

We would like to thank the reviewers for their comments which helped us to improve the quality of the manuscript. Please find below our responses to the reviewer's comments.

Reviewer 3

#####

Title: Calibration of sea ice drift forecasts using random forest algorithms

Authors: Cyril Palerme and Malte Müller

This papers present short-term (1-10 days) forecasts of sea ice drifts using a random forest (AI) algorithm and a comparison of the AI forecasts with those of the operational ice-ocean prediction system TOPAZ. The models were trained using buoy or radarsat-derived sea-ice drifts. Predictors include short-term forecast ice speed and angle, wind speed and angle and ice thickness. Results show that the AI forecasts are more skillful than those of TOPAZ irrespective of the training data set. Furthermore, the model trained using sea ice buoys is more skillful in predicting sea ice drift for all lead-time when compared with the model trained with radarsat ice drifts.

The paper addresses an interesting question. The use of AI in sea-ice forecasting is relatively new and for this reason, this is a welcome contribution. The paper however is not well written, the introduction is succinct and does not place the work in the context of previous effectively, the model section is entirely missing and there is relatively little discussion of the pre-processing of the input data and its impact on the forecast skill (a factor that is at least equally important as the AI algorithm in producing a skillful model).

I recommend that the paper be accepted for publication after the comments below have been addressed (i.e. not rebutted).

Major Points:

1- The paper must be substantially edited/restructured.

I) The introduction is vague, there is a lot of name-dropping but it does not present an in-depth description of the previous work that is required to fully appreciate the content of the paper. I suggest that the authors review the literature more in-depth and revise the introduction substantially, or add a third co-author that works more closely in the field of sea ice forecasting.

We have completely changed the introduction in order to add a more in-depth description of previous works. We hope that the new introduction meets the expectations of the reviewer. The new introduction:

“Passive microwave observations of sea ice concentration have been available for more than 40 years, and have shown negative trends in Arctic sea ice extent since the beginning of the satellite era (e.g., Cavalieri and Parkinson, 2012; Comiso et al., 2017), with particularly strong trends during the summer (e.g., Comiso et al., 2017). There have been less satellite observations of sea ice thickness, and these retrievals have mainly been restricted to the winter due to issues related to surface melting during the summer (Ricker et al., 2017; Petty et al., 2020). Nevertheless, long-term negative trends in sea ice thickness have also been assessed by comparing retrievals from satellite altimeters (ICESat and CryoSat-2) with submarine measurements during the period 1958 - 2000 (Kwok and Rothrock, 2009;

Kwok, 2018). Furthermore, an acceleration of sea ice drift has been observed using drifting buoys and satellite observations (Rampal et al., 2009; Spreen et al., 2011; Tandon et al., 2018; Tschudi et al., 2020), and has been suggested as being a consequence of decreases in sea ice thickness and concentration due to reduced sea ice strength (Rampal et al., 2009; Olason and Notz, 2014; Tandon et al., 2018).

As a result of these changes, the Arctic ocean is becoming more accessible to marine operations, and there is an increase in maritime traffic (Eriksen and Olsen, 2018; Berkman et al., 2020). In order to ensure maritime safety, it is essential that accurate sea ice information is delivered to marine end-users. National ice services manually produce high-resolution sea ice charts using retrievals from various satellites such as passive microwave radiometers, optical instruments, and synthetic aperture radars (SAR). In addition to sea ice charts, short-term sea ice forecasts are also necessary for planning activities and providing up-to-date information to end-users. However, the spatial resolution of current sea ice models is often too coarse compared to user needs.

