Proceedings of the 26th International Conference on
Port and Ocean Engineering under Arctic Conditions
June 14-18, 2021, Moscow, Russia

Iceberg Drift Forecasting Using Machine Learning

Renat Yulmetov[1]
[1]C-CORE, St. John's, Newfoundland and Labrador, Canada

ABSTRACT

A machine learning approach has been developed to improve iceberg drift forecasting on a tactical timescale to mitigate potential impacts between icebergs and offshore facilities in the North Atlantic. The problem of iceberg forecasting involves many uncertain terms. Although iceberg drift trajectories can be measured with sufficient accuracy, the iceberg shapes are generally unknown. Moreover, the corresponding metocean forecasts, in particular ocean current forecasts, carry their own uncertainties. The forecasting error for this highly uncertain drift process can be minimized using a hybrid approach that calculates Coriolis acceleration explicitly and estimates combined hydrodynamic and wind drag accelerations using machine learning. The latter is an implementation of gradient boosted trees algorithm that predicts and integrates the unknown accelerations. The training dataset consists of iceberg GPS tracks obtained during 2019 iceberg data collection campaign offshore Labrador. The trained model is cross-validated using the k-folds approach. The evolution of the average error between the observed and forecast tracks is used as the measure of performance. The resulting error curve is promising and can be improved even further if more, and more accurate in-situ data are collected. The trained model can be used operationally, together with conventional dynamic approach, as an additional source of information in order to determine an adequate response to iceberg threat.

KEYWORDS: Icebergs, machine learning, forecasting, Newfoundland, iceberg management

INTRODUCTION

The 2019 iceberg season was busy for the offshore industry and tourism in Newfoundland, with 1515 reported icebergs crossing 48° latitude (International Ice Patrol, 2019). Based on previous years' statistics, the 2019 iceberg season severity was characterized as "extreme". The oil and gas industry relied on extensive iceberg management operations near offshore facilities on the Grand Banks.

Iceberg management operations on the Grand Banks usually include, but are not limited to, iceberg surveillance, tracking and drift forecasting. Surveillance is performed by means of reconnaissance flights, satellite imagery analysis and on-site vessel observations. The tracking is done from the radar-equipped vessels and offshore facilities, however, GPS tracking may be involved in special cases. The iceberg tracks are usually logged in a table containing information on iceberg locations, size, towing attempts, and associated weather conditions.

Operational drift forecasting for the observed icebergs is carried out using International Ice Patrol (IIP) and Canadian Ice Service (CIS) drift models (Murphy and Carrieres, 2010). One is usually better at estimating distance and the other at estimating direction of iceberg drift tracks. These models use large-scale gridded ocean and atmosphere circulation products for wind and current input data. The input data carry high level of uncertainty, yet they are often the only source of data, because local Acoustic Doppler Current Profiler (ADCP) or other local metocean measurements are unavailable.

Several models have been recently developed to deal with these uncertain metocean inputs and to generally advance short-term iceberg drift forecasting. Some of the approaches are based on the use of inferred currents (Turnbull et al., 2015), ocean current statistics (Andersson et al., 2019), state-estimation and filtering (Andersson et al., 2016), and machine learning (Yulmetov and Ralph, 2019).

The initial attempt to apply machine learning (ML) to iceberg drift prediction was based on a neural network (NN) approach (Yulmetov and Ralph, 2019). It was a purely statistical model that predicted new iceberg velocities using observed iceberg velocities and additional metocean inputs. The NN approach is prone to input normalization issues, and is inferior in performance to modern ML regression algorithms.

This study presents the next generation of the approach undertaken in Yulmetov and Ralph (2019). The major improvements are: implementation of gradient-boosted trees instead of NN, and a hybrid approach instead of a "black box" model. In addition, the model is trained on new GPS iceberg tracks collected during an iceberg tagging campaign off the coast of Labrador in 2019 (Briggs et al., 2020).

The paper starts with a short description of the approach and the setup of the forecasting model. It is followed by the input data description. The forecasting performance is presented in the next section and discussed afterwards. A short conclusion is given in the final section.

