Interactive comment on "September Arctic Sea Ice minimum prediction – a new skillful statistical approach" by Monica Ionita et al.
Anonymous Referee #1

The authors use a statistical model to skillfully predict the Arctic September sea ice extent and the regional East Siberian Sea ice extent up to 4 months ahead. They combine several oceanic and atmospheric parameters and sea ice extent itself from previous months and perform a multiple linear regression. Variables and regions are selected based on stable teleconnections between the predictors and the predictand. This study is an important contribution for seasonal Arctic and regional sea ice predictions. The predicted skill is higher compared with previous studies. The identification of relevant regions and parameters is useful for understanding processes and changes in climate models. I reviewed previous versions of this manuscript and I am pleased with the current version. The focus on de-trended time series and the separation into a calibration and validation period increases the robustness of the results. I strongly recommend publication and would like to make a few minor comments, only.

We thank the reviewer for the comments and useful feedback regarding our manuscript. Please find below our responses to the reviewer's concerns. Comments will carefully be included in the revised version, as they will help to improve the clearness and scientific content.

Minor comments:
1. Section 2.2: Give reference to stability figures and remove sentence about colors.

The text has been modified following the aforementioned suggestion.

2. Section 3.1: Would be nice to get some information about the impact of the individual predictors. It is surprising to see that only March ice extent is used for prediction of pan-Arctic sea ice extent based on June and July data. Is there no additional benefit from April, May and June ice extent?

The revised version of the manuscript has been substantially modified trying to take into account all the reviewers suggestions.

3. Section 3.2: Would recommend to rename header from "Robustness of the methodology" to "Regional September ice prediction". From Table 2, only Lag 0 results are discussed.

The title of Section 3.2 has been renamed from "Robustness of the methodology" to "Application of the methodology for regional SSIE prediction". We discussed just the Lag 0 because our aim was to identify the region which shows the highest correlation with the pan-Arctic sea ice extent in September.

4. Conclusion: "Moreover, our statistical model is able to properly reproduce the years with extreme low / high sea ice extent, both at pan-Arctic level as well as at regional scale (e.g., 2007 and 2012 – low SSIE and 1996 – high SSIE; see Figure 4 and Figure 5)." Given that only 2012 is within the validation period, it is questionable how robust this statement is.

We agree with the reviewer's comment and we took his recommendation into account and we have modified the text accordingly.

5. Figure 5: Use same legend for all sub-figures.

Modified as suggested.

6. Missing reference: Petty et al. 2017,
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016EF000495

The reference has been added in the revised version of the manuscript.

Interactive comment on "September Arctic Sea Ice minimum prediction – a new skillful statistical approach" by Monica Ionita et al.
Anonymous Referee #2

In this paper, the authors propose a statistical model to predict the Arctic September sea ice extent (SSIE) and East Siberian Sea ice extent (ESSIE) up to 4 months ahead with a high predictive skill compared to previous studies. Stability maps and stepwise multiple regression analysis are applied to find the optimal predictors for the model in regions and variables respectively. The results of prediction here are excellent and reliable. I believe the approach to build statistical prediction model between the predictors and predictand could be widely used in more climate predictions. I recommend to publish this paper and would like to make a few minor comments.

We thank the reviewer for the comments and useful feedback regarding our manuscript. Please find below our responses to the reviewer's concerns. Comments will carefully be included in the revised version, as they will help to improve the clearness and scientific content.

Minor comments:
1. In Figure 1-3 and S1-S3, those regions inside the black boxes are used for SSIE. However, besides those regions, there are also other regions with significant correlation coefficients. Some regions are even more significant than those regions you choose. Why do you only choose those regions in the black boxes? Could you give an explanation?

An explanation why do we only choose those regions in the black boxes has been added in the revised version of the manuscript.

2. In Section 3.2, I recommend giving the definition or boundary of East Siberian Sea as well as other areas mentioned in Table 2.

Modified as suggested.

3. In Section 3.2, there is a writing mistake in the sentence ". . . coefficient between the observed and forecasted ESS SSIE is r = 0.94 (r = 77) . . .". I think "r = 77" should be corrected as "r = 0.77".

Modified as suggested.

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 parameter combination to predict 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.

We thank the reviewer for the comments and useful feedback regarding our manuscript. Please find below our responses to the reviewer's concerns. We have tried to answer those detailed and thoroughly including references to underline our arguments. Comments have been carefully included in the revised version, as they helps us to improve the clearness and the scientific content.

General Comments:

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

The introduction part has been modified and up to date references have been added.

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).

Introduction part has been modified according the reviewers suggestions. Therefore, the introduction starts with an overview of how Arctic sea ice is changing and which role an accurate prediction of the sea ice situation plays in terms of economics. Then we turn to the current possibilities of predicting the sea ice situation. This is followed by a paragraph discussing the different approaches how they perform and which are there limitations. This leads to the final paragraph where we highlight the motivation and why we try to come along with the above-mentioned challenges by using a statistical approach that is well documented in the literature. Unfortunately, these references where not included in the submitted manuscript. We included them in the revised version. By this we can emphasize why the method performed is based on a state of the art and good statistical practice and

we can demonstrate the high expertise we have in regard to the statistical method we are presenting here. The same statistical approach is used in pre-operational mode for the streamflow prediction for Elbe and Rhine streamflow (Ionita et al., 2009, 2014, 2017; Meißner et al., 2017) and we strongly feel that the method as well as our results have both the physical and statistical meaning to be published in the context of sea ice prediction.

We have to mention that the aim of this paper is to show the advantages of our statistical approach. It is meant in a more technical direction. We agree that the psychical mechanism behind our stable regions have to be properly described and we have tried to do so in the revised version, but we have to keep in mind that a full overview for each region would require a much more longer manuscript and it is beyond the scope of our paper.

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.

We comment on this below in the specific comments in order to avoid any repetition.

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

We comment on this below in the specific comments in order to avoid any repetition.

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.

The description of the multiple linear regression has been improved in the revised version of the manuscript (see Methods section).

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.

Following the reviewer's suggestion we have modified the manuscript substantially, especially the introduction and the conclusions part, in order to take into account the aforementioned concerns.

#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.

We agree with the reviewer's concern, thus we have removed "sea ice thickness" from the text.

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.

We followed the reviewer's suggestion and added more recent references in this part of the manuscript.

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.

We have changed to introduction to take into account the aforementioned suggestions and references.

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

We added these references 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 predictand 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.

Although we agree with this comment, we have to keep in mind that each model (statistical or dynamical) offers room for improvement. In this paper, we want to show a different statistical approach, compared to the existing ones, and its potential advantages. By using running correlations, we try to avoid, at least partially, the non-stationarity issue between different variables. As mentioned above the same statistical approach is used in pre-operational mode for the streamflow prediction for Elbe and Rhine streamflow (Ionita et al., 2009, 2014, 2017; Meißner et al., 2017) and we strongly feel that the method as well as our results have both the physical and statistical meaning required to work in context of sea ice prediction. The method was also used to investigate the "Moisture transport and Antarctic sea ice: austral spring 2016 event" (Ionita et. al. 2018).

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).

If one looks at the stability maps, you can see that the stable regions are located over different regions for SLP, USURF and VSURF. In some cases, SLP is not even selected as a potential predictor. It is true that SLP, USURF and VSURF are interconnected, but in the current case, the stable regions are independent of each other, thus we choose to use all there variables.

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).

The data section has been modified following the reviewer's comment. For the forecast purposes in our manuscript we are using just the Arctic sea ice extent index. Thus, the other data have been removed from the revised version. We are not using the sea ice concentration, so we have deleted all the information about this part. The revised version of the manuscript has included a new Table 1, which summarizes all data sets used with source, temporal and spatial resolution as well as the references. Therefore, we think that the reader can easily catch up with the database of the analysis performed in this study.

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 motivate 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 accurate 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.

Again we are thankful for the reviewer comments on our data base. We have to point out that the choice of our datasets is mainly based on their availability. Although ERA-interim (HadSST) would be a better option compared to NCEP (ERSST) data, ERA-interim (HadSST) unfortunately are not updated in real time, thus making it impossible to be used in "operational" mode. We are participating in the Sea Ice Prediction Network forecast

challenge, thus we need the data to be publically available at the beginning of each month. We agree that it's optimal to have the proper reanalysis dataset, but we have to take into account also their availability in near real time. Regarding the ERSST data, there was a written mistake from our side. We have of course used the last version of the ERSST data for the forecast, namely the ERSSTv5, not the previous version. We have corrected this in the revised version of the manuscript. To make it easier for the reader, we also added an overview table (Table 1), listing all the datasets used in this study, their proper references and the link where the data can be downloaded.

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. –

Modified as suggested.

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?

The text has been modified accordingly to answer this question.

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%.

This section was rephrased in the revised version of the manuscript to clarify the questions raised. We described the time periods and we referred to the references where the method is used in order to underline the used moving window of 21 years. We have added also a table (table 2) in which we show the time lags used for each variable. The grid of each data set is now mentioned in the new table 1. For the statistical significance we used a two-sided student t-test – we have inserted a small paragraph stating this in the methods section. As we already mentioned we used 80% and 85% level as buffer zones. The data from these regions are not taken into account in our forecast exercise.

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.

We think concerning our method and expertise that parts of this comment are not applicable (e.g. the use of 94%, 92%, 87%,and 82% significance level). We use a well-established and published methodology. For the final predictors we select the regions where the correlation is stable above the 90% significance level. Everything below this level

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.

The detrending procedure has been included in the revised version of the manuscript in the section data and methods. We also make reference to Figure 1 in the description of the stability maps procedure.

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

Modified as suggested.

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?

In the revised version of the manuscript we added more information regarding the stepwise regression procedure. We also add a reference for the applied method.

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.

We removed parts of this paragraph, but we prefer to keep it in it's current location because we use it as an introduction to explain why we opt for the predictors we are using.

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 straightforward to only refer to the oceanic components in Lines 11/12 and leave out snow cover and ice cover?

Modified as suggested.

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 September SIE based on spring atmospheric conditions?

We deleted "if predictable" in the revised version of the paper.

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."

Parts of this paragraph have been removed and some part have been modified as suggested.

Page 5, Line 26 to Page 6, Line 4: - This paragraph belongs to either the introduction or the methods section.

The paragraphs has been moved in 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?

The text has been modified to properly capture the real meaning behind SSIE index. We are not using a gridded field, we are using just one time series, the September Sea Ice Extent, that's why we call it an index. Nevertheless, in the revised version of the manuscript we use the same definition throughout the manuscript, namely SSIE. Moreover, we have included also a table (Table 2) in which we clearly specify which time lags are involved for each variable and which kind of data (monthly and/or seasonal. This frames the analysis and helps to follow the method performed.
All the predictors used in the regression model are calculated based on the significance criterion. All the grid point in which the correlation coefficient between SSIE and large-scale data are significant at a significance level >90%, for more than 80% of the case, are used when defining an index (by averaging the gridded data over the black boxes in the stability maps).
In addition Table S1. showing the skill parameters (see supplementary file for definition) based on different statistical methods for the observed and predicted pan-Arctic sea ice extent in September with different time lags adds much information to understand the approach.

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.

The method and approach is not all "arbitrary" as mentioned by the reviewer. Nevertheless, we think the reviewer is correct that we have to present the paragraph more structured and clearer concerning the analysis. In the revised version of the manuscript, we have added Figure S2 in the supplement, which shows the workflow of our methodology and how we end up with the "optimal model".

-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."?

The captions of Figure 2 and 3, as well as for ESS, have been modified following the reviewer's suggestion. Also the text has been modified in the revised version of the manuscript.

