Supplement of The Cryosphere, 15, 3989–4004, 2021
https://doi.org/10.5194/tc-15-3989-2021-supplement

The Cryosphere Open Access EGU

Supplement of

Calibration of sea ice drift forecasts using random forest algorithms

Cyril Palerme and Malte Müller

Correspondence to: Cyril Palerme (cyril.palerme@met.no)

1 Sensitivity tests

For the sensitivity tests (figures S1, S3, S4, S5, and S7), the random forest models were trained using data from about 80 % of the forecast start dates (randomly selected) within the training periods. Then, the data from the remaining forecast start dates were used for evaluating the forecast performances. This selection prevents using neighboring grid points with very similar conditions in the training and validation data sets, and was repeated 10 times in order to obtain robust results. Furthermore, the algorithms were evaluated using the same product as the one used for training the random forest models (CMEMS SAR MOSAIC product for those trained with SAR observations, and IABP buoys for those trained with buoy observations).

2 Bias correction of TOPAZ4 forecasts

In order to compare the random forest models developed in this study with more simple calibration methods, we have developed calibrated forecasts by correcting the biases from TOPAZ4 forecasts. The biases from TOPAZ4 forecasts have been evaluated for each grid point and each lead time during the period 2018 - 2019 using SAR observations as reference (figures S9, S10, and S11). Only the grid points containing at least 20 SAR observations during the period 2018 - 2019 have been used for this calibration and for the evaluation presented in figure S12. For the direction of sea ice drift, the bias from the period 2018 - 2019 has been subtracted from the TOPAZ4 forecasts. For the speed of sea ice drift, the TOPAZ4 forecasts have been multiplied by the ratio of the SAR observations over the TOPAZ4 forecasts during the period 2018 - 2019.

3 Random forest models using only three predictors (figure S12)

Random forest models using only three predictor variables have been developed and compared to the other calibration methods. Only the x and y coordinates, as well as the drift direction from TOPAZ4 have been used for the models predicting the direction of sea ice drift. For the models predicting the speed of sea ice drift, the drift speed from TOPAZ4 has been used with the x and y coordinates. Note that the number of predictors randomly selected at each node has been fixed at two for these random forest models.

4 Random forest models predicting sea ice drift along the x and y axes of TOPAZ4 grid

We developed 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). 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 locations of the sea ice for comparing those models with the ones directly predicting the direction and speed of sea ice drift.

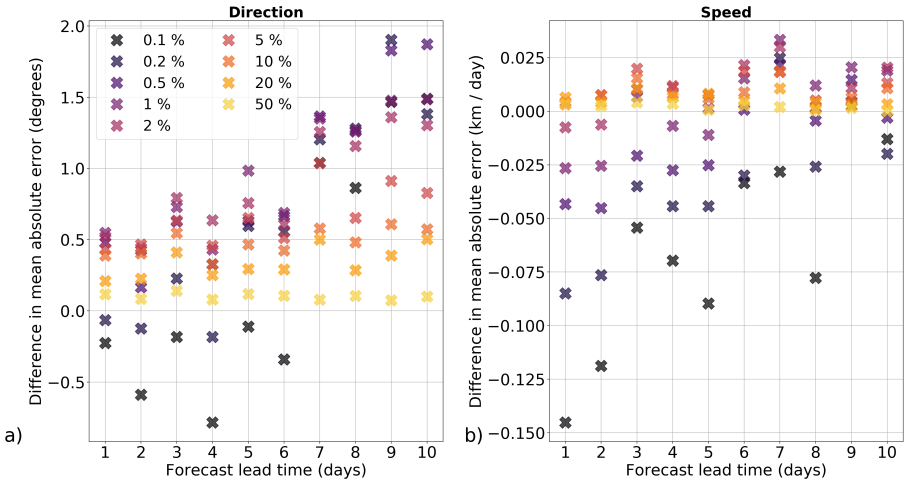

Figure S1. Difference in mean absolute errors between the algorithms trained using all the grid points with SAR observations and the algorithms trained using only a fraction of the grid points with SAR observations for the direction (a) and the speed (b) of sea ice drift. SAR observations have been used as reference. Positive values mean that the algorithms using only a fraction of the available grid points for training outperform the algorithms using all the available grid points. The legend from figure a) shows the fraction of grid points randomly selected for training the random forest algorithms.

