Forecasting Ice Thickness on the Churchill River and Lake Melville, Labrador Using Machine Learning, 2023-2025

Ian D. Turnbull[1], Michael Lynch[1], Maria Yulmetova[1], and Pradeep Bobby[1]

[1]C-CORE, Captain Robert A. Bartlett Building, 1 Morrissey Road, St. John's, NL A1B 3X5

Keywords: River ice monitoring, Ice forecasting, Machine learning.

Abstract

During the winters of 2023-2024 and 2024-2025, machine learning (ML) based models were implemented to predict ice thickness at eight sites on the Churchill River and Lake Melville, Labrador for one- and three-day horizons. The forecast ice thicknesses were fed into the Churchill River Flood Forecasting System (CRFFS) operated by the Newfoundland and Labrador (NL) provincial government's Water Resources Management Division (WRMD). The models were trained on measured ice thickness data from 2017-2023, with the 2024-2025 models additionally trained with data from the 2023-2024 ice season. The 2023-2024 models were deep learning models that used Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNNs), and the 2024-2025 models were ML models that used a simpler gradient boosting regression (GBR) algorithm. The LSTM (2023-2024) models used a running time-series of local meteorological observations as predictor variables to directly forecast ice thickness, and the GBR (2024-2025) models mainly used forecast surface energy balance variables to predict changes in ice thickness. The average performance of the models across the eight sites was comparable between the two ice seasons; however, the 2024-2025 season models improved performance at key sites on the Churchill River that are critical to ice jam flood forecasting. This paper describes the development of the models and their operation and comparative performance over the 2023-2025 ice seasons.

1. Introduction

During 2023-2025, daily forecast reports of ice thickness were provided along the Churchill River and on Lake Melville in Labrador to the Water Resources Management Division (WRMD) of the provincial Government of Newfoundland and Labrador (NL). The forecast ice thicknesses were then fed into WRMD's Churchill River Flood Forecasting System (CRFFS, 2025). The forecasts were particularly important during the melt season as break-up approached. The WRMD considered the forecast ice thicknesses as critical input to the CRFFS. During break-up season, floods are primarily due to ice jams and can cause significant damage to local communities and force evacuations. The community of Mud Lake (see Figure 1), located where the Churchill River meets Lake Melville, is particularly vulnerable to ice jam floods.

River ice monitoring and forecasting are critical due to their wide-ranging environmental, infrastructural, and socioeconomic impacts. River ice influences water flow regimes, aquatic ecosystems, sediment transport, and seasonal hydrology, and the formation of ice jams can lead to sudden and severe flooding that endangers infrastructure and communities in cold-region basins. Seasonal ice also affects navigation, hydropower operations, and winter transportation infrastructure such as ice roads in Arctic and sub-Arctic regions, making accurate observations and forecasts essential for planning and hazard mitigation. Satellite remote sensing, in situ sensors, and modelling approaches have all been developed to monitor ice cover extent, thickness, and dynamics, with more recent advances incorporating radar altimetry, synthetic aperture radar (SAR), and machine learning (ML) techniques to improve spatial and temporal coverage and prediction capabilities. Despite these advances, forecasting river ice processes remains challenging due to dynamic ice behaviour and data limitations, hence, additional research into integrating multi-source observations with numerical and ML-enhanced models is needed to support operational early warning systems.

The ice thickness forecasts discussed in this paper were intended to replace real-time estimates of thickness that had been provided by temperature thermistor strings installed in the ice in previous seasons. The thermistor strings had to be retrieved prior to break-up when the ice was still safe for helicopters to land on it, but knowledge of ice thickness much closer to break-up was critical input for the CRFFS model after the thermistor retrieval. The ice thickness forecast models filled that critical gap in the daily thickness timeseries.

Ice thickness was forecast at eight selected points along the Churchill River and on Lake Melville (see Figure 1 and Table 1) at one- and three-day horizons. The eight points were selected based on locations at which WSP (a Canadian engineering and consulting firm) took biweekly ice thickness measurements. Only sites N8 and N9 were on Lake Melville while the other six sites were along the Churchill River.

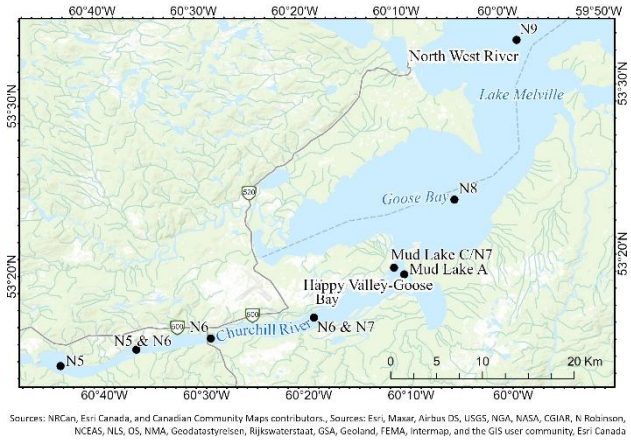

Figure 1. Locations of the eight Churchill River and Lake Melville sites for ice thickness prediction.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-61-2026 | © Author(s) 2026. CC BY 4.0 License.

For the 2023-2024 ice season, a set of eight separate machine learning based models were developed for each of the eight sites, and in 2024-2025, a single model was developed for all sites. The models developed for 2023-2024 were Long Short-Term Memory (LSTM), which are a type of Recurrent Neural Network (RNN). LSTM models are well-suited to time-series based prediction problems because they have "memory cells" that "remember" long-term variable dependencies (Shertinsky, 2020). They use the recent and present state of a set of predictor variables, or "features" to predict the predictand variable in the future at some specified lag at which the model is trained.

