

Response to reviewers

We thank the reviewers for their careful reading of our manuscript and their feedback. Below we include our response and changes to the manuscript. The original comments are quoted in gray.

Reviewer 1

General comments

General comment 1

The authors quantify the importance of initial conditions on seasonal prediction by comparing two versions of ACCESS model. However, the data assimilation scheme of these two versions is different in multiple aspects. This raises the question of which components of the initial conditions are truly responsible for the skill differences. In this manuscript, the authors mostly examine the role of sea ice initial condition but the potential influence of other variables appears to be underexplored and should be discussed more explicitly.

We thank the reviewer for their comment and agree the sea ice initial conditions are not the sole difference between the two systems; the ocean initial conditions and ensemble generation also differ slightly. Unfortunately, the raw ACCESS-S1 hindcast is no longer available, so we cannot directly compare these aspects between the systems.

We have tried to address the difference of ocean initial conditions through previous literature. In their assessment of ACCESS-S2, Wedd et al. (2022) suggests that initial conditions in the Southern Ocean are similar between the two systems, so we think that this should be a minor difference. The ensemble generation would mainly affect ensemble spread, which according to Fig 14 (Fig A15 in the revised manuscript), is comparable between the two systems after just a few days.

We have added a section pointing out these limitations (Lines 290–320) and mention them explicitly in the conclusions.

General comment 2

It remains unclear whether the interannual variability of SIE increases with lead time in ACCESS-S2, while this behavior is not evident in ACCESS-S1. The authors suggest that the thinner ice in ACCESS-S2 is responsible for this difference; however, sea ice becomes thinner with lead time in both systems. The authors then propose that data assimilation may lead to an imbalanced state. Could the authors provide a clearer conclusion or more direct evidence to explain the contrasting behavior between ACCESS-S1 and ACCESS-S2?

In the revised manuscript we have revised this section to improve clarity (Lines 218–237). We have also added a figure showing the correlation between sea-ice thickness and sea-ice concentration amplitude (Fig. 8), which supports our assertion that thinner sea ice is more variable than thicker sea ice. A more thorough understanding of the differences between the two modelling systems is beyond the scope of this study.

General comment 3

Many climate models suffer from systematic bias, and bias correction is often applied prior to forecast evaluation. Have the authors examined whether bias correction would lead to substantial differences in the reported results? Clarifying the role of model bias in the skill assessment would strengthen the robustness of the conclusions.

Since we compute anomalies relative to each model's own lead-time-dependent climatology, this applies a first-order bias correction. Hence we cannot determine the separate impact of bias correction on the results.

Specific comments

P5 Line146, I have a careful reading on the textbook by Murphy and Daan (1985) and found that the skill score S is defined based on the MSE rather than the RMSE (see equation 23 in the textbook), despite I believe these two definitions won't lead to dramatic difference in conclusion.

Yes, Murphy and Daan (1985) use MSE to define skill in equation 23; however, the general definition of skill score in section 4.3 states that it can be derived from any error measure. You are correct that it does not make a difference to the results since it is a monotonic transformation from one to the other.

P7 Line 175, according to my observe, the period should be from April to September?

Thank you, we have corrected the error.

P12 Line 200, here the authors suggest that the decrease in interannual variability with increasing lead month in ACCESS-S2 is responsible for the improvement in RMSE. However, I found that for some months (e.g., July, August, September) the interannual variability increase with the lead time. Could the authors clarify how this explanation applies to these months?

The reviewer’s finding is correct in some months interannual variability increases with lead time. However, in those months RMSE also increases with lead time instead of decreasing. We added Line 212: “This effect is seen in all months except for July to September.” and a supplementary figure showing RMSE of sea-ice extent as a function of lead time for each month.

P12 Line 206, how does the sea ice thickness adjust when assimilating the sea ice concentration?

