers. CGU HS Committee on River Ice Processes and the Environment (CRIPE), Québec, Canada (12 pp.).

Chokmani, K., Ouarda, T.B.M.J., Hamilton, S., Ghedira, M.H., Gingras, H., 2008. Comparison of ice-affected streamflow estimates computed using artificial neural networks and multiple regression techniques. Journal of Hydrology, 349(3–4): 383-396. DOI:http://dx.doi.org/10.1016/j.jhydrol.2007.11.024

Chu, T., Lindenschmidt, K.-E., 2016. Integration of space-borne and air-borne data in monitoring river ice processes in the Slave River, Canada. Remote Sensing of Environment, 181: 65-81. DOI:http://dx.doi.org/10.1016/j.rse.2016.03.041

Demuth, H., Beale, M., 2009. Matlab neural network toolbox user's guide version 6. The MathWorks Inc.

Deo, R.C., Downs, N., Parisi, A.V., Adamowski, J.F., Quilty, J.M., 2017. Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. Environmental Research, 155: 141-166. DOI:https://doi.org/10.1016/j.envres.2017.01.035

Deo, R.C., Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, 153: 512-525.

Deo, R.C., Samui, P., Kim, D., 2016a. Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. Stochastic Environmental Research and Risk Assessment, 30(6): 1769-1784.

Deo, R.C., Tiwari, M.K., Adamowski, J.F., Quilty, J.M., 2016b. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stochastic Environmental Research and Risk Assessment: 1-30. DOI:10.1007/s00477-016-1265-z

Dornan, L., 2005. Development of site specific ice growth models for hydrometric purposes, 13th Workshop on the Hydraulics of Ice Covered Rivers. CGU HS Committe on River Ice Proccesses and the Environment (CRIPE), Hanover, NH, USA, pp. 101-131.

Duguay, C.R., Bernier, M., Gauthier, Y., Kouraev, A., 2015. 12 Remote sensing of lake and river ice.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. The annals of Statistics: 1-26.

Erdal, H.I., Karakurt, O., 2013. Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. Journal of Hydrology, 477: 119-128.

Gill, M.K., Asefa, T., Kaheil, Y., McKee, M., 2007. Effect of missing data on performance of learning algorithms for hydrologic predictions: Implications to an imputation technique. Water Resour. Res., 43(7): n/a-n/a. DOI:10.1029/2006WR005298

Govindaraju, R.S., 2000. Artificial neural networks in hydrology. II: Hydrologic applications. Journal of Hydrologic Engineering, 5(2): 124-137. DOI:10.1061/(ASCE)1084-0699(2000)5:2(124)

Greve, R., Blatter, H., 2009. Dynamics of ice sheets and glaciers. Springer Science & Business Media.

Hejazi, M.I., Cai, X., 2009. Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRMR) algorithm. Advances in Water Resources, 32(4): 582-593. DOI:https://doi.org/10.1016/j.advwatres.2009.01.009

Helsel, D.R., Hirsch, R.M., 2002. Statistical methods in water resources, 323. US Geological survey Reston, VA.

Hicks, F., Cui, W., Andres, D., 1997. Modelling thermal breakup on the Mackenzie River at the outlet of Great Slave Lake, NWT. Canadian Journal of Civil Engineering, 24(4): 570-585.

Hong, W.-C., Pai, P.-F., 2006. Predicting engine reliability by support vector machines. The International Journal of Advanced Manufacturing Technology, 28(1-2): 154-161.

Huang, G.-B., Chen, L., Siew, C.K., 2006a. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans. Neural Networks, 17(4): 879-892.

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006b. Extreme learning machine: Theory and applications. Neurocomputing, 70(1–3): 489-501. DOI:https://doi.org/10.1016/j.neucom.2005.12.126

Huang, J.J., Xiang, W., 2015. Investigation of point source and non-point source pollution for Panjiakou reservoir in North China by modelling approach. Water Quality Research Journal of Canada, 50(2): 167-181. DOI:10.2166/wqrjc.2014.019

Huusko, A. et al., 2007. Life in the ice lane: the winter ecology of stream salmonids. River Research and Applications, 23(5): 469-491. DOI:10.1002/rra.999

Jeffries, M.O., Morris, K., Duguay, C.R., 2012. Floating ice: Lake ice and river ice. Satellite Image Atlas of Glaciers of the World-State of the Earth's Cryosphere at the Beginning of the 21st Century: Glaciers, Global Snow Cover, Floating Ice, and Permafrost and Periglacial Environments.

Karl-Erich, L., Gerry, S., Robert, H., 2010. Measuring ice thicknesses along the Red River in Canada using RADARSAT-2 satellite imagery. Journal of Water Resource and Protection, 2010.

