

The Cryosphere Discuss., referee comment RC2
<https://doi.org/10.5194/tc-2021-24-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on tc-2021-24

Anonymous Referee #2

Referee comment on "Calibration of sea ice drift forecasts using random forest algorithms" by Cyril Palerme and Malte Müller, The Cryosphere Discuss., <https://doi.org/10.5194/tc-2021-24-RC2>, 2021

Review of Calibration of sea ice drift forecasts using random forest algorithms.

The manuscript describes a new method that post-processes numerical forecasts of sea ice drift using either in situ drifting buoys or satellite images for the training of a random forest algorithm. The results are evaluated against ice drift observations but in a different period, posterior to the training data. The results reveal that there is a systematic component of the ice drift forecast error that can be corrected by machine learning, although the reduction of error remains often less than 10%. The ML algorithms learns more efficiently from the buoys data than from the satellite images, highlighting the problem of temporal averaging.

The drift direction can mostly be improved in the short forecast range, likely because of the unpredictability of wind directions, but interestingly the algorithm is more often able to correct drift speed at longer forecast horizons, which I did not expect. The authors could spice up their article by analysing what their algorithm does to the sea ice drift speed that improves the skills at a 10 days range: are the drifts made systematically faster or slower? This kind of analysis can - if understood - lead to improvements of the forecast systems. More generally, not seeing what the algorithm does to the forecast is a little frustrating. An example of comparison of original to postprocessed and to observed sea ice drifts could be more convincing than cold-blooded skills scores.

One general remark pertains to the Lagrangian nature of sea ice drift. The variable influencing the drift at a lead time of several days may not be at the same location as the sea ice drift value. This issue is not addressed in the paper, what do the authors expect to be the effect of considering both the predictor and the target at the same location?

The authors have also neglected the seasonal changes of the forecast model performance, as well as the long-term model drift (or rather the absence of sea ice acceleration) as

pointed out originally by Rampal et al. (2011) and then Xie et al. (2017) using an almost identical model. Can the algorithm learn the seasonality of the errors or could it be improved if trained separately on summer and winter data?

The manuscript cites the relevant literature and is original in its goals. I am not aware of any similar study carried out elsewhere. The article is logically structured and reads quite well. The figures are generally nice and clear. Exceptions are noted in detailed comments below.

Based on the above, I recommend the manuscript is published with minor corrections.

Detailed comments:

- P1, l21: The relationship is complex and nonlinear in the ice pack where the rheology is active, but for low ice concentrations, the ice is in free drift and should be a linear function of the winds (the Nansen relationship).
- P2, l29: "but they obtained": false opposition. Is there any reason why RF or CNNs would have an advantage for sea ice concentrations?
- P3, l78: The overestimation of sea ice drift was reported in reanalysis, but since the decadal acceleration of sea ice drift is not reproduced by the model, the bias should be smaller in recent times, as can be seen in the TOPAZ4 validation pages: https://cmems.met.no/ARC-MFC/V2Validation/timeSeriesResults/year-day-01/SItimeSeries_year-day-01.html#drift (accessed 2nd March 2021)
- l108: "different algorithms were used": "models" should not be synonymous with "algorithm" (the Random Forest is one algorithm, from which you can build several models). Maybe use "distinct models were developed to..."?
- L148: At which point is the averaging used? Is it related to the averaging of each prediction tree?
- L148 If the predictive variable is a complex number, isn't it similar to predict normalised u and v components (with a norm of 1)? In that case, this choice is apparently contradictory with the assertion line 90: "In order to predict independent variables, it has been chosen to forecast the direction and speed of sea-ice drift rather than the eastward and northward components"
- Section 3.2: It is very positive that sensitivity studies are detailed. The algorithms were tuned against the size of the training set (period for buoys, subsampling rate for SAR), size of the forest (number of trees), other parameters of the RF. It is not clear to me which criteria were used for this tuning. On which dataset the error has been computed to evaluate the tuning? Is it the one used to evaluate the results (buoys in June-November 2020) or the one used to evaluate the importance of predictor variables (section 4.3)?
- L183: The period chosen for evaluating the model is mostly in the summer season (June-November)? Do you expect it to be representative of the winter? The link above shows a seasonal signal in the drift bias, though not a large one.
- L206-2018. It is fair to note the absence of data where the performance deteriorates. This however deserves an explanation as to how the random forest algorithm extrapolates the training data spatially. Does it find the most analogous situations where and when training observations are available? The authors explain that the

random forest does provide the average of an ensemble but it would be good to have insights about the values returned, for example, in places of intermittent landfast ice.

- L224-226 "The selection of the data sets used for training and evaluating the random forest models is a random process according to the forecast start date to avoid the influence of neighboring grid points with very similar conditions," this point of correlations between training and validation data (leading to data leakage) is essential to avoid correlation between training/validation data that could lead to data leakage and overfitting. It would be beneficial for the community to give more details (even if it is given in appendix) about your selection procedure.
- I 236: Intuitively one may expect that the areas of thicker ice drift slower than thin ice due to the increased resistance to stress.
- Section 4.3. This sensitivity study is important. But I am surprised not to see the standard "Importance variable" diagnostic available in any random Forest algorithm? Even if the results are redundant with your study, it would have offered another point of view of variable importance.
- L258: It is correct to mention the changes in operational systems but the authors should note that even with unchanged reanalysis systems, the gradual acceleration of ice drift is not reproduced by the models and may also affect the training over long periods.
- L261: I may have misunderstood this point. I do not expect any 7-days frequency signal in sea ice drift so Thursdays are representative of the rest of the week.
- Code availability: I would like to point out that there is not enough details given on the results so it can be reproduced. It is said that "the codes used for this analysis can be made available upon request." but without the code, it is not possible to reproduce the results as the RF models are not detailed.
- Figures 2 and 10: the crosses colours are not colourblind-friendly. Try a simpler scale - a gradient - that can easily distinguish the high from the low percentages. The general tendency is more interesting to me than the exact values.
- Figures 4 and 5: do we need to see both the MAE and the RMSE?

References:

Rampal, P., Weiss, J., Dubois, C. and Campin, J.-M.: IPCC climate models do not capture Arctic sea ice drift acceleration: Consequences in terms of projected sea ice thinning and decline, *J. Geophys. Res.*, 116, C00D07, doi:10.1029/2011JC007110, 2011.

Xie, J., Bertino, L., Counillon, F., Lisæter, K. A. and Sakov, P.: Quality assessment of the TOPAZ4 reanalysis in the Arctic over the period 1991–2013, *Ocean Sci.*, 13(1), 123–144, doi:10.5194/os-13-123-2017, 2017.