Responses to Reviewer's Comments for Manuscript egusphere-2024-4092

# Improving Seasonal Arctic Sea Ice Predictions with the Combination of Machine Learning and Earth System Model

Addressed Comments for Publication to

The cryosphere

by

Zikang He, Yiguo Wang, Julien Brajard, Xidong Wang, Zheqi Shen

### Authors' Response to Reviewer #1

**Comment 1**

The authors have improved the seasonal prediction skill of Arctic sea ice in the NorCPM model using a machine learning method. Specifically, they experimented with two approaches: (a) an online method, which involves correcting the model's initial fields after each time step, and (b) an offline method, where the model runs freely, and the results are uniformly corrected afterward. Both methods are based on simple concepts but have been proven to be effective.
I acknowledge that the approach of integrating machine learning with dynamical models is quite advanced and interesting; however, I have some concerns regarding the broader scientific implications of this study. Please see the following general comments.

**Response:** We very much appreciate that the reviewer found this study advanced and interesting. We thank the reviewer for providing insightful comments that helped significantly improve the manuscript. We carefully addressed each concern and revised the manuscript. Below, we provide our detailed point-by-point responses to the reviewer's comments. To enhance the legibility of this response letter, the reviewer's comments are typeset in blue boxes. Rephrased or added statements in the revised version of the manuscript are indicated in gray boxes.

### General Comments

**Comment 2**

The discussion section is currently somewhat underdeveloped. I would suggest that the authors discuss their results from the following perspectives:
a) How do the findings of this study inform physical enhancements to predictive systems? For instance, which sea ice processes benefit from online correction of sea ice concentration? And thus improved prediction skills?

**Response:** We thank the reviewer for the suggestion. In this study, all sea ice errors associated with thermodynamic and dynamic processes were incorporated into the error correction process. Still, we did not explicitly attribute sea ice errors to specific physical processes. Specifically, the online correction method simultaneously adjusts sea ice concentration (SIC), sea surface temperature (SST), and sea surface salinity (SSS). While updating SST primarily addresses thermodynamic errors, changing SST and SSS would change the water density and influence the ocean circulation.

For clarity, we have revised one paragraph in the discussion and conclusion section as follows (L406-L412 in the manuscript):

> It is important to note that the proposed approaches still have room for improvement. In this study, we only use ocean and sea ice variables as input features. Including atmospheric variables would help to address errors due to both dynamic and thermodynamic processes and further improve the performance. Increasing the frequency of online correction could help enhance its effectiveness (Gregory et al., 2024), but this is challenging in practice since analysis increments in NorCPM are currently available only every month. An alternative strategy is to train hybrid models that combine ML with dynamical models, which has been shown to be effective in other systems (Farchi et al., 2021). However, this approach relies on external constraints to compute the gradient of the dynamical model, which are not available in NorCPM.

**Comment 3**

b) As a hybrid approach integrating dynamical models with machine learning, what advantages does this study offer compared to purely data-driven machine learning methods (e.g., Andersson et al., 2021; Ren et al., 2024; Kim et al., 2025)? For instance, does it demonstrate enhanced scalability, such as in applications to ice thickness prediction corrections?

**Response:** The primary objective of the approaches used in this study is to reduce the intrinsic prediction errors of the dynamical model itself. The dynamical model can provide predictions for a large number of variables and ensures physical consistency among the variables. On the other hand, purely data-driven models, typically relying on predictors from observations or reanalysis data, aim for certain specific variables and often lack physical consistency among these variables.
As suggested, we have added one paragraph to the discussions and conclusions section as follows (L388-L392 in the manuscript):

> The approaches proposed in this study integrate ML with a dynamical modeling framework, with the primary objective of reducing the intrinsic prediction errors of the dynamical model itself. Unlike purely data-driven models (e.g., Andersson et al., 2021; Ren et al., 2024; Kim et al., 2025), which are typically designed for statistical prediction of specific sea ice properties, ML here aims to improve the overall performance of the dynamical prediction system that ensures physical consistency among a large number of predicted variables.

**Comment 4**

c) The current training framework employs a relatively short rolling forecast window (10-month input $\longrightarrow$ 1-month output). However, would a simple MLP's nonlinear approximation capacity remain effective when applied to longer windows required for daily sub-seasonal timescale predictions (e.g., 90-day input $\longrightarrow$ 1-day output)? Alternatively, might this would require more comprehensive deep learning methods? I recommend expanding the discussion to explicitly address the method's generalizability across varying temporal scales.

