Response to comments by Reviewer #1

We would like to thank the reviewer for the helpful comments on the manuscript. Please find below our responses to the comments.

The authors have addressed my previous comments. The ms is clearly improved. I still have a few editorial comments.

1、 In the comparison of predictive skill between ANTSIC-UNet and HIS-V, how does ANTSIC-UNet perform in terms of ACC?

Thanks for this question and concern about the ACC of DL models. Figure R1 shows the ACC of HIS-V trained by historical data without incorporating the future 6 months linear trend predictions of SIC, and the difference in ACC between HIS-V and ANTSIC-UNet. For the Pan-Antarctic, ANTSIC-UNet shows higher ACC from February to July and October to December at short lead times, and lower ACC as lead time increases, with contributions from all five sectors. Specifically, lower ACC is found in the Weddell Sea, Indian Ocean, and Pacific Ocean from December to April as the lead time exceeds 3 months. Higher ACC is observed in the Ross Sea from January to March, and the Amundsen and Bellingshausen Seas show a broad coverage of relatively high ACC. Additionally, the Pacific Ocean consistently exhibits higher ACC from July to September across all lead times. These differences in the interannual variability of SIE anomalies may be linked to the different inherent sea ice trends in these regions. For instance, the Indian Ocean experiences significant interannual fluctuations, with total sea ice area reaching its maximum in October 2010, followed by a decline to a record low in 2016, and subsequent recovery. Therefore, incorporating the linear trend prediction of SIC may reduce the predictive performance of the deep learning model in most seasons of the Indian Ocean. Furthermore, when incorporating the linear trend predictions of SIC and considering the interactions between sea ice and other climate variables, the ANTSIC-UNet shows improved skill in capturing the interannual variability of SIE anomalies throughout the year in the Ross Sea, Amundsen and Bellingshausen Seas, and during summer in the Pacific Ocean.

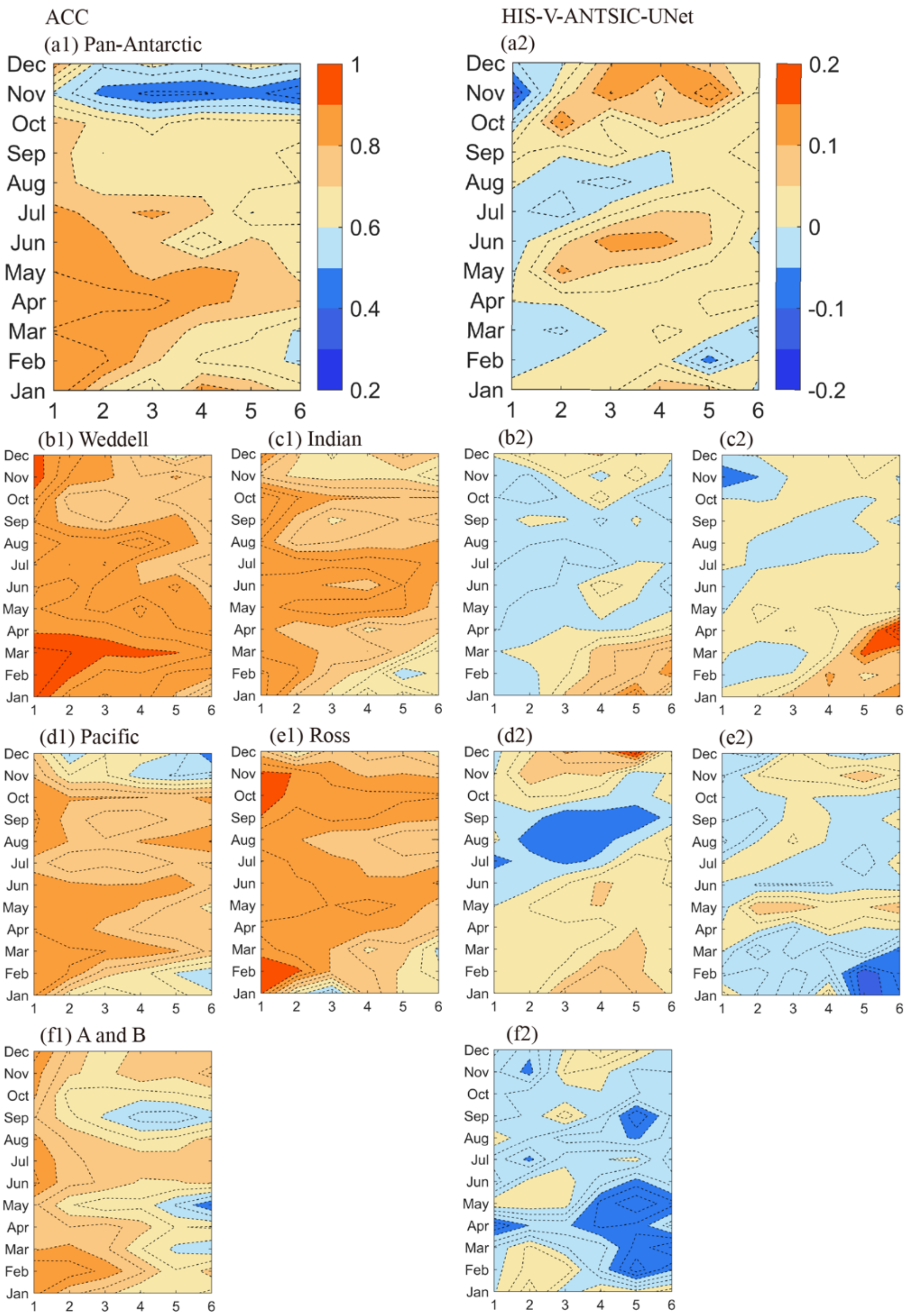

Figure R1. The ACC (a1-f1) between the observed and HIS-V (DL model trained by historical data, without incorporating the future 6 months linear trend predictions of sea ice concentration) predicted regional SIE anomalies for different target months and forecast lead times during 1981-2023. (a2-f2) as (a1-f1) but for the ACC difference between HIS-V and ANTSIC-UNet.

2、 The authors use the permutation feature-importance method to explain model variance, which is primarily based on the distance between predictions and observations. Given this, are the variable importances consistent with RMSE in terms of the prediction of the SIC variability?

