

Seasonal Forecasting of Pan-Arctic Sea Ice with State Space Model

Wei Wang¹, Weidong Yang^{1*}, Lei Wang^{2,3}, Guihua Wang²,
Ruibo Lei⁴

^{1*}School of Computer Science, Fudan University, Shanghai, China.

²Department of Atmospheric and Oceanic Sciences & Institute of Atmospheric Sciences, Fudan University, Shanghai, China.

³Key Laboratory of Polar Atmosphere-ocean-ice System for Weather and Climate, Ministry of Education, Fudan University, Shanghai, China.

⁴Key Laboratory of Polar Science, MNR, Polar Research Institute of

*Corresponding author(s). E-mail(s): wdyang@fudan.edu.cn;
Contributing authors: wang_wei23@m.fudan.edu.cn;
wanglei_ias@fudan.edu.cn; wanggh@fudan.edu.cn; leiruibo@pric.org.cn;

Abstract

The rapid decline of Arctic sea ice resulting from anthropogenic climate change poses significant risks to indigenous communities, ecosystems, and the global climate system. This situation emphasizes the immediate necessity for precise seasonal sea ice forecasts. While dynamical models perform well for short-term forecasts, they encounter limitations in long-term forecasts and are computationally intensive. Deep learning models, while more computationally efficient, often have difficulty managing seasonal variations and uncertainties when dealing with complex sea ice dynamics. In this research, we introduce IceMamba, a deep learning architecture that integrates sophisticated attention mechanisms within the state space model. Through comparative analysis of 25 renowned forecast models, including dynamical, statistical, and deep learning approaches, our experimental results indicate that IceMamba delivers excellent seasonal forecasting capabilities for Pan-Arctic sea ice concentration. Specifically, IceMamba outperforms all tested models regarding average RMSE and anomaly correlation coefficient (ACC) and ranks second in Integrated Ice Edge Error (IIIE). This innovative approach enhances our ability to foresee and alleviate the effects of sea ice variability, offering essential insights for strategies aimed at climate adaptation.

³⁴ 1 Introduction

³⁵ Arctic sea ice plays a critical role in the global climate system, profoundly influencing
³⁶ the atmosphere, oceans, and global weather patterns [1–3]. As global warming
³⁷ intensifies, the retreat of Arctic sea ice has become a notable characteristic of cli-
³⁸ mate change. The decline in sea ice alters the surface albedo, reducing the reflection
³⁹ of solar radiation, leading to increased heat absorption by the exposed ocean, further
⁴⁰ accelerating the warming in the region [4]. This warming can disrupt atmospheric cir-
⁴¹ culation patterns [5] and may contribute to an increase in extreme weather events in
⁴² mid-latitude [6]. Additionally, sea ice loss affects the Arctic’s role in regulating global
⁴³ ocean circulation, particularly the meridional overturning circulation (AMOC) [7],
⁴⁴ which has profound implications for global climate stability. Therefore, understand-
⁴⁵ ing and forecasting the seasonal variations of Arctic sea ice is not only important
⁴⁶ for regional climate assessments but also for grasping its global climatic implications.
⁴⁷ Accurate seasonal forecasts can help Arctic policymakers address climate change, pro-
⁴⁸ mote sustainable resource development, and protect wildlife [8]. They are also critical
⁴⁹ for understanding long-term climate dynamics, as the changes in the Arctic climate
⁵⁰ underscore the importance of precise monthly to seasonal projections for global climate
⁵¹ systems [9]. However, the rapid changes in sea ice under global warming conditions pose
⁵² challenges for forecasts, making the development of forecast methods both scientifically
⁵³ and practically significant.

⁵⁴ Arctic sea ice forecasting research can be divided into physics-driven and data-
⁵⁵ driven approaches. Physics-based methods utilize numerical models integrating ocean,
⁵⁶ atmosphere, land, and ice interactions to understand the involved physical mech-
⁵⁷ anisms [10–12]. These models offer high precision but require significant computational
⁵⁸ resources. They are sensitive to initial conditions, and structural model physics errors
⁵⁹ can lead to systematic biases [13]. In contrast, data-driven approaches, including sta-
⁶⁰ tistical and machine learning models, provide computationally lighter alternatives.
⁶¹ Historically, data-driven sea ice forecasts have primarily targeted predictions of inte-
⁶² grated Pan-Arctic September sea ice extent, achieving notable success in seasonal
⁶³ outlooks [14–16]. While these extent-based forecasts demonstrate high skill metrics,
⁶⁴ their limited operational utility, stemming from coarse spatial resolution and inabil-
⁶⁵ ity to resolve regional ice dynamics, has motivated a paradigm shift toward spatially
⁶⁶ explicit sea ice concentration (SIC) and probabilistic forecasts. Statistical models such
⁶⁷ as vector autoregressive and Markov models are used in sea ice forecasting but strug-
⁶⁸ gle with capturing nonlinear dynamics [17–22]. Machine learning, particularly deep
⁶⁹ learning (DL), has shown greater effectiveness in modeling the nonlinear evolution of
⁷⁰ sea ice. Current DL models forecast sea ice concentration (SIC) at both monthly [23–
⁷¹ 28] and daily [29–36] scales, offering forecasting capabilities for different time horizons.
⁷² Despite these advancements in DL models, challenges remain in capturing multiscale
⁷³ dependencies in sea ice dynamics, which exhibit complex spatiotemporal patterns.
⁷⁴ Existing DL models, such as Convolutional Neural Networks (CNNs) and Long Short-
⁷⁵ Term Memory networks (LSTMs) [37], struggle with spatial variability due to fixed
⁷⁶ receptive fields, which becomes particularly problematic in modeling complex, mul-
⁷⁷ tiscale phenomena like sea ice dynamics. While Transformers [38] capture a global
⁷⁸ receptive field, their quadratic complexity makes them inefficient for processing such

79 large-scale, heterogeneous datasets. Moreover, the large volume of sea ice, ocean, and
80 atmospheric data from diverse sources—such as observations, reanalysis, and modeling—introduces significant variability in data quality, which is strongly correlated with
81 forecast accuracy. This heterogeneity in data sources often increases model uncertainty,
82 complicating predictions.

83 To address these challenges, we introduce IceMamba, a novel DL architecture for
84 seasonal SIC forecasting, combined with the State space model. To our knowledge,
85 this is the first study to apply state space models to sea ice forecasting. IceMamba
86 processes historical sea ice concentration (SIC) data along with reanalysis variables
87 from ERA5 [39] and ORAS5 [40] to forecast monthly averaged SIC maps over the next
88 several months at a 25 km resolution. We design the Residual Efficient State Space
89 Block (RESSB) to enhance the model’s performance by focusing on temporal and vari-
90 able domains, enabling efficient extraction of correlations between climate and sea ice
91 variables. In the spatial domain, RESSB inherits the global receptive field from the
92 Vision state space block (VSSB) [41], enabling IceMamba to efficiently capture the
93 spatiotemporal dependencies of sea ice dynamics across different spatial regions, sur-
94 passing traditional DL models with fixed receptive fields. Moreover, IceMamba also
95 inherits the linearly increasing computational complexity from VSSB [41], making it
96 more efficient than Transformers, which exhibit quadratic complexity as the input size
97 grows. In the temporal and variable domains, RESSB effectively enhances IceMamba’s
98 attention to various input variables and their durations through the Efficient Chan-
99 nel Attention (ECA) module [42]. This enables the IceMamba efficient extraction of
100 correlations between different climate variables, thereby improving the forecast perfor-
101 mance and stability of the model. Experimental results demonstrate that incorporating
102 oceanic heat content and mixed-layer depth measurements from the ORAS5 reanalysis
103 dataset significantly enhances IceMamba’s sea ice forecasting performance, establish-
104 ing subsurface oceanographic parameters as critical predictors beyond surface-level
105 sea surface temperature. Under the seasonal forecast benchmark [43], IceMamba out-
106 performs all tested models in average RMSE and ACC metrics, while ranking second
107 in IIEE. Through comprehensive evaluation of extreme September sea ice events, Ice-
108 Mamba demonstrates robust SIC forecast capabilities under critical conditions in the
109 seasonal forecast benchmark comparisons. Besides, the permutation-based explain-
110 ability framework demonstrates IceMamba’s ability to dynamically prioritize key
111 mechanisms across seasonal timescales: short-term forecasts depend predominantly on
112 recent SIC, reflecting persistence, whereas longer lead times shift toward historical SIC
113 from the same month in prior years, indicating an implicit seasonal alignment with-
114 out explicit temporal encoding. The model also captures ice-albedo feedback through
115 heightened sensitivity to upward solar radiation during summer melt, contrasting with
116 weaker responses to downward radiation, while distinguishing atmospheric drivers
117 from delayed oceanic influences. Detrending analyses suggest that apparent strato-
118 spheric linkages via 10 hPa zonal wind speed primarily arise from anthropogenic
119 trends [44, 45] rather than direct dynamical coupling, as sensitivity to detrended 10
120 hPa zonal wind speed declines sharply. These results highlight IceMamba’s capacity to
121 disentangle timescale-dependent interactions from rapid atmospheric forcing to slow
122 oceanic adjustments, while maintaining physical consistency in feature attribution.

124 **2 Result**

Table 1 Comparison of mean MAE, RMSE, IIEE, and ACC for IceMamba variants and baseline models over non-land regions in the Pan-Arctic.

	MAE (%)	RMSE (%)	IIEE ($\times 10^6 \text{ km}^2$)	ACC
IceMamba-1	1.8065	6.9297	0.7865	0.9834
IceMamba-1-ERA5	1.8283	7.0564	0.7929	0.9829
IceMamba-4	2.3024	8.8384	1.0244	0.9731
IceMamba-4-ERA5	2.3258	8.9152	1.0388	0.9727
IceMamba-6	2.4431	9.2994	1.0899	0.9706
IceMamba-6-ERA5	2.5025	9.4414	1.1123	0.9691
IceMamba-6-VSSB	2.4674	9.435	1.0945	0.9696
Anomaly Persistence	3.9940	11.1576	1.4653	0.9579

125 **Overall Performance of IceMamba.** To evaluate the overall performance of the
 126 IceMamba architecture, we trained three variants (IceMamba-1, IceMamba-4, and
 127 IceMamba-6) using a consistent training scheme. Each variant is designed to fore-
 128 cast the monthly mean SIC for the next 1, 4, and 6 months, representing short-term
 129 (IceMamba-1), seasonal (IceMamba-4), and long-term seasonal forecasts (IceMamba-
 130 6), respectively. For each variant, we tested multiple input combinations (as shown
 131 in Supplementary Tables 2 to 4) and the data combination with the best forecast
 132 performance is selected as the final training set for IceMamba-1, IceMamba-4, and
 133 IceMamba-6. Additionally, we trained IceMamba-1-ERA5, IceMamba-4-ERA5, and
 134 IceMamba-6-ERA5 using only ERA5 data to compare the forecast performance of
 135 models only trained with SIC and ERA5 reanalysis variables. This comparative anal-
 136 ysis delivers a comprehensive assessment of model performance with various forecast
 137 windows, while also delineating the specific contributions of ORAS5 reanalysis data
 138 in enhancing predictive capabilities at different temporal scales.

139 The optimal data configurations for these variants revealed a dependence on
 140 the scale of the forecast window and the oceanic processes involved. Specifically,
 141 IceMamba-1 achieved the best performance with a combination of ERA5, ohc300, and
 142 mld001 reanalysis data, reflecting the dominance of surface ocean dynamics in short-
 143 term predictions. Meanwhile IceMamba-4, optimized for seasonal predictions, performs
 144 better with a combination of ERA5, ohc300, and a deeper mixed layer (mld003),
 145 indicating the increasing importance of vertical oceanic processes over a longer fore-
 146 cast horizon. For IceMamba-6, the combination of ERA5, ohc700, and both shallow
 147 and deep mixed layer depths (mld001 and mld003) emerged as the optimal configu-
 148 ration, suggesting that deep ocean heat content and vertical mixing become critical
 149 for long-term seasonal predictions. These findings highlight the model's ability to cap-
 150 ture the shift from surface-driven to deeper ocean-driven processes as the forecast
 151 window lengthens. The scale-dependent shift in optimal input variables underscores
 152 the model's sensitivity to oceanic climate mechanisms and provides insight into how
 153 different ocean layers influence sea ice variability at different temporal scales.