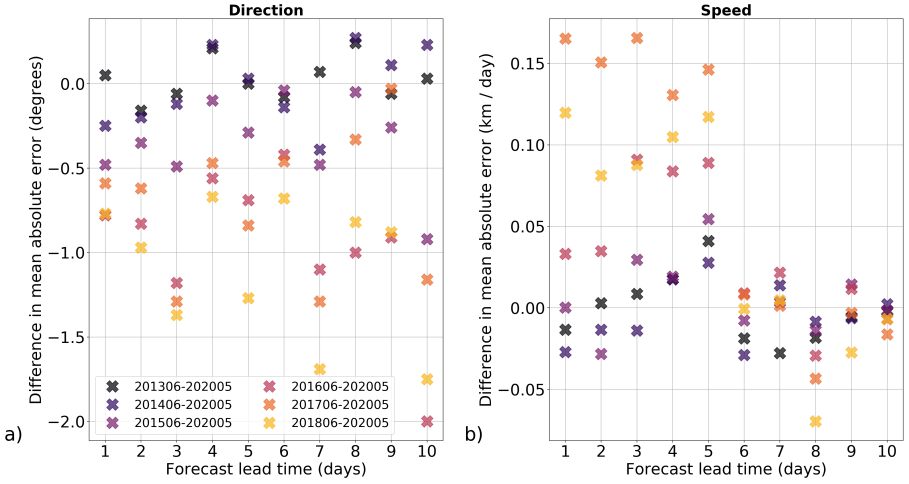

Figure S2. Difference in mean absolute errors during the period June 2020 - May 2021 for the direction (a) and the speed (b) of sea ice drift between the models trained with the period June 2012 - May 2020 and the models trained with the period indicated in the legend of figure a). Buoy observations have been used as reference. Positive values mean that the models trained with the period indicated in the legend outperform the models trained with the period June 2012 - May 2020. The legend from figure a) shows the different training periods.

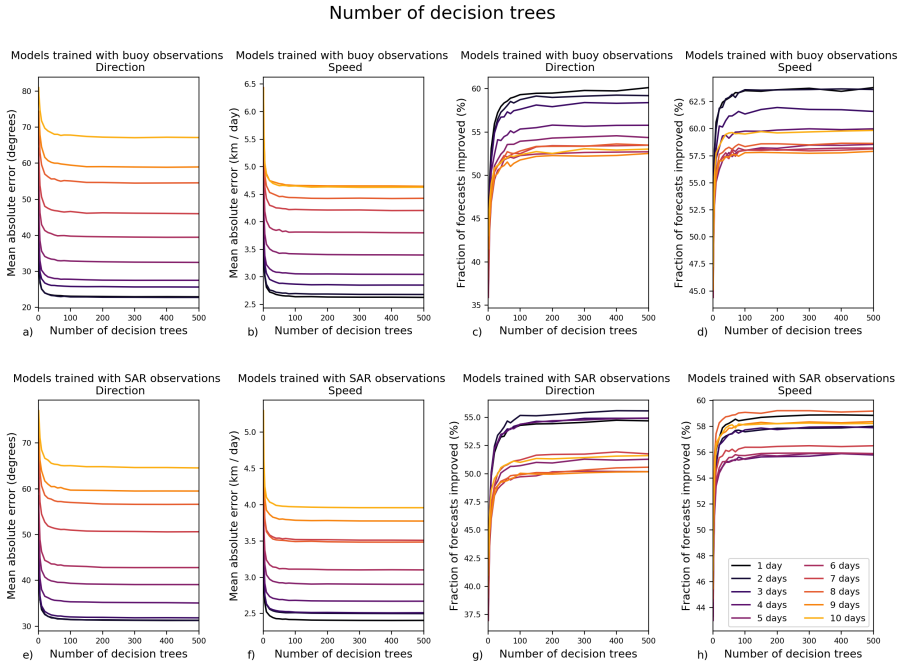

Figure S3. Performances of the random forest models depending on the number of decision trees used in the algorithms. The lead times are indicated in the legend of figure h).