Khalil, B., Adamowski, J., 2014. Comparison of OLS, ANN, KTRL, KTRL2, RLOC, and MOVE as Record-Extension Techniques for Water Quality Variables. Water, Air, & Soil Pollution, 225(6): 1966. DOI:10.1007/s11270-014-1966-1

Khalil, B., Ouarda, T.B., St-Hilaire, A., 2012. Comparison of record-extension techniques for water quality variables. Water Resour. Manage., 26(14): 4259-4280.

Kisi, O., 2012. Modeling discharge-suspended sediment relationship using least square support vector machine. Journal of Hydrology, 456: 110-120.

Kişi, Ö., 2008. River flow forecasting and estimation using different artificial neural network techniques. Hydrology Research, 39(1): 27-40.

Kisi, O., Parmar, K.S., 2016. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. Journal of Hydrology, 534: 104-112.

Kumar, M., Kar, I., 2009. Non-linear HVAC computations using least square support vector machines. Energy Conversion and Management, 50(6): 1411-1418.
Li, H., Zhang, K., Jiang, T., 2005. The regularized EM algorithm, AAAI, pp. 807-812.
Liu, X., Lin, S., Fang, J., Xu, Z., 2015. Is extreme learning machine feasible? A theoretical assessment (Part I). IEEE Transactions on Neural Networks and Learning Systems, 26(1): 7-20.
Lock, G.S.H., 1990. The growth and decay of ice. Cambridge University Press.
Ma, X., Fukushima, Y., 2002. A numerical model of the river freezing process and its application to the Lena River. Hydrol. Processes, 16(11): 2131-2140. DOI:10.1002/hyp.1146
Mahabir, C., Hicks, F., Fayek, A.R., 2005. Neuro-fuzzy logic model for breakup forecasting at Fort McMurray, AB, Proceedings of 13th Workshop on the Hydraulics of Ice Covered Rivers, Hanover, NH. Citeseer, pp. 55-67.
Martynov, A., Sushama, L., Laprise, R., Winger, K., Dugas, B., 2012. Interactive lakes in the Canadian Regional Climate Model, version 5: the role of lakes in the regional climate of North America. Tellus A, 64.
Massie, D.D., White, K.D., Daly, S.F., 2002. Application of neural networks to predict ice jam occurrence. Cold Regions Science and Technology, 35(2): 115-122. DOI:http://dx.doi.org/10.1016/S0165-232X(02)00056-3
Mermoz, S. et al., 2014. Retrieval of river ice thickness from C-band PolSAR data. IEEE Transactions on Geoscience and Remote Sensing, 52(6): 3052-3062.
Michel, B., 1971. Cold Regions Science and Engineering Monograph 3, Section B1a: Winter Regime of Rivers and Lakes, DTIC Document.
Nghiem, S.V., Leshkevich, G.A., 2007. Satellite SAR remote sensing of Great Lakes ice cover, Part 1. Ice backscatter signatures at C band. Journal of Great Lakes Research, 33(4): 722-735.
Peters, D.L., Atkinson, D., Monk, W.A., Tenenbaum, D.E., Baird, D.J., 2013. A multi‐scale hydroclimatic analysis of runoff generation in the Athabasca River, western Canada. Hydrol. Processes, 27(13): 1915-1934.
Prowse, T. et al., 2011a. Past and future changes in Arctic lake and river ice. AMBIO: A Journal of the Human Environment, 40(sup 1): 53-62.
Prowse, T. et al., 2011b. Arctic freshwater ice and its climatic role. AMBIO: A Journal of the Human Environment, 40(sup 1): 46-52.
Prowse, T., Bonsal, B., Duguay, C., Lacroix, M., 2007. River-ice break-up/freeze-up: a review of climatic drivers, historical trends and future predictions. Annals of Glaciology, 46(1): 443-451.
Prowse, T.D., Culp, J.M., 2003. Ice breakup: a neglected factor in river ecology. Canadian Journal of Civil Engineering, 30(1): 128-144. DOI:10.1139/l02-040
Quilty, J., Adamowski, J., Khalil, B., Rathinasamy, M., 2016. Bootstrap rank‐ordered conditional mutual information (broCMI)—A nonlinear input variable selection method for water resources modeling. Water Resour. Res.
Sapankevych, N.I., Sankar, R., 2009. Time series prediction using support vector machines: a survey. IEEE Computational Intelligence Magazine, 4(2).
Schneider, T., 2001. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. Journal of Climate, 14(5): 853-871.
Seidou, O. et al., 2006. Modeling ice growth on Canadian lakes using artificial neural networks. Water Resour. Res., 42(11).
She, Y., Hicks, F., 2006. Modeling ice jam release waves with consideration for ice effects. Cold Regions Science and Technology, 45(3): 137-147. DOI:https://doi.org/10.1016/j.coldregions.2006.05.004
Shen, H.T., 2003. Research on river ice processes: progress and missing links. Journal of Cold Regions Engineering, 17(4): 135-142.

