The Cryosphere Discussions

The Cryosphere Discuss., referee comment RC1
https://doi.org/10.5194/tc-2021-21-RC1, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on tc-2021-21

Anonymous Referee #1

Referee comment on "Accelerated decline of Svalbard coasts fast ice as a result of climate change" by Jacek A. Urbański and Dagmara Litwicka, The Cryosphere Discuss., https://doi.org/10.5194/tc-2021-21-RC1, 2021

Review of TC-2021-21 by J. UrbaÅ□ski and D. Litwicka

Review:

Although this manuscript presents some promise as a demonstration of harnessing the power of Machine Learning (ML) for analysis of time series, it falls short in a number of key areas. These are detailed below in "Major issues". I also list many minor issues following this. I feel there is considerable work required to address the major comments, unfortunately.

Major issues:

1) The structure of the manuscript is OK up until L155, but quite poor from then onward. For example, there is no "Results" section. Results are, instead, scattered throughout the remainder of the manuscript, and even right up until the final paragraph of the conclusion. A revised manuscript would strongly benefit from containing all new results within a results section, then sticking to the tried-and-true Discussion and Conclusions following this. I find that the Elsevier 11 step guide to formatting a scientific paper helps here: https://www.elsevier.com/connect/11-steps-to-structuring-a-science-paper-editors-will-take-seriously

2) I have some serious reservations about the method underlying the conclusions in this paper. These are expanded upon in three parts:

a) In line 127 you say that for Random Forests, "extrapolation beyond the range of values in the training set is not possible" - however this is exactly what you do, both back in time (where you have shown the Freezing Degree Day (FDD) values are higher than in the training set), and forward in time (where FDDs are much lower). Ideally you'd want to split training/testing toward both the end and start of your validation time series to see if the model performs similarly at both ends - this might then give some confidence for extrapolation, but without this, it's unconvincing. In fact, I believe the discrepancy between observed and modeled fast ice in Fig 10 may be evidence of this poor performance when extrapolating.

b) You indicate that the default hyperparameters generally achieve good results, but i) there is no attempt to verify this in any way for this study, and ii) these default hyperparameters are not given anywhere in the manuscript. You don't even really state that you used the default hyperparameters. Also, without a sensitivity test for hyperparameter selection, you're essentially "breaking" the paradigm of using a training/cross-validation/testing split (because you aren't able to use the cross-validation properly). You also don't indicate how temporal and spatial autocorrelation are avoided in selecting the training dataset. From your description on Line 135, I half suspect you didn't even try to avoid this (e.g., neighbouring pixels are going to be highly correlated. Choosing one pixel for the training and the neighbouring pixel for the testing is obviously (and inappropriately) going to confer high skill to the model). In light of these suspicions, I suspect your very high claimed R^2 of 0.95 is a symptom of this, and that the "true" performance of the model, (e.g., when the validation data are not very highly correlated with the training data) is much lower.

3) The manuscipt suffers from a lack of clarity and an excess of unnecessary jargon - particularly in the Methods sections 3 and 4 (e.g., the definition of OOB which is neither used again nor even referred to). I also don't need to know details like the fact that your vector layers were zipped.

4) Figures are clear, but their presentation quality is not ideal. Many figures suffer from a lack of proper sentence case; lack of labels; lack of units; lack of sufficient caption; or strange red underlines of text suggestive of a screenshot of a word processor.

5) A major reference is completely missing from consideration: Yu et al., 2014, Journal of Climate (doi:10.1175/JCLI-D-13-00178.1). In this work, the authors (of which I am not one) digitise pan-Arctic fast ice charts back to 1976, not only including Svalbard, but also analysing the decrease in fast ice around Svalbard. Thus, your sentence in Line 55 "Unfortunately, this distribution in the last quarter of the 20th century is unknown" is not accurate. This might have considerable consequences for the justifaction of your work. However, I still believe there is value in your technique. But I don't think your work, which essentially attempts to model fast ice extent in this area, can be published without some kind of comparison to the Yu dataset. To be clearer, I think your work should be re-formed around a comparison with the Yu dataset back in time - essentially treating the Yu dataset as truth (NB - as a general statement, in my experience, ice charting of fast ice can be, at times, quite far from the truth! But there's certainly nothing better without re-interpreting the satellite imagery yourself).

6) I'm unconvinced that your choice of 0 C is appropriate as a Freezing Degree Days threshold. I was under the opinion that 0 C should only be used for fresh ice. I checked the Lepparanta 1993 paper and, although it's not explicitly stated, close reading reveals that it's clear that a value of around -2 C should be used for sea ice. I realise the Polar Science Center mentions a value of 0 C on their website, but the intended ice salinity is ambiguous on that site and still think this value should only be used for studies of formation/melt of fresh ice.

7) Fig 9 has a couple of major problems: Panel a): Your color scale shows fast ice time difference, but doesn't take into account that sometimes this difference is negative (e.g., around the northeast of Edge Island) - i.e., negative and positive differences are represented in the same color - not sure this is appropriate. Panel b): The diagonal lines going to the Isfjord higher-resolution domain imply that this is an enlargement, but it's not the case, as can be seen by the clearly different colours. This implies that your result is strongly dependent on the resolution of your hexagonal grid. But why is this the case? There are only three inputs to your model: FDD, TDD and day number of the ice season. And only one output: the ice coverage, which is trained by the ice charts. Which of these 4 things is different? (This brings me to another point: How was a spatially-complete field of

TDD and FDD generated? Is this what's attempted to be explained at L79? Linear interpolation between three observer stations? Even assuming the grid scale change results in a slightly different FDD and TDD across the two different grids, it should still be very similar). Anyway, I really can't understand why grid scale has such a huge effect on the result given the inputs are so simple.

