**CGU HS Committee on River Ice Processes and the Environment**
22nd Workshop on the Hydraulics of Ice-Covered Rivers
*Canmore, Alberta, Canada, July 9-12, 2023.*

# Extending a stochastic modelling framework for ice-jam flood predictions with machine learning

**Apurba Das[1], and Karl-Erich Lindenschmidt[2]**

*[1]Global Institute for Water Security, Water Security Agency*
*11 Innovation Boulevard, Saskatoon, Saskatchewan, S7N 3H5*
*apurba.das@usask.ca*
*[2]Global Institute for Water Security, School of Environment and Sustainability*
*11 Innovation Boulevard, Saskatoon, Saskatchewan, S7N 3H5*
*karl-erich.lindenschmidt@usask.ca*

Ice jams can cause severe flooding along many rivers in cold regions. Ice-jam-associated flooding can harm human and natural systems, such as damage to infrastructure, houses, and navigation systems and fatal to inhabitants. Predicting the occurrence and severity of ice jam floods is vital for mitigating their impacts. In recent years, machine learning has been used in river ice research, and these models have shown promising results in predicting trends, breakup and freeze-up timing. However, there are still challenges associated with using a machine learning approach for ice jam flood prediction, such as the limited availability of historical data and the difficulty in accounting complex and dynamic nature of ice jam formation. In this study, we tried to address these challenges and developed a machine learning-based ice-jam flood prediction model. A stochastic framework was applied to generate thousands of probable ice-jam scenarios to generate a machine-learning modelling data set. In this stochastic framework, a river ice hydrodynamic model, RIVICE, was embedded into a Monte-Carlo Analysis (MOCA) framework to generate thousands of river ice scenarios. The simulated parameter set and associated backwater level elevation outputs were then used to develop a machine-learning prediction model. The study presented the potential of using stochastic outcomes to develop a machine-learning model and improved the capability of ice-jam flood prediction approaches for the ice-jam vulnerable communities. The Athabasca River, Canada, at Fort McMurray was used as a test site to develop this machine learning prediction framework.

## 1. Introduction

River ice hydraulic models have been increasingly used to predict ice-jam flood forecasting (Huokuna et al. 2021; Lindenschmidt et al., 2020; Lindenschmidt et al. 2021). These models often require a high level of expertise and users, extensive input parameters, and high computational costs. Hence limiting their application in ice-jam flood forecasting and management. On the other hand, advanced machine learning (ML) models need a low level of expertise and provide a fast solution to a forecasting problem.

The ML models can recognize the complex pattern and correlations between hydrological inputs and associated outputs without explicitly describing the underlying multifaceted hydraulic mechanisms and processes. These models can be trained using a collection of a large and diverse dataset that includes historical observation, records, and characteristics to predict ice-jam conditions (Mosavi et al. 2018). In recent times, many ML models have been developed and successfully applied for predicting various ice-jam conditions, such as ice-jam backwater level, breakup dates, and climate change impacts.

One of the biggest challenges of developing a ML model is that it requires many observations to recognize an accurate pattern and trends of predictors and hydrological responses. However, the dataset related to river ice events is often short and challenging to collect, particularly in remote areas and extreme weather conditions. Moreover, ice-jam formation can cause damage to the hydrometric gauge stations, and the equipment is often inoperable to collect data; thus, important data records can be lost from the dataset (Kovachis et al. 2017). These limitations can be resolved by simulating ice-jam events by applying a stochastic modelling (SM) framework (Lindenschmidt et al. 2019) using a river ice hydraulic model. An SM framework to simulate hundreds of ice-jam scenarios has been applied in many rivers used in this study to generate many training sets for the ML model.

This study applied the SM framework to generate hundreds of ice-jam events. The parameter set and simulated output from the SM framework were used for training and validating the model. The main purpose of this study is to extend the SM framework using a ML algorithm to predict ice-jam flooding. The specific objectives are to (1) introduce a strategy to apply modelling results to develop a ML model; (2) assess the effectiveness of the use of simulated results in the ML model, and (2) improve the prediction capability of the ice-jam flooding.

## 2. Study area

The Athabasca River (Figure 1) is one of the major rivers in Alberta, Canada, which originates at the Columbia Icefield in the Canadian Rockies. The river flows approximately 1538 km before draining into Lake Athabasca. Numerous islands, sandbars, a series of rapids and abrupt changes in river slopes characterize the geomorphological settings of the river.