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.

The AMO paragraph has been moved in the discussion part.

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 different 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.

Figure 4 and Figure 5 (now Figure 8) have been modified following the suggestions. Regarding the "calibration" and "validation" wording, we like to stress that this is the state of the art wording in forecasting. All the papers mentioned in the references, and not only, use the terms "calibration" and "validation" to train and evaluate their model. Thus, we choose to keep this wording also in the revised version of the manuscript. We also feel that by adding two new figures of the difference between observation and prediction in relative terms will not add much to the paper.

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).

Modified as suggested.

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.

We have modified the title of subsection 3.2 following the reviewer's suggestion. Moreover, we have now included all the former supplementary figures into the regular paper. In the actual supplementary document we have added new figures, for a better understanding of our stability regions determination.

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).

We have added a figure, which emphasizes the workflow of the selection of the optimal parameter combination (Figure S2 and additional text in the supplementary file).

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

The conclusion part has been modified substantially and now most of the discussion are integrated in the new part of the paper called Discussion and conclusions.

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.

Modified as suggested.

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.

The reference of Parkinson et al. (2006) is added as one example. We are aware the OHC is not a new highlight in term of predictability, but can be used to discuss potential issues.

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.?

As stated before, the conclusion part has been substantially modified and we added the discussion in this sub-section. We discuss the different issues raised here by the reviewer throughout the whole "Discussion and conclusion" sub-section.

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

This wording has been removed from the revised version of the manuscript.

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?

The text has been modified to properly capture what is found in the stability maps. As is stated in the description of the methodology we used all the regions which are above the 90% significance level. Moreover, we have added as supplementary file the original figures where we show all the stable regions used in the regressions model. What is shown is Figures 1 - 3 and Figures 5 – 8 are just the stable region that are used for the SSIE forecast after we applied the linear regression model.

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).

To take into account this suggestion we have added a new section in the revised manuscript, namely Discussion (section 4), where we try to incorporate the reviewers comments/issues regarding a more detailed explanation of the physical mechanisms and some issues related to our methodology.

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?

The text has been modified according to the suggestion.

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.

We have modified the revised version of the manuscript substantially trying to incorporate the suggestions of the reviewer. We have added a new section for discussion in which we discuss some of the issues as well as advantages of our methodology.

Typos: - Please decide whether you use the American or British way of writing "skillful".
Page 2, L12: "flows" –> "floes"
Modified as suggested.

Page 3, Lines 20 & 24: "extracted". I suggest to write "obtained" or "downloaded".
The respective line has been removed from the revised version of the manuscript.

Page 4 Line 14: "level" –> "significance level" Line 24: "forecast m all" –> "m"? Line
Modified as suggested.

27: "each ... parameters" –> "each ... parameter"
Modified as suggested.

Page 5 Line 18: "substantial" –> "substantially"
Modified as suggested.

Page 7 Line 23: "EES" –> "ESS"
Modified as suggested.

Page 9 Line 11: "EES" –> "ESS" Line 12: "form" –> "from"
Modified as suggested.

September Arctic Sea Ice minimum prediction – a new skillful statistical approach

Monica Ionita[1,2], Klaus Grosfeld[1], Patrick Scholz[1], Renate Treffeisen[1], Gerrit Lohmann[1,2]

[1] Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, 27570, Germany

[2] MARUM – Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

Correspondence to: Monica Ionita (Monica.Ionita@awi.de)

Abstract. Sea ice in both Polar Regions is an important indicator for the expression of global climate change and its polar amplification. Consequently, a broad interest exists on sea ice coverage, variability and long term change. However, its predictability is complex and it depends strongly on different atmospheric and oceanic parameters. In order to provide insights into the potential development of a monthly/seasonal signal of sea ice evolution, we applied a robust statistical model based on different oceanic and atmospheric parameters to calculate an estimate of the September sea ice extent (SSIE) on monthly time scale. Although previous statistical attempts of monthly/seasonal SSIE forecasts show a relatively reduced skill, when the trend is removed, we show here that the September sea ice extent has a high predictive skill, up to 4 months ahead, based on previous months' oceanic and atmospheric conditions. Our statistical model skillfully captures the interannual variability of the SSIE and could provide a valuable tool for identifying relevant regions and oceanic and atmospheric parameters that are important for the sea ice development in the Arctic and for detecting sensitive/critical regions in global coupled climate models with focus on sea ice formation.

1 Introduction

Arctic sea ice plays an important role in modulating the global climate system by influencing the atmospheric and oceanic circulation in Polar Regions. Moreover, it has a strong impact also on the global economic system through changes in marine and natural resources development. The sea ice extent over the Arctic region has undergone an extraordinary decline during the last decades that can be linked to climate change (Allison et al., 2009; Kay et al., 2011; Notz and Marotzke, 2012, Stroeve and Notz, 2018). The trends in the Arctic sea-ice extent are negative for all months, with the largest trend recorded at the end of the melt season in September (Serreze et al., 2007), with an average decline of 12.9% per decade relative to the long-term mean of 1981-2010 September average (Cavalieri and Parkinson, 2012; Comiso et al., 2017). These negative trends, with their environmental and economic implications as well as its impacts on human society, have led to a rising demand for accurate sea ice predictions at monthly, seasonal up to decadal time scales, which in turn will be able to address the growing demands from different stakeholders and the scientific community (Meier et al., 2014). As such, an accurate sea ice prediction plays a crucial role for ecosystems, coastal communities, planning for new shipping ports, oil and gas exploration and marine transportation. The ten lowest September sea ice extents all occurred in the past 10 years and climate projections indicate that the Arctic Ocean could be ice free (sea ice less than $1x 10^6$ km² for at least five consecutive years) in September in the second half of the 21st century (IPPC, 2013) . As a result, the ship traffic and Arctic resources extraction have already increased (Pizzolato et al., 2014). For example, the exploitation of shipping via the Northwest Passage or Northeast Passage could reduce the

navigational distance between Europe and Asia by ~40% compared to the route via the Suez canal (Schøyen and Bråthen, 2011). The reduction in distance compared to the Suez and/or Panama Canal routes could result in large cost saving due to reduced fuel consumption and an increase in the number of ships (Lassere, 2015). Melia et al. (2016) have shown that by mid-century, the frequency of navigable period will double and the routes across the central Arctic will become available. For example, for a high-emission scenario they have shown that by late century trans-Arctic shipping might become a commonplace, with the shipping season ranging from four to eight months. Overall, the summertime use of these routes by different vessels (i.e. cargo ship and tanks) has increased (Eguíluz et al., 2016), thus the need for a proper forecast for the Arctic sea ice conditions has become imperative. Currently, forecasting the open water route through the Arctic basin is accurate within 200 km when the predictions are initialized in July (Melia et al., 2016). As such, an early knowledge on the potential opening of the maritime Arctic routes could allow a better management for the shipping companies to optimize (in terms of time and costs) shipping routes between the Atlantic and the Pacific Oceans (Hassol, 2004; Smith and Stephenson, 2013). However, the opening of the Northeast and Northwest Passages does not guarantee ice free transects along the passages at all times and can always include the possibility of drifting ice floes, which for conventional ships poses high risks and potential environmental danger when getting damaged in case of accidents. Pizzolato et al. (2016) have shown, that despite the persistence of low sea ice condition since 2007, very little shipping activities has been recorded within the northern route of the Northwest Passage. This might be attributed to the multiyear ice concentrations in the Canadian Arctic waters, which strongly influences the shipping activity. Hence, a proper forecast does not imply a dangerous free transect as long as the Arctic Ocean is ice covered with thick multiyear ice for its larger parts over the significant times of the year.

Although the evolution of Arctic sea ice physical properties has been extensively studied, the prediction of detrended Arctic sea ice extent, with lead times of 3 months and longer, have not been very promising (Lindsay et al., 2008; Blanchard- Wrigglesworth et al., 2011). From a forecasting point of view, the evolution of autumn Arctic sea ice is closely associated with initial conditions in previous winter and spring. Different studies have emphasized that some parameters contribute significantly to the improvement of the seasonal sea ice forecast skill at different time lags (Holland and Stroeve, 2011; Lindsay et al., 2008). For example, sea surface temperature and sea ice concentration in spring are highly relevant predictors for the minimum Arctic sea ice extent (Drobot et al., 2006). Some studies suggested that accurate sea ice thickness could increase the forecast skill 2 months ahead (Day et al., 2014; Dirkson et al., 2017). Also, the spring melt pond fraction has been employed to improve the forecast skill of the Arctic minimum sea ice extent (Schröder et al., 2014).

Currently, there are different approaches used to make sea ice forecast: ice-ocean-atmosphere coupled models, statistical models, best guesses model and mixed models (Stroeve et al., 2014; Hamilton and Stroeve, 2016). From a statistical point of view, Drobot et al. (2006) showed that 46% of the pan-Arctic minimum sea ice extent would be predictable as early as February based on monthly sea ice concentration, surface albedo, downwelling long-wave radiation and surface skin temperature. Lindsay et al. (2008) have shown that their statistical model based on a wide range of predictors (e.g., atmospheric circulation indices, sea ice extent and sea ice concentration, ocean temperature at different levels) exhibited a greater skill in predicting the September sea ice extent (SSIE) than those by Drobot et al. (2006). The forecasts based on the state-of-the-art coupled atmosphere-ocean sea ice models (Chevallier et al., 2013; Sigmond et al., 2013) do not show better results when compared

with the statistical models (Kapsch et al., 2014; Schröder et al., 2014; Zhan and Davies, 2017). These caveats indicate that our understanding regarding the controlling factors of Arctic sea ice may still be insufficient. Overall, skillful forecasts extend only two to five months ahead, for the summer months (Stroeve et al., 2015; Schröder et al., 2014), regardless of the type of the model used for the forecast (dynamical or statistical). The results and error margins based on these different approaches have highlighted how difficult it is to make skillful prediction for the SSIE. This is particular true for the years with extreme low September sea ice concentrations (e.g., 2012 or 2007), with both, the dynamical and the statistical models showing similar limitations (Stroeve et al., 2015; Schröder et al., 2014; Stroeve et al., 2014; Hamilton and Stroeve, 2016). Stroeve et al. (2014) have shown that seasonal predictions of the SSIE are most accurate in years when the sea ice extent is near the long-term trend, but skillful sea ice extent prediction appear challenging in years when the weather plays a larger role (Hamilton and Stroeve, 2016).

In order to improve the monthly/seasonal prediction skill of the sea ice extent one possibility would be to identify stable predictors (the correlation coefficient between the predictor and the predictand does not change in time) and to develop a statistical forecast model based on these predictors. Following this idea, here we analyze the oceanic and atmospheric conditions associated to the SSIE in order to identify potential predictors based on a simple statistical methodology and placed them in a longer temporal context. Our statistical model takes into account different atmospheric and oceanic variables following the approach in Ionita et al. (2008, 2014, 2017, 2018). These parameters are: sea level pressure (SLP), air temperature (TT), precipitable water content (PWC), surface zonal wind (USURF), surface meridional wind (VSURF), the ocean heat content integrated over the first 700m (OHC), sea surface temperature (SST) and water temperature integrated over the first 100m (OT100), in order to calculate an estimate of SSIE. The paper is structured as follows: the data and methods used in this study are presented in Section 2, while the main results of our analysis are shown in Section 3. The discussion and concluding remarks are presented in Section 4 and 5.

2 Data and methods

2.1 Data

The monthly sea ice extent has been extracted from the National Snow and Ice Data Center ftp server (ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/north/) (Fetterer et al., 2016).