During data assimilation, if sea-ice concentration innovations are positive, these are added to the first ice category with a fixed thickness of 50 cm. If they are negative, they are removed from the thinnest category first and from thicker categories if needed. Therefore, although this system only assimilates sea-ice concentration data, sea-ice thickness is nonetheless modified.

We have added this clarification to the revised manuscript (Lines 88–91).

P12 Line 210, please consider examining the role of the ocean in the bias of sea ice concentration magnitude. I understand that the ocean conditions are nudged toward a reference dataset only when SST exceeds 0 degree. It is therefore possible that, in the regions of interest, SST remains below this threshold, in which case the nudging wouldn’t effectively remove ocean-related model biases.

Yes, that is indeed possible. As we explained in the answer to General comment 1, ACCESS-S1 and ACCESS-S2 seem to have similar ocean initial conditions even at relatively high latitudes so it is possible that this effect is not too significant. Unfortunately we can’t explore this in depth due to the aforementioned lack of data for ACCESS-S1.

P12 Line 211, which doesn’t assimilate

We have corrected this.

P12 Line 214, sea ice concentration anomaly

We have corrected this.

P15 Line 245, should it be the June that cannot be forecasted better than the benchmarks?

Yes. This is now corrected.

In addition, the metric employed in Libera et al (2022) was ACC, which is slightly different from the RMSE used in this study.

Yes, you are correct. Based on reviewers' comments we realised that there are several methodological differences that prevent a direct comparison between our results and those of Libera et al. (2022) and modified our discussion about their work in the revised manuscript. See response to R3's Major comment 4.

P17 Line 255, Is the eastward propagation feature also evident in December? Are there any criteria used to determine whether eastward propagation is present?

We are not aware of any objective criterion to derive propagation of error. We follow Holland et al. (2013), who based their conclusions on visual inspection of a similar Hovmoller plot.

P22 Line 289, subseasonal to seasonal.

We have corrected this.

Figures

Figure 1: Could the author clarify why the maximum lead time in ACCESS-S1 varies between 213 and 216 days. Was this choice intended to align with the ACCESS-S2?

The triangles show the mean forecasted extent for the first of the month using the maximum lead time possible. That lead time is always about 7 months, but the exact number of days depends on the month due to the different number of days in each month. We have now clarified this in the caption.

Figure 5: what does the word in the bracket represent, e.g., (CI: 0.04...0.27). The confidence interval? Then which confidence level? I recommend adding the asterisk to show the result is significant or not.

The numbers in brackets represent the 95% confidence interval. We have now clarified this in the caption. We added asterisks to correlation values statistically different from zero at the 95% confidence level.

Figure 6, how do you compute the 95% confidence level? Is it based on an assumption of normal distribution for the ratio of SIE between the prediction and the reference?

We assume a normal distribution for sea-ice extent, so the standard deviation has a chi-squared distribution with $N - 1$ degrees of freedom (where N is the number of samples). We then divide all the values by the observation standard deviation. We have modified the caption to make this clearer.

Figure 9, I found that the RMSE of ACCESS-S2 mostly overlay with that of climatology. Does this imply that the oceanic and atmosphere data assimilation provide limited benefit to the prediction?

Yes, that is correct that the RMSE overlays the climatology. However, with our data, it is not possible to say to what extent the oceanic and atmospheric data assimilation improve sea-ice forecasting. It might very well be the case that if it were not for data assimilation, the forecasts would have been even worse as it is evident that some information is being transferred from the atmosphere or ocean to the ice in some months (Figure 1).

What role does the model bias play in this prediction error?

Since we compute RMSE on anomalies with respect of the model climatology, we are correcting (at least at first order) for model bias.

Figure 10-11, please consider clarifying whether this skill score is significant or not

We have added stippling showing statistical significance.