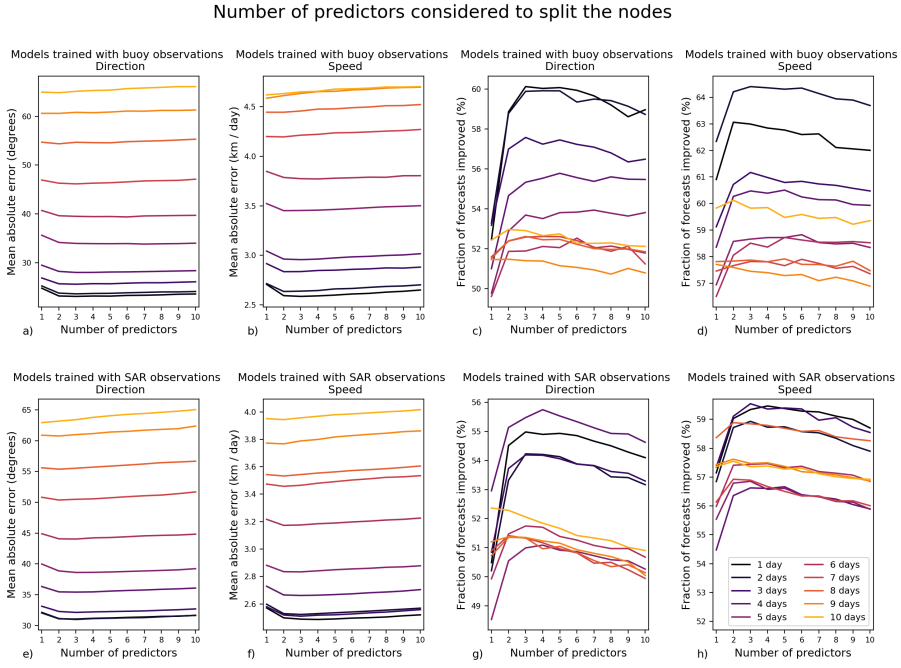

Figure S4. Performances of the random forest models depending on the number of predictor variables considered to split the nodes. The lead times are indicated in the legend of figure h).

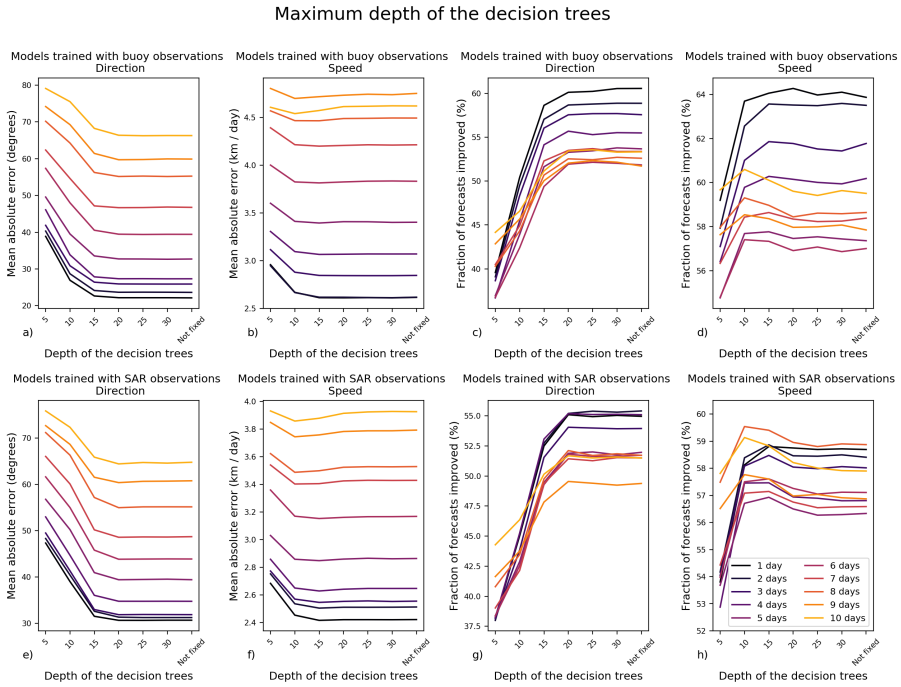

Figure S5. Performances of the random forest models depending on the maximum depth of the decision trees. The lead times are indicated in the legend of figure h).

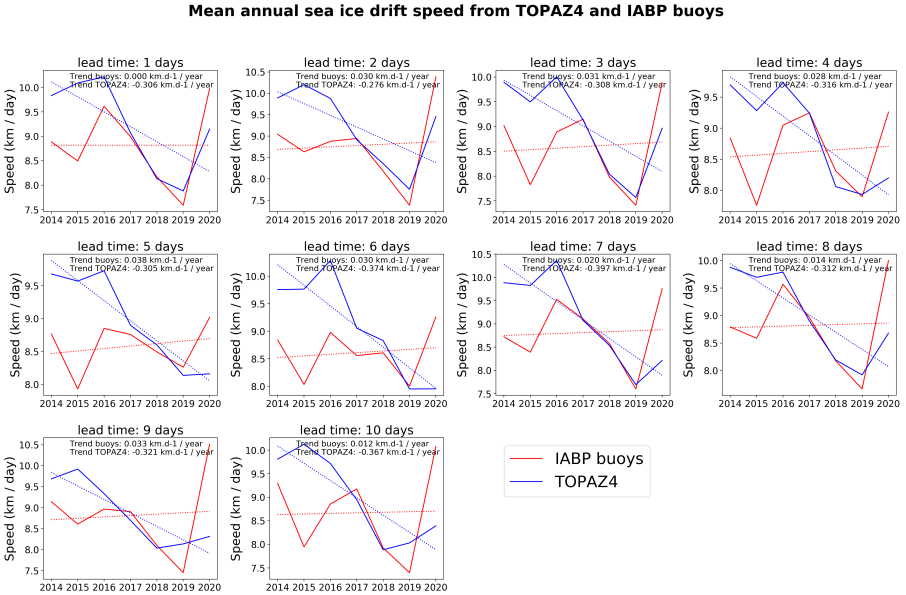

Figure S6. Mean annual sea ice drift speed from collocated IABP buoy observations and TOPAZ4 forecasts. The solid lines show the mean annual sea ice drift speed (km / day) from buoy observations and TOPAZ4 forecasts. The dashed lines show the linear trends. The annual sea ice drift speeds have been calculated from monthly mean values.