For the Northern Hemisphere temperature and atmospheric circulation, we use the monthly means of air temperature at 2m (TT), downward longwave radiation flux (DW), zonal wind (USURF), meridional wind (VSURF), precipitable water content (PWC) and the mean sea level pressure (SLP) from the NCEP/NCAR 40-year reanalysis project (Kalnay et al., 1996) on a 2.5º x 2.5º grid. Global sea surface temperature (SST) is extracted from the Extended Reconstructed Sea Surface Temperature data (ERSSTv4b) (Huang et al., 2014). This dataset covers the period 1854 – present and has a spatial resolution of 2° x 2°. The global heat content data in the first 700m (OHC) and the ocean temperature integrated over the first 100m (OT100) is extracted from the Global Ocean Heat and Salt Content database (Levitus et al., 2012; Boyer et al., 2013).

The monthly Atlantic Multidecadal Oscillation (AMO) index The AMO index has been calculated as the average of monthly SST anomalies with respect to the mean over the North Atlantic north of 25∘N (75∘W–7∘W, 25∘N–60∘N). For the AMO index computation, we used the RRSSTv5 data set (Huang et al., 2014). In this study, we

3

use the yearly mean of AMO index. Table 1 gives an overview of all the data sets included in the study. All used data sets have been detrended before the analysis by computing the linear trend for the entire time series/gridded fields in question. This trend was then subtracted from the initial time series/gridded data set. The linear trend was estimated using a least-square linear regression.

5

2.2 Stability Maps

The statistical model used in this study for the estimation of SSIE is based on a methodology successfully used to make monthly/seasonal streamflow predictions for the central European rivers (e.g., Elbe river, Rhine river, Danube river, Ionita et al., 2008, 2014, 2017, 2018; Meißner et al., 2017). Furthermore, for identifying the drivers

10 of the Antarctic sea ice variability (Ionita et al., 2018). The basic idea of this method is to identify regions where the spatio-temporal distribution of the predictors is stable correlated with the Pan-Arctic SSIE. The SSIE has been correlated with the potential predictors from previous months (Table 2) in a moving window of 21 years and the statistical significance of the correlation coefficient was tested using a two-sided *Student t-test*. The correlation is considered *stable* for those grid-points where SSIE and the large-scale predictors (e.g., OHC,

15 OT100, SST, SLP, TT, PWC, DW, USURF and VSURF) are significantly correlated at 95%, 90%, 85% and 80% significance level for more than 80% of the 21-year windows, covering the period 1979-2007. We choose the period 1979-2007 as calibration period, as both extreme years of sea ice extent, namely 1996 and 2007, were included and it provides a climate relevant period of nearly 30 years. The area where the correlation coefficient is stable and positive are represented as dark red (95%), red (90%), orange (85%) and yellow (80%), while the

20 regions where correlation coefficient is stable and negative are represented as dark blue (95%), blue (90%), green (85%) and light green (80%). Such maps are referred in our study as *stability maps* and their spatial structures remain qualitatively the same if the significance levels that define the stability of the correlation vary within reasonable limits and if the length of the moving window varies between 15 to 25 years. The optimal predictors are defined as the average values over the stable regions for each gridded parameter. For the current analysis only

25 regions where the correlation is above 90% significance level, are retained for further analysis (Figure 1). The raw stability maps between SSIE (pan-Arctic and regional) and the potential predictors are shown in Figures S3 – S15. Although the length of our time series is relatively short (40 years) the methodology proved to work also in cases of timer series <40 years (Ionita et al., 2018). Moreover, we use the same methodology, with the same number of years (40 years), for the prediction of September Arctic and Antarctic Sea Ice

30 (https://www.arcus.org/sipn/sea-ice-outlook/2017/post-season).

As a further main contributor to our forecast model, we use persistence, defined here as the sea ice extent from previous months (e.g., January, February up to August). Persistence of sea ice anomalies stands as the first source of predictability for sea ice (Guemas et al., 2016; Walsh et al., 1979; Blanchard-Wrigglesworth et al., 2011).

2.3 Multiple Linear Regression

35 For the forecast all datasets were separated into two parts: 1) the calibration period (1979–2007) and 2) the validation period (2008 – 2017). The optimal predictors are identified by employing stepwise multiple regression analysis (e.g., von Storch and Zwiers, 1999). Although the "*stability maps*" methodology (Figure 1) identifies multiple stable regions for each atmospheric/oceanic parameter (Figures S3 – S15), after applying the stepwise

multiple regression, the optimal/final prediction model is based just on the regions shown in Figures 1 - 3 and 5 - 7. To forecast the September Sea Ice Extent we have used a multiple linear regression model with the regression equation:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

where Y represents the SSIE, β_0, β_1, β_2,...β_n are constants determined by the least squares procedure, x_1, x_2,...x_n the predictors used (e.g., OHC, OT100, etc) and ε the error.

In this study, we choose stepwise regression. Thus, each predictor was prioritized based on its correlation coefficient with the SSIE and was added to the model in that order. As we added more predictors to the model, the F statistic was used to determine whether the added predictors were significant in the regression equation. Entrance and exit criteria for the F statistic were set to 0.05 and 0.1, respectively. Stepwise regression was used because it prioritizes predictors based on the partial correlation and it is likely that high and significant correlations will reflect underlying physical processes. In order to estimate possible over fitting, we make use of the Akaike information criterion (AIC) (von Storch and Zwiers, 1999), the explained variance, R^2 and the residual standard error. A workflow of the selection of the optimal model for the SSIE prediction is shown in the supplementary file and Figure S2.

3 Results

3.1 Pan-Arctic September sea ice prediction

The skill of a long-range forecast for the Arctic SSIE is associated with the predictors that represent the slow varying components of the climate system that are able to integrate the climate information such as ocean heat content and SST (Guemas et al., 2016, Lindsay et al., 2008). These variables can be used as potential predictors for months and even seasons in advance due to their long-term memory. Thus, here we investigate the potential link between the Arctic SSIE (Fetterer et al., 2016) and OHC, OT100 (Levitus et al., 2012; Boyer et al., 2013) and SST (Huang et al., 2014) as long-term predictors (lags ~4 years (AMO index) up to 2 months in advance, see Table 2 for a detailed description of all the lags used in the study). On shorter time-scales (2 − 4 months) the atmospheric circulation, especially during the summer months, plays a major role in driving the Arctic sea ice variability (Guemas et al., 2016). The atmospheric circulation can substantially contribute to the skill of the sea ice predictions. As such, for the SSIE prediction we have also tested the skill of atmospheric variables (up to 4 months in advance), e.g., SLP, TT, PWC, USURF and VSURF (Kalnay et al., 1996). 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.

For the final forecast, based on data available at the end of May (4 months ahead forecast) we have retained all identified stable regions shown as black boxes in Figure 1. For the forecast based on June data, we have included also the stable regions based on all June stability maps (Figure 2). We have applied the same technique for the July data (Figure 3). For SSIE prediction based on the end of May data, the optimal model is based on a combination of: OHC SON, SST MAM, PWC Apr, VSURF MAM and SLP May (Table 3). Together with these identified stable regions, the optimal model includes also the persistence of sea ice extent (here the sea ice extent from previous March (SIE Mar), as well as the annual Atlantic Multidecadal Oscillation index, with a lag 4 of

years (AMO L4). The highest correlation between SSIE and the annual AMO index was found at a time lag of 4 years (AMO leads SSIE). The time lag identified in our analysis is in line with previous studies (Day et al., 2012; Mahajan et al., 2011). The observed and forecasted values based on the May data are shown in Figure 4a. The explained variance of the model, over the calibration (*validation*) period, is 81% (*71%*) and the correlation coefficient between the observed and forecasted SSIE is r = 0.90 (*r = 84*) (99.9% significance level). To better assess the skill of the SSIE prediction, the root mean square error (RMSE), the Nush-Sutcliffe efficiency (NSE) and the index of agreement (d) are calculated, among other statistical tests (see Table S1 and supplementary file for a definition of all the metrics used to test the skill of the model). The forecasted model based on May data shows a very good skill (Table S1) NSE = 0.82 (*0.68*) (NSE = 1 means perfect model) and d = 0.95 (*0.88*) (d = 1 indicates a perfect match between the observed and forecasted values, d = 0 indicates no agreement at all).

Following the same steps as in the case of May data, for the model based on June data, the parameters contributing to the optimal forecast model are shown in Figure 2. As additional predictors, on top of those for May (Figure 1), we have: VSURF Jun, USURF Jun and TT Jun (Table 3). The observed and forecasted values of SSIE based on June data are shown in Figure 4b. The overall explained variance of the June-based model, over the calibration (*validation*) period, is 85% (*79%*) and the correlation coefficient between the observed and forecasted SSIE values is r = 0.92 (*r = 0.89*) (99.9% significance level). The June-based model exhibits also a very good skill and shows slight improvements compared to the May – based model (NSE = 0.85 (*0.78*) and d = 0.96 (*0.93*). For the model based on July data, the parameters contributing to the optimal forecast model, on top of those based on May (Figure 1) and June (Figure 2), are shown in Figure 3 and Table 3. The observed and predicted values of SSIE based on July data are shown in Figure 4c. The overall explained variance of the June-based model, over the calibration (*validation*) period, is 86% (*81%*) and the correlation coefficient between the observed and forecasted SSIE values is r = 0.93 (*r = 0.90*) (99.9% significance level). The July-based model exhibits also a very good skill and shows also slight improvements compared to the May and June – based models (NSE = 0.86 (*0.80*) and d = 0.96 (*0.94*)).

3.2 Application of the methodology for regional SSIE prediction

To test the robustness of our statistical model and to move towards stakeholder-relevant regions, in this study we are investigating also the skill of our model at regional scale. Thus, we have repeated the same analysis as in the previous section but for the sea ice extent averaged over the East Siberian Sea (ESS) (Figure S1). In this study, we focus on the ESS because in September 2007 and 2012, negative ice concentration anomalies were particularly pronounced over this region of the Arctic Ocean (Figure S1a and S1b, respectively) and the highest variability of the SSIE is recorded here (Figure S1c).In addition, since 2011 the eastern ESS has been nearly ice-free (<10% SSIE) at the end of summer (Polyakov et al., 2017). Moreover, when looking at the correlation coefficients between the pan-Arctic SSIE and regional September SIE, the highest correlation, at lag 0, is found with the ESS-SIE (r = 0.72, Table 4).

The stability maps between the detrended ESS SSIE and the large scale oceanic and atmospheric fields are shown in Figure 5 (stability maps based on May and previous months data), Figure 6 (stability maps based on June and previous months data) and Figure 7 (stability maps based on July and previous months data), respectively. For ESS SSIE prediction based on the end of May data, the optimal model is based on a combination of: annual

OT100, SST MAM, SLP Jan, VSURF MAM, PWC May, TT May and DW MAM (Table 5). The observed and forecasted values based on the May data are shown in Figure 8a. The explained variance of the model, over the calibration (*validation*) period, is 88% (*58%*) and the correlation coefficient between the observed and forecasted ESS SSIE is r = 0.94 (*r = 0.77*) (99.9% significance level). The forecasted model based on the May shows a very good skill (Table S2) NSE = 0.88 (*0.57*) (NSE = 1 means perfect model) and d = 0.97 (*0.86*) (d = 1 indicates a perfect match between the observed and forecasted values, d = 0 indicates no agreement at all).