Site	Latitude (°N)	Longitude (°W)
N5	53.244503	60.736855
Between N5 and N6	53.258071	60.612499
N6	53.266477	60.489923
Between N6 and N7	53.283239	60.32027
Mud Lake C/N7	53.329139	60.186266
Mud Lake A	53.322331	60.170255
N8	53.393533	60.083033
N9	53.547369	59.970317

Table 1. Coordinates of the eight sites.

For the 2024-2025 season, a new ice thickness prediction approach was created, combining physical modelling with ML. The ML component employed a Gradient-Boosted Regression (GBR) algorithm, as described by (Prettenhofer and Louppe, 2014). This hybrid model contrasts with the one used during the 2023-2024 season, which relied solely on a deep learning LSTM architecture. Unlike the 2023-2024 model, which predicted ice thickness directly as a time-series, the 2024-2025 model was designed to estimate changes in ice thickness over one to three-day periods. Full ice thickness was then updated simply by adding the one- to three-day predicted change in ice thickness to the last daily measured or forecast ice thickness value. In contrast to the 2023-2024 season in which a one- and three-day ice thickness prediction model was trained for each site, a single one-day and a single three-day GBR ice thickness prediction model was trained and applied to all sites. This updated modelling strategy was implemented to streamline the model architecture, based on evidence that simpler machine learning models often yield better performance when training data are limited (e.g., see Söylemez, 2024). Additionally, it was adopted to incorporate a hybrid physical-ML framework, as physics-informed approaches have been shown to outperform purely data-driven models in problems grounded in physical systems (e.g., see Bhasme et al., 2021).

This paper describes the development of the models over the 2023-2025 ice seasons, compares their overall performance, and discusses further model development for future ice seasons. Section 2 discusses the models' training data, Section 3 discusses the models' architecture and training processes, Section 4 discusses the relative performance of the models during the two ice seasons, and Section 5 discusses conclusions of this work and future model development.

2. Model Training Data

2.1 Measured Ice Thickness

Models were trained using measured ice thicknesses collected by WSP over the 2017-2024 ice seasons (only through 2023 for the 2023-2024 season). The data were provided in biweekly ice thickness and water sampling reports generated for the eight sites along the Churchill River and on Lake Melville over each ice season. Ice thickness data were linearly interpolated to daily time-series for each site. Ice typically begins to form on the Churchill River and Lake Melville during mid-December to early January and breaks up during mid-April to mid-May. Peak seasonal ice thicknesses are typically in the range of 90-130 cm (see Figure 2).

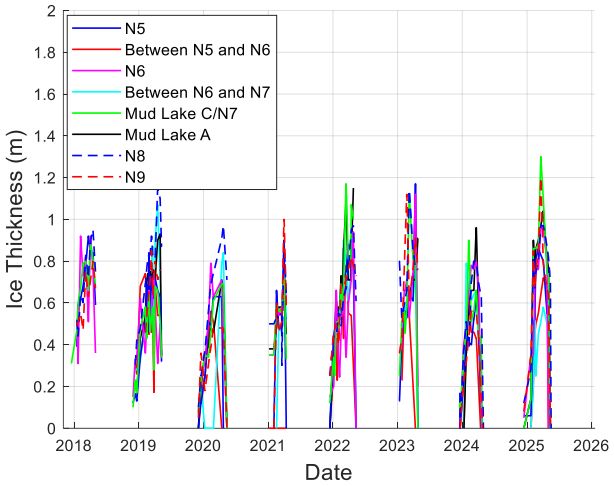

Figure 2. Measured ice thickness on Churchill River and Lake Melville.

2.2 2023-2024 Season Model Training Features

The one- and three-day prediction models were trained for each of the eight sites, resulting in a total of 16 separate LSTM models. This was done because the way the ice thickness growth and melt respond to the weather variables could be different at each site due to local river dynamics.

Daily mean predictor variables were derived from observed Happy Valley-Goose Bay (HVGB) Environment and Climate Change Canada (ECCC, 2024) automatic weather station data for 2017-2023, and included the following variables:

- 2 m air and dew point temperature,
- 10 m wind speed and direction,
- atmospheric pressure, and
- Freezing Degree Days (FDD) and Positive (or Thawing) Degree Days (PDD).

The FDD and PDD were computed as follows:

$$FDD = (T_{fr} - T_a) \text{ if } T_a < T_{fr}, \text{ and}$$
$$PDD = (T_a - T_{fr}) \text{ if } T_a \geq T_{fr}. \quad (1)$$

Here, T_{fr} is the reference freezing temperature, set at -5°C for river ice following (Bilello, 1980), and T_a is the air temperature in degrees Celsius.

The input features to the LSTM models were preprocessed using the standard scaling method. This method involves subtracting the mean of all samples in a dataset from each value for a given variable, and then dividing by the standard deviation of all samples in the dataset for that variable, as follows:

$$X_{scaled_i} = \frac{X_i - \mu_x}{S_x}. \quad (2)$$

In Equation 2:

X_{scaled_i} = the scaled value of a given variable at the i^{th} sample index,

X_i = the unscaled value of the variable at the i^{th} sample index,
μ_x = the mean of the variable over all time-steps, and
S_x = the standard deviation in the variable over all time-steps.

