

The Cryosphere Discuss., referee comment RC3
<https://doi.org/10.5194/tc-2021-282-RC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on tc-2021-282

Anonymous Referee #3

Referee comment on "Probabilistic spatiotemporal seasonal sea ice presence forecasting using sequence-to-sequence learning and ERA5 data in the Hudson Bay region" by Nazanin Asadi et al., The Cryosphere Discuss., <https://doi.org/10.5194/tc-2021-282-RC3>, 2021

Asadi et al.:
Probabilistic Gridded Seasonal Sea Ice Presence Forecasting using Sequence to Sequence Learning

The authors present a new approach for forecasting sea ice presence in the Hudson Bay area using machine learning techniques. The study presents models which use the Sequence-to-Sequence Learning framework to predict probabilities of sea ice presence for up to 90 days lead time. The authors suggest two somewhat different models, which are applied in hindcasting experiments, where they exhibit slightly more skill than "climate normal"-predictions especially in the breakup season. The models are also evaluated for their ability to predict freeze-up dates and breakup dates.

The study has a clear motivation, is well structured, and applies new (as to my knowledge) methods in a promising way. The text is short and precise. The general setup of the experiments is described well, however, as I'm not an expert in ML, I cannot judge the parts of the paper that go into technical details of the ML process. The results are presented clearly, but I miss a broader discussion of the results and which conclusions can be drawn from them. Especially as the motivation of the study is to develop new methods to support maritime users with new operational forecast products, a comparison to existing products would be valuable. Also a short assessment/discussion of the applicability to operational forecasting is missing in my opinion.

Please find here some general comments, followed by comments related to specific line numbers. Text in quotation marks after the given line number refers to the original text of the manuscript. Text in quotation marks in the following line is my suggestion of how to replace the original text.

%%%%%%%%%
General comments
%%%%%%%%%

A)
The captions of Figures 1-3 do not only explain the figure but also contain statements about the shown results. This is a new approach to me. If the journal allows for that, I would not object, but I don't think it is common style.

B)
For Figure 4 you compare predicted ice presence with observed ice presence. The observations are calculated from SIC from ERA5 by using a threshold of 15 %, while the ice presence from the forecasted probabilities is calculated with a threshold of 50 %. As the model is based and trained on SIC from ERA5, why do you use a threshold of 50 % and not of 15 %? Or at least the same threshold for both?

C)
I understand that the motivation of your study is (at least partly and in the long run) to improve operational forecasting of sea ice conditions in the Hudson Bay area. In line 254-257 you explain the technical advantages of the ML approach compared to standard numerical models (=reduced computational costs). The paper would benefit from also looking into the results/skill of the ML models compared to standard models. Is your approach not only faster but also better than currently used forecasting systems? Or is it so much faster that it is useful despite of a possibly lower quality? Or is it worse/better only for some lead times? It would be interesting to see how your model compares e.g. to the S2S-forecast of ECMWF (up to 60 day forecast at 1/4 degree resolution).
<https://www.ecmwf.int/en/forecasts/dataset/sub-seasonal-seasonal-prediction>
<https://apps.ecmwf.int/datasets/data/s2s-realtime-daily-averaged-ecmf/levtype=sfc/type=cf/>
Comparing your models to the climate normal is a very good and valid first step. The comparison to a numerical forecast would however be a very interesting addition.

D)
With the Basic and Augmented models you introduce two approaches, which you compare with each other throughout the paper. However, I don't find a conclusion/discussion about which of the two models you would suggest in the end. Is it worth the effort of the Augmented model, which needs more input data, or is the Basic model sufficient for the purpose? Or are both needed, for different purposes? Maybe you can here also explain/speculate why the Augmented model is considerably worse than the Basic model in Figure 2c. [This text could extend the summary given in lines 258-265.]

E)
For your hindcasts you use input data from ERA5, which is a reanalysis product that is not available in real time. Hence, when one wants to apply your method for forecasting *future* conditions, other input data need to be used. It would be good to elaborate on this topic in a paragraph in the Discussion or Outlook sections. Is it difficult/problematic to switch to

other input data? Can the trained monthly models be applied if the forecast is started based on other data for the 3 historical days? And/Or what else is still needed before your models can be used for operational forecasting? This would be a good topic for an Outlook-section/paragraph.

