Manuscript tc-2017-122

Mechanisms influencing seasonal-to-interannual prediction skill of sea ice extent in the Arctic Ocean in MIROC

Jun Ono, Hiroaki Tatebe, Yoshiki Komuro, Masato I. Nodzu, and Masayoshi Ishii

October 9, 2017

Response to Anonymous Referee #2

We deeply appreciate the reviewer's kind remarks about our paper. Detailed comments from reviewer are numbered consecutively and cited in italics, followed by our reply in bold face.

Summary

The authors present results on Arctic sea-ice extent prediction skill obtained with a MIROC-based forecast system. Further, they explore possible reasons for differences in skill in different times of the year based on lagged correlation and regression pat- terns, focussing on preceeding states of the (subsurface) ocean heat content and of the sea-ice itself.

In general, The paper is generally well-written and provides interesting results that merit publication. However, there are some points that in my view need further scrutiny. For example, the conclusion that the advection of subsurface water masses from the Altantic Ocean into the Barents Sea, though plausible, is in my view not sufficiently supported by the results shown. Also, the definition of the subsurface ocean heat content and how it's interpreted deserves additional attention, and the rationale behind performing the lagged correlation/regression analysis primarily based on the hindcasts rather than on the control run, and what might cause differences between them, needs clarification. In addition, there is quite a number of minor issues, listed below.

Therefore I recommend the manuscript should be reconsidered after major revisions.

Thank you very much for your summary comments and suggestions. We respond to specific comments as below.

Specific comments

1. P1L8: The term "seasonal-to-interannual" should be shifted in front of "predictions".

As suggested, we corrected it (P1L8).

2. P1L10: "of up three years" - here is a word ("to") missing.

Thank you. We replaced "of up" with "up to" (P1L10).

3. P1L12: "December SIE_AO can be predicted up to 1 year ahead" - I suggest that this statement should be made more quantitative, e.g., by providing the ACC, and maybe also substantiated with the corresponding p-value.

As suggested, we added "(anomaly correlation coefficient is 0.42)" to the text (P1L13).

4. P1L13-15: The role of advection as indicated here is in my view insufficiently supported by the results shown; see details below.

Please see our response to the referee's comment 20.

5. P1L23: "problem" - just as a side remark, I think this judgmental term adds an unnecessary political dimension to this observation.

According to your advice, we removed "An even more serious problem is the decline in Arctic sea ice thickness (Kwok et al., 2009), which has decreased by around 65% from 1975 to 2012 (Lindsay and Schweiger, 2015)", and added "Moreover, Arctic sea ice thickness has decreased by around 65 % from 1975 to 2012 (Kwok et al., 2009; Lindsay and Schweiger, 2015)" to the text (P1L23-24).

6. P2L1: "or" - I think I know what is meant, but using "or" here seems illogical.

We replaced "two- or five-year" with "two and five years" (P2L1).

7. P2L2: "the potential predictability for sea ice extent is continuously one to two years" - I think this statement again needs some numbers; theoretically, marginal (but pratically meaningless) potential predictability should be out there for very long lead times, whereas *pratically meaningful potential predictability survives much shorter lead times. At least, something like an ACC threshold which is considered to distinguish "meaningful" from "no" skill should be provided. (Note that "statistical significance" is not necessarily the correct concept needed here.)*

According to [1]Blanchard-Wrigglesworth et al. (2011), predictability is considered to be significant when the root mean square deviation of the ensemble of prediction experiments is less than that of the reference based on an F-test (for example, please see Figure 1 of Blanchard-Wrigglesworth et al. (2011)). However, the specific value that is considered to distinguish "meaningful" from "no" skill is not found in the paper. It might be overlooked, but we did not add any number to the text.

1. Blanchard-Wrigglesworth, E., Bitz, C. M., and Holland, M. M.: Influence of initial conditions and climate forcing on predicting Arctic sea ice, Geophys. Res. Lett., 38, L18503, doi:10.1029/2011GL048807, 2011.