Short-term sea ice drift forecasts are operationally produced by numerical prediction systems, but are affected by biases despite the numerous efforts for improving the models (Hebert et al., 2015; Schweiger and Zhang, 2015; Rabatel et al., 2018; Williams et al., 2019). Hebert et al. (2015) evaluated sea ice drift speed forecasts from the U.S. Navy's Arctic Cap Now-cast/Forecast system. They found that the predicted ice drift speed was slower than drifting buoys in the summer months, and that a persistence forecast was generally better than the forecasts from the prediction system during the summer. In contrast, the forecasts produced by the U.S. Navy's Arctic Cap Nowcast/Forecast system outperformed persistence forecasts during the winter months. Schweiger and Zhang (2015) evaluated forecasts of sea ice drift speed from the Marginal Ice Zone Modeling and Assimilation System (MIZMAS) and found root mean square errors from 4.5 to 8 km per day for lead times of 1 and 9 days, respectively. These forecasts outperform a climatological reference for all lead times (up to 9 days). Sea ice drift forecasts from the neXtSIM-F system have been evaluated by Rabatel et al. (2018) and Williams et al. (2019), and root mean square errors of about 3 and 4 km per day have been reported for lead times of 1 and 4 days, respectively (Williams et al., 2019).

Sea ice drift is influenced by various sea ice characteristics such as concentration and thickness, as well as by near-surface wind and ocean currents (Rampal et al., 2009; Spreen et al., 2011; Olason and Notz, 2014; Yu et al., 2020). Though sea ice drift is mainly driven by the wind in areas with a low sea ice concentration, the relationships between these variables and sea ice drift are complex and not linear in most of the ice-covered areas (Yu et al., 2020). In order to improve the accuracy of sea ice drift forecasts, we have developed two calibration methods using random forest algorithms (Breiman, 2001), which is a supervised machine learning technique suitable for assessing nonlinear relationships between a set of predictor variables and a target variable.

While random forest methods have been widely used in sea ice remote sensing (Miao et al., 2015; Han et al., 2016; Lee et al., 2016; Gegiuc et al., 2018; Park et al., 2020), as well as in weather forecasting (Gagne et al., 2014; Ahijevych et al., 2016; Herman and Schumacher, 2018; Loken et al., 2019; Mao and Sorteberg, 2020), there has been less interest in using random forests in sea ice forecasting. Recently, Kim et al. (2020) developed and compared 1-month sea ice concentration forecasts based on random forests and convolutional neural networks. They obtained more accurate results using convolutional neural networks probably due to the larger learning capacity of convolutional neural networks compared to random forests, in particular to extract spatial features from the predictors (Kim et al., 2020). Furthermore, other machine learning and statistical methods have been used for sea ice forecasting, particularly for predicting the sea ice concentration and extent. Wang et al. (2019) used a

vector autoregressive model and a vector Markov model to predict sea ice concentration at subseasonal timescales, and obtained the best results using the vector Markov model. The vector Markov model also significantly outperformed the National Centers for Environmental Prediction (NCEP) Climate Forecast System, version 2 (NCEP CFSv2) for lead times between 2 and 6 weeks. Comeau et al. (2019) used a method based on analog forecasting for predicting Arctic sea ice area and volume anomalies at seasonal time scales, and obtained improvements compared to damped persistence forecasts. Moreover, various neural networks have been used for predicting sea-ice concentration, and found to be skillful for 1 and 12 month forecasts (Chi and Kim, 2017; Kim et al., 2020), but only slightly better than persistence forecasts for short-term prediction (Fritzner et al., 2020). Nevertheless, there has not been any attempt to calibrate short-term sea ice drift forecasts using advanced statistical methods.

The random forest models developed in this study are based on predictor variables from sea ice forecasts produced by the Copernicus Marine Environment Monitoring Service's (CMEMS) TOPAZ4 prediction system (Sakov et al., 2012), wind forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), and sea ice satellite observations from the Ocean and Sea Ice Satellite Application Facility (OSI-SAF). While all the models use the same predictor variables, two sets of models were developed using either drifting buoy displacements or SAR observations for the target variables. The data and methods used in this study are presented in sections 2 and 3, respectively. In section 4, the daily SAR observations used for analyzing the spatial variability of the forecast errors, as well as for training some of the random forest algorithms, are evaluated using buoy observations. Then, the performances of the calibrated forecasts are evaluated and compared to those from the TOPAZ4 forecasts in section 4. The discussion and conclusion of this study are presented in section 5.”

ii) The use of the present perfect to describe the data is odd. “Satellite sea-ice drift observations ... have been used...” (Line 47, 56, etc.). It sounds as though the authors are speaking of previous work by other authors when they are speaking of their own work being presented. The use of the present tense is much more engaging for the reader, or at least the simple past. I.e. We use (used) satellite sea-ice drifts observations...” . These are two examples; there are many more in the paper.