**Response:** We thank the reviewer very much for this comment. As stated in the manuscript, we used the latest 11 years of data for training and validation for each month (1-month input $\longrightarrow$ 1-month output, L153-L155 in the manuscript):

> Considering the seasonality of the error of the sea ice state, we build one error correction model for each calendar month. Also, we employ a running training strategy and use the most recent 11 years of data before the prediction month (the first 10 years for training and the last year for validation).

As suggested, we have revised the relevant paragraph in the discussions and conclusions section (L406-L422 in the manuscript):

> It is important to note that the proposed approaches still have room for improvement. In this study, we only use ocean and sea ice variables as input features. Including atmospheric variables would help to address errors due to both dynamic and thermodynamic processes and further improve the performance. Increasing the frequency of online correction could help enhance its effectiveness (Gregory et al., 2024), but this is challenging in practice since analysis increments in NorCPM are currently available only every month. An

alternative strategy is to train hybrid models that combine ML with dynamical models, which has been shown to be effective in other systems (Farchi et al., 2021). However, this approach relies on external constraints to compute the gradient of the dynamical model, which are not available in NorCPM.

Furthermore, the current ML model (MLP) is trained independently at each grid point and thus cannot capture spatial correlations. This limits its ability to correct spatially coherent errors, particularly in regions where NorCPM already performs well and only subtle adjustments are needed. As a result, the hybrid model often struggles to reproduce the reanalysis, which are treated as the "truth" in this study. While it is unrealistic to expect the model to perfectly replicate analysis increments, the discrepancy is closely related to the ML-based model's learning capacity and the nature of the underlying errors. Possible contributing factors include: (1) the lack of spatial dependencies due to pointwise training and (2) the tendency of models trained on long-term data to learn systematic biases rather than instantaneous random errors, the latter of which tend to be averaged out over time. Therefore, there is still room to improve the ML-based error correction framework. Future studies could explore spatially-aware architectures, such as CNNs and U-Net, and incorporate additional predictors to capture complex error structures and enhance correction performance (Palerme et al., 2024).

### Comment 5

d) Comparing online and offline methods based on predictive skill metrics is necessary but may be insufficient. Could the analysis be extended to incorporate additional dimensions to better delineate their applicable scenarios? This expanded discussion would provide more actionable guidance for researchers applying machine learning to calibrate model forecasts in operational settings.

**Response:** We agree with the reviewer on this comment. We have added some discussions on physical consistency and computational cost and have revised the relevant paragraph of the discussions and conclusions section as follows (L398-L412 in the manuscript):

By comparing the two error correction approaches, we find that the offline approach yields smaller errors than the online approach. The online error correction approach corrects instantaneous model errors only on the 15th day of the month, and the effect of this correction gradually weakens during model integration due to the accumulation of errors in the other model components. Consequently, the impact of the correction becomes less evident when computing monthly-averaged outputs. Nevertheless, the online error correction can reduce errors in SST and SSS (Figures S7 and S8). Moreover, the online correction approach maintains better physical consistency among the predicted variables through dynamical model integration. The offline error correction approach directly corrects the model outputs without requiring model integration. As a result, it is computationally more efficient and easier to integrate into operational sea ice prediction systems than the online approach.

It is important to note that the proposed approaches still have room for improvement. In this study, we only use ocean and sea ice variables as input features. Including atmospheric variables would help to address errors due to both dynamic and thermodynamic processes and further improve the performance. Increasing the frequency of online correction could help enhance its effectiveness (Gregory et al., 2024), but this is challenging in practice since analysis increments in NorCPM are currently available only every month. An alternative strategy is to train hybrid models that combine ML with dynamical models, which has been shown to be effective in other systems (Farchi et al., 2021). However, this approach relies on external constraints to compute the gradient of the dynamical model, which are not available in NorCPM.

**Specific Comments**

**Comment 6**

The language needs to be polished, and the logical flow between sentences should be strengthened. The paper's subtitles do not clearly convey the intended meaning and would benefit from revision.