For the model based on June data, the parameters contributing to the optimal forecast model in addition to the May variables are shown in Figure 6 and Table 5. As additional predictors, on top of May data (Figure 5), we have: SIE Jun, and TT Jun (Table 5). The observed and forecasted values of ESS-SSIE based on June data are shown in Figure 8b. The overall explained variance of the June-based model, over the calibration (*validation*) period, is 91% (*71%*) and the correlation coefficient between the observed and forecasted SSIE values is r = 0.95 (*r = 0.84*) (99.9% significance level). The June-based model exhibits also a very good skill and shows slight improvements compared to the May – based model (NSE = 0.91 (*0.69*) and d = 0.98 (*0.91*). For the model based on July data, the parameters contributing to the optimal forecast model, on top of May data (Figure 5) and June data (Figure 6), are shown in Figure 7 and Table 5. The observed and predicted values of SSIE based on July data are shown in Figure 8c. The overall explained variance of the July-based model, over the calibration (*validation*) period, is 94% (*81%*) and the correlation coefficient between the observed and forecasted SSIE values is r = 0.97 (*r = 0.90*) (99.9% significance level). The July-based model exhibits also a very good skill and shows slight improvements compared to the May and June – based models (NSE = 0.94 (*0.78*) and d = 0.98 (*0.93*)).

4 Discussion

The results of this study demonstrate that statistically based models are able to predict SSIE with high skill, if the accurate drivers and their regional localizations (herein stable regions) are identified via various statistical techniques. In this paper, our analysis was focused on a single month - September, but the same methodology has been be successfully applied also for other months/seasons and also for the Antarctic region (Ionita et al., 2018).

Our results highlight the potential for skillful prediction of SSIE, both at pan-Arctic level as well as for ESS, based on large-scale drivers from stable regions. The ocean drivers (OHC, TT100 and SST) from the identified stable regions are strongly related with the Atlantic inflow or with the SST variability over regions strongly influenced by decadal modes of variability (e.g., Pacific Decadal Oscillation (PDO) in the central and north Pacific) to multidecadal modes of variability (e.g., Atlantic Multidecadal Oscillation (AMO) in the Atlantic Ocean region). The Atlantic inflow, AMO and PDO play a significant role in driving the Arctic sea ice variability (Polyakov et al., 2017; Miles et al., 2014; Ionita et al., 2016; Screen et al., 2016). For example, the North Atlantic might act as a source for the OHC anomaly identified over the Kara Sea, Laptev Sea and ESS (Figure 1 and Figure 5), thus contributing to the skill of our forecast. The OHC anomalies form the North Atlantic flow into the Arctic basin, via advection, affect the sea ice distribution (Polyakov et al., 2017, Ono et al., 2018). In a recent study, Yu et al., (2017) have shown that the leading mode of variability of global sea-ice concentration is positively correlated with the AMO and negatively correlated with the PDO. Furthermore, two thirds of the total global sea ice trend can be explained by a combination of these two modes of variability. Over-imposed on the interannual variability, the temperature and salinity of the Atlantic inflows to the Arctic Ocean shows also

pronounced decadal to multidecadal variability (Zhang, 2015). This aligns with the concept of different previous studies, which suggest that the decreasing trend in the Artic sea ice is partially driven by AMO (Park and Latif, 2008; Lindsay et al., 2005; Ding et al., 2014; Yu et al., 2017). Moreover, starting at the beginning of 1990's the AMO has switched to a positive phase, at the same time when the Arctic sea ice extent started its abrupt decline.

5   Thus, in this study we have tested previous years AMO index as a potential driver of the Arctic sea ice extent.

The stability maps based on the predictors related to the atmospheric variables (Figures 1-3) show significant and stable correlations with regions restricted to the Artic basin, indicating a much regional connection between the September sea ice variability and large-scale atmospheric circulation. The state of the Arctic SSIE depends both on the state of the ice in spring as well as on the atmospheric condition during summer (Ding et al., 2017). In this

10   respect, the precipitable water content and air temperature in spring and early summer were found to show significant predictive skill for the SSIE both at pan-Arctic as well as regional level. This is also in agreement with previous studies (Kapsch et al., 2013; 2014) which have shown a significantly increased cloudiness and humidity over the Arctic region in spring, thus accelerating the sea ice retreat in the upcoming summer.

Overall, such a methodology can be valuable also for the modelling community. If the coupled models, used for

15   forecasting purposes, face problems to simulate the ocean and/or the climate background over the areas that play a significant role in driving the SSIE variability (stable regions), one expects a relatively small forecast skill. The opposite case is also valid: a good representation of the key regions that drive SSIE could imply a good forecast skill. For example, Parkinson et al. (2006) determined that many climate models tend to simulate more winter sea ice in the Barents Sea compared to observations. One hypothesis for this overestimation is that the models

20   underestimate the heat content in the Atlantic Basin (which has proved to be one of the main contributors for a skillful prediction for SSIE in our model). By using a simple and computationally inexpensive statistical approach, one can anticipate more than 80% of SSIE up to four months in advance, based on the antecedent atmospheric and oceanic conditions over stable regions. Moreover, our statistical model is able to properly reproduce the years with extreme low / high sea ice extent, both at pan-Arctic level as well as at regional scale

25   (e.g., 2007 and 2012 – low SSIE and 1996 – high SSIE; see Figure 4 and Figure 8). The predictability of these extreme years poses big challenges for the sea ice prediction community (Hamilton and Stroeve, 2016).

For example, one of the most unpredictable years was 2012. Most of the models (statistical and dynamical) were unable to properly forecast the extremely low value of the sea ice extent in September 2012 (Stroeve et al., 2014). Overall, the statistical predictions came closer to the unexpected low sea ice extent in September 2012 than the

30   dynamical-based predictions. In this respect, our statistical model was able to capture the overall decline in the SSIE and we forecasted the lowest sea ice extent since the observational period (Figure 4). Nevertheless, in terms of amplitude, our forecast has underestimated the observed values (Figure 4). One of the reasons for this underestimation could come from the fact that in August 2012 a strong storm prevailed over the Arctic basin, which triggered extreme sea ice melt by bringing heat and moisture from the south towards the central Arctic

35   (Parkinson and Comiso, 2013). Because the atmosphere is mostly unpredictable beyond 1 or 2 weeks, we were not able to accurately predict, in terms of amplitude, the sea ice conditions that developed because of the Arctic storm in August 2012.

Another challenge for the sea ice community was the predictability of SSIE in 2013. Sea ice extent in September 2013 was characterized by a revival compared to the low values recorded in September 2012 (SSIE in 2013 was

1.69 million square kilometers above the record minimum extent in September 2012). Most of the models, involved in the SIPN, have underestimated the September 2013 sea ice extent, despite the fact that this was not an extreme low sea ice year like 2012. The observed September 2013 sea ice extent lied outside the intervals given with 13 out of 16 predictions, but the modelling methods performed better than the statistical ones (Stroeve et al., 2014). For September 2013, our statistical model performed almost perfectly, giving one of the best predictions (in terms of amplitude) over the validation period. The revival of the sea ice extent in 2013 was mostly due to a colder summer over the Arctic basin, compared to 2012 and no storms prevailing throughout the summer months. Summer 2013 was characterized by an unusual low pressure system over much of the Arctic Ocean, which acted as a limiting factor for the heat transport from the south. Both the SLP and air temperature over the Arctic basin were part of our final predictors for the sea ice extent in 2013 (Figure 2 and Figure 3). As such, the accurate predictions based on our statistical model for 2013 may arise from the fact that no extreme weather events were occurring throughout the summer months over the Arctic region. In addition, we had persistent negative temperature anomalies and a long lasting low pressure system prevailing in June and July over the Arctic basin, variables which were used in our forecast model. A high/low skill in the predictability of extreme September sea ice can be the results of extreme spring preconditioning (e.g., very low ice thickness) and/or the results of extremely anomalous summer weather systems, independent of the spring preconditioning. In observation not all extremes are the results of the same forcing, thus implying that different extremes events will have a different level of predictability.

5 Conclusions

In this study we have developed a statistical method based on different oceanic and atmospheric variables to estimate the monthly signal and variability of the Arctic sea ice extent. Based on stepwise multi-regression analysis optimal predictors are identified in terms of stability maps to forecast SSIE on pan-Arctic or regional scale. We have demonstrated that our well-established statistical approach can be used as a promising tool to improve the skill of sea ice extent prediction. In the future, the same methodology will be applied to test the potential predictability, up to two years ahead, by taking into account variables with long-term memory (e.g., heat content and water temperature integrated over different depths) for the whole Arctic. For other regions prone to extreme decrease in the sea ice extent (e.g., Chukchi Sea, Beaufort Sea, Barents Sea) as well as for Antarctica the method will also be adopted. Finally, since the concept can be used as an early warning system for September sea ice extent, both at pan-Arctic level as well as regionally, the potential environmental and economic benefits can be very high.

5

10

15

20

25

30

35

40

References

Allison, I., Bindoff, N.L., Bindschadler, R.A., Cox, P.M., de Noblet, N., England, M.H., Francis, J.E., Gruber, N., Haywood, A.M., Karoly, D.J., Kaser, G., Le Quéré, C., Lenton, T.M., Mann, M.E., McNeil, B.I., Pitman, A.J., Rahmstorf, S., Rignot, E., Schellnhuber, H.J., Schneider, S.H., Sherwood, S.C., Somerville, R.C.J., Steffen, K., Steig, E.J., Visbeck, M., and Weaver, A.J. The Copenhagen Diagnosis: Updating the World on the Latest Climate Science. Sydney, Australia: University of New South Wales, Climate Change Research Centre. Available at http://www.ccrc.unsw.edu.au/Copenhag en/Copenhagen_Diagnosis_LOW.pdf, 2009.

Blanchard-Wrigglesworth, E., Armour, K.C., Bitz, C.M., and DeWeaver, E. Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. J. Clim. 24: 231–250, 2011.

Boyer, T.P., Antonov, J. I., Baranova, O.K., Coleman, C., Garcia, H.E., Grodsky, A., Johnson, D.R., Locarnini, R.A., Mishonov, A.V., O'Brien, T.D., Paver, C.R., Reagan, J.R., Seidov, D., Smolyar, I.V., and Zweng, M.M. World Ocean Database 2013, NOAA Atlas NESDIS 72, S. Levitus, Ed., A. Mishonov, Technical Ed.; Silver Spring, MD, 209 pp., 2013.

Cavalieri, D. J. and Parkinson, C. L.: Arctic sea ice variability and trends, 1979–2010, The Cryosphere, 6, 881-889, https://doi.org/10.5194/tc-6-881-2012, 2012.

Chevallier, M., D. Salas y Mélia, A. Voldoire, M. Déqué, and G. Garric. Seasonal Forecasts of the Pan-Arctic Sea Ice Extent Using a GCM-Based Seasonal Prediction System. *J. Climate,* **26**, 6092–6104, https://doi.org/10.1175/JCLI-D-12-00612.1, 2013

Comiso, J. C., W. N. Meier, and R. Gersten. Variability and trends in the Arctic Sea ice cover: Results from different techniques, *J. Geophys. Res. Oceans*, 122, 6883–6900, doi:10.1002/2017JC012768, 2017.

Day, J.J., Hargreaves, J.C., Annan, J.D., and Abe-Ouchi, A. Sources of multi-decadal variability in Arctic sea ice extent, *Environ. Res. Lett.*, 7, 034011, doi:10.1088/1748-9326/7/3/034011, 2012.

Day, J., Hawkins, E., and Tietsche, S.: Will Arctic sea ice thickness initialization improve seasonal forecast skill?, Geophys. Res. Lett., 41, 7566–7575, 2014.