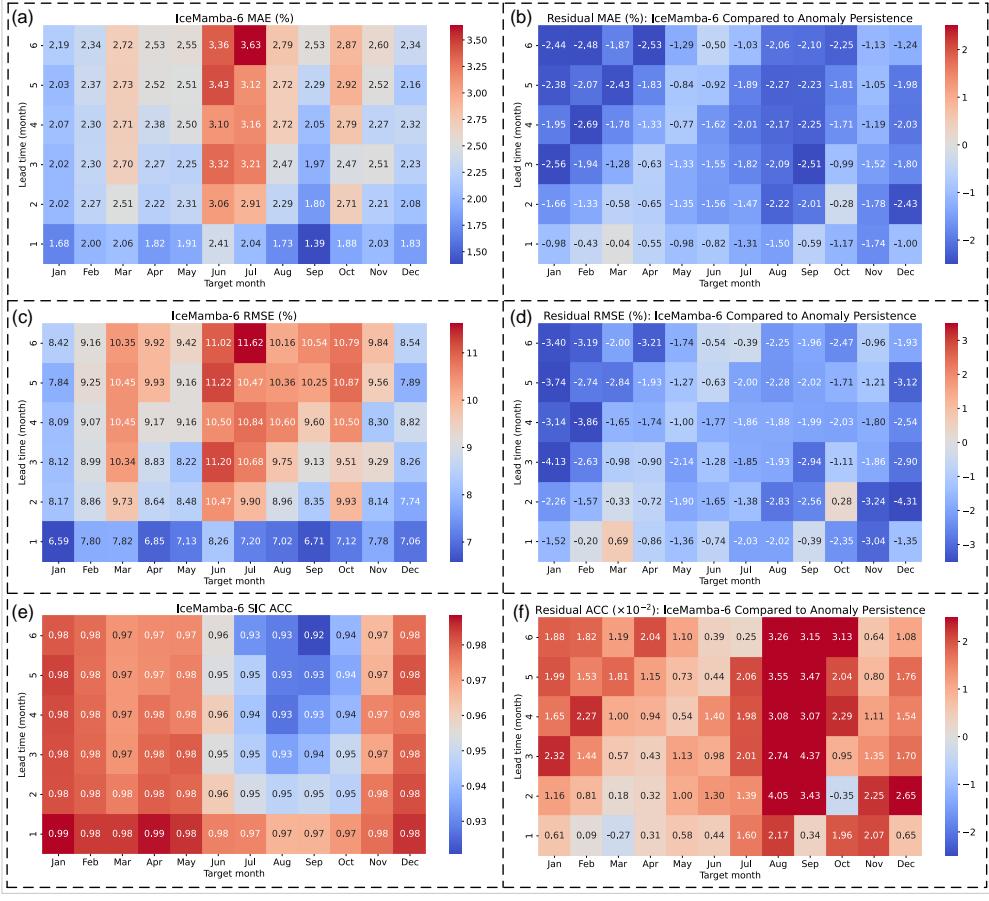


Fig. 1 Comparison of IceMamba-6 and Anomaly Persistence. (a), (c), and (e): MAE, RMSE, and ACC of IceMamba-6, averaged over the test years (2016–2022), presented for each target month (month of prediction) and lead time, with heatmap values in each grid cell. (b), (d), and (f): Heatmaps showing the differences in MAE, RMSE, and ACC between IceMamba-6 and Anomaly Persistence. All metrics are calculated over non-land regions in the Pan-Arctic.

Table 1 presents the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Integrated Ice Edge Error (IIIE), and Anomaly Correlation Coefficient (ACC) for the test sets of each model, with all metrics calculated for the pan-Arctic non-land region. Comprehensive descriptions of these metrics are provided in the Supplementary Information. As shown in Table 1, IceMamba variants incorporating ocean reanalysis data exhibit consistent yet modest superiority across all metrics compared to their ERA5-only counterparts. Supplementary Figure 1 presents the error differences between the IceMamba variants and their corresponding ERA5 models. By comparing the various IceMamba variants with their corresponding ERA5 models, we notice that the benefit of including deep ocean parameters becomes more pronounced with longer forecast windows. While IceMamba-1 and IceMamba-4 also

show improvements over their ERA5 counterparts in several months, the most consistent and significant improvement is observed for IceMamba-6 over the majority of the calendar year. This suggests that for longer-range forecasts, the additional deep ocean data provide enhanced initial conditions and better capture the evolving ocean-ice interactions. Moreover, IceMamba-6 demonstrates a notable reduction in forecast errors during the August–October period compared to other months. This suggests the presence of key ocean-ice coupling mechanisms during the critical transition from the melt season to freeze-up in the Arctic. During this seasonal window, enhanced ocean heat flux from subsurface warm layers and deepening of the mixed layer play a crucial role in influencing sea ice dynamics. These processes are more effectively captured by the oceanographic parameters in the ORAS5 reanalysis, which contribute to the improved model performance during this transitional phase. This result aligns with previous studies [46–48], which demonstrated that the heat capacity of a deep mixed layer enables summer and spring SST anomalies to reemerge in late fall, thereby improving the predictability of SIE anomalies in this region. Given the strong linkage between SIE and SIC, our findings suggest a similar impact on SIC predictability.

To evaluate RESSB, we established IceMamba-6-VSSB as the baseline by substituting RESSB with two VSSBs in IceMamba-6. Both architectures maintained identical experimental configurations and datasets for comparative validity. Furthermore, the Anomaly Persistence forecast is chosen as a baseline comparison model. Anomaly Persistence is defined as the sum of the SIC anomaly and the climatology at each lead time. The climatology is computed using a 10-year sliding window preceding the forecast to account for its temporal changes [49–51]. Compared to the baseline models, all IceMamba variants outperform the Anomaly Persistence model, exhibiting lower MAE, RMSE, and higher ACC, indicating improved model fitting and reduced mean forecast errors. Furthermore, all IceMamba variants demonstrate lower IIEE, suggesting superior sea ice edge forecasting capabilities relative to the baseline models. Notably, IceMamba-6 outperforms IceMamba-VSSB across all metrics, highlighting that RESSB more effectively captures relevant variables and spatiotemporal features from the input data than VSSB.

Fig. 1(a), (c), and (e) display the average MAE, RMSE, and ACC for the IceMamba-6 model across each lead time and target month, respectively. As lead time increases, forecast errors for each target month gradually rise, while the ACC declines. Compared to other target months, the forecast errors and ACC from June to October exhibit more pronounced fluctuations. We hypothesize that this increase in errors is primarily due to the melting season of Arctic sea ice (June to September). During this period, rising temperatures lead to rapid ice melting, resulting in significant changes in ice coverage and dynamics. The pronounced variability in sea ice makes accurate forecasts more challenging. Moreover, the transition to autumn brings about complex interactions between melting ice, ocean currents, and atmospheric conditions, further exacerbating uncertainties in model forecasts. While incorporating ORAS5 subsurface ocean data reduces August–October forecast errors compared to ERA5-only models (Supplementary Fig.1), the error levels during this period remain higher than in other months.

As shown in Fig. 1 (b) and Fig. 1 (d), IceMamba-6 exhibits superior forecast performance, as evidenced by its consistently lower average MAE and average RMSE in almost all lead times and target months compared to the anomaly persistence model. Fig. 1 (f) shows the ACC residuals between IceMamba-6 and the Anomaly Persistence model. The IceMamba model exhibits a more pronounced ACC decline with increasing lead time in August and September. However, IceMamba shows a significantly higher ACC during the melting season, indicating that IceMamba retains a significant predictive advantage in this period, which may reflect its ability to capture complex meteorological patterns that are not as effectively represented by simpler persistence models.

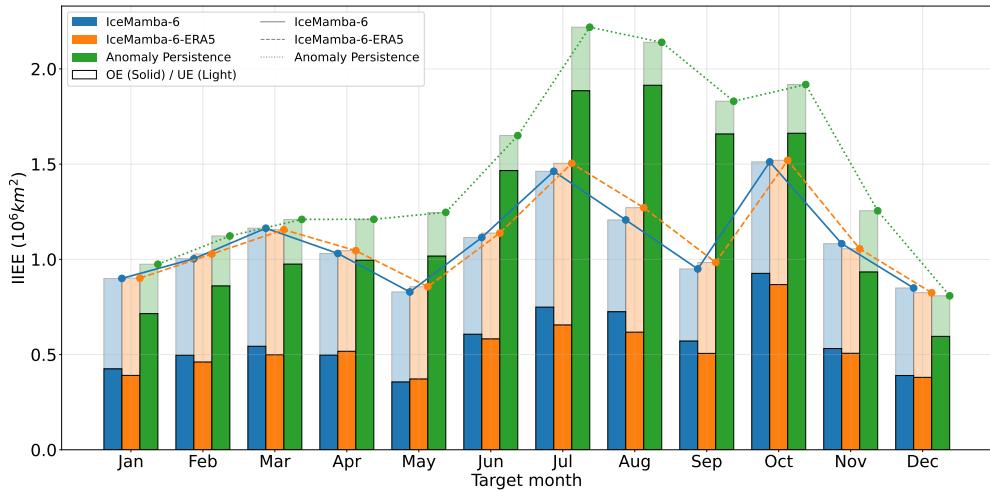


Fig. 2 Seasonal cycle of IIEE and its two components, overestimated error (OE) and underestimated error (UE), during the test period from sea ice forecast of IceMamba-6, IceMamba-6-ERA5, and Anomaly Persistence, averaged over six lead times.

Performance in Sea Ice Edge Forecast. Figure 2 illustrates the seasonal cycle of IIEE and its components (OE and UE) for IceMamba-6, IceMamba-6-ERA5, and Anomaly Persistence models, averaged across six lead times, with each value corresponding to the target month. Compared to the period from November to April, the IIEE of Anomaly Persistence has increased significantly from July to October. It may be caused by the complexity of the sea ice system from June to October. Sea ice is thinner between June and September and is more susceptible to atmospheric influences, significantly increasing uncertainty about sea ice changes.[52, 53]. From December to May, the sea ice is thicker and more stable, offering greater resistance to dynamical and thermodynamical effects, resulting in relatively minor interannual changes in the sea ice edge, making the IIEE of Anomaly Persistence relatively lower during this period. The IceMamba framework (IceMamba-6 and IceMamba-6-ERA5) shows a similar seasonal forecast pattern to Anomaly Persistence. However, IceMamba framework

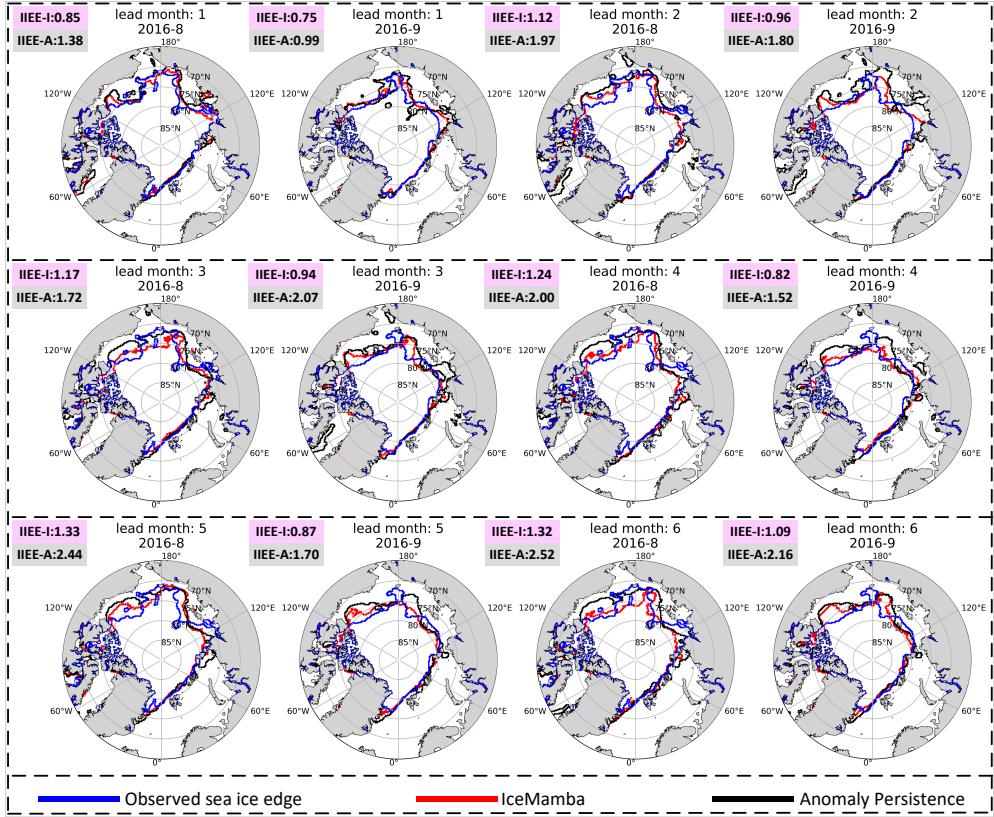


Fig. 3 Comparative visualization of predicted sea ice edge boundaries (colored contours) and Integrated Ice-Edge Error (IIEE) for August-September 2016 forecasts generated by IceMamba-6 (red) and the Anomaly Persistence benchmark (black), evaluated across 1 to 6 month lead times.

demonstrates significantly lower mean IIEE than Anomaly Persistence, particularly in summer and autumn. This improvement suggests that IceMamba framework effectively learns the complex, nonlinear relationships between SIC and reanalysis variables, providing a more accurate forecast of the sea ice edge during periods of higher variability. Moreover, the OE and UE distributions differ notably between the two models. Anomaly Persistence consistently shows significantly higher OE than UE throughout all 12 months, indicating a tendency to overestimate SIE (a limitation that could be partially mitigated by incorporating a damped anomaly persistence approach [43]). In contrast, the IIEE of IceMamba framework aligns with seasonal variations, with OE and UE balancing more closely and adjusting to seasonal ice changes, providing a less biased but seasonally variable error distribution.