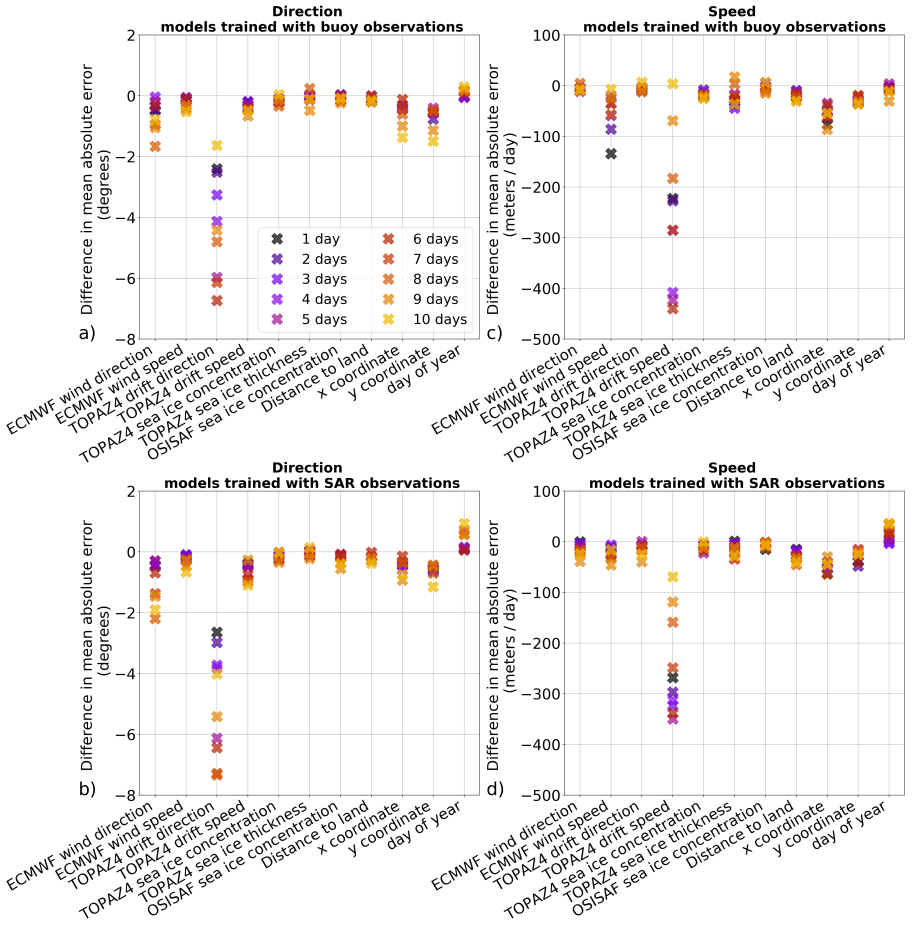

Figure S7. Same as figure 9, but with the day of year as an additional predictor variable.

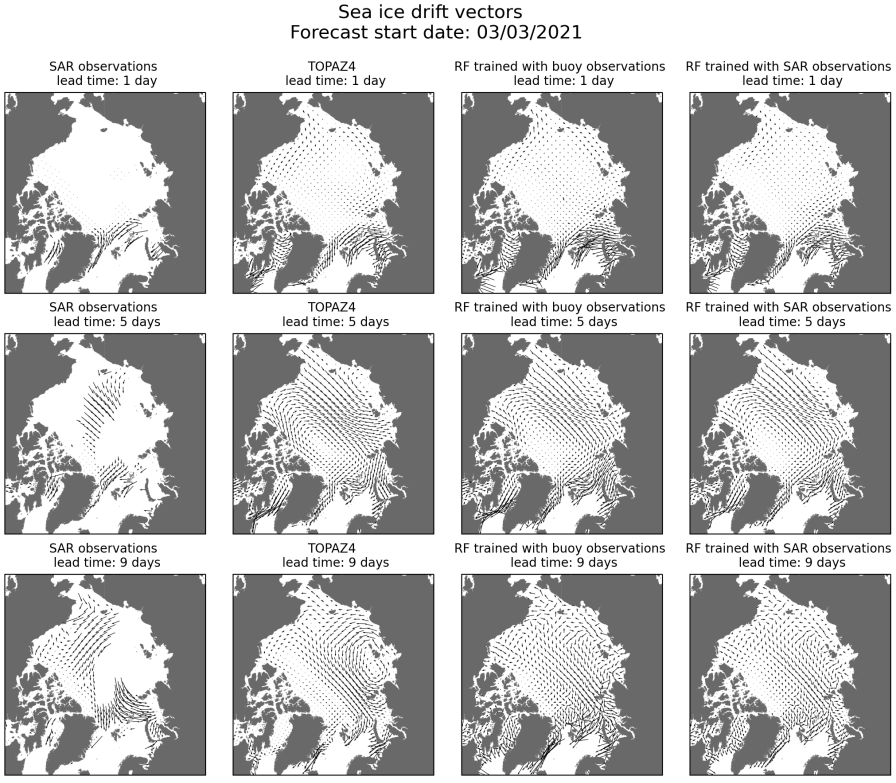

Figure S8. Example of calibration with the random forest (RF) algorithms for the forecasts which started on 03/03/2021 and for lead times of 1, 5, and 9 days.