Reviewer 2

General comments

General comment 1

Improper methodology for addressing the stated objective The manuscript emphasizes the importance of sea ice initialization (e.g., L28–29, L56–57, and L85). While ACCESS-S1 and ACCESS-S2 share identical model configurations (e.g., L54 and L65–73) and both use atmospheric initial conditions derived from ERA-Interim (L74), their ocean and sea ice initial conditions differ substantially in both assimilation methods and observational datasets (L74–84). Consequently, a direct comparison between ACCESS-S1 and ACCESS-S2 does not isolate the impact of sea ice initialization alone, as ocean initialization is known to play a critical role in sea ice predictability. Furthermore, the ensemble generation strategies differ between the two systems (L93–100), which may also affect the comparison of ensemble-mean results. I therefore suggest revising the stated objective of the manuscript and explicitly analyzing the combined impacts of ocean, atmosphere, and sea ice initialization, rather than attributing the results solely to sea ice initialization, given the fully coupled nature of the system

You are correct in pointing out that there are other differences between the two forecasting systems. The two other differences besides sea-ice initial conditions are the ocean initial conditions and the ensemble generation.

Ideally we would compare the ocean initial conditions – Southern Ocean SSTs in particular and recreate a time-lagged ensemble with ACCESS-S1. Unfortunately, the raw ACCESS-S1 hindcast is no longer available, so we cannot do either.

However, on their initial assessment of ACCESS-S2, Wedd et al. (2022) suggests that initial conditions in the Southern Ocean are similar between models, so we think that this should be a minor difference. The ensemble generation would mainly affect ensemble spread, which is comparable between systems after just a few days.

Due to the likely minor impacts of the other factors and our inability to analyse these other factors directly, we feel our stated objective remains the same.

We have added a section pointing out these limitations (Lines 290–320) and mention them explicitly in the conclusions.

General comment 2

Quality of figures I have several specific comments regarding the figures. Improving their quality and clarity would substantially enhance the overall quality of the manuscript and improve its readability.

We improved the figures based on reviewer feedback.

Specific comments

L64, Section 2.1 (ACCESS-S1 and ACCESS-S2): Please ensure consistency between the section title and its content. Clarify whether this section sufficiently describes the key differences between ACCESS-S1 and ACCESS-S2 that are relevant to this study.

We modified the section title and made it clear that ocean and sea-ice initial conditions and ensemble generation procedure are the only differences between ACCESS-S1 and ACCESS-S2.

L125, Section 2.3: The current version of Section 2.3 primarily describes RMSE and skill scores. However, bias is extensively used in Section 3 (Results) and should therefore be formally introduced and documented in the methodology section. In addition, correlation is also employed (e.g., Fig. 5) and is a widely used metric for evaluating seasonal sea ice forecasts; it should be described as well. Furthermore, the authors should report the statistical significance or confidence intervals for RMSE and correlation

Thank you for the suggestion. We have added bias and correlation to the Methods section. We added confidence intervals when possible (Figures 5, 6, 8, 9, 14) and significant p-values to Figures 11-13.

L126–127: It is unclear how the ensemble mean sea ice extent (SIE) is calculated. Do the authors first compute the ensemble mean of sea ice concentration (SIC) and then derive SIE from the mean SIC, or do they compute SIE separately for each ensemble member and then analyze the ensemble mean of SIE? This should be explicitly clarified in the text.

The hindcast ensemble mean sea-ice extent is defined as the average of the sea-ice extent of individual ensemble members. We added this clarification in the Methods section.

Figures 3 and 4: These figures are difficult to read. Adding latitude–longitude grid lines would improve readability. In addition, narrowing the colorbar range (e.g., from -1 to 1 to -0.5 to 0.5) may help highlight relevant spatial patterns

Agreed. We have changed the colour scale and added lat-lon lines.

L194: The term “skillful” is ambiguous here. Please clarify whether it refers to skill assessed using RMSE, correlation, or another metric.

We have clarified that we are talking about RMSE.

L214–215: It appears that the magnitude of sea ice anomalies in ACCESS-S2 is too small, rather than too large, as currently stated. Please verify and revise this interpretation

There was an error in the labeling of the plots where observations were labeled as ACCESS-S2. This is now fixed.