Comparative analysis between IceMamba-6 and IceMamba-6-ERA5 shows that IceMamba-6, which integrates mixed layer depth (mld) and ocean heat content (ohc), shows a more noticeable reduction in the IIEE during July–October compared to other

months. Although IceMamba-6 shows slightly lower IIEE values than IceMamba-6-ERA5 in a few months, its overall average IIEE remains lower. This performance gap highlights the crucial regulatory role of subsurface oceanic processes in Arctic sea ice forecast: Autumn mixed-layer deepening facilitates vertical heat transport from subsurface reservoirs, and accumulated ohc in marginal ice zones during summer creates delayed thermal forcing on basal ice melt. In contrast, the sst-restricted model fails to capture sub-mixed-layer heat flux processes, resulting in a systematic overestimation of sea ice extent during the late melt season. These findings quantitatively validate the necessity of incorporating subsurface ocean thermodynamics in sea ice forecasting systems.

Fig. 3 compares the performance of IceMamba-6 and Anomaly Persistence in forecasting sea ice boundaries for August and September 2016. The results indicate that IceMamba significantly reduces the mean Integrated Ice Edge Error (IIEE) during peak melting periods, demonstrating enhanced accuracy in capturing ice edge position and morphological changes. The IIEE of both methods increases with lead time; however, the error growth rate of Anomaly Persistence is noticeably higher. The figure also reveals that for extended forecast periods, Anomaly Persistence often exhibits substantial deviations of the ice edge from observations, whereas IceMamba more effectively tracks the actual ice retreat, leading to lower overall errors. In summary, the experimental results in Fig. 3 confirm IceMamba’s advantage in reducing mean IIEE and highlight its ability to maintain lower errors in long-term forecasts and capture local details demonstrating improvements in sea ice boundary prediction.

Skill comparison with dynamical and statistical models. To comprehensively assess the forecast performance of the IceMamba architecture, we adopted the evaluation benchmark proposed by [43], facilitating a systematic evaluation of various dynamical, statistical (including DL models) for September Arctic sea ice. Evaluation metrics for assessing SIC forecast performance include RMSE, ACC, and IIEE. RMSE and ACC metrics focus on regions where the standard deviation of September SIC exceeds 10%, as this region exhibits significant variability, making it ideal for evaluating model performance.

To match the benchmark, IceMamba-4 and a retrained version of IceMamba-1, named IceMamba-1-only-SIC (using only SIC as input), were chosen for skill evaluation. IceMamba-1-only-SIC employs a recurrent-based forecast method. Specifically, with a forecast initiation date of June 1, 2001, the monthly mean SIC from June 2000 to May 2001 is used to predict the monthly mean SIC for June 2001. This prediction is then incorporated into the input data to forecast the monthly mean SIC for July 2001, and the process is repeated sequentially to forecast the monthly mean SIC for September 2001. While IceMamba-4 can directly predict the monthly mean SIC for each of the next four months in a single step. Moreover, both models undergo annual recalibration through a temporally segregated rolling-window framework. Let Y denote the target prediction year. The training set spans January 1979 to December of year $Y-5$ (e.g., 1979-1996 for $Y=2001$), while the validation set occupies January $Y-4$ to December $Y-1$ (1997-2000 for $Y=2001$). When forecasting for year $Y+1$, the training window extends to December $Y-4$ (1979-1997 for $Y+1=2002$), with the validation window advancing to span January January $Y-3$ to December Y (1998-2001). Crucially, the

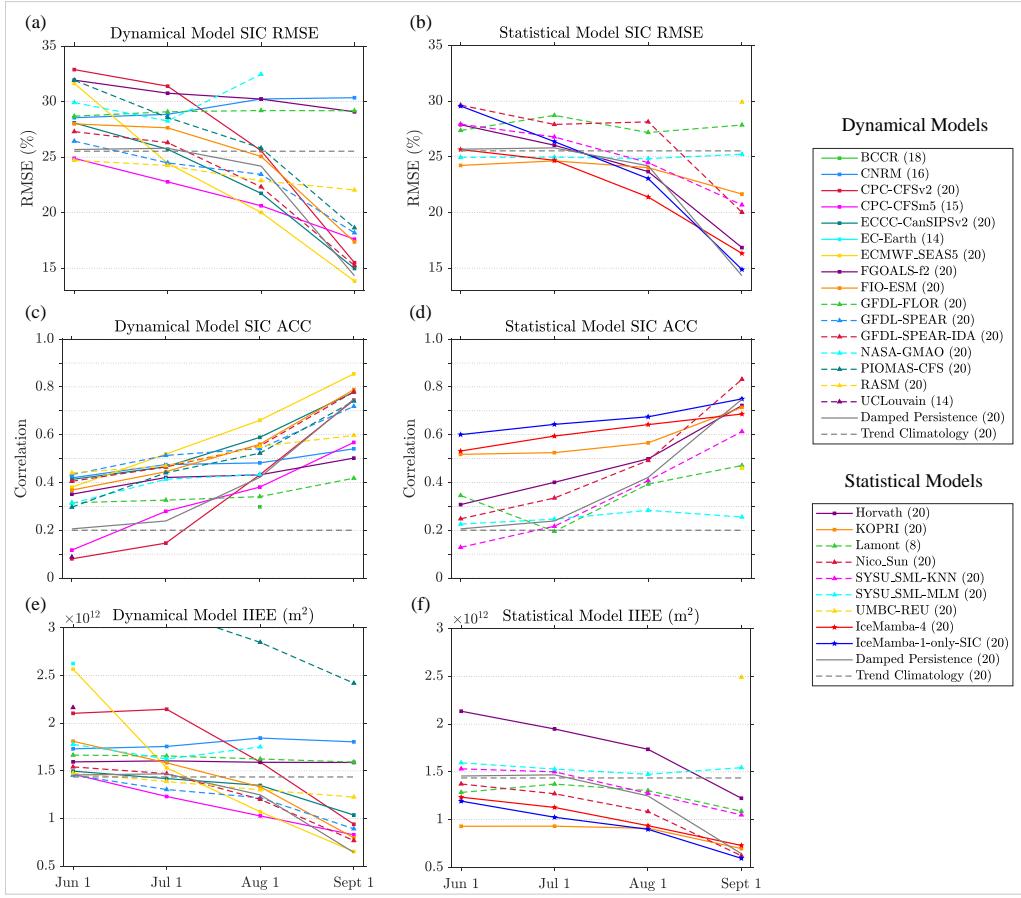


Fig. 4 RMSE, ACC, and IIEE for September SIC forecast (2001–2020). (a), (c), (e) show results for dynamical models, and (b), (d), (f) for statistical models. RMSE and ACC are averaged over regions where SIC standard deviation $> 10\%$. Models are color-coded, with reference forecasts in grey. Skill metrics are shown for each initialization from June 1 to September 1. Bracketed numbers in the legend indicate the years of data each model contributed over 20 years.

291 training window exhibits progressive expansion (annual increments), whereas the vali-
292 dation window maintains a fixed four-year span, ensuring both temporal isolation and
293 evaluation consistency.

294 IceMamba-4 and IceMamba-1-only-SIC are benchmarked against 16 dynamical
295 models [40, 46, 54–74], 7 statistical models [18, 23, 26, 43, 75–78], and 2 reference
296 predictions (Damped Persistence [79], Trend Climatology). Since the statistical models
297 include several DL components, we also categorize IceMamba as part of the statistical
298 models. The evaluation results for all tested models are illustrated in Fig. 4. Bracketed
299 numbers in the legend indicate the years of data each model contributed over 20
300 years. All specific data of evaluation results for RMSE, ACC, and IIEE are recorded
301 in Supplementary Tables 2, 3, and 4, respectively.

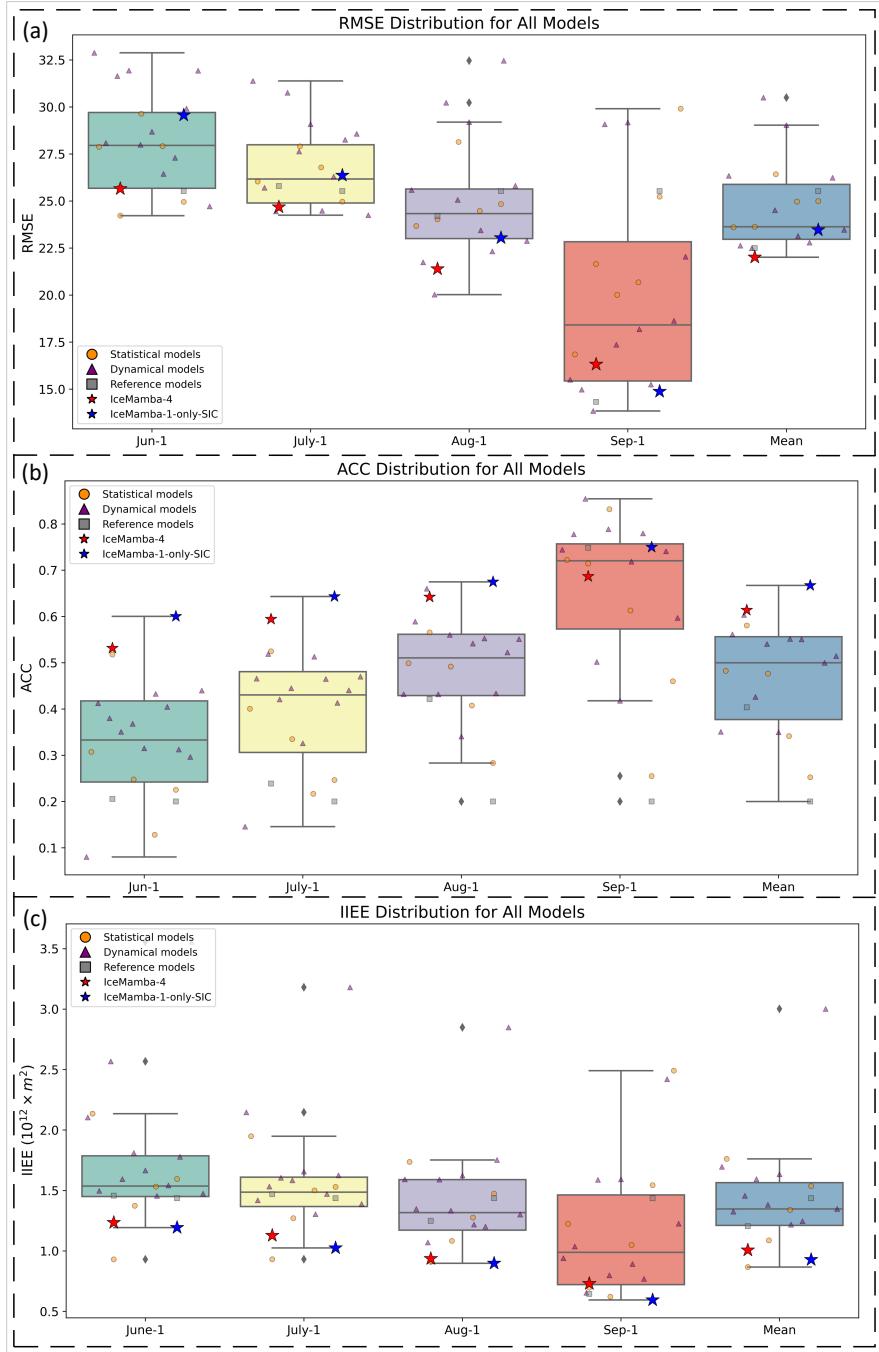


Fig. 5 Box plot illustration: RMSE, ACC, and IIEE for September SIC forecast (2001–2020) from models that contribute a full 20-year forecast. Panels (a), (b), and (c) show RMSE, ACC, and IIEE across models, averaged over regions with SIC standard deviation $> 10\%$. Models are color-coded: grey (reference), orange (statistical), and purple (dynamical). IceMamba-4 and IceMamba-1-only-SIC are highlighted in red and blue, respectively.