8. P2L5-6: "The observed Arctic sea ice extent based on ensemble hindcasts can be predicted up to 2–7 and 5–11 months ahead for summer and winter" - see my previous remark.

As you pointed out, the specific value like an ACC threshold should be provided in the text. Predictability up to 2-7 and 5-11 months are based on the several results by previous studies (e.g., [2]Chevallier et al., 2013; [3]Sigmond et al., 2013; [4]Wang et al., 2013; [5]Msadek et al., 2014; [6]Peterson et al., 2015; [7]Guemas et al., 2016; [8]Sigmond et al., 2016). For example, Chevallier et al. (2013) have provided values for correlations and bootstrap test in Table 1. Also, in the study of Sigmond et al. (2016), forecast skill is considered to be significant when anomaly correlation coefficient exceeds to 0.296. However, such a value is not necessarily described in the previous all papers, although the assessment methods for forecast skill are described. Thus, we would like to avoid providing something like an ACC threshold to the text.

2. Chevallier, M., Salas-Mélia, D., Voldoire, A., and Déqué, M.: Seasonal forecasts of the Pan-Arctic sea ice extent using a GCM-based seasonal prediction system, J. Clim., 26, 6092-6104, doi:10.1175/JCLI-D-12-00612.1, 2013.

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

4. Wang, W., Chen, M., and Kumar, A.: Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system, Mon. Weather Rev., 141, 1375-1394, doi:10.1175/MWR-D-12-00057.1, 2013.

5. Msadek, R., Vecchi, G. A., Winton, M., and Gudgel, R. G.: Importance of initial conditions in seasonal predictions of Arctic sea ice extent, Geophys. Res. Lett., 41, 5208-5215, doi:10.1002/2014GL060799, 2014.

6. Peterson, K. A., Arribas, A., Hewitt, H. T., Keen, A. B., Lea, D. J., and McLaren, A. J.: Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system, Clim Dyn., 44, 147-162, doi:10.1007/s00382-014-2190-9, 2015.

7. Guemas, V., Chevallier, M., Déqué, M., Bellprat, O., and Doblas-Reyes, F.: Impact of sea ice initialization on sea ice and atmosphere prediction skill on seasonal timescales, Geophys. Res. Lett., 43, 3889-3896, doi:10.1002/2015GL066626, 2016.

8. Sigmond, M., Reader, M. C., Flato, G. M., Merryfield, W. J., and Tivy, A.: Skillful seasonal forecast of Arctic sea ice retreat and advance dates in a dynamical forecast system, Geophys. Res. Lett., 43, 12457-12465, doi:10.1002/2016GL071396, 2016.

9. P2L16: Again, I think that the term "seasonal-to-interannual" needs to be relocated, this time in front of "predictability".

As suggested, we correct (P2L17).

10. P3L7-8: "eight ensemble members produced by perturbing the sea surface temperature based on the observational errors" - I am wondering whether these perturbations are able to generate any meaningful spread, given that the 3D ocean and atmosphere are assimilated *towards the same, gap-free, reanalyses. Or, are the differences just very small (and all "assimilations" thereby very similar; note that Fig.2 also shows just one single "assimilation"), but of course sufficient to trigger subsequent divergence during the free forecast/hindcast runs due to atmospheric chaos, so that the same effect could have been obtained with quasi arbitrary small initial perturbations? Maybe the authors can comment.*

Thank you for your comments. As for assimilation experiments, the ensemble spreads for detrended SIE_{AO} are range from 10^2 to 10^3 km^2 (not shown) and therefore the time series of SIE for each member appear to a single curve. As you pointed out, the spread is very small and therefore the same effect could be obtained with small initial perturbations. However, in the present study, we have not conducted any hindcasts with small initial perturbations, for example, by the lagged averaged forecast (LAF; [9]Hoffman and Kalnay, 1983) method. Thus we cannot evaluate whether the initial SST perturbations are an effective method for producing the ensemble members or not, which will be remained as future works. At least, the time series of the ratio of ensemble spread for hindcasts to the corresponding RMSE indicates that the ensemble spread for hindcasts have values close to the RMSE (Figure B1), although are small for September. The initial SST perturbation methods seem to produce the meaningful spread to some extent.