The model predictions were finally inversely scaled to produce the predicted surface temperatures. The models were trained for each season on 80% of the data and tested on the remaining 20% according to the Pareto Principle (Macek, 2008 and Ozdemir, 2016).

Figure 3 shows the time-series of daily mean air and dew point temperature recorded at the ECCC automatic weather station at HVGB airport during December 18, 2023, through April 29, 2024. This covers the full period from which the first ice thickness measurements were made until break-up occurred at the last measurement and forecasting site (N8). Figure 3 shows daily mean air temperatures fluctuated significantly during the season, with three instances of temperatures reaching above 0°C from late February through the end of March prior to the beginning of the normal break-up season. Daily mean air temperatures began to reach more consistently above 0°C during April when break-up occurred at all sites.

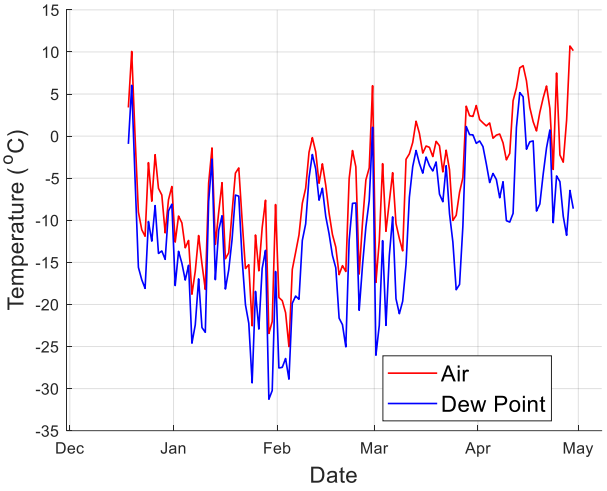

Figure 3. Measured air and dew point temperatures at HVGB during the 2023-2024 ice season.

2.3 2024-2025 Season Model Training Features

For the 2024-2025 season, a total of 11 input features were chosen for use in the ice thickness prediction models:

- net shortwave (solar) radiation at the surface,
- net longwave radiation at the surface,
- latent heat flux from the surface,
- sensible heat flux from the surface,
- surface (skin) temperature,
- air temperature at two meters above the surface,
- snow depth (meters),
- ice thickness change estimated from a one-dimensional (1D) thermodynamic ice growth/melt model (Equation 3),
- conductive heat flux through the snow/ice layer (Equation 5), and
- FDD and PDD (Equation 1).

The surface radiation and heat flux variables, specifically, the net shortwave and longwave radiation, and the latent and sensible heat fluxes, represent components of the surface energy balance (with downward flux considered positive). These variables, along with air and surface temperatures, were sourced from the ERA5-Land reanalysis dataset (e.g., see Muñoz Sabater, 2019), which offers hourly data at a 0.1° spatial resolution (approximately 11 km) over global land areas for 1950-present.

To align with the prediction windows, the ERA5-Land variables were averaged using one- and three-day periods. As such, the models predicted ice thickness changes as a function of the average values of the input variables over the corresponding time periods. While the ERA5-Land were used to train the ML models, ECCC weather forecast data from the Global Deterministic Prediction System (GDPS, 2025). The GDPS is a numerical weather prediction model that provides three-hourly forecasts of variables out to 240 hours (10 days) at 15 km spatial resolution, issued twice daily at 00:00 and 12:00 Universal Time Coordinated (UTC). The GDPS data were averaged over the following one and three days to provide predictor variables for the one- and three-day forecast changes in ice thickness.

The change in ice thickness (in meters), estimated using a 1D thermodynamic model of ice growth and melt, was calculated following the approach of (Liston and Hall, 1995) as:

$$Dz = \left(\frac{\frac{T_f - T_{s0}}{\frac{z_i}{k_i} + \frac{z_s}{k_s}} - h_w(T_w - T_f)}{\rho_i L_f} \right) \Delta t. \quad (3)$$

In Equation 3:

k_i = the thermal conductivity of ice (2.034 W·m^{-1}·K^{-1}),
L_f = the latent heat of fusion for ice (3.36 × 10^5 J·kg^{-1}),
ρ_i = the density of freshwater ice (917 kg·m^{-3}),
h_w = the convective heat transfer coefficient at the ice-water interface (0.56 W·K^{-1}·m^{-2}),
T_f = the freezing point temperature (-0.1096°C which was commonly the approximate reading from the thermistor at the ice base from the SIMBAs),
T_{s0} = the daily mean surface (skin) snow or ice temperature (°C) at the present time-step,
T_w = the water temperature (assumed to be 0°C),
z_i = the daily ice thickness (meters) at the present time-step,
z_s = the daily mean snow depth (meters) at the present time-step,
k_s = the snow thermal conductivity (W·m^{-1}·K^{-1}), and
Δt = the time step, set to 172,800 seconds for one-day changes in ice thickness (equivalent to the number of seconds over two full days, the present and next day) and 345,600 seconds for three-day- changes in ice thickness (equivalent to the number of seconds over four full days, the present day and the next three days).

The snow thermal conductivity was calculated by (Liston and Hall, 1995):

$$k_s = 0.074 + 2.576 \times 10^{-6} \rho_s^2, \quad (4)$$

where ρ_s is the snow density (kg·m^{-3}) which was obtained from ERA5-Land for model training and the GDPS for operational forecasting.