Shen, H.T., 2010. Mathematical modeling of river ice processes. Cold Regions Science and Technology, 62(1): 3-13. DOI:http://dx.doi.org/10.1016/j.coldregions.2010.02.007
Shen, H.T., Bjedov, G., Daly, S.F., Wasantha Lal, A., 1991. Numerical model for forecasting Ice Conditions on the Ohio River, DTIC Document.
Shen, H.T., Liu, L., 2003. Shokotsu River ice jam formation. Cold Regions Science and Technology, 37(1): 35-49. DOI:https://doi.org/10.1016/S0165-232X(03)00034-X
Shen, H.T., Wang, D.S., Lal, A.W., 1995. Numerical simulation of river ice processes. Journal of Cold Regions Engineering, 9(3): 107-118.
Shen, H.T., Yapa, P.D., 1985. A unified degree-day method for river ice cover thickness simulation. Canadian Journal of Civil Engineering, 12(1): 54-62.
Shouyu, C., Honglan, J., 2005. Fuzzy Optimization Neural Network Approach for Ice Forecast in the Inner Mongolia Reach of the Yellow River/Approche d'Optimisation Floue de Réseau de Neurones pour la Prévision de la Glace Dans le Tronçon de Mongolie Intérieure du Fleuve Jaune. Hydrological sciences journal, 50(2).
Shu, C., Burn, D.H., 2004. Artificial neural network ensembles and their application in pooled flood frequency analysis. Water Resour. Res., 40(9).
Strategy, I.G.O., 2007. Integrated Global Observing Strategy Cryosphere Theme Report—For the Monitoring of our Environment from Space and from Earth; WMO/TD-No. 1405. World Meteorological Organization: Geneva, Switzerland.
Suykens, J.A., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural processing letters, 9(3): 293-300.
Tao, W., Kailin, Y., Yongxin, G., 2008. Application of artificial neural networks to forecasting ice conditions of the Yellow River in the Inner Mongolia Reach. Journal of Hydrologic Engineering, 13(9): 811-816.
Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis, part 3, Proceedings of Koninalijke Nederlandse Akademie van Weinenschatpen A. Vol. 53.
Theil, H., 1992. A rank-invariant method of linear and polynomial regression analysis, Henri Theil's contributions to economics and econometrics. Springer, pp. 345-381.
Tiwari, M.K., Chatterjee, C., 2010a. Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. Journal of Hydrology, 394(3): 458-470.
Tiwari, M.K., Chatterjee, C., 2010b. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). Journal of Hydrology, 382(1–4): 20-33. DOI:https://doi.org/10.1016/j.jhydrol.2009.12.013
Tuthill, A.M., 1999. Flow control to manage river ice, COLD REGIONS RESEARCH AND ENGINEERING LAB HANOVER NH.
Unterschultz, K., Van der Sanden, J., Hicks, F., 2009. Potential of RADARSAT-1 for the monitoring of river ice: Results of a case study on the Athabasca River at Fort McMurray, Canada. Cold regions science and technology, 55(2): 238-248.
Wang, J., Sui, J., Guo, L., Karney, B., 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, W.-C., Chau, K.-W., Cheng, C.-T., Qiu, L., 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Journal of Hydrology, 374(3–4): 294-306. DOI:http://dx.doi.org/10.1016/j.jhydrol.2009.06.019
Williams, G., Layman, K.L., Stefan, H.G., 2004. Dependence of lake ice covers on climatic, geographic and bathymetric variables. Cold Regions Science and Technology, 40(3): 145-164. DOI:https://doi.org/10.1016/j.coldregions.2004.06.010

Yaseen, Z.M. et al., 2016. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. Journal of Hydrology, 542: 603-614. DOI:https://doi.org/10.1016/j.jhydrol.2016.09.035

Zaier, I., Shu, C., Ouarda, T.B.M.J., Seidou, O., Chebana, F., 2010. Estimation of ice thickness on lakes using artificial neural network ensembles. Journal of Hydrology, 383(3–4): 330-340. DOI:http://dx.doi.org/10.1016/j.jhydrol.2010.01.006

Zhao, L., Hicks, F.E., Fayek, A.R., 2012a. Applicability of multilayer feed-forward neural networks to model the onset of river breakup. Cold Regions Science and Technology, 70: 32-42. DOI:http://dx.doi.org/10.1016/j.coldregions.2011.08.011

Zhao, L., Jin, J., Wang, S.Y., Ek, M.B., 2012b. Integration of remote‐sensing data with WRF to improve lake‐effect precipitation simulations over the Great Lakes region. Journal of Geophysical Research: Atmospheres, 117(D9).