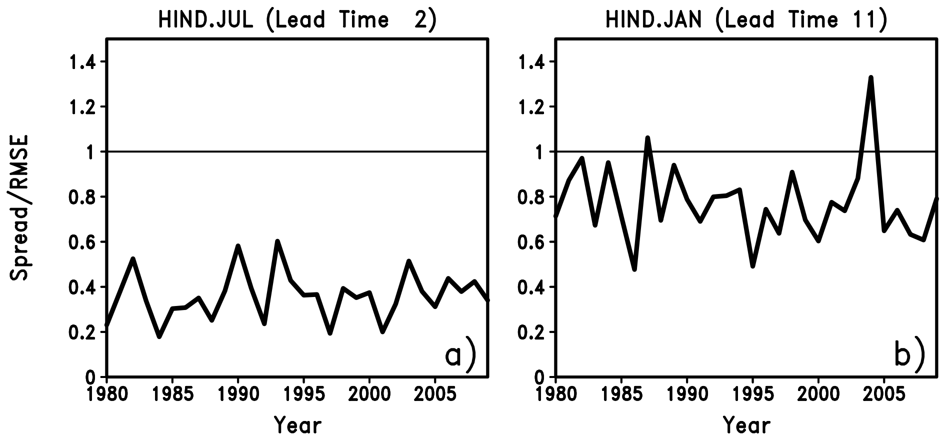

Figure B1. Time series of the ratio of prediction ensemble spread to the RMSE for (a) September started in July 1st and (b) December started in January 1st.

9. Hoffman, R. N., and Kalnay, E.: Lagged average forecasting, an alternative to Monte

Carlo forecasting, Tellus, 35A, 100-118, doi:10.1111/j.1600-0870.1983.tb00189.x.

11. P3L17-18: "the detrended components were calculated by subtracting monthly linear trends during 1980–2009 from the original monthly data, and anomalies are defined as deviations from the climatology from 1980–2009" - are not the "detrended components" mentioned at the beginning of this sentence already the "anomalies"?

The "detrended components" are not anomalies. Firstly, anomalies are calculated by the definition described in the text, and then the linear trend is removed from anomalies.

12. P4L6-7: "September SIE_AO can be dynamically predicted from the previous July" - again, I think this statements needs some quantification; the same holds for the subsequent sentence.

As suggested, we added the values of ACC to the text (P4L16, P4L17, and P4L19).

13. P4L8-9: "The ACC is also significant for the winter SIE_AO, in particular for December, except for the hindcasts started from April 1st, indicating the potential use of dynamical forecasts up to 1 year ahead" - the fact that December SIE_AO is more skillfully predicted by the January hindcasts than by the April hindcasts, also visible in Fig.2, deserves more explanation. While such "reemergence" of skill is often encountered when simple statistical relations - like persistence - are used, in situations with strong seasonal cycles like given for sea ice, to my understanding this is not to be expected for dynamical forecasts: the closer to the target date they are initialised (taking into account current as well as past observations!), the better should the dynamical forecasts become. To be specific, the OHC content anomalies put into the January hindcasts should also make it into the April hindcasts, although subject to some advection etc. Instead, could this unexpected drop of forecast skill be a mere matter of sampling uncertainty?

In the present study, the December SIE_{AO} can be predicted from January 1st but not from April 1st. To provide more explanation, here we considered two possibilities for the reasons. Firstly, we created the same figure as Figure 4 for the control experiment (Figure S3) and the April hindcasts (Figure S5). As in Figure 4, significant regression and correlation patterns appear in Figures S3 and S5. This suggests that the same physical

mechanism occurs in the hindcasts started from April 1st. Thus the sampling uncertainty may not be the main reason for difference between the January hindcasts and the April hindcasts. Secondly, we compared the SIC RMSE between the observations and the hindcasts. In the Barents Sea contributing to the skill of the December SIE_{AO}, the SIC RMSE in April is larger in the April hindcasts than the January hindcasts (Figure B2). Possibly, the larger RMSE at 0 month lead time in the Barents Sea is the reason why the December SIE_{AO} cannot be predicted by the April hindcasts. In the revised manuscript, we added "In contrast, the December SIE_{AO} cannot be predicted from April 1st (Fig. 2a), although significant regression and correlation patterns appear in the results for the April hindcasts (Fig. S5). This may be because the RMSE for April SIC in the Barents Sea is larger in the April hindcasts than the January hindcasts (not shown)." to the text (P6L6-8).