We agree with this comment and we have replaced the present perfect by the preterit many times in the revised version of the paper.

iii) Line 67: “Section 2.2: Data used for the predictor variable”. The model used as a reference for the evaluation of the AI models (i.e. TOPAZ) is included here, yet it is not a predictor variable. The forecasted sea ice thickness, concentration, drift speed and angle are all predictor variables but are not described in this section. Only the 10-m wind speed is discussed.

This section was initially about the data sets used in this study, and not about the predictor variables. However, we have decided to describe the predictor variables in this section in the revised version of the paper:

“The list of predictor variables is the same for all the models developed in this study, and can be divided into three different categories. First, some geographical information is used with the Cartesian coordinates of the grid points (x and y in the stereographic projection from the TOPAZ4 system), and the distance of the grid point to the nearest coastline in the TOPAZ4 system. Then, the sea ice concentration from passive microwave observations during the day preceding the forecast start date is also used as predictor variable. The variables from sea ice and wind forecasts during the predicted lead time can be considered as the last category. These variables are the wind direction and speed from

ECMWF forecasts, as well as the sea ice concentration, thickness, drift speed and direction from TOPAZ4 forecasts. Furthermore, the sea ice drift and concentration observations, as well as ECMWF wind forecasts, were projected onto the grid used in the TOPAZ4 prediction system using nearest-neighbor interpolation before developing the random forest models.”

iv) TOPAZ is described only very succinctly. It does not say which sea ice model is used, whether there is an ice thickness distribution included, the grid on which the equation are solved, etc.

We agree with this comment and we have added a more detailed description of TOPAZ4 in section 2.2 of the revised version of the paper:

“TOPAZ4 is a coupled ice-ocean model for the North Atlantic and the Arctic which provides 10-day forecasts at a spatial resolution of 12.5 km, as well as a reanalysis (Sakov et al., 2012). It uses the version 2.2 of the Hybrid Coordinate Ocean Model (HYCOM; Bleck, 2002; Chassignet et al., 2006) coupled with a one thickness category sea ice model using an elastic-viscous-plastic rheology (Hunke and Dukowicz, 1997) derived from the version 4.1 of the Community Ice Code (CICE). The model native grid created using conformal mapping has a spatial resolution between 12 and 16 km in the whole domain. An ensemble Kalman filter is used to assimilate satellite sea ice and oceanic observations such as sea ice concentration and drift, along-track sea-level anomalies, sea-surface temperature, as well as in-situ temperature and salinity profiles. Moreover, TOPAZ4 is forced by ECMWF high-resolution weather forecasts at the ocean surface. While TOPAZ4 forecasts are produced daily, data assimilation is only performed on Thursdays, and only the forecasts starting on Thursdays are stored in the long-term archive. Though the TOPAZ4 system provides forecasts with hourly time steps, the forecasts with daily outputs were used here due to the 24-hour span of SAR observations. Previous studies have reported that the speed of sea ice drift is overestimated in the TOPAZ4 system compared to buoy observations from the IABP (Sakov et al., 2012; Xie et al., 2017)”

2- The model is section is entirely missing. A mathematical description of the random forest model must be given because AI is relatively new in the field of short-term sea ice forecasting and more simply for the sake of completeness. The reader should not have to read other papers about random forest in order to fully appreciate the content of the current work.

We agree that the description of the random forest technique was missing, and we have added a longer description of the random forest technique in the section 3.1 of the revised version of the paper:

“Random forest algorithms consist of an ensemble of decision trees used for regression or classification tasks (Breiman, 2001). In order to avoid overfitting (meaning that the models learn from noise in the training data), independent decision trees must be developed. The independence of decision trees is ensured by using different subsets of the training data set for developing each decision tree, as well as by randomly selecting a fraction of the predictor variables at each node (the node is then split using the variable maximizing a dissimilarity metric among the selected predictors). Each decision tree is trained with a data set created using the bootstrap method, which consists of randomly selecting samples from the original training data with replacement for creating a new data set of the same size as the original one. This results in using about 63 % of the samples from the original data set for training each decision tree.

