Earth System Dynamics

Open Access

Discussions

ESDD

Interactive
comment

Interactive comment on "September Arctic Sea Ice minimum prediction – a new skillful statistical approach" *by* Monica Ionita et al.

Anonymous Referee #3

Received and published: 15 November 2018

Review of

September Arctic sea ice minimum prediction - a new skillful statistical approach

by

Ionita, M., et al.

Summary: Multi-annual time series of sea-ice extent and various oceanic and atmospheric parameters are used to establish so-called stability maps. These maps inform about location, degree and significance of linear correlations between sea-ice extent and the other parameters where these correlations are particularly stable over time using a 21-year long time window. Based on the information provided by these maps a step-wise multiple linear regression analysis is carried out to find the optimal param-

eter combination to predice pan-Arctic and Eastern Siberian Sea (ESS) September sea-ice extent (SSIE). Different start months (May to July) are tested. The result of this regression analysis is then used to predict pan-Arctic and ESS SSIE based on the optimal parameter combination. The predicted time series of SSIE are compared to the observed ones. Evaluation of the prediction skill is carried out using various statistical metrics, suggesting that the method has a) high predictive skill and that b) the skill improves the more parameters are used and c) the closer one moves in time to September.

The manuscript might be a useful addition to existing knowledge. However, the lack of detail in description of data and methodology, the lack of transparency in how certain choices in the methodology were made, and the lack of a critical discussion of the methodology and its potential limitations as well as of the results themselves require that a lot more work needs to be put into the manuscript before it can be accepted for publication. It cannot be published in its current form.

In the following I have listed my main concerns in the general comments which are followed by specific comments, which are either a specification of the general comments or point to other weaknesses / ways to improve the manuscript; finally there is a small typos section.

General Comments: GC1: The introduction requires a revision of cited literature which partly seems to be outdated and partly is simply missing.

GC2: The introduction should work out better the choice of the statistical method used and of the atmospheric and oceanographic parameters used. In my view the attempts listed do not provide a clear motivation of the methodology and set of data chosen by you. I suggest to better motivate the choice of method & data by providing more quantitative evidence about which methods failed so far (and why) and which parameters failed so far (and why).

GC3: The description of the used data lacks quite some information and motivation.

It appears that some of the used data sets are outdated. See the respective specific comments.

GC4: The description of the generation of the stability maps should be improved. There are some open questions. See the respective specific comments.

GC5: The description of the stepwise multiple linear regression analysis should be improved. It is very short, some information is missing and, in particular, it is not possible to understand why and how you ended up with the optimal parameters shown in Figures 1 through 3 and Table 1 as well as the supplementary material.

GC6: A discussion of the results is missing completely. Elements could be, for instance: - critical reflections on the method: How physically meaningful are the parameters found to the prediction? Only very briefly mentioned is that a change in the period used for the "calibration" between 15 & 25 years does not make much of a difference. To give this 1-sentence statement a better basis one could show similar figures as Figure 4 using development periods of, e.g., 17 and 25 years (a symmetric change around the chosen 21 years) and discuss how the prediction skill changes and how the improvement in prediction skill from May to July changes. If there is not too much change then this is not necessarily a good sign of the credibility of the method, is it? - critical reflection on the quality of the data used - reflections on the increase in prediction skill the closer one gets to September + an explanation why this increase is more pronounced for the ESS. - reflections on the years where the prediction is particularly good or bad because of which reason? - reflections on your skill metrics: Have others used these metrics to assess sea-ice prediction skill as well already? If yes, which results did they obtain? If not: Which other measures were used and how comparable are your results to those of the others? - reflections in the practical side: Optimal skill is achieved when using a quite large suite of parameters for July, i.e. just two months ahead of the September sea-ice cover minimum. Is this what we aim for? Aren't we after a minimum set of parameters to be used already in May to predict the September sea-ice cover minimum, i.e. four months in advance? The more parameters you

need the more uncertainties (from the larger number of parameters required) might be loaded into your prediction. In that sense it would have been extremely useful to see a discussion about what is the minimum set of (which?) parameters in May to achieve a prediction skill of X.

Specific Comments:

Page 1, Line 23: You include sea-ice thickness here albeit our knowledge about sea-ice thickness decline seems much less solid than our knowledge about sea-ice extent reduction - and to my knowledge there is not yet scientifically proven evidence that the observed sea-ice thickness reduction is also caused by climate change. It is very likely from physical principles that this is the case, but I am not aware that there has been a paper like Notz and Marotzke, 2012, where this issue has been discussed.