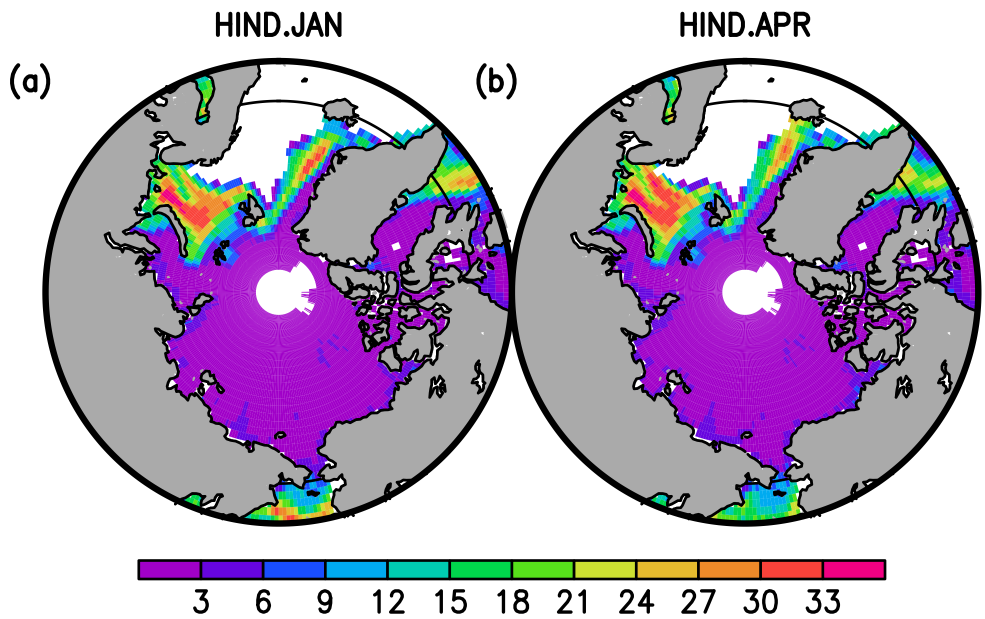

Figure B2. April RMSE for sea ice concentration (%) at (a) 3 months lead time from the January hindcasts (HIND.JAN) and (b) 0 month lead time from the April hindcasts (HIND.APR).

14. P4L10: "The RMSE for all 10 hindcasts increases throughout the melting and early freezing seasons (July–October), before decreasing in November–June" - to be precise, it appears that *the RMSE does not "increase" and "decrease" during those periods, but that it "is larger" and "is smaller" (with the change happening in between).*

The sentence in the previous manuscript was not precise. As suggested, we rewrote this part as follows. "The RMSE for all hindcasts is larger throughout the melting and early freezing seasons (July-October), before smaller values in November-June." (P4L20-21).

15. P4L18-19: See again my comment on P4L8-9!

Please read our response to referee's comment 13.

16. P4L19: "those started from July 1st, in which only the September SIE_AO is significant" - this statement seems to contradict Fig.2 where there are many "significance stipples" for other target months as well.

Your comment is quite right. We removed "only" from this sentence (P5L3).

17. P4L22-23: "the SIV_AO is defined as the sum of the grid cell volumes obtained by multiplying the sea ice thickness (SIT) by the SIC for grid cells with SIC greater than 15 %" - if I am not mistaken, the multiplication by the grid-cell areas is missing here, no?