In this study, random forest models were developed for regression using the Python library Scikit-learn-0.23.2 (Pedregosa et al., 2011), and the mean squared error was used to measure the quality of the splits.”

3- Some pre-processing was done to the data. E.g. the authors used speed and angle rather than latitudinal and meridional components; two different models for speed and angle were proposed. All these decisions leads to improvements in the forecast. Was there any more pre-processing done to the data to improve skill? What was the improvement in the forecast skill using these pre-processing techniques? A few sentences should be included in the discussion about this in section 4.3. I would call this section “Pre-processing of the data”.

The reason why we develop different models for predicting the speed and direction of sea ice drift is that random forest algorithms can only be used to predict one variable. We hope that the new section describing the random forest technique will help to understand this better.

We have added the following statement at the beginning of the new section 3.3 called “Pre-processing of the data”:

“In order to avoid overfitting, it is better to use predictor variables that are not highly correlated. This is why the speed and direction of sea-ice drift, as well as the wind speed and direction, have been used as predictor variables instead of the eastward and northward components. ”

Furthermore we have added the following paragraph at the end of the same section:

“We also tested random forest models predicting the sea ice drift along the x and y axes of the TOPAZ4 grid using a different set of predictor variables (figure S12 of the supplementary material). For these models, the northward and eastward components of the ECMWF wind forecasts were used as predictors instead of the wind speed and direction, as well as the sea ice drift along the x and y axes from TOPAZ4 forecasts (which are provided by TOPAZ4 outputs) instead of the sea ice drift speed and direction. The direction and speed of sea ice drift were then calculated using the start and end location of the sea ice for comparing those models with the ones directly predicting the direction and speed of sea ice drift. Relatively similar performances were achieved by these models for predicting the direction of sea ice drift, but these models had significantly worse performances for predicting the speed of sea ice drift (larger mean absolute errors of about 12.2 % and 13.7 % on average for the models trained with buoy and SAR observations, respectively). ”

4- Line 215: A model trained within the Arctic Ocean proper should not be used to predict sea-ice drift in the land-lock sea ice of the Canadian Arctic Archipelago. This is an entirely different dynamical regime. This results and associated discussion should be removed from the paper and from the abstract. Or at least not given such an important presence.

We agree with this comment, though random forest algorithms could, in principle, be able to identify regions with particular conditions with the spatial coordinates. Nevertheless, we have decided to exclude the Canadian Arctic Archipelago in the revised version of the paper, and we have added the following statement:

“The Canadian Arctic Archipelago is excluded from our study due to the different characteristics of sea-ice drift in this region (largely influenced by the presence of narrow channels and landfast ice) compared to the rest of the Arctic. Therefore, no data located in the Canadian Arctic Archipelago were used for training and evaluating the random forest models.”

Minor Points:

Line 13-14: Sea ice conditions in the Arctic do not change increasingly faster because of increase in ice drift speed. Increase sea ice drift speed is one such change associated with arctic climate change, but it is not the cause. The cause is thinning of sea ice associated

with warmer air temperature, change in cloud phase and its impact on the radiative fluxes at the surface, increased ocean heat flux that interacts closely with sea ice on the shallow arctic shelves, increased storminess in the Arctic, etc

We have modified the introduction and this statement has been removed.

Line 61: Why only use sea ice drift speed lower than 5km per day? The mean speed in the Arctic Ocean is 5km /day or ~5cm/sec. It seems that a large amount of data is being ignored without acknowledging it or without providing a rationale for doing so.

It seems that there has been a misunderstanding here. In the discussion paper, it was written (lines 59-62):

“While all buoy observations located in an area with a sea-ice concentration higher than 10 % (in the OSI-SAF product described in the next section) were used for training the random forest algorithms, only the buoys with a speed between 0.5 and 100 km per day, located in an area with a sea-ice concentration higher than 10 %, and further than 50 km from the coastlines were used for verification.”