As shown in Fig. 4, among the 25 evaluated models, 5 models did not provide complete forecasts from 2001 to 2020. To ensure a fair comparison, the RMSE, ACC, and IIEE distributions for the models that contributed complete 20-year forecasts are displayed in the box plots of Fig. 5. As shown in Fig. 5 (a), IceMamba-4 achieved the lowest average RMSE (22.0086%) in all tested models. Moreover, only ECMWF_SEAS5 (dynamic model, 22.4916% RMSE) and IceMamba-4 exhibited lower errors than the damped persistence reference, highlighting their forecasting capabilities. For short-term forecasts (September 1 initiation), IceMamba-1-only-SIC attained the lowest RMSE among statistical models (14.8792%), slightly surpassed by ECMWF_SEAS5 (13.8444%). Although dynamical models generally perform better in short-term forecasts compared to statistical models, IceMamba-1-only-SIC achieves competitive performance (15.2924% RMSE), approaching ECMWF_SEAS5 (13.8444%). However, its reliance on a recurrent-based method leads to significant error accumulation as lead time extends, causing RMSE to rise sharply. Despite this limitation, its average RMSE (24.0335%) remains competitive against non-IceMamba statistical models.

The Anomaly Correlation Coefficient (ACC) evaluation reveals that IceMamba-1-only-SIC achieves superior performance (mean ACC=0.6604), exceeding all comparative models (Fig. 5b). Notably, while IceMamba-4 incorporates additional climatic inputs, it demonstrates reduced ACC values (0.6126) relative to both IceMamba-1-only-SIC across all initialization dates, suggesting that multi-month SIC forecasting in a single step can compromise ACC. For IIEE, IceMamba-1-only-SIC achieves the lowest IIEE at September 1 ($0.6152 \times 10^{12} m^2$) and August 1 ($0.9265 \times 10^{12} m^2$) initializations, yet exhibits the second-lowest mean IIEE. Notably, IceMamba-4 maintains the third-lowest mean IIEE despite showing marginally higher values than IceMamba-1-only-SIC at all lead times.

Overall, IceMamba achieves competitive performance compared to other tested models, particularly the IceMamba-4 variant, which achieves the lowest average RMSE in all evaluated models. For short-term forecasts initiated on September 1, IceMamba-1-only-SIC achieves the lowest RMSE among all statistical models, second only to the dynamical model ECMWF_SEAS5. In terms of the ACC and IIEE evaluations, IceMamba-1-only-SIC performs well in the one-month lead forecasts and IceMamba-4 maintains a clear advantage in overall performance. These findings collectively indicate that the IceMamba framework represents a significant advancement in the field of sea ice forecast, particularly in addressing the complexities and variability of Arctic climates.

Forecast performance in extreme September sea ice events. Arctic sea ice variability, especially during extreme low conditions at the end of summer, is a critical indicator of climate change. To better capture the early evolution of these anomalies, we include forecasts initialized on June 1, in addition to those on September 1, thus providing a more comprehensive evaluation of model performance during extreme sea ice events. Supplementary Tables 8 and 9 present a comprehensive comparison of IceMamba-4, IceMamba-1-only-SIC, statistical models, and dynamical models in predicting SIC during the extreme years (2012, 2016, 2019) based on the initialization date of September 1 and June 1, respectively. Since ACC requires data from multiple

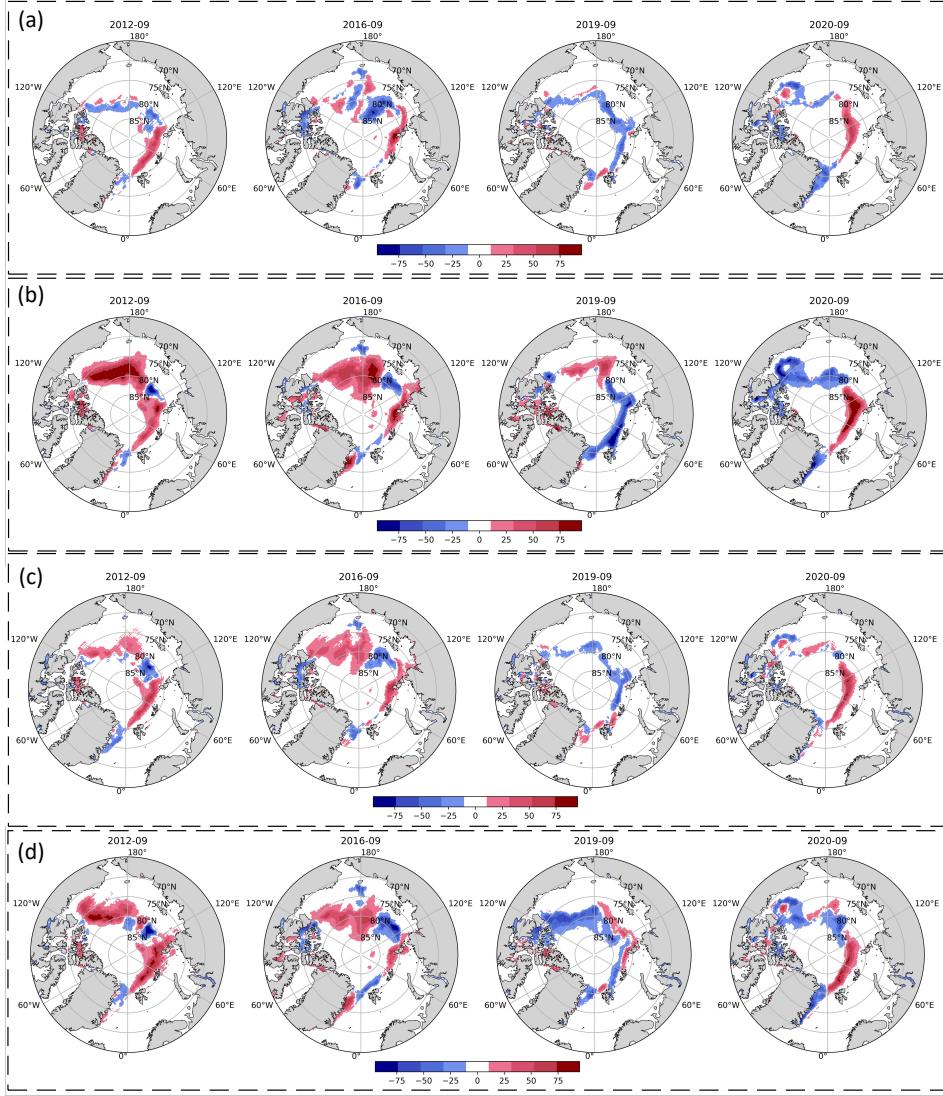


Fig. 6 Residuals of SIC predictions. (a-b) IceMamba-1-only-SIC at initialization dates September 1 and June 1; (c-d) IceMamba-4 at initialization dates September 1 and June 1; Residuals are calculated as model predictions minus NSIDC observational values. Color bars indicate residual magnitude, with blue/red tones representing systematic underprediction/overprediction respectively.

time points to effectively assess the correlation of anomaly patterns, calculating ACC at a single time point does not reflect the overall trend and consistency of model performance. Therefore we only compare the metrics of RMSE and IIEE in this Section. A further visualization of the performance of all evaluated models is provided by the box plot in Supplementary Fig. 2 and Supplementary Fig. 3, which illustrates the distribution of the RMSE and IIEE across the evaluated models.

353 For the forecasts initialized on September 1, the results in Supplementary Fig. 2
354 (a) highlight the outstanding performance of the IceMamba models. IceMamba-1-
355 only-SIC achieved the lowest RMSE in all non-IceMamba statistical models during
356 each extreme year. Compared with dynamical models, only a few dynamical models
357 reached lower RMSE than IceMamba-1-only-SIC. Notably, Dynamical models excel
358 in short-term Arctic sea ice prediction primarily due to their capacity to assimilate
359 near-real-time observational data into physically consistent frameworks. This synergy
360 between data assimilation and physical principles ensures high-fidelity initial condi-
361 tions, whereas purely data-driven approaches rely on historical training data that may
362 not capture rapidly evolving initial states. In September 2019, IceMamba-1-only-SIC
363 recorded an RMSE of 11.72%, the best performance observed across all evaluated
364 models. As shown in Supplementary Fig. 2 (a) (b), the IIEE metrics corroborate
365 the effectiveness of the IceMamba. IceMamba-1-only-SIC achieved the lowest IIEE in
366 most extreme years, significantly outperforming many other models, thereby reinforc-
367 ing its reliability in predicting the sea ice edge. In comparison, IceMamba-4, which
368 forecasts monthly averaged SIC maps over the next four months in one step, demon-
369 strates higher IIEE than IceMamba-1-only-SIC. This difference can be attributed to
370 the increased uncertainty inherent in long-term forecasts.

371 For forecasts initialized on June 1, a notable performance divergence emerged
372 between the two model architectures. As shown in Supplementary Fig. 3 (a) and (b),
373 IceMamba-1-only-SIC exhibited marked degradation in predictive skill. In contrast,
374 IceMamba-4 maintained stable forecasting capabilities without significant performance
375 decay. Particularly noteworthy are its achievements: attaining the lowest RMSE
376 (18.73%) in 2019 and simultaneously achieving both minimal RMSE and IIEE met-
377 rics in 2020 (20.92%) among all dynamical and statistical models, demonstrating
378 reliable forecast performance. Notably, the Great Arctic Cyclone of August 2012
379 preconditioned extreme sea ice loss prior to the September initialization date. This
380 implies that the initial conditions fed to the models already encapsulated extreme
381 anomalies. While IceMamba-1-only-SIC struggled to maintain accuracy under such
382 preconditioned extremes - likely due to error accumulation in its recursive frame-
383 work - IceMamba-4's superior performance suggests its architectural design effectively
384 decouples initial condition dependencies through advanced feature extraction.

385 The residual of predicted and observed SIC is shown in Fig. 6. In general, the fore-
386 cast errors in Sep 2012, 2019, and 2020 are mainly concentrated on marginal regions.
387 Since 2012 and 2020 are the years with the lowest and second lowest SIE, respectively,
388 our model generally overestimated SIC in these years. Comparative residual analysis
389 between the two models shows nuanced temporal dependencies. When initialized on
390 September 1, IceMamba-1-only-SIC demonstrates superior short-term predictive skill,
391 evidenced by its narrower residual distribution range compared to IceMamba-4. This
392 advantage, however, diminishes with extended lead times. Under June 1 initialization
393 conditions, IceMamba-1-only-SIC exhibits significantly amplified residuals (indicated
394 by intensified color gradients), while IceMamba-4 maintains relatively stable error
395 patterns. This divergence is particularly pronounced in the extreme ice loss years of
396 2012 and 2020. Overall, these results indicate that the IceMamba models are profi-
397 cient in forecasting SIC under extreme climate conditions. This capability is crucial

398 for enhancing our understanding of Arctic dynamics. By accurately forecasting SIC,
399 the IceMamba models provide a vital tool for researchers and policymakers aiming to
400 formulate effective strategies in response to the evolving climate landscape.

401 **Explainability of IceMamba.** Achieving Explainability in DL models is crucial for
402 understanding the mechanisms behind their forecasts. We implement a permutation-
403 based Explainability framework. For each variable, we randomly shuffle its 2D input
404 fields across temporal dimensions while keeping other variables unchanged. The degra-
405 dation in IceMamba’s forecast skill (quantified by MAE increase) after permutation
406 systematically reveals each variable’s relative importance across forecast months and
407 lead times. Fig. 7 (a) and (b) present heat maps showing the changes in MAE after the
408 permutation of specific variables. The y-axis denotes the types of permuted variables,
409 with numbers in brackets indicating the input lag.

410 Among all variables, SIC (1), reflecting the SIC with one lag month, exerts the most
411 substantial influence in one-month forecasts, highlighting the predictive significance of
412 sea ice persistence for short-term changes. As lead time increases, reliance on recent
413 SIC data diminishes, with the model increasingly utilizing last year’s SIC value for
414 the same target month corresponding to the forecast month, demonstrating a clear
415 seasonal alignment. The alignment is evident in the influence shift with lead time.
416 At one-month and two-month lead times, SIC (1) is the most influential variable,
417 while SIC (12) and SIC (11) are the second most influential variables, corresponding
418 to the forecast target month for each respective lead time from the previous year.
419 By three months, SIC (10) becomes the primary influence, while the effect of SIC
420 (1) decreases. This pattern continues with longer lead times, where each step reflects
421 a significant impact from the same month’s SIC in the previous year. This pattern
422 demonstrates the model’s capacity to implicitly capture and utilize the seasonal cycle
423 of SIC by focusing on historical data relevant to the forecasted month. Notably, this
424 seasonal alignment occurs without explicit temporal coding, suggesting the model has
425 developed an internalized understanding of annual SIC variations.