Thank you. As you pointed out, this sentence was not precise, but the SIV itself was correctly calculated. We added "and the area" to the text (P5L7).

18. P4L23-25: "the OHC_AO is the vertically integrated temperature multiplied by the density and specific heat capacity of seawater from the mixed layer depth (MLD) to a depth of 200 m, in the area north of 65◦ N" - (i) The way it's defined here, temperature is vertically integrated instead of averaged, so the distance from the MLD to 200m directly enters the "OHC" and should thereby dominate variations in "OHC" instead of temperature variations, which seems odd. Please clarify.

In the previous manuscript, we did not consider the heat content in the mixed layer, in order to remove the direct effects due to the atmospheric heating and cooling. However, as

you pointed out, our previous definition of the OHC is affected by seasonal changes in the distance from the MLD to a depth 200 m (i.e., water volume). According to suggestions from referee #1 and referee #2, in the revised manuscript, we recalculated the OHC from the surface to a depth of 200 m and rewrote the text using new Figures 3 and 4. For comparison, we also added Figures 3d-3f and Figures 4c-4f in the previous manuscript to supplement as new Figure S4.

19. (ii) Why is not the same area used as for SIE_AO, that is, excluding Hudson Bay and Baffin Bay?

As you pointed out, we should calculate in the same region used as for SIE_{AO}. In the revised manuscript, we recalculated the SIV and OHC in the domain north of 65^{o}N excluding Hudson Bay and Baffin Bay. Please see new Figure 3.

20. P5L7-18 and Fig.4c-f: I am not convinced that the "advection and emergence hypothesis" constructed here is sufficiently supported by the results shown. Firstly, some of the apparent propagation of ("subsurface") OHC anomalies from off the Scnadinavian western coast to the eastern Barents Sea might be simply due to a slight shift of the area with a mixed layer deeper than 200m (areas with quite deep convection): a larger part of the Barents Sea is thereby effectively "masked" in March compared to December in Fig.4 Secondly, the sea-ice edge extends further into the Barents Sea in March compared to December (I assume this is true also in these simulations), and ocean temperatures under ice are subject to weaker variability (with the surface being tied to the freezing point). Thirdly, the rather narrow stripe of anomalies off the Scandinavian coast in March - an important part of the presented explanation - is not present in the control run (Fig. S6). Maybe some clarification could be provided if Fig.S5 was also provided for lags -3, -6, and -9 months? It might also help to clarify things if the integration/averaging was done between fixed depths, so that nothing is masked and the MLD changes do not superimpose temperature anomaly changes. Even more simply, showing just SST anomalies might help.

Thank you for your suggestions. As you pointed out, ocean is masked when the mixed layer depth become deeper than a depth of 200 m. Firstly, we recalculated the OHC from the surface to a depth of 200 m in the revised manuscript, as mentioned in our response to referee's comment 18. Next we reconstructed new Figure 4 using new OHC and partly

rewrote the Section 4 (P5L1-P6L21).

21. P5L15-16: "The above features are also found in the control run, suggesting that the advection processes of the OHC in the hindcasts are not due to processes distorted by the influence of initialization or climate drift in MIROC5" - In fact, I do not quite understand the reason why the main figures related to the lagged correlation and regression alaysis are not based on the control run in the first place. Maybe it's just me, but I am somewhat confused why this should be done primarily for the hindcasts, where also the statistical sampling is much worse. If the main analysis was based on the control run, however, it would make sense to show corresponding results for the hindcasts as a supplement, to prove that the shown relations still hold, no?

Referee comments may be correct. However, the main analysis using data from the hindcast experiments appear to be natural, as the first step, in order to investigate the physical processes contributing to the prediction skill of SIE_{AO}. Meanwhile, since the hindcast data may be influenced by climate drift or initialization, a control experiment without initialization and anthropogenic effects is complementally used to interpret the analyzed results.

22. P5L24: "the persistence of sea ice states initialized in July persists" - the first word maybe should be "anomalies" or similar?