The conductive heat flux through the snow/ice layer (W·m^{-2}) was determined by (Liston and Hall, 1995):

$$Q_c = -\frac{k_{is} \partial T}{\partial z}, \quad (5)$$

where the temperature gradient $\frac{\partial T}{\partial z}$ is calculated as the difference between the snow surface and basal ice temperatures, divided by the snow and ice combined thickness, yielding units of K·m^{-1} (Liston and Hall, 1995). The thermal conductivity, k_{is}, was determined as a thickness-weighted average of the snow and ice thermal conductivities.

The FDD and PDD were computed using the ERA5-Land 2m air temperature data (for model training and using the forecast GDPS air temperature in operational forecasting).

Figure 4 shows the time-series of daily mean air temperature over the ice season from December 13, 2024, through May 16, 2025, at the HVGB airport ECCC weather station and the GDPS forecast from site N6, the closest ice thickness site to the weather station (7.6 km away). The GDPS air temperature forecast was overall highly accurate throughout the ice season. Multiple periods of freeze-thaw are apparent during the 2024-2025 ice season. Daily mean air temperatures began to reach more consistently above 0°C beginning in mid-April.

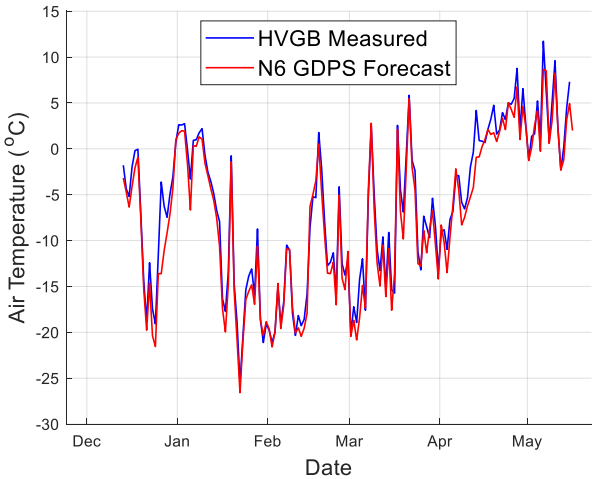

Figure 4. HVGB measured and site N6 GDPS forecast air during the 2024-2025 ice season.

Figure 5 shows the daily mean GDPS forecast surface energy balance variables at site N6 that served as predictor variables for the ML model over December 13, 2024, through May 16, 2025. A net positive total surface energy balance generally began to dominate beginning in the latter half of March that was mainly due to the increase in daily mean surface solar radiation.

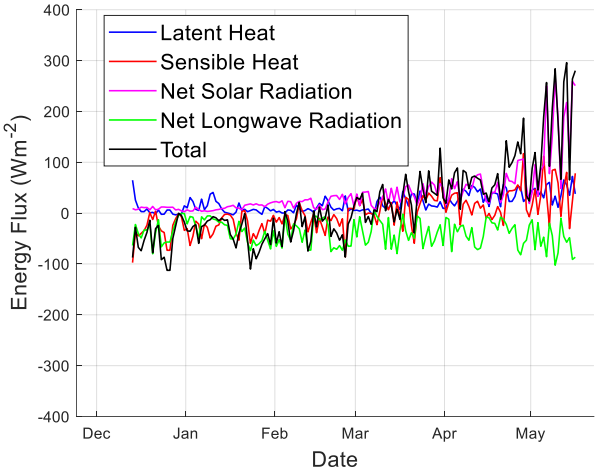

Figure 5. Forecast surface energy fluxes at site N6 during the 2024-2025 ice season.

3. Model Architecture and Training

This section discusses the model architecture and training processes for the LSTM models developed for the 2023-2024 season (Section 3.1) and the GBR models developed for the 2024-2025 season (Section 3.2).

3.1 2023-2024 Season Models

The 2023-2024 season LSTM models were developed in the open-source TensorFlow library (TensorFlow, 2015) in the Python computer programming language. The recent and present ice thickness and observed weather variables were used to forecast the ice thickness at one and three days into the future (e.g., two models were trained at one- and three-day lag periods for each of the eight sites). Ice thickness and weather conditions further in the past have decreasing influence on ice thickness predictions as time progresses. Figure 6 shows how the LSTM model works.

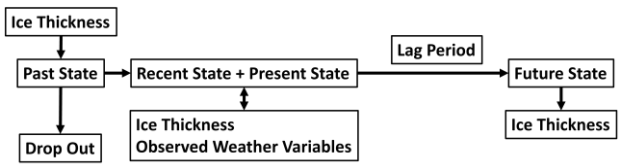

Figure 6. Schematic of how an LSTM model works for ice thickness prediction.

A simple model architecture proved to be the best overall for model development due to the relatively limited amount of training data and number of features. Figure 7 shows the optimized model architecture determined from hyperparameter tuning (e.g., experimenting with different model architectures in terms of numbers of layers, units, regularization methods, and activation functions). The predictor features (e.g., the daily mean meteorological variables and ice thickness) were fed into a bidirectional LSTM model with 256 units or "nodes," followed by a dropout layer (0.4) and then a Dense output layer with a single node.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-61-2026 | © Author(s) 2026. CC BY 4.0 License.

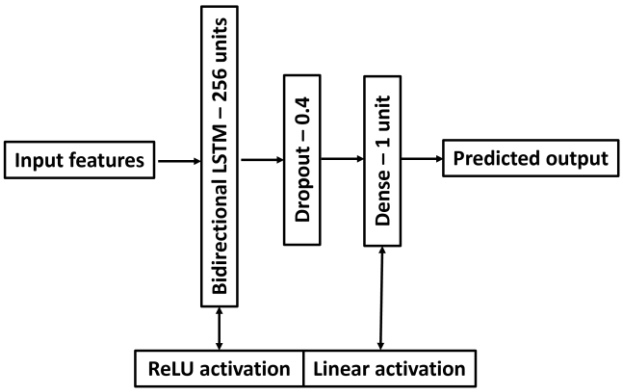

Figure 7. LSTM model core architecture.