L245: “June cannot be forecasted...”?

Fixed

Figures 9 and 10: Figure 10 largely overlaps with the information presented in Figure 9. It may not be necessary to include Figure 10, or its added value should be better justified.

We have moved Figure 10 to the supplementary material, and replaced it with a figure highlighting the lead time at which ACCESS-S1 and ACCESS-S2 forecasts are statistically indistinguishable.

Figures 11–13: These figures are difficult to interpret. Consider using an alternative colorbar, narrowing the colorbar range, and adding longitude grid lines to improve clarity

We altered the colour scale, added a vertical grid, and smoothed values to increase readability.

Caption of Figure 14: Please clarify the lead time shown (e.g., 1, 3, and 30 days?).

We have added “days” in the figure to make the unit clearer.

L277–278: As shown in Fig. 14c1, the two systems clearly differ.

Yes, the spread is slightly higher in S1 between Feb and April. Lines 303–304 now point that out.

L279–280: How is the standard deviation computed? Is it calculated across years, ensemble members, or both? Please clarify.

Figure 14: The methodology used to decompose the forecast error spread into the mean standard deviation and the standard deviation of the ensemble-mean error is not sufficiently clear and should be explained in more detail

We have reworked this figure and have removed the decomposition of variance information based on reviewer comments.

L307–308: Please clarify how ACCESS-S1 updates sea ice states during the assimilation step. Specifically, does the assimilation update both sea ice concentration and volume across different thickness bins, or does it retain the prior thickness distribution and adjust sea ice volume proportionally during the post-processing of the assimilation step?

During data assimilation, if sea-ice concentration innovations are positive, these are added to the first ice category with a fixed thickness of 50 cm. If they are negative, they are removed from the thinnest category first and from thicker categories if needed. Therefore, although this system only assimilates sea-ice concentration data, sea-ice thickness is nonetheless modified.

We have added this clarification to the revised manuscript (Lines 88–91).

L313–314: I kindly disagree with the authors’ statement here; please refer to my general comments for further explanation

We agree that this statement is too strong and ignores the limitations of these study. However, as we argue in the response to the general comments above, we believe that the difference in ocean data assimilation and ensemble generation have a smaller impact on the quality of the sea-ice forecast than the lack of sea-ice data assimilation. We qualified this sentence and added an explicit discussion of these limitations in the conclusions.

Additional references: I recommend citing the following two studies, which are highly relevant to the discussion of Antarctic sea ice initialization and predictability: <https://journals.ametsoc.org/view/journals/clim/34/15/JCLI-D-20-0965.1.xml> <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2024MS004382>

We thank you for the recommendations. Xiu et al. (2025) was particularly useful since it uses a similar methodology.

Reviewer 3

Major comments

Major comment 1

The simulation periods (S1: 1990–2012; S2: 1981–2018) and forecast lead times (S1: 212 days; S2: 273 days) of the two systems are inconsistent. Please clarify the rationale for these differences and verify whether the overlapping period (1990–2013) can fully represent the variability of Antarctic sea ice (e.g., covering extreme events and long-term trends). It is recommended to standardize lead times for fairer comparison if feasible.

We apologise that it wasn't clear that all computations were done using only the overlapping period between hindcasts. We have now clarified this in the Methods section, including the clarification that this period does not cover the recent decade of low sea-ice extent. We cut Figures 11 to 13 in the original manuscript to use the same maximum lead time for easier comparison. Figures 6 and 8 in the original manuscript also use the longer maximum lead time for ACCESS-S2, but in those cases we believe that the extra two months are useful for understanding ACCESS-S2 behaviour and the message does not depend on a common maximum lead time.

Major comment 2

Excessive monthly subplots (e.g., Figures 6 and 8) lead to redundant information and lack narrative progression. For results reflecting common characteristics across months, aggregating or simplifying subplots is suggested to avoid distracting readers from core findings.