Page 1, Lines 27/28: "Grosfeld et al., 2016" –> Albeit I appreciate the activities at AWI to enhance its service with regard to informing scientists and the wider public about topics relevant in polar research, I doubt that in a publication like this manuscript the citation of Grosfeld et al., 2016, should be one of the major citations when it comes to the observed sea-ice decline in the Arctic. Also "Serreze et al., 2007" is about ten years old a citation. Please look for more recent citations of Arctic sea ice decline in the white literature ... Cavalieri and Parkinson, 2012; Comiso et al., 2017; to mention two.

Page 2, Line 4-15: Since your background and motivation to carry out this study seemingly is driven by shipping activities I would have expected a more thorough review of what is available in the literature towards this topic. I would have liked to see, for example, Melia et al., and Pizzolato et al., both 2016, both Geophysical Research Letters; citing the AMAP report is fine (I have it in my shelf as well), but this is from 2004 = almost 15 years old and there have been revisions of this material. In addition, when it comes to potential challenges, then you could have looked into papers such as Lasserre, F., 2014, Polar Record and Larsen, L.-H., et al., 2016, Polar Research and I

bet there are many more papers you could use to support your statements in the last lines of this paragraph.

Page 2, Line 26: Chevallier et al., 2013 and Sigmond et al., 2013 are missing in the reference list.

Page 3, Line 8: This sounds like a valuable approach, however, as you wrote earlier in the introduction, the Arctic system is a very complex system and in the light thereof I find this statement puzzling. It can well be that the correlation between one pair of predictor and predictant changes due to changes in the correlation between another pair which is related to pair one. The influence of one particular parameter to SSIE can grow relative to the influence of another parameter. While correlation with this latter parameter might stay the same, correlation to the former parameter might change (e.g. increase) such that its skill to forecast SSIE increases as well. This approach might therefore a-priori exclude skill-improvements over time.

Page 3, Lines 11-15: This is a fairly larger number of parameters. I am wondering whether, e.g. with the SLP distribution one does not already have enough information about USURF and VSURF since the latter are primarily driven by the pressure gradient (provided) and the air temperature gradient (provided). Or in other words, what is the motivation of your choice of parameters (see GC2).

Page 3, Lines 21-25: - This paragraph lacks information about the time period, grid type, grid resolution, temporal resolution, and satellite from and for which the SIC is obtained. - Please also correct "NSIDC bootstrap algorithm" with "Comiso bootstrap algorithm"; NSIDC is simply the organization hosting and distributing this data. But see further below. - While it is clear from the introduction that you will use certain atmospheric and oceanographic parameters it is not clear why you require SIC and in particular the sea-ice index. This needs to be pointed out either in the introduction or at the beginning of section 2. - Please check whether it is "sea ice extent index" or "sea ice index". - You might also want to mention that the sea-ice index is just one

number per month while in case of the SIC we are talking about maps. - I do not understand why you are using the Comiso Bootstrap algorithm sea-ice concentration from the SIC data set you are citing. If I go to the doi cited behind the Meier et al. 2013 reference then I end up with the NOAA/NSIDC climate data record (CDR) of sea-ice concentrations, version #2. I would like to know why i) you did not use the main CDR, i.e. the combination of Comiso Bootstrap and NASA-Team algorithms, and ii) you did not use version #3 of this data set: https://nsidc.org/data/G02202/versions/3 Please note also, independent of whether you are using #2 or #3 you would need to cite the Peng et al., 2013 paper (see "Citing These Data" under the just above given URL).

Page 3, Line 26-30 / Page 4, Line 1-3: - What is missing here is the time period. For which time period did you actually obtain the reanalysis and also the other data sets listed in this paragraph? - What is your motivation to use a coarse resolution, outdated atmospheric reanalysis when i) there is an updated, finer resolved reanalysis of the same kind available from the same provider: NCEP-DOE and when ii) there are other reanalyses available which might perform better than NCEP/NCAR in the Arctic and the parameters chosen (e.g. JRA-55, ERA-Interim, MERRA-2, etc.)? Please moti-vate your choice. I guess for your approach it is quite important to have a re-analysis data set which is as accurate and realistic as possible. - What is the advantage of using the ERSST data set over using other, finer resolved SST data sets? While you are using Version #4b there has been a version #5 out for a while, see: Huang, B., Peter W. Thorne, et. al, 2017: Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5), Upgrades, validations, and intercomparisons. J. Climate, doi: 10.1175/JCLI-D-16-0836.1 Would you mind checking whether version #5 is more ac-curate in the Arctic Ocean than version #4? I guess it is worth it. It might be good as well to cite a paper or two in which you motivate your choice of using this SST data set over other SST data sets. - Finally, for the global ocean heat content data sets it would also be very good to have a statement pointing towards why this particular data set is chosen.