Yes, the permutation feature importance is consistent with the RMSE of SIC spatial variability. We quantify the importance of each variable by calculating the change in the model evaluation metric (RMSE between the predicted SIC by the trained model and observed SIC) before and after permuting the particular variable.

Response to comments by Reviewer #2

We would like to thank the reviewer for the helpful comments on the paper. Please find below our responses to the comments.

This is a well written paper exploring the development and application of a convolutional neural network (CNN), ANTSIC-UNet, for seasonal predictions of Antarctic sea ice concentration (SIC). The paper demonstrates how ANTSIC-UNet outperforms two benchmark models, as well as the SEAS5 numerical sea ice forecasting model. This paper also explores variable importance via the use of the explainable AI tool, permute and predict. I recommend this paper for publication subject to the major revisions outlined below.

Whilst the research presented here is of high quality, the abstract, introduction and discussion need to emphasise the novelty brought by this paper. This is currently not clear to the reader. The introduction highlights that fewer studies have predicted SIC in the Antarctic compared with the Arctic. Although the application of sea ice forecasts to the Antarctic provides some novelty, greater clarification of the methodological novelty provided by this study is also required. For example, previous studies have already applied CNNs for sea ice forecasting, undertaken analysis of feature importance and compared ML model outputs to SEAS5. This clarification of methodological novelty will make it easier for the reader to follow the paper. Linked to this point, whilst it is true that far fewer papers have forecast Antarctic SIC, some key publications are missing. Please cite these and contextualise the findings of this paper to these manuscripts:

Dong, X., Yang, Q., Nie, Y., Zampieri, L., Wang, J., Liu, J. and Chen, D., 2024. Antarctic Sea Ice Prediction with A Convolutional Long Short-Term Memory Network. Ocean Modelling, p.102386.

Lin, Y., Yang, Q., Li, X., Dong, X., Luo, H., Nie, Y., Wang, J., Wang, Y. and Min, C., 2025. Ice-kNN-South: A lightweight machine learning model for Antarctic sea ice prediction. Journal of Geophysical Research: Machine Learning and Computation, 2(1), p.e2024JH000433.

Wang, Y., Yuan, X., Ren, Y., Bushuk, M., Shu, Q., Li, C. and Li, X., 2023. Subseasonal prediction of regional Antarctic sea ice by a deep learning model. Geophysical Research Letters, 50(17), p.e2023GL104347.

Thank you for your comment. We have revised the abstract, introduction and discussion to more clearly emphasize the novelty and contributions of our study. In the abstract and discussion, we have clarified the distinctions and advantages of our deep learning model compared to previous studies on Antarctic sea ice prediction. Specifically, we address key challenges in both deep learning models and dynamical models, particularly their limited representation of air-ice-sea interactions and lack of interpretability, by training our deep learning model (ANTSIC-UNet) using multiple climate variables. In addition, we explore the relative importance of these variables using the permutation feature importance approach to enhance the interpretability of our model. Moreover, we place significant emphasis on the model's performance for extended seasonal predictions (i.e., longer lead times) and conduct a systematic evaluation during extreme sea-ice years, which have both received little attention in previous studies.

In the introduction, we have added the recent key publications on Antarctic SIC prediction, as highlighted by the reviewer, to provide a comprehensive context for our work. For example, we revised the introduction as follows:

Original: "Recently, Wang et al. (2023) developed a SIPNet model with encoder-decoder structure for subseasonal Antarctic sea ice concentration prediction, which outperforms some dynamical models and advanced linear statistical models. Nevertheless, these DL methods were trained by pure historical sea ice concentration data without considering underlying physical processes governing the variation of Antarctic sea ice."

Revised: "Recently, Wang et al. (2023) developed a SIPNet model with encoder-decoder structure for subseasonal Antarctic sea ice concentration prediction, which outperforms some dynamical models and advanced linear statistical models at lead times of 1-8 weeks. Dong et al. (2024) employed a convolutional long short-term memory (ConvLSTM) network to predict Antarctic SIC up to 60 days ahead, which shows skillful predictions within 30 days and accurately forecasts annual maximum and minimum sea ice extents from 2017 to 2022. However, ConvLSTM demands significant computational resources during training, and relies on iterative forecasting which leads to error accumulation over time and requires a trade-off between accuracy and prediction length. Lin et al. (2025) proposed Ice-KNN-South, a lightweight machine learning model for predicting daily Antarctic SIC at lead times of 1-90 days. While these studies have made significant contributions, they primarily rely on historical

SIC data without considering underlying physical processes governing the variation of Antarctic sea ice. Furthermore, they focus on shorter prediction horizons, and their skillfulness in extended seasonal forecasting remains unknown."

As a general point, I also believe this paper requires some restructuring. The comparison of ANTSIC-UNet to SEAS5 is not mentioned until the discussion section on line 395. The use of SEAS5 requires mentioning in the introduction, methods, and results. Further, it is not common for the discussion section to provide new results and figures. I suggest Figures 10, 11 and 12, alongside the supporting text and equations describing these results, are moved to the results section. This will allow the discussion section to focus on the relevance and contextualisation of the results, making the paper easier to follow for the reader.

Thank you for your comment. We agree that reorganizing the paper would improve its clarity and flow. In response to the comment, we have moved the introduction of SEAS5 to the methods section. We now include the comparison of ANTSIC-UNet with the statistical models (a linear trend model and an anomaly persistence model) and a dynamical model (SEAS5) in the results section. Additionally, we have added a new Section 3.5: "Physical constraints" to describe the results with the incorporation of physical constraints into ANTSIC-UNet. Finally, we have revised the abstract and discussion sections to align with these changes. We hope this restructuring addresses the reviewer's concern.

Overall, the paper reads very well with very few typographical errors, I suggest these further minor corrections:
Line 77 – 79: please provide detail on the algorithm used to convert from passive microwave brightness temperatures to sea ice concentration values.

Thank you for your comment. The monthly SIC data are derived using the Bootstrap algorithm, which utilizes brightness temperature observations from the 37H, 37V, and 19V channels to estimate sea ice concentration (Comiso et al., 1997; Comiso and Nishio, 2008). We modified the sentence as follows:

Original: "In this study, monthly Antarctic sea ice concentration (SIC) data obtained from the National Snow and Ice Data Center (NSIDC) (https://nsidc.org/data/nsidc-0079/versions/3) are used as the input of ANTSIC-UNet, and are derived from brightness temperature of the

Scanning Multichannel Microwave Radiometer (SMMR), the Special Sensor Microwave/Imager (SSM/I) sensors, and the Special Sensor Microwave Imager/Sounder (SSMIS)."