Ding, Q., Schweiger, A., L'Heureux, M., Battisti, D.S., Po-Chedley, S., Johnson, N.C., Blanchard-Wrigglesworth, E., Harnos, K., Zhang, Q., Eastman, R., and Steig, E.J. Influence of high-latitude atmospheric circulation changes on summertime Arctic sea ice, Nature Climate Change, doi:10.1038/nclimate3241, 2017.

Ding, Q., Wallace, J.M., Battisti, D.S., Steig, E.J., Gallant, A.J.E., Kim, H.-J., and Geng, L. Tropical forcing of the recent rapid Arctic warming in northeastern Canada and Greenland. Nature 509, 209–212, 2014.

Dirkson, A., Merryfield, W. J., and Monahan, A.: Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions, J. Climate, 30, 1001–1017, 2017.

Drobot, S.D., Maslanik, J.A., and Fowler, C. A long-range forecast of Arctic summer sea ice minimum extent. *Geophys. Res. Lett.* **33**: L10501, doi:10.1029/2006GL026216, 2006.

Eguíluz V. M., Fernández-Gracia, J., Irigoien ,X., and Duarte, C. M. A quantitative assessment of Arctic shipping in 2010–2014. Sci. Rep. 6, 30682, doi:10.1038/srep30682, 2016.

Fetterer, F., Knowles, K., Meier, W. and Savoie, M. Updated daily Sea Ice Index, Version 2. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: http://dx.doi.org/10.7265/N5736NV7, 2016.

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J.J., Déqué, M., Doblas-Reyes, F.J., Fučkar, N.S., Germe, A., Hawkins, E., Keeley, S., Koenigk, T., Salas y Mélia, D., and Tietsche, S. A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. Q.J.R. Meteorol. Soc., 142: 546-561. doi:10.1002/qj.2401, 2016.

Hamilton, L., and Stroeve, J.C. 400 Predictions: the SEARCH Sea Ice Outlook 2008–2015. Polar Geography, 39:4, 274-287, 2016.

Hassol, S,J. Impacts of a Warming Arctic: Arctic Climate Impact Assessment. Cambridge University Press*:* Cambridge, UK. http://www.amap.no/arctic-climate-impact-assessment-acia, 2004.

Holland, M. M. and Stroeve, J.: Changing seasonal sea ice predictor relationships in a changing Arctic climate, Geophys. Res. Lett., 38, L18501, https://doi.org/10.1029/2011GL049303, 2011.

Huang, B., Banzon, V.F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T.C., Smith, T.M., Thorne, P.W., Woodruff, S.D., and Zhang, H.-M. Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4): Part I. Upgrades and intercomparisons. Journal of Climate, 28, 911–930, doi:10.1175/JCLI-D-14-00006.1, 2014.

Ionita, M., Dima, M., Lohmann, G., Scholz, P., and Rimbu, N. Predicting the June 2013 European Flooding based on Precipitation, Soil Moisture and Sea Level Pressure. J. Hydrometeorology, 16, 598–614., doi: http://dx.doi.org/10.1175/JHM-D-14-0156.1, 2014.

Ionita, M., Lohmann, G. and Rimbu, N. Prediction of Elbe discharge based on stable teleconnections with winter global temperature and precipitation, Journal of Climate, 21, 6215–6226, doi:10.1175/2008JCLI2248.1, 2008.

Ionita, M., Scholz, P., Lohmann, G., Dima, M. and Prange, M. Linkages between atmospheric blocking, sea ice export through Fram Strait and the Atlantic Meridional Overturning Circulation, Scientific Reports, 6: 32881, 2016.

Ionita, M. Mid range forecasting of the German Waterways streamflow based on hydrologic, atmospheric and oceanic data, Alfred Wegener Institute for Polar and Marine Research. Bremerhaven, 711, 2017.

Ionita, M., Scholz, P., Grosfeld, K., and Treffeisen, R.: Moisture transport and Antarctic sea ice: austral spring 2016 event, Earth Syst. Dynam., 9, 939-954, https://doi.org/10.5194/esd-9-939-2018, 2018.

IPCC: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324, 2013.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. The NCEP/NCAR 40-year reanalysis project Bull. Am. Meteorol. Soc. 77 437–71, 1996.

Kapsch, M.-L., Graversen, R.G., and Tjernström, M. Springtime atmospheric energy transport and the control of Arctic summer sea-ice extent. Nature Clim. Change, 3, 744–748, doi:10.1038/nclimate1884, 2013.

Kapsch, M.-L., Graversen, R.G., Economou, T, and Tjernström, M. The importance of spring atmospheric conditions for predictions of the Arctic summer sea ice extent. *Geophys.* Res. Lett., 41, 5288–5296, doi:10.1002/2014GL060826, 2014.

Kay, J. E., Holland, M. M., and Jahn., A. Inter-annual to multi-decadal Arctic sea ice extent trends in a warming world. *Geophysical Research Letters* 38:L15708, doi:10.1029/2011GL048008, 2011.

Koenigk, T., and Mikolajewicz, U. Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. *Clim. Dyn.* **32**: 783–798, doi: 10.1007/s00382-008-0419-1, 2009.

Lasserre, F. Simulations of shipping along Arctic routes: Comparison, analysis and economic perspectives. *Polar Record, 51*(3), 239-259. doi:10.1017/S0032247413000958, 2015

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., Mishonov, A. V., Reagan, J. R., Seidov, D., Yarosh, E. S., and Zweng, M. M. World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. Geophysical Research Letters 39(10). doi:10.1029/2012GL051106, 2012.

Lindsay, R.W., and Zhang, J. The thinning of Arctic sea ice, 1988–2003: Have we passed a tipping point? J. Clim. 18(22), 4879–4894, 2005.

Lindsay, R.W., Zhang, J., Schweiger, A.J., and Steele, M.A. Seasonal predictions of ice extent in the Arctic Ocean. J. Geophys. Res. 113, C02023, 2008.

Mahajan, S., Zhang, R., and Delworth, T.L. Impact of the Atlantic Meridional Overturning Circulation (AMOC) on Arctic Surface Air Temperature and Sea Ice Variability. J. Climate, 24, 6573–6581, https://doi.org/10.1175/2011JCLI4002.1 , 2011.

Meier, W.N., Hovelsrud, G., van Oort, B., Key, J., Kovacs, K., Michel, C., Granskog, M., Gerland, S., Perovich, D., Makshtas, A. P., and Reist, J. Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity, Rev. Geophys., 51,185–217, doi:10.1002/ 2013RG000431, 2014.

Meißner, D., Klein, B., and Ionita, M. Development of a monthly to seasonal forecast framework tailored to inland waterway transport in Central Europe, Hydrol. Earth Syst. Sci. Discuss., 2017, 1-31, doi:10.5194/hess-2017-293, 2017.

Melia, N., Haines, K., and Hawkins, E. Sea ice decline and 21st century trans-Arctic shipping routes, *Geophys. Res. Lett.*, 5 43, 9720–9728, doi:10.1002/2016GL069315, 2016.

Miles, M.W., Divine, D.V., Furevik, T., Jansen, E., Moros, M., and Ogilvie, A.E.J. A signal of persistent Atlantic multidecadal variability in Arctic sea ice, Geophys. Res. Lett., 41, 463–469, doi:10.1002/2013GL058084, 2014.

10 Notz, D., and Marotzke, J. Observations reveal external driver for Arctic sea-ice retreat. *Geophysical Research Letters* 39:L08502, doi:10.1029/2012GL051094, 2012.

Ono, J., Tatebe, H., Komuro, Y., Nodzu, M. I., and Ishii, M. Mechanisms influencing seasonal to inter-annual prediction skill of sea ice extent in the Arctic Ocean in MIROC, The Cryosphere, 12, 675-683, https://doi.org/10.5194/tc-12-675-2018, 15 2018.

Park, W., and Latif, M. Multidecadal and multicentennial variability of the meridional overturning circulation. Geophys. Res. Lett. 35, L22703, doi: 10.1029/2008GL035779, 2008.

20 Parkinson, C.L., Vinnikov, K.Y., and Cavalieri, D. J. Evaluation of the simulation of the annual cycle of Arctic and Antarctic sea ice coverages by 11 major global climate models, J. Geophys. Res., 111, C07012, doi:10.1029/2005JC003408, 2006.

Parkinson, C. L., and Comiso, J.C. On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm, *Geophys. Res. Lett.*, 40, 1356–1361, doi:10.1002/grl.50349, 2013.
25
Petty, A. A., Schröder, D. , Stroeve, J. C., Markus, T. , Miller, J. , Kurtz, N. T., Feltham, D. L. and Flocco, D. Skillful spring forecasts of September Arctic sea ice extent using passive microwave sea ice observations. Earth's Future, 5: 254-263. doi:10.1002/2016EF000495, 2017.

30 Pizzolato, L., Howell, S.E.L., Derksen, C., Dawson, J., and Copland, L. Changing sea ice conditions and marine transportation activity in Canadian Arctic waters between 1990 and 2012, Clim. Change, 123, 161–173, doi:10.1007/s10584-013-1038-3, 2014.

Pizzolato, L., Howell, S.E.L., Dawson, J., Laliberté, F., and Copland, L. The influence of declining sea ice on shipping 35 activity in the Canadian Arctic, *Geophys. Res. Lett.*, 43, 12,146–12,154, doi:10.1002/2016GL071489, 2016.

Polyakov, I.V., Pnyushkov, A.V., Alkire, M.B., Ashik, I.M., Baumann, T.M. , Carmack, E.C., Goszczko, I., Guthrie, J., Ivanov, V.V., Kanzow, T., Krishfield, R., Kwok, R., Sundfjord, A., Morison, J., Rember, R., and Yulin, A. Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean, Science, 356 (6335), pp. 285-291 . doi: 40 10.1126/science.aai8204, 2017.

Schröder, D., Feltham, D. L., Flocco, D., and Tsamados M. September Arctic sea-ice minimum predicted by spring melt-pond fraction, Nat. Clim. Change., *4*, 353–357, 2014.

45 Schøyen H., and Bråthen, S. The Northern Sea route versus the Suez Canal: cases from bulk shipping. J. Transp. Geogr. 19, 977–983, 2011.

Screen, J.A., and Francis, J.A. Contribution of sea-ice loss to Arctic amplification is regulated by Pacific Ocean decadal variability. Nature Clim. Change 6, 856–860, 2016.
50
Serreze, M.C., Holland, M.M., and Stroeve, J. Perspectives on the Arctic's shrinking sea-ice cover. Science 315(5818):1533–1536, 2007.

Sigmond, M., J. C. Fyfe, G. M. Flato, V. V. Kharin, and W. J. Merryfield. Seasonal forecast skill of Arctic sea ice area in a 55 dynamical forecast system, *Geophys. Res. Lett.*, 40, 529–534, doi:10.1002/grl.50129, 2013.

Smith, L.C., and Stephenson, S.R. New Trans-Arctic shipping routes navigable by midcentury. Proc. Natl. Acad. Sci. U.S.A. 110: 4871–4872, doi: 10.1073/pnas.1214212110, 2013.

60 Stroeve, J. C., Hamilton, L., Bitz, C., and Blanchard-Wigglesworth, E. Predicting September sea ice: Ensemble skill of the SEARCH sea ice outlook 2008–2013, Geophys. Res. Lett., 41(7), 2411–2418, doi:10.1002/2014GL059388, 2014.

Stroeve, J., Blanchard-Wrigglesworth, E., Guemas, V., Howell, S., Massonnet, F., and Tietsche, S. Improving predictions of Arctic sea ice extent, Eos, 96, doi:10.1029/2015EO031431, 2015.