The town of Fort McMurray lies in the northeastern region of Alberta, and the town's downtown area is located at the confluence of the Athabasca and Clearwater rivers. The site is very vulnerable to ice-jam flooding, as the ice jams along the Athabasca River significantly impact the flow and rise in backwater levels along the Clearwater River channel to flood the low-lying areas. Over the years, Fort McMurray has experienced devastating ice jam flooding events that have caused widespread damage to infrastructure, homes, and the local community. Historically, severe ice jam events were recorded in 1977, 1978, 1979, 1997, and 2020.

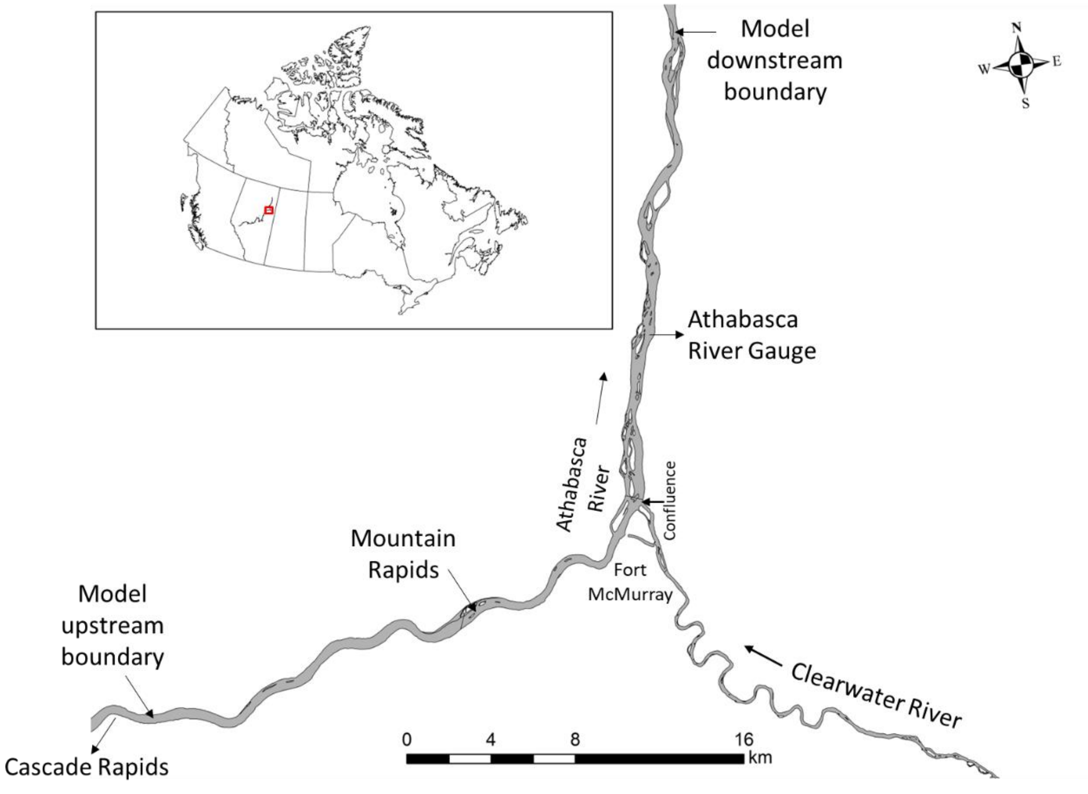


Figure 1 Model domain along the Athabasca River at Fort McMurray

## 3. Stochastic Modelling Approach

The calibrated hydraulic model, RIVICE, was embedded into a stochastic method to simulate thousands of ice-jam scenarios along the model domain. RIVICE is a one-dimensional river ice hydrodynamic model to simulate various river ice processes, including frazil ice generation, intact ice cover hydraulics and ice-jam formation. The model requires various river ice and hydraulic parameters and boundary conditions as inputs to simulate ice-jam processes. The parameters include the porosity of the ice-cover *PC* and rubble ice *PS*, the thickness of the ice-cover front *FT*, the thickness of the intact ice-cover *h* and ice floes *ST*, the roughness coefficient of the ice $n_8$ and riverbed $n_b$, the friction between the ice cover and the riverbanks *K1* and the stability of the ice-jam through its thickening *K2*, and the ice erosional threshold velocity $v_{er}$, and deposition velocity $v_d$. The boundary condition inputs are the incoming volume rate of ice $V_{ice}$, upstream discharge *Q*, and toe of ice-jam locations *x* (Figure 2).