Revised: "In this study, monthly Antarctic sea ice concentration (SIC) data obtained from the National Snow and Ice Data Center (NSIDC) (https://nsidc.org/data/nsidc-0079/versions/3) are used as the input of ANTSIC-UNet, and are derived from brightness temperature of the Scanning Multichannel Microwave Radiometer (SMMR), the Special Sensor Microwave/Imager (SSM/I) sensors, and the Special Sensor Microwave Imager/Sounder (SSMIS). SIC is retrieved using the Bootstrap algorithm, which utilizes brightness temperature observations from the 37H, 37V, and 19V channels to estimate sea ice concentration (Comiso et al., 1997; Comiso and Nishio, 2008)."

Reference:

Comiso, J. C., Cavalieri, D. J., Parkinson, C. L., and Gloersen, P.: Passive microwave algorithms for sea ice concentration: A comparison of two techniques, Remote Sensing of Environment, 60, 357–384, https://doi.org/10.1016/S0034-4257(96)00220-9, 1997.

Comiso, J. C. and Nishio, F.: Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data, Journal of Geophysical Research: Oceans, 113, https://doi.org/10.1029/2007JC004257, 2008.

Section 2.2. Please justify the use of a CNN with UNet architecture. Some recent papers have shown generative models or other AI approaches to outperform UNets. Were other ML algorithms and architectures considered?

Thank you for your comment. The primary goal of this study was to explore the feasibility of using complex climate variables to predict Antarctic sea ice concentration (SIC) and to investigate the interpretability of deep learning models. The fully convolutional neural network (FCN) based on the U-Net architecture, known for its simplicity and effectiveness in handling spatial data, was chosen as a useful tool to achieve this objective. Our results show that ANTSIC-UNet based on a relatively simple U-Net architecture, can effectively capture the complex relationships between climate variables and sea ice dynamics, and outperform benchmark models and state-of-the-art dynamical models (e.g., SEAS5) in predicting Antarctic sea ice.

While generative models, such as Generative Adversarial Networks (GANs), have shown promise in certain applications, they often require significantly more computational resources and training time. Given the exploratory nature of this study and the need for efficient experimentation, we opted for the U-Net architecture, which strikes a balance between performance and computational efficiency. In future work, we plan to explore the use of these AI-based models to assess whether they can provide additional predictive improvements. We will also continue to investigate the interpretability of these models and their ability to incorporate physical constraints to advance our understanding of Antarctic sea ice change.

Line 81 – 84: "A linear least-squares trend was fit…." This information does not fit under the subsection 2.1, as these lines are describing a method applied to the passive microwave data, rather than the data itself. Please create a new subsection in the methods section on the benchmark models.

Thank you for your comment. We agree that the information in lines 81–84 should be moved to a more appropriate section. As suggested, we have removed these sentences and created a new subsection (Section 2.3) to describe the benchmark models, and subsequent section numbers have been updated accordingly to maintain the proper structure.

"2.3 Benchmark models

In this study, the linear trend and anomaly persistence predictions are used as benchmarks to assess the predictive skill of ANTSIC-UNet. The linear trend model involves fitting a linear least-squares trend to observed SIC over the past 30 years at each grid cell for each calendar month. This trend is then used to predict SIC values for the corresponding calendar month in the following year. Additionally, these SIC predictions from this linear trend model are also used as the input to ANTSIC-UNet.

The anomaly persistence prediction is calculated as follows:

$$\boldsymbol{SIC_{pred}(t+\tau) = SIC_{clim}(t+\tau) + SIC_{anom}(t)} \quad \textbf{(1)}$$

where $\boldsymbol{SIC_{pred}}$ is the target month predicted ice concentration at the lead time $\boldsymbol{\tau}$, $\boldsymbol{SIC_{clim}}$is the climatogy ice concentration at the target month, and $\boldsymbol{SIC_{anom}}$ is the observed ice concentration anomaly relative to the climatology at the initial time. The climatology for each month is computed for the period of the training data (1979-2011). The anomaly persistence works by preserving the deviations from the climatological anomalies and assuming these anomalies will persist into the future. For example, if a particular region currently has more sea ice than average, this positive anomaly will continue as time progresses. This statistical method has been widely used as a benchmark for predicting sea ice concentration on seasonal timescales, since sea ice conditions often change gradually rather than abruptly (Wayand et al., 2019; Bushuk et al., 2021; Niraula and Goessling, 2021). While this method is effective for short-term forecasts, its accuracy declines over longer lead times as the influence of initial anomalies weakens."

Line 95 - 105: Please refer to Table 1 when listing all the variables.

Thank you for your comment. We revised the MS to refer to Table 1 when listing all the variables.

Original: "These variables include 2m air temperature (T2), 500-hPa air temperature (T500), sea surface temperature (SST), ocean temperature (PT), ocean heat content for the upper 300m (OHC300), downwelling solar radiation (DSR), upwelling solar radiation (USR), sea level pressure (SLP), 500-hPa geopotential height (H500), 250-hPa geopotential height (H250), 10m u-component of wind (U10), 10m v-component of wind (V10), and 10-hPa zonal wind (U10hPa)."

Revised: "These variables are listed in Table 1 and include 2m air temperature (T2), 500-hPa air temperature (T500), sea surface temperature (SST), ocean temperature (PT), ocean heat content for the upper 300m (OHC300), downwelling solar radiation (DSR), upwelling solar radiation (USR), sea level pressure (SLP), 500-hPa geopotential height (H500), 250-hPa geopotential height (H250), 10m u-component of wind (U10), 10m v-component of wind (V10), and 10-hPa zonal wind (U10hPa)."

Line 110: Please make clear here or somewhere else the temporal resolution of the forecasts. Is it monthly, daily, seasonal or some other resolution?

Thank you for your comment. To clarify the temporal resolution of the forecasts, we revised the sentence as follows:

Original: “The final output provides the 6-month forecast of Antarctic sea ice concentration.”

Revised: “The final output provides the 6-month forecast of monthly Antarctic sea ice concentration.”