426 For most non-SIC climate parameters, the model exhibits maximum sensitivity to
427 data from one lag month. However, u10 (zonal wind at 10 hPa) stands as an excep-
428 tion. Unlike near-surface wind parameters (u10m and v10m), which directly influence
429 sea ice through surface drag, u10 cannot physically interact with sea ice. Despite this,
430 u10 shows a strong and sustained influence across three lag months, particularly from
431 July to December, and impacts all lead times. This unique sensitivity to u10 sug-
432 gests the model’s potential recognition of stratospheric drivers. However, this apparent
433 stratospheric linkage raises the question of whether it is primarily driven by physical
434 mechanisms or by the climate change signal, with anthropogenic sea ice loss since 1979
435 weakening the stratospheric polar vortex [44, 45].

436 To answer this question, We retrain the model using the original training data
437 with only u10 detrended to remove long-term climate trends. The results in Sup-
438 plementary Fig. 4 showed that IceMamba’s sensitivity to u10 significantly decreased
439 after removing the trend, indicating that the relationship could be driven by shared
440 anthropogenic trends, rather than direct causal effects of stratospheric wind. Notably,
441 training on detrended anomaly ua10 led to a decline in model performance (Supple-
442 mentary Table. 10), indicating that long-term climate trends in ua10 provide valuable

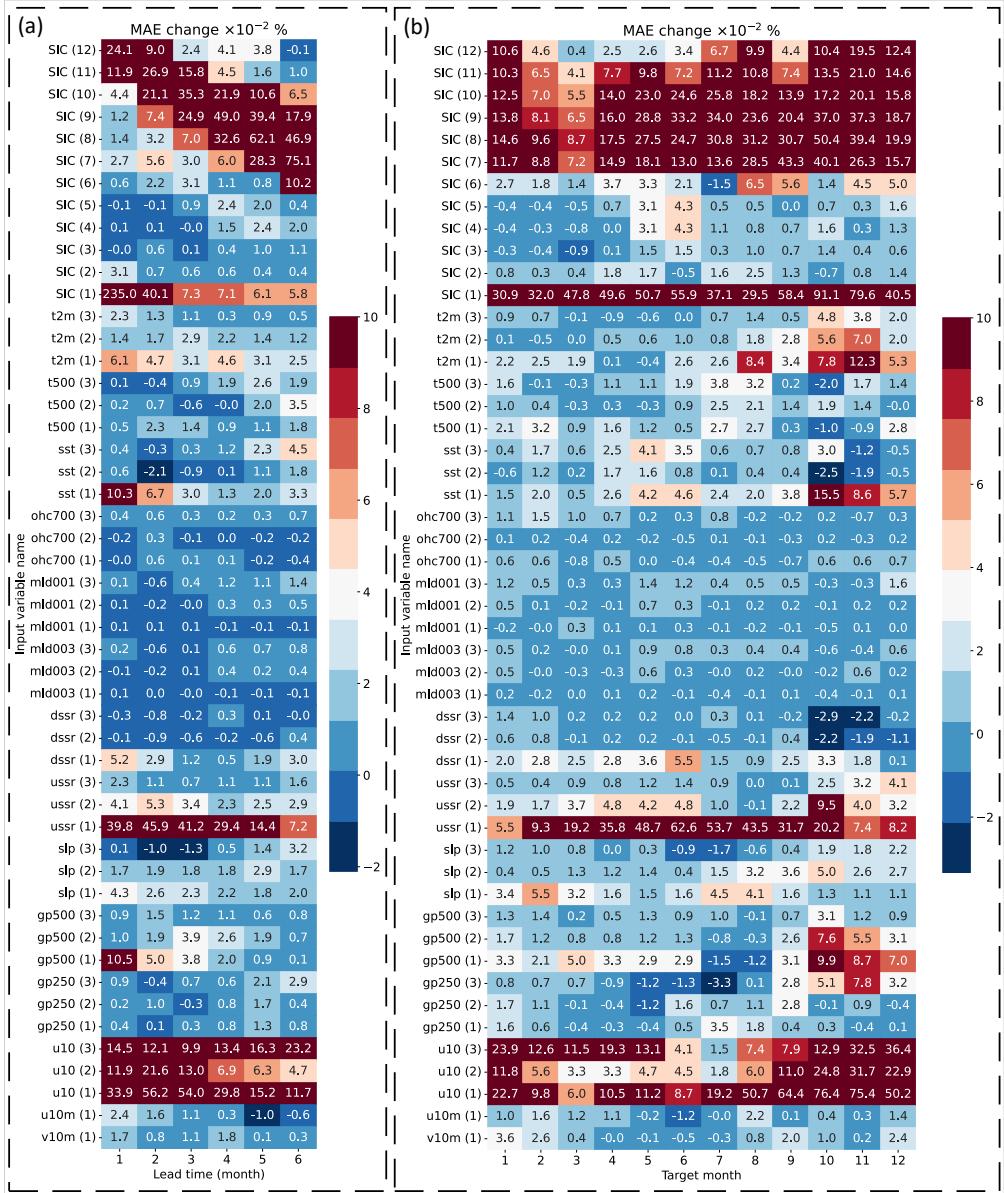


Fig. 7 Heatmaps from the permute-and-predict method: (a) Mean MAE change for each input variable in every lead time. (b) Mean MAE change for each input variable in every target month. MAE averaged over 10 random seeds and 84 forecast months (2016-2022). Bracketed numbers indicate the input lag (month)

predictive signals. Overall, The sharp reduction in sensitivity to detrended u10 and the decline in predictive skill when using anomalies underscores the role of long-term human-driven sea ice loss and its cascading impacts on the stratospheric polar vortex. While the model successfully identifies shared trends between u10 and sea ice variability, this relationship mainly reflects a co-evolution of anthropogenic forcing rather than a mechanistic stratosphere-ice coupling. This highlights the critical importance of disentangling climate change signals from intrinsic dynamical processes in attribution studies. Furthermore, the retained predictive value of long-term trends suggests that anthropogenic forcing provides a key "memory" for seasonal sea ice forecasts, even as it complicates the interpretation of causal drivers in coupled climate systems.

ussr (1), representing upward surface solar radiation with one lag month, demonstrates a noticeable and consistent impact across all lead times, with a peak influence specifically during the Arctic summer months. This period corresponds to the seasonal peak in solar radiation and enhanced ice melt, suggesting that the model captures the role of radiative fluxes in driving the summer sea ice melt through the albedo effect. The model's sensitivity to ussr (1) also reflects an understanding of the seasonal dependency of radiative forcing and its critical role in Arctic ice melt dynamics. In contrast, dssr representing downward surface solar radiation, has a weaker influence on sea ice prediction compared to ussr. This difference likely arises from how these parameters interact with ice-albedo feedback. While dssr provides the initial solar energy input, ussr captures the net radiative balance after reflection by ice. In the Arctic, high-albedo ice and snow reflect most of the incoming solar radiation. The ussr accurately measures this reflection process. Thus, the model's heightened sensitivity to ussr suggests it recognizes the critical role of the albedo-driven feedback loop in amplifying seasonal ice melt.

For the atmospheric temperature variable t2m (1), representing 2m air temperature with a one-month lag period. its predominant influence on August-November forecasts demonstrates critical regulatory effects on autumn predictability. This aligns with the Arctic's seasonal transition from ice-melt to freeze-up, during which delayed surface heat fluxes—driven by declining albedo and amplified ice-ocean-atmosphere feedbacks—propagate thermal anomalies into the boundary layer. In contrast, the model exhibits lower sensitivity to t500 relative to t2m, reflecting the model's limited sensitivity to upper atmosphere temperature.

For ocean variables, sst (1), representing sea surface temperature with one lag month, has the strongest influence, particularly from October to December, suggesting that the model captures the thermal inertia of the ocean in delaying ice formation beyond the melt season. While the model has lower sensitivity to ohc and mld compared to sst and other atmospheric variables. Interestingly, the sensitivity to ohc and mld is mainly concentrated in the third lag month, which is significantly different from other variables. This delayed response reflects the deeper ocean memory, where subsurface processes, such as the release of accumulated heat from the deeper ocean layers and mixed-layer dynamics, gradually influence ice-ocean-atmosphere interactions. The model successfully distinguishes between the dynamics of subsurface processes and those of other variables, capturing the delayed effects of deeper oceanic memory. However, its relatively low sensitivity to these processes suggests that while the

488 model acknowledges their importance, it faces challenges in fully capturing the slow,
489 multi-month thermodynamic adjustments involved.

490 For pressure variables, gp500 (1), representing 500 hPa geopotential height with
491 one lag month, demonstrates a notable effect across the one and two-month lead time
492 during autumn and early winter (October to December). This pattern suggests that
493 the model may implicitly recognize the role of mid-tropospheric pressure systems in
494 modulating atmospheric circulation patterns, which can impact ice formation and
495 stability during these months. In contrast, slp (1) and gp250 (1), representing sea
496 level pressure and 250 hPa geopotential height with one lag month, respectively, show
497 relatively weaker impacts, reflecting the model's limited sensitivity to surface and
498 upper-level pressure variations in shaping seasonal sea ice changes.

499 3 Discussion

500 In this study, we introduce IceMamba, a novel DL architecture designed for the sea-
501 sonal forecast of SIC over the pan-Arctic region. Experimental results indicate that
502 the state space model is particularly well-suited for the field of sea ice forecast. The
503 inherent capacity of Vision state space block (VSSB) to capture long-range dependen-
504 cies enables them to effectively represent the complex dynamics of sea ice, which are
505 influenced by an array of climatic factors over extended periods. This characteristic
506 is essential for understanding the fluctuations and variability of sea ice in the Arctic,
507 allowing for a more comprehensive analysis of the mechanisms driving these changes.
508 The integration of the Residual Efficient State Space Block (RESSB) further enhances
509 the forecast performance. By incorporating Efficient Channel Attention (ECA) and a
510 residual branching structure, RESSB optimizes feature selection, enabling the model
511 to prioritize the most relevant climatic signals. This optimization aligns well with the
512 unique challenges of sea ice forecasting. To our knowledge, this is the first study to
513 apply state space models to sea ice forecasting.

514 In a comprehensive comparison with dynamic and statistical models, IceMamba-4
515 achieved an average RMSE of 22.0086%, positioning it as the leading model among
516 all evaluated models. For the short-term forecast, IceMamba-1-only-SIC recorded the
517 lowest RMSE of 15.2924% under the initialization date of September 1. ACC analy-
518 sis revealed that IceMamba-1-only-SIC outperformed all models with a mean ACC of
519 0.6604, whereas IceMamba-4 had a slightly lower ACC of 0.6126, indicating the inher-
520 ent trade-offs in forecasting multiple months consecutively. For IIEE, IceMamba-4 and
521 IceMamba-1-only-SIC have lower average IIEE than all dynamical models. IceMamba-
522 1-only-SIC had the lowest values at initialization dates of September and August, but
523 its performance deteriorated significantly with longer lead times, reflecting a mean
524 IIEE of $0.9459 \times 10^{12} m^2$, which was the second lowest among all evaluated models.

525 IceMamba demonstrates robust forecasting capabilities for extreme Arctic Septem-
526 ber sea ice, with IceMamba-4 achieving the lowest RMSE (11.72% in 2019) and stable
527 performance under June-initialized long-term predictions, while IceMamba-1 excels in
528 short-term accuracy. The ability to accurately forecast SIC, even during significant sea
529 ice decline, makes IceMamba a vital tool for understanding Arctic dynamics and aiding

530 researchers and policymakers in responding to climate change. Moreover, The interpretability experiments demonstrated that IceMamba retains a clear understanding of
531 the relationships between input climate features and output predictions, confirming
532 the model's efficacy in revealing the underlying mechanics driving sea ice dynamics.
533 This transparency enhances the model's utility in both research and policy-making
534 contexts.

535 Despite its strengths, this study still has some limitations. One limitation is that
536 IceMamba adopts a channel fusion approach, which couples all input data into one
537 dimension. Although RESSB can effectively capture the correlation between data from
538 different channels, channel fusion requires all data to use the same grid method. This
539 re-gridding process may introduce inherent biases in the data preparation process.
540 Channel fusion also fixes the spatial range of climate data in the Arctic region, but
541 the impact of atmospheric variables is global, which may also limit the prediction
542 effect of IceMamba. Additionally, the spatial resolution of the input data (25 km)
543 might not capture fine-scale processes that could influence sea ice dynamics, such as
544 localized ocean currents and small-scale atmospheric phenomena. Another limitation
545 pertains to the generalizability of the model. While IceMamba has shown commendable
546 performance in the pan-Arctic region, its applicability to other regions, such as the
547 Antarctic, remains to be tested.

