The Cryosphere Discussions

The Cryosphere Discuss., referee comment RC2
https://doi.org/10.5194/tc-2021-282-RC2, 2021

Comment on tc-2021-282

Anonymous Referee #2

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-RC2, 2021

Asadi et al. 2021 - Probabilistic Gridded Seasonal Sea Ice Presence Forecasting using Sequence to Sequence Learning

The authors present a fascinating application of machine learning techniques to better predict ice presence/absence within Hudson Bay, using ERA5 data as an input. The results show promise in helping plan shipping operations around the ice-free season, however, the clarity of these results is lost in lengthy wording. It is recommended that the authors read through the document for grammatical errors and places where the wording of sentences can be made more succinct. This article can become much more impactful and easier to read with more 'straight to the point' sentences.

General comments:

- Ensure you are consistent using 'freeze-up' with a hyphen throughout the document, and choose either 'breakup' or 'break up' to use throughout the document

- I am aware that it is difficult to phrase sentences when discussing the number of lead days and the two models, however, I found most sentences discussing these topics hard to read. For example, line 149:

'For example, the top row of Fig 1b shows the accuracy of forecasts launched in January

using Basic model for forecast lead days of 1 to 90. E.g., the first top-left box in this figure (Fig 1(b)) corresponds to the average accuracy after 1 day forecast for all forecasts launched between January 1 and January 31, ending in January 2 to April 1 and the second box corresponds to average accuracy of forecasts launched between January 1 and January 31 ending in January 3 to April 2.'

I think it would be easier if you use articles when you are referencing lead days or models. For example: 'the Basic model' or 'a 1 day forecast'. This would make your sentences flow better while reading them, which would communicate your results more efficiently.

- The results section has some statements that are more suited towards the discussion section, however I see your discussion and conclusion section are combined. I'm unsure if the section headers are pre-determined by the journal, but if they are not I would suggest making section 6 'Results and Discussion', and section 7 'Conclusion'. This would allow you to discuss your results more in depth as you present them, as I feel like some of your results could be discussed more in depth.

- Throughout the document, you abbreviate some month names and use the full name for others. You should pick one method and stick to it throughout (i.e. always abbreviate or always use the full word).

- There is a comment in the specific comments regarding this, but you should include some discussion regarding the resolution of your results, and how this may impact the use of your results for port-specific operations. I am a little wary of how the land mask may impact how 'close' the pixel you use to represent the port is to the actual port in question. A figure representing this may add some clarity.

Specific comments:

Line 3 – You may be limited on word count in your abstract, but I think it would be helpful if you stated the type of data you are feeding into your ML system to derive these predictions.

Line 3 – recommend changing to "Given the recent observations **of** the declining trend"

Line 6 – recommend changing to 'within a 7-day time period', unless you define why a

7-day time period is ‘valid’ in the manuscript?

Line 8 – The introductory sentence needs a little bit of work. I would recommend removing ‘northern communities’ as you do not speak of them in the rest of the introduction. Maybe focus more on the topic of shipping and why ice forecasting is vital for shipping operations in this introductory sentence. OR add in reference to communities, and why they rely on ice.

Line 12 – Could you expand on what you mean by ‘Typical approaches are usually statistical or dynamical in nature.’? Maybe add a reference to examples of these? I see that you go more in depth in the next paragraph into dynamical forecasts, but what about statistical like you mentioned earlier?

Line 15 – I would recommend splitting this up into two sentences, breaking it up at one of the commas

Line 16 – remove ‘the summer of’ before 2008, as you have already indicated that this study was in the spring and summer

Line 18 – I am not too sure what you mean by ‘skill’. Do you mean the forecasts ability to predict ice? There may be a better way to word this to avoid ambiguity.

Line 20 – It might be nice to list some environmental controlling factors in brackets, like: (i.e. wind speed and direction, tides)

Line 24 – Recommend to change to ‘Both of these approaches determine…’

Line 28 – Change to ‘composed of sea ice concentration data…’

Line 30 – Remove ‘good’

Line 30 – Would help the reader if you included where the mean September sea ice extents from 2017 came from (ice charts? Passive microwave data?)

Line 43 – ‘calibrated probability of ice’: presence or concentration?

Line 64 – Need to define ‘SST’

Line 65 – Doesn’t ERA5 have a 31km resolution? I would state this plainly so the reader knows what resolution your results are.

Line 74 – Would recommend shuffling around this sentence: ‘Shipping traffic is also generated by mining, fishing, tourism and research activities, being mostly confined to the ice-free and shoulder season’.