Line 135: Please describe the hyperparameter selection and tuning process you employed.

Thank you for your comment. We added the description of the hyperparameter selection and tuning process as follows:

“Here, we use typical hyperparameters for the deep learning model. The kernel size for the convolutional layers is set to (3,3). Due to memory constraints, we set the batch size to 2. The loss function applied is the mean squared error (MSE), with a learning rate of 0.0001 and a weight decay of 0. The Adam optimizer is used for training.”

Figure 2: Please provide some background in the introduction on how and why the Southern Ocean is split into these five regions.

Thank you for your comment. We added more details in the introduction regarding the division of the Southern Ocean into five sectors as follows:

“Sea ice in different regions exhibits complex spatial patterns of change in growth, retreat, and duration (Liang et al., 2023). The Southern Ocean sea ice region is divided into five sectors: the Weddell Sea, Indian Ocean, Pacific Ocean, Amundsen and Bellingshausen Seas, and Ross Sea. These regions are characterised by their unique climatic, oceanographic, and geographical characteristics (Zwally et al., 2002; Grieger et al., 2018; Josey et al., 2024). This division has been widely used in studying the regional dynamics and prediction of Antarctic sea ice (e.g., Eayrs et al., 2019; Bushuk et al., 2021; Liang et al., 2023).”

Reference:

Bushuk, M., Winton, M., Haumann, F. A., Delworth, T., Lu, F., Zhang, Y., Jia, L., Zhang, L., Cooke, W., Harrison, M., Hurlin, B., Johnson, N. C., Kapnick, S. B., McHugh, C., Murakami, H., Rosati, A., Tseng, K.-C., Wittenberg, A. T., Yang, X., and Zeng, F.: Seasonal Prediction and

Predictability of Regional Antarctic Sea Ice, Journal of Climate, 34, 6207–6233, https://doi.org/10.1175/JCLI-D-20-0965.1, 2021.

Eayrs, C., Holland, D., Francis, D., Wagner, T., Kumar, R., and Li, X.: Understanding the Seasonal Cycle of Antarctic Sea Ice Extent in the Context of Longer-Term Variability, Reviews of Geophysics, 57, 1037–1064, https://doi.org/10.1029/2018RG000631, 2019.

Grieger, J., Leckebusch, G. C., Raible, C. C., Rudeva, I., and Simmonds, I.: Subantarctic cyclones identified by 14 tracking methods, and their role for moisture transports into the continent, Tellus A: Dynamic Meteorology and Oceanography, 70, 2018.

Liang, K., Wang, J., Luo, H., and Yang, Q.: The Role of Atmospheric Rivers in Antarctic Sea Ice Variations, Geophysical Research Letters, 50, e2022GL102588, https://doi.org/10.1029/2022GL102588, 2023.

Zwally, H. J., Comiso, J. C., Parkinson, C. L., Cavalieri, D. J., and Gloersen, P.: Variability of Antarctic sea ice 1979–1998, Journal of Geophysical Research: Oceans, 107, 9-1-9–19, https://doi.org/10.1029/2000JC000733, 2002.

Figure 2 caption: typo: "based on the same calendat month".

The typo has been corrected.

Figure 3 a) and e), please flip the colour ramp so white is ice and water is blue, or use a separate colour ramp altogether. Same for Figure 7.

Thank you for your comment. We modified the figures and used a separate colour ramp. Additionally, we included the SEAS5 predictions for comparison.