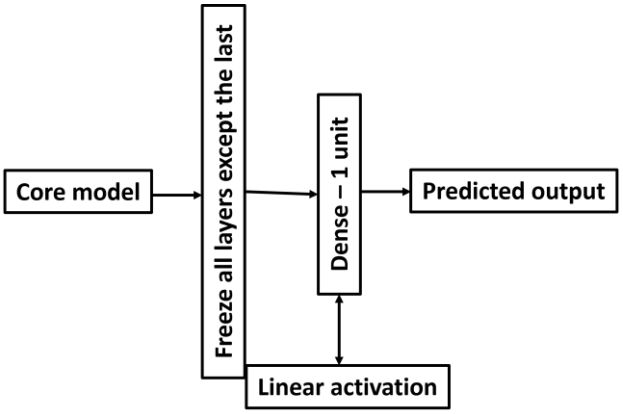

Figure 8. LSTM model transfer learning architecture.

Bidirectional LSTM models process the sequential data in both forward and backward directions during the training phase. In this way, the model captures both past and future context of the input sequence. Dropout layers are a method of regularizing an ML model to mitigate overfitting to the training dataset, so that model can generalize more effectively to unseen data and still make good predictions. The dropout layer of 0.4 means that around 40% of the eight training features (e.g., three features) were randomly selected to be dropped from a given training epoch (cycle) to prevent certain features from becoming too heavily weighted over the full set of training epochs.

The models were permitted to train for up to 80 epochs, but the training was stopped early if after five consecutive epochs the Mean Absolute Error between the observed and predicted ice thickness did not decrease. “Early Stopping” is another method of ML model regularization to mitigate overfitting to the training data. A Rectified Linear Unit (ReLU) activation function was used in the bidirectional LSTM layer, and a linear activation function was used in the Dense output layer. Activation functions help the neural network learn complex patterns in the data, and “decide” on the variable weights to be passed onto the next model layer. The ReLU function is a commonly used activation function in deep learning models that introduces nonlinearity and solves the vanishing gradients issue. The vanishing gradient problem occurs when model weights undergoing updating via gradient descent across training epochs become so small such that weights no longer update properly. The ReLU function returns zero if it receives negative input and returns the value it received if the input is positive. The linear activation function is used in the output layer for regression models trained to return a quantitative predictand and simply returns the weighted sum of its input.

The models for each site were initially trained on the 2017-2018 ice season using the model architecture shown in Figure 7. Each model was subsequently retrained on the successive ice seasons up through the 2022-2023 season using the architecture shown in Figure 8. For each season from 2018-2019 onward, the models for each site were retrained by freezing all layers except the last and allowing only the last Dense layer to be retrained on the new seasonal data. This is a process known as “transfer learning” in which an ML model retains its original deepest layers, and the outermost layer is “tweaked” based on new training data.

3.2 2024-2025 Season Models

The 2024-2025 season models were implemented in Python using the scikit learn library (Pedregosa et al., 2011). Gradient-boosted regression is an ensemble learning method that constructs a series of decision trees in a step-by-step fashion, with each new tree aiming to fix the prediction errors made by the previous ones. It begins with a basic initial model, usually a shallow decision tree, and progressively improves performance by adding trees that model the residuals, using gradient descent for optimization.

Key hyperparameters that influence the behaviour and performance of a GBR model include the:

- learning rate,
- maximum depth of each tree,
- minimum samples required at a leaf node,
- minimum samples needed to split a node, and
- the total number of estimators (trees).

The learning rate determines how significantly each new tree adjusts the model, affecting both convergence and the likelihood of overfitting. Tree depth influences the model’s complexity; deeper trees can capture more intricate patterns but also raise the risk of overfitting. Setting a minimum number of samples per leaf helps regulate prediction detail, while the minimum samples per split prevents nodes from being split too early, maintaining a balance in model complexity. The number of estimators controls how many trees are included in the model, with more trees generally improving predictions but increasing computation cost. Together, these hyperparameters allow fine-tuning of the model to achieve a good trade-off between bias and variance. Unlike many other regression techniques, GBR does not require input features to be scaled, simplifying preprocessing. Figure 9 illustrates a schematic where multiple trees are combined, with final tree outputs shown in blue at the leaf nodes.

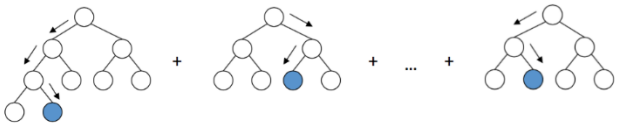

Figure 9. Gradient-boosted regression model architecture (image reproduced from Awasthi, 2023).

Both the one-day and three-day ice thickness prediction models underwent a detailed hyperparameter tuning process to enhance their accuracy and reliability. Table 2 outlines the candidate hyperparameters explored during model development and

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-61-2026 | © Author(s) 2026. CC BY 4.0 License.

validation, along with the final values selected for each model. This tuning process was designed to achieve a balance between predictive accuracy, the ability to generalize to new data, and computational efficiency.