We appreciate the reviewer's concern about monthly subplots. While we agree that some months do not show strong signals, those months change depending on the figure. We also feel that readers may be interested in seeing the complete evolution of results over the year. Hence for consistency across all figures and to show the evolution of results over the year we have decided to keep monthly subplots.

Major comment 3

The explanation for the anomalous phenomenon (S2's RMSE decreasing with increasing lead time) is confusing due to disorganized logic and irrelevant information. The proposed causal chain—"systematic thin sea ice in S2 → over-sensitivity to atmospheric/oceanic forcing → overestimated interannual variability at short

lead times \rightarrow convergence to model climatology at long lead times \rightarrow reduced RMSE”—needs to be clearly structured, with key links (e.g., thin sea ice vs. variability) supported by direct evidence (e.g., spatial correlation between ice thickness and variability).

In the revised manuscript we have revised this section to improve clarity (Lines 218–237). We have also added a figure showing the correlation between sea-ice thickness and sea-ice concentration amplitude (Fig. 8), which supports our assertion that thinner sea ice is more variable than thicker sea ice. A more thorough understanding of the differences between the two modelling systems is beyond the scope of this study.

In addition, Figure 9 also presents the RMSE for sea ice concentration anomalies (SICA). Does the same error trend apply to the RMSE of SICA as well?

We thank you for the question. While we haven’t computed RMSE of monthly sea-ice concentration anomalies, we can approximate that by computing the monthly mean RMSE of the daily RMSE. Figure 1 shows this mean RMSE either decrease or stays constant with lead time for most months in ACCESS-S2 hindcasts. The pattern is broadly consistent with the pattern observed in monthly sea-ice extent RMSE.

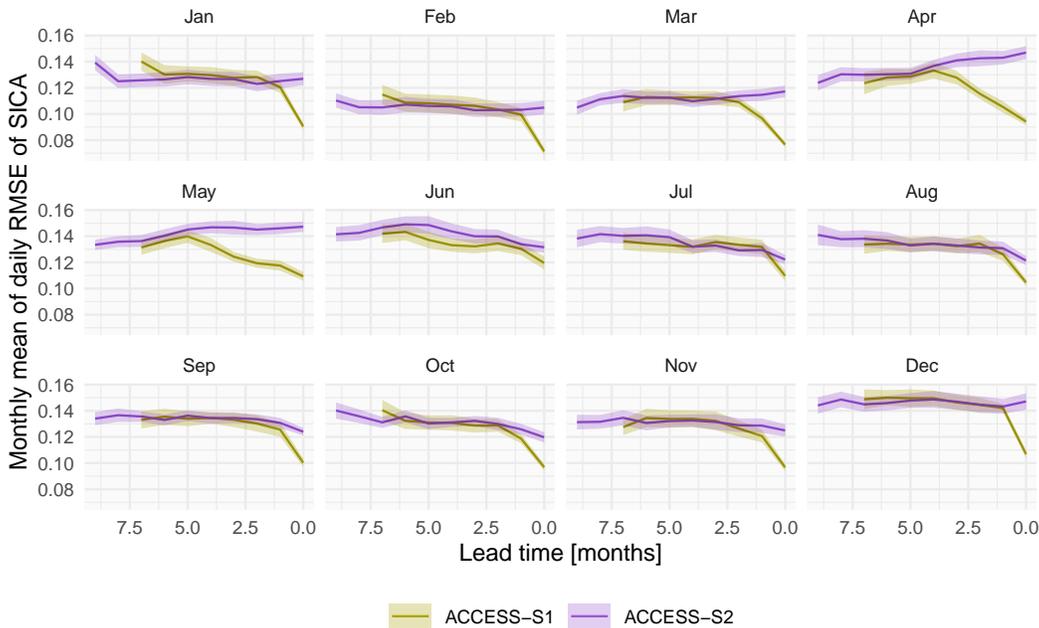


Figure 1: Mean RMSE of mean daily RMSE of sea-ice concentration anomalies. Each panel indicates the target month. Note the reverse horizontal axis.