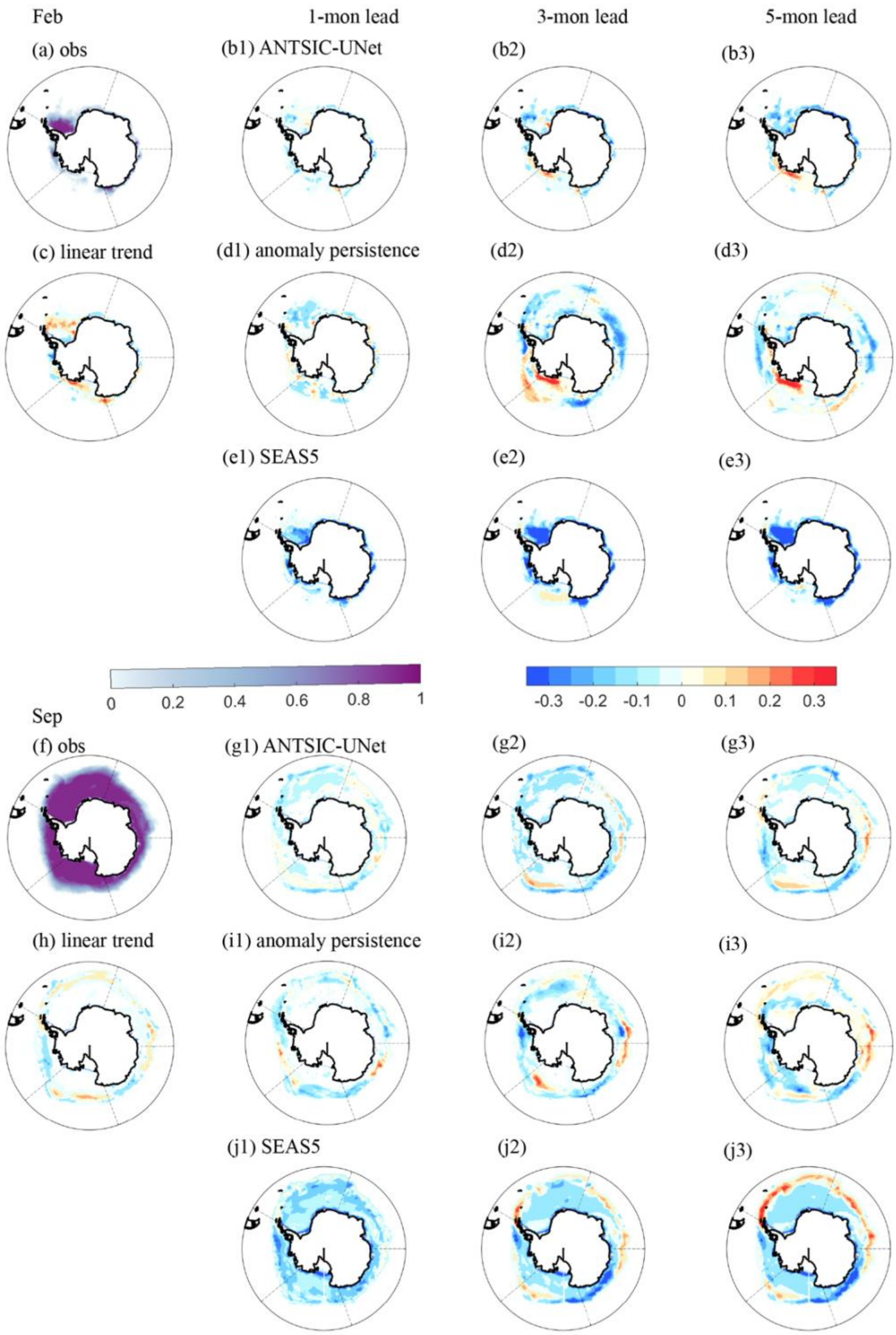

Figure R2. The monthly mean sea ice concentration of the NSIDC observations for (a) February and (f) September, and the errors in predicting by ANTSIC-UNet (b1-b3, g1-g3), the linear trend model (c and h), anomaly persistence model (d1-d3, i1-i3) and SEAS5 (e1-e3, j1-j3) at lead time of 1, 3, and 5 months for February (upper panel) and September (lower panel) during the testing years.

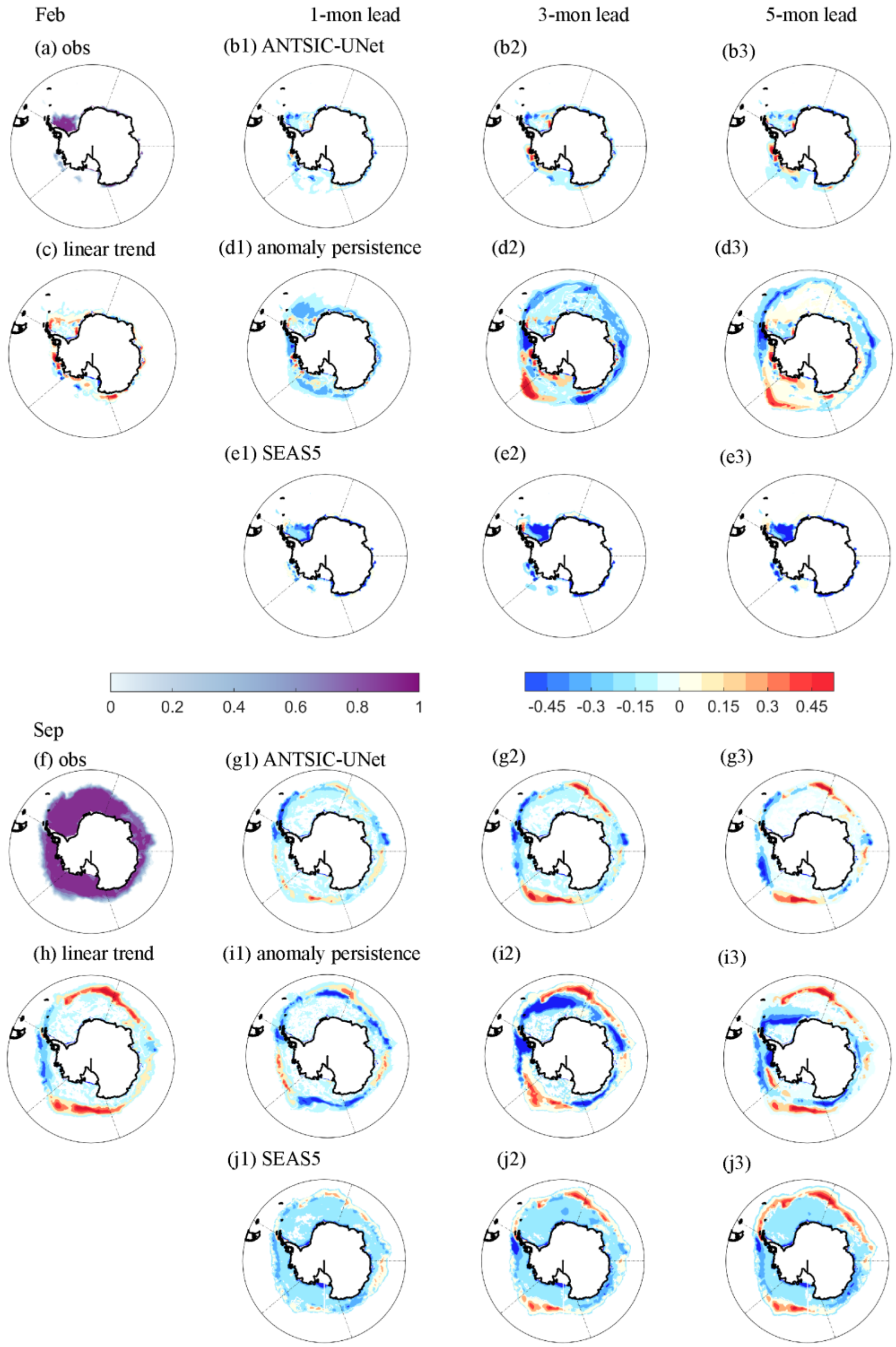

Figure R3. February and September 2023 SIC of NSIDC observations (a, e) and errors predicted by ANTSIC-UNet (b1-b3, g1-g3), the linear trend model (c and h), anomaly persistence model (d1-d3, i1-i3) and SEAS5 (e1-e3, j1-j3) at lead time of 1, 3 and 5 months (lowest sea ice extent on record).

Figure 3: Please make clear whether these are February and September means for a particular year or the whole date range.

Thank you for your comment. We indicated in the figure caption that the data represent the mean for February and September for the testing years.

Figure 4 f1 and f2: please make clear that A and B stand for Amundsen and Bellingshausen.

Thank you for your comment. We modified the caption of Figure 4 as follows:

Original: "Figure 4. The predictive skill of sea ice concentration (spatially and temporally averaged during the testing years) in terms of RMSE and IIEE (units: million square kilometers) between the ANTSIC-UNet predictions and NSIDC observations for different target months and forecast lead times."