METHOD

Iceberg drift prediction can be approached using dynamics-based models that integrate the equation of motion for an iceberg, or using statistical models that fit the iceberg motion based on local currents, winds and waves. There are several hybrid models that are able to deal with uncertain inputs statistically while still relying on the equation of motion to ensure physical validity. The resulting performance for those models tend to be superior (Andersson et al., 2018).

A hybrid model is presented here. The drift model predicts uncertain drag accelerations using ML while the Coriolis acceleration is calculated accurately using iceberg location and velocity according to:

$$\frac{d\vec{U}}{dt} = -f\vec{k} \times \vec{U} + \vec{a}\left(\vec{U}, \overrightarrow{V_a}, \overrightarrow{V_w}\right) \tag{1}$$

where $\vec{U}$ is iceberg velocity, f is Coriolis parameter, $\vec{k}$ is an outward unit normal vector to the Earth surface, $\vec{a}$ is acceleration caused by hydrodynamic force acting from water and air drag, $\overrightarrow{V_a}$ is wind velocity, and $\overrightarrow{V_w}$ is ocean current velocity. The ML part of the model uses observed iceberg velocities, predicted wind velocity and predicted ocean current velocity to predict the combined hydrodynamic and wind drag acceleration.

Machine learning is a term that covers a wide range of algorithms and methods aiming to classify, regress, predict, or analyze data. An approach called supervised learning is used in this model. In supervised learning a function that maps a set of inputs onto the set of outputs is determined based on the known data. For each known input/output pair the function predicts the output and is adjusted depending on the error metric between the known and predicted outputs. This adjustment process is called "training". In this case the model is trained to predict the unknown acceleration based on iceberg velocity, wind velocity and ocean velocity. A sufficient amount of accurate data is the key to ML performance.

Neural networks were used in the first generation of the drift model, but generally better-performing algorithms were discovered. The gradient boosting (GB) technique was chosen based on better performance for general regression problems and a lack of normalization issues to which NNs are prone. In GB the algorithm combines multiple decision trees, improving prediction. The original work combining multiple regression trees into an optimal regression function was published by Friedman (2001). An open-source implementation called Catboost was used here (Prokhorenkova et al., 2018). It supports GPU architecture, has a wide range of optimization parameters and tools to perform cross-validation routines.

Depending on the discretization of (1), two mathematical formulations were implemented. An explicit form can be written as:

$$\begin{aligned} \frac{U_x^{i+1} - U_x^{i-1}}{2\Delta t} &= fU_y^i + a_x^i \\ \frac{U_y^{i+1} - U_y^{i-1}}{2\Delta t} &= -fU_x^i + a_y^i \end{aligned} \tag{2}$$

where the upper index denotes the time step, and the lower index denotes the projection. In the training process the model is trained to estimate the acceleration projections (a_x^i, a_y^i) using known iceberg velocities at time steps $(i-1, i, i+1)$, wind velocity at time step i and ocean current velocity at time step i. In the actual prediction process, the algorithm calculates the unknown acceleration using the trained model, and then finds an unknown position of the iceberg at time step $i+1$.

The implicit form can be written as:

$$\begin{aligned} \frac{3U_x^{i+1} - 4U_x^i + U_x^{i-1}}{2\Delta t} &= fU_y^{i+1} + a_x^{i+1} \\ \frac{3U_y^{i+1} - 4U_y^i + U_y^{i-1}}{2\Delta t} &= -fU_x^{i+1} + a_y^{i+1} \end{aligned} \tag{3}$$

In the implicit case the training and prediction processes require wind velocity and ocean current velocities at time step $i+1$, instead of i as in the explicit case.

Thus, a total of eight inputs are required at each time step to make a prediction: two velocity vectors, wind velocity, and ocean current velocity, each having two projections. The model is currently trained to make single-step predictions and integrate multiple single-step predictions iteratively. The drift forecast duration does not exceed the tactical timescale, which is 24 hours long in this study.