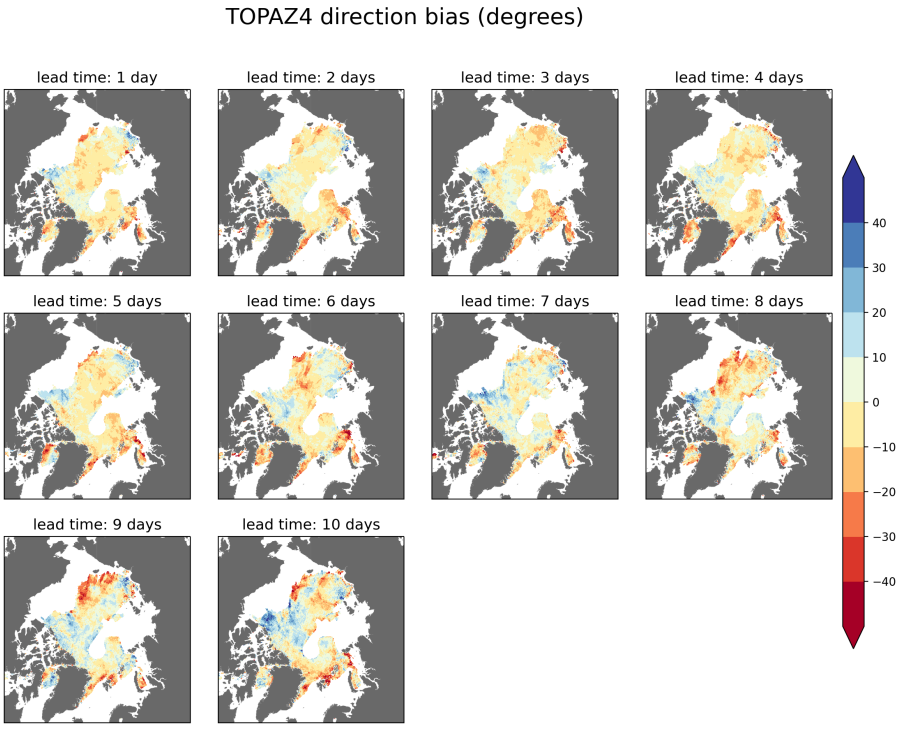

Figure S9. Mean direction bias (degrees) from TOPAZ4 forecasts during the period 2018 - 2019 compared to SAR observations. Only the grid points containing at least 20 SAR observations during the period 2018 - 2019 have been taken into account.

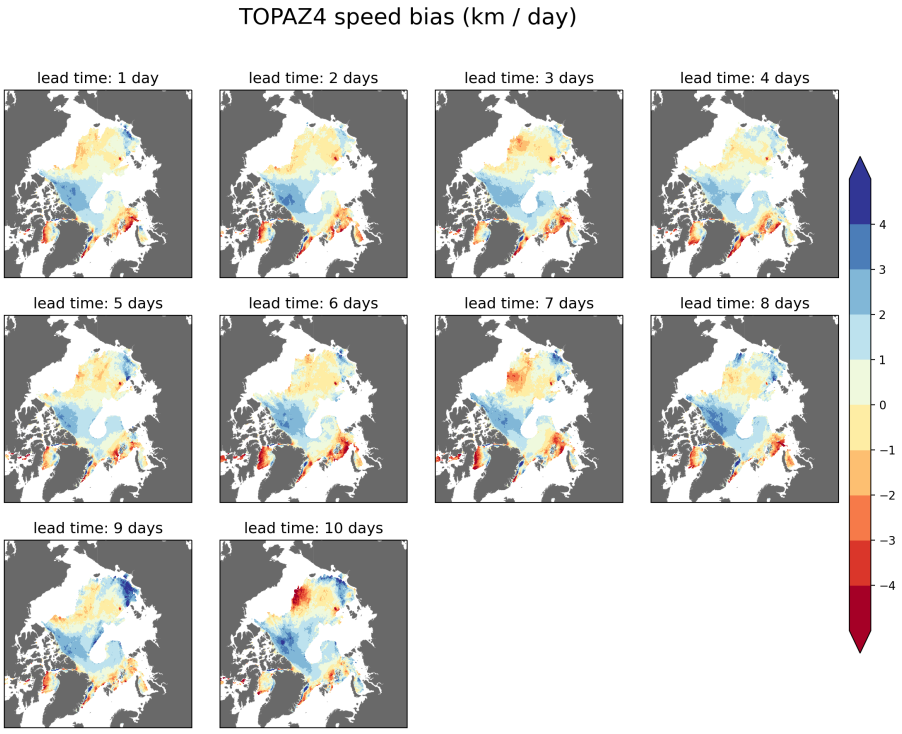

Figure S10. Mean speed bias (km / day) from TOPAZ4 forecasts during the period 2018 - 2019 compared to SAR observations. Only the grid points containing at least 20 SAR observations during the period 2018 - 2019 have been taken into account.

