AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2025WR040635

Special Collection:
Integrating In Situ, Remote Sensing, And Physically Based Modeling Approaches to Understand Global Freshwater Ice Dynamics

Long Short-Term Memory Model to Forecast River Ice Breakup Throughout Alaska USA

Russ Limber[1,2], **Forrest M. Hoffman**[2], **Jon Schwenk**[3], **and Jitendra Kumar**[2]

[1]The University of Tennessee, Knoxville, TN, USA, [2]Oak Ridge National Laboratory, Oak Ridge, TN, USA, [3]Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract The annual breakup of river ice in Arctic regions poses significant risk of ice jam flooding, causing property damage, altering ecosystems, and jeopardizing inhabitants. Predicting the timing of the annual breakup is crucial for residents to prepare for potential flooding and assess the safety of rivers for transportation. This analysis develops a deep learning algorithm using widely available meteorological and geospatial data products to forecast river ice breakup. We selected 33 locations along eight major rivers across Alaska, USA, and Western Canada, leveraging annual breakup dates from the Alaska-Pacific River Forecast Center database. Daily meteorological data from Daymet and static watershed attributes from the pan-Arctic catchment database were used to develop a Long Short-Term Memory (LSTM) model for predicting river ice breakup. Of the 33 locations, 23 were used for tuning, training and testing the LSTM. The model demonstrated high efficacy, predicting the annual breakup date with a mean absolute error (MAE) of 5.40 days, standard deviation of 4.03 days and mean absolute percentage error (MAPE) of 4.37%. The spatial generalizability of the LSTM was evaluated using the remaining 10 locations as holdouts, with most locations showing MAPE <8% over the entire time series (1980–2023). Additionally, we retrieved 51 long-range seasonal forecast ensembles from the Copernicus Climate Data Store and applied the trained model to them to showcase the capability of the LSTM to predict future river ice breakup using operational weather forecasts. LSTM was able to predict the breakup dates within 5–14 days of observed breakup.

Plain Language Summary We developed and applied a specialized deep learning model designed to forecast the timing of river ice breakup across eight major rivers in Alaska, USA, and Western Canada. The model uses commonly available meteorological data to accurately predict the breakup date within an average margin of error of 5.40 days. To test the model's practical application, we evaluated its performance using weather forecast data for the breakup season. The model maintained its accuracy, making accurate predictions even with inputs projected up to 5 months in advance. Notably, the model does not require specific location details to make predictions, demonstrating its reliability across 23 diverse locations and successful performance on nine out of 10 holdout locations (locations not used for model training). Results showcase the model's robust ability to generalize across both space and time. To determine the contribution individual variables had on the model's predictions, Shapley values were derived. Shapley values are a technique originating in game theory to assess variable contribution to a model per observation. A correction was applied to the Shapley values allowing them to be better suited to preserving information over time.

Key Points:

- The annual breakup date of river ice in Alaska for several major rivers can be modeled effectively using a LSTM with relatively few meteorological inputs
- The LSTM can generalize over space and is effective for future predictions using long-range forecast data
- Timely predictions of breakup dates can help residents prepare for potential ice jam floods and safely use rivers for transportation

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

R. Limber,
limberrl@ornl.gov

Citation:

Limber, R., Hoffman, F. M., Schwenk, J., & Kumar, J. (2025). Long short-term memory model to forecast river ice breakup throughout Alaska USA. *Water Resources Research*, *61*, e2025WR040635. https://doi.org/10.1029/2025WR040635

Received 27 MAR 2025
Accepted 2 SEP 2025

Author Contributions:

Conceptualization: Russ Limber, Jon Schwenk, Jitendra Kumar
Data curation: Russ Limber
Formal analysis: Russ Limber
Funding acquisition: Forrest M. Hoffman, Jitendra Kumar
Investigation: Russ Limber, Jitendra Kumar

1. Introduction

In high latitude cold regions of North America, the annual breakup of river ice leads to significant ecological impacts (Beltaos, 1995; Church, 1974; Mackay & MacKay, 1973; Prowse & Gridley, 1993) that take on a wide variety of forms from flooding, disturbance of sediments and erosion (Beltaos et al., 1994; Chambers, 1994; Gray & Prowse, 1993) to alterations in the gross primary productivity of ecosystems (Thellman et al., 2021). Beyond its ecological consequences, the thawing of river ice significantly impacts the residents of high latitude regions. Throughout the winter season, frozen rivers serve as vital transportation routes facilitating the delivery of essential goods to remote communities (Ian et al., 2016). Escalating global temperatures pose potential challenges for inhabitants of high-latitude communities, as the safe travel window afforded by river ice coverage has diminished over the past century (Brown et al., 2018). Perhaps the most significant impact river ice breakups have on the inhabitants of high latitude regions is that they lead to ice jam flooding (IJF). IJF has the potential to occur when broken ice shards from the once frozen river accumulate, leading to a damming effect (Beltaos et al., 1994).

Methodology: Russ Limber, Jitendra Kumar
Resources: Jitendra Kumar
Software: Russ Limber
Supervision: Forrest M. Hoffman, Jitendra Kumar
Validation: Russ Limber
Visualization: Russ Limber
Writing – original draft: Russ Limber, Jitendra Kumar
Writing – review & editing: Russ Limber, Jon Schwenk, Jitendra Kumar

In 2017 alone, it was estimated that across North America IJF contributed to ~$300 million in damage and losses (French, 2018).

Predicting the timing and location of river ice breakup remains an important and challenging problem. A number of meteorological and hydrological factors influence and drive the timing of river ice breakup. Air temperature, precipitation and snow accumulation over time play a critical role in regulating river ice processes at diurnal and seasonal scales (Lindenschmidt, 2020), including formation of river ice at the beginning of the ice season in early winter, growth and thickness of the ice during winter months and timing of breakup in spring (Ashton, 1986; Lindenschmidt, 2020). Mechanisms initiating breakup of river ice can either be thermal (due to deterioration of ice cover through warming without any significant increase in discharge) or mechanical (driven by increased flow without requiring deterioration of ice cover). The distinction between thermally and mechanically driven breakup events is simple to define but challenging to identify in application. Often, breakup events possess aspects from both mechanical and thermal breakups making the classifications difficult to disentangle. Thermal breakup events driven by warming temperatures and solar radiation occur consistently close to the spring season. Mechanical breakup tends to be more variable in space and time. In the regions with moderately cold winter, mechanically driven mid-winter breakup events can happen, often triggered by above freezing temperatures and rain-on-snow (Beltaos, 2002, 2003). The lack of consistency and abruptness of mid-winter breakup events makes them often more dangerous than breakups that occur during the spring thaw and more challenging to predict (De Coste et al., 2022).

Our study focuses primarily on thermal river ice breakup processes occurring during the spring thaw. While some of the rivers included in this study have experienced mid-winter breakups, analysis and modeling of mid-winter breakup events were not included in the current study (Newton et al., 2017).

Past studies have explored numerical modeling as well as machine learning approaches for predicting river ice breakup and formation. Numerical modeling approaches leverage physical processes that drive the formation and decay of river ice. For example, using conservation of energy to approximate the heat flux between the surface of rivers and the ambient temperature have been conducted extensively (Ashton, 1986; Hicks et al., 2008; Paily et al., 1974). Shen (2010) expanded on previous works (Shen, 2003; Shen & Chiang, 1984) to create a comprehensive numerical model, outlining specific ice formation possibilities based on meteorological and hydrological conditions. Their model took into consideration different breakup scenarios such as dynamic versus static breakup events. Due to the high degree of danger that comes with IJF, some organizations representing communities at high latitudes have designed their own models paired with impact assessment plans. For example, the Water Resources Section of the Department of Indian and Northern Affairs in Canada designed its own river ice breakup and flood monitoring model using a combination of heuristics and statistical analysis (Jasek, 1999). The model has been expanded upon since 1976 but was scrutinized for being unable to anticipate the severity of a hundred year flood that occurred in 1979 in Dawson City Canada that left much of the city submerged under 1–2 m of water (Jasek, 1999). The proposed model predicts river ice breakup as well as the likelihood and severity of ice jam flooding but requires regular site assessments (INAC, 2008). Using the same or similar breakup records to the ones used in our analysis, studies of the Tanana River in Alaska from 1917 to 2000 (Sagarin & Micheli, 2001) and the Yukon River in Canada from 1896 to 1998 (Jasek, 1999) found that river ice breakup seems to be occurring 5 days earlier between centuries for these locations (Prowse et al., 2007).

The complexity and nonlinearity of river ice breakup renders deep learning approaches potentially useful for modeling breakup events. A number of statistical and machine learning frameworks have been explored by past studies to address the challenge of river ice breakup modeling. Madaeni et al. (2022) developed a convolutional Long Short-Term Memory model that proved highly effective at predicting the occurrence of IJF on an annual basis across Quebec Canada. However, this model was specific to IJF rather than the breakup date of river ice. A fully connected neural network (FCNN) was implemented to model freeze up and breakup dates along the Yellow River in Mongolia (Tao et al., 2008). River ice breakup along the Yellow River in China was modeled using a FCNN paired with backpropogation and particle swarm optimization (Hu et al., 2008). Zhou et al. (2009) developed a support vector machine optimized using a multiobjective shuffled complex evolution metropolis algorithm, which was also used to predict breakup dates on the Yellow River in Mongolia. In North America, Sun and Trevor (2018) found that a Bayesian regularization back-propagation FCNN was effective at predicting the breakup dates for a single location near Fort McMurray on the Athabasca River in Alberta Canada.

Considering all machine learning frameworks, Long Short-Term Memory models (LSTMs) have been proven to be effective for modeling time series data as they incorporate contextual information from past inputs to learn complex non-linear patterns, while maintaining automatic feature extraction (Bouktif et al., 2018). LSTMs are also well equipped at parsing sequential patterns to recognize long-term trends while simultaneously weighing more heavily toward recent observations (Yu et al., 2019), making them an excellent choice for river ice breakup modeling given the nature of the drivers leading to the spring thaw.

While there has been considerable efforts to forecast river ice breakup, past analyses contain recurring shortcomings that this analysis sets out to address. Many of the previous studies were performed for only one river or even one location along a river. Our primary contribution is a predictive, accurate model, that is generalizable across different rivers. With different rivers and locations having a unique set of geomorphological parameters, it is highly unlikely that a model trained on one river or a single location can be generalized to rivers in various environments. Generalizability of a single site is even more unlikely when time-varying hydrologic variables are used as predictors.