%%%%%%%%
Specific comments
%%%%%%%%

Title

In the introduction you mention that the novelty is that your forecast is "spatiotemporal". Hence I wonder why you don't use this word in the title.

Abstract
#####

2-3
Would ML approaches be less important without global warming? Suggestion: Remove "Given ... global warming".

4 "a daily spatial map"
Isn't the clue of your study that you provide several "daily spatial mapS", namely 90 for a 90-day forecast?

1 Introduction
#####

9, 10, 12, 14, 82
Be more clear on the terms "short-term", "longer term", "seasonal" and "medium-term" forecasting.

16
Maybe mention the lead-time used in Zhang et al. (2008)

21 "governing"
Do the equations govern the physics? I'd suggest "describing"

24 "This is a key advantage of..."
Suggestion: "This disadvantage can be overcome by using..."

27 "to perform"
"for"

29 "Their results are"
"The results were"

29, 31
In line 29 you write that the model predicts sea ice concentration but in line 31 you present results for ice extent. I'm not sure if one can assume that everyone knows the relation between SIC and sea ice extent.

32 "September sea ice minimum."
minimum extent? minimum thickness?

33 "This study"
This is misleading because it could mean your own study. Better use "They" or "Hovath et al. (2020)".

34 "was found the uncertainty"
"was found that the uncertainty"

38 "that is closer to what is proposed here"
As we don't know yet, what you will propose, this information is not very useful here.

49 "probability of ice at"
"probability of ice presence at"

2 Data
#####

56 "data from 1985-2017 is"

"data from 1985-2017 are"

65 "the following input variables"
"the following 8 input variables". This helps explaining the number 8 in line 96.

67 "V-Component"
"V-component"

67
replace "and" before "landmask" by comma

3 Study region

Here would be a good place for a map which also indicates the ports used later on.

76
Remove the parenthesis if Foxe Basin can be shown in a map.

79 "Recent decades"
For me, 1985-2017 includes several "recent decades" and one could get the impression that lines 77-79 are not valid for "recent decades". So maybe consider re-phrasing "recent decades" to "In recent years" or using the term "trend"?

4 Forecast model architecture
#####

85/86 "sequence of inputs"/"sequence of outputs"
It would be helpful if you could mention (maybe in a new sentence) some examples for "input" and "output" for the application in this study. I guess input includes SST, t2m, winds, etc. and output is ice presence probability?

86 "consist"
"consists"

88
Does "desired length" in your application mean number of variables or number of grid cells

or number of forecasted days? It is good with a general explanation of the Seq2Seq method like you do here, but for someone not from the ML field, it would also be nice to directly get examples about how the method can be understood for the application of sea ice forecasting.

89-90

I first understood this sentence such that the encoder part would be called Basic model and the decoder part would be the Augmented model. Can you phrase it differently to make it more clear also for non ML-experts?

94 "three days of environmental conditions"

Shouldn't it read "environmental conditions of the last three days"?

96

Maybe explain why you call the number of input variables "C":

"and C is the number of channels, in this case the total number of input variables (here 8)."

98 "sequence of extracted feature grid"

Are the feature grids what was called "environmental patterns" in the previous sentence? If so, could you use the same term? If not, could you explain how to get from one to another?

98 "the sequence are"

"the sequence is"

97-113

As I don't have a background in AI/ML, I unfortunately don't understand the setup of the model in detail. However, as I can follow the general concept, e.g. what is input and what is output, I think it is OK to keep the text as it is if the targeted audience is AI/ML experts more than sea ice modellers.

115

I miss a sentence about why you suggest an additional model. What is the (expected) problem with the Basic model or which benefits do you expect from the Augmented model?

116 "(e.g., 60 or 90 days)"

remove comma

117 "t2m, u10 and v10"

Why do you use exactly these variables? I could imagine that climate normals of e.g. sea ice concentration or sea surface temperature could also be beneficial to correctly predict sea ice presence.

5 Description of Experiments

#####

124 "required"
"requires"

125 "assess"
"assesses"

130 "the model from year i-1"
"the model for year i-1"

130
End the sentence after "i-2" and start a new one.

136 and 140
Are "ML models" (line 136) and "neural network model" (line 140) something different? If not, use the same word.