Line 84 – ‘In Seq2Seq learning, which **has** successful applications in machine translation’

Line 87 – Recommend to spell out ‘two’

Line 88 – suggest removing ‘part’

Line 92 – In line 54 you use the double wavy equal sign, but here you use a single wavy line. I would recommend picking one and keeping it consistent throughout.

Line 94 – Recommend to change to: ‘The encoder section of the Basic model takes the last three days of environmental conditions as an input’

Line 97 – Remove ‘so as’

Line 99 – May be better to spell out ‘LSTM’ in full form

Line 101 – Recommend rewording the last sentence for clarity: ‘The output to the encoder is a single raster with the same height and width as the input, but a higher number of channels to represent the fully encoded system state.’

Line 115 – Remove ‘so as’ (try and write sentences as simply as possible, i.e. with as little

unnecessary words)

Line 128 – Just verify that your quotation is facing the correct way before ‘April’

Line 137 – How did you determine what learning rater and momentum to use?

Line 142 – Suggest to remove ‘coming’, or replace with ‘derived’

Line 151 – I would recommend changing the formats of your dates here: ‘forecasts launched between 1-31 January, ending in 2 January to 1 April, and the second box corresponds to average accuracy of forecasts launched between 1 – 31 January, ending in 3 January to 2 April.

Line 154 – This sentence needs a lot of work: suggest removing ‘very’ and changing ‘on January’ to ‘of January’. As well, are you indicating that the accuracy is close to 100% for both January and the span of January – March (this is not clear)? It would be helpful if you stated the actual accuracies.

Line 155 – This sentence struggles with the same structural problems as the first, I would recommend rewording to something like: ‘In contrast, for forecasts at the beginning of the open water season (June and July), the climate normal struggles to accurately capture the ice cover for lead times of 1 to 50 days likely due to inter-annual variability and the impact of climate change’. You might also want to indicate what climate change has to do with this (i.e. ‘lengthening of the open water period due to climate change’)

Line 165 – This sentence also needs to be reworded, I have underlined grammatical errors: ‘Using additional climate variables for the input of the Augmented model is showing its impact here where in the periods that Basic model is worse than climate normal (Fig 1d), the Augmented model has better accuracy and is closer accuracy to climate normal.’

Line 169 – Double check if it should be ‘the climate normal’ or ‘climate normal’

Line 179 – Spell out ‘April’ fully, as you have spelled out every other month

Line 202 – ‘Observations’ should not be capitalized

Figure 4 – Include units for Latitude and Longitude, and capitalize the words in your legend

Figure 5 – Units for lat and long

Line 212 – 'Figure 5 and 6 show the overall…'

Line 215 – 'The freeze-up accuracy maps at Fig 5 show that except the Basic model's prediction at 30 lead day (Fig 6b), other maps are showing similar patterns of accuracy.' This sentence needs reworking – would recommend flipping the sentence, so you are presenting the positive results first, then adding on the Basic model's prediction after.

Line 221 – 'compared' instead of 'comparing'

Line 222 – Capitalize 'fig 6a'

Figure 6 - Units for lat and long

Line 227 – I would recommend changing all of your dates to the format: '1 Oct to 31 Jan'. This is a more standard way of presenting dates and is more simplistic.

Line 233 – 'Compared' instead of 'comparing'

Line 234 – Change to 'its accuracy over the breakup season…'

Line 235 – Since you discuss the break up at three sample ports, and present the results in Figures 8 and 9, I think it would be important to include a map of these three locations, indicating which pixels you use to extract this data. I am curious how the land mask affects the data, i.e. how close are the pixels you are using to the actual port? Since you are using 31km ERA5 data, I would suspect that the pixel you chose to represent each port is actually a distance away from the actual dock. In the end, I guess I am a little wary of how applicable your results are to local communities, as they are likely more impacted by ice break up on a smaller scale along the coast (for hunting and travel), whereas shipping operations are more concerned of the large scale ice break up along

shipping corridors. Some discussion of how the scale of your results impacts how they are used by different groups may help address this.

Line 237 – Capitalize ‘figures’

Line 242 – Would recommend moving the figure reference to the end of the sentence, and putting it in brackets OR starting the sentence with ‘In figure 8, 30 lead day predictions for freeze-up are more…’

Line 242 – Any idea why this is? I am curious why the predictions varied at the different town ports and would think a discussion of this would add to your paper.

Line 252 – If you have space in your word count, I would recommend listing the 8 variables used in the Basic model, and the other variables added to the Augmented. This would help refresh the reader’s memory as to how these two models vary.

Figures 8 and 9 – If possible, the font size should be increased, particularly for your axis labels. This might take some reorganizing of your figure boxes – maybe you could rotate the ‘model’ and ‘day’ labels on the far left of your figures?