The objective of this analysis was to (a) develop a time series deep learning model that effectively predicts river ice breakup timing on a daily temporal frequency and (b) adds to our understanding of the relationship between controlling variables. To do this, we implement a LSTM to forecast river ice breakup events and analyze variable contribution using temporally adjusted Shapley values.

2. Materials and Methods

Our analysis encompasses 33 locations, spread out over a far larger geographical region than previous works, along eight major Alaskan rivers. Additionally, our analysis only uses meteorological data as time-varying inputs and does not include any hydrologic or hydraulic time-varying inputs since such data is sparse and often does not exist for all possible sites, thus limiting the applicability to a developed model. In addition to generalizability over space, the LSTM in this analysis is better suited to generalize over time. This is because recurrent neural networks, in particular LSTMs, are well suited to capture temporal trends. LSTMs provide more weight to observations that occur closer to the breakup event, while information from long-term seasonal patterns is still conserved (Yu et al., 2019). The implementation of a neural network utilizing a lookback window to assign breakup likelihood on a per day basis is advantageous for a warning system, and is effective for finding temporal patterns within inputs that dictate the timing of breakup events. Since long-term breakup timing will likely be impacted by changing river ice regimes (Brown et al., 2018), leveraging underlying seasonality in the input data is necessary to project breakup events into the future.

To define the contribution of features input into the LSTM, we compute Shapley values to assess variable significance. Shapley values are a metric based on variable contribution to define the influence of a variable on a given cost function (or similar dependent variable) on an observation by observation basis. The principle originates from game theory (Shapley, 1967). A temporal adjustment to this metric was made to account for time series analysis during the Monte Carlo simulation step to approximate marginal contribution. In our adjustment, only time steps that correspond to one another seasonally are used to compute the permutations leveraged to find marginal contribution. The results of the Shapley analysis reveals the capability of the LSTM to recognize key moments in the annual timeline of inputs that drive river ice breakup.

2.1. Study Area and Breakup Dates

Observations for annual river ice breakup dates were collected from the Alaska-Pacific River Forecast Center database (APRFC; (APRFC, 2024)) for 394 locations across Alaska USA and extending into Western Canada (Figure 1). The study region is sparsely inhabited with a highly diverse topographic landscape and located between 60.80° N to 67.15° N latitude and −164.56° W to −139.35° W longitude. Locations were selected based on the total number of breakup event observations available within the period 1980 to 2023, which corresponds to the time span of Daymet meteorological data availability at the time of this writing. While breakup events can occur more than once per year in nature, for the recorded APRFC observations breakups were considered to be an annual event corresponding to the date at which a visible break in the river ice can be seen from one bank of the river to the other–typically as a result of spring thaw. This is also referred to as a "transverse break" in the ice. While there have been.

Figure 1. 33 selected observations locations were distributed across Alaska USA and Western Canada across eight rivers (Copper, Koyukuk, Kuskokwim, Susitna, Tanana, Yukon, Buckland, and Kobuk). The purple colored points represent locations used to train, validate and test the LSTM, while light blue colored points are locations where the data was held out for additional evaluation and test of transferability of the LSTM.

Most locations within the APRFC database had gaps in observations with years missing records of breakup dates–many containing only one to five breakup records. We defined and selected a threshold to be a minimum 35 years of available annual records during our study period 1980 to 2023 (80% of the years during the study period). The threshold was selected to ensure that there were sufficient number of observations to train our machine learning models while still capturing spatial variability across the study region. This reduced the number of locations from 394 in the APRFC database to 33 locations that met the threshold criteria (Figure 2a).

These 33 locations were spatially distributed across Alaska and Western Canada across eight major rivers: Copper, Koyukuk, Kuskokwim, Susitna, Tanana, Yukon, Buckland and Kobuk Rivers. Kuskokwim and Yukon were best sampled with 10 and 14 locations respectively. Kobuk and Koyukuk had 2 and 3 locations respectively, while Copper, Susitna, Tanana and Buckland each had one location. Timing of annual breakup shows large inter-annual variability within and across the eight rivers and cyclical patterns of breakup dates during period of record as shown in Figure 3. These cyclic patterns seen across different locations and rivers are likely attributed to Pacific decadal oscillation, which determines temperature shifts in the Northern Pacific Ocean. This has been shown to affect Arctic freshwater processes including river ice breakup timing (Pavelsky & Smith, 2004). While this work does not focus on the analysis of trends in annual river ice breakup, previous studies have investigated and reported on such trends (Arp et al., 2020; Brown et al., 2018; Wang & Feng, 2024).

2.2. Meteorological Data

The meteorological data for our study was obtained from Daymet (Thornton et al., 2022), which provides estimates of daily weather parameters. Daymet provides daily surfaces of minimum and maximum temperature (°C), precipitation (mm), vapor pressure (Pa), shortwave radiation $\left(\frac{\text{watts}}{\text{m}^2}\right)$, snow water equivalent $\left(\frac{\text{kg}}{\text{m}^2}\right)$, and day length

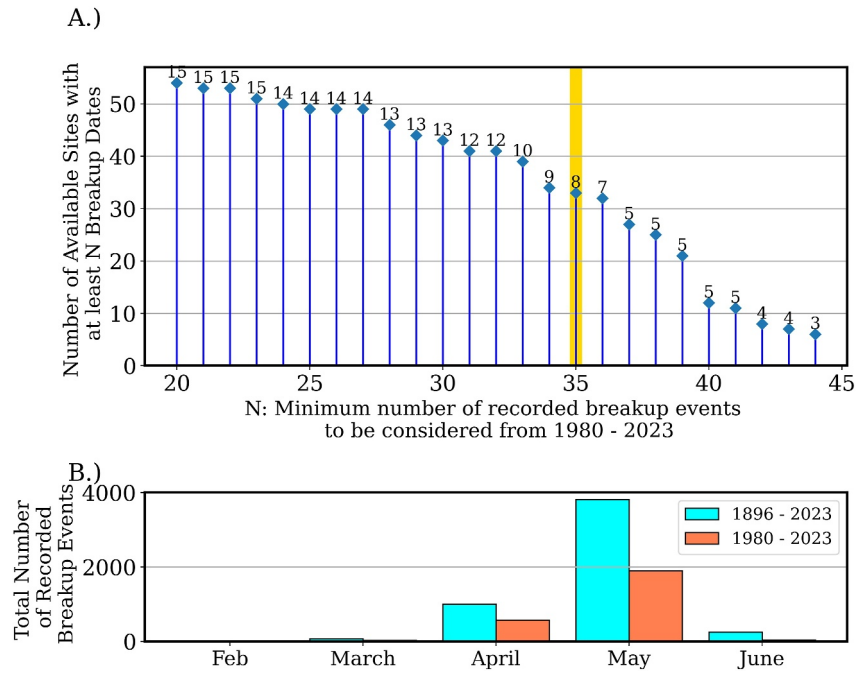

Figure 2. (a) Stem plot showing the number of available sites with *N* breakup dates from the APRFC database (vertical axis), where *N* is the threshold that is, the minimum number of years of available data (horizontal axis). The number above each stem plot shows the numbers of distinct major rivers that are available given the selected threshold and sites. The gold highlighted line shows the threshold ($N = 35$) selected in the analysis. (b) Bar plot of the total number of breakup events recorded by month in the APRFC database from 1896 to 2023 (blue) and the total number of breakup events recorded by month between 1980 and 2023 (orange).

(seconds) on a 1 km × 1 km gridded surface. Locations of the selected 33 sites were derived from the APRFC database, and the meteorological time series at those location for the period 1980 to 2023 was downloaded from Oak Ridge National Laboratory (ORNL) Distributed Archive and Analysis Center (DAAC) using their Single Pixel Extraction Tool.

2.3. Pan-Arctic Watershed Properties Data

In addition to the meteorological conditions, characteritics of the local watershed for the breakup observaton site plays an influential role in regulating the hydrological processes that may have an impact on river ice dynamics. We leverage the pan-Arctic catchment database (ARCADE; (Speetjens, 2022; Speetjens et al., 2023)) that include major drainage basins across the pan-Arctic with a Strahler order of 5 or greater (Strahler, 1957). The ARCADE data set is available as a multipolygon shapefile with a course spatial resolution (Figure 4) and includes 103 geospatial, environmental, climatic, and physiographic catchment properties. We selected five statics variables from ARCADE characterizing the watersheds around our breakup sites: (a) fraction of watershed area covered by surface water, (b) mean watershed topographic wetness index, (c) catchment perimeter (km), (d) gravelius index, (e) mean catchment slope. These variables were selected to capture the information about size and shape of the watershed, topographic conditions and surface water conditions.

2.4. Long-Range Seasonal Forecast Data

Ensemble of long-range seasonal forecasts were obtained from Copernicus Climate Change Service (C3S) Climate Data Store (CDS) at 1° × 1° spatial resolution. The C3S provides a multi-system seasonal forecast service, where data produced by state-of-the-art seasonal forecast systems developed, implemented and operated at forecast centers in several European countries is collected, processed and combined to support user-relevant applications. We used 51 members ensembles of "Seasonal forecast daily and subdaily data on single levels" product from the European Center for Medium-Range Weather Forecasts (ECMWF) provided by CDS (C3S, 2018). Forecast products are released once per month. Seasonal forecasts are provided in the form of