The SM approach was implemented using a Monte-Carlo Analysis (MOCA) framework, where frequency distributions of parameter and boundary conditions values are determined from observed, calibrated and previously attained data. Upstream flow frequency distributions were estimated using observed flows during ice-jam events. A relationship was established between $V_{ice}$ and *Q* data to determine the incoming volume rate of ice (see Lindenschmidt, 2023 for details). The ice jam toe locations were derived from a uniform distribution based on historical ice-jam information. The values for the parameters were randomly selected from a uniform distribution.

The range of these distributions was determined through calibration using multiple ice-jam events, as conducted by Rokaya and Lindenschmidt (2020). After extracting the parameter and boundary condition values from each of their corresponding distributions and functions, the simulations were carried out to generate thousands of ice-jam scenarios.

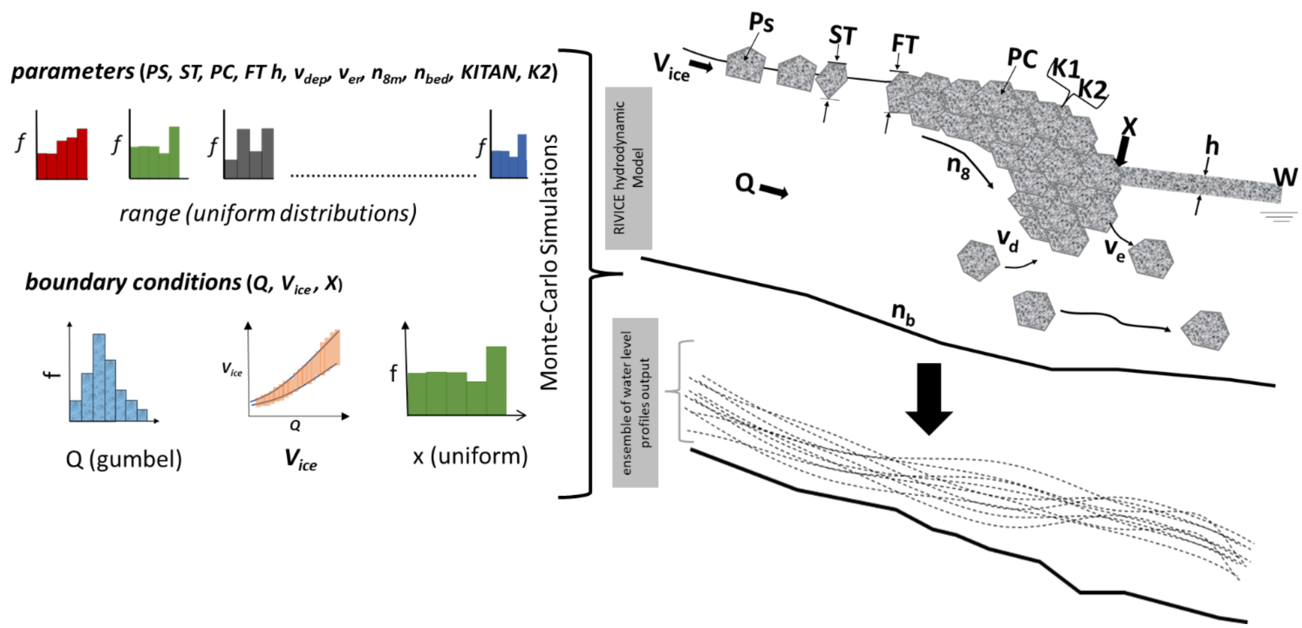


Figure 2 Stochastic approach and parameters and boundary conditions used to simulate river ice processes in RIVICE hydrodynamic model.

## 4. Machine Learning Approach

A decision tree regression (DTR) algorithm was applied to predict ice-jam water level elevation using stochastic model variable sets and results. The DTR is a supervised learning algorithm which creates a tree-like model by splitting the data based on the feature values (Figure 3). To classify data instances, the model follows a path starting from the root of the tree and ending at a leaf node. At each node, a test is performed on a particular attribute, with the branches representing the possible outcomes of the test. The leaf nodes provide the predicted group.