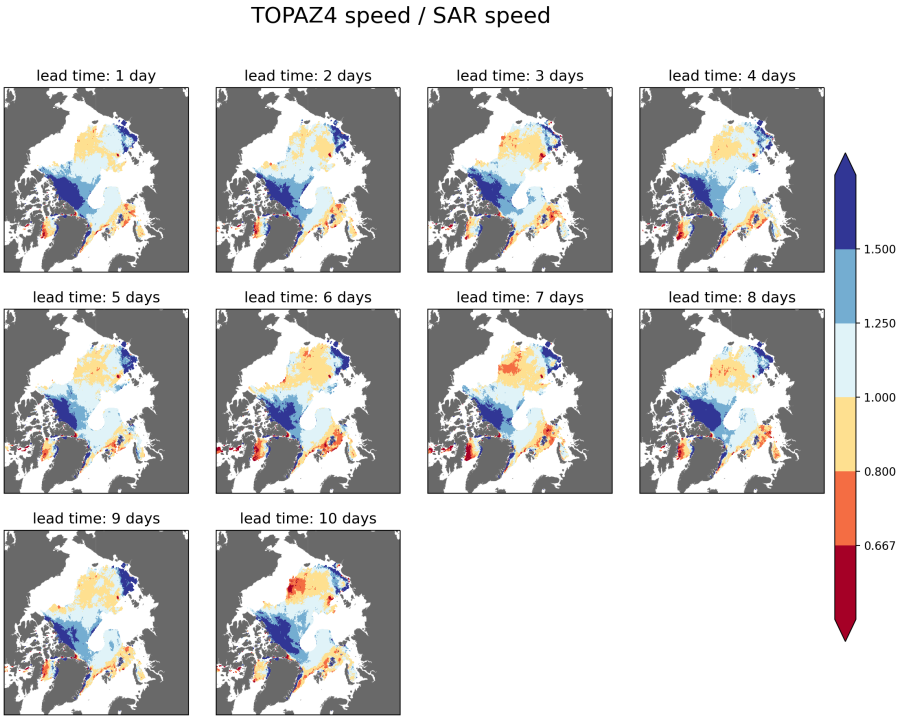

Figure S11. Ratio of the drift speed from TOPAZ4 forecasts over the drift speed from SAR observations during the period 2018 - 2019. Only the grid points containing at least 20 SAR observations during the period 2018 - 2019 have been taken into account.