Major comment 4

Line 253 references a “winter predictability barrier” in the Weddell Sea, but the study’s results show the most prominent barrier in the King Haakon Sea (0° – 120° E). Please explain this discrepancy, considering factors such as regional circulation differences or model representation of ocean-sea ice processes.

We thank the reviewer for this comment. We have thought about the discrepancy further and come to the conclusion that there are many differences between our work and Libera et al. (2022) that actually prevent a direct comparison.

First, by analysing sea-ice area, Libera et al. (2022)’s results are insensitive to the distribution of sea ice within the Weddell Sea region, while our use of RMSE of sea ice concentration anomalies takes this spatial distribution into account. Figure 2 illustrates how the choice of method can lead to different results.

Second, Libera et al. (2022) used lagged autocorrelation as the definition of predictability – essentially measuring persistence – while we are analysing predictions from a fully coupled model.

We have decided to change our wording to make the distinction clearer (Lines 364–367).

Major comment 5

The key conclusion “ACCESS-S1 and ACCESS-S2 forecast errors are statistically indistinguishable after just two weeks” is repeatedly mentioned in the abstract and conclusion but lacks explicit textual support. The only relevant evidence is in Figure 9 (parenthetical values), which should be discussed in the main text.

You are absolutely right. We have added a figure (Fig. 10 in the revised manuscript) and a paragraph highlighting the key result (Lines 257–262).

Major comment 6

Figures 11–13 are information-dense but low in informativeness, with cumbersome correspondence between text and subplots. Alternative visualization methods (e.g., spatial aggregation, key region highlighting, or summary statistics) are recommended to convey core messages (e.g., regional skill differences) more directly.

Similarly to the answer to Major comment 2, we believe that in order to show how the results depend on seasonality, we want to show as much information as possible. Other researchers might be interested in what results look at different months of the year as well.