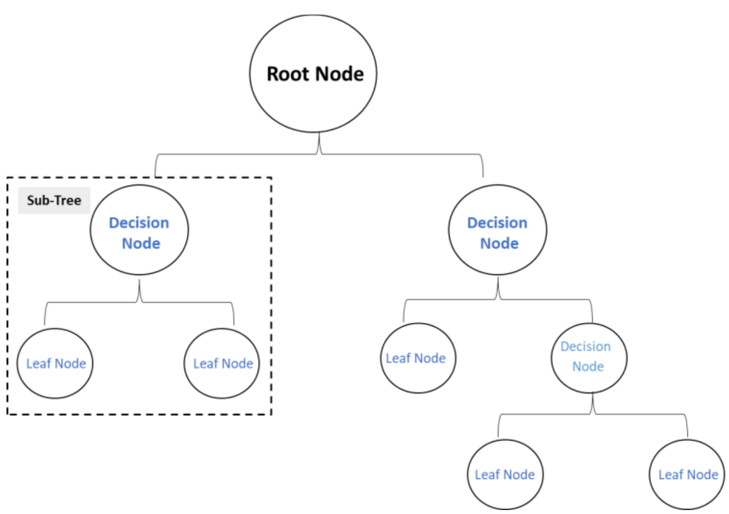

Figure 3: Schematic diagram of the decision tree regression approach.

The input sets from the stochastic approach were refined based on the observed data to select the training and test feature sets for the DTR algorithm. A threshold envelope was developed using the relationship between flows and water level elevation at the hydrometric gauge station downstream of Fort McMurray during ice-jam events (Figure 4). The envelope was then used to select the input feature set for training and testing the model. The parameter set that produced the results outside of this envelope was considered an outlier and removed from the machine-learning model input features.

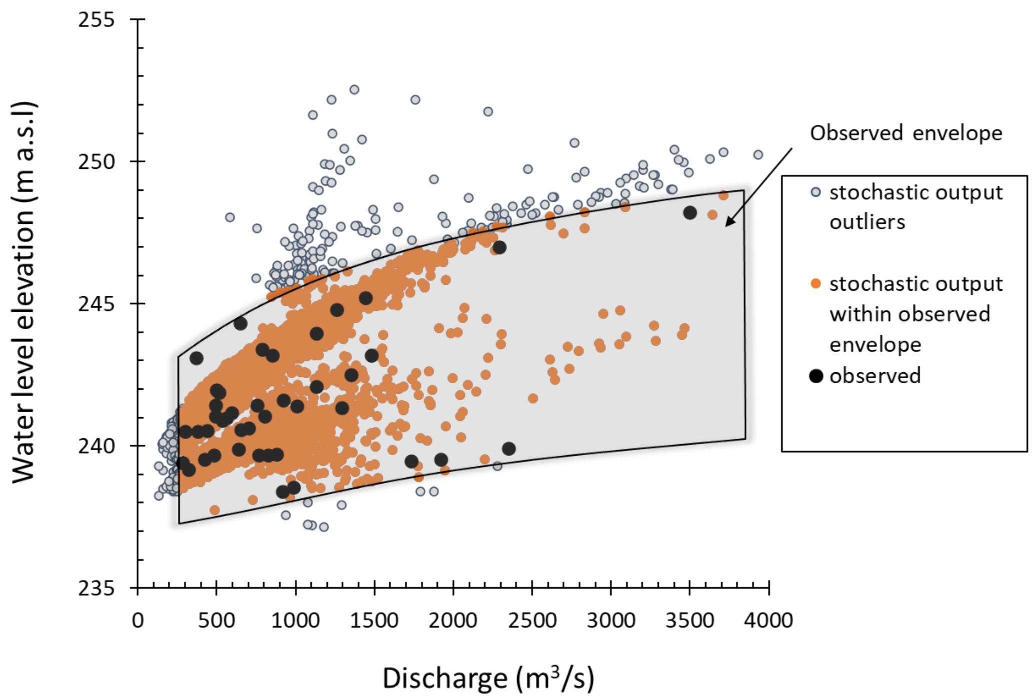


Figure 4: Relationship between ice-affected water level elevation and discharge data recorded at the hydrometric gauge station downstream of Fort McMurray during ice-jam events. The overall approach of the ML model is illustrated in Figure 5. The stochastic model was applied to produce hundreds of ice jam scenarios and associated backwater level profiles along the model domain. Then backwater level at the confluence of the Athabasca and Clearwater rivers was extracted to create an input dataset for the machine-learning model. The DTR model was then applied to train the model for the prediction of the ice-jam associated backwater level at the same location.