Page 4, Lines 6-22_ - Line 8/9: Please place the publications related to streamflow predictions behind the "... Danube river)" to ensure easier association of topic and citation. - Line 9/10: When you are referring to "region" then you mean regions where the spatiotemporal distribution of the predictors is stably correlated with the Pan-Arctic SSIE? - Lines 10-12: "from previous months (years)" –> unclear. I assume your time step is a month, because you are using monthly data. Why years? Further, since you did not specify yet the length of the time period you are investigating mentioning of "moving window of 21 years" remains unclear. What is the maximum or minimum time lag with which you compute the correlation? Why 21 years? - Did you carry out any pre-processing before you perform the correlation analysis? From the data section it becomes vaguely clear that the involved data sets potentially have a different grid resolution and are also on different grids. - Is the Student t-test used a two-sided one? - What is the motivation to go down to significance levels of 80%? Usually, researchers are using significance levels of 99% or 95%. - Line 14: "for more than 80% of the 21-year windows" –> It is not entirely clear what you did here. Please be more accurate in the description of how you performed the correlation analysis. If "1979-2007" is the length of your entire data set (I assume now from Line 15) and your moving window is 21 years long, then you have 9 years (=108 months or 9 seasonal cycles) in total for which you compute the correlation. If the correlation is above a certain significance level in 80% of these 9 years, i.e. 7.2 years, then the correlation is considered stable? Why? What is scientific rationale behind this choosing this threshold? - "significance levels that define the stability of the correlation vary within reasonable limits" –> I don't understand this. What are "reasonable limits"? Do you mean that if you are using other significance levels (SL), say 94%, 92%, 87%,and 82% instead of the mentioned ones, doesn't change the result, i.e. the general pattern in the stability maps? Why should it? You are showing a whole suite of SLs already anyways and at the end you take the 90% SL (Line 21) as the one to base you analysis on. I find this statement confusing. I'd rather ask why you did not take a SL of 95% instead of asking whether the stability maps would change with different SL. What is the motivation to vary the length of the

moving window the way you wrote, i.e. 15-25 years? After all, the main contributor to the correlation possibly is the seasonal cycle in the predictors and the predictand. - Lines 21/22: How is the detrending done? Did you compute average seasonal cycles - for each grid cell - for which time period? - It would not hurt to refer to Figure 1 already in this paragraph since you are describing the significance levels in a quite detailed way already.

Page 4, Lines 24-28: - Line 26/27: A good place to again refer to Figure 1.

Page 5, Lines 1-7: - This multiple linear regression is one of the key ingredients of your paper. I therefore suggest that you give more details, like a) How did you technically implement the stepwise linear regression? What is the step size? Are we talking about temporal steps or about steps in terms of parameters used for the linear regression? How is the prioritizing of the predictors quantified? Are the partial correlations in this step wise approach as high as those shown in the stability maps? b) What is the error? How did you derive / quantify the error? c) I assume that Y is dimensionless since you term it an index? Otherwise it has unit square kilometers. d) How is ensured that the equation in Line 3 provides a correct physical unit at the end?

Page 5, Lines 10-25: - This paragraph belongs to either the introduction or the methods section. - Since you have given the citations of the data sets used already in the data set subsection there is not need to repeat them here. - Lines 11/12 vs. lines 14/15: You state that sea-ice cover and snow cover belong to the long-term memory components but then use OHC, OT100 and SST as long-term predictors. Wouldn't it be more straightfoward to only refer to the oceanic components in Lines 11/12 and leave out snow cover and ice cover? - Line 18: "if predictable" –> I don't understand this statement in this context. Why has the atmospheric circulation to be predictable if you aim for the prediction of Septemnber SIE based on spring atmospheric conditions? - Line 21: What do you mean by "advective parameters"? - Lines 22-25: These two sentences should perhaps be re-formulating for clarity and to avoid repetition along the lines: "Atmospheric moisture content, e.g. clouds, water vapor content, has an impact

on the net surface radiation balance and hence also on the SSIE (Kapsch et al., 2013, 2014). As a measure for this impact we use the precipitable water content (PWC) as an additional predictor."