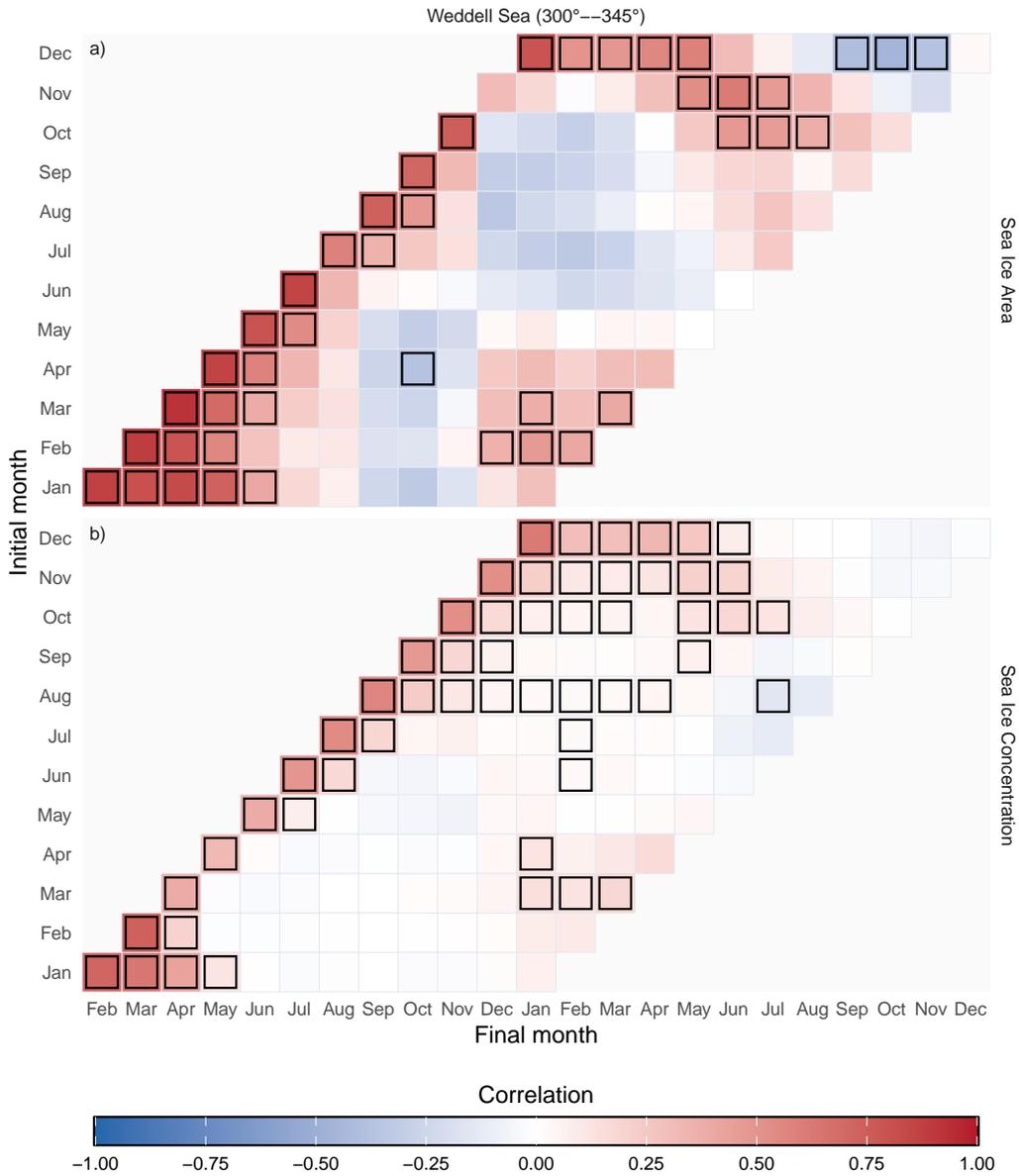


Figure 2: Autocorrelation of monthly sea ice area at different lags (panel a) and mean lagged sea ice concentration anomaly pattern correlation at different lags (panel b) in the Weddell Sea. Values statistically different from zero at the 0.05 level are highlighted with black edges.

Minor comments

Line 30 has an ambiguity: clarify whether sea ice initial conditions are important in “autumn” (as per Arctic studies) or “spring” (relevant to Antarctic seasons).

We clarified that that result is relevant for the Arctic.

Section titles 3.1 (“Bias”) and 3.2 (“RMSE”) are overly simplistic. Revise them to reflect core content (e.g., “Bias Characteristics of Sea Ice Extent and Concentration” or “RMSE and Skill Score Analysis of Forecast Anomalies”).

We changed the section titles.

In Figures 3 and 4, flip the color bar to follow conventional standards (blue for underestimation, red for overestimation) and standardize color bars across all relevant figures.

It is not uncommon to use the cold colour to denote higher sea ice concentrations (which generally imply colder temperature) and the warm colour to denote lower sea ice concentrations (associated with higher temperatures) (e.g. Espinosa, Blanchard-Wrigglesworth, and Bitz 2024; Massonnet et al. 2023; Morioka et al. 2022). We prefer to use this convention.

In Figure 5, align the layout with other figures (place S1 on the left, S2 on the right) and add subfigure labels (e.g., a), b)) for clarity.

We have changed the column order and added labels to each panel.

In Figure 6, add subfigure labels and specify the unit of “lead time” (e.g., “months”). Synchronize these revisions across similar figures.

We have added units to that and other figures.

Consistently use “sea ice” (instead of “sea-ice”) throughout the text.

We chose to use “sea ice” as an open compound noun when used on its own (“Accurately modelling Antarctic sea ice is essential...”) and “sea-ice” as a hyphenated adjective (“...progress in Antarctic sea-ice forecasting system has lagged behind...”) (Merriam-Webster n.d.).