DATA

The iceberg GPS tracks were collected during a tagging campaign offshore Newfoundland and Labrador in 2019. Unmanned Aerial Vehicles (UAVs) were used to deliver the GPS trackers from the vessel to the icebergs (Briggs et al., 2020). Each of the trackers consisted of a GPS unit attached to a rubber mat with toothed metal plates. This maximized friction, preventing the trackers from sliding down the icy surface when placed on a slope or as it tilts during a partial iceberg roll.

After the UAV with the payload reached an iceberg, the deployment process consisted of two stages in which the tracker would be first released free-hanging on a short tether, and then it would be released completely when the tracker was suitably positioned on an iceberg. The two-stage release was developed to avoid risks of crashing associated with the UAV's landing and to provide the sense of depth to the pilot because of the bright white iceberg surface.

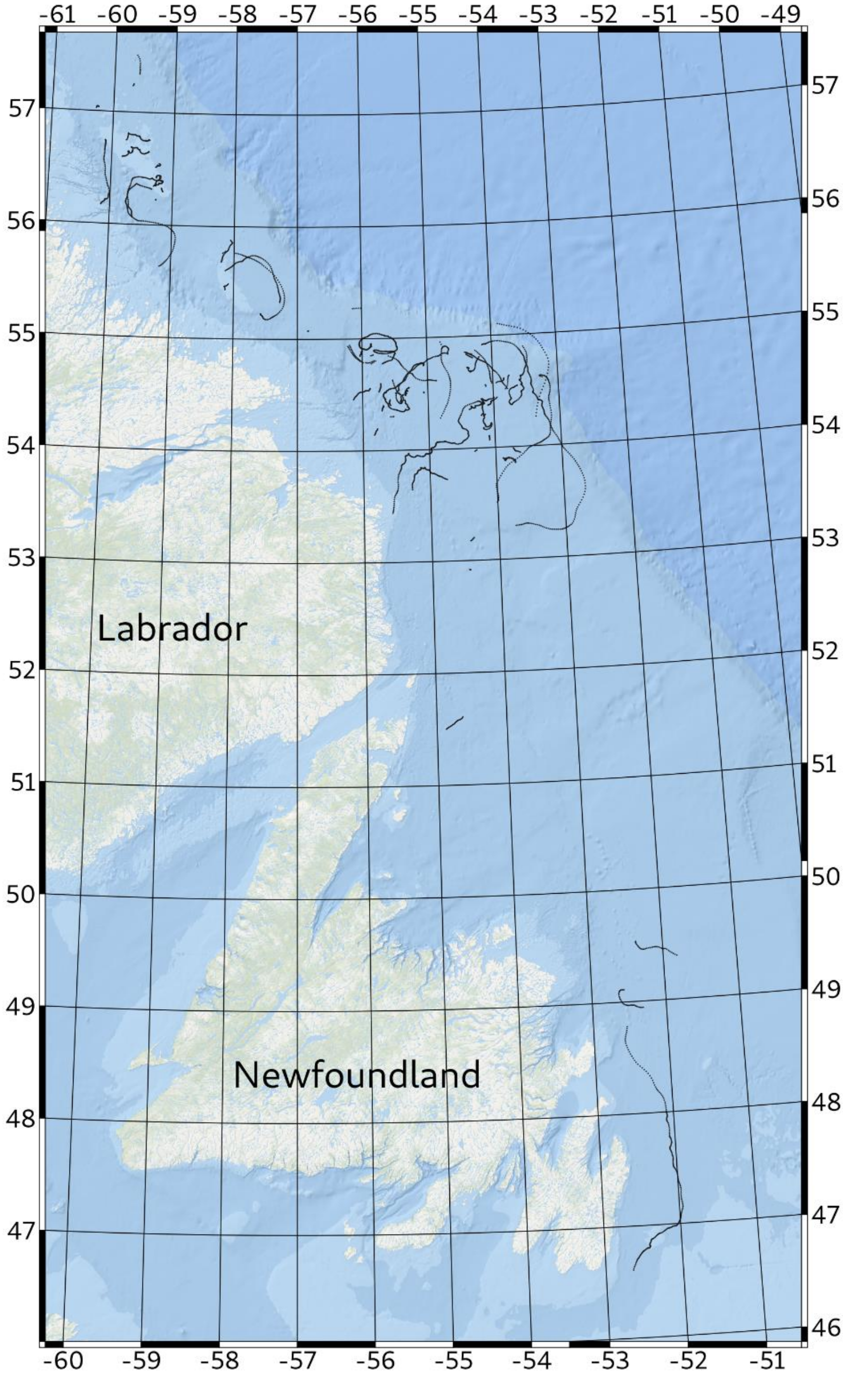

Figure 1. Iceberg drift tracks used for the model training

The trackers were set to transmit the GPS location every 30 minutes. At total of 119 GPS trackers were deployed, however many of the trackers got lost within just a few hours, because of waves washing over some of the smaller icebergs and frequent iceberg rolling over in warm weather. Multiple icebergs were excluded from consideration because they were grounded or scouring, which would affect the training of the model. Finally only 77 iceberg tracks were left, and used for the model training (Figure 1). The average track duration was approximately 2 days, while the longest track lasted more than two weeks.

It is worth noting that the icebergs were profiled using multibeam echosounder for keels, and laser scanner for sails (Bruce et al., 2021). The majority of the icebergs that were tagged can be classified as small/medium size, however, there were a few larger tabular icebergs. The smallest iceberg was less than 30 m waterline length, while the largest was longer than 300 m.

The wind data were taken from the CFSv2 operational analysis (Saha et al., 2014), which is a gridded global-scale circulation model. Its spatial resolution is 0.2°, and its temporal resolution is one hour. The ocean current data were taken from HYCOM (Chassignet et al., 2007), which is a gridded global-scale ocean model. It has 0.08° spatial resolution, 32 vertical layers in the water column and three-hourly temporal resolution. The ocean currents were averaged for the top 50 m, which was characteristic of the tracked icebergs' keel depth. The gridded metocean data have been linearly interpolated over the iceberg trajectories (Figure 2).