13

Von Storch, H., and Zwiers, F.W.: Statistical Analysis in Climate Research. Cambridge University Press, 1999.

Walsh, J.E., and Johnson, C.M. An analysis of Arctic sea ice fluctuations. J. Phys. Oceanogr. 9: 580–591, 1979.

5   Yu, L., Zhong, S., Winkler, J.A., Zhou, M., Lenschow, D.H., Li, B., Wang, X., and Yang, Q. Possible connections of the opposite trends in Arctic and Antarctic sea-ice cover, Nature Scientific Reports, 7, doi:10.1038/srep45804, 2017.

Zhan, Y., and Davies, R. September Arctic sea ice extent indicated by June reflected solar radiation, J. Geophys. Res. Atmos., 122, 2194–2202, doi: 10.1002/2016JD025819, 2017.

10 Zhang, R. Mechanisms for low-frequency variability of summer Arctic sea ice extent. Proc. Natl Acad. Sci. USA **112**, 4570–4575, 2015.

15

20

25

30

35

40

45
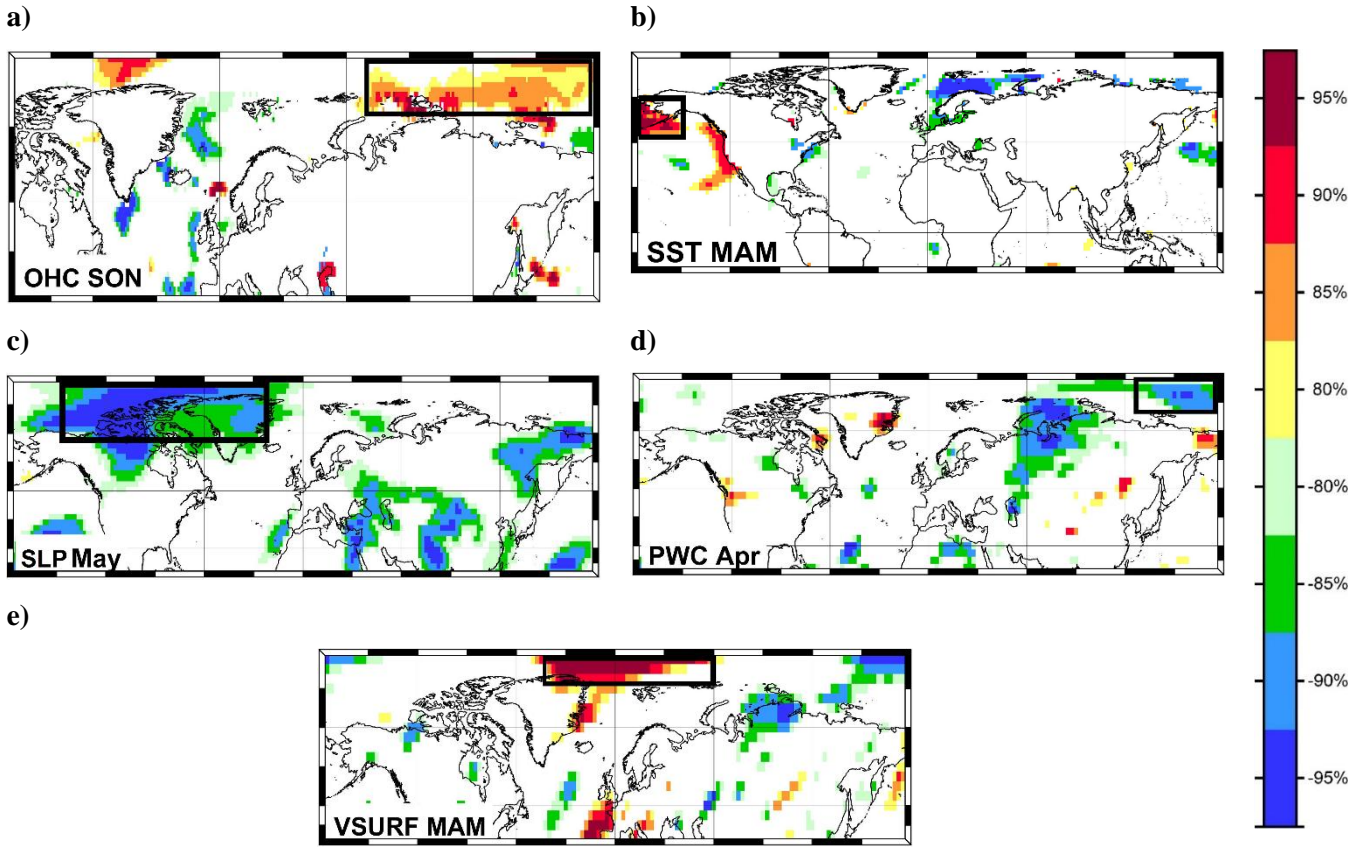
Figure 1. Stability map of the correlation between September Sea Ice Extent and a) OHC SON, b) SST MAM, c) SLP May, d) PWC Apr, and e) VSURF MAM. Regions where the correlation is stable, positive and significant for at least 80% of the 21-year windows are shaded with dark red (95%), red (90%), orange (85%) and yellow (80%). The corresponding regions where the correlation is stable, but negative, are shaded with dark blue (95%), blue (90%), green (85%) and light green (80%). The black boxes indicate the regions used for the September sea ice extent at the end of May.
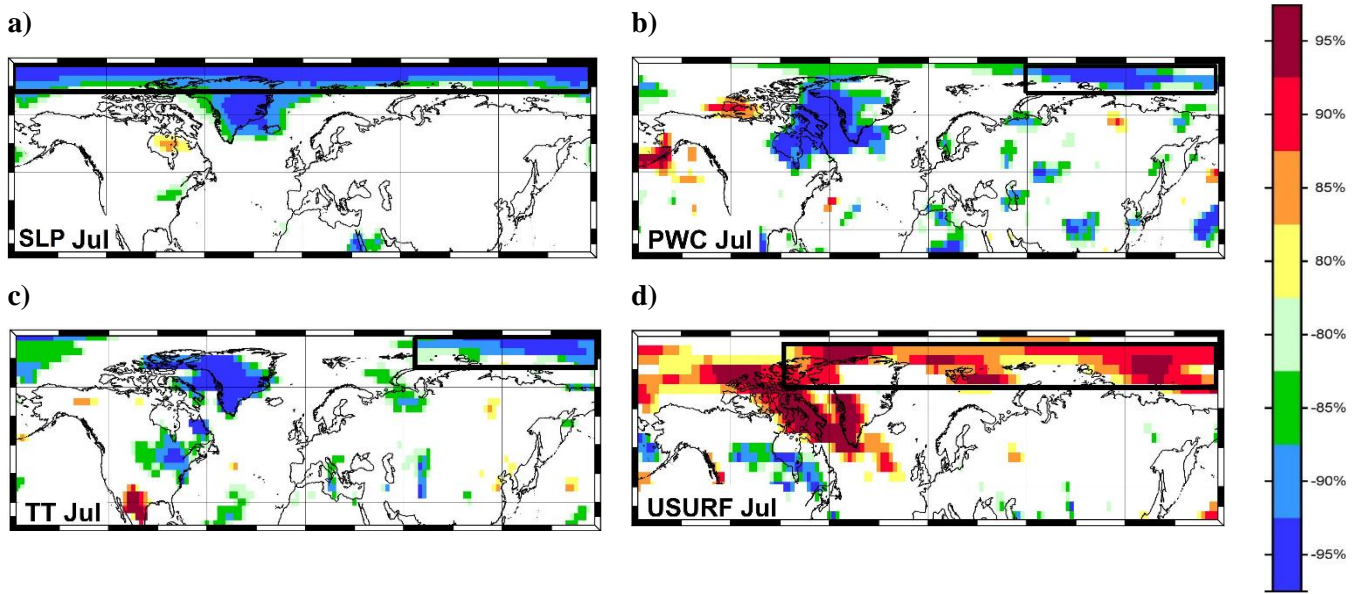
5

10

15

20
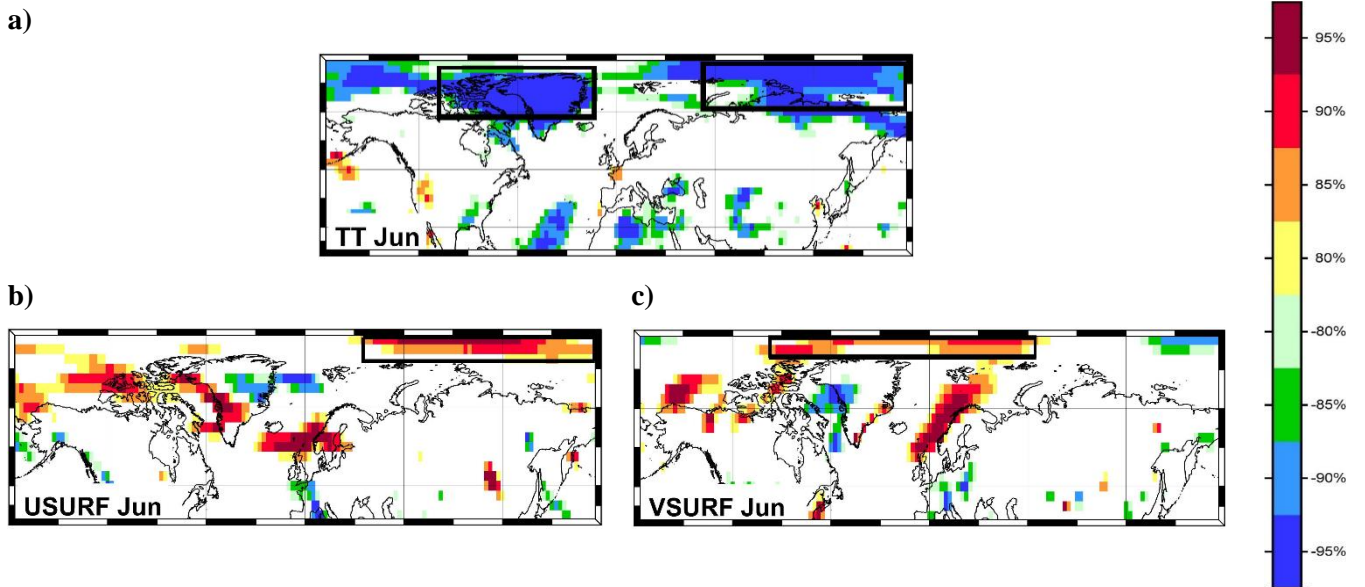
Figure 2. Stability map of the correlation between September Sea Ice Extent and a) TT Jun, b) USURF Jun and c) VSURF Jun. Regions where the correlation is stable, positive and significant for at least 80% of the 21-year windows are shaded with dark red (95%), red (90%), orange (85%) and yellow (80%). The corresponding regions where the correlation is stable, but negative, are shaded with dark blue (95%), blue (90%), green (85%) and light green (80%). The black boxes indicate the regions used for the September sea ice extent at the end of June in addition to the variables of May.
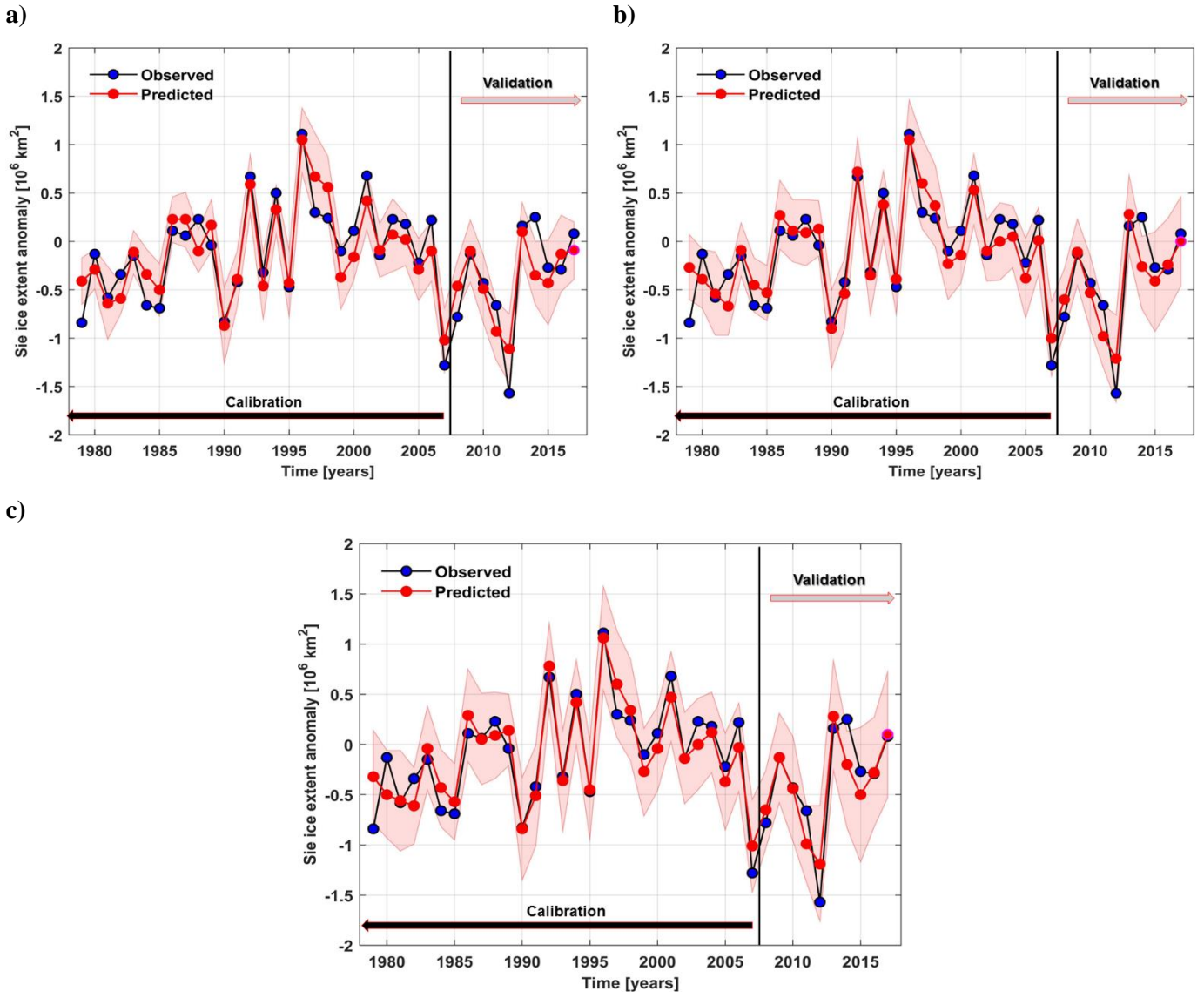
5