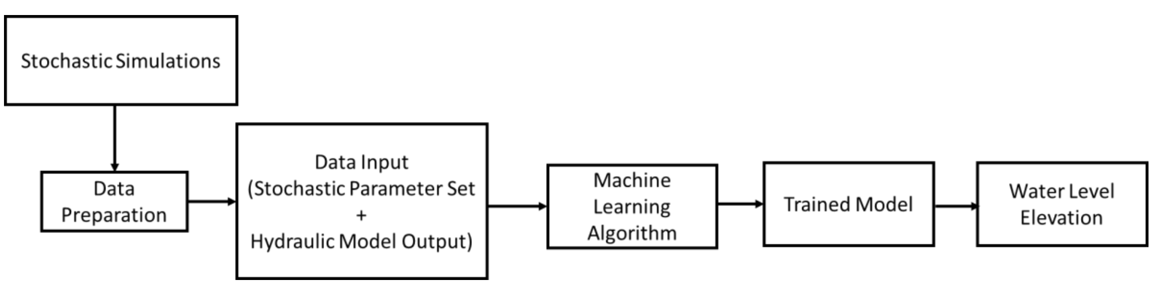

Figure 5: The overall approach of the machine learning model.

## 4. Result and Discussion

The stochastic model framework was extended to a machine-learning model to predict backwater level elevation in the vicinity of Fort McMurray at the confluence of the Athabasca and Clearwater rivers. A total of 6500 ice-jam simulated scenarios were used to develop the machine-learning model. The model was trained and tested by splitting the dataset into two groups. First, seventy percent of the data was used for model training, and the rest of the data was used for testing. The accuracy of the test model was estimated to be $R^2 = 0.91$ and RMSE = 0.229 m (Table 1).

Table 1 Performance statistics of the model during training and testing periods.

| **Prediction** | $\mathbf{R^2}$ | **RMSE (m)** |
|---|---|---|
| Test set | 0.91 | 0.229 |
| Validation set | 0.90 | 0.565 |

A new stochastic dataset was used to validate the model results. The result also produces a high accuracy with $R^2 = 0.9$ and RMSE = 0.565 m. RMSE values in the validated model are relatively higher only a small number of the stochastic dataset was applied to validate the model. The result shows that the model successfully predicted the backwater level elevation for most of the ice-jam scenarios (Figure 6). This result can be very helpful in assessing the pattern of the stochastic parameter set before applying the framework. Since a ML prediction model requires a large number of data, the SM approach has the capability to produce thousands of ice-jam scenarios that can be used to develop a prediction model.

Although the ML model was able to predict most of the ice-jam backwater levels with reasonable accuracy, some of the high backwater level elevations were underestimated (Figure 6). Since the ML model requires a large number of datasets to learn and predict precise values, the number of extreme ice-jam events in the training dataset was relatively low, limiting the accurate prediction of these extreme ice-jam water level elevations. The SM approach can help to overcome this limitation by simulating large numbers of ice-jam events and supplementing small datasets to train the ML model.