548 To address these limitations, future research could explore alternative methods for
549 data fusion that mitigate the biases introduced by re-gridding. Additionally, future
550 work could investigate the incorporation of global atmospheric variables more explicitly
551 into the IceMamba model. By integrating data that captures global atmospheric
552 patterns and their influences on the Arctic, the model could potentially improve its
553 forecast performance. Expanding the model to include physical processes explicitly,
554 through hybrid modeling approaches that combine physical and data-driven methods
555 [80–83], thereby enhancing its robustness and interpretability. Testing IceMamba
556 in different geographical contexts, such as the Antarctic, would provide insights into
557 its versatility and potential adaptations required for different sea ice regimes.

559 4 Method

560 **Experimental data.** The datasets used in this research comprise observational SIC
561 and observational climate reanalysis data. The SIC data is obtained from NOAA/N-
562 SIDC Climate Data Record of Passive Microwave Sea Ice Concentration version
563 4 product, which focuses on pan-Arctic region covering 448×304 grids (about
564 39.36°N – 89.84°N , 180°W – 180°E) with a spatial resolution of 25 km. This type of grid
565 format is also known as the Equal Area Extensible Earth 2 (EASE2) Grid. The prod-
566 uct provides passive-microwave-derived SIC estimates conforming to NOAA Climate
567 Data Record (CDR) standards, combining the NASA Team (NT) [84] and NASA Boot-
568 strap [85] (BT) algorithms, which estimate SIC from passive microwave brightness
569 temperatures at various frequencies and polarizations. The CDR product refines these
570 estimates by adjusting algorithm coefficients for each sensor, ensuring consistency in
571 daily and monthly SIC time series. It began with NASA Nimbus-7 SMMR data in
572 1978, and brightness temperature input data are processed and incorporated into the

record every three to six months. The reanalysis variables utilized in this study are derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis at a resolution of 0.25° [39]. ERA5 covers the global climate from January 1940 up to the present day, offering detailed insights into atmospheric, land, and oceanic reanalysis variables on an hourly basis. We utilized ERA5 monthly averaged datasets spanning from 1979 to 2022 for surface variables on single levels and for upper air variables on pressure levels [86, 87]. Moreover, we incorporated ocean reanalysis data from the ORAS5 product to further investigate subsurface oceanic processes. ORAS5 provides a global ocean reanalysis, delivering monthly averaged fields over the period from 1958 to the present. The inclusion of ORAS5 data enables a more comprehensive analysis of the ocean's thermal structure and its interactions with the atmosphere [40].

To focus on sea ice grid points, before fitting the SIC data into IceMamba, the SIC values in points of land are set to 0. In part of the SIC data, information around the North Pole is absent because of the "polar hole" phenomenon. The "polar hole" refers to a certain region around the geographic North Pole where satellite data coverage is limited by satellite orbits. As technology advances, the coverage of polar holes continues to decrease. The SMMR instrument exhibited a pole hole from November 1978 to June 1987, characterized by an area of $1.19 \times 10^6 km^2$ and a radius measuring 611 km, situated at a high latitude of $84.5^\circ N$. Subsequently, the SSM/I instrument featured a more modestly-sized pole hole, active from July 1987 to December 2007. This hole, with a radius of 311 km and a latitude of $87.2^\circ N$, covered an area of $0.31 \times 10^6 km^2$. Presently, the SSMIS instrument boasts the smallest pole hole to date, with a radius of 94 km, a latitude of $89.18^\circ N$, and an area of $0.029 \times 10^6 km^2$, in operation since January 2008 [88]. The CDR addresses the polar "pole hole" issue by employing spatial interpolation. Initially, a temporal interpolation is applied using data from up to five days before or after to fill the pole hole locations. If such data points are unavailable, the pole hole is filled with the average SIC from surrounding grid cells.

All reanalysis datasets are re-gridded from a latitude-longitude grid to the EASE2 grid using bilinear interpolation. Due to the strong periodicity of meteorological data, some reanalysis variables were calculated as abnormal values to highlight the differences from climate values. These anomaly variables were determined during the training phase by deducting the climatological average for each respective target month from the observed values. In addition, we normalize the reanalysis variables by subtracting the mean and using the standard deviation calculated during the training year (1979-2010) to keep the values of each variable within a similar range, thereby improving the training stability of the model. The details of all input variables are shown in Supplementary Table 1.

Input variables of IceMamba. Inspired by the parameter selection and climate preprocessing in [27], we adopted a subset of their methods, but IceMamba's input data excludes linear trends, land mask, and initialization date encoding (cosine and sine of the initialization date index). To address potential subsurface ocean influences, we further incorporated mixed-layer depth (mld) and ocean heat content (ohc) anomalies from reanalysis datasets, which effectively reflect heat fluxes below the mixed layer,

critical for autumn sea ice variability. IceMamba uses 12 historical monthly mean SIC from NSIDC, and historical mean climate reanalysis sequences from ERA5 [39] and ORAS5 [40] for 1-month or 3-month lead times as input, capturing the spatiotemporal and ice-atmosphere-ocean coupling relationships within the sequence and generating predictions for the SIC over the next several months.

The input data includes six key groups: sea ice, atmospheric temperature, ocean, radiation, pressure, and wind. SIC represents the sea ice group. The atmospheric temperature group contains 2-metre air temperature (t2m) and 500 hPa air temperature (t500). The ocean group contains sea surface temperature (sst), Ocean heat content for the upper 300m (ohc300), Ocean heat content for the upper 700m (ohc700), Mixed layer depth 0.01 (mld001), and Mixed layer depth 0.03 (mld003). The radiation group contains upwards surface solar radiation (ussr) and downwards surface solar radiation (dssr). The pressure group contains 500 hPa geopotential height (gp500) and 250 hPa geopotential height (gp250). Finally, the wind group contains 10-metre X-direction wind speed (u10m), 10-metre Y-direction wind speed (v10m), and 10 hPa zonal wind speed (u10). Details of all input data are shown in Supplementary Table 1.

These climate variables are chosen to capture key processes affecting sea ice dynamics, which encompass both thermodynamic and dynamic processes. Thermodynamic processes involve the exchange of heat between the sea ice, atmosphere, and ocean, driving the growth and melting of sea ice. For example, the atmospheric temperature, radiation, and ocean groups capture the energy balance critical for seasonal melt and growth cycles. Dynamic processes involve mechanical forces that cause sea ice to move, deform, and redistribute. For instance, u10m and v10m in the wind group directly influence sea ice drift, while the 10 hPa zonal wind is included to account for teleconnections between the stratospheric polar vortex and Arctic sea ice anomalies. The pressure group captures large-scale atmospheric circulation patterns that drive ice motion and deformation. These input variables allow IceMamba to model the key mechanisms that influence sea ice.

To avoid the effect of higher uncertainties in the late 1978 observation, only SIC data from 1979 to 2022 are used in this study. The temporal coverage for the training, valid, and test sets, respectively, spans from Jan 1979 to Dec 2010 (32 years), Jan 2011 to Dec 2014 (4 years), and Jan 2015 to Dec 2022 (8 years). Note that IceMamba uses the monthly average SIC of the previous 12 months as input, the test period corresponding to the test set is 2016-2022.

IceMamba: A state space model for sea ice forecasting. Recently, state space models (SSMs) have become a powerful tool for modeling continuous long sequences of data by representing system dynamics through latent variables. Mamba [41] significantly advances SSMs with a selection mechanism that enhances efficiency and performance, enabling them to handle global contextual information effectively. Mamba models long-range dependencies by formalizing discrete state-space equations into a recursive form and combining them with a specially designed structured reparameterization, along with hardware-aware optimizations, making it a promising alternative to the Transformer [38]. Building on Mamba, Vmamba [89] is tailored for computer vision tasks. It addresses the direction-sensitive problem[89] by using the Cross-Scan Module (CSM) to capture interrelations among image patches, significantly

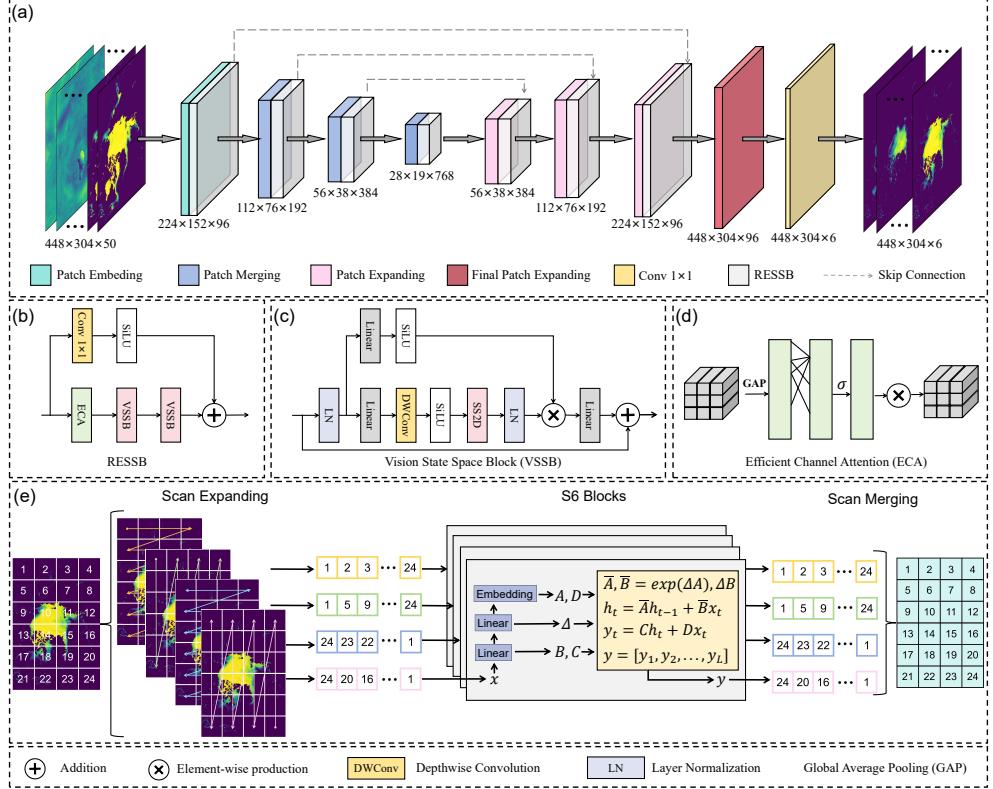


Fig. 8 Schematic diagram of IceMamba framework: (a) Structure of IceMamba-6. (b) Structure of Residual Efficient State Space Block (RESSB). (c) Structure of Visual State Space Blocks (VSSB) in RESSB. (d) Structure of Efficient Channel Attention (ECA) module in RESSB. (f) Illustration of the 2D-Selective-Scan (SS2D) used in VSSB.

enhancing performance in object detection [90, 91], segmentation [92, 93], and classification [94, 95]. VMamba excels in the efficient management of long-sequence data and the extraction of spatial and temporal features. Unlike CNNs, which primarily focus on local spatial information and struggle to capture long-range dependencies, Mamba seamlessly integrates these long-range dependencies with its advanced sequence processing capabilities. Furthermore, Mamba surpasses the traditional Transformer [38] model by optimizing computational efficiency, reducing memory usage, and maintaining high performance even with large datasets. Given that satellite and climate data are often presented in gridded formats akin to images, this makes VMamba particularly well-suited to address the complexities and demands of climate prediction tasks.

Leveraging the potential of SSMs in satellite data processing, we propose IceMamba, a novel DL framework employing an Advanced SSM to forecast monthly mean SIC for the next several months. As illustrated in Fig. 8(a), IceMamba adopts an encoder-decoder framework with multiple Residual Enhanced State Space Blocks (RESSB). Each RESSB (Fig. 8(b)) integrates two Vision State Space Blocks (VSSB), a

678 1×1 convolutional layer, and an Efficient Channel Attention (ECA) module (Fig. 8(d)).
679 ECA is added to RESSB to enhance the learning of different channel representations
680 and select key channels to prevent redundancy. This enables IceMamba to focus on the
681 key climate variables in the input data and prioritize the most influential lagged month
682 of the corresponding climate variable. In addition, a residual branch with 1×1 convo-
683 lution further improves pattern capture and information fusion. The VSSB (Fig. 8(c))
684 combines Layer Normalization [96], Linear layers, Depthwise convolution [97], and a
685 2D-Selective-Scan (SS2D) block [89]. While standard Mamba struggles with 2D visual
686 data, the SS2D overcomes this through its Cross-Scan Module (CSM) (Fig. 8(e)),
687 which scans image patches in four directions (top-left, bottom-right, top-right, bottom-
688 left) to generate independent sequences. These sequences are processed through the
689 Selective Scan Space State (S6) model [41] before fusion, enabling comprehensive
690 global receptive fields with minimal computational overhead. IceMamba employs patch
691 embedding to convert inputs into numerical representations, dynamically capturing
692 spatial relationships through CSM’s ordered sequence transformation and S6’s paral-
693 lel cyclic processing, eliminating positional encoding requirements. Spatial resolution
694 adjustments use patch merging/expanding instead of traditional pooling/upsampling:
695 merging concatenates adjacent patches to halve resolution while doubling channels,
696 whereas expanding splits patches to increase resolution, proportionally reducing chan-
697 nels (Sect. 4 details RESSB implementation). This design reduces computational
698 complexity while enhancing generalization across visual perception tasks.