8) Even after reading the relevant sections a couple of times, I can't quite figure out what exactly you asked your RF model to produce. My interpretation is that you fed your RF time series (daily? Or at the resolution of the ice charts?) of TDD, FDD and day of ice season, and asked it to produce time series of fast ice coverage, as trained by ice charts. Is that the case? Then you compared this time series with ice charts (in the validation set, many times). So you end up with two (daily?) time series of fast ice cover for each grid cell: One from the RF and one from the validation data (the ice charts). Are these both binary? Or is the RF-derived prediction a smoothly-varying value between 0 and 1? From these, as detailed in L135, you "evaluated the error using RMS". RMS probably isn't a good metric to use if both outputs are binary. If the RF output is smoothly varying it might be OK though. However, I have a major concern in that a majority of these values require absolutely no skill to predict (i.e., on ice season day 1, a blanket prediction of 0 fast ice is probably a good idea everywhere). Indeed, as the vast majority of the grid cells in your domain never have fast ice (as shown in Fig. 8), a blanket prediction of "0" for every cell and for every time step is probably not bad. Basically, I am suspicious that your RMS statistic is artificially lowered by the fact that almost all grid cells never have fast ice. Similarly, your R^2 metric may also be affected. Even for those cells with fast ice cover, there is a skew toward lower fast ice coverage, so this warrants a more appropriate consideration of errors, such as using precision/recall/F1 score instead of a simple RMS.

9) Your abstract tells a story of an accelerating decline of fast ice extent. However, you never show a fast ice time series in any of your figures. The closest we get to see is a reduction in FDD in Fig 7 - however this isn't a paper about a reduction in FDD! Given you asked your RF to recreate fast ice extent, I think we need to see a time series of the primary output. In fact, in Line 188 you draw conclusions about fast ice directly from this plot of FDD - but would be so much more believable if you plotted a time series of fast ice extent.

Minor comments:

8: I don't think ice charts count as in situ observations.

13: It's not clear why you compare a 14 y period with a 6 y period. These seem quite arbitrary.

14: Avoid "now" - especially since your most recent data are already 2 years old.

14: What time period does a further two degrees warming correspond to?

17: Using the 66 degrees 33 min N definition of the Arctic, this doesn't seem to be true (much more land area in Scandanavia).

17: I don't believe that the location of Svalbard is strongly influenced by any current.

27: Needs a reference.

31: Needs a reference.

34: Extraneous space in this line (and a few other places in the manuscript).

46: Citation style inconsistence with Pavlov ref. Also extraneous hyphen.

Fig 1 caption: Missin degrees and minutes.

62: Most of these datasets are not in situ.

65: Missing circle over capital A (sorry, I don't know the correct name) in Alesund. Elsewhere in manuscript too.

75: We don't need to know the details of the libraries used.

Fig 2 caption: No mention of the date of this chart.

121: Should be just "Breiman (2001)"

171-177: Much of this is repeated - and it also reads like a conclusion.

Fig 7 caption: I think "since" should be "prior to"

186: Repeat of earlier.

191: Significance never tested.

192: Using units of deg C.day feels so imappropriate here, despite being numerically correct. K.day would be preferable.

Fig 8: Why are the two timescales in the comparison so different? In general, throughout the manuscript, the timescales being considered are not well-justified.

Fig 8 caption: Wait a second. The first paragraph of the discussion says you don't use a geographically-weighted random forest - but here and in Fig 10 you say that you do.

221: Typo in Conclusion.

228: "a bit" is too colloquial. Also, why are these points not annotated on Fig. 10?

233: What years to +2 and +4 degrees correspond to?

241: "wits" typo.

242: "rice" typo. Also this whole paragraph is not appropriate in the conclusions section.

275: double comma.

277: "influence of"

314: Capital B for Bay needed.

Does the paper address relevant scientific questions within the scope of TC?

Yes.

Does the paper present novel concepts, ideas, tools, or data?

Yes.

Are substantial conclusions reached?

Yes.

Are the scientific methods and assumptions valid and clearly outlined?

No.

Are the results sufficient to support the interpretations and conclusions?

No - questionable results stemming from choice of FDD threshold and extrapolation of random forest technique. Questionable methodology in the choice to not use cross-validation, nor to test effect of hyperparameters. Questionable methodology for selection of training data (suspect data independence not guaranteed)

Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

No - Random forest hyperparameters not given. Methods section readability is low due to focus on unimportant details, e.g., file formats.

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

Yes.

Does the title clearly reflect the contents of the paper?

Yes.

Does the abstract provide a concise and complete summary?

No.

Is the overall presentation well structured and clear?

No - structure is poor from line 155 onward (e.g., no "Results" section; results scattered throughout discussion and conclusion).

Is the language fluent and precise?

Not always - but not a major problem.

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

Yes.

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

Yes - extensive clarification and reduction of duplicate information required.

Are the number and quality of references appropriate?

No - Important reference missing completely (Yu et al., 2014. doi:10.1175/JCLI-D-13-00178.1).

Is the amount and quality of supplementary material appropriate?

N/A.