Revised: "Figure 4. The predictive skill of sea ice concentration (spatially and temporally averaged during the testing years) in terms of RMSE and IIEE (units: million square kilometers) between the ANTSIC-UNet predictions and NSIDC observations for different target months and forecast lead times. 'A and B' in (f1) and (f2) refer to the Amundsen Sea and Bellingshausen Seas, respectively."

Discussion section: Due to the large similarities between this paper and the IceNet model published in Andersson et al. (2021), please provide detailed discussion on the relative performance of ANTSIC-UNet and IceNet.

Thank you for your comment. Although ANTSIC-UNet and the IceNet model proposed by Andersson et al. (2021) have similarities in their underlying U-Net architecture, the two models differ in design, objectives, and application domains, making direct comparisons difficult.

IceNet was designed for Arctic sea ice classification, aiming to predict three discrete SIC categories: open water (SIC<=15%), marginal ice (15%<SIC<80%), and full ice (SIC>=80%). In contrast, ANTSIC-UNet is developed for Antarctic sea ice concentration (SIC) regression prediction. This difference in task leads to different loss functions being used during training: classification models like IceNet use categorical loss functions (e.g., cross-entropy), while

ANTSIC-UNet employs regression-based loss functions (e.g., mean squared error) to predict continuous SIC values.

Moreover, the Arctic and Antarctic have different geographical features, which lead to major differences in oceanic and atmospheric circulation patterns (Maksym, 2019). As a result, the sea ice in the Antarctic and Arctic shows completely different trends and behaviors. The Antarctic and Arctic also experience extreme sea ice events in different ways, which are driven by different atmospheric and oceanic factors.

Deep learning models such as IceNet and ANTSIC-UNet exhibit strong nonlinear learning capabilities, which are particularly valuable in predicting extreme events that deviate from climatological norms. Figure R4 in Andersson et al. (2021) shows IceNet's predictive skill for seasonal September forecasts in the Arctic. The 2012–2020 period contains three anomalous September Arctic SIEs: 2012 (lowest extent on record), 2013 (anomalously high extent), and 2020 (second lowest extent on record). IceNet shows skillful predictions for these extreme events, outperforming the linear trend predictions and SEAS5, except in September 2013, when its error slightly exceeded that of SEAS5 at 2-3 months lead. In our study, we placed particular emphasis on evaluating model performance during three extreme summer sea ice events in the Antarctic (2017, 2022 and 2023). As shown in Figure R5, ANTSIC-UNet outperforms SEAS5 and linear trend predictions for sea ice edge error in all extreme summer years. Therefore, despite bring designed for different hemispheres and sea ice prediction tasks, both IceNet and ANTSIC-UNet highlight the strength of deep learning models in capturing nonlinear changes, particularly in extreme sea ice years.

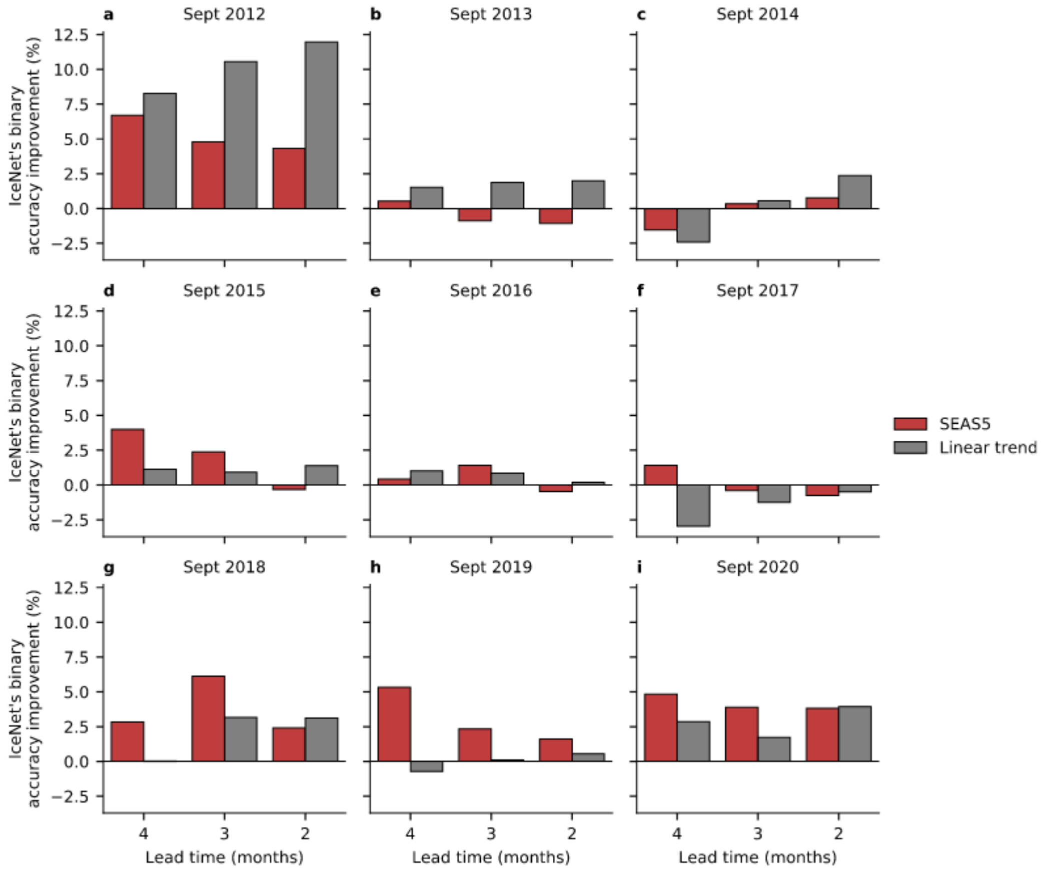

Figure R4. Comparing IceNet with SEAS5 and the linear trend for seasonal September forecasts. a–i IceNet's improvement in binary accuracy relative to SEAS5 and the linear trend models for September forecasts at 4- to 2-month lead times for the validation and test years (2012–2020) (From Andersson et al., 2020; Figure 4).