We corrected some instances in which we used the hyphenated version when used as a noun.

Define abbreviations for key terms at their first appearance: sea ice concentration (SIC), sea ice thickness (SIT), and sea ice extent (SIE).

We don’t use abbreviations for these terms.

In Figure 10’s caption, correct “larger than” to “lower than” (consistent with the logic of RMSE exceeding the 95% confidence interval lower bound of persistence forecast RMSE).

We have corrected the caption

Lines 279–280: Since the persistence forecast is listed separately, present the climatology forecast as a separate reference category for clarity.

We have removed this panel in the revised version.

Section 3.3 (“Conclusions”) should be moved to an independent Section 4 to serve as the full text’s comprehensive conclusion.

We have moved the conclusions to its own section instead of a subsection.

References

- Espinosa, Zachary I., Edward Blanchard-Wrigglesworth, and Cecilia M. Bitz. 2024. “Understanding the Drivers and Predictability of Record Low Antarctic Sea Ice in Austral Winter 2023.” *Communications Earth & Environment* 5 (1): 1–9. <https://doi.org/10.1038/s43247-024-01772-2>.
- Holland, Marika M., Edward Blanchard-Wrigglesworth, Jennifer Kay, and Steven Vavrus. 2013. “Initial-Value Predictability of Antarctic Sea Ice in the Community Climate System Model 3.” *Geophysical Research Letters* 40 (10): 2121–24. <https://doi.org/10.1002/grl.50410>.
- Libera, Stephy, Will Hobbs, Andreas Klocker, Amelie Meyer, and Richard Matear. 2022. “Ocean-Sea Ice Processes and Their Role in Multi-Month Predictability of Antarctic Sea Ice.” *Geophysical Research Letters* 49 (8): e2021GL097047. <https://doi.org/10.1029/2021GL097047>.
- Massonnet, François, Sandra Barreira, Antoine Barthélemy, Roberto Bilbao, Edward Blanchard-Wrigglesworth, Ed Blockley, David H. Bromwich, et al. 2023. “SIPN South: Six Years of Coordinated Seasonal Antarctic Sea Ice Predictions.” *Frontiers in Marine Science* 10 (May). <https://doi.org/10.3389/fmars.2023.1148899>.
- Merriam-Webster. n.d. “Should That Word Have a Hyphen? Merriam-Webster.” Accessed February 26, 2026. <https://www.merriam-webster.com/grammar/hyphen-rules-open-closed-compound-words>.
- Morioka, Yushi, Doroteaciro Iovino, Andrea Cipollone, Simona Masina, and Swadhin K. Behera. 2022. “Decadal Sea Ice Prediction in the West Antarctic Seas with Ocean and Sea Ice Initializations.” *Communications Earth & Environment* 3 (1): 189. <https://doi.org/10.1038/s43247-022-00529-z>.
- Murphy, Allan H., and Harald Daan. 1985. “Forecast Evaluation.” In *Probability, Statistics, and Decision Making in the Atmospheric Sciences*. CRC Press.

- Wedd, Robin, Oscar Alves, Catherine de Burgh-Day, Christopher Down, Morwenna Griffiths, Harry H. Hendon, Debra Hudson, et al. 2022. “ACCESS-S2: The Upgraded Bureau of Meteorology Multi-Week to Seasonal Prediction System.” *Journal of Southern Hemisphere Earth Systems Science* 72 (3): 218–42. <https://doi.org/10.1071/ES22026>.
- Xiu, Yongwu, Yiguo Wang, Hao Luo, Lilian Garcia-Oliva, and Qinghua Yang. 2025. “Impact of Ocean, Sea Ice or Atmosphere Initialization on Seasonal Prediction of Regional Antarctic Sea Ice.” *Journal of Advances in Modeling Earth Systems* 17 (2): e2024MS004382. <https://doi.org/10.1029/2024MS004382>.