Page 5, Line 26 to Page 6, Line 4: - This paragraph belongs to either the introduction or the methods section. - Line 1: "SSIE index" –> why index? I thought you are after the sea-ice extent? - Lines 2/3: "with different lags, depending on the variable" –> It seems you are using different time lags for different parameters in the correlation analysis? Please specify why. Please detail the time lags associated with which parameter. Line 3/4: "The optimal predictors are defined as the average values over the stable regions for each gridded parameter." –> unclear. Did you take a stability map between, say SSIE and an arbitrary parameter, e.g. SST, and average over the SST within the region defined as having a correlation at 90% significance? And then? Then you have an average SST value ... fine ... and next? Why is this the "optimal predictor"? Further, when doing this averaging, do you take into account the actual correlation value as well or do you only use the significance as a criterion to select over which SST values you are averaging?

Page 6, Lines 5-30: - It is not clear how you end up with the variables and time periods shown in Figures 1 through 3 and Table 1. It is not clear why in particular in some cases optimal predictors either are based on monthly values (of one or even two different months) or are based on seasonal averages; you have not introduces seasonal averages at all yet. This all seems quite arbitrary. - It it not clear what the criterion is to place black boxes in Figures 1 through 3 and what their meaning is. It is not clear in particular how the location of the black boxes go along with the notion that only in these stability maps only regions with > 90% significance are used. This all seems quite arbitrary. - Figures 1 through 3: If the 90% significance level is as important as I think based on what you wrote, then I suggest to mark the 90% also by a change in color in the stability maps. Currently, 90% significance is right in the middle of (regular) red or blue. - In the captions of Figures 2 and 3 you need to state that these

are the ADDITIONAL parameters (additional with respect to May, Figure 1) on which the prediction of the September sea-ice extent is based. - Lines 5/6: What is meant by "...we have retained all the stable regions ... for all variables based on previous months' data."? - Lines 11-22: The entire issue about AMO needs to be put into the methods section where you can introduce this as an important additional parameter. It should be introduced in the RESULTS section.

Figure 4: I suggest to change the y-axis title to "Sea-ice extent anomaly []"; currently it is "Sie ice extent []". This applies also to Figure 5. For Figure 5, I in addition note that the range of the y-axis differs between a) and b)&c); I suggest to make this consistent as there is no physical reason to have differrent ranges. Another comment, maybe a matter of taste, but I would not call your comparison of the last 7 years' skill of the prediction a "validation" - this is an evaluation. I would not call your 21-year period "calibration". Instruments like a pyranometer, pyrgeometer, anemometer, etc., are calibrated. You use that period to develop your method. Hence, I suggest to speak of a "development phase" or "training phase" and an "evaluation phase" It would be good, finally, to also show a graph of the difference between observation and prediction in relative terms; a difference of 300 000 kmˆ2 with respect to the sea-ice extent of the entire Arctic is different in relative terms than a difference of 100 000 kmˆ2 with respect to the ESS sea-ice extent. That way you can better quantify the skill of your method.

Page 7, Lines 1-13: - Line 8: Please make clear that these are again additional parameters, i.e. on top of those for May (Fig. 1) and June (Fig. 2).

Heading of section 3.2: I would reformulate the title of this subsection into: "Application of the methodology for regional SIE prediction" Also, my general view is that you should include all the supplementary figures and table of this subsection into the regular paper.

- Lines 21-25: I don't think that Table 2 and these lines starting with "Moreover, when looking ..." are required to motivate your look at a specific region of the Arctic Ocean.

- A link / hint towards how you ended up at the optimal parameter combination would

be an asset here as well for the correct understanding of your results (see also my comments with respect to Figures 1 through 3 and Table 1 further up).

Page 8, Line 19 3/4: A discussion of the results is missing completely. See GC6

Page 8, Line 21 to Page 9, Line 2: - Line 23: "Although" ... –> Why? - I suggest to move Line 24 - end of this paragraph to a later place in the conclusions. Generally, starting the conclusion with a short summary about what has been done in this study would put whatever information coming later in the conclusion into a better context. - L28-30: I am not sure what the mentioning of the Parkinson et al (2006) paper and the excurse on modelling has to do with your work. After all, inconsistencies between model and observation with regard to sea-ice cover can have MANY reasons. The fact that OHC is one of these is neither new nor would I use that as one of the highlights to show why your study is valuable.

