

# Response to reviewers

## R1

### General comments

#### General comment 1

We agree that the comparison between the two systems is not perfect. 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. Unfortunately, the ACCESS-S1 hindcast is no longer available (we only have access to the postprocessed sea-ice hindcast) so we cannot test how similar ACCESS-S1 and ACCESS-S2 ocean initial conditions are. However, on their initial assessment of ACCESS-S2, Wedd et al. (2022) suggests that initial conditions in the Southern Ocean are similar between models.

Similarly, since we don't have access to the raw sea-ice hindcast, we cannot produce equivalent time-lagged ensembles for ACCESS-S1. We believe this should only entail minor differences and mostly affect ensemble spread.

We added a section pointing out these limitations.

#### General comment 2

We reworked that part of the manuscript to hopefully make it clearer. We added a figure showing the correlation between sea-ice thickness and sea-ice concentration amplitude, which supports our assertion that thinner sea ice is more sensitive than thicker sea ice.

#### General comment 3

Since we compute anomalies relative to each model's own lead-time-dependent climatology, this applies a first-order bias correction.

## **Specific comments**

### **Specific comment 1**

Section 5.3.2 in Murphy and Daan (1985) gives examples of skill scores in terms of Mean Absolute Deviation and Mean Squared Error, but Root Mean Square Error is also a valid measurement from which to derive skill scores:

Finally, a skill score based on the root mean square error (RMSE) measure can be defined by replacing MSEFs and MSERs in (24) with RMSEFs and RMSERS respectively

The general definition of skill score is given in section 4.3.

### **Specific comment 2**

We corrected the error.

### **Specific comment 3**

That is true. However, in those months RMSE increases with lead time. We added Line XX: “This effect is seen in all months except for July to September.” and a supplementary figure showing RMSE as a function of lead time for each month.

### **Specific comment 4**

Yes, indirectly. Lines XXX now read: “The way sea-ice thickness is handled is that sea-ice concentration innovations are either added to the first ice category with a fixed thickness of 50 cm or removed from the thinnest category first and from thicker categories if needed.”

### **Specific comment 5**

See response to General comment 1.

### **Specific comment 6**

We corrected this.

### **Specific comment 7**

We corrected this.

### **Specific comment 8**

Yes. This is now corrected. We modified our discussion about Libera et al. (2022). See response to R3.

### **Specific comment 9**

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.

### **Specific comment 10**

We corrected this.

## **Figures**

**Figure 1** 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 clarified this in the caption.

**Figure 5** The numbers in brackets represent the 95% confidence interval. We clarified this in the caption. All correlations are statistically different from zero, but we do not think that's a meaningful comparison.

**Figure 6** 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 modified the caption to make it clearer.

**Figure 9:** 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. Clearly some information is being transferred from the atmosphere or ocean to the ice in some months (Figure 1).

**Figure 10-11** We added stippling showing statistical significance.

**Additional references** We thank you for the recommendations. (**xiu2025?**) was particularly useful since it uses a similar methodology.

## R2

### General comments

#### Improper methodology for addressing the stated objective

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.

Regarding ocean initial conditions, unfortunately we cannot directly compare ACCESS-S1 and ACCESS-S2 initial conditions because that data is currently not available. However, we believe there are reasons to consider that the effect would be of second-order importance.

First, while the DA process is different, the underlying observations are similar, especially in the high latitudes, where there are few direct observations and the small correlation length-scale (Waters et al. 2015) reduces the influence of observations in more distant locations. The similarity of ocean initial conditions is evident in Wedd et al. (2022)'s Figure 9, where they show the correlation between ocean heat content in the upper 300 m of the ocean between EN4 and ACCESS-S2 and ACCESS-S1 reanalysis. In both systems, correlations are relatively high around 60° and very low along the Antarctic coast, and they exhibit very similar patterns. Second, even though the ocean initial conditions are very similar between the two systems, the SST forecast was shown by Wedd et al. (2022) to be extremely degraded in the Southern Ocean even at zero months lead time. Correlation of SST anomalies in the ice-covered regions is much lower for ACCESS-S2 (Wedd et al. (2022) Figure 13) and with a higher positive bias (their Figure 12), which is consistent with the greater negative sea-ice concentration bias in ACCESS-S2. This suggests that it is the lack of sea-ice data assimilation that is degrading SST forecasts and not the other way around.