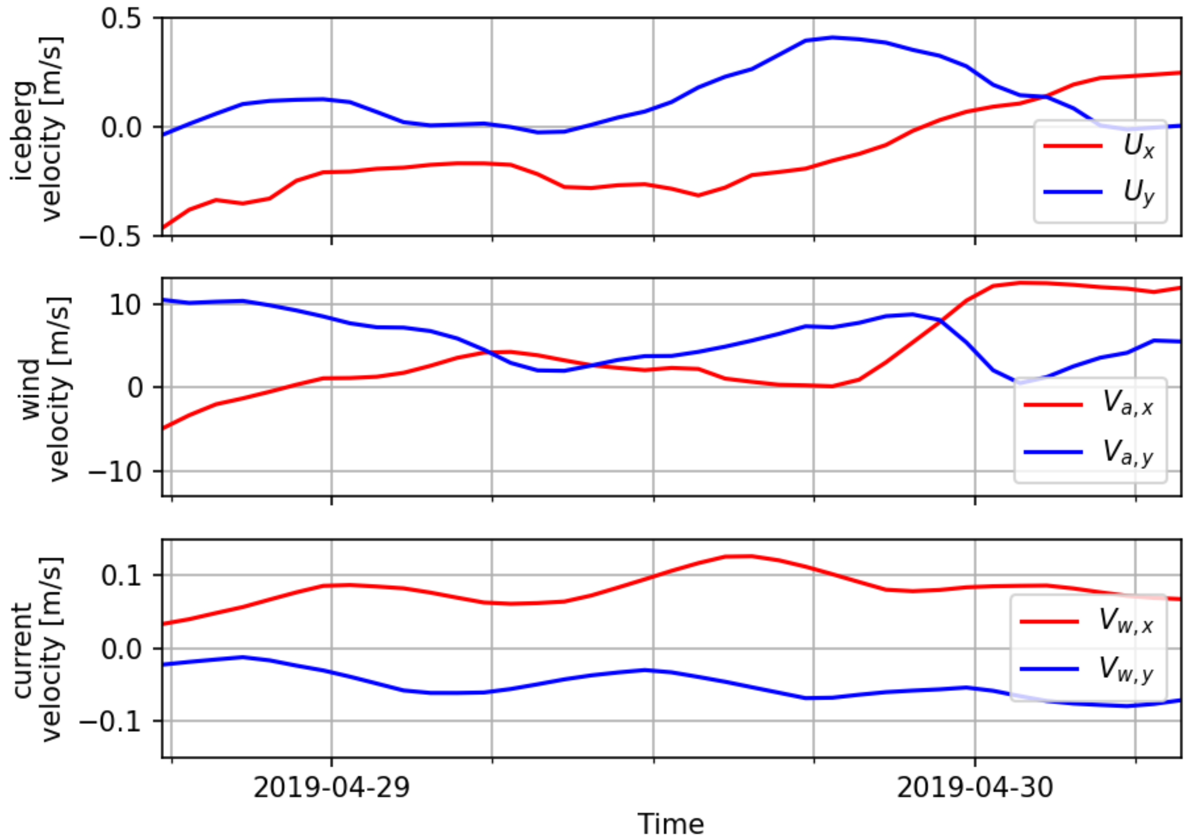

Figure 2. An example of iceberg velocity and corresponding wind velocity and 50m-average ocean current velocity

Large scale gridded models are generally considered insufficiently accurate for reliable iceberg drift modelling, however those are often the only data sources available. Current operational models are using gridded metocean products as inputs. Local current profiles are required for accurate drift hindcasts, and it is expected that real-time access to ADCP and weather station data would help to correct the gridded products, and thus deliver more accurate drift forecasts.

RESULTS

The model has been used to forecast every 24 hour trajectory span starting every hour of drift data for each of the icebergs in the database. Shorter forecasts were issued where remaining measured track duration was shorter than 24 hours. An example of measured and predicted iceberg tracks is shown in Figure 3a. The corresponding drift velocity and metocean data are shown in Figure 2. It can be seen that in the first 12 hours, the predicted iceberg drifted with approximately same speed, but then it slowed down significantly, while the real iceberg accelerated northwest. An increasing eastward ocean current component provided by HYCOM model likely decelerated the forecasted iceberg.

The forecasting error is characterized by the evolution of distance between actual and predicted iceberg locations in time for a single forecast. A more accurate forecast would result in lower error curve. The corresponding forecast error curve for this particular example is shown in Figure 3b. The error stayed below 3 km in the first 12 hours, while the forecasted iceberg was able to catch up. But then the error jumped up to 12 km in 24 hours, when the forecasted iceberg decelerated.