Page 9, Lines 3-24: This is part of the discussion - see GC6. - I suggest to split this paragraph into two in Line 16. the first one could talk about the teleconnections with, e.g. AMO and the impact the AMO has on which variable used in your method - if you find that relevant. Frankly speaking, if Yu et al. (2017) found that a strong link between these large scale pattern - why didn't you use correlations of between sea-ice cover and the strength of these patterns for your predictions? The second one could talk and refer to the regionally / locally driven feedbacks between ocean, sea-ice and atmosphere. You refer to SLP under these more regional forcings (Line 17) - but isn't particularly the SLP distribution the one which is directly linked to PDO, AMO, AO, NAO, etc.?

- Line 4: What do you mean by "climate variables"?

Lines 4-6: - "(the stable regions over the western European coast)" –> I do not find these in your figures. - OHC: For most of the areas shown in Figure 1 a) the correlation with OHC of SON of the previous year (?) is stable with than 90% significance and hence does not enter your prediction according to your description of the method. Does

this change for June & July as starting months of the prediction? - SST: Figure 1b) shows two areas of significant correlation with SST of MAM, a positive one in the Bering Sea and a negative one in the Barents Sea. You only used the one in the Bering Sea. Why?

- Line 10/11/12: Why "anomaly"? - Line 17 ++: I would about the "and so on" and I would try to be as explicit as possible here, referring to the respective Figure(s) for better understanding. It is crucial to discuss these influences in the light of which SLP distribution, associated VSURF and USURF distributions and PWC and TT patterns belong together physically and whether these are reflected in a consistent way in your stability maps. There is a lot to see in these stability maps and potentially also a lot to discuss albeit with the danger that one over-interprets correlations. In that respect you could improve the value of your paper and the discussion / conclusion in particular. A discussion could and should include more information about your choice of regions used for the prediction (black boxes in Figures 1-3). - Line 20-23: You mention PWC and air temperature in Lines 20/21 but then refer to "transport" in Line 23 - so how about the meridional transport of PWC?

Page 9, Line 25 - Page 10, Line 5: - Line 28: 1996 and 2007 lie within your development or training phase while only 2012 lies outside it. Hence one could argue that the only true year for comparison is 2012 for which the agreement is not as good as for 1996, for example. You could include 2013 as well, for which the prediction is quite good. It is further interesting to see that for ESS a start of the prediction in July results in a substantially better agreement for the two years of extreme pan-Arctic minima 2007 and 2012 (compare Fig. 5c with Fig. 5 a). In contrast, for the entire Arctic it does not really make a difference whether you start in May (Fig. 4a) or July (Fig. 4c): the discrepancy between observations and prediction remains unchanged. But this of course belongs to the discussion, in which you need to critically assess your approach to here, in the conclusions, give explicit recommendations about which starting month and which parameters are suited best to achieve the best prediction. Lines 4-5: Yes,

the concept can be used but how much better is it compared to other systems and concept. This you did not demonstrate and/or discuss. Therefore I would delete this last sentence. Another argument for deleting this sentence is that, e.g. shipping, might not be too much interested in the pan-Arctic sea-ice extent. Sea-ice area or, even better, the sea-ice distribution will be a better measure. As an example: The SSIE of the ESS is unchanged compared to winter but the sea-ice area reduced to 1/2 of the winter value if half of the ESS is ice covered by 75% sea ice and half of the ESS is ice covered by 25% sea ice. Hence the value of regional sea-ice extent prediction for shipping remains questionable.

Typos: - Please decide whether you use the American or British way of writing "skillful".

Page 2, L12: "flows" –> "floes"

Page 3, Lines 20 & 24: "extracted". I suggest to write "obtained" or "downloaded".

Page 4 Line 14: "level" –> "significance level" Line 24: "forecast m all" –> "m"? Line 27: "each ... parameters" –> "each ... parameter"

Page 5 Line 18: "substantial" –> "substantially"

Page 7 Line 23: "EES" –> "ESS"

Page 9 Line 11: "EES" –> "ESS" Line 12: "form" –> "from"

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2018-61, 2018.