Similarly, since we don't have access to the raw sea-ice hindcast, we cannot produce equivalent time-lagged ensembles for ACCESS-S1. We believe this should only entail minor differences and mostly affect ensemble spread. Our analysis of ensemble spread shows that ACCESS-S1 sea-ice starts underdispersed but matches ACCESS-S2's spread after 2 to 6 days depending on the initial date.

We added a section pointing out these limitations.

### Quality of figures

We improved the figures based on reviewer feedback.

## **Specific comments**

**L64, Section 2.1** We modified the section title and made it clear that ocean and sea-ice initial conditions are the only differences between ACCESS-S1 and ACCESS-S2.

**L125, Section 2.3** We 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** The hindcast ensemble mean sea-ice extent is defined as the mean sea-ice extent of individual ensemble members. We added this clarification in the Methods section.

**Figures 3 and 4** We changed the colour scale and added lat-lon lines.

**L194** We clarified that we are talking about RMSE.

**L214–215** There was an error in the labeling of the plots where observations were labeled as ACCESS-S2. This is now fixed.

**L245** Fixed

**Figures 9 and 10** We moved Figure 10 as a supplemental and replaced it with a figure highlighting the lead time at which ACCESS-S1 and ACCESS-S2 forecasts are statistically indistinguishable.

**Figures 11–13** We altered the colour scale, added a vertical grid, and smoothed values to increase readability.

**Caption of Figure 14** Added “days” in the figure to make the unit clearer.

**L277–278** Yes, the spread is slightly higher in S1 between Feb and April. Lines XXX now point that out.

**L279–280 and Figure 14** We reworked this figure and removed the decomposition of variance.

**L307–308** Lines XX now read: The way sea-ice thickness is handled is that sea-ice concentration innovations are either added to the first ice category with a fixed thickness of 50 cm or removed from the thinnest category first and from thicker categories if needed.

## **R3**

### **Major comments**

#### **Major comment 1**

We made it clear that all computations were done using only the overlapping period between hindcasts.

## **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.

## **Major comment 3**

We reworked that part of the manuscript to hopefully make it clearer. We added a Supplementary Figure showing the correlation between sea-ice thickness and sea-ice concentration amplitude, which supports the assertion that thinner sea ice is more sensitive than thicker sea ice.

## **Major comment 4**

We thought about the discrepancy further and came to the conclusion that there are many differences between our work and Libera et al. (2022) that actually prevent a direct comparison.

First, Libera et al. (2022) looked at sea ice area and we looked at RMSE of sea ice concentration anomalies. This means that their results are insensitive to the distribution of sea ice within the Weddell Sea region, while ours do take it into account.

Second, they use lagged autocorrelation as the definition of predictability – essentially measuring persistence – while we are analysing predictions from a fully coupled model.

We decided to change our wording to make the distinction clearer.

## **Major comment 5**

You are absolutely right. We added a figure and a paragraph highlighting the key result.

## **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.

## Minor comments

1. We clarified that that result is relevant for the Arctic.
2. We changed the section titles.
3. It is customary 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). Some references use this convention Morioka et al. (2022).
4. We changed the column order.
5. We added units to that and other figures.
6. As with warm/cold colours used to represent sea-ice anomalies, different publications adopt different conventions of how to spell “sea ice”. According to [The Cryosphere Observing System blog](#) “Strictly speaking, it should be spelled without hyphenation when it’s a noun, and hyphenated when it is an adjective”. This is supported by [Merriam-Webster’s article on compound nouns](#):

For example, when a compound that is usually left open is used to modify another noun, that compound will usually take a hyphen. So you can speak of a person who loves science fiction going to a science-fiction convention.