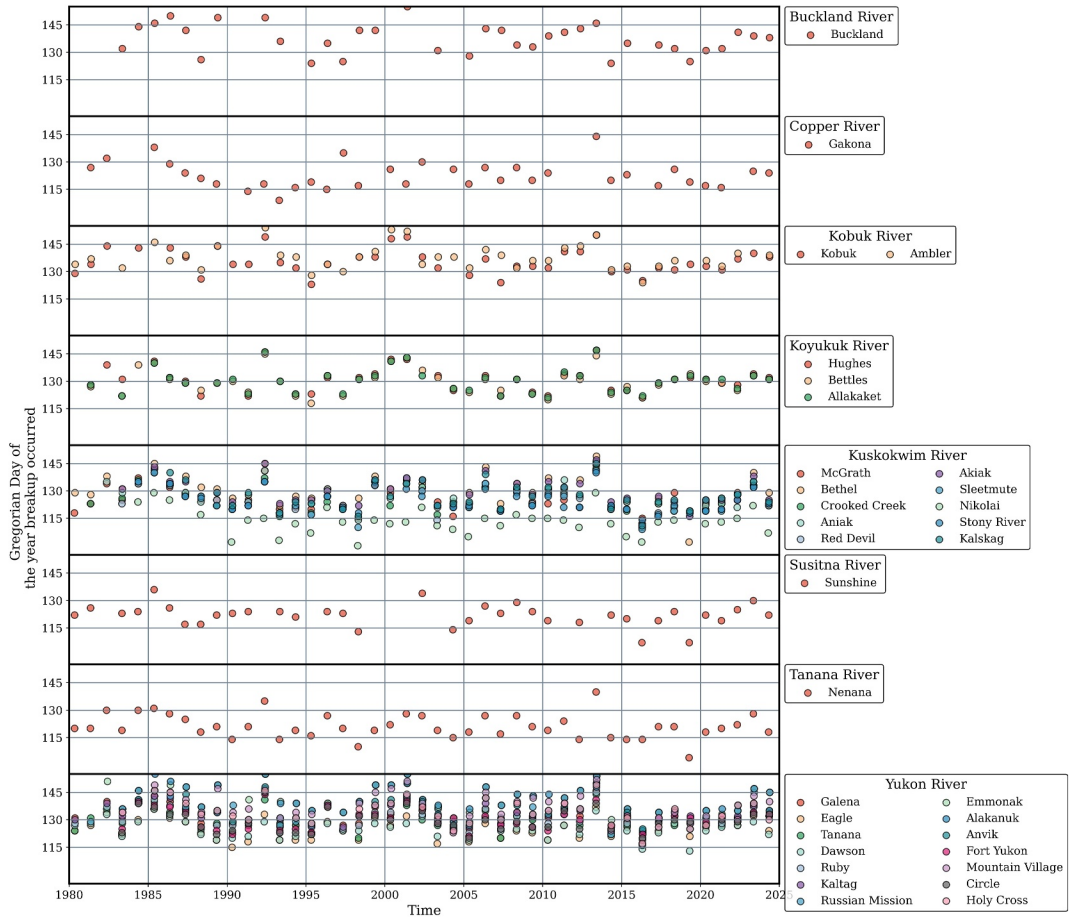

Figure 3. The Gregorian day of the year when the breakup occurred for each location separated by river. Discontinuity in the time series represents years where gaps exist in the APRFC record of breakup events.

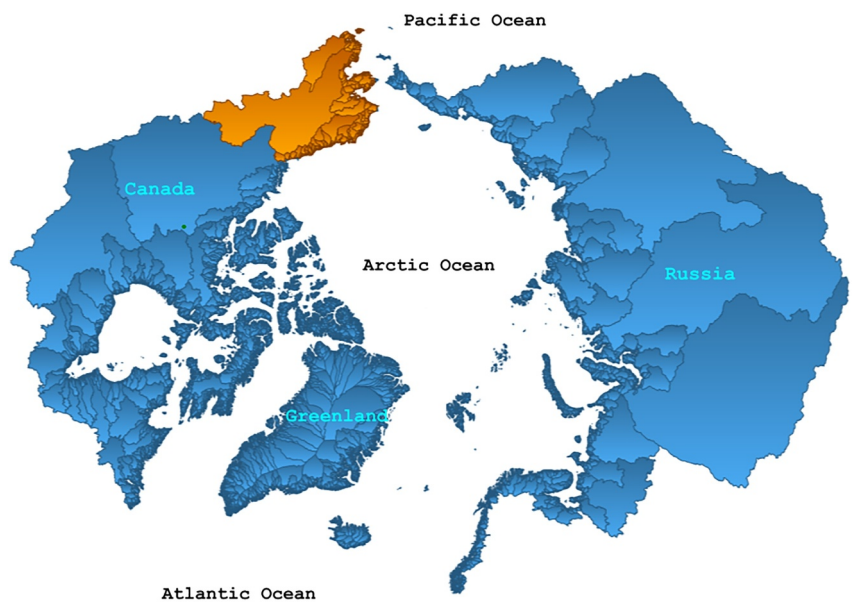

Figure 4. Draingage basins in ARCADE data set (native coordinate reference system (EPSG:6931)). Orange color highlights the study region of interest for this analysis.

ensembles in an effort to account for changes in parameterization within the ECMWF framework. While more computationally rigorous, leveraging forecast ensembles help account for uncertainty within the future predictions which can be high with long-range projections (Rabier, 2023). Each member was collected starting on the first of 5 months: February, March, April, May, June and extending to the first of July. This created five forecast ranges with lead times of 5, 4, 3, 2 and 1 month(s) respectively. Forecast for breakup date were continually updated each month using newly available forecasts products. These months were chosen to conservatively represent the annual breakup season of our region of interest as shown in Figure 2b. While the forecast data product does not include vapor pressure, we followed the vapor pressure estimation method used by Daymet, which includes Murray's adjustment (Murray, 1967) to Teten's equation for temperatures below freezing. Equation 1 descibes the calculation of vapor pressure (vp) in Pascals where T is the daily minimum temperature in °C used as a proxy for minimum dew point temperature, consistent with Daymet's estimation method (Thornton et al., 2022). This forecast data set was processed for the years 2020–2023.

$$vp = 610.78e^{\left(\frac{21.875T}{T+265.5}\right)} \tag{1}$$

2.5. Data Processing

We implemented a number of data processing steps to prepare the data for tuning and training the LSTM.

Gregorian day of the year (DOY_i), which is a discrete variable, was transformed to a continuous temporal variable to be included as input features to the LSTM. This removed discontinuity during changes from one year to the next. Days of the year were normalized to be between 0° and 360° (Equation 2), converted to radians, and the cos and sin of said radians (Equation 3) was calculated to create two temporal variables representing the day. This allows for the LSTM to recognize the day of the year without creating discontinuity in the features. All dates used in the analysis were standardized to a no leap year calendar to align with the no leap year calendar used by Daymet, which included dropping the last day of the year (31 December) in the leap years.

$$h(DOY_i) = (DOY_i - 1)\left(\frac{360}{364}\right)\left(\frac{\pi}{180}\right) \tag{2}$$

$$\left[\text{Temporal}_x,\ \ \text{Temporal}_y\right] = \begin{cases} \cos\left(h(DOY_i)\right) \\ \sin\left(h(DOY_i)\right) \end{cases} \tag{3}$$

The input variables ($\mathbf{X}$) for the LSTM were processed as a daily time series consisting of the Gregorian day of the year (represented as sin/cos transforms, Equations 2 and 3), seven Daymet variables (time varying), and five ARCADE variables (static).

Breakup dates (derived from APRFC) were the target variable for the LSTM, and were cast as a sparse one dimension binary vector with 1 representing the annual breakup event and 0 representing a non-breakup day. Each element of this vector, which from here onward will be referred to as dependent variable $\mathbf{y}$, maps to the daily timestep over the time series. The labels of $\mathbf{y}$ are highly unbalanced as the expected frequency of 1 to 0 is 1 : 364 since there is only one annual breakup event per year.

Of the total 33 observation locations from APRFC used in this study, 23 were used to create the *main data set* (for model training, validation, and testing). Standard schemes for splitting the data among training, validation and testing sets ensures appropriate evaluation of model performance and prevents overfitting, however, does not provide any quantification of model generalizability and transferability to completely new locations. Thus, the remaining 10 locations were used to create a *holdout data set* for independent testing of the trained model for transferability. Holdout locations were selected to have wide geographical variability and were dispersed across different rivers (Figure 1). Holdout locations were selected randomly along each river, with the constraint that no two holdout locations were adjacent to one another. This spatial separation was implemented to capture maximum along-river variability for our holdout evaluation. This was done to test the skill of the LSTM to generalize and transfer to new locations not seen in the main data set. The *long-range forecast data set* from C3S CDS was processed in a consistent fashion for application of the trained model.

A LSTM model is a sequential neural network, therefore the data was concatenated sequentially with each day sorted by year, then sorted by location, Equation 4:

$$\mathbf{X}[x_i|\text{location}_1]_{t=1980},\ \mathbf{X}[x_i|\text{location}_2]_{t=1980},\ \ldots,\ \mathbf{X}[x_i|\text{location}_1]_{t=1981} \tag{4}$$

where $\mathbf{X_t}$ is an array representing all daily observations within a given year t, for each location.

A temporal window was then applied to the concatenated data set to create a three dimensional array with dimensions: [number of observations, length of lookback window, features]. The lookback window was set to a relatively large 457 days. This was done to combat the sparsity of vector $\mathbf{y}$ to ensure that a breakup event was present within every lookback window. The length of the lookback window was determined as 365 days in a year plus the length of the river ice breakup season (based on historical records) plus 1 day. Thus, if there were adjacent breakup seasons with one breakup happening early in the season and the other with a breakup happening at the end of the season, the lookback window would still include at least a single breakup event. Since none of the locations in our filtered data set have a breakup event that occurs in February or March (Figure 2b) we selected April, May, and June as our temporal buffer. The number of days in April, May, and June are 30, 31, 30 days respectively thus the lookback window can be derived as: $365 + 30 + 31 + 30 + 1 = 457$ days.

There are no missing values in the Daymet data set, however, the breakup events provided by the APRFC database has years with missing breakup records. Each location has at least 35 years of annual breakup dates recorded out of 43 years of possible breakup dates (number of years from 1980 to 2023). A mask was created to conceal observations (applied to $\mathbf{X}$ and $\mathbf{y}$) within the breakup season months (April, May, and June) for each missing year. This allows information from the rest of the year to still be utilized by the LSTM to make predictions for subsequent years. All data sets used in this analysis (i.e., main, holdout, long-range seasonal forecast) were prepared follwing the scheme in Equation 4.

The main data set was partitioned into training, validation and testing data sets using a proportion of 0.6, 0.2 and 0.2 respectively. These data sets from here onward are referred to as *training*, *validation* and *testing* data sets. Data sets were normalized between 0 and 1 using min-max scaling on a by variable basis (Equation 5).

$$\mathbf{X}_{\text{normed}} = \frac{\mathbf{X} - \mathbf{X}_{\text{min}}}{\mathbf{X}_{\text{max}} - \mathbf{X}_{\text{min}}} \tag{5}$$

where $\mathbf{X}$ is an array corresponding to the input data; while $\mathbf{X}_{\text{min}}$ and $\mathbf{X}_{\text{max}}$ are conformable vectors representing minimum and maximum statistics over the features, respectively.