**Response:** We carefully checked the language and logical flow between sentences and believe that the manuscript has been improved. For clarity, we have revised the title of this study as follows:

> Correcting Errors in Seasonal Arctic Sea Ice Prediction of Earth System Model with Machine Learning

**Comment 7**

Lines 9-10: If possible, please use one or two sentences to further elaborate on why the offline error correction approach performs better than the online error correction approach.

**Response:** We thank the reviewer for the comment. As suggested, we have revised the abstract as follows (L10-L13 in the manuscript):

> This is primarily because the online approach targets only instantaneous model errors on the 15th of each month, while errors can grow during the subsequent one-month model integration due to interactions among the model components, damping the error correction in monthly averages.

**Comment 8**

Line 14: Is there a newer paper to cite? This one is not "recent" enough.

**Response:** We have added some other recent references into the manuscript as follows (L16-L17 in the manuscript):

> According to satellite observations, the Arctic sea ice extent (SIE) rapidly declines throughout all calendar months during the recent decades (e.g., Serreze et al., 2007; Onarheim et al., 2018; Wang et al., 2022; Heuzé and Jahn, 2024).

**Comment 9**

Line 19: I suppose the word "compared" could be removed.

**Response:** As suggested, we have removed " and compared" from the text.

**Comment 10**

Line 58: The abbreviation "SIC" appears for the first time without a definition.

**Response:** We thank the reviewer for the comment. We have revised the text as follows (L24-L26 in the manuscript):

> They found that dynamical and statistical models are overall comparable in predicting the Pan-Arctic SIE, and dynamical models generally outperform statistical models in predicting the regional SIE and sea ice concentration (SIC, i.e., local quantities).

**Comment 11**

Line 69: The title should be changed to "Data and Methods" as this section introduces the model and data parts first.

**Response:** As suggested, we have revised the section title to "Data and Methods" (L74 in the manuscript).

**Comment 12**

Lines 94: "the summer sea ice extent" should be revised to "the summer SIE".

**Response:** As suggested by the reviewer, we have modified the text as follows (L99-L100 in the manuscript):

> Consequently, the summer SIE in the Arctic has large positive biases, contributing to an underestimation of global temperatures (Bentsen et al., 2013; Bethke et al., 2021).

**Comment 13**

Lines 102-103: Maybe I missed something, but why not directly use observations as the "truth"? How much discrepancy is there between the reanalysis of NorCPM and observations?

**Response:** We appreciate the reviewer's comment. The main reason is that NorCPM employs anomaly-field assimilation (Kimmritz et al., 2019; Wang et al., 2019; Bethke et al., 2021), which keeps the model close to its attractor and helps to reduce model drift during the monthly model integration (Carrassi et al., 2014; Weber et al., 2015). The online error correction approach based on the analysis increments of the anomaly-field assimilation is not able to reduce model biases. For clarity, we have revised section 2.1 as follows (L97-L110 in the manuscript):

> NorESM1 tends to overly produce thick sea ice, especially in the polar oceans adjacent to the Eurasian continent. This is partly due to factors such as weaker winds across the polar basin and overestimated Arctic cloudiness, which leads to little summer snowmelt. Consequently, the summer SIE in the Arctic has large

positive biases, contributing to an underestimation of global temperatures (Bentsen et al., 2013; Bethke et al., 2021).
NorCPM uses the EnKF to update unobserved ocean and sea ice variables by leveraging state-dependent covariance from the simulation ensemble (Kimmritz et al., 2018, 2019). The EnKF allows the assimilation of observations of various types while accounting for observational errors, spatial coverage, and the evolving covariance with the climate state. The EnKF accounts for uncertainties in initial conditions to generate ensemble predictions, which evolve in time and provide time- and space-dependent error estimates.
NorCPM employs anomaly-field assimilation (Kimmritz et al., 2019; Wang et al., 2019; Bethke et al., 2021) in which the climatology of the observations is replaced by the model climatology calculated from the ensemble mean of the model historical simulation (without assimilation). While the anomaly-field assimilation keeps the model close to its attractor and helps to reduce the model drift during the monthly model integration (Carrassi et al., 2014; Weber et al., 2015), it does not significantly change model biases.

In section 2.2, we have modified the text to provide the reasons for using the reanalysis as the "truth" (L112-L119 in the manuscript):

In this study, we use the reanalysis of NorCPM as the "truth" to assess the improvement achieved by the ML-based error correction approaches. First, it is because NorCPM performs anomaly-field assimilation. The large model biases are not corrected by DA (section 2.1) and thus the analysis increment of the reanalysis used to build the online error correction model (section 2.3) does not take into account model biases. Second, the online error correction approach needs to consistently update SIC in each category, sea surface temperature (SST), and sea surface salinity (SSS) under sea ice, which are often not observed. The reanalysis of NorCPM is a physically consistent construction of the Earth system (Counillon et al., 2016; Kimmritz et al., 2019) and provides a reasonable and physically consistent estimation of these variables. Finally, the reanalysis combining observations with NorESM represents the upper limit of the sea ice predictability of NorCPM.

**Comment 14**

Line 130 & Table 1: Why consider latitude only but not longitude? Can any explanation be provided? And how about the relative importance of these input features?

**Response:** In this study, our error correction approaches take SST, SSS, SIC, and SIT—along with additional latitude information. Longitude is excluded primarily due to technical considerations. Geographically, 0° and 360° represent the same location, but in the ML context, they are treated as two extreme values. To avoid introducing artificial discontinuities or distortions, we chose not to include longitude as an input variable during ML model training.

We apologize to the reviewer for not providing the relative importance of each variable during training, as our ML model training did not track feature importance explicitly. However, Gregory et al. (2023) offered valuable insights into the relative contributions of the predictors, with additional consideration of sea ice velocity components (SIU and SIV). Their findings indicate that SIC alone accounts for approximately 66% of the total prediction skill, making it the most influential input. SST, together with SIU and SIV, contributes an additional 20%, with SST itself estimated to provide around 6–8%. The remaining variables—SIT, shortwave radiation (SW), ice-surface skin temperature (TS), and SSS—collectively account for the final 14%, with SIT and SSS contributing approximately 3–5% and 2–3%, respectively.

**Comment 15**

Lines 142-143: I am curious about the exact post-processing of physically inconsistent fields. Could you give a concise and clear description rather than simply citing a paper?

**Response:** For clarity, we have revised the text as follows (158-L166 in the manuscript):

> Before restarting the model after applying online error correction, it is essential to ensure that the updated variables remain within physical limits (e.g., SIC between 0% and 100%) and maintain consistency with non-updated variables. If unphysical values or inconsistencies arise, they can lead to model instability. To prevent these issues, we apply a post-processing method specifically designed for NorCPM (Kimmritz et al., 2018):
>
> – If SIC in any thickness category falls below 0% or exceeds 100%, it is set to 0% or 100%, respectively.
>
> – If the total SIC across all thickness categories exceeds 100%, SIC values in each category are proportionally scaled to ensure the total does not surpass 100%.
>
> – Sea ice volume in each category is adjusted proportionally to changes in SIC while preserving the ice thickness.
>
> This approach ensures physical constraint and model stability after the error correction.

**Comment 16**

Line 205: "Pan-arctic". Please make the capitalization of this term consistent throughout the paper. "Fig. 2" or "Figure 2"? Please also make this consistent throughout the paper.

**Response:** We thank the reviewer for the suggestions. We have gone through the whole manuscript to consistently use "Pan-Arctic" and "Figure X".

**Comment 17**

Lines 217-218: "We define the IIEE as the area where the prediction and the truth disagree on the ice concentration being above or below 15%:" It would be better to rephrase as the IIEE metric has been defined by Goessling et al. (2016). Are the authors themselves defining a new metric called IIEE?

**Response:** We followed the definition of Goessling et al. (2016). For clarity, we have revised the text as follows (L263-L264 in the manuscript):

> Following the definition of Goessling et al. (2016), the IIEE is computed as the area where the prediction and the "truth" disagree on the SIC being above or below 15%

**Comment 18**

Lines 211 & 220-221: Why are consistent subscripts not used in these two equations?

**Response:** Sorry for the confusion. To be consistent in subscripts, we have revised the manuscript as follows (L255-L259 in the manuscript):

> To evaluate the sea ice prediction skill, we employ the root mean square error (RMSE) as follows:
>
> $$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}_{\mathrm{p}} - \mathbf{X}_{\mathrm{t}})^2}, \tag{1}$$
>
> where $\mathbf{X}_{\mathrm{p}}$ represents the prediction and $\mathbf{X}_{\mathrm{t}}$ represents the "truth" (i.e., the reanalysis in this study). In this study, $\mathbf{X}$ can refer to either the integrated ice-edge error (IIEE) on a Pan-Arctic scale, the SIE on a Pan-Arctic/regional scale, or the SIC at a specific grid point. $N$ represents the number of hindcasts, spanning from 2003 to 2021.

**Comment 19**

Line 223: What is the meaning of "squared errors"? Do you mean "RMSE"?

**Response:** The reviewer is right. We have replaced "squared errors" by "RMSE" in the manuscript (L271 in the manuscript).

**Comment 20**

Line 253: "NorCPM overestimates the Arctic cloudiness, and its summer-season snowmelt is too slow." I am not clear which figure I can draw such a conclusion from.

**Response:** We agree with the reviewer on this comment. The conclusion was drawn from Bentsen et al. (2013). For clarity, we have revised the text as follows (L299-L304 in the manuscript):

> Figure 5 presents a comparative analysis of the RMSE for SIE prediction and the IIEE for ice edge prediction in the Pan-Arctic across the three hindcast sets. The Reference hindcast shows higher RMSE in September and October (Figure 5a), primarily due to several factors that have been documented in Bentsen et al. (2013). NorCPM overestimates the Arctic cloudiness, and its summer-season snowmelt is too slow. In addition, NorCPM has slightly too weak winds across the polar basin. These factors lead to too thick sea ice in the polar oceans and excessive Arctic SIE, in particular in summer (Bentsen et al., 2013).

**Comment 21**

Line 256: "Both the OnlineML and OfflineML hindcasts exhibit similar behaviors regardless of the seasonality." This sentence is somewhat confusing, please rephrase it.

**Response:** For clarity, we have revised the relevant text as follows (L305-L306 in the manuscript):

> Both the OnlineML and OfflineML hindcasts exhibit a small error reduction from January to July and a large error reduction from August to December (Figure 5b and 5c).

**Comment 22**

Lines 276-277: Why choose to analyze/present the reanalysis initialized in July? Could you provide some explanation? The later analysis should also clarify whether the results shown in Figure 6 depend on the initialization month.

**Response:** For clarity, we have revised the first paragraph of section 3.2.2 as follows (L327-L331 in the manuscript):

> The previous section highlighted significant improvements in predictions, primarily evident from September to January regardless of the initialized month. In this section, we focus on analyzing the hindcasts initialized in July, and we show the performance for different regions and both SIE and SIC. It is mostly because summer sea ice prediction serves as a critical climate change indicator, affects ecosystems and human activities, and presents a significant scientific challenge due to its high variability (Figure 5a). For validation on the other initialization months, please refer to Figures S1-S4.

Also, we have modified the summary paragraph as follows (L362-L365 in the manuscript):

> In summary, while the error correction performance varies by region and target month, overall, both approaches improve the sea ice prediction. In addition, the offline approach is more efficient than the online approach in reducing the SIE RMSE for both Pan-Arctic and subregions. These conclusions also hold for seasonal predictions initialized in the other seasons. For details, please refer to Figures S1-S4.

**Comment 23**

Lines 289-293 (Figure 6d & 6e): Why is the result of the Online ML hindcast in August worse than the Reference hindcast in the Alaskan and Canadian regions?

**Response:** From the time series of SIE in both regions (Figure R1), the OnlineML consistently underestimates SIE compared to the "truth" and the Reference hindcast. However, the underlying mechanisms responsible for this bias remain unclear at this stage and warrant further investigation.

We also added a description in the manuscript as follows (L350-L353 in the manuscript):

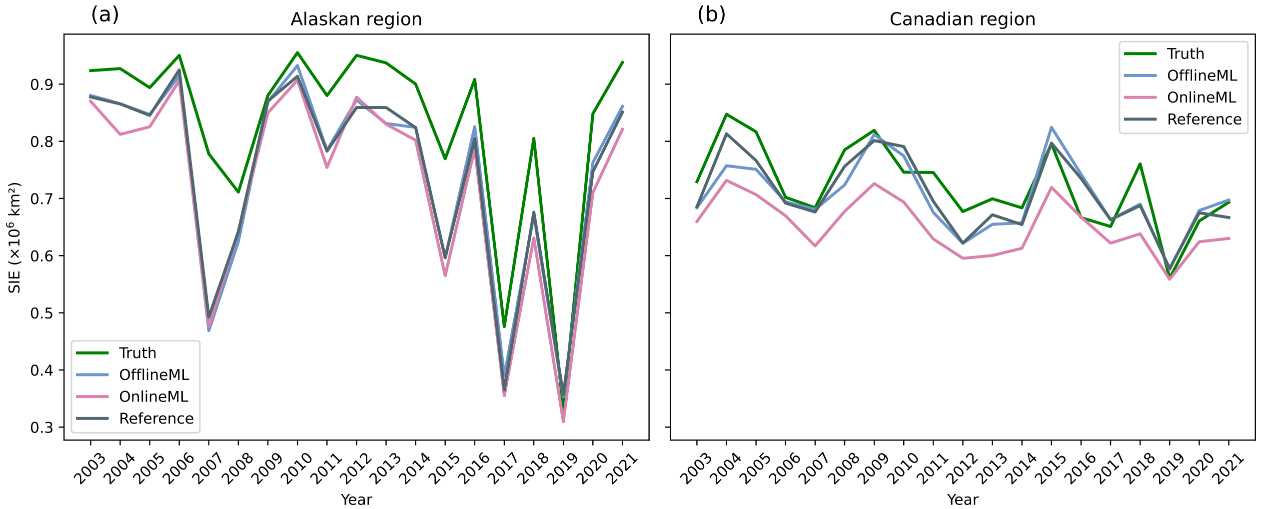


**Figure R1.** Time series of SIE in the Alaskan and Canadian regions.

> Notably, in August, the RMSE of the OnlineML hindcast exceeds that of the Reference hindcast in both the Alaskan and Canadian regions. This is primarily due to the systematic underestimation of SIE by the OnlineML hindcast relative to both the Reference hindcast and the "truth" in these regions (Figure S5). The underlying causes of this systematic underestimation, however, warrant further investigation.

### Comment 24

Lines 297-299: As the author mentioned, the different performance of these two approaches (OnlineML and OfflineML) comes from the way they are constructed. Therefore, these two methods should be intended for different purposes. Is this comparison appropriate? Maybe rephrasing it would be better.

**Response:** As suggested, we have revised the text as follows (L354-L361 in the manuscript):

> The offline approach outperforms the online approach across all regions, primarily because the online correction targets instantaneous model errors (i.e., those on the 15th day of each month). These corrected errors may reemerge through interactions with the other components of the coupled model system, thereby diminishing the overall impact of the online error correction when evaluated using monthly-averaged outputs. In contrast, the offline approach directly adjusts the monthly model outputs, which aligns closely with the evaluation metrics used in this study. Moreover, the offline approach does not need to run the dynamical model and is computationally cheaper than the online approach. However, the online approach not only reduces SIC errors but also propagates corrections through the model integration to the other variables (e.g., sea ice thickness and sea ice drift), ensuring physical consistency between the predicted variables.

**Comment 25**

Line 302: Does the error correction performance vary with the initialization month (as mentioned above)?

**Response:** As shown in Figures 5, 6, and S1-S4 of the manuscript, the effectiveness of the correction is more sensitive to the target month than to the initialization month. For clarity, we have modified the summary paragraph as follows (L362-L365 in the manuscript):

> In summary, while the error correction performance varies by region and target month, overall, both approaches improve the sea ice prediction. In addition, the offline approach is more efficient than the online approach in reducing the SIE RMSE for both Pan-Arctic and subregions. These conclusions also hold for seasonal predictions initialized in the other seasons. For details, please refer to Figures S1-S4.

**Figures**

**Comment 26**

Figure 1: The colors in Figure 1 are somewhat confusing (especially the purple and pink, which may cause difficulty for readers in distinguishing them). I recommend to use more distinguishable colors.

**Response:** We thank the reviewer for the comment. As suggested, we have replotted Figure 1 with more distinguishable colors. The revised figure is also presented below as Figure R2.

**Comment 27**

Figure 2: "Regional domain definitions for Central Arctic, Atlantic, Siberian, Alaskan, Canadian, and Regions based on sea area definitions in Kimmritz et al. (2019)." The "Regions" should be corrected to "regions". I think it seems a bit crude to combine the Bering and the Sea of Okhotsk into the "Pacific Region", is there any literature to support this approach?

**Response:** We thank the reviewer for the comments. We have corrected "Regions" to "regions" and revised the caption of Figure 2 as follows:

> Regional domain definitions for central Arctic, Atlantic, Siberian, Alaskan, Canadian, and Pacific regions are based on sea area definitions in Kimmritz et al. (2019) and are similar to those used in Bushuk et al. (2024). Atlantic region: Greenland, Ice, Norwegian, Barents and Kara Seas; Siberian region: Laptev and East Siberian Seas; Alaskan region: Chukchi and Beaufort Seas; Canadian region: Canadian archipelago, Hudson Bay, Baffin Bay, and Labrador Sea; Pacific region: Bering Sea and the Sea of Okhotsk.

We decided to group the Bering Sea and Sea of Okhotsk as one region mostly due to similar sea ice characteristics in the Bering Sea and Sea of Okhotsk (e.g., strong seasonality) and their geographic locations. In addition, we aimed to reduce the number of the regions and thus the number of sub-figures (e.g., those shown in Figure 6 of the manuscript). Please note that the other regions are very similar to those used in Bushuk et al. (2024).

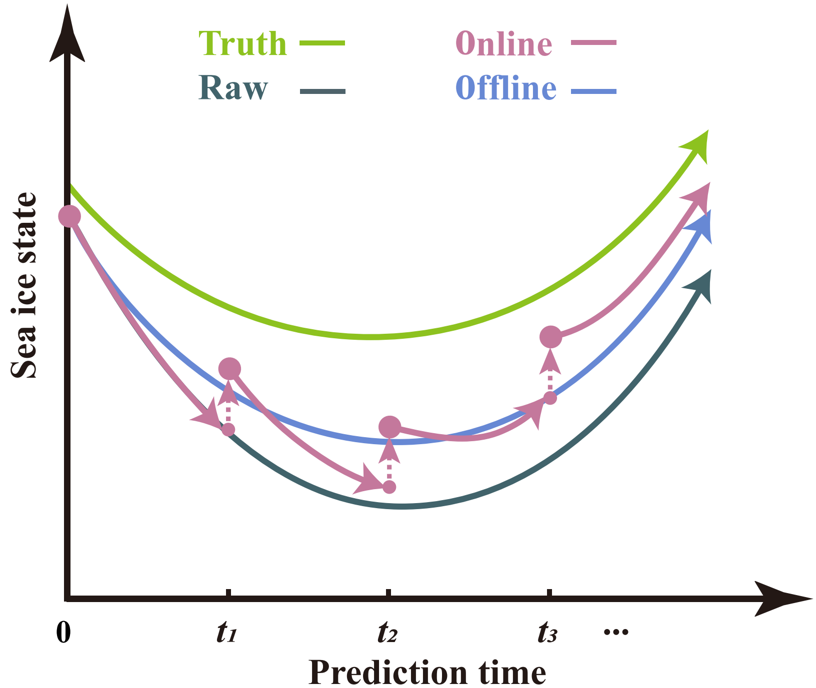


**Figure R2.** Schema for the online and offline ML-based error correction approaches. The green line represents the "truth". The gray line represents dynamical prediction without error correction. The purple (blue) line represents prediction with online (offline) ML-based error correction. The purple dashed arrows indicate pauses during the prediction production, facilitating correction to the instantaneous model states.

**Comment 28**

Figure 3: Please indicate in the figure caption that this is the result of reanalysis minus the model (as in Figure 4's caption).

**Response:** As suggested, we have modified the caption of Figure 3 as follows:

> Top row: "true" errors of SIC in the middle of the month based on the analysis increments (i.e., the changes thanks to monthly DA in the reanalysis). Bottom row: the errors predicted by the online error correction model. These errors are averaged over the period 2003-2021. Values in the bottom row are the MAE between the "true" and predicted errors across space.

## References


Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., et al.: Seasonal Arctic sea ice forecasting with probabilistic deep learning, Nature communications, 12, 5124, 2021.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., et al.: The Norwegian Earth System Model, NorESM1-M–Part 1: description and basic evaluation of the physical climate, Geoscientific Model Development, 6, 687–720, 2013.

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., et al.: NorCPM1 and its contribution to CMIP6 DCPP, Geoscientific Model Development Discussions, 2021, 1–84, 2021.

Bushuk, M., Ali, S., Bailey, D. A., Bao, Q., Batté, L., Bhatt, U. S., Blanchard-Wrigglesworth, E., Blockley, E., Cawley, G., Chi, J., et al.: Predicting September Arctic Sea Ice: A Multi-Model Seasonal Skill Comparison, Bulletin of the American Meteorological Society, 2024.

Carrassi, A., Weber, R., Guemas, V., Doblas-Reyes, F., Asif, M., and Volpi, D.: Full-field and anomaly initialization using a low-order climate model: a comparison and proposals for advanced formulations, Nonlinear Processes in Geophysics, 21, 521–537, 2014.

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., and Bentsen, M.: Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model, Tellus A: Dynamic Meteorology and Oceanography, 68, 32 437, 2016.

Farchi, A., Laloyaux, P., Bonavita, M., and Bocquet, M.: Using machine learning to correct model error in data assimilation and forecast applications, Quarterly Journal of the Royal Meteorological Society, 147, 3067–3084, 2021.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T.: Predictability of the Arctic sea ice edge, Geophysical Research Letters, 43, 1642–1650, https://doi.org/https://doi.org/10.1002/2015GL067232, 2016.

Gregory, W., Bushuk, M., Adcroft, A., Zhang, Y., and Zanna, L.: Deep learning of systematic sea ice model errors from data assimilation increments, Journal of Advances in Modeling Earth Systems, 15, e2023MS003 757, 2023.

Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., and Zanna, L.: Machine learning for online sea ice bias correction within global ice-ocean simulations, Geophysical Research Letters, 51, e2023GL106 776, 2024.

Heuzé, C. and Jahn, A.: The first ice-free day in the Arctic Ocean could occur before 2030, Nature Communications, 15, 1–10, 2024.

Kim, Y. J., Kim, H.-c., Han, D., Stroeve, J., and Im, J.: Long-term prediction of Arctic sea ice concentrations using deep learning: Effects of surface temperature, radiation, and wind conditions, Remote Sensing of Environment, 318, 114 568, 2025.

Kimmritz, M., Counillon, F., Bitz, C., Massonnet, F., Bethke, I., and Gao, Y.: Optimising assimilation of sea ice concentration in an Earth system model with a multicategory sea ice model, Tellus A: Dynamic Meteorology and Oceanography, 70, 1–23, 2018.

Kimmritz, M., Counillon, F., Smedsrud, L. H., Bethke, I., Keenlyside, N., Ogawa, F., and Wang, Y.: Impact of ocean and sea ice initialisation on seasonal prediction skill in the Arctic, Journal of Advances in Modeling Earth Systems, 11, 4147–4166, 2019.

Onarheim, I. H., Eldevik, T., Smedsrud, L. H., and Stroeve, J. C.: Seasonal and regional manifestation of Arctic sea ice loss, Journal of Climate, 31, 4917–4932, 2018.

Palerme, C., Lavergne, T., Rusin, J., Melsom, A., Brajard, J., Kvanum, A. F., Macdonald Sørensen, A., Bertino, L., and Müller, M.: Improving short-term sea ice concentration forecasts using deep learning, The Cryosphere, 18, 2161–2176, 2024.

Ren, Y., Li, X., and Wang, Y.: SICNet season V1. 0: a transformer-based deep learning model for seasonal Arctic sea ice prediction by integrating sea ice thickness data, Geoscientific Model Development Discussions, 2024, 1–20, 2024.

Serreze, M. C., Holland, M. M., and Stroeve, J.: Perspectives on the Arctic's shrinking sea-ice cover, science, 315, 1533–1536, 2007.

Wang, X., Liu, Y., Key, J. R., and Dworak, R.: A new perspective on four decades of changes in Arctic sea ice from satellite observations, Remote Sensing, 14, 1846, 2022.

Wang, Y., Counillon, F., Keenlyside, N., Svendsen, L., Gleixner, S., Kimmritz, M., Dai, P., and Gao, Y.: Seasonal predictions initialised by assimilating sea surface temperature observations with the EnKF, Climate Dynamics, 53, 5777–5797, 2019.

Weber, R. J., Carrassi, A., and Doblas-Reyes, F. J.: Linking the anomaly initialization approach to the mapping paradigm: a proof-of-concept study, Monthly Weather Review, 143, 4695–4713, 2015.