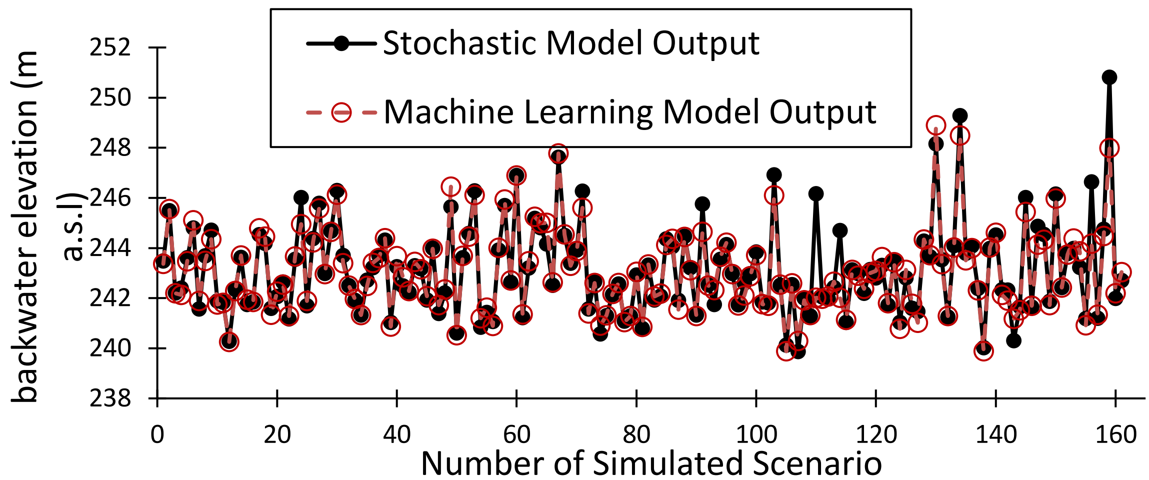


Figure 6: Machine learning model vs stochastic model outputs.

## 5. Conclusion

The SM output was successfully applied to develop a ML model to predict ice-jam backwater level elevations along the Athabasca River at Fort McMurray. This ML model is simple and easier to execute; therefore, it can be used as a pre-assessment tool for selecting the MOCA parameter set for the ice-jam stochastic framework. It could remove any unrealistic parameter set that can be generated during random variable selection based on the preliminary result of the backwater level elevation of a particular location.

Future work will be focused on identifying the most important parameters and boundary conditions to predict the ice-jam-associated backwater level elevation using a sensitivity analysis. The sensitivity analysis can also help to reduce the number of input variables for the ML model and increase the performance accuracy of the model. Moreover, ML model output will be tested using observed flood hazard data, such as predicting the backwater level for a specific return period.


## Acknowledgments

We thank Global Water Future's core modelling for funding this study.


## References


Huokuna, M., Havu, P., Triipponen, J.-P., Koivisto, T. and Aaltonen, J. (2021) Ice forecasting model for the lower Kokemäenjoki River. CGU HS Committee on River Ice Processes and the Environment. 21st Workshop on the Hydraulics of Ice Covered Rivers, Saskatoon, Saskatchewan, Canada, 29 August 29 - 1 September 2021.

Kovachis, N., Burrell, B. C., Huokuna, M., Beltaos, S., Turcotte, B., & Jasek, M. (2017). Ice-jam flood delineation: Challenges and research needs. Canadian Water Resources Journal/Revue canadienne des ressources hydriques, 42(3), 258-268.

Lindenschmidt, K. E., Rokaya, P., Das, A., Li, Z., & Richard, D. (2019). A novel stochastic modelling approach for operational real-time ice-jam flood forecasting. Journal of Hydrology, 575, 381-394.

Lindenschmidt, K.-E. (2023) Extension and refinement of a stochastic modelling approach to assess ice-jam flood hazard. *Hydrology Research* **54**(2): 149–164 https://doi.org/10.2166/nh.2023.042

Lindenschmidt, K.-E., Brown, D., Hahlweg, R., Luo, B., Propp, C. and Weiss, A. (2021) Preparing ice-jam flood outlooks for the lower reach of the Red River. CGU HS CRIPE 21st Workshop on the Hydraulics of Ice Covered Rivers, Saskatoon, Saskatchewan, 29 Aug. - 1 Sep. 2021.

Lindenschmidt, K.-E., Brown, D., Khan, A.A., Khan, H., Khayer, M., McArdle, S., Mostofi, S., Naumov, A., Pham, T. and Weiss, A. (2020) A novel fully-operational fully-automated real-time ice-jam flood forecasting system. 25th IAHR International Symposium on Ice, Trondheim, 23 – 25 November 2020. https://www.iahr.org/library/infor?pid=8570

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. Water, 10(11), 1536. https://doi.org/10.3390/w10111536

Rokaya, P., & Lindenschmidt, K. E. (2020). Correlation among parameters and boundary conditions in river ice models. Modeling Earth Systems and Environment, 6, 499-512.