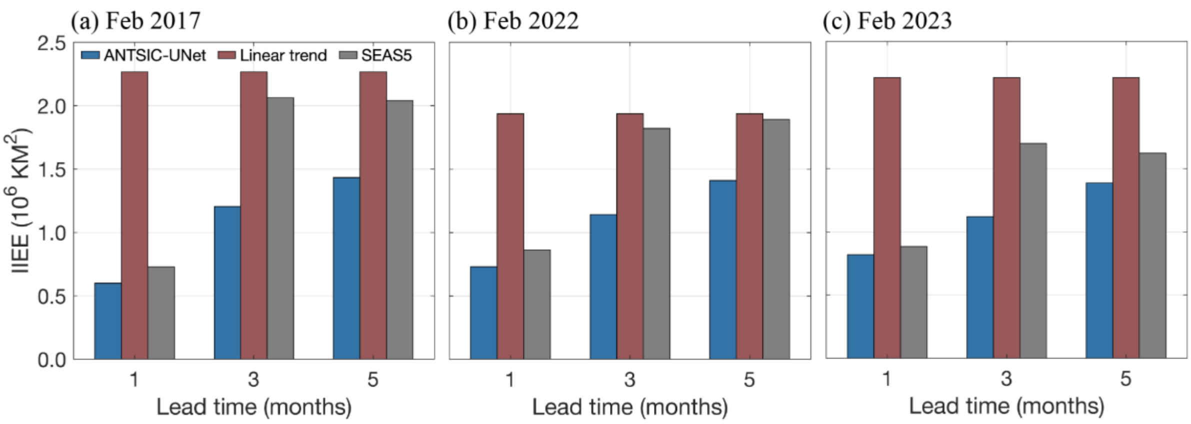

Figure R5. Integrated ice-edge error (IIEE) of ANTSIC-UNet, the linear trend forecast and SEAS5 for February forecasts at lead time of 1, 3, and 5 months for the extreme summer years. (a) 2017, (b) 2022 and (c) 2023.

Reference:

Maksym, T.: Arctic and Antarctic Sea Ice Change: Contrasts, Commonalities, and Causes, Annual Review of Marine Science, 11, 187–213, https://doi.org/10.1146/annurev-marine-010816-060610, 2019.

Figure 6 line 262-263: "ANTSIC-UNet (anomaly persistent model) at different lead times up to 6 months…." How are different lead times represented in this figure?

Thank you for your comment. In Figure 6, the different lead times are represented on the x-axis, which ranges from 1 to 6 months, corresponding to the lead times up to 6 months as mentioned in the caption of Figure 6.

Section 3.3: As a general discussion point, what is the suitability of using the anomaly persistence model as a benchmark model for forecasting extreme events? Isn't it always destined to underrepresent these extremely anomalous cases? Are there more appropriate benchmark models that could be used for these circumstances?

Thank you for your comment. For shorter lead times, the anomaly persistence model does not always underestimate such cases due to the inherent characteristics of sea ice. The persistence of sea ice anomalies can often continue in the short term, errors tend to become more significant because the anomaly persistence model fails to capture the longer-term variability and more complex interactions of extreme events as lead time increases. Using more complex benchmark models, such as multiple linear regression and random forest, which can incorporate additional predictors (e.g., atmospheric and oceanic variables) and partially capture the nonlinear relationship, may provide a more appropriate reference for evaluating the predictive ability of deep learning models.

Table 3 line 298: Typo "Here, Observed.." change to lower case.

The typo has been corrected.

Figure 8: Help the reader- in which month when did the extreme event(s) occur.

Thank you for your comment. We modified the caption for Figure 8 as follows:

Original: "Figure 8. Seasonality errors of the Pan- and regional Antarctic monthly mean SIE (SIC > 15%) between NSIDC observations and ANTSIC-UNet (a-f) and anomaly persistence model (g-l) predictions at different lead times for 2023 (lowest sea ice extent on record). The black lines show the seasonality SIE errors between observations and linear trend model. (units: million square kilometers)"

Revised: "Figure 8. Seasonality errors of the Pan- and regional Antarctic monthly mean SIE (SIC > 15%) between NSIDC observations and ANTSIC-UNet (a-f) and anomaly persistence model (g-l) predictions at different lead times for 2023 (lowest sea ice extent on record in February). The black lines show the seasonality SIE errors between observations and linear trend model. (units: million square kilometers)"

Section 3.4. There are some points that are more appropriate for the discussion, particularly where references are made to other papers. For example from line 320 "previous studies….."

Thank you for your comment. Regarding the sentence, "Previous studies suggested that the evaluation metrics of model's predictive skill, particularly for models with strong generalization ability, correlate closely with feature importance (FI) (Andersson et al., 2021; Molnar, 2019)." we would like to clarify that the principle behind the permutation feature importance method we use is consistent with the method described in the referenced studies. We intend to retain this content as it provides the necessary background and theoretical support, helping readers understand how this method has been applied in previous research and its relevance to our study.

However, to better contextualize our findings in relation to other studies, we expanded on this in the discussion section. For example, as the reviewer suggested, we associated our feature importance results with those of Uebbing et al. (2025).

Line 334: Typo: Circulation. (Raphael…) – remove full stop.

The typo has been corrected.

Discussion: Please contextualise your findings on feature importance for sea ice forecasting

with the following paper that also carried out a similar study:
Uebbing, L., Joakimsen, H.L., Luppino, L.T., Martinsen, I., McDonald, A., Wickstrøm, K.K., Lefèvre, S., Salberg, A.B., Hosking, S. and Jenssen, R., 2025, January. Investigating the Impact of Feature Reduction for Deep Learning-based Seasonal Sea Ice Forecasting. In Northern Lights Deep Learning Conference 2025.

Thank you for your comment. Following line 368 in the Discussion section, we have added the following to link our feature importance results with the study by Uebbing et al. (2025):

"Our feature importance findings can be associated with recent work by Uebbing et al. (2025) investigating the impact of feature reduction on seasonal Arctic sea ice forecasting by using the state-of-the-art IceNet model (Andersson et al., 2021) combined with explainable AI (XAI) techniques. Their study showed that using only a subset of key features (such as historical sea ice concentration, linear trend forecasts, and seasonal encoding), high predictive accuracy under general scenarios was still obtained. However, their research also highlighted that for extreme events, such as anomalous sea ice extents, models incorporating additional climate variables perform better. This suggests that further studies might benefit from exploring different XAI methods for estimating feature importance and investigating the extent to which the reduction of the number of features affects deep learning model predictions for Antarctic sea ice.

