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.
   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.
   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.

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.

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.

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".

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.

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

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.

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?

Line 79: Which ocean observations are assimilated?

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

Line 95: No new paragraph here. "… where R is the Earth's radius, lamda and phi are the…"

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

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

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

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

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.

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?

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

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

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.

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.

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

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

Bruno Tremblay
McGill University