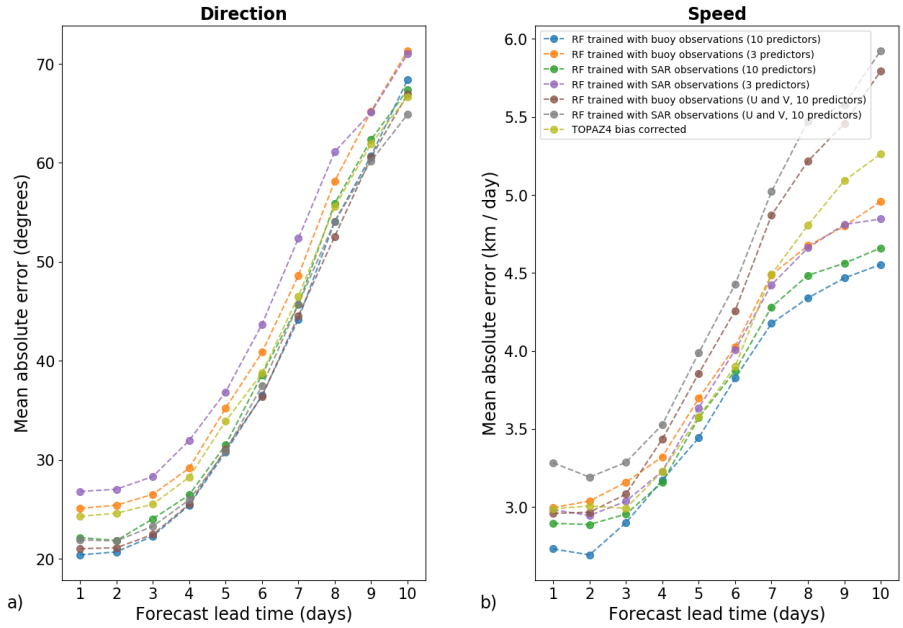

Figure S12. Mean absolute errors of different calibration methods for the period June 2020 – May 2021. Buoy observations have been used as reference. The random forest (RF) models using all the predictors (10 predictors) described in the main paper are shown by the blue and green curves. The random forest models using only 3 predictors (x and y coordinates, as well as the drift direction from TOPAZ4 for the models predicting the direction, and the drift speed from TOPAZ4 for the models predicting the speed of sea ice drift) are shown by the orange and purple curves. The TOPAZ4 forecasts which are bias corrected (using the period 2018-2019 for calculating TOPAZ4 biases) are shown by the yellow curves. The random forest models predicting the sea ice drift along the x and y axes of TOPAZ4 grid are shown by the brown and gray curves.

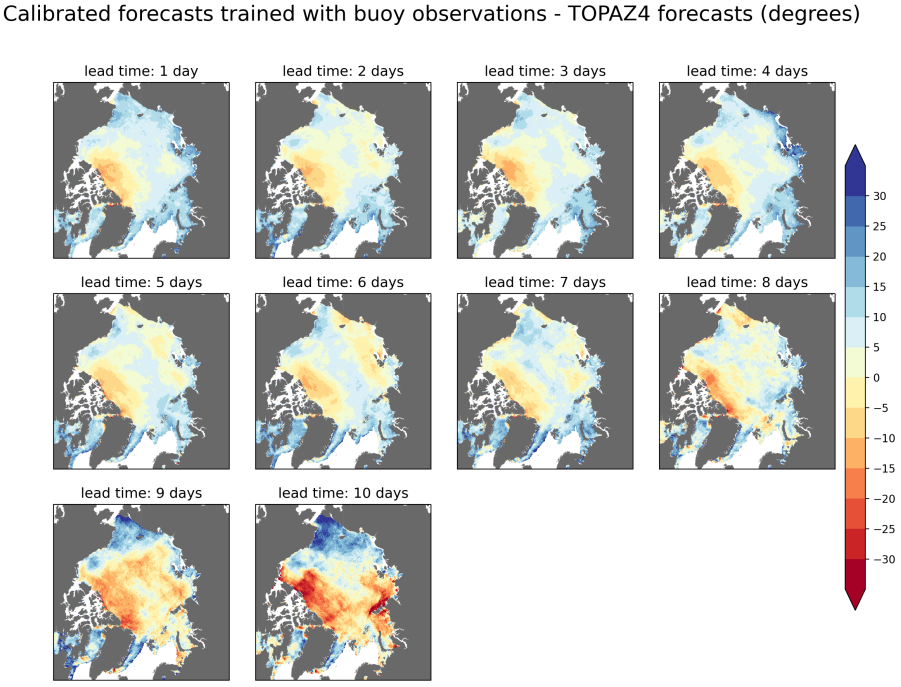

Figure S13. Difference between the random forest models trained with buoy observations and the TOPAZ4 forecasts for the direction of sea ice drift (degrees) during the period June 2020 - May 2021.

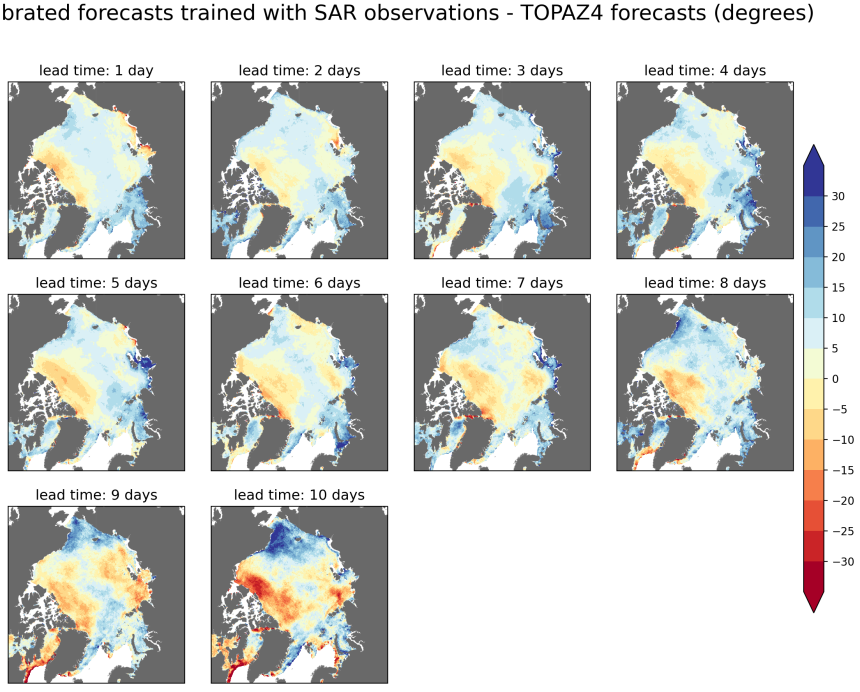

Figure S14. Difference between the random forest models trained with SAR observations and the TOPAZ4 forecasts for the direction of sea ice drift (degrees) during the period June 2020 - May 2021.