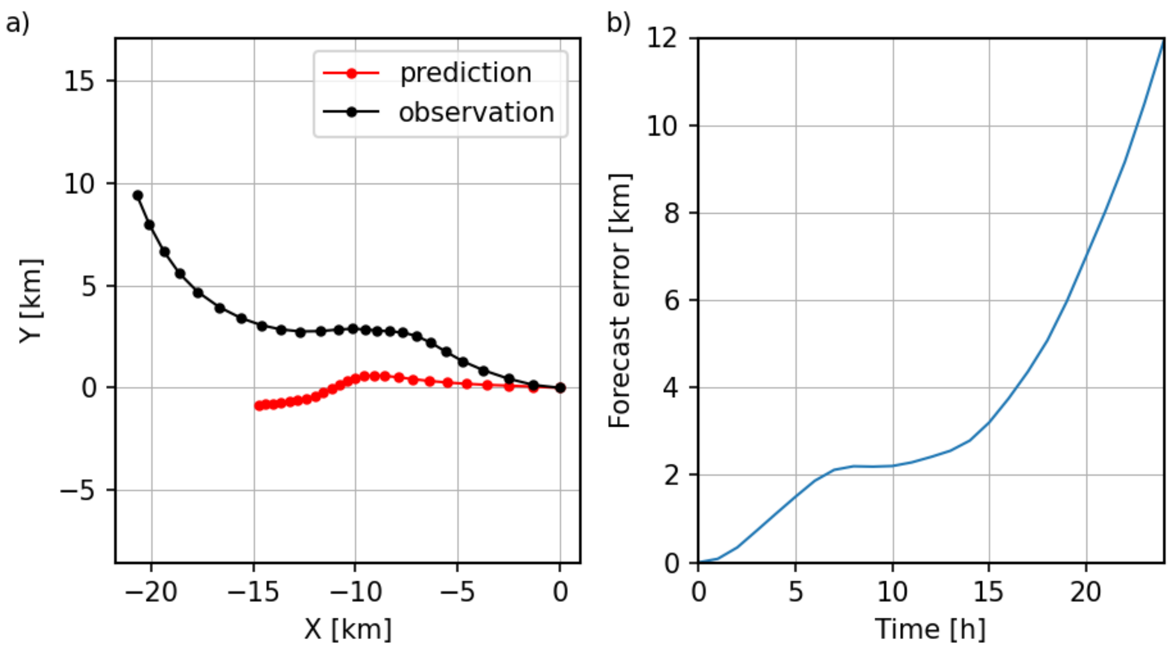

Figure 3. a) An example of observed and predicted iceberg tracks; b) the corresponding forecast error.

Error curves averaged between all of the available forecasts indicate model performance. The average forecasting error curve for the explicit model given by equation (2) trained using the whole dataset shows approximately 11.2 km error at 24 hours of forecast (Figure 4). The corresponding standard deviation is 7.2 km. The 24 hour error for the implicit model is

10.1±6.6 km. For the implicit implementation the minimal 24 hour forecast error achieved is 79 m, the worst forecast has been 35.6 km off. These numbers, however, are biased because of the same data used for training and testing the forecasting performance.

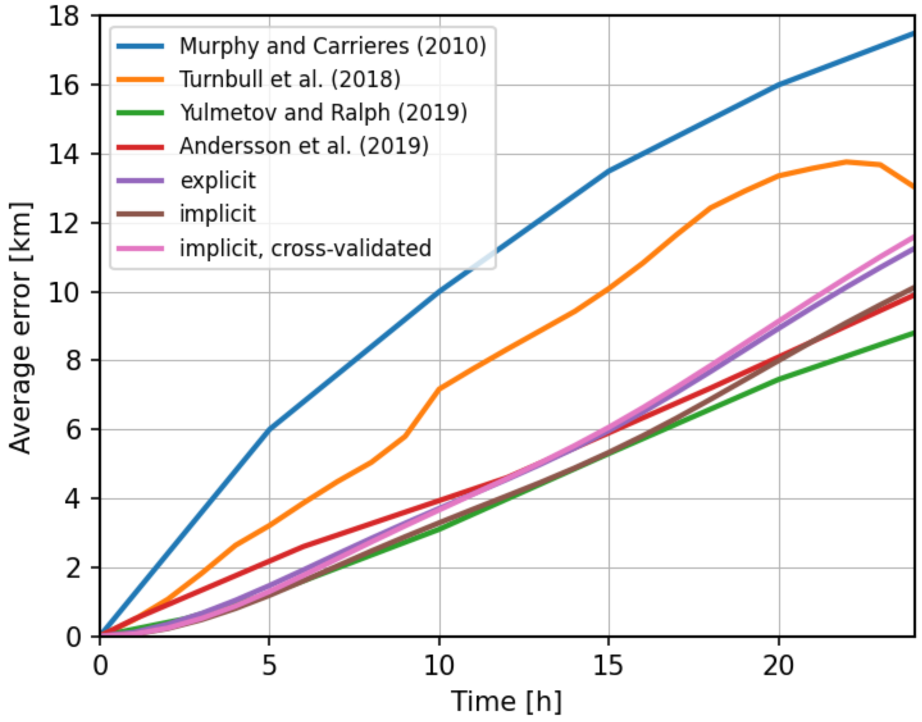

Figure 4. Average error curves compared for various iceberg drift forecasting models.

K-folds cross-validation is a statistical method that allows to estimate model performance on new 'unfamiliar' data. In the *k*-folds cross-validation, the data is divided into *k* blocks of equal size. By iterating over the blocks, the model is trained using the corresponding remaining *k-1* blocks, and its performance is measured using that 'unfamiliar' *k*-th block. The average performance calculated between the *k* blocks can be considered a more accurate estimate of expected model performance, than pure training performance. The cross-validated implicit model's average error is about 11.5±7.3 km error at 24 hours.

