

Evaluation: Probabilistic Gridded Seasonal Sea Ice Presence Forecasting using Sequence to Sequence Learning

Overall evaluation

Principal criteria	Excellent (1)	Good (2)	Fair (3)	Poor (4)
<i>Originality (novelty)</i>		X		
<i>Scientific quality (rigour)</i>			X	
<i>Significance (impact)</i>		X		
<i>Presentation quality</i>			X	

TC review questions

- 1. Does the paper address relevant scientific questions within the scope of TC?**
Yes, absolutely.
- 2. Does the paper present novel concepts, ideas, tools, or data?**
The manuscript uses innovative approaches in ice forecasting.
- 3. Are substantial conclusions reached?**
More work needs to be done on this matter. Even though the tool is solid and technically pertinent, more has to be said on the overall conclusions of its applications and usefulness.
- 4. Are the scientific methods and assumptions valid and clearly outlined?**
On this point, more clarity must be achieved in the manuscript. Research hypotheses and objectives could be more precisely outlined. Details about the different tests and configurations should be brought to the attention of the readers.
- 5. Are the results sufficient to support the interpretations and conclusions?**
Details are lacking in this matter. A lot of attention has been put on the technical matters but too few on the applications of the model itself. This comparison and “real world application test” part will prove the pertinence of the tool, not only the technical justifications.
- 6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?**
No. More details need to be provided. Especially the ML architecture and configuration that needs to be explained and also, supported by a schematic representation.

7. **Do the authors give proper credit to related work and clearly indicate their own new/original contribution?**

More could be done, especially in terms of comparison with existing models. However, when speaking about the model and ML approaches, the authors gave proper credit to existing work and clearly stated their contribution to the field.

8. **Does the title clearly reflect the contents of the paper?**

Not entirely, it could be more precise. At first glance, I thought it was based on remote sensing time series and had no idea of what “seasonal” referred to.

9. **Does the abstract provide a concise and complete summary?**

It is quite short and could provide more details. More could be said about the applicability and practicality of the developed tool as well as about the limitations.

10. **Is the overall presentation well structured and clear?**

Yes. The authors tend to go straight to the point when needed.

11. **Is the language fluent and precise?**

Yes absolutely.

12. **Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?**

Yes.

13. **Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?**

Both methodology and discussion should go into more details. Especially the discussion (see comments).

14. **Are the number and quality of references appropriate?**

There are little to no reference listed about the known Hudson Bay sea ice spatiotemporal patterns and dynamics. As I suggested more application examples and comparison with existing methods and data, I suppose a few references will be appended.

15. **Is the amount and quality of supplementary material appropriate?**

Since it is not the final submitted version, I understand that at this point the code isn't shared. However, this would be an excellent asset to help readers and potential users appreciate the authors work.

Reviewer comments

This manuscript presents an innovative forecast tool for sea ice conditions (presence) based on a machine learning approach using sequence to sequence learning for both short term (7 days) and long term (up to 90 days) predictions.

The presented tool is, without a doubt, something that is of interest to the sea ice expert community and is based on novel methods of machine learning that make analysis of massive information datasets possible nowadays.

Even though the pertinence of the presented tool, several improvements must be done on the manuscript. The research design in itself has to be described into more details, especially by providing a more complete description of the different tests and protocols followed in the experiments and model calibration part.

In addition, many paragraphs, especially in the methodology part, could be supported by figures and schematic representations, for example, the ML design and architecture.

Also, as the Hudson Bay region is highly documented and studied, the results obtained by your approach could be compared to data provided in the sea ice atlas from the Canadian Ice Service or results from other probabilistic/modelling approaches applied on the Hudson Bay (Saucier et al. 2004, Hochheim and Barber 2014, Kowal et al. 2017, Gignac et al. 2018, Dirksen et al. 2021.). Even though the comparison may not be quantitative, a qualitative assessment, outlining the differences between the approaches and the strategic advantages you provide using ML would be relevant.