2.6. Long Short-Term Memory Model

A LSTM is a type of recurrent neural network, that leverages sequential autocorrelation within input observations. The LSTM architecture (Figure 5) was first proposed in 1997 as a means for solving the vanishing/exploding gradient problem (Hochreiter & Schmidhuber, 1997). The unique arrangement and function of gates within each LSTM neuron (i.e., forget, input and output gates) allows weights and biases to be updated relative to how well inputs can predict the outputs, while also accounting for the sequential order of observations. The input variables are organized using a lookback window (in this analysis that lookback window was 457 days) with each sequential observation within that lookback window read into the LSTM neuron. Observations occurring further back in time relative to the iteration being assessed at dependent variable y_i, tend to have less impact on altering weights and biases than observations occurring closer in the sequence to iteration i. However, memory is still conserved in the system, allowing for long-term patterns to still be maintained. This is a central reason why LSTMs have been so effective at predicting temporal events (Yu et al., 2019).

Due to the unbalanced nature of $\mathbf{y}$ an initial bias was calculated and applied to the input layer using the validation portion of the main data set, Equation 6:

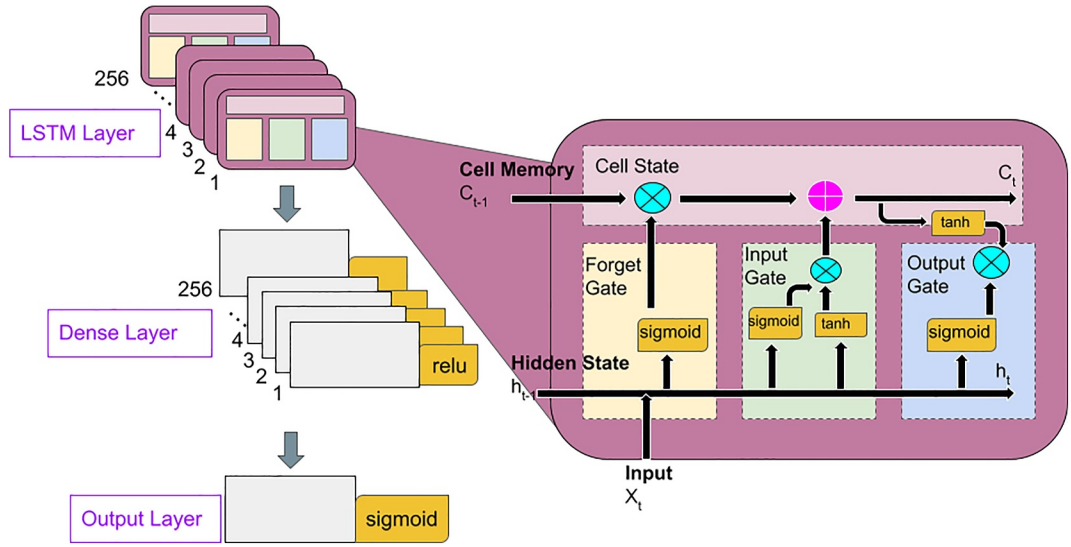

Figure 5. Schematic of the architecture of the LSTM used in this analysis. The dark pink cells represent the 256 LSTM neurons comprising the first layer. A zoom in of the LSTM neuron to the right, shows the steps taken as inputs are read into the gates and cell state of each neuron. Blue ⊗ represent multiplication of arrays while light pink ⊕ represents summation. Below in gray are the dense neurons which make up a single layer of 256 neurons. Orange tabs accompanying each cell provide the activation function used.

$$\mathrm{P(Breakup)} = \frac{\mathrm{TP}}{\mathrm{TN}+\mathrm{TP}} = \frac{1}{1+e^{-\mathrm{Bias}_0}}$$
$$\mathrm{Bias}_0 = -\ln\left(\frac{1}{\mathrm{P(Breakup)}-1}\right) = \ln\left(\frac{\mathrm{TP}}{\mathrm{TN}}\right) \quad (6)$$

where TP and TN are the number of true positives and true negatives within the validation data set respectively; Bias_0 is equal to the initial bias of the unbalanced classification space; and P(Breakup) is the expected probability of a river ice breakup event occurring on an annual basis.

Hyperparameter tuning and model training was conducted using the training and validation data sets and was carried out on the Perlmutter supercomputer at the National Energy Research Center (NERSC). The model was tuned using Bayesian selection, an extension of the KerasTuner API (O'Malley et al., 2019). This method provides a more optimal approach for selecting hyperparameters compared to a naïve random grid search. This is because Bayesian selection calculates the upcoming decision from a surrogate model or probabilistic model that uses Bayes Theorem to select hyperparameters that minimize the loss function of the LSTM (Dewancker et al., 2016). The area under the receiver operating characteristic curve metric was used to assess the epochs within each tuning trial for the validation data set. There were 75 tuning trials conducted each with as many as 50 epochs, however with an early stopping patience threshold of 10 epochs none of the trials ever actually reached 50 epochs. After a model architecture was selected using this framework, the model was trained until a local minimum was identified within the binary cross-entropy loss function, using the Adam optimizer (Kingma & Ba, 2017), with a patience of 20 epochs. The final model architecture as well as the parameters tuned using Bayesian selection are shown in Table 1.

Whereas past models designed to predict river ice breakup used the Gregorian day of the year as the dependent variable (Hu et al., 2008; Sun & Trevor, 2018; Tao et al., 2008; Zhou et al., 2009), in this analysis **y** is encoded as a binary vector acting as a boolean mask for the true breakup dates over the time axis. Using a large lookback window, daily information from the entire year can be utilized by treating the problem as a sparse classification problem rather than a regression problem. Using binary cross-entropy as the loss function, the outputs form a likelihood function with likelihood assigned to each day of the year. These likelihood curves are then overlaid with the time axis where maximum likelihood estimation is applied to calculate the annual breakup dates. Figure

Table 1
The Bayesian Hyperparameter Optimization Applied to the LSTM

Hyperparameter	Range of values	Step	Final selection
Number of LSTM Layers	1–3	1	1
Number of LSTM Neurons	32–256	32	256
Recurrent Dropout	0–0.6	0.05	0.0
Dropout	0–0.6	0.05	0.0
Number of Dense Layers	1–5	1	1
Number of Dense Neurons	32–256	32	256
Dense Neurons Activation Function	{elu, relu, tanh, leaky-relu}	N/A	relu
Learning Rate	0.001–0.1	0.01	0.1
Decay Rate	1e−6–1e−4	0.01	1e−6
Batch Size	64–1344	128	1216[a]

Note. For each hyperparameter assessed using the validation data set, the range of values that could have been selected for that hyperparameter, the step size for each trial and the final hyperparameter selected for the LSTM, are shown. [a]partitioned over 4 NVIDIA A100 40GB GPUs.

S1 in Supporting Information S1 shows how the likelihood function produced by the LSTM forms distinct peaks which are used to determine the predicted breakup date. Predictions were made annually using the testing data set. After the prediction was made for a given year, the validation data shifted forward 1 year to include observed meteorological time series and breakup date and the LSTM was trained (starting at the previous configuration) for an additional 25 epochs, with a patience of seven (the LSTM never reached the full 25 epochs). This iterative evaluation was carried out for each year of the testing data set to update the LSTM without introducing information leakage, while allowing meteorological inputs from the past year to inform estimates made on the current year. The bias in Equation 6 was applied at each iteration to the initialization of both the LSTM layer and the output layer of the updated model.

2.7. Naïve Model

To benchmark the forecasting capability of the LSTM, a naïve autoregressive model was implemented. The naïve model uses the date of the previous year's breakup to predict the current breakup date, $\hat{y}_{t|\text{location}}$ (Equation 7). However, when a breakup date is unavailable for the previous year, the naïve model is unable to make a prediction of the current year's breakup date.

$$\hat{y}_{t|\text{location}} = y_{t-1|\text{location}} \tag{7}$$

To maintain generalizability over space, the LSTM was constrained to only use meteorological inputs and watershed basin characteristics. The LSTM does not explicitly model on a location by location basis nor does it use past breakup dates in the input matrix. However, the naïve model utilizes this information exclusively making the comparison between the LSTM and the naïve model more compelling and revealing than conventional naïve (sometimes called null) model evaluations.

2.8. Shapley Values for Model Explainability

We computed Shapley values to add explainability to the LSTM model and gain insights into the role of input variables in breakup prediction. They provide interpretability similar to feature importance commonly associated with decision tree based machine learning models. A Shapley value is the average marginal contribution of a variable to predictions of a model f (Sterbak, 2022), as derived by Lloyd S. Shapley in 1967 (Shapley, 1967), across all combinations of feature variables. Since the number of variable combinations becomes prohibitively large across all test observations, approximation methods such as Equation 8 have been derived (Štrumbelj & Kononenko, 2013):

$$\phi_j = \frac{1}{M}\sum_{i=1}^{M} \hat{f}\left(u_i\left(j \in \mathbf{Q}_z\right)\right) - \hat{f}\left(u_i\left(j \notin \mathbf{Q}_z\right)\right) \tag{8}$$

Where ϕ_j is the Shapley value for variable j; M is the number of Monte Carlo simulations; u_i is the test observation that the Shapley value is computed for under the stochastic conditions at iteration i whereby $\mathbf{Q}$ is the set of features being created from a random mixing with training observation z as shown in Equations 9 and 10.

$$\mathbf{h} \subseteq \mathbf{n}_z \quad \text{where } ||\mathbf{h}|| = ||\mathbf{n}_z||p \text{ and } \forall\, \mathbf{h} \in \mathbf{n}_z \tag{9}$$

$$\mathbf{Q}_z = \{\mathbf{h}, (\mathbf{n}_u \notin \mathbf{h})\} \tag{10}$$

Where $\mathbf{n}_u$ and $\mathbf{n}_z$ are sets of all variables for observations u_i and z_i respectively; $\mathbf{h}$ is the set of variables after a random selection of $||\mathbf{n}_z||p$ variables without replacement, assuming p is a majority proportion (in this case 0.8).

Conventional approximation methods select z_i randomly from the training data (Štrumbelj & Kononenko, 2013). Here we introduce a temporal correction that selects z_i randomly but ensures that z_i is always on the same day of the year as u_i. This way when variables are rearranged (Equations 9 and 10) the comparison at observation u_i, as well as the lookback window preceeding it, align on regions of the time series that have a similar seasonality. This Shapley value analysis was carried out using $p = 0.8$ and $M = 500$ for each location.

3. Results

3.1. LSTM Model Trained Using APRFC Observations

We developed a LSTM trained using optimal hyperparameters (Table 1) identified through Bayesian hyperparameter optimization. The trained model was applied to the testing data set spanning 2014 to 2023 and predicted the river ice breakup dates with an overall mean absolute error (MAE) of 5.40 days, standard deviation of the absolute model residuals (STD) equal to 4.03 days and a mean absolute percentage error (MAPE) of 4.37%, across the 23 locations. The accuracy of prediction vary by location and are shown in Table 2.