699 The encoder module analyzes input data, exploiting spatiotemporal and ice-
700 atmosphere-ocean coupling relationships at different scales. This process produces a
701 set of feature maps, each capturing a different aspect of the input sequences. After
702 that, the decoder module gradually restores the scale of the feature maps, ultimately
703 outputting the monthly SIC for the next several months. The feature maps captured
704 by the encoder and decoder at the corresponding levels are fused by skip connection
705 to integrate the spatial dependencies at these scales. The final patch-expanding layer
706 increases the spatial resolution without reducing the number of channels. It ensures
707 that the output tensor retains the full channel information, which is important since
708 the channels encode temporal information. Following this layer, a 1×1 convolution is
709 applied to generate the final output. Reducing the number of channels at this stage
710 could result in the loss of crucial temporal features. Tanh activation is used in the
711 1×1 convolution layer. We find that using the Tanh function leads to better model
712 performance compared to the no activation function and the Sigmoid activation func-
713 tion in prediction. We hypothesize that this is because the Tanh function provides
714 a zero-centered distribution. This helps maintain gradient stability during training,
715 preventing issues like vanishing and exploding gradients. Additionally, the Tanh func-
716 tion better captures non-linear features, leading to improved model performance when
717 learning complex data patterns.

Residual efficient state space blocks. We proposed the Residual Efficient State Space Blocks (RESSB) as a modification of the VSSB for sea ice forecast. The RESSB

is formulated as:

$$\begin{aligned}\mathbf{F}_E^{N+1} &= \text{VSSB}(\text{VSSB}(\text{ECA}(\mathbf{F}_R^N))), \\ \mathbf{F}_C^{N+1} &= \text{SiLU}(\text{Conv}_{1 \times 1}(\mathbf{F}_R^N)), \\ \mathbf{F}_R^{N+1} &= \mathbf{F}_E^{N+1} \oplus \mathbf{F}_C^{N+1},\end{aligned}\tag{1}$$

where \mathbf{F}_R^N represents the feature map of the n-layer. \mathbf{F}_E^{N+1} and \mathbf{F}_C^{N+1} are the output features of the two branches in RESSB respectively. \mathbf{F}_R^{N+1} represents the final output feature of RESSB. $\text{VSSB}(\cdot)$, and $\text{ECA}(\cdot)$ denote the VSSB, and ECA operations as shown in Fig. 8(c) and Fig. 8(d), respectively. $\text{Conv}_{1 \times 1}(\cdot)$ represents a convolutional layer with a kernel size of 1. The ECA module incorporates an efficient channel attention mechanism that starts with Global Average Pooling (GAP) to aggregate features across each channel without any dimensionality reduction, thereby preserving original correspondence between channels. Subsequently, the ECA module utilizes 1D convolution to capture the interactions between local cross-channels. Finally, The ECA module computes the weights for each channel in a parameter-efficient manner and then activates them using a Sigmoid function to obtain the final channel attention weights.

Given the interactions between sea ice, atmosphere, and ocean data at different time scales, we leverage the powerful modeling capabilities of VSSB to capture long-range dependencies. To this end, we integrate all SIC, ocean, and atmosphere data into the input channels of IceMamba. However, it's worth noting that while VSSB excels at retaining long-range dependencies, it requires the introduction of a large number of hidden states. This tendency often leads to apparent channel redundancy [89]. Hence, we incorporate ECA into RESSB, which allows the block to focus on learning diverse channel representations, while the channel attention mechanism selects the most critical channels to prevent redundancy. By integrating the ECA, we can alleviate this issue and improve the block's overall performance. Moreover, we incorporated a residual branch [98] with a 1×1 convolutional layer to enhance the block's capability to capture complex patterns. The residual branch enables the module to learn residual information, while the 1×1 convolutional layer enables the block to better fuse the information from the residual branch.

Principles and application of the State space model for SIC forecast. The state space model (SSM) is inspired by linear time-invariant systems. As shown in Fig. 8 (e), the Cross-Scan Module (CSM) maps input data to four types of vectors, and each vector will be input into a corresponding Selective Scan Space State Sequential (S6) model. Assume that one of the vectors is $x(t) \in \mathbb{R}^L$, representing the input data to the S6 model, then an intermediate state $h(t) \in \mathbb{R}^N$ is used to map the input data to an output response $y(t) \in \mathbb{R}^L$, representing the output feature. This can be mathematically expressed using a linear ordinary differential equation (ODE):

$$\begin{aligned}h'(t) &= \mathbf{A}h(t) + \mathbf{B}x(t), \\ y(t) &= \mathbf{C}h(t) + \mathbf{D}x(t),\end{aligned}\tag{2}$$

752 where N is the state size, $\mathbf{A} \in \mathbb{C}^{N \times N}$ is the state transition matrix, $\mathbf{B} \in \mathbb{C}^N$ and
 753 $\mathbf{C} \in \mathbb{C}^N$ are the input and output projection parameters, and $\mathbf{D} \in \mathbb{C}^1$ is the skip
 754 connection.

755 For practical implementation in DL algorithms, the continuous-time SSMs can be
 756 discretized by the zero-order hold (ZOH) method, which allows for the conversion of
 757 continuous parameters \mathbf{A} and \mathbf{B} into discrete-time parameters $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$:

$$\begin{aligned}\bar{\mathbf{A}} &= e^{\Delta \mathbf{A}}, \\ \bar{\mathbf{B}} &= (e^{\Delta \mathbf{A}} - \mathbf{I})\mathbf{A}^{-1}\mathbf{B} \approx (\Delta \mathbf{A})(\Delta \mathbf{A})^{-1}\Delta \mathbf{B} = \Delta \mathbf{B},\end{aligned}\tag{3}$$

758 where \mathbf{I} is the identity matrix and Δ is the timescale parameter to transform the
 759 continuous parameters \mathbf{A} and \mathbf{B} to their discrete counterparts $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$. For the
 760 S6 model, time-variability and nonlinearity are introduced to the state-space model
 761 (SSM) by mapping the input data \mathbf{x} to the parameter Δ through a nonlinear function.
 762 Specifically, \mathbf{B} and \mathbf{C} also depend on the input \mathbf{x} , generated dynamically via the
 763 nonlinear mappings. This design allows the S6 model to have time-varying system
 764 parameters at each time step, enabling the state-space model to handle both time-
 765 variability and nonlinear relationships.

766 The discretized version of Eq. 2 can be rewritten in recurrent neural network (RNN)
 767 form as:

$$\begin{aligned}h_k &= \bar{\mathbf{A}}h_{k-1} + \bar{\mathbf{B}}x_k, \\ y_k &= \mathbf{C}h_k + \mathbf{D}x_k,\end{aligned}\tag{4}$$

768 where k represents the initialization month for prediction. x_k represents the input vec-
 769 tor at the k -th initialization month, which includes the feature for SIC and reanalysis
 770 variables for the past 12 months leading up to the initialization month. The out-
 771 put y_k represents the output feature for the predicted SIC for the k -th initialization
 772 month. The state h_k captures the temporal dependencies and system dynamics up to
 773 the initialization month k , serving as the state at the initialization month for predic-
 774 tion. Moreover, Eq. 2 can be transformed into a convolutional neural network (CNN)
 775 structure resulting in the following formulation [41]:

$$\begin{aligned}\mathbf{y} &= \mathbf{x} * \bar{\mathbf{K}}, \\ \bar{\mathbf{K}} &= [\mathbf{CB}, \mathbf{CAB}, \dots, \mathbf{CA}^{L-1}\bar{\mathbf{B}}],\end{aligned}\tag{5}$$

776 where $\bar{\mathbf{K}} \in \mathbb{R}^L$ is a structured convolutional kernel and $*$ denotes the convolution
 777 operation. This CNN-based representation enables parallel training of recurrent-based
 778 models and allows the model to learn and capture the spatial and temporal patterns
 779 in the SIC and climate data effectively.

780 **The detail of training sheme.** In the training phase, the network receives input
 781 from randomly selected batches of training data. The adaptive moment estimation

(Adam) [99] optimizer is employed to minimize the loss function. To ensure the model concentrates on sea ice within the pan-Arctic region, the loss function (MAE) is computed exclusively for non-land areas. A batch size of 1 is used with an initial learning rate of 0.001. To ensure the convergent and robust performance of the model, a learning rate schedule is defined for training. Specifically, the learning rate is decreased by 0.5 every 10 epochs, which helps the model avoid oscillations and overfitting in the later training stages. Additionally, an early stopping strategy of 10 epochs is implemented to prevent overfitting and enhance the model’s generalization. This strategy allows the model to stop training when the validation loss stops improving, which helps to prevent the model from overfitting the training data. IceMamba is developed in Python 3.7, utilizing the PyTorch deep learning framework. The entire computational process is executed on an Nvidia A100 GPU, which enables the IceMamba training to be completed in approximately four hours.

Data availability. The datasets used in this study include observational SIC data and climate reanalysis data, both accessible online. The SIC data is sourced from the NSIDC, available at <https://nsidc.org/data/g02202/versions/4>. The reanalysis data is derived from the ERA5 and ORAS5 datasets, with ERA5 single-level variables retrievable at <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means>, ERA5 pressure-level variables at <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressurelevels-monthly-means>, ORAS5 variables at <https://cds.climate.copernicus.eu/datasets/reanalysis-oras5?tab=overview>.

Code availability. The source code used for the design, training, and evaluation of the IceMamba model, as well as the scripts for data preprocessing and analysis, are available on <https://github.com/WeiWang31/IceMamba.git>. All experimental data and model weights are publicly available at <https://zenodo.org/records/14926245>.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (grant No. 42325604). The computations in this research were performed using the CFFF platform of Fudan University.

Author contributions. Wei Wang designed and trained IceMamba, and wrote the manuscript under the guidance of Lei Wang and Weidong Yang. Wei Wang handled the climate data downloading and preprocessing under the supervision of Guihua Wang and Ruibo Lei. Wei Wang conducted the multi-model comparison with the assistance of Lei Wang. Weidong Yang managed the project. All authors reviewed and provided feedback on the manuscript.

Competing interests. The authors declare no competing interests.

References

- [1] Liu, Z., Risi, C., Codron, F., He, X., Poulsen, C.J., Wei, Z., Chen, D., Li, S., Bowen, G.J.: Acceleration of western arctic sea ice loss linked to the pacific north american pattern. *Nature communications* **12**(1), 1519 (2021)

- 822 [2] Smith, D.M., Eade, R., Andrews, M., Ayres, H., Clark, A., Chripko, S., Deser,
823 C., Dunstone, N., García-Serrano, J., Gastineau, G., *et al.*: Robust but weak
824 winter atmospheric circulation response to future arctic sea ice loss. *Nature*
825 communications **13**(1), 727 (2022)
- 826 [3] Chripko, S., Msadek, R., Sanchez-Gomez, E., Terray, L., Bessières, L., Moine, M.-
827 P.: Impact of reduced arctic sea ice on northern hemisphere climate and weather
828 in autumn and winter. *Journal of Climate* **34**(14), 5847–5867 (2021)
- 829 [4] Cai, Q., Wang, J., Beletsky, D., Overland, J., Ikeda, M., Wan, L.: Accelerated
830 decline of summer arctic sea ice during 1850–2017 and the amplified arctic warm-
831 ing during the recent decades. *Environmental Research Letters* **16**(3), 034015
832 (2021)
- 833 [5] Chung, P.-C., Feldl, N.: Sea ice loss, water vapor increases, and their interac-
834 tions with atmospheric energy transport in driving seasonal polar amplification.
835 *Journal of Climate* **37**(8), 2713–2725 (2024)
- 836 [6] Overland, J.E., Ballinger, T.J., Cohen, J., Francis, J., Hanna, E., Jaiser, R., Kim,
837 B.-M., Kim, S.-J., Ukita, J., Vihma, T., *et al.*: How do intermittency and sim-
838 taneous processes obfuscate the arctic influence on midlatitude winter extreme
839 weather events? *Environmental Research Letters* **16**(4), 043002 (2021)
- 840 [7] Westen, R.M., Jacques-Dumas, V., Boot, A.A., Dijkstra, H.A.: The role of sea-
841 ice insulation effects on the probability of amoc transitions. *Journal of Climate*
842 (2024)
- 843 [8] Lannuzel, D., Tedesco, L., Van Leeuwe, M., Campbell, K., Flores, H., Delille, B.,
844 Miller, L., Stefels, J., Assmy, P., Bowman, J., *et al.*: The future of arctic sea-ice
845 biogeochemistry and ice-associated ecosystems. *Nature Climate Change* **10**(11),
846 983–992 (2020)
- 847 [9] Jung, T., Gordon, N.D., Bauer, P., Bromwich, D.H., Chevallier, M., Day, J.J.,
848 Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H.F., *et al.*: Advancing polar
849 prediction capabilities on daily to seasonal time scales. *Bulletin of the American*
850 *Meteorological Society* **97**(9), 1631–1647 (2016)
- 851 [10] Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C.O.,
852 Dunne, J.P., Griffies, S.M., Hallberg, R., Harrison, M.J., *et al.*: The gfdl global
853 ocean and sea ice model om4. 0: Model description and simulation features.
854 *Journal of Advances in Modeling Earth Systems* **11**(10), 3167–3211 (2019)
- 855 [11] Horvat, C., Jones, D.R., Iams, S., Schroeder, D., Flocco, D., Feltham, D.: The
856 frequency and extent of sub-ice phytoplankton blooms in the arctic ocean. *Science*
857 *advances* **3**(3), 1601191 (2017)
- 858 [12] Smith, G.C., Roy, F., Reszka, M., Surcel Colan, D., He, Z., Deacu, D., Belanger,