DISCUSSION

It is hard to compare various models' performance on different data sources because, although iceberg tracks can be measured with rather consistent accuracy, the metocean input data accuracy varies. Figure 4 compares average error curves calculated for various drift models. The operational IIP/CIS models deliver more than 17 km of error at 24 hours on the Grand Banks (Murphy and Carrieres, 2010). A purely statistical drift model with ocean currents predicted using vector-autoregression achieves less than 10 km error (Andersson et al., 2019). The inferred current approach applied to 18 profiled icebergs offshore Labrador resulted in 13 km average error (Turnbull et al., 2018). Based on historical data from Labrador exploratory drilling sites in 1980s, the currents and winds measured locally provided more accurate drift forecasts (Yulmetov and Ralph, 2019). Due to this local data it was possible to achieve less than 9 km error at 24 hours using the previous generation of the ML model.

This model offers competitive performance, achieving 10.1 km error when using full dataset for training and about 11.5 km error when cross-validated. The model uses only iceberg velocities, gridded winds and currents and no additional information on size and shape so far.

It should be noted that the performance for all of the studies mentioned above was determined using different sets of iceberg tracks. Nevertheless, accurate data is the key to ML model performance. It is expected that a real-time stream of accurate local metocean data will significantly improve the reliability for iceberg management applications.

It is possible to further expand the model and formulate it using iceberg velocity relatively water to include water accelerations into account. As seen in Figure 2, the water accelerations are on the order of iceberg accelerations and will contribute in the form of the Froude-Krylov hydrodynamic force (Turnbull et al., 2015, Marchenko et al, 2019). In this case equations (2) and (3) will include an additional term related to water acceleration that will be determined from the ocean current forecast data.

In addition to the metocean inputs, size and shape information can be used in the ML part of the model. The 2019 dataset contained accurate information on size (measured by laser scanner and multibeam), and the iceberg waterline length will be used in the next generation of the model. The shape, however, has been hard to characterize. There have always been long discussions and disagreements whether a particular iceberg is wedged or blocky, pinnacled or dry dock. It is simply hard to describe the above water shape of and iceberg in one word.

The approach to forecasting performance metrics is nor trivial either. The forecasting error measured as distance between the forecasted and actual iceberg positions is not always a fair measure. For example, faster icebergs cover larger distances, and the corresponding forecasting error might become larger as a result. At the same time, the forecast usually predicts the direction for faster icebergs much better than for slower drifting icebergs. To tackle this effect, the forecasting error may be considered in relation to the distance traveled by the iceberg, similar to the relative performance index used by Andersson et al. (2018). But minimizing the average relative performance index will result in non-optimal forecasting error. Thus, the choice of the performance metric is determined by the priority for the forecast: more accurate drift direction or more accurate drift distance.

The model can be reformulated for longer-term forecasting too. It is expected that the RADARSAT Constellation Mission (RCM) will provide consistent twice-daily coverage of the region of interest. There is possibility to constantly track icebergs and generate a significant amount of data using automated satellite imagery analysis. A large number of iceberg tracks is a perfect case for ML application, even though those tracks are of low temporal resolution. Twelve hour resolution iceberg tracks are not able to capture the higher-frequency iceberg dynamics, but may be well correlated with lower-resolution circulation model inputs. Machine learning applied to satellite data may be used for longer-term forecasts exceeding short-term tactical scale and setting a bridge to upstream and seasonal forecasting time-scales.

Finally, the next step in the ML model development is reprogramming it into the form that will use the integral of the multi-step forecasting error during training instead of the single-step error. It is expected that it will significantly improve the model performance. Additional improvements will be the introduction of iceberg size to the training database, and further expansion of the training database via data collection and assimilation.