The breakup timing of Allakaket on the Koyukuk River was captured the most accurately of any location in the testing data set with the lowest MAE by location ($\text{MAE}_{\text{location}}$) equal to 2.70 days, a $\text{STD}_{\text{location}}$ of 2.36 days and a $\text{MAPE}_{\text{location}}$ of 2.15%. The LSTM had the highest error of any location within the testing data set when predicting the breakup timing at Nenana on the Tanana River with a $\text{MAE}_{\text{location}}$ of 9.70 days, a $\text{STD}_{\text{location}}$ of 6.06 days and a $\text{MAPE}_{\text{location}}$ of 8.49%. Of the 23 locations used in the testing data set, 19 had the annual breakup of river ice predicted within less than a week (7 days) on average. The boxplots shown in Figure 6 suggest that the LSTM tends to overpredict, that is, predict that the breakup will occur later than it actually does, more than it under predicts (more negative residuals than positive residuals). Of the 225 predictions made, 165 were within 1 week of the actual breakup event (73%) and only 4 predictions were outside of 2 weeks (1.8%). Figure 6 shows that the accuracy and precision of the LSTM vary over the years. For example, 2016 was consistently overpredicted (all residuals are negative) with a relatively higher MAE compared to the other years, while 2017 consistently predicted close to the true breakup date and had the lowest MAE of any year. The spatial variability of LSTM accuracy also vary over years with 2017 having the highest precision with the least spread around the predicted breakup dates while 2022 had the lowest precision with the highest variability of predicted values. We also applied the naïve model to the testing data, the naïve model consistently underperforms compared to the LSTM with an MAE = 7.06 days, STD = 6.09 days and a MAPE = 5.52%. LSTM provides an increase in MAPE by about 1.15%, and reduction in MAE by about 23% compared to the naïve model, demonstrating the superior performance of the LSTM.

3.2. Independent Evaluation of LSTM Using the Holdout Data Set

To test the ability of the LSTM to generalize and transfer to completely new and unseen locations, we applied the trained LSTM to the holdout data set comprised of 10 APRFC locations. The LSTM was neither tuned nor trained using information from these 10 locations which span from 1980 to 2023. Overall, the MAE of the LSTM when

Table 2
Statistics for Absolute Residuals of LSTM Model Prediction of Breakup Dates Were Calculated for Each of 23 Sites in the Main Data Set

Data set	Location	River	Min. (days)	Max. (days)	STD (days)	MAE (days)	MAPE (%)
Testing Data set (2014–2023)	Buckland	Buckland	1	10	3.22	5.89	4.51
	Gakona	Copper	2	13	4.06	8.75	7.31
	Kobuk	Kobuk	1	13	3.61	4.20	3.13
	Allakaket	Koyukuk	0	7	2.36	2.70	2.15
	Hughes	Koyukuk	1	9	2.74	3.20	2.54
	Akiak	Kuskokwim	0	10	3.37	4.50	3.71
	Aniak	Kuskokwim	0	12	3.53	4.70	3.92
	Bethel	Kuskokwim	0	23	6.83	6.70	5.85
	Crooked Creek	Kuskokwim	2	10	2.79	6.70	5.55
	McGrath	Kuskokwim	1	9	2.79	4.30	3.54
	Red Devil	Kuskokwim	1	12	3.16	6.80	5.75
	Sleetmute	Kuskokwim	0	13	3.75	6.40	5.45
	Nenana	Tanana	2	23	6.06	9.70	8.49
	Circle	Yukon	0	7	2.26	3.00	2.35
	Dawson	Yukon	1	15	5.02	5.60	4.71
	Emmonak	Yukon	2	14	4.92	7.33	5.62
	Fort Yukon	Yukon	0	10	2.74	3.80	2.93
	Galena	Yukon	1	9	2.35	3.80	2.96
	Holy Cross	Yukon	0	16	4.73	5.20	3.97
	Kaltag	Yukon	0	12	3.19	4.80	3.68
	Mountain Village	Yukon	0	10	3.74	5.20	4.01
	Russian Mission	Yukon	2	13	4.09	7.22	5.52
	Tanana	Yukon	1	8	2.79	4.70	3.73
Holdouts Data set (1980–2023)	Ambler	Kobuk	0	12	3.01	4.45	3.22
	Kalskag	Kuskokwim	0	17	4.36	5.43	4.37
	Nikolai	Kuskokwim	10	45	8.01	20.50	18.65
	Stony River	Kuskokwim	0	19	5.08	6.72	5.53
	Bettles	Koyukuk	1	19	5.27	9.48	7.47
	Sunshine	Susitna	1	20	5.66	8.43	7.09
	Alakanuk	Yukon	0	13	3.98	5.97	4.26
	Anvik	Yukon	0	10	2.80	5.06	3.72
	Eagle	Yukon	2	23	4.97	10.95	8.90
	Ruby	Yukon	1	15	3.92	6.37	4.94

Note. Shown below are minimum (Min.), maximum (Max.), standard deviation (STD), mean absolute error (MAE), and mean absolute percentage error (MAPE) for each APRFC site location.

applied to the holdouts data set was 8.46 days with a STD of 6.64 days and a MAPE of 6.92% (Table 2). The best performing holdout location was Ambler on the Kobuk River with a $MAE_{location}$ of 4.45 days, a $STD_{location}$ of 3.01 days and a $MAPE_{location}$ of 3.22%. Nikolai on the Kuskokwim River had the largest $MAE_{location}$ recorded by far at 20.50 days, a $STD_{location}$ of 8.01 days and a $MAPE_{location}$ of 18.65% (Table 2). When comparing the LSTM to the naïve model on the holdout locations the naïve model demonstrated a lower error than the LSTM. Overall, the MAE of the naïve model when applied to the holdout data set was 6.70 days with a STD of 5.29 days and a MAPE of 5.25%. The best performing holdout location for the naïve model was Stony River on the Kuskokwim River with a $MAE_{location}$ of 5.79 days, a $STD_{location}$ of 4.77 days and a $MAPE_{location}$ of 4.57%. Nikolai on the Kuskokwim River was again the most challenging location within the holdouts for the naïve model to predict with a $MAE_{location}$ of 7.57 days, a $STD_{location}$ of 6.23 days and a $MAPE_{location}$ of 6.83%.

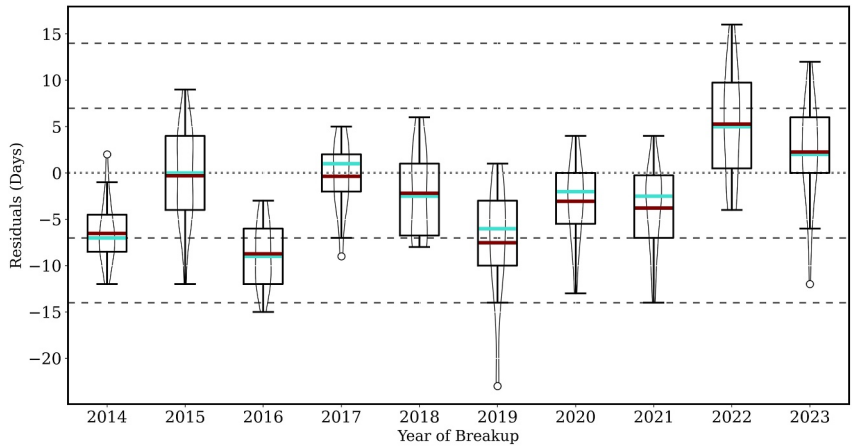

Figure 6. Boxplots overlayed with distribution plots showing the model residuals for predictions using the last 10 years of Daymet meteorological data (the testing data set from 2014 to 2023). Each plot represents residuals across the 23 locations. The maroon lines represent the mean; light blue lines represent the median; gray horizontal dotted line at Residual (days) = 0 represents perfect predictions; gray horizontal dashed lines delineat one and 2 weeks from a perfect prediction.

3.3. Forecasting River Ice Breakup Using the Trained LSTM

While we trained the LSTM model and evaluated its efficacy using observation-based gridded Daymet meteorological products, forecasting of breakup events ahead of time in spring season is desired by the local communities. We leveraged long-range seasonal weather forecast ensembles from ECMWF, available from C3S CDS (described in Section 2.4) to test the applicability of the LSTM trained using Daymet meteorological data, to forecast river ice breakup from a long-range forecast. The Daymet-based LSTM was not tuned or trained using forecast data sets and was applied as is. The LSTM was applied over the period 2020 to 2023, with 5 monthly updated forecasts windows starting on February, March, April, May and June and ending on first of July, applied to each year. During the forecast period, LSTM model was applied daily to a moving 457 days window that was composed of Daymet meteorological time series up till the start of forecast period (first of Feb/March/April/May/June) and appended with C3S CDS forecasts based meteorological time series up until the date of forecast. The LSTM performance on the long-range seasonal weather forecast data set was slightly below or at par in terms of accuracy metrics with the testing data set performed in Sections 3.1 and 3.2. Accuracy of the LSTM shows large variability across locations and years. Table 3 shows the overall performance of the LSTM applied to 23 APRFC locations for five forecasting time windows over the period 2020 to 2023. Location specific accuracies are detailed in the Tables S1–S4 of Supporting Information S1. River ice breakups at all 23 locations during this period occurred in late April or May. Overall, the 5 month forecasting window starting on February 1st (February 1st—July 1st) performed well for most years and locations, and the forecasting accuracy often declined for

Table 3
Aggregated Mean Absolute Error (MAE) ± the Standard Deviation of Absolute Residuals (STD) (Days) Using ECMWF Seasonal Forecast Data

Forecast time span	2020	2021	2022	2023
Feb 1st to July 1st	7.12 ± 5.17	5.16 ± 4.27	4.94 ± 3.91	6.37 ± 5.12
	n = 1,173	n = 1,122	n = 1,122	n = 1,173
March 1st to July 1st	7.99 ± 5.37	5.81 ± 4.73	5.41 ± 4.37	6.26 ± 4.73
	n = 1,173	n = 1,122	n = 1,122	n = 1,173
April 1st to July 1st	14.05 ± 6.24	10.77 ± 6.71	6.49 ± 4.60	6.46 ± 5.02
	n = 1,173	n = 1,122	n = 1,122	n = 1,173
May 1st to July 1st	9.92 ± 6.05	8.86 ± 6.08	6.85 ± 4.70	6.59 ± 5.09
	n = 918	n = 918	n = 1,122	n = 1,173

Note. Location specific detailed statistics are presented in Tables S1–S4 of Supporting Information S1.

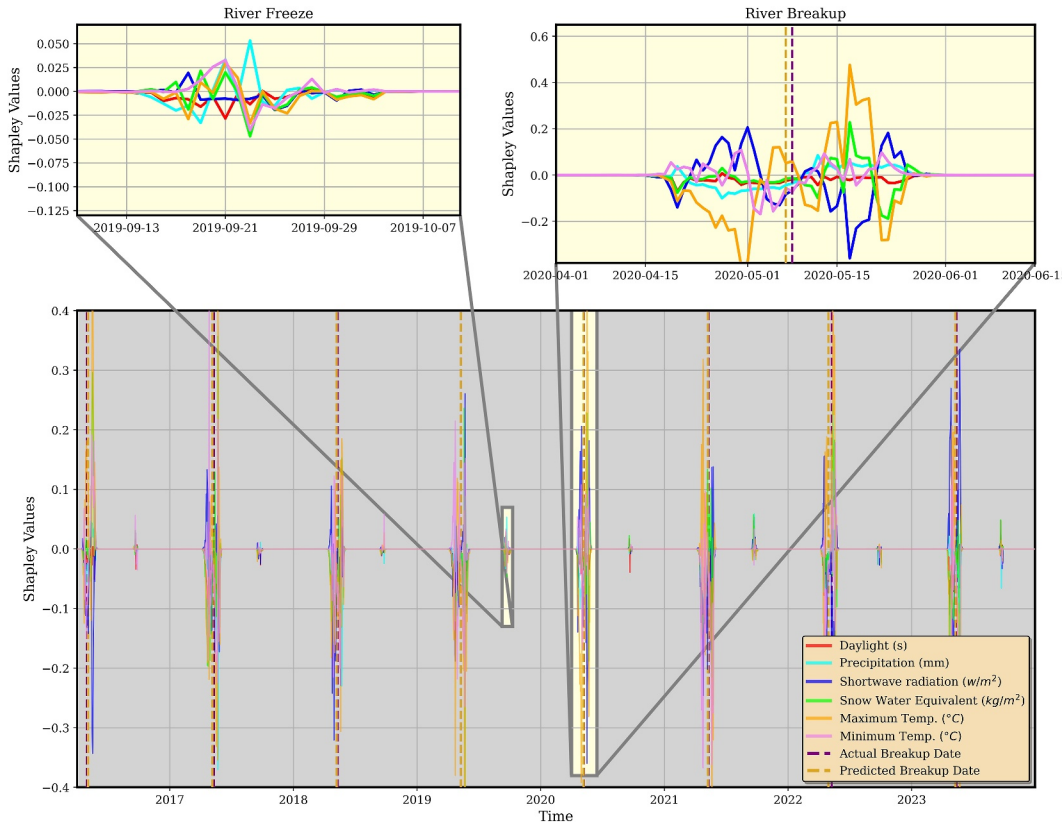

Figure 7. Shapley values for the seven temporal Daymet meteorological variables from the testing data set for Fort Yukon on the Yukon River. Subplots above (light yellow) show zoom in plots of the image at two key points in time (i.e., river ice freeze and breakup) using the 2019/2020 hydrological season as an example.

forecasting periods closer to the breakup dates. These declines were especially pronounced for year 2020 and 2021. Variability across 51 member ensembles showed similar patterns. ECMWF long-range forecast data were available at significantly lower resolution of 1° (compared to Daymet 1 km), however, the LSTM was able to forecast breakup dates at all locations with under 2 weeks of error.

3.4. Marginal Contribution of Variables Over Time Using Shapley Values

The Shapley values provide a metric for marginal contribution or feature importance over time, relative to each variable. Increases in Shapley values signify a greater contribution toward a positive score (a breakup event), negative values contribute toward a negative score (non-breakup event) while values close to zero signify that information at that time period has little to no contribution. Using Fort Yukon on the Yukon River as an example site for illustration, Figure 7 shows the Shapley values for the seven Daymet meteorological time varying variables, relative to the breakup dates. The time series of Shapley values show a high degree of activity close to both the river freeze (or slightly before it in winter) and river breakup (in spring) periods. In comparison to spring breakup period, the activity around winter freeze is lower in magnitude and persist for a shorter period of time. The remaining part of the year has little to no contribution to the predictability of breakup dates as the Shapley values tend to be close to zero. A similar seasonal patterns in Shapley values were observed at all of the sites included in our study. However, relative strengths of the Shapley values, and hence marginal contributions, associated with specific variables show variability across locations and years. At Fort Yukon (Figure 7), during the period before the freeze showed a strong contribution from precipitation and temperature related variables (i.e., precipitation, snow water equivalent minimum and maximum air temperature). However, during the breakup period, maximum air temperature was the strongest driver for the prediction with very little contribution from minimum air temperature. Shortwave radiation was also influential in determining predictions from the LSTM closer to the breakup event. Meteorological conditions in spring right before the breakup have a strong influence

on the initiation of the breakup and the predictions of the LSTM. Virtually all breakup events demonstrated a "dipole" pattern, whereby there is a spike in activity, followed by a sharp reduction in the signal, followed by a spike in activity again.

4. Discussion

4.1. Effectiveness of the LSTM for Modeling and Predicting River Ice Breakup

In this study, we designed and trained a LSTM using Daymet meteorological data (Thornton et al., 2022), watershed characteristics from ARCADE (Speetjens, 2022, 2023) and river ice breakup observations from APRFC (APRFC, 2024). Comprehensive testing of the developed LSTM demonstrates promising results and effectiveness for predicting the timing of river ice breakup across our study region in Alaska, USA. Availability of historical river ice breakup records across Alaska is often spatially limited and temporally sparse constraining the effectiveness of the LSTM trained using available data sets. Our results demonstrate that the LSTM was able to generalize the patterns across the study region and was effective in its ability to transfer and be applied to new locations for predicting breakup events. Location specific LSTM prediction errors ($MAE_{location}$) were higher for rivers with fewer observation locations in the APRFC database. Buckland, Copper, Kobuk and Tanana Rivers each had single locations available and were predicted by the LSTM with $MAE_{location}$ of 5.89, 8.75, 4.20, 6.40 days respectively (Table 2). Two locations on the Koyukuk River were predicted with $MAE_{location}$ of 2.70 and 3.20 days. The Kuskokwim River had seven locations available that were predicted with $MAE_{location}$ range of 4.30–6.80 days. Ten locations on the Yukon River were predicted with $MAE_{location}$ of 3.00–7.33 days.

When applied to the holdout data set to test transferability of the LSTM, the test locations on the Kobuk and Koyukuk Rivers were predicted with an $MAE_{location}$ of 4.45 and 9.48 days respectively (Table 2). Accuracy of the LSTM at four locations on the Yukon were slightly lower than the testing data set ($MAE_{location}$ 5.06–10.95 days). The Kuskokwim River was well represented within the training data set, and the three holdout locations along this river (Kalskag, Stony River and Nikolai) were predicted with $MAE_{location}$ of 5.43, 6.72 and 20.5 days respectively (Table 2). Nikolai on the Kuskokwim River, performed poorly with a $MAE_{location}$ of 20.50. This was by far the highest error $MAE_{location}$ recorded in our analysis. The location at Nikolai has consistently earlier breakup dates than any other location included in our analysis by almost a month, and thus was an outlier compared to all other locations Figure 3. This likely explains the poor performance of the LSTM and overprediction of the breakup date by a large error at this location. This discrepancy of timing between Nikolai on the Kuskokwim River and the other locations in the analysis is likely attributed to differing river characteristics at that location, which again, are processes that require additional data to capture. However, the LSTM was overall able to learn from and generalize well across the training locations. Application of the model to Sunshine on the Susitna River, which was completely unrepresented in the training data set and was only modeled as holdout data, produced a prediction with a $MAE_{location}$ of 8.43 days. The drop in accuracy of the LSTM on the holdout data set, compared to the testing data set, points to the limits of its generalizabilty which can be attributed to the spatially sparse and limited training data set that it was set to learn and generalize from. We expect that inclusion of breakup records from additional locations and data sources will improve the representation of underrepresented rivers in future work and will improve the accuracy and robustness of the LSTM predictions.

While the consistently late predictions at Nicolai can be attributed to the exceptionally early ground truth breakup timing observed at this location Figure 3, the contrasting performance between Nenana (poorest) and Allakaket (best) within the testing data set requires further examination. Nenana was the sole location representing the Tanana River, which inherently contributed to reduced model performance. Rivers with greater spatial representation typically exhibit superior model performance, as the LSTM regularizes toward more autocorrelated predictions. Additionally, the Tanana River represents the longest river system characterized by a single monitoring location (Figure 1), likely introducing hydraulic and hydrologic complexities inadequately captured by this singular spatial representation. Regarding Allakaket's superior performance, examination of Figure 3 reveals that this location maintains close alignment with the average temporal signal and exhibits substantial overlap with neighboring locations along the Koyuk River. However, similar patterns are observed for multiple Yukon River locations, which benefit from considerably denser spatial representation. Notably, Allakaket demonstrates strong correspondence with mean meteorological variables, consistent with its geographical positioning. The superior predictive accuracy at Allakaket likely results from the combination of average breakup timing and average meteorological conditions, which are factors that are inherently related. This convergence toward mean

conditions effectively simplified the prediction task for the LSTM, as the location represents the central tendency of the input signals rather than extremes requiring extrapolation.

When applied to the testing data set, the LSTM (MAE 5.40 days, STD 4.03 days) outperformed the naïve autoregressive model (MAE 7.06 days, STD 6.09 days), however, when applied to the holdout data set the LSTM (MAE 8.46 days, STD 6.64 days) underperformed the naïve model (MAE 6.70 days, STD 5.29 days). While the naïve model leveraged past breakup information at the particular location, to achieve generalizability and transferability, the LSTM was specifically designed to not use information about the past breakup dates allowing it to be applied to previously unseen locations.

Accurate forecasting of river ice breakup events is of great significance for local communities at high latitude cold regions that rely on seasonally frozen rivers for subsistence living, supply logistics, transportation etc. We explored the application of our trained LSTM to forecast river ice breakup leveraging ensembles of long-range seasonal forecasts from ECMWF provided by C3S CDS (C3S, 2018). The LSTM predicted the breakup for the years 2021–2023 with a MAE of 7.12 ± 5.17, 5.16 ± 4.27, 4.94 ± 3.91, 6.37 ± 6.12 days respectively with the highest accuracy achieved using forecast products released in Februrary. ECMWF seasonal forecast products are available at 1° spatial resolution, and thus the meteorological signals specific to a site is expected to have lower accuracy, impacting in turn the skill of the breakup forecast. Forecast accuracy also decreased for forecasting windows closer to the breakup dates. ECMWF seasonal forecasts are known to have issues of "jumpiness" where previous forecasts are better than the latest ones especially at long lead times (Hewson, 2018; Richardson et al., 2020; Zsoter et al., 2009). However, detailed investigation of the skill of the long-range weather forecast products was out of scope for the current study. Additionally, the analysis of variable contribution shows that interdependence of data closer to the breakup has a greater influence on predicting the breakup than observations from prior months. This pattern of autocorrelation suggests that having a transition in meteorological timeseries, from Daymet to C3S CDS forecast, close to the breakup will have greater leverage on likelihood outcomes for subsequent estimates.

4.2. Explainability of LSTM Model

We calculaed Shapley values to gain insight into the role input variables play in predicting the breakup event and their patterns over time. While they are unable to reveal specific patterns embedded within complex the architecture of neural networks, Shapley values shed light on variable significance at various steps in the time series. Figure 7 demonstrates that the LSTM is utilizing not only the information closer to the breakup period itself but also leveraging information from data close to the river freeze period. Meteorological conditions during early winter determines the freeze up and formation of the river ice which in turn has implications on the breakup during the following spring season. While the LSTM was trained with a long lookback window, including the past winter, it was never explicitly provided any information about the freeze up process or its timing. The LSTM architecture, however, is able to identify, extract and leverage, information contained within relevant variables at certain times, for predictions.

The temperature and precipitation during the coldest months of the year has the greatest correlation with breakup timing, while temperature and precipitation occurring closer to the period prior to the breakup date show weaker correlation (Limber et al., 2024). Figure 7 shows the temperature and precipitation variables as the strongest drivers during the freeze up period. Increases in the magnitude of Shapley values occur about 2–3 weeks before the freeze up coinciding with air temperature dropping close to or below freezing.

4.3. Limitations Within the Data

Data sets used in this study have several inherent properties and limitations that impact the accuracy of the LSTM. Making improvements to these data sets has the potential to improve the accuracy and robustness of the LSTM for predicting river ice breakup.

The greatest limitation of this analysis was the availability of river ice breakup records. Many locations had to be omitted from the analysis as they did not contain enough annual breakup records from 1980 to 2023 to be used. This also limited the number of rivers that could be analyzed. Perhaps the second biggest limitation of this work was the lack of high resolution information pertaining to geomorphological characteristics which are known to play a role in river freezing (Beltaos, 1995) and breakup (De Munck et al., 2017). While this additional

information (Zhou et al., 2009) would improve the LSTM, these data are challenging to collect at a large scale while gridded meteorological data is far more ubiquitious. Choosing not to use such in situ measurements allows for greater generalizability. Nonetheless, had this data been available for our study area it would likely have made a significant improvement in performance.

Static watershed characteristics included in our study were derived from the ARCADE database (Speetjens, 2022; Speetjens et al., 2023) that provides a comprehensive set of watershed properties across the pan-Arctic region. However, the data set is available at a fairly coarse spatial resolution with large basins, that at times, were inadequate to capture the local watershed characteristics for locations in our study. Several of the locations used in our study were located within a single large watershed within ARCADE, thus providing ineffective static properties to differentiate them. Deriving watershed characteristics at higher resolution, to resolve and represent the local watershed conditions for APRFC sites would help further improve the LSTM.

To develop the LSTM, time varying meteorological conditions from Daymet were sampled at the APRFC locations based on their coordinates. However, the hydrological processes in the river are impacted by the conditions experience by the larger watershed the site is located in. Incorporating the heterogeneous distribution of conditions across the watershed and not just the location would potentially improve the model. While not used in our study, hydrological flow and thermal conditions upstream of the river ice observation site can provide valuable information to better resolve and model river ice conditions.

To test the applicability of the trained LSTM for seasonal forecast we leveraged a monthly updated ECMWF ensemble product from C3S CDS (C3S, 2018). The product is available at a coarse spatial resolution, and thus does not accurately reflect the conditions experienced by the locations of interest. Medium-range weather forecasts, that are often available at higher spatial resolution, and shorter lead times of 6–10 days and better forecast skill (Wagner, 1989) may provide higher accuracy for LSTM forecast.

As part of continued research, in the future we plan to address these limitations to improve our ability to model and predict the dynamics of river ice breakup in seasonally frozen rivers of Alaska.

5. Conclusion

In this study we developed a LSTM-based approach to model and predict river ice breakup in seasonally frozen rivers of Alaska, USA. Results from the presented analysis show promise in improving our ability to forecast river ice breakup over high latitude regions. Our study makes a number of contributions including: (a) designed and developed a LSTM capable of predicting river ice breakup with a high degree of accuracy leveraging easily available meteorological data sets; (b) LSTM architecture able to generalize over multiple locations and transfer to new locations across a vast study area (although further improvement is needed); (c) demonstrated the ability to apply the LSTM for long-range seasonal forecasts of river ice breakup; (d) developed a new scheme for temporally adjusted Shapley values to interpret and understand the temporal contribution of meteorological data on the LSTM predictions over time.

Our results show that the LSTM is able to predict the annual breakup date on average within 5.40 days with a standard deviation of 4.03 days. We also demonstrate the practical application of the LSTM to predict the breakup date within less than a week error on average using long-range seasonal forecast data (Table 3). We demonstrate that the model is able to generalize over a large geographical region, across different locations, along different rivers, leveraging primarily meteorological inputs. The Shapley analysis was able to provide insights into variables (typically precipitation and temperature) and time periods (the freeze and breakup periods) that exert the primary control on river ice breakup prediction.

A key contribution of our work is the observation-trained model capable of generalizing river ice breakup timing over space. The vast expanse of the State of Alaska represents large topographic complexity and heterogeneous microclimate, where many regions remain completely ungauged and even well-monitored locations have significant data gaps due to observer availability, access constraints, or safety concerns during breakup periods. A spatially generalizable model offers the promise of better forecast of breakup timing, beyond reliance on observations from nearest monitoring location. Observation-based reanalysis and weather forecasts products provide consistent spatio-temporal coverage fo meteorological data, without the observational gaps inherent in community-based monitoring, that can be leveraged by our LSTM-based model to forecast river ice breakup even in remote regions of Alaska. As demonstrated by the results of the naïve model, even a simplistic surrogate

model, leveraging site specific information, is able to capture valuable predictive information pertaining to river ice breakup timing. Used in conjunction with surrogate models, our meteorology-based model can be applied for a retrospective analysis and serve as a valuable gap-filling tool for missing observational records, and provide uncertainty bounds for community observations.

In our future study, we aim to integrate our LSTM-based modeling approach with operational long-range and high resolution medium-range weather forecasts to predict when river ice breakup in high latitude regions will occur with improved forecast skill and lead times. Forecast results from the seasonal ensemble provided by ECMWF shows that the model is able to predict the river ice breakup comparably well to when using observation-based Daymet data exclusively, for a 5 month forecast leading to the first of July used in this study. Results from our study are encouraging and demonstrates the potential for seasonal and short-term forecast of river ice breakup timing. River ice dynamics processes, such as freezeup and breakup are also completely missing in most Earth system models, however, data driven models such as the LSTM developed in this study that are capable of predicting such processes using meteorological information offers opportunities to develop hybrid physics-based and machine learning models to fill that void in Earth system models. We hypothesize that this general methodology, combining LSTMs with maximum likelihood estimation to identify infrequent events in nature, could have greater applications for the cryosphere community through the prediction of other hydrological processes such as freezeup, ice jams and more.

Data Availability Statement

Daily meteorological data used to train, tune and test the LSTM were obtained from Daymet (Thornton et al., 2022) and are publicly available from NASA ORNL DAAC at https://daymet.ornl.gov/. River ice breakup records were collected from the Alaska-Pacific River Forecast Center (APRFC, 2024) at https://www.weather.gov/aprfc/breakupDB. Static variables representing watershed characteristics were obtained from the pan-Arctic Catchment Database (ARCADE) (Speetjens, 2022; Speetjens et al., 2023) and are available at https://doi.org/10.34894/u9hspv. ERA5 reanalysis meteorological variables used to evaluate the model on forecasted data were obtained from the Copernicus Climate Data Store (C3S, 2018), and are openly available under the Copernicus products license at https://cds.climate.copernicus.eu/datasets/seasonal-original-single-levels?tab=overview. All figures were created using matplotlib version 3.5.1 (Caswell et al., 2021; Hunter, 2007), openly available under the Matplotlib license at https://matplotlib.org/. The deep learning architecture was created using TensorFlow version 2.15.0 (Abadi et al., 2015) with Bayesian hyperparameter tuning carried out using the KerasTuner API (O'Malley et al., 2019). All of the codes needed to run the analysis and reproduce this work are available on GitHub at https://github.com/Russtyhub/river_ice_Daymet_LSTM_code.git.

Acknowledgments
This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Regional and Global Model Analysis (RGMA) program, as part of The Interdisciplinary Research for Arctic Coastal Environments (InteRFACE) project and the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Science Focus Area. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy Office of Science User Facility under project m3958 (2024). This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. [Software]. Retrieved from https://www.tensorflow.org/

APRFC. (2024). Alaska-Pacific River Forecast center—break up database [Dataset]. Retrieved from https://www.weather.gov/aprfc/breakupDB

Arp, C. D., Cherry, J. E., Brown, D. R. N., Bondurant, A. C., & Endres, K. L. (2020). Observation-derived ice growth curves show patterns and trends in maximum ice thickness and safe travel duration of Alaskan lakes and rivers. *The Cryosphere*, *14*(11), 3595–3609. https://doi.org/10.5194/tc-14-3595-2020

Ashton, G. (1986). *River and lake ice engineering*. Water Resources Publications. Retrieved from https://books.google.com/books?id=xglYVjAsnt8C

Beltaos, S. (Ed.) (1995). *River ice jams*. Water Resources Publications.