- 859 J.-M., Skachko, S., Liu, Y., Dupont, F., *et al.*: Sea ice forecast verification in
 860 the canadian global ice ocean prediction system. Quarterly Journal of the Royal
 861 Meteorological Society **142**(695), 659–671 (2016)
- 862 [13] Blanchard-Wrigglesworth, E., Cullather, R., Wang, W., Zhang, J., Bitz, C.: Model
 863 forecast skill and sensitivity to initial conditions in the seasonal sea ice outlook.
 864 Geophysical Research Letters **42**(19), 8042–8048 (2015)
- 865 [14] Schröder, D., Feltham, D.L., Flocco, D., Tsamados, M.: September arctic sea-ice
 866 minimum predicted by spring melt-pond fraction. Nature Climate Change **4**(5),
 867 353–357 (2014)
- 868 [15] Petty, A., Schröder, D., Stroeve, J., Markus, T., Miller, J., Kurtz, N.T., Feltham,
 869 D., Flocco, D.: Skillful spring forecasts of september arctic sea ice extent using
 870 passive microwave sea ice observations. Earth's Future **5**(2), 254–263 (2017)
- 871 [16] Gregory, W., Tsamados, M., Stroeve, J., Sollich, P.: Regional september sea
 872 ice forecasting with complex networks and gaussian processes. Weather and
 873 Forecasting **35**(3), 793–806 (2020)
- 874 [17] Wang, L., Yuan, X., Ting, M., Li, C.: Predicting summer arctic sea ice concen-
 875 tration intraseasonal variability using a vector autoregressive model. Journal of
 876 Climate **29**(4), 1529–1543 (2016)
- 877 [18] Yuan, X., Chen, D., Li, C., Wang, L., Wang, W.: Arctic sea ice seasonal prediction
 878 by a linear markov model. Journal of Climate **29**(22), 8151–8173 (2016)
- 879 [19] Wang, Y., Yuan, X., Bi, H., Bushuk, M., Liang, Y., Li, C., Huang, H.: Reassessing
 880 seasonal sea ice predictability of the pacific-arctic sector using a markov model.
 881 The Cryosphere **16**(3), 1141–1156 (2022)
- 882 [20] Wang, L., Yuan, X., Li, C.: Subseasonal forecast of arctic sea ice concentration
 883 via statistical approaches. Climate Dynamics **52**, 4953–4971 (2019)
- 884 [21] Wang, Y., Yuan, X., Bi, H., Ren, Y., Liang, Y., Li, C., Li, X.: Understanding
 885 arctic sea ice thickness predictability by a markov model. Journal of Climate
 886 **36**(15), 4879–4897 (2023)
- 887 [22] Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J.J., Déqué, M.,
 888 Doblas-Reyes, F.J., Fučkar, N.S., Germe, A., Hawkins, E., Keeley, S., *et al.*: A
 889 review on arctic sea-ice predictability and prediction on seasonal to decadal time-
 890 scales. Quarterly Journal of the Royal Meteorological Society **142**(695), 546–561
 891 (2016)
- 892 [23] Chi, J., Kim, H.-c.: Prediction of arctic sea ice concentration using a fully data
 893 driven deep neural network. Remote Sensing **9**(12), 1305 (2017)

- 894 [24] Kim, J., Kim, K., Cho, J., Kang, Y.Q., Yoon, H.-J., Lee, Y.-W.: Satellite-based
 895 prediction of arctic sea ice concentration using a deep neural network with multi-
 896 model ensemble. *Remote Sensing* **11**(1), 19 (2018)
- 897 [25] Kim, Y.J., Kim, H.-C., Han, D., Lee, S., Im, J.: Prediction of monthly arctic
 898 sea ice concentrations using satellite and reanalysis data based on convolutional
 899 neural networks. *The Cryosphere* **14**(3), 1083–1104 (2020)
- 900 [26] Chi, J., Bae, J., Kwon, Y.-J.: Two-stream convolutional long-and short-term
 901 memory model using perceptual loss for sequence-to-sequence arctic sea ice
 902 prediction. *Remote Sensing* **13**(17), 3413 (2021)
- 903 [27] Andersson, T.R., Hosking, J.S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell,
 904 C., Law, S., Jones, D.C., Wilkinson, J., Phillips, T., *et al.*: Seasonal arctic sea ice
 905 forecasting with probabilistic deep learning. *Nature communications* **12**(1), 5124
 906 (2021)
- 907 [28] Zhu, Y., Qin, M., Dai, P., Wu, S., Fu, Z., Chen, Z., Zhang, L., Wang, Y., Du,
 908 Z.: Deep learning-based seasonal forecast of sea ice considering atmospheric con-
 909 ditions. *Journal of Geophysical Research: Atmospheres* **128**(24), 2023–039521
 910 (2023)
- 911 [29] Choi, M., De Silva, L.W.A., Yamaguchi, H.: Artificial neural network for the
 912 short-term prediction of arctic sea ice concentration. *Remote Sensing* **11**(9), 1071
 913 (2019)
- 914 [30] Fritzner, S., Graversen, R., Christensen, K.H.: Assessment of high-resolution
 915 dynamical and machine learning models for prediction of sea ice concentration
 916 in a regional application. *Journal of Geophysical Research: Oceans* **125**(11),
 917 2020–016277 (2020)
- 918 [31] Liu, Q., Zhang, R., Wang, Y., Yan, H., Hong, M.: Short-term daily prediction of
 919 sea ice concentration based on deep learning of gradient loss function. *Frontiers
 920 in Marine Science* **8**, 736429 (2021)
- 921 [32] Ren, Y., Li, X., Zhang, W.: A data-driven deep learning model for weekly sea
 922 ice concentration prediction of the pan-arctic during the melting season. *IEEE
 923 Transactions on Geoscience and Remote Sensing* **60**, 1–19 (2022)
- 924 [33] Ren, Y., Li, X.: Predicting the daily sea ice concentration on a sub-seasonal scale
 925 of the pan-arctic during the melting season by a deep learning model. *IEEE
 926 Transactions on Geoscience and Remote Sensing* (2023)
- 927 [34] Zheng, Q., Li, W., Shao, Q., Han, G., Wang, X.: A mid-and long-term arctic sea
 928 ice concentration prediction model based on deep learning technology. *Remote
 929 Sensing* **14**(12), 2889 (2022)

- 930 [35] Zheng, Q., Wang, R., Han, G., Li, W., Wang, X., Shao, Q., Wu, X., Cao, L., Zhou,
 931 G., Hu, S.: A spatio-temporal multiscale deep learning model for subseasonal
 932 prediction of arctic sea ice. *IEEE Transactions on Geoscience and Remote Sensing*
 933 (2024)
- 934 [36] Grigoryev, T., Verezemskaya, P., Krinitkiy, M., Anikin, N., Gavrikov, A., Trofi-
 935 mov, I., Balabin, N., Shpilman, A., Eremchenko, A., Gulev, S., *et al.*: Data-driven
 936 short-term daily operational sea ice regional forecasting. *Remote Sensing* **14**(22),
 937 5837 (2022)
- 938 [37] Graves, A., Graves, A.: Long short-term memory. Supervised sequence labelling
 939 with recurrent neural networks, 37–45 (2012)
- 940 [38] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N.,
 941 Kaiser, L., Polosukhin, I.: Attention is all you need. *Advances in neural*
 942 *information processing systems* **30** (2017)
- 943 [39] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater,
 944 J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., *et al.*: The era5 global reanal-
 945 ysis. *Quarterly Journal of the Royal Meteorological Society* **146**(730), 1999–2049
 946 (2020)
- 947 [40] Zuo, H., Balmaseda, M.A., Tietsche, S., Mogensen, K., Mayer, M.: The ecmwf
 948 operational ensemble reanalysis–analysis system for ocean and sea ice: a descrip-
 949 tion of the system and assessment. *Ocean science* **15**(3), 779–808 (2019)
- 950 [41] Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state
 951 spaces. *arXiv preprint arXiv:2312.00752* (2023)
- 952 [42] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient chan-
 953 nel attention for deep convolutional neural networks. In: *Proceedings of the*
 954 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.
 955 11534–11542 (2020)
- 956 [43] Bushuk, M., Ali, S., Bailey, D.A., Bao, Q., Batté, L., Bhatt, U.S., Blanchard-
 957 Wrigglesworth, E., Blockley, E., Cawley, G., Chi, J., *et al.*: Predicting september
 958 arctic sea ice: A multi-model seasonal skill comparison. *Bulletin of the American*
 959 *Meteorological Society* (2024)
- 960 [44] Kim, B.-M., Son, S.-W., Min, S.-K., Jeong, J.-H., Kim, S.-J., Zhang, X., Shim,
 961 T., Yoon, J.-H.: Weakening of the stratospheric polar vortex by arctic sea-ice loss.
 962 *Nature communications* **5**(1), 4646 (2014)
- 963 [45] Xu, M., Screen, J.A., Tian, W., Zhang, J., Zhang, C., Yu, H.: Influence of regional
 964 sea ice loss on the arctic stratospheric polar vortex. *Journal of Geophysical*
 965 *Research: Atmospheres* **129**(15), 2023–040571 (2024)

- 966 [46] Bushuk, M., Msadek, R., Winton, M., Vecchi, G.A., Gudgel, R., Rosati, A., Yang,
 967 X.: Skillful regional prediction of arctic sea ice on seasonal timescales. *Geophysical*
 968 *Research Letters* **44**(10), 4953–4964 (2017)
- 969 [47] Blanchard-Wrigglesworth, E., Bitz, C., Holland, M.: Influence of initial conditions
 970 and climate forcing on predicting arctic sea ice. *Geophysical Research Letters*
 971 **38**(18) (2011)
- 972 [48] Ordoñez, A.C., Bitz, C.M., Blanchard-Wrigglesworth, E.: Processes controlling
 973 arctic and antarctic sea ice predictability in the community earth system model.
 974 *Journal of Climate* **31**(23), 9771–9786 (2018)
- 975 [49] Zampieri, L., Goessling, H.F., Jung, T.: Bright prospects for arctic sea ice predic-
 976 tion on subseasonal time scales. *Geophysical Research Letters* **45**(18), 9731–9738
 977 (2018)
- 978 [50] Liu, Y., Bogaardt, L., Attema, J., Hazeleger, W.: Extended-range arctic sea ice
 979 forecast with convolutional long short-term memory networks. *Monthly Weather
 980 Review* **149**(6), 1673–1693 (2021)
- 981 [51] Wei, J., Hang, R., Luo, J.-J.: Prediction of pan-arctic sea ice using attention-based
 982 lstm neural networks. *Frontiers in Marine Science* **9**, 860403 (2022)
- 983 [52] Holland, M.M., Bitz, C.M., Tremblay, B.: Future abrupt reductions in the summer
 984 arctic sea ice. *Geophysical research letters* **33**(23) (2006)
- 985 [53] Stroeve, J.C., Serreze, M.C., Holland, M.M., Kay, J.E., Malanik, J., Barrett, A.P.:
 986 The arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic change*
 987 **110**, 1005–1027 (2012)
- 988 [54] Kimmritz, M., Counillon, F., Smedsrød, