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.

This article introduces a deep learning model called ANTSIC-UNet for predicting the extended seasonal variations in Antarctic sea ice concentration, with the ability to forecast up to 6 months in advance. The study utilized a rich set of climate variables for model training and compared it against two benchmark models (linear trend and anomaly persistence models). The results demonstrate that ANTSIC-UNet exhibits superior predictive skills in sea ice concentration and integrated ice-edge error, especially in forecasting extreme events in recent years. The strengths of the article include the consideration of both sea ice and related atmospheric and oceanic variables enhances the accuracy of the predictions. The results are interesting and the work could be published after moderate revision. My comments are intended to improve the presentation of the paper and require clarifying unclear points.

Comments

1. *L76"57 is the dimension of the variables" However, when we calculate 12+1+14*3+1, it equals 56. So, what is the extra one?*

The dimension of 57 includes sea ice concentration for the past 12 months, the linear trend prediction of sea ice concentration for the following 6 months, 12 climate variables for the past 3 months, 2 climate variables for the past 1 month, and the land mask. Therefore, the calculation is 12 + 6 + 12*3 + 2 + 1 = 57. For the details of climate variables, please refer to Table 1, which provides the variable names along with their respective lead or lag times. We clarified this in the revision.

2. *L184 For September, compared to anomaly persistence, ANTSIC-UNet shows a larger negative bias in the sea ice edge region. What could be the possible reasons for this error?*

Thanks for your comment. The larger negative bias in the sea ice edge region in September for the ANTSIC-UNet prediction relative to the anomaly persistence as the lead time increase is due to the limited number of years used for calculating the average of sea ice errors, which only includes the testing years of 2017, 2020 to 2023 (anomalously low ice extents), and 2014 (record high). Specifically, this averaging results in large positive and negative anomalies in different years offsetting each other for the anomaly persistence prediction. To demonstatrate this, we selected three sub-regions that show larger negative bias in the sea ice edge region in September for ANTSIC-UNet at 5-month lead compared to the anomaly persistence prediction (see Figure R1), including the Weddell Sea, the Pacific Ocean, and the Amundsen and Bellingshausen Seas. Here we used the mean absolute error (MAE) as the evaluation metrics (Figure R2). ANTSIC-UNet shows smaller prediction errors in the sea ice edge across all regions compared to anomaly persistence, except for the Weddell Sea as the lead time exceeds 4 months.

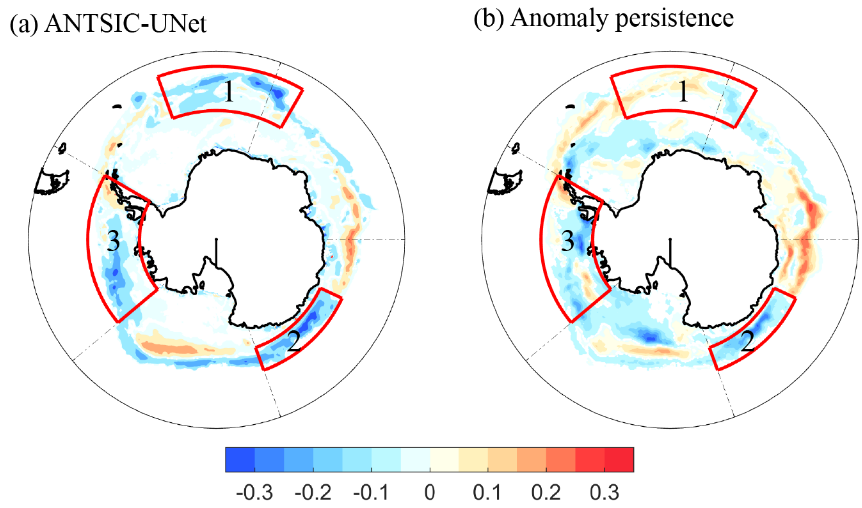

Figure R1. September mean sea ice concentration errors predicted by (a) ANTSIC-UNet and (b) anomaly persistence model at 5-month lead for the testing years. The red boxes indicate the three regions where ANTSIC-UNet shows larger negative bias compared to the anomaly persistence model: region 1 – eastern Weddell Sea (53°-63°S, 20°W-30°E), region 2 – eastern Pacific Ocean (60°-65°S, 115°-160°E) and region 3 - Amundsen and Bellingshausen Seas (62°-72°S, 130°-60°W).

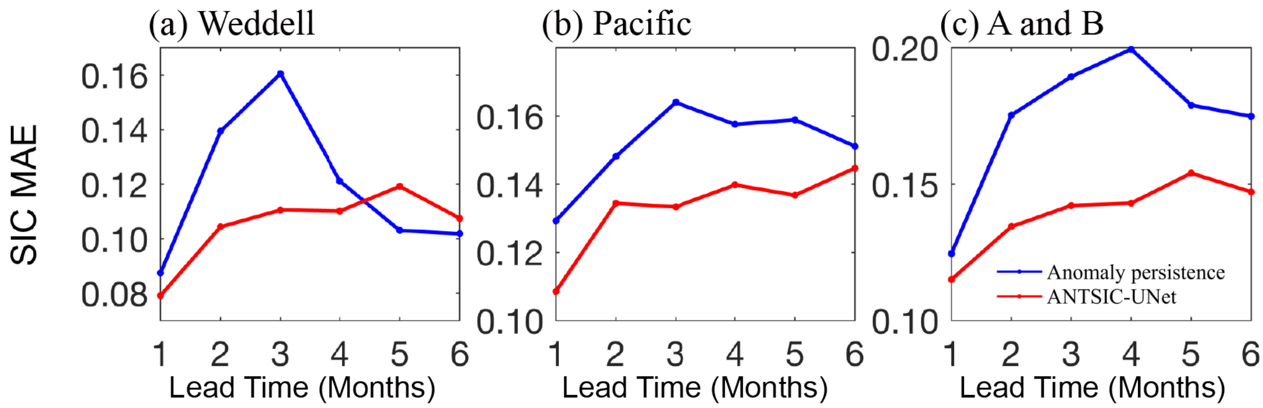

Figure R2. September sea ice concentration mean absolute error (SIC MAE) between the predictions and NSIDC observations for (a) eastern Weddell Sea, (b) eastern Pacific Ocean, and (c) Amundsen and Bellingshausen Seas for the testing years. (ANTSIC-UNet: red line; anomaly persistence model: blue line)

3. *L186 Is the lower RMSE in September compared to February related to the size of the area considered during the calculation? Are the regions used for calculating each indicator consistent with the respective months?*

Thanks for your comment. The RMSE is calculated based on the area where sea ice concentration is more than 15% in observations or predictions. The IIEE is the sum of overestimated and underestimated sea ice extent where sea ice concentration is more than 15%.

The area size varies in different months. Both RMSE and IIEE with the respective months are measured in units of area. Figure R3 shows the percentage of the sea ice edge error relative to the actual sea ice extent. In February, although the Antarctic sea ice extent reaches its seasonal minimum, the relative percentage of the sea ice edge error is large and increases as lead times increase. In September, the Antarctic experiences extensive sea ice coverage, but the relative percentage is smaller for all lead months, resulting in the overall low RMSE.

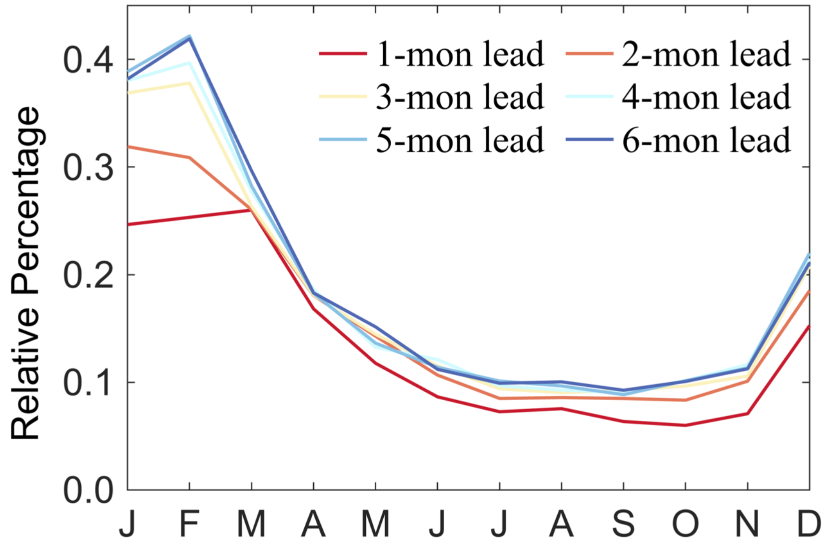

Figure R3. Percentage of the sea ice edge error relative to the actual sea ice extent (where sea ice concentration more than 15% in observations or predictions).

4. *L274 Is the high importance of variables in the* model due to the seasonal cycle? Does the importance of variables change for SIC anomaly?

Thanks for your comment. The importance of relevant climate variables is independent of the seasonal cycle. All non-SIC variables, were converted to anomalies (by subtracting the climatological mean for each calendar month during 1979-2011) before being input into ANTSIC-UNet. SIC has a pronounced seasonal cycle, which serves as an important reference for predicting future changes.

Variable importance changes across different seasons in the context of SIC anomaly. For example, Table R1 (see below) gives the variable importance ranking for the target months of January and June at 1-month lead. For January, ANTSIC-UNet relies mostly on the upwelling solar radiation and 10-hPa zonal wind in the stratosphere. For June, sea surface temperature and initial sea ice state are more important. Additionally, the linear trend predictions of SIC at the target month are important for both months though it ranks as the third.

Table R1. Variable importance ranking for the target months of January and June at 1-month lead averaged for the testing years 2020-2023.

Rank	(a) For Jan forecasts	(b) For Jun forecasts
1	Dec USRA (0.90%)	May SSTA (0.79%)
2	Dec U10hPaA (0.55%)	May SIC (0.46%)
3	Jan SIC trend(0.46%)	Jun SIC trend (0.35%)

5. *The main improvement of this article compared to other DL methods is the inclusion of relevant variables that affect sea ice in the training data of the model. How significant is the impact of these variables compared to a model trained solely using historical data?*

Thanks for this question and concern about the improvement made by incorporating relevant variables for training. We did compare the performance of DL models trained by three sets of input variables (Table R2).

As shown in Table R2, compared to HIS-V which is trained by historical data without incorporating the future 6 months linear trend predictions of sea ice concentration, ANTSIC-UNet shows relatively reduced RMSE and notable improvement in IIEE during all testing years and extreme years. This suggests that incorporating future sea ice trends enhances the deep-learning model's predictive accuracy, particularly at the sea ice edge.

Furthermore, compared to SIC-V, which is trained by only sea ice data, including the future 6 months linear trend predictions and past 12 months of historical sea ice concentration, Both ANTSIC-UNet and HIS-V show significant improvement of IIEE, which indicates that using enriched climate variables as inputs allows ANTSIC-UNet to effectively capture the complex nonlinear relationships in air-ice-sea interactions and enhance the predictive skill for Antarctic sea ice concentration.

Table R2. The averaged predictive skill of ANTSIC-UNet (the original DL model trained by 57 variables, see Table 1 in the manuscript for the details of all input variables), HIS-V (DL model trained by historical data, without incorporating the future 6 months linear trend predictions of sea ice concentration), and SIC-V (DL model trained by pure SIC data, including the future 6 months linear trend predictions of sea ice concentration and past 12 months of historical sea ice concentration). (RMSE: root-mean-square error; IIEE: integrated ice-edge error.)

		ANTSIC-UNet	HIS-V	SIC-V
All testing years	RMSE	0.21	0.22	0.22
	IIEE	1.68	1.75	1.95
2017	RMSE	0.21	0.22	0.22
	IIEE	1.80	1.92	2.27
2022	RMSE	0.21	0.22	0.22
	IIEE	1.68	1.77	1.98
2023	RMSE	0.24	0.25	0.24
	IIEE	1.99	2.07	2.57

6. *The section on the importance of each variable is very insightful. The author presents some viewpoints that are inconsistent with statistical models, such as the minimal impact of variables like temperature and wind speed in DL methods. Does this suggest that DL methods have not learned the underlying mechanisms of these variables to some extent?*

Thank you for your comment. Our study showed that ANTSIC-UNet had been trained to learn the nonlinear and indirect relationships among climate variables that contribute to improved accuracy of Antarctic sea ice prediction. The variable importance results from ANTSIC-UNet are generally consistent with known causal links between climate variables and sea ice, suggesting that physically plausible statistical relationships have been learned. For example, sea surface temperature and air temperature play a crucial role in Antarctic sea ice predictions at 1-2 month lead, influencing sea ice formation and melting through thermodynamic processes. 10m meridional wind is also important at short lead times, affecting sea ice variation through sea ice advection, air-sea heat flux, and ocean mixing. As the lead time increases, the influence of these variables tends to be reduced, and the 10-hPa zonal wind in the stratosphere becomes more important. This is consistent with previous studies showing that the changes in stratospheric zonal circulation affect sea ice variability by influencing the circumpolar westerly winds in the troposphere through downward propagation (Wang et al., 2019; Cordero et al., 2023). The relatively not very significant importance of tropospheric variables (i.e., H500A) may be related to the inherent structure of the deep learning model that still has not learned all underlying mechanisms, which requires further investigation in future research.

When a variable shows small or even negative importance, as Andersson et al. (2021) suggested the DL model might be overlooking that feature or has not yet fully captured the intrinsic relationships involving that variable.

Reference:
Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., Byrne, J., Tietsche, S., Sarojini, B. B., Blanchard-Wrigglesworth, E., Aksenov, Y., Downie, R., and Shuckburgh, E.: Seasonal Arctic sea ice forecasting with probabilistic deep learning, Nature Communications, 12, 5124, https://doi.org/10.1038/s41467-021-25257-4, 2021.

Cordero, R. R., Feron, S., Damiani, A., Llanillo, P. J., Carrasco, J., Khan, A. L., Bintanja, R., Ouyang, Z., and Casassa, G.: Signature of the stratosphere–troposphere coupling on recent record-breaking Antarctic sea-ice anomalies, The Cryosphere, 17, 4995–5006, https://doi.org/10.5194/tc-17-4995-2023, 2023.

Wang, G., Hendon, H. H., Arblaster, J. M., Lim, E.-P., Abhik, S., and van Rensch, P.: Compounding tropical and stratospheric forcing of the record low Antarctic sea-ice in 2016, Nat Commun, 10, 13, https://doi.org/10.1038/s41467-018-07689-7, 2019.