10

15

20

25

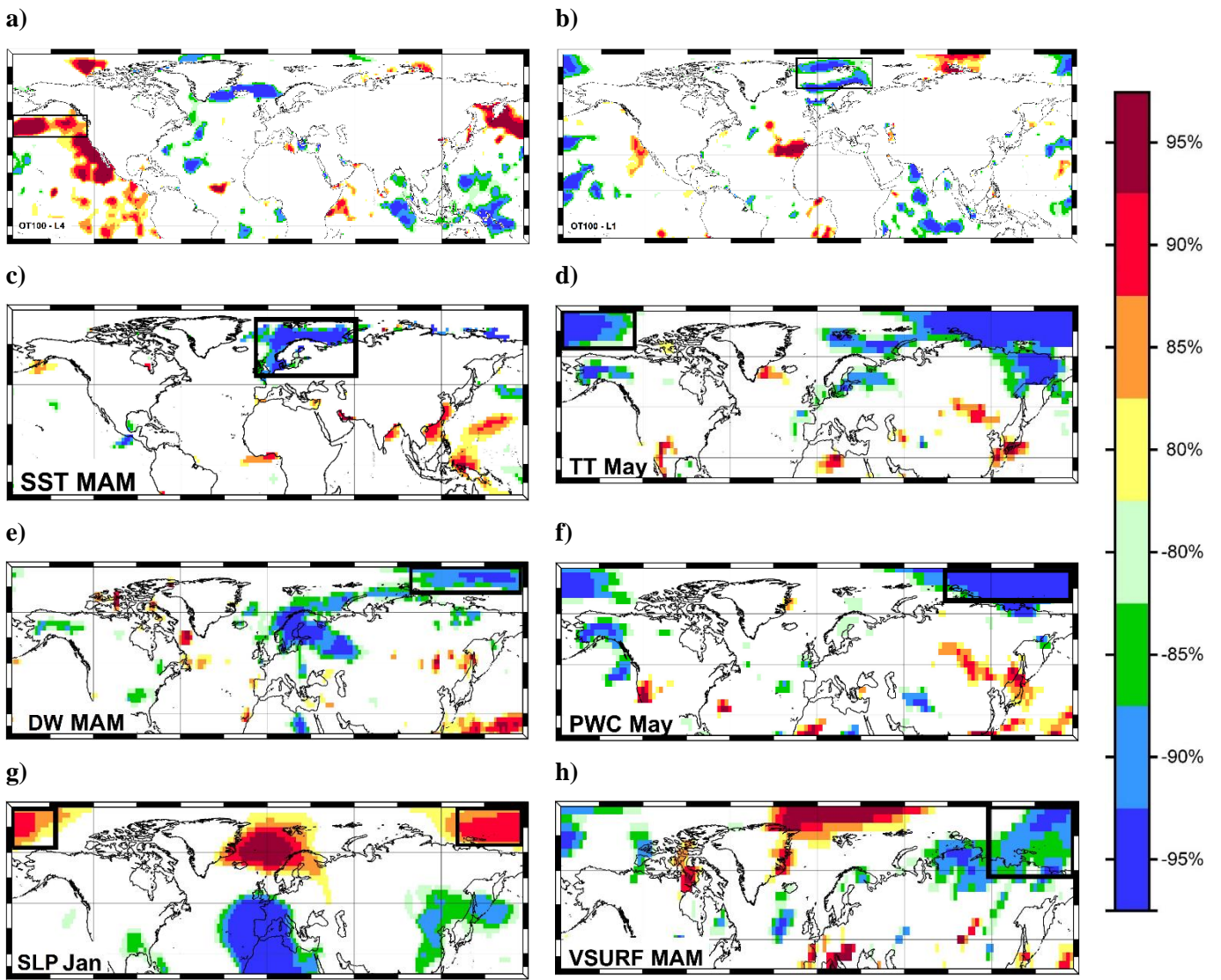**Figure 3.** Stability map of the correlation between September Sea Ice Extent and a) SLP Jul, b) PWC Jul, c) TT Jul and d) USURF Jul. Regions where the correlation is stable, positive and significant for at least 80% of the 21-year windows are shaded with dark red (95%), red (90%), orange (85%) and yellow (80%). The corresponding regions where the correlation is stable, but negative, are shaded with dark blue (95%), blue (90%), green (85%) and light green (80%). The black boxes indicate the regions used for the September sea ice extent at the end of July in addition to the variables of May and June.

Figure 4. Observed (black) and predicted (red) September Sea Ice Extent detrended anomalies over the period 1979-2017 based on a) May, b) June and c) July predictors from the stable regions. The shaded area represents the 95% uncertainty bounds.

5

10

15

Figure 5. Stability map of the correlation between East Siberian September Sea Ice Extent and a) OT100 Annual (L4), b) OT100 Annual (L1), c) SST MAM, d) TT May, e) DW MAM, f) PWC May, g) SLP Jan and h) VSURF MAM. Regions where the correlation is stable, positive and significant for at least 80% of the 21-year windows are shaded with dark red (95%), red (90%), orange (85%) and yellow (80%). The corresponding regions where the correlation is stable, but negative, are shaded with dark blue (95%), blue (90%), green (85%) and light green (80%). The black boxes indicate the regions used for the September sea ice extent at the end of May.

5

19
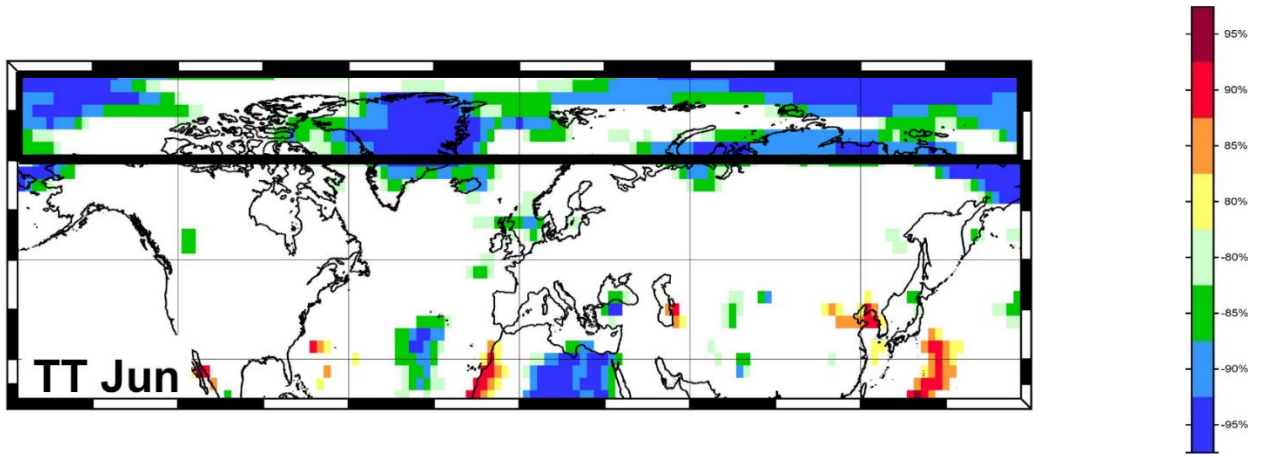
Figure 6. Stability map of the correlation between East Siberian September Sea Ice Extent and TT Jun. Regions where the correlation is stable, positive and significant for at least 80% of the 21-year windows are shaded with dark red (95%), red (90%), orange (85%) and yellow (80%). The corresponding regions where the correlation is stable, but negative, are shaded with dark blue (95%), blue (90%), green (85%) and light green (80%). The black boxes indicate the regions used for the September sea ice extent at the end of June in addition to the variables of May.
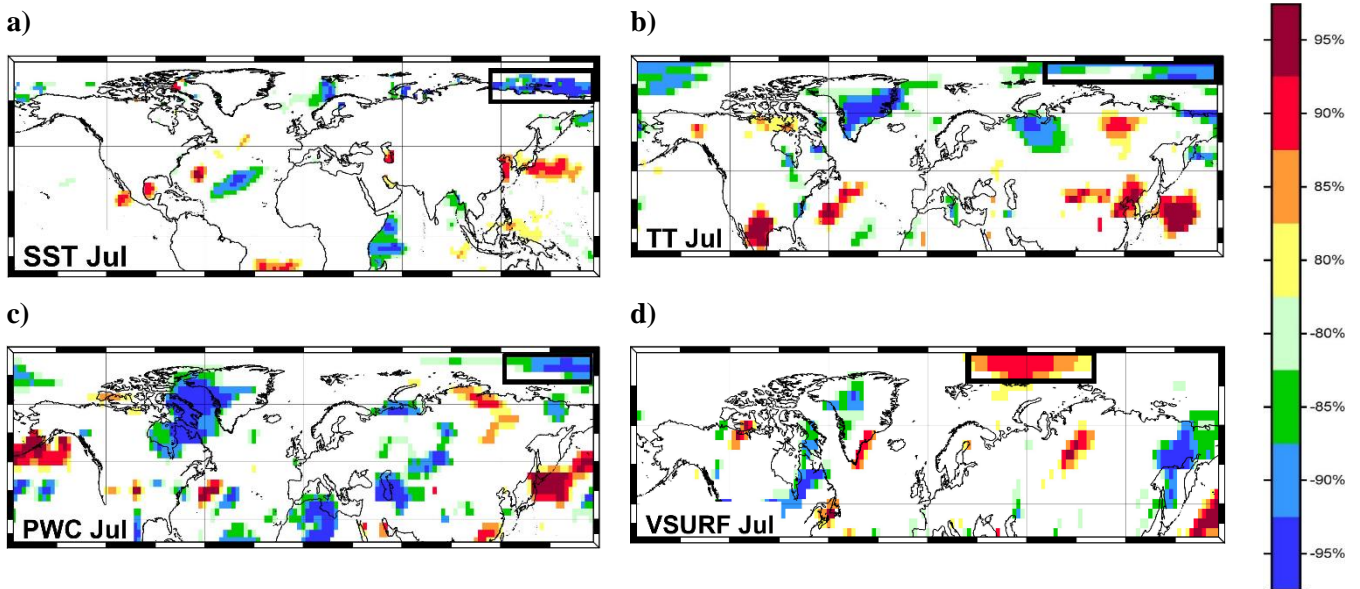
5