Discussion: Why does the performance differ between the different regions of the Southern Ocean? For example, the disparities between 4 b1 – f1. There is mention of this on lines 367 – 369, but please expand further. Also, please comment on the better predictive performance of the tool in the Austral summer.

Thank you for your comment. The differences in model performance across regions could be attributed to regional variability due to oceanographic conditions, sea ice dynamics, and the influence of atmospheric and oceanic circulation patterns. We expanded on this in discussion section. As observed in our results, ANTSIC-UNet shows better predictive performance relative to the two benchmark models and SEAS5 during the Austral summer, particularly in the sea ice edge zone. We added the further discussion of the spatial forecasting performance as follows:

"Our findings are consistent with those of Marchi et al. (2019) and Bushuk et al. (2021) that sea ice concentration prediction tends to be more accurate in the winter months but less so in the summer due to rapid and irregular changes in the ice edge during that season. Inspiringly, ANTSIC-UNet shows lower summer sea ice edge error and SIC RMSE compared to both the two benchmark models and SEAS5, especially during extreme years. The differences in model performance across regions could be attributed to regional variability due to oceanographic conditions, sea ice dynamics, and the influence of atmospheric and oceanic circulation patterns. Regional seas in the West Antarctic, including the Ross Sea, Amundsen Sea, Bellingshausen Sea, and Weddell Sea, exhibit larger interannual variability in sea ice concentration compared to the East Antarctic (Cavalieri and Parkinson, 2008). These regions are influenced by the Circumpolar Deep Water (CDW), with warm-shelf regions such as the Amundsen and Bellingshausen Seas being particularly sensitive to climate changes, with sea ice concentration and the position of the ice edge strongly driven by wind forcing (Stammerjohn et al., 2003; Saenz et al., 2023). The ice flux driven by wind in the Weddell Sea along the Antarctic Peninsula and the Pacific Ocean plays a crucial role in modulating sea ice dynamics, with the dynamical influence being more pronounced in the Pacific sector (Holland and Kwok, 2012). The sea ice increase (decrease) in the Ross Sea (Bellingshausen Sea) is linked to the Amundsen Sea Low (ASL) which is a key climate feature of these regions (Hosking et al., 2013; Turner et al., 2016). In contrast to other regions of Antarctica, sea ice expansion in the Indian Ocean sector is significant throughout all seasons and is associated with surface cooling and ocean renewal processes that stabilize the ocean and limit the intrusion of warmer subsurface waters into the surface layer (Bintanja et al., 2013; Purich et al., 2018). Additionally, seasonal variability in sea ice in the Indian Ocean sector is closely linked to the Southern Annular Mode (SAM) (Yadav et al., 2022)."

Reference:

Bintanja, R., van Oldenborgh, G. J., Drijfhout, S. S., Wouters, B., and Katsman, C. A.: Important role for ocean warming and increased ice-shelf melt in Antarctic sea-ice expansion, Nature Geosci, 6, 376–379, https://doi.org/10.1038/ngeo1767, 2013.

Cavalieri, D. J. and Parkinson, C. L.: Antarctic sea ice variability and trends, 1979–2006, Journal of Geophysical Research: Oceans, 113, https://doi.org/10.1029/2007JC004564, 2008.

Marchi, S., Fichefet, T., Goosse, H., Zunz, V., Tietsche, S., Day, J. J., and Hawkins, E.:

Reemergence of Antarctic sea ice predictability and its link to deep ocean mixing in global climate models, Climate Dynamics, 52, 2775–2797, https://doi.org/10.1007/s00382-018-4292-2, 2019.

Holland, P. R. and Kwok, R.: Wind-driven trends in Antarctic sea-ice drift, Nature Geosci, 5, 872–875, https://doi.org/10.1038/ngeo1627, 2012.

Hosking, J. S., Orr, A., Marshall, G. J., Turner, J., and Phillips, T.: The Influence of the Amundsen–Bellingshausen Seas Low on the Climate of West Antarctica and Its Representation in Coupled Climate Model Simulations, Journal of Climate, 26, 6633–6648, https://doi.org/10.1175/JCLI-D-12-00813.1, 2013.

Marchi, S., Fichefet, T., Goosse, H., Zunz, V., Tietsche, S., Day, J. J., and Hawkins, E.: Reemergence of Antarctic sea ice predictability and its link to deep ocean mixing in global climate models, Climate Dynamics, 52, 2775–2797, https://doi.org/10.1007/s00382-018-4292-2, 2019.

Purich, A., England, M. H., Cai, W., Sullivan, A., and Durack, P. J.: Impacts of Broad-Scale Surface Freshening of the Southern Ocean in a Coupled Climate Model, Journal of Climate, 31, 2613–2632, https://doi.org/10.1175/JCLI-D-17-0092.1, 2018.

Saenz, B. T., McKee, D. C., Doney, S. C., Martinson, D. G., & Stammerjohn, S. E. (2023). Influence of seasonally varying sea-ice concentration and subsurface ocean heat on sea-ice thickness and sea-ice seasonality for a 'warm-shelf' region in Antarctica. Journal of Glaciology, 69(277), 1466–1482. https://doi.org/10.1017/jog.2023.36

Stammerjohn, S. E., Drinkwater, M. R., Smith, R. C., and Liu, X.: Ice-atmosphere interactions during sea-ice advance and retreat in the western Antarctic Peninsula region, Journal of Geophysical Research: Oceans, 108, https://doi.org/10.1029/2002JC001543, 2003.

Turner, J., Hosking, J. S., Marshall, G. J., Phillips, T., and Bracegirdle, T. J.: Antarctic sea ice increase consistent with intrinsic variability of the Amundsen Sea Low, Clim Dyn, 46, 2391–2402, https://doi.org/10.1007/s00382-015-2708-9, 2016.

Yadav, J., Kumar, A., Srivastava, A., and Mohan, R.: Sea ice variability and trends in the Indian Ocean sector of Antarctica: Interaction with ENSO and SAM, Environmental Research, 212, 113481, https://doi.org/10.1016/j.envres.2022.113481, 2022.

There is no conclusion section. Please check if this is required.

Thank you for your comment. In the manuscript, we have included a section titled “Discussion and Conclusion”, which combines the interpretation of findings with the main conclusion of the study.