Therefore, we have never excluded the buoy observations with a speed higher than 5 km / day, but only the buoy observations with a speed lower than 0.5 km / day and higher than 100 km / day. In the revised version of the paper, we have changed the threshold of 0.5 km / day to 0.1 km / day. However, we agree that the fraction of observations excluded by this selection was missing, and we have added the following statement in the revised version of the paper:

“In order to avoid inaccurate and unrealistic values, only the buoys with a speed between 0.1 and 100 km per day, located in an area with a sea ice concentration higher than 10 %, and further than 50 km from the coastlines were used for verification. While only the buoys with a speed between 0.1 and 100 km per day were used for training the random forest models predicting the direction of sea ice drift, all the buoys with a speed lower than 100 km per day were used for training the models predicting the speed of sea ice drift in order to make them able to predict very low speed. During the period from June 2013 to May 2020, about 4.5 % and 0.1 % of the buoys had a speed lower than 0.1 km per day and higher than 100 km per day, respectively. ”

Line 63: “...have been projected onto the grid used in the TOPAZ4 system”. This is not useful information. What grid is used in TOPAZ4? Tri-polar? Curvi-linear? Cube-sphere? I see now that this has been defined later in the paper on Line 103. The grid must be defined when it is first discussed. Is it a Cartesian grid? Or Lat/Lon?

We have added the projection in the following sentence: *“The drift vectors from buoy observations were then projected onto the polar stereographic grid used in the TOPAZ4 system.”* However, we have described the other information in section 2.2 in the description of the TOPAZ4 prediction system (see our response to major point iv).

Line 79: Which ocean observations are assimilated?

We have added this information in the revised version of the paper:

“An ensemble Kalman filter is used to assimilate satellite sea ice and oceanic observations such as sea ice concentration and drift, along-track sea-level anomalies, sea-surface temperature, as well as in-situ temperature and salinity profiles.”

Line 86: When did the switch to higher resolution happened?

We have changed this sentence in the revised version of the paper: *“These forecasts have lead times up to 10 days, and the model’s spatial resolution changed from about 16 km to 9 km in March 2016 (<https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model>).”*

Line 95: No new paragraph here. “... where R is the Earth’s radius, lamda and phi are the...”

We agree with this comment and we have modified this. The new sentence:
“where arctan2 represents the 4-quadrant inverse tangent function, R is the Earth’s radius, φ and λ represent the latitude and the longitude, and the subscripts "start" and "end" indicate the start and end locations”

Equ 4: Unusual notation. arctan(v/u)?

It is true that “arctan2” was not defined in the text. We have added the following statement in the text:
“where arctan2 represents the 4-quadrant inverse tangent function”.

Line 121: Should it be “data points” instead of “data sets”?

We agree with this comment, and we have replaced “data sets” by “data points”.

Line 165, Equ. 5: Why Case #3 in Equ. 5? Don’t Case #1 and #2 above cover all cases?

There are also cases where the difference between two directions is between -180 and 180°. However, there was an error in the discussion paper (but not in the analysis). We have corrected this error in the revised version of the paper. When $\Delta D > 180 \Rightarrow \text{Error} = \Delta D - 360$ (and not $360 - \Delta D$). Note that this error only affected the direction error, but not the absolute error.

Line 169-171: This is “Method” material that was already covered earlier. It should be moved to the method section.

We have removed this statement in the revised version of the paper.

Line 191. “Moreover the fraction of forecasts improved by the calibration is, on average, larger for the models trained with buoy observations (57.0 %) than for the models trained with SAR observations (54.8 %)”. Is this really statistically significant? Errors are provided throughout the paper but it does not transpire in the discussion. The errors should used to assess whether the improvements are significant or not.

We have added an analysis of the statistical significance using the Wilcoxon signed-rank test, which is suitable for non-parametric data and paired observations (see paragraph below). Note that we have used the Wilcoxon signed-rank test to assess if the difference in absolute errors are significant. The fraction of forecasts improved does not have any statistical distribution, and it is therefore more difficult to assess the statistical significance for this metric.