As with almost everything in the English language, this is more a guideline than a rule. In the case of “sea ice”, Google Ngram Viewer shows that although “sea ice extent” is more frequent than “sea-ice extent”, both phrases are in common use (Figure 1).

In the scientific literature, both terms can be found, even in some of the references in this manuscript. (e.g. “Impact of observed sea-ice concentration on the Southern Hemisphere extratropical atmospheric circulation in summer” vs. “Decadal Sea Ice Prediction in the West Antarctic Seas with Ocean and Sea Ice Initializations”). Amusingly, we find [a review comment](#) on this very journal where the authors originally used the open compound version in the manuscript and the reviewer suggested to use the hyphenated form.

We chose to follow Merriam-Webster’s guidelines and 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...”).

7. We corrected the caption
8. We removed this panel in the revised version.
9. We moved Conclusions to its own section instead of a subsection.

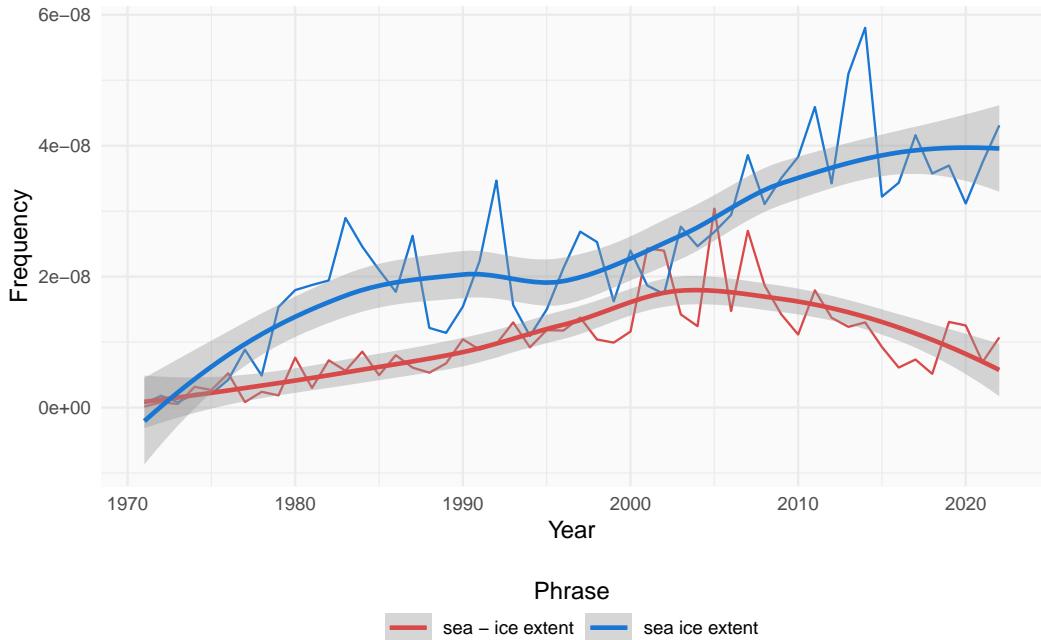


Figure 1: Frequency of the phrases “sea-ice extent” and “sea ice extent” from 1970 according to Google’s Ngram viewer.

## 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, Steph, Will Hobbs, Andreas Klocke, 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 Barthélémy, 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>.
- 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.
- Waters, Jennifer, Daniel J. Lea, Matthew J. Martin, Isabelle Mirouze, Anthony Weaver, and James While. 2015. “Implementing a Variational Data Assimilation System in an Operational 1/4 Degree Global Ocean Model.” *Quarterly Journal of the Royal Meteorological Society* 141 (687): 333–49. <https://doi.org/10.1002/qj.2388>.
- 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>.