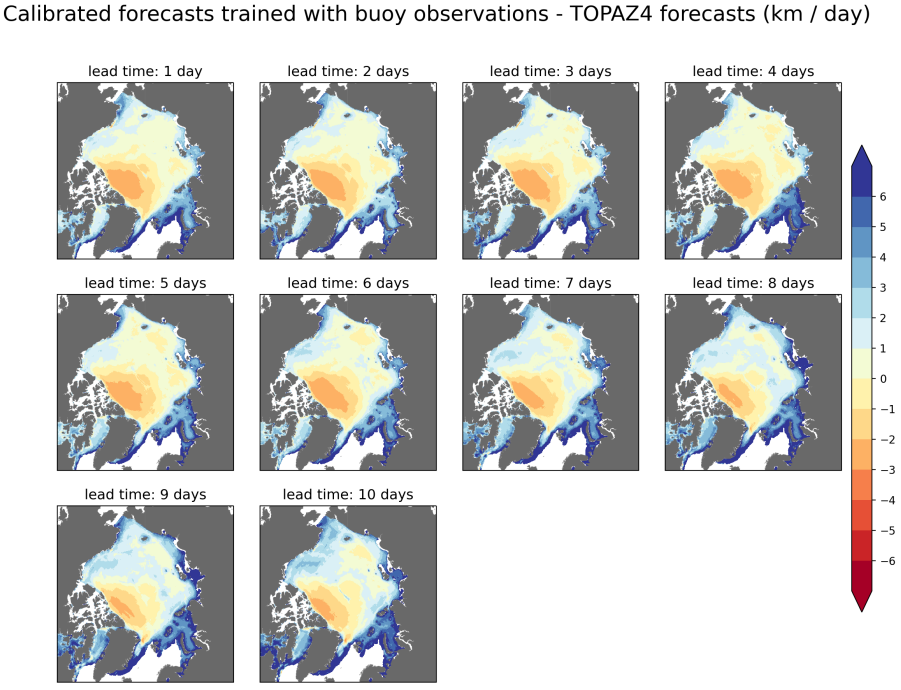

Figure S15. Difference between the random forest models trained with buoy observations and the TOPAZ4 forecasts for the speed of sea ice drift (km / day) during the period June 2020 - May 2021.

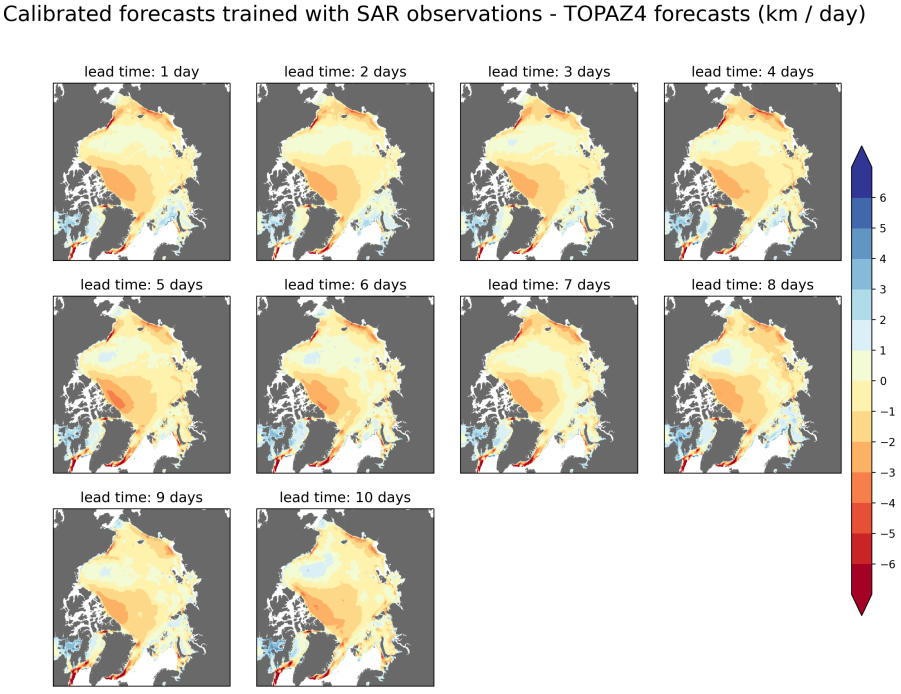

Figure S16. Difference between the random forest models trained with SAR observations and the TOPAZ4 forecasts for the speed of sea ice drift (km / day) during the period June 2020 - May 2021.