We have added the following paragraph in the method section:

“In this study, we used the Wilcoxon signed-rank test to assess the statistical significance of the differences between the absolute errors due to its suitability for non-parametric data (the absolute errors are not normally distributed) and paired observations (the same data set was used for evaluating the different models). We performed this analysis using the two-tailed hypothesis test and the significance level of 0.05.”

And we have describe the statistical significance of the results in the section “4.2 Evaluation of the calibrated forecasts”:

“The performances of the calibrated forecasts have been evaluated and compared to those from the TOPAZ4 prediction system during the period from June 2020 to May 2021 using buoy observations (figures 3). For predicting the direction of sea ice drift, the models trained with buoy observations significantly outperform the TOPAZ4 prediction system and the models trained with SAR observations for all lead times, except 10 days. On average, the calibrated forecasts produced by these models have a mean absolute error about 8.0 % lower than TOPAZ4 forecasts. The models trained with SAR observations significantly outperform the TOPAZ4 prediction system for lead times up to 5 days, and reduce the mean absolute errors by 3.3 % compared to TOPAZ4 forecasts. However, the TOPAZ4 prediction system slightly outperform the models trained with SAR observations for lead times from 8 to 10 days, though the differences are not statistically significant. Moreover, the fraction of forecasts improved by the calibration is, on average, larger for the models trained with buoy observations (55.7 %) than for the models trained with SAR observations (52.9 %). Furthermore, the correlation between the forecasts and the buoy observations is improved by both calibration methods for lead times up to 7 days, and deteriorated for longer lead times.

For the speed of sea ice drift, the models trained with buoy observations have the best performances for all lead times. They significantly outperform the TOPAZ4 system and the models trained with SAR observations for all lead times, except 4 days for which the difference with the TOPAZ4 system is not statistically significant. The forecasts from the models trained with SAR observations have slightly larger mean absolute errors than TOPAZ4 forecasts for lead times up to 5 days, but significantly outperform TOPAZ4 forecasts for longer lead times. On average, the mean absolute error is reduced by 7.1 % and 2.5 % by the calibration for the models trained with buoy and SAR observations, respectively. The fraction of forecast improved is, on average, slightly larger for the models trained with buoy observations (53.4 %) than for the models trained with SAR observations (53.1 %). Moreover, the correlation between the buoy observations and the forecasts is improved by both calibration methods.

The spatial variability of the fraction of forecasts improved by the calibration has been analyzed using SAR observations as reference in order to use as many observations as possible (figures 4, 5, 6, 7), though the grid points with less than 20 SAR observations during the period from June 2020 to May 2021 have been excluded from this analysis. The number of SAR observations per grid cell used for this comparison has been mapped in figure 1 d). Overall, both calibration methods perform relatively well for predicting the direction of sea ice drift in the Central Arctic for lead times up to 5 days (figures 4 and 5). However, the fraction of forecasts improved decreases with increasing lead times, and both calibration methods have relatively poor performances in the Beaufort, Chukchi, and East Siberian seas. Furthermore, the models trained with buoy observations perform better than the models trained with SAR observations in most of the area taken into account in this analysis.

For the speed of sea ice drift, the models trained with SAR observations perform better than the models trained with buoy observations in most of the area analyzed. The models trained with buoy observations have particularly poor performances compared to TOPAZ4 near the Greenland and Russian coastlines (figure 6), while the models trained with SAR observations perform better in these

areas (figure 7). It is worth noting that most of the buoys taken into account for evaluating the forecasts in figure 3 are not located in the areas where the models trained with buoy observations have poor performances, which likely explains the better performances of the models trained with buoy observations compared to the models trained with SAR observations in figure 3.”

Line 197: “The fraction of forecast improved is, on average, slightly larger for the models trained with SAR observations (55.3 %) than for the models trained with buoy observations (54.9 %).” Again, is this statistically significant?

We have answered to this comment in our previous response.

Line 222: The fraction of data used in the training and validation of the model belongs to the Method section.

We have moved this statement in the Method section in the revised version of the paper.

Line 225-230: Repetitive. This was already mentioned in the Method section.

We have moved this section in the Method section in the revised version of the paper.