10

15

20

25

Figure 7. Stability map of the correlation between East Siberian September Sea Ice Extent and a) SST Jul, b) TT Jul, c) PWC Jul and d) VSURF Jul. Regions where the correlation is stable, positive and significant for at least 80% of the 21-year windows are shaded with dark red (95%), red (90%), orange (85%) and yellow (80%). The corresponding regions where the correlation is stable, but negative, are shaded with dark blue (95%), blue (90%), green (85%) and light green (80%). The black boxes indicate the regions used for the September sea ice extent at the end of July in addition to the variables of May and June.
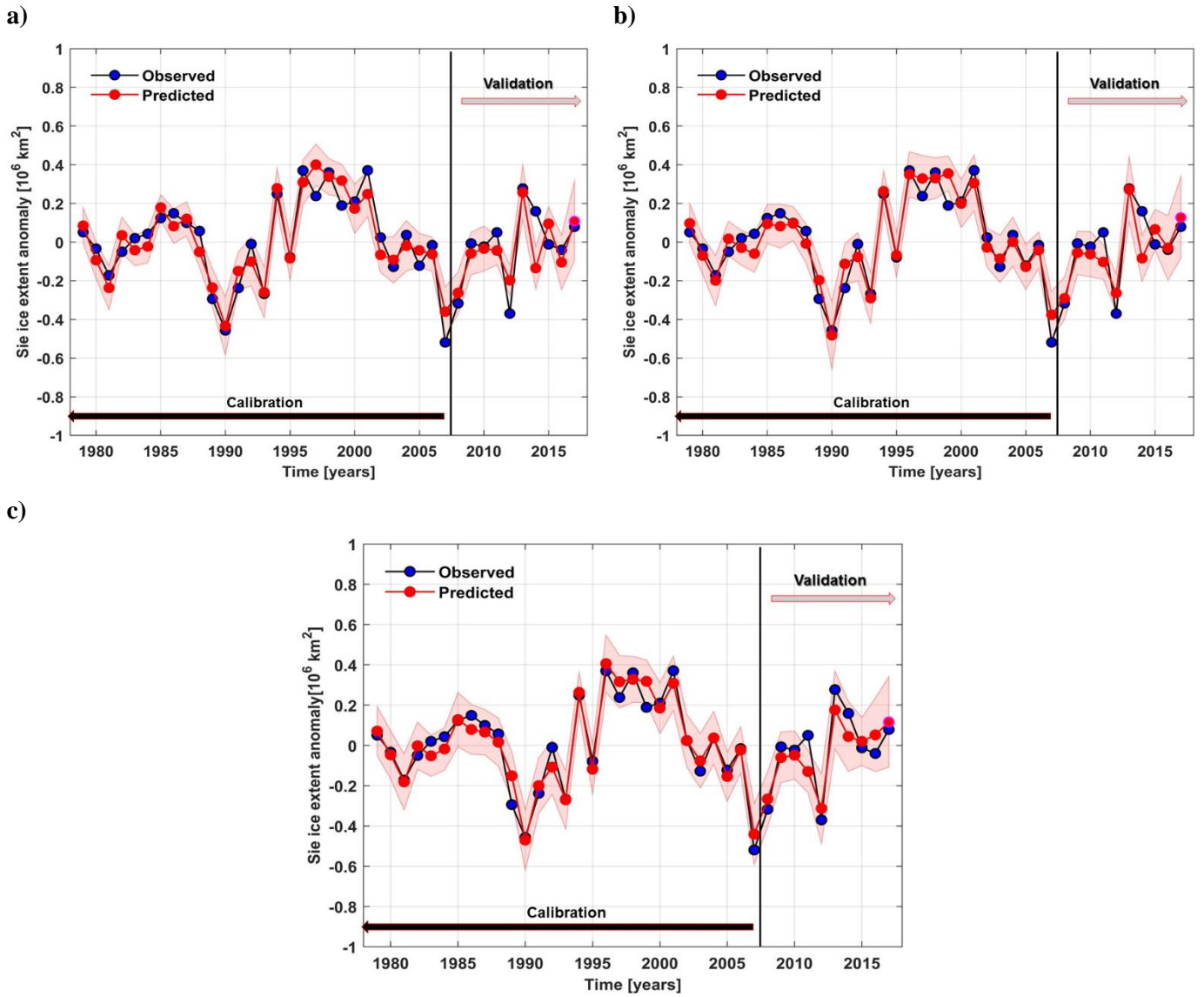
5

10

15

20

25

a)

b)

c)

Figure 8. Observed (black) and predicted (red) East Siberian Sea Ice Extent detrended anomalies over the period 1979-2017 based on a) May, b) June and c) July predictors from the stable regions. The shaded area represents the 95% uncertainty bounds.

5

Table 1. Name, acronym, source, spatial and temporal resolution of the data sets used in this study

Name	Source	Temporal resolution	Spatial resolution	Reference
Arctic sea ice extent	ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/north/monthly/	1979 - 2017		Fetterer et al., 2016
AMO index	https://climexp.knmi.nl/data/iamo_ersst.dat	1979 - 2017		Huang et al., 2014
Mean air temperature at 2m (TT)	ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/surface_gauss/	1979 -2017	2.5° X 2.5°	Kalnay et al., 1996
Downward longwave radiation (DLR)	ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/surface_gauss/	1979 - 2017	2.5° X 2.5°	Kalnay et al., 1996
Zonal surface wind (USURF)	ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/surface/	1979 - 2017	2.5° X 2.5°	Kalnay et al., 1996
Meridional surface wind (VSURF)	ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/surface/	1979 - 2017	2.5° X 2.5°	Kalnay et al., 1996
Precipitable water content (PWC)	ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/surface/	1979 - 2017	2.5° X 2.5°	Kalnay et al., 1996
Sea level pressure (SLP)	ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/surface/	1979 - 2017	2.5° X 2.5°	Kalnay et al., 1996
Sea surface temperature (ERSSTv5)	ftp://ftp.ncdc.noaa.gov/pub/data/cmb/ersst/v5/netcdf/	1979 - 2017	2.0° X 2.0°	Huang et al., 2014
Ocean heat content in the first 700m (OHC)	https://www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT/	1979 - 2017	2.5° X 2.5°	Levitus et al., 2012 Boyer et al., 2013
Ocean temperature in the first 100m (OT100)	https://www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT/	1979 - 2017	2.5° X 2.5°	Levitus et al., 2012 Boyer et al., 2013

5

Table 2. Time lags used for the forecast of SSIE. Seasonal averages are indicated as winter (December/January/February – DJF), spring (March/April/May – MAM), summer (JJA – June/July/August) and autumn (September/October/December – SON).

Variable	Time lag	Month	Season
TT, DLR, USURF, VSURF, PWC, SLP	1 – 7 months, 1 – 2 seasons	January - July	DJF, MAM
ERSSTv5	1 – 7 months, 1 – 2 seasons	January - July	DJF, MAM
OHC, OT100	1 – 4 seasons, 1- 4 years		Annual, DJF, MAM, JJA, SON
AMO index	1- 4 years		Annual mean

5

10

15

20

25

24

Table 3. Variables retained for the September pan-Arctic sea ice extent forecast (black boxes in Figure 1, 2 and 3). Single months are abbreviated with the first three letters of the month.

	May Data	**June Data**	**July Data**
Persistence	SIE Mar	SIE Mar	SIE Mar
Ocean variables	OHC SON	OHC SON	OHC SON
	SST MAM	SST MAM	SST MAM
	AMO – L4	AMO – L4	AMO – L4
Atmospheric variables	SLP May	SLP May	SLP May
			SLP Jul
	VSURF MAM	VSURF MAM	VSURF MAM
		VSURF Jun	VSURF Jun
		USURF Jun	USURF Jun
			USURF Jul
	PWC Apr	PWC Apr	PWC April
			PWC Jul
		TT Jun	TT Jun
			TT Jul

5

10

15

Table 4. The correlation coefficients between the detrended pan-Arctic September sea ice extent and the regional September sea ice extent. A detailed description about the definition of each region is given here: ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/seaice_analysis/

	Lag 4	Lag 3	Lag 2	Lag 1	Lag 0
Baffin	0.07	0.09	0.34	0.40	0.39
Barents	0.20	0.16	0.27	0.13	0.14
Beaufort	0.15	0.24	0.37	0.51	**0.60**
Bering	-0.30	-0.02	0.14	0.00	-0.04
Canadian	0.07	-0.16	0.01	0.52	0.49
Chukchi	-0.26	0.03	0.09	0.53	**0.60**
East Siberian	0.19	0.24	0.39	0.61	**0.72**
Greenland	0.04	0.06	0.22	0.16	-0.07
Hudson	0.44	0.51	0.46	0.38	0.47
Kara	0.09	-0.03	0.05	-0.08	-0.07
Laptev	0.34	0.32	0.40	0.37	0.53

5

10

15

20

Table 5. Variables retained for the September East Siberian sea (ESS) ice extent forecast (black boxes in Figure 5, 6 and 7). Seasonal averages are indicated as spring MAM (March, April, May); single months are abbreviated with the first three letters of the month.

	May Data	**June Data**	**July Data**
Persistence		SIE Jun	SIE Jun
			SIE Jul
Ocean variables	OT100 – L4, L1	OT100 – L4, L1	OT100 – L4, L1
	SST MAM	SST MAM	SST MAM
Atmospheric variables	SLP Jan	SLP Jan	SLP Jan
	VSURF MAM	VSURF MAM	VSURF MAM
			VSURF Jul
	PWC May	PWC May	PWC May
			PWC Jul
	TT May	TT May	TT May
		TT Jun	TT Jun
			TT Jul
	DW MAM	DW MAM	DW MAM

5

10

15