Hyperparameters Tested		Selected Hyperparameters	
		1-Day Forecast	3-Day Forecast
Learning Rate	0.01, 0.05, 0.1, 0.2	0.05	0.05
Maximum Tree Depth	3, 5, 10	10	10
Minimum Samples per Leaf	1, 2, 4	4	4
Minimum Samples per Split	2, 5, 10	5	10
Number of Estimators	10, 50, 100, 200	100	100

Table 2. Ice thickness ML models hyperparameter tuning set for the 2024-2025 season.

In these GBR models, the learning rate adjusts the influence of each tree in the ensemble, striking a trade-off between stable, gradual improvement and faster training. The maximum depth of the trees dictates how complex the model can become; deeper trees allow more detailed representations of the data but may lead to overfitting. The minimum number of samples required at a leaf node helps ensure that predictions are based on sufficient data, promoting generalization while retaining model adaptability. Requiring a minimum number of samples to split a node helps guard against unnecessarily complex trees by enforcing a threshold for node division. The number of estimators, or trees, influences both prediction quality and runtime, with more estimators typically improving performance at the cost of increased computation.

This selected set of hyperparameters supports a well-rounded optimization strategy, aiming to balance performance, robustness, and efficiency. A comprehensive grid search using five-fold cross-validation (k = 5) was used to evaluate all possible combinations, following the 80/20 training-validation split in line with the Pareto Principle (Macek, 2008 and Ozdemir, 2016). For each model, the configuration that minimized the mean squared error between predicted and observed ice thickness changes was selected. Table 2 provides the set of hyperparameters tested and the sets selected for the one- and three-day ice thickness forecast models. For the 2023-2024 season, the 2017-2023 seasons provided the training and validation dataset, and 2023-2024 served as the test set. For the 2024-2025 season, the 2017-2024 seasons provided the training and validation dataset, and 2024-2025 served as the test set.

4. Model Performance

This section discusses the models' performance during the 2023-2025 ice seasons. It is beyond the scope of this paper to explore the model's performance at all eight sites in detail; hence, examples are provided for one key site on the Churchill River that is critical for ice jam flood forecasting in the CRFFS, N6, and one of the Lake Melville sites, N8.

Figure 10 through Figure 13 show the measured and predicted ice thicknesses over the two seasons for sites N6 and N8. The black circles mark the dates on which the biweekly measurements were taken, and the blue lines represent the daily measured ice thicknesses linearly interpolated between measurements. The vertical dashed lines mark the dates on which WSP measurement reports were received, and the models could be reinitialized with the updated ice thickness data. The fact that WSP ice thickness reports were received anywhere from two to 19 days after the fact meant that ice thickness forecasts were in essence forecasts up to almost three weeks.

4.1 Churchill River Site N6

Figure 10 shows the measured and forecast ice thicknesses for site N6 on the Churchill River during the 2023-2024 season. The LSTM model forecasting could not begin until two measurements were taken. The LSTM models captured the overall growth pattern of ice thickness through the season.

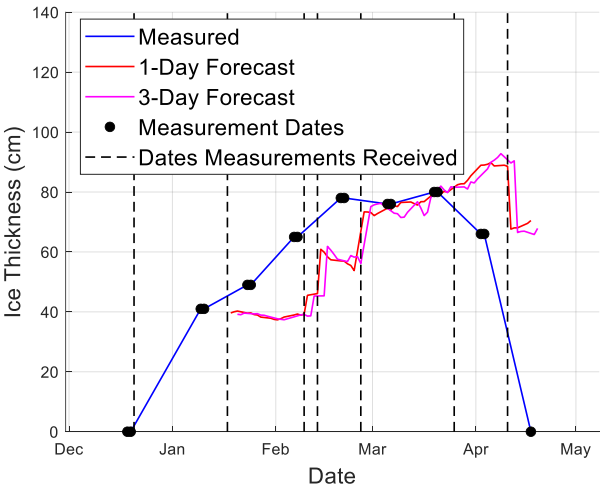

Figure 10. Measured and forecast ice thickness at Churchill River site N6 during 2023-2024.

Figure 11 shows the measured and forecast ice thicknesses for site N6 during the 2024-2025 season. Since the 2024-2025 season models predicted one- and three-day changes in ice thickness as opposed to a direct time-series of ice thickness as in the 2023-2024 season, the forecasting could begin immediately upon receipt of the first ice thickness report. The GBR models again captured the overall growth pattern of ice thickness through the season, with a small improvement in average accuracy of the one-day forecasts (two cm, see Table 3) over the 2023-2024 season LSTM models for this river site.

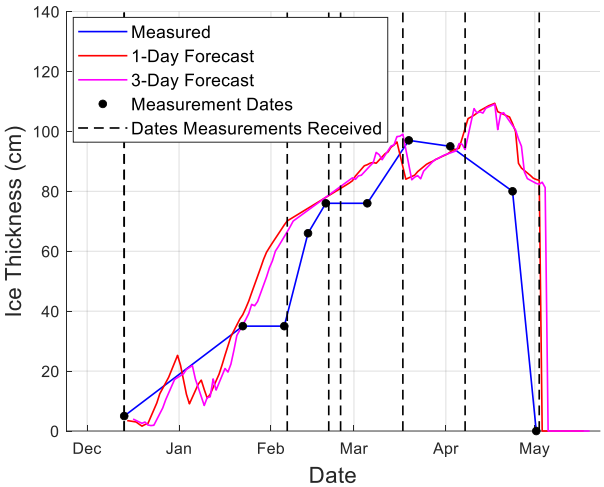

Figure 11. Measured and forecast ice thickness at Churchill River site N6 during 2024-2025.