Overall, the presented method and tools appear as scientifically sound and clear. They should, however, be described more carefully and more examples of applications of the model shall be presented to the readers.

It is a work of great interest and I hope my comments will guide and help you in improving your manuscript.

Major comments

1. As aforementioned, comparisons have to be made and applications examples provided. Especially in areas of high variability or in the presence of particular entities such as polynyas or narrower Bays (Frobisher Bay or Hall Beach polynya, for example).
2. Limitations of the approach shall be discussed. The model is providing forecasts on a ~31 km grid. How does this affect the usage capabilities for the principal expected users (the mariners)? This should definitely be discussed more in depth.
3. Sensibility to the input "sea ice normal condition" wasn't discussed, when speaking about the augmented model. Were variable time spans tested? If so, did they generate similar forecasts? If not, how would you explain this situation? In other words, a certain "sensibility analysis" would be convincing about the model capabilities.
4. I strongly suggest that you add a map of your validation sites and provide a short description of each. For example, Quaqtq is located in a bay, narrower than 31km. How this does affects the results? Also, why were these 3 sites chosen? Have you found any irregularities in the ERA-5 sea ice concentration dataset you used to calibrate your model?
5. Nowhere in the manuscript have I found a justification on why it is the presence of ice that is modeled and not the concentration. That should be discussed since the OSI-SAF OSI409 (which are SIC) data are ingested into ERA-5.

Specific comments

- Line 10 : Define "high spatial resolution" for the reader. Depending on your field, it differs.
- Line 46 : Define sea ice presence (SIC > 15%).

- Line 51 : This information should be provided way before, otherwise, some will think, as I did, that you use remote sensing data.
- Line 57 : Why not starting in 1979 ?
- Line 67 : Remove last “and”.
- Line 87 : A schematic representation of the encoder and decoder parts would be useful.
- Line 92 : What time bin(-s) were used as input (12:00, 00:00, a daily average ?)
- Line 112 : How does extending to a longer input affects the forecast quality ?
- Line 126 : Can you describe these “extensive experimentations” ?
- Line 129 : “Chosen to be 10 years”. Why is it so ?
- Line 135 : This processing logic should definitely be represented in a figure.
- Line 166 – 167 : The end of the sentence doesn’t make sense. Consider reformulating.
- Line 181 : It seems counterintuitive. Can you explain why ?
- Figure 1 : Y-Axes for subfigures d-e-f should be Accuracy differences or Δ Accuracy.
- Line 220 : What would you link the lower accuracy in “central region” ? Is it the higher variability of the freeze-up pattern or to climate variables that are, given the distance to stations, less reliable in such areas ?
- Lines 236, 246 & 249 + Figures 8 & 9 : From what I know, it should be written Quaqtaq, not Quataq (<https://www.makivik.org/quaqtaq/>).

References

Dirkson, Arlan, et al. "Development and Calibration of Seasonal Probabilistic Forecasts of Ice-free Dates and Freeze-up Dates." *Weather and Forecasting* 36.1 (2021): 301-324.

Gignac, C., Bernier, M., & Chokmani, K. (2019). IcePAC—a probabilistic tool to study sea ice spatio-temporal dynamics: application to the Hudson Bay area. *The Cryosphere*, 13(2), 451-468.

Hochheim, K. P. and Barber, D. G.: An update on the ice climatology of the Hudson Bay system, *Arct. Antarct. Alp. Res.*, 46, 66–83, 2014.

Kowal, S., Gough, W. A., and Butler, K.: Temporal evolution of Hudson Bay Sea Ice (1971–2011), *Theor. Appl. Climatol.*, 127, 753–760, 2017.

Saucier, F., Senneville, S., Prinsenber, S., Roy, F., Smith, G., Gachon, P., Caya, D., and Laprise, R.: Modelling the sea ice-ocean seasonal cycle in Hudson Bay, Foxe Basin and Hudson Strait, Canada, *Climate Dynam.*, 23, 303–326, 2004.