Here we would like to state that initialized sea ice states persist until September. In the revised version, we changed "the persistence of sea ice states" to "the sea ice states". (P6L13).

23. P26-27: "possible mechanisms or sources cannot be detected in the hindcasts started from April 1st (Fig. S8)" - I'd like to repeat my points that this might be partly due to sampling, and that important regions are "masked" due to the MLD-related OHC definition. I would argue that Fig.S6d, based on the control run (implying better sampling, although showing March instead of April), supports the notion that the April state should be at least as informative as the January state to predict September SIE_AO.

In the revised manuscript, we recalculated the OHC from the surface to a depth of 200 m, according to suggestions from referee #1 and referee #2, and then reconstructed Figure S8 in the previous supplement as new Figure S6. In the hindcasts started from April 1st, the September SIE_{AO} shows similar lagged correlation patterns to the July hindcasts for SIV_{AO} (Figure S6a) and OHC_{AO} (Figure S6b). Thus, the same physical processes as the July hindcasts are expected to work in the April hindcasts. However, the positive regression and correlation patterns for SIC and SIT are weaker than those for the July hindcasts, particularly in the Pacific Sector of the Arctic Ocean (Figures S6c and S6d). In addition, the same figures based on the control experiment as Figure S6c and S6d are shown in Figure S7. Similar positive correlation and regression patterns for SIC and SIT clearly appear in the Pacific sector of the Arctic Ocean, as in Figure 5. As you pointed out, the sampling uncertainty may be one reason for unclear signals in the hindcasts started from April 1st.

In the revised manuscript, we removed "In contrast, possible mechanisms or sources cannot be detected in the hindcasts started from April 1st (Fig. S6), at least from the lagged correlation and regression analyses, although the September SIE_{AO} is weakly correlated with the SIV_{AO} and the OHC_{AO}." (P5L26-28 in the previous manuscript), and then newly added "In the hindcasts started from April 1st, the September SIE_{AO} shows similar lagged correlation patterns to the July hindcasts for SIV_{AO} (Fig. S6a) and OHC_{AO} (Fig. S6b). Thus, the same physical processes as the July hindcasts are expected to work in the April hindcasts. However, the positive regression and correlation patterns for SIC and SIT are weaker than those for the July hindcasts, particularly in the Pacific Sector of the Arctic Ocean (Figs. S6c and S6d). In contrast, similar patterns to Fig. 5 clearly appear in the Pacific sector of the Arctic Ocean for the control experiment (Fig. S7). These results suggest that the persistence of sea ice contributes to the skill of September SIE_{AO} started from April 1st, but the sampling uncertainty may lead to unclear signals in Fig. S6." to the text (P6L15-21).

24. P6L4-6: "Numerical experiments to confirm whether the subsurface OHC anomalies 5 originating from the North Atlantic control the December sea ice extent in the BS and eventually in the Arctic Ocean will be explored in future work." - I am actually quite curious to see results of such interesting experiments!

Thank you for your interest. As mentioned in the text, we will conduct such an experiment in future works.

25. P6L7-9: The first two sentences of this paragraph seem to contradict each other.

As you and referee #1 pointed out, these two sentences were contradictory. We removed the second sentence "Nevertheless, we note that the forecast skill of summer SIE_{AO} is not necessarily low, because the hindcasts initialized in January and April have significant skills for SIE_{AO} in August and September" (P6L7-9 in the previous manuscript) from the revised text.

26. P6L20: "Further improvements in the predictability of sea ice" - here I would recommend to avoid the term "predictability (of)" because in my view "skill to predict" is more accurate.

As suggested, we replaced "predictability" with "skill to predict" (P7L14).

27. Fig.2: I would find it helpful if the situations shown in panels c) and d) could be highlighted in panels a) and b), e.g., by black boxes around the corresponding fields of the heat maps. Also, do I understand correctly that panel c) corresponds to a 3 months lead time, whereas panel d) corresponds to a 11 months lead time? That could be stated more clearly in the caption.

In the revised manuscript, we reconstructed Figure 2 following your suggestions. Please see new Figure 2.