4.2 Lake Melville Site N8

Figure 12 shows the measured and forecast ice thicknesses for site N8 on Lake Melville during the 2023-2024 season, and Figure 13 shows the measured and forecast ice thicknesses for site N8 during the 2024-2025 season. The GBR models in 2024-2025 showed moderate improvement in average accuracy (five cm, see Table 3) over the 2023-2024 LSTM models for this lake site.

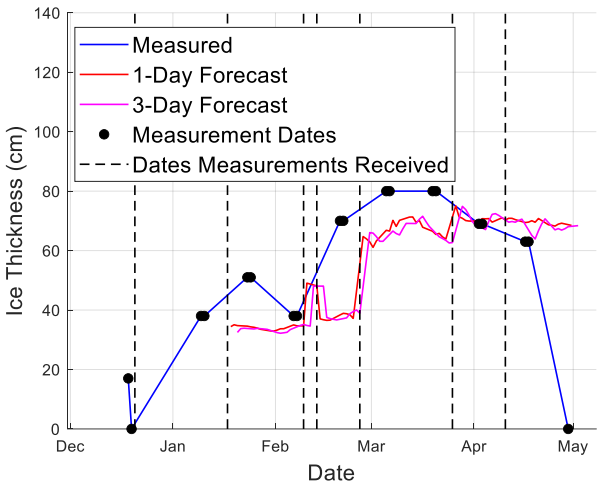

Figure 12. Measured and forecast ice thickness at Lake Melville site N8 during 2023-2024.

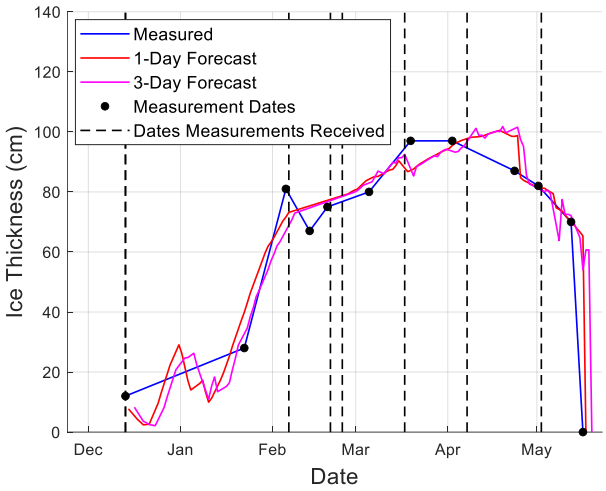

Figure 13. Measured and forecast ice thickness at Lake Melville site N8 during 2024-2025.

4.3 Seasonal Comparison

To evaluate the accuracy of the ice thickness prediction models, the Mean Absolute Error (MAE) performance metric was used. The MAE quantifies the differences between the observed and predicted values and is defined as follows (Hou et al., 2022):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - M_i|. \quad (6)$$

In Equation 6, n represents the total number of observations, $i = 1, 2, \ldots, n$ denotes the index of each sample, M_i is the measured (observed) value, and P_i is the corresponding predicted value. Smaller MAE values, especially those approaching zero, indicate more accurate model predictions. Table 3 shows the MAE for the one-day forecast minus measured ice thicknesses for each site on the measurement dates, for the 2023-2024 and 2024-2025 ice seasons.

Site	2023-2024 Forecast - Measured (cm)	2024-2025 Forecast - Measured (cm)
N5	17	24
Between N5 and N6	12	16
N6	14	12
Between N6 and N7	25	16
Mud Lake C/N7	20	20
Mud Lake A	14	13
N8	12	7
N9	12	19
Average	**16**	**16**

Table 3. Mean absolute average one-day forecast minus measured ice thickness errors for 2023-2024 vs. 2024-2025 ice seasons.

Half of the sites showed improved forecast accuracy for 2024-2025 compared to 2023-2024, with three of the sites showing worse performance and one site (Mud Lake C/N7) showing no change. The largest improvement of nine cm was at site Between N6 and N7, which had the worst performance in 2023-2024. It should be noted that two of the sites where average prediction error increased from 2023-2024 to 2024-2025, N5 and Between N5 and N6, had missed measurements, which may have

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-61-2026 | © Author(s) 2026. CC BY 4.0 License.

artificially inflated the apparent model prediction errors. There were also more predictions issued for the 2024-2025 season compared with 2023-2024 since forecasting could begin immediately after the first set of measurements in 2024-2025.

The varying model results between the eight sites can also be attributed to different hydrological dynamics across the sites. Specific river sites can have local channel geometry-induced eddies that affect heat mixing in the water column, as well as groundwater inflow that can introduce additional heat into the water column. These variables cannot be accounted for in the models and could significantly impact ice thickness growth and melt rates. Without such variables explicitly included as training features, the ML algorithms can implicitly account for them to only a limited extent.

5. Conclusions

This paper describes the development of the ML models and their operational use during the 2023-2025 ice seasons on the Churchill River and Lake Melville to predict ice thickness for input to the WRMD's river forecast model (CRFFS, 2025). The one and three-day forecast ice thicknesses were generally within 10-20 cm of the measured ice thicknesses at the eight sites when averaged over the ice seasons and interpolated to a daily timeseries. However, given the biweekly time between ice thickness measurements and the often days to weeks delay in receiving the WSP reports, many of the forecasts were predictions out many days beyond one to three days (up to 19 days in some cases).

The ice thickness prediction errors during 2024-2025 using the GBR models were comparable to the 2023-2024 season LSTM models in terms of absolute averages (MAE). The largest improvement of nine cm was at site Between N6 and N7, which had the worst performance in 2023-2024 (e.g., see Table 3). However, future model development will focus on improving the GBR models given their simpler architecture and training process compared with the deep learning LSTM models, and the fact that the GBR models have already been hybridized with physical thermodynamic models of ice growth and melt. The physical part of the GBR model used only 1D thermodynamic ice growth and melt and conductive heat flux algorithms typically used for lake ice. The ML part of the model was intended to bridge the gap between these physics and the more complex physics of river ice, which can include dynamic changes in thickness, and could not be explicitly modelled with physics in the present work.

Additional future work will focus on potentially improving model performance not only by using an expanded training set as more seasonal ice thickness data become available, but also by improving training feature selection. Since GBR models can account for complex interactions and non-additive effects, it is important to use methods like feature importance scores from the trained model (such as gain-based metrics from tree splits) or permutation importance to reflect nonlinear relationships and feature interactions. A key part of the next stage in developing the ice thickness prediction ML models will be to analyse and document the importance of input features. This step, which falls under the umbrella of Explainable Artificial Intelligence (XAI), can help improve model performance by identifying and removing redundant or uninformative features, reducing overfitting and enhancing generalization.

In addition to considering snow depth on the river ice (as was done in the current project), future model development will also consider the thickness of the water-saturated slush mixture which sits just above the ice surface. Separate models for the river (sites N5, Between N5 and N6, N6, Between N6 and N7, Mud Lake C/N7, and Mud Lake A) versus lake sites (N8 and N9) will be developed given the potentially different responses of ice to the dynamics of relatively faster flowing river water versus slower moving lake water. Rainfall will also be tested as an additional model feature in ice thickness prediction given its importance in melt processes. Finally, another ML model will be developed to correct for biases in GDPS forecast snow depth versus WSP measured snow depth on the ice, given the importance of snow depth in insulating the ice cover and modulating its growth and melt rates.

Acknowledgements

We thank the Water Resources Management Division (WRMD) of the provincial government of Newfoundland and Labrador (NL) and Transport Canada through the National Trade Corridors Fund (NTCF) for providing funding for this work to develop and deploy the ice thickness prediction models. We also thank and acknowledge WSP and Newfoundland Hydro for providing the ice thickness measurements that were critical to training, testing, and developing the models.

References

Awasthi, R., 2023: "Gradient Boosted Decision Tree – Clearly Explained," Medium (blog), https://medium.com/@ruchi.awasthi63/gradient-boosted-decision-tree-clearly-explained-bd1d8c7d9923.

Bhasme, P., Vagadiya, J., and Bhatia, U., 2021: "Enhancing Predictive Skills in Physically Consistent Way: Physics Informed Machine Learning for Hydrological Processes," https://doi.org/10.48550/arXiv.2104.11009.

Bilello, M.A., 1980: "Maximum Thickness and Subsequent Decay of Lake, River and Fast Sea Ice in Canada and Alaska," 80-6, Hanover, NH: Cold Regions Research and Engineering Laboratory (CRREL).

CRFFS, 2025: "Churchill River Flood Forecasting System," https://www.gov.nl.ca/ecc/files/Churchill-River-Flood-Forecasting-System-1.pdf.

ECCC, 2024: "Environment and Climate Change Canada – Climate Data," https://climate.weather.gc.ca/.

GDPS, 2025: "Global Deterministic Prediction System," https://eccc-msc.github.io/open-data/msc-data/nwp_gdps/readme_gdps_en/.

Hou, J., Wang, Y., Zhou, J., and Tian, Q., 2022: "Prediction of Hourly Air Temperature Based on CNN-LSTM," *Geomatics, Natural Hazards and Risk*, vol. 13, no. 1, pp. 1,962-1,986, https://doi.org/10.1080/19475705.2022.2102942.

Liston, D.E., and Hall, D.K., 1995: "An Energy-Balance Model of Lake-Ice Evolution," *Journal of Glaciology*, vol. 41, no. 138, pp. 373-82.

Macek, K., 2008: "Pareto Principle in Datamining: An above-Average Fencing Algorithm," *Acta Polytechnica*, vol. 48, pp. 55-59.

Muñoz Sabater, J., 2019: "ERA5-Land Hourly Data from 1950 to Present," Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.e2161bac.

Ozdemir, S., 2016: *Principles of Data Science*, 1st Edition, Birmingham, UK: Packt Publishing Ltd.

Pedregosa, F., and et al., 2011: "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2,825-2,830.

Prettenhofer, P., and Louppe, G., 2014: "Gradient Boosted Regression Trees," https://www.slideshare.net/PyData/gradient-boosted-regression-trees-in-scikit-learn-gilles-louppe.

Shertinsky, A., 2020: "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network," *Physica D*, vol. 404, no. 132306, https://doi.org/10.1016/j.physd.2019.132306.

Söylemez, A.O., 2024: "Why Simpler Machine Learning Models Can Outperform Deep Learning in Financial Forecasting," Medium (blog), https://medium.com/%40ariforcunsoylemez/why-simpler-machine-learning-models-can-outperform-deep-learning-in-financial-forecasting-302924f5b0e0.

TensorFlow., 2015: "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," available at: https://www.tensorflow.org/, accessed on: March 18, 2025.