Beltaos, S. (2002). Effects of climate on mid-winter ice jams. *Hydrological Processes*, *16*(4), 789–804. https://doi.org/10.1002/hyp.370

Beltaos, S. (2003). Threshold between mechanical and thermal breakup of river ice cover. *Cold Regions Science and Technology*, *37*(1), 1–13. https://doi.org/10.1016/s0165-232x(03)00010-7

Beltaos, S., Burrell, B. C., & Ismail, S. (1994). Ice and sedimentation processes in the Saint John River, Canada. In *Proceedings of the international association for hydraulic research, international ice symposium* (Vol. 1, pp. 11–21). Trondheim.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches †. *Energies*, *11*(7), 1636. https://doi.org/10.3390/en11071636

Brown, D. R., Brinkman, T. J., Verbyla, D. L., Brown, C. L., Cold, H. S., & Hollingsworth, T. N. (2018). Changing river ice seasonality and impacts on interior Alaskan communities. *Weather, Climate, and Society*, *10*(4), 625–640. https://doi.org/10.1175/WCAS-D-17-0101.1

C3S. (2018). Seasonal forecast daily and subdaily data on single levels. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [Dataset]. *Climate Data Store*. https://doi.org/10.24381/cds.181d637e

Caswell, T., Droettboom, M., Lee, A., De Andrade, E., Hoffmann, T., & Hunter, J. (2021). matplotlib/matplotlib [Software]. *Zenodo*. Retrieved from https://pypi.org/project/matplotlib/3.5.1/

Chambers, P. A., Prowse, T. D., & Culp, J. M. (1994). Ecological effects of river ice break-up: A review and perspective. *Freshwater Biology*, *32*(2), 261–275. https://doi.org/10.1111/j.1365-2427.1994.tb01125.x

Church, M. (1974). Hydrology and permafrost with reference to northern North America. In *Permafrost hydrology, proceedings of a workshop seminar* (pp. 7–20).
De Coste, M., Li, Z., & Dibike, Y. (2022). Assessing and predicting the severity of mid-winter breakups based on Canada-wide river ice data. *Journal of Hydrology*, *607*, 127550. https://doi.org/10.1016/j.jhydrol.2022.127550
De Munck, S., Gauthier, Y., Bernier, M., Chokmani, K., & Légaré, S. (2017). River predisposition to ice jams: A simplified geospatial model. *Natural Hazards and Earth System Sciences*, *17*(7), 1033–1045. https://doi.org/10.5194/nhess-17-1033-2017
Dewancker, I., McCourt, M., & Clark, S. (2016). Bayesian optimization for machine learning: A practical guidebook. https://doi.org/10.48550/arXiv.1612.04858
French, H. (2018). *The periglacial environment*. (4th ed.). Permafrost and Periglacial Processes. https://doi.org/10.1002/ppp.2009
Gray, D., & Prowse, T. (1993). Snow and floating ice. In D. Maidment (Ed.), *Handbook of hydrology* (pp. 7–1–7–58). McGraw-Hill.
Hewson, T. (2018). Forecast jumpiness: An introduction. *ECMWF*. Retrieved from https://www.ecmwf.int/node/18120
Hicks, F., Cui, W., & Ashton, G. D. (2008). Heat transfer and ice cover decay. In S. Beltaos (Ed.), *River ice breakup* (pp. 67–123). Highlands Ranch, CO: Water Resources Publications, LLC.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
Hu, J., Liu, L., Huang, Z., You, Y., & Rao, S. (2008). Ice breakup date forecast with hybrid artificial neural networks. In *2008 Fourth International Conference on natural computation* (Vol. 2, pp. 414–418). https://doi.org/10.1109/ICNC.2008.169
Hunter, J. D. (2007). Matplotlib: A 2d graphics environment [Software]. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55
Ian, J., Todd, B., Kelda, B., James, K., Kris, H., Bryce, L., & Verbyla, D. (2016). Quantifying rural hunter access in Alaska. *Human Dimensions of Wildlife*, *21*(3), 240–253. https://doi.org/10.1080/10871209.2016.1137109
INAC. (2008). *Standardized protocols for collection of monitoring information for the nwt cumulative impact monitoring program working group and valued components advisory teams*. Inuvik, NT: Indian and Northern Affairs Canada.
Jasek, M. (1999). 1998 break-up and flood on the Yukon River at Dawson – Did El Niño and climate change play a role? In H. Shen (Ed.), *Ice in surface waters: Proceedings of the 14th international symposium on ice* (pp. 761–768). A.A. Balkema.
Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. https://doi.org/10.48550/arXiv.1412.6980
Limber, R., Massoud, E. C., Guan, B., Hoffman, F. M., & Kumar, J. (2024). Influence of atmospheric rivers on Alaskan river ice. *Geophysical Research Letters*, *51*(23), e2024GL111544. https://doi.org/10.1029/2024GL111544
Lindenschmidt, K.-E. (2020). *River ice processes and ice flood forecasting*. Springer.
Mackay, J. R., & MacKay, D. K. (1973). Break-up and ice jamming on the Mackenzie River, N.W.T. Hydrologic aspects of northern pipeline development. (Tech. Rep. Nos. 73–3). *Saskatoon, Saskatchewan: Canada, Environmental-Social Committee, Northern Pipelines, Task Force on Northern Oil Development*.
Madaeni, F., Chokmani, K., Lhissou, R., Homayouni, S., Gauthier, Y., & Tolszczuk-Leclerc, S. (2022). Convolutional neural network and long short-term memory models for ice-jam predictions. *The Cryosphere*, *16*(4), 1447–1468. https://doi.org/10.5194/tc-16-1447-2022
Murray, F. W. (1967). On the computation of saturation vapor pressure. *Journal of Applied Meteorology and Climatology*, *6*(1), 203–204. https://doi.org/10.1175/1520-0450(1967)006⟨0203:OTCOSV⟩2.0.CO;2
Newton, B., Prowse, T., & Rham, L. (2017). Hydro-climatic drivers of mid-winter break-up of river ice in western Canada and Alaska. *Hydrology Research*, *48*(4), 945–956. https://doi.org/10.2166/nh.2016.358
O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Kerastuner [Software]. https://github.com/keras-team/keras-tuner
Paily, P. P., Macagno, E. O., & Kennedy, J. F. (1974). *Winter-regime surface heat loss from heated streams*. Research Report (Report No. PB-241941; IIHR-155). Iowa Institute of Hydraulic Research. Retrieved from https://www.osti.gov/biblio/7179276
Pavelsky, T. M., & Smith, L. C. (2004). Spatial and temporal patterns in arctic river ice breakup observed with Modis and Avhrr time series. *Remote Sensing of Environment*, *93*(3), 328–338. https://doi.org/10.1016/j.rse.2004.07.018
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*, 443–451. https://doi.org/10.3189/172756407782871431
Prowse, T., & Gridley, N. (1993). *Environmental aspects of river ice*. National Hydrology Research Institute.
Rabier, F. (2023). *ECMWF 50: Focus on ensembles* (Vol. 176). ECMWF Newsletter. Retrieved from https://www.ecmwf.int/en/newsletter/176/editorial/focus-ensembles
Richardson, D. S., Cloke, H. L., & Pappenberger, F. (2020). Evaluation of the consistency of ECMWF ensemble forecasts. *Geophysical Research Letters*, *47*(11), e2020GL087934. https://doi.org/10.1029/2020gl087934
Sagarin, R., & Micheli, F. (2001). Climate change in nontraditional data sets. *Science*, *294*(5543), 811. https://doi.org/10.1126/science.1064218
Shapley, L. S. (1967). *Utility comparison and the theory of games* (Tech. Rep.). Rand Corporation.
Shen, H. T. (2003). Research on river ice processes: Progress and missing links. *Journal of Cold Regions Engineering*, *17*(4), 135–142. https://doi.org/10.1061/(ASCE)0887-381X(2003)17:4(135)
Shen, H. T. (2010). Mathematical modeling of river ice processes. *Cold Regions Science and Technology*, *62*(1), 3–13. https://doi.org/10.1016/j.coldregions.2010.02.007
Shen, H. T., & Chiang, L. (1984). Simulation of growth and decay of river ice cover. *Journal of Hydraulic Engineering*, *110*(7), 958–971. https://doi.org/10.1061/(ASCE)0733-9429(1984)110:7(958)
Speetjens, N. J. (2022). Arcade: The pan-arctic catchment database [Dataset]. *DataverseNL*. https://doi.org/10.34894/U9HSPV
Speetjens, N. J., Hugelius, G., Gumbricht, T., Lantuit, H., Berghuijs, W. R., Pika, P. A., et al. (2023). The pan-Arctic Catchment Database (ARCADE). *Earth System Science Data*, *15*(2), 541–554. https://doi.org/10.5194/essd-15-541-2023
Sterbak, T. (2022). How to calculate Shapley values from scratch. Retrieved from https://www.depends-on-the-definition.com/shapley-values-from-scratch/
Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union*, *38*(6), 913–920. https://doi.org/10.1029/TR038i006p00913
Štrumbelj, E., & Kononenko, I. (2013). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*, 647–665. https://doi.org/10.1007/s10115-013-0679-x
Sun, W., & Trevor, B. (2018). A stacking ensemble learning framework for annual river ice breakup dates. *Journal of Hydrology*, *561*(January), 636–650. https://doi.org/10.1016/j.jhydrol.2018.04.008
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. [Article]. *Journal of Hydrologic Engineering*, *13*(9), 811–816. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:9(811)

Thellman, A., Jankowski, K. J., Hayden, B., Yang, X., Dolan, W., Smits, A. P., & O'Sullivan, A. M. (2021). The ecology of river ice. *Journal of Geophysical Research: Biogeosciences*, *126*(9), e2021JG006275. https://doi.org/10.1029/2021JG006275

Thornton, M., Shrestha, R., Wei, Y., Thornton, P., Kao, S.-C., & Wilson, B. (2022). Daymet: Annual climate summaries on a 1-km grid for North America, version 4 r1 [Dataset]. *ORNL Distributed Active Archive Center*. https://doi.org/10.3334/ORNLDAAC/2130

Wagner, A. J. (1989). Medium- and long-range forecasting. *Weather and Forecasting*, *4*(3), 413–426. https://doi.org/10.1175/1520-0434(1989)004⟨0413:malrf⟩2.0.co;2

Wang, X., & Feng, L. (2024). Patterns and trends in northern hemisphere river ice phenology from 2000 to 2021. *Remote Sensing of Environment*, *313*, 114346. https://doi.org/10.1016/j.rse.2024.114346

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, *31*(7), 1235–1270. https://doi.org/10.1162/neco_a_01199

Zhou, H., Li, W., Zhang, C., & Liu, J. (2009). Ice breakup forecast in the reach of the Yellow River: The support vector machines approach. *Hydrology and Earth System Sciences Discussions*, *6*, 3175–3198. https://doi.org/10.5194/hessd-6-3175-2009

Zsoter, E., Buizza, R., & Richardson, D. (2009). "Jumpiness" of the ECMWF and met office eps control and ensemble-mean forecasts. *Monthly Weather Review*, *137*(11), 3823–3836. https://doi.org/10.1175/2009mwr2960.1