142 "3 months of year"
"3 months of each year"

141 "thresholded at 15%"
This is unclear to me. Do you apply the threshold to convert to ice presence?

131, 132, 142
In line 142 you talk about "test procedure". Is this the same as "validation" mentioned above?

6.1 Presence of Ice Forecasts
#####

145 "6.1 Presence of Ice Forecasts"
Shouldn't it be "Forecasts of Ice Presence"?

147 "test set"
What is this? I don't think you have introduced this term before.

148
How do you calculate accuracy from the binary forecast map?

148, 149, 150 etc.
I would put a period after the abbreviated "Fig" -> "Fig."

150 "in this figure (Fig 1(b))"
"in Fig. 1b"

150 "the first top-left"
"the first (top-left)"

151 "after 1 day forecast"
"after a 1-day forecast"

152-153
Why April 1 and April 2? Wouldn't a 1-day forecast started on January 31 end on February 1, and a 2-day forecast on February 2?

154 "month on January"
"month of January"

155 "consistently"
"constantly"

156

Mention the sub-figure number you are talking about.

158 "Fig 1d"

"Fig. 1d and 1e"

158 "significantly"

Did you do a statistical test whether it is significant? Otherwise maybe remove the word.

160 "early lead times"

"short lead times"

163 "Climate normal"

Stay consistent with capital C or not.

163-164

Comparing the Augmented model with the climate normal (Fig. 1e), I don't see an improvement for March/April. I see the Augmented model is better than the Basic model, but actually it is just 'less bad' compared to the climate normal. The accuracy of the Augmented model is not higher than of the climate normal. (You explain this later, so maybe make clear that this sentence only deals with Fig. 1f and not with Fig. 1e.

166 "(Fig 1d)"

"(dark areas in Fig. 1d)"

167

Remove "accuracy" at the beginning of the line.

167 "90 lead day"

"90 lead days"

167

The "For example..."-sentence is not complete, there is no verb.

169

Consider to start a new paragraph for the Brier score.

169

(related to comment for line 148): What is "probabilistic accuracy" compared to "accuracy"?

173

Remove "Also"

174 "Pt is the model prediction"

Maybe add "... of ice presence probability"

174 "represents"

"presents"/"shows"

177 "The pattern observed"

"observed can easily be mixed with observations, so maybe say "The resulting pattern"

178

End sentence after "both models" and start a new one for the differences.

178 "(2c)"

"(Fig. 2c)"

179 "longer lead days"

"longer lead times"

184 "For early lead days"

"For short lead times"

185 "lead day"

"lead days" (make sure to do it consistently throughout the paper, e.g. line 211 and in caption of Figure 3)

184-187

I find this sentence is too long. Also, there is inconsistency of the used terms "forecasted

probability" and "forecasted probabilities".

190

When first reading the sentence, it sounds like monthly averaging would make it impossible to provide information on a map. Maybe you can clarify it with "Monthly averaged and domain-integrated accuracy values ..."

194-195

To simplify the explanation of the different dates, I suggest to add the dates in the caption of the subfigures 4a-4c, e.g. "(a) 5 June 2014 (after 30 days)"

194 "given data"
"given date"

198 "and and"
"and the"

200-201

I don't see in the figures that the Basic model would have "increased ice presence probability in the northern part of the domain". If it is important, highlight the area in the plots.

202 "Observations"
"observations"

201-202

"the agreement ... to be in good agreement" Too many agreements.

Figure 1 and 2:

I would suggest to use a diverging colormap when differences are displayed. This makes it easy to see where zero is located and it makes it more clear which plots display differences and which plots display absolute values.

Caption of Figure 1:

"Model performance and improvements": Why not call it "Accuracy"?

Caption of Figure 2:

"Brier score of the Basic model (a) and the Augmented model (b) as a function of lead time. Their score difference is shown in (c). Most differences are observed in breakup and freeze-up seasons."

Figure 3:

- It would be nice if the aspect ratio of x and y axis was 1, so that the dashed line would be at 45 degree.
- for the text in the legend I suggest "xx lead days" instead of "xx Lead Day"

Caption of Figure 3:

- Would be nice to remind the reader (especially those who only look at the figures and don't read the text) that you are talking about ice presence probabilities/frequencies.