CONCLUSIONS

This study presents the next generation of ML iceberg drift model. The model is a hybrid that predicts combined acceleration caused by hydrodynamic force and wind drag based on iceberg velocities and metocean inputs from large-scale circulation models. A gradient-boosted trees implementation was used in the ML part of the model. The model was trained and tested using new iceberg track data collected during an iceberg tagging program offshore Labrador in 2019.

The training performance of the implicit implementation is 10.1±6.6 km of forecasting error in 24 hours. The cross-validated performance is 11.5±7.3 km of error for the same forecasting period. This is significantly lower than the average error level of the models that are currently operational for the Grand Banks (e.g. Murphy and Carrieres (2010) curve in Figure 4).

The model requires minimal data and no additional optimization once trained. This makes it a perfect candidate for operational applications. The model can be further developed in a number of ways including the addition of iceberg size information to the training data, switching to the multi-step cost function, and extending the forecast duration using satellite data products for training. Obviously, the model will benefit from expansion of the training database through data collection. Simultaneous measurements of drift tracks and local metocean data would be the most valuable.

ACKNOWLEDGEMENTS

The author is grateful to the United States Department of Homeland Security Office of Science and Technology under an Interagency Agreement with the United States Army Combat Capabilities Development Command C5ISR Center NVESD for permission to use the data. Additional thanks to current and former colleagues at C-CORE for supporting this work during all stages, from data collection to internal review.

REFERENCES

Andersson, L.E., Scibilia, F., Copland, L., Imsland, L., 2018. Comparison of statistical iceberg forecast models. *Cold Regions Science and Technology,* 155, pp.69–89. https://doi.org/10.1016/j.coldregions.2018.07.003

Andersson, L.E., Scibilia, F., Imsland, L., 2019. An iceberg forecast approach based on a statistical ocean current model. *Cold Regions Science and Technology*, 158, pp.128–142. https://doi.org/10.1016/j.coldregions.2018.11.016

Andersson, L.E., Scibilia, F., Imsland, L., 2016. An estimation-forecast set-up for iceberg drift prediction. *Cold Regions Science and Technology,* 131, pp.88–107. https://doi.org/10.1016/j.coldregions.2016.08.001

Briggs, R., Thibault, C., Mingo, L., King, T., 2020. Usage of UAVs for Surveying and Monitoring Icebergs. *Journal of Ocean Technology*, 15, pp.46–57.

Bruce, J., Yulmetov, R., King, T., Ralph, F., Younan, A., 2021. Development of iceberg profiling technology. In: *Proceedings of the ASME 2021 40th International Conference on Ocean, Offshore and Arctic Engineering, OMAE2021*. Virtual, Online.

Chassignet, E.P., Hurlburt, H.E., Smedstad, O.M., Halliwell, G.R., Hogan, P.J., Wallcraft, A.J., Baraille, R., Bleck, R., 2007. The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *Journal of Marine Systems, Marine Environmental Monitoring and Prediction,* 65, pp.60–83. https://doi.org/10.1016/j.jmarsys.2005.09.016

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics,* 29, pp.1189–1232.

International Ice Patrol, 2019. *Report of the International Ice Patrol in the North Atlantic. 2019 Season.* (No. CG-188-74).

Marchenko, A., Diansky, N. and Fomin, V., 2019. Modeling of iceberg drift in the marginal ice zone of the Barents Sea. *Applied Ocean Research*, *88*, pp.210-222.

Murphy, D.L. and Carrieres, T., 2010. CIS-IIP iceberg model inter-comparison. *Report prepared for the North American Ice Service*, p.27.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18. Curran Associates Inc., Red Hook, NY, USA, pp.6639–6649.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., Dool, H. van den, Zhang, Q., Wang, W., Chen, M., Becker, E., 2014. The NCEP Climate Forecast System Version 2. *Journal of Climate,* 27, pp.2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Turnbull, I.D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. *Cold Regions Science and Technology,* 110, pp.1–18. https://doi.org/10.1016/j.coldregions.2014.10.006

Yulmetov, R., Ralph, F., 2019. Machine learning for tactical iceberg drift forecasting. In: *The 25th International Conference on Port and Ocean Engineering under Arctic Conditions*, Delft, The Netherlands.