Line 236: Sea ice thickness does not change very much in 10 days. I suspect the ice thickness at $t=0$ would be equally skillful. This should be mentioned.

We agree that using sea ice thickness during the initialization of the forecasts should provide a relatively similar information to the algorithms. However, because Pan-Arctic sea ice thickness observations are not available during the summer, it is not possible to use sea ice thickness observations in our random forest models which are used all year round. Therefore, we consider that the best available information is the sea ice thickness forecasts at the predicted lead time, and we do not think that the low temporal variability of sea ice thickness should be mentioned here.

Section 4.3: The discussion does not present a quantitative assessment of the predictive skill of each predictor. A more quantitative discussion should be provided.

We agree with this comment, and we have modified this section to present the results more quantitatively. The new section:

“For both calibration methods, the most important variable for predicting the drift direction is the sea ice drift direction from TOPAZ4 forecasts, followed by the wind direction from ECMWF forecasts (figure 8). On average, the relative importance of sea ice drift direction forecasts is about 1.4 and 1.5 times larger than the one from wind direction forecasts for the models trained with buoy and SAR observations, respectively (figure 8). The sum of the relative importances of these two variables represent, on average, about 46 and 41 % of the sum of all relative importances for the models trained with buoy and SAR observations, respectively. However, the relative importances of these two variables decrease with increasing lead times.

Similarly, the sea ice drift speed from TOPAZ4 is the most important variable for predicting the speed of sea ice drift, followed by the wind speed from ECMWF forecasts. On average, the relative importance of sea-ice drift speed forecasts is about 1.7 and 2.2 larger than the one from wind speed forecasts for the models trained with buoy and SAR observations, respectively (figure 8). For the models predicting the speed of sea ice drift, the sum of the relative importances of these two variables represent, on average, about 40 % of the sum of all relative importances for both calibration methods.

Furthermore, the relative importances of these two variables also decrease with increasing lead times. On average, the mean absolute errors are reduced by all predictors for the direction and speed of sea ice drift in both calibration methods (figure 9), though some predictor variables do not improve the forecast accuracy for all lead times. While the sea ice concentration observations during the initialization of the forecasts and the sea ice concentration forecasts from TOPAZ4 are correlated, removing one of these variables decreases the accuracy of most random forest models. Therefore, we decided to keep both variables, even if the importances of these variables are probably underestimated due to this correlation. Furthermore, we also tested using the day of year as an additional predictor variable (figure S7 of the supplementary material), but adding this variable tends to deteriorate the forecast accuracy for most models, so we decided to discard this variable.

For the models predicting the direction of sea ice drift, removing the drift direction from TOPAZ4 forecasts increases the mean absolute error between 1.1 and 6.7 degrees depending on the lead time and the observations used for the target variable. This is much larger than the differences in mean absolute error when the wind direction from ECMWF forecasts is removed (between 0.1 and 2.2 degrees). For the models predicting the speed of sea ice drift, removing the drift speed from TOPAZ4 forecasts increases the mean absolute error between 0.041 and 0.444 km / day depending on the lead time and the observations used for the target variable. This is also much larger than the differences in mean absolute error when the wind speed from ECMWF forecasts is removed. Surprisingly, removing the wind speed forecasts slightly reduces the mean absolute error (difference of 0.005 km / day) for the model predicting the speed of sea ice drift for a lead time of 4 days trained with SAR observations. For the other models predicting the speed of sea ice drift, removing the wind speed forecasts increases the mean absolute error between 0.001 and 0.127 km / day. Furthermore, the mean absolute errors for the speed of sea ice drift are also considerably reduced by adding the sea ice thickness forecasts from TOPAZ4 (between 0.011 and 0.098 km / day), probably due to the anti-correlation between sea ice thickness and sea ice drift speed (Yu et al., 2020).”

Figure 1: Colorbar for the d panel should be changed to avoid saturation.

The colorbar has been changed.

Figure 4: Units for sea ice drift should be km/day or ideally cm/sec. It should not be m/day.

We have changed the unit for sea ice drift speed in the revised version of the paper, and km / day is now used.

Bruno Tremblay
McGill University