Figure 4:

- In order to compare probability maps with ice presence maps it would be nice (if possible) if the plots would have the same size, i.e. smaller plots for those which have no colorbar.
- I would prefer if the height of the colorbar was the same as the height of the map-plot.
- Make sure to use "Climate normal" or "Climate Normal" consistently.

Caption of Figure 4:

- The figure does not illustrate "the May models" but the forecasted conditions. Hence a suggestion for rephrasing: "Ice/water distribution in the model domain as observed and forecasted by Basic and Augmented models for a forecast started on 6 May 2014 and lasting for 30 (a), 50 (b), and 70 (c) days, respectively."

6.2 Assessment of operational capability

#####

209 "15 continuous days in a row"

Either "continuous" or "in a row" is enough.

212

It can be a bit confusing that "accuracy" here is related to freeze-up/breakup while the same word is used in section 6.1 for ice presence. (Well done in line 215.)

214 "...prediction is correct."

The reader has to infer that 'correct' is translated to 1 and 'not correct' is translated to 0.

215-216

Check the grammar of the sentence.

219-212

It is surprising to me that a model can have that much more skill on a 30 day longer lead time. Could you elaborate on possible reasons and/or why the Augmented model is doing a better job? (This should probably go to the Discussion section)

222 "breakup prediction ability"

Why not call it "breakup accuracy" in analogy to line 215?

222 "are presented"

"is presented"

222 "fig 6a"

Fig. 6a

227 "variability"

Would "interannual variability" be more clear?

227 "models' accuracy"

As climate normal is not really a model, I would remove the word "models".

228 "represented"

"presented"

228 "For each prediction... same color."

Suggestion: "The respective trends are shown by dashed lines."

229,233 "freeze-up season accuracy"

"season" is not necessary.

229 "both lead days"

"both lead times"

229 "breakup plots (...)"
"breakup accuracy (Fig. 7c and 7d)

230
Move "accuracy" directly after "2%"?

232-233
Is this because the Augmented model uses climate information as input data and hence tends to predict more similarly to the climate normal than the (more independent) Basic model does?

240 "is different."
"is different (i.e. different x and y axes)."

236
Refer here to the (new) overview map for locations of the ports.

245 "both models both lead days"
"both models and both lead times"

248 "as freeze-up for breakup"
"for breakup as for freeze-up"

Figures 5 and 6:
- Add a space between (a), (b), ... and the subfigure caption text.
- Why did you choose to use a diverging colormap even though the plot does not display differences?
- Remove the grid lines which are drawn around each grid cell in order to make the plot less busy.

Caption of figure 6:
- End with a period.

Figure 7:
- Add a space between (a), (b), ... and the subfigure caption text.
- You could specify "Freeze-up accuracy" and "Breakup accuracy" also in the y-labels.

- Red and green lines are probably difficult to distinguish for color-blind persons. What about using black for the climate normal and e.g. red and cyan for the two models?

Caption of figure 7:
"Dashed lines" instead of "Dotted lines"

Figures 8 and 9:
- The little red arrows next to the dots are hard to see and the written year numbers don't allow for getting a quick overview of the distribution of the year (e.g. whether the skill is getting better or worse over time). Did you try to plot the dots using a colormap which represents the different years (i.e. colorful dots, and the "valid area" as gray)? Then you don't need the text labels anymore.
- "freeze-up" instead of "Freeze-up" in x-labels and y-labels.

Caption of figure 8 and 9:
- Mention that each dot represents one year.

7 Discussion
#####

251
In my opinion you should mention somewhere that your model is not (yet) used for forecasting future condition but rather for hindcasting.

254 "where it takes"
"which takes"

261 "Augmented model showing better scores comparing to"
"Augmented model shows better scores compared to"

262 "analysis on"
"analysis of"

267 "less disperse"
less disperse than what?

267 "to accepted"
"to the accepted"

References

#####

281 "shipping"
"Shipping"

293, 305, 308
Why do you cite preprints of papers that are several years old?

297 "S., R., G.,"
The co-author is called "Graversen R"

297 "high resolution"
"high-resolution"

298, 322
missing volume/issue/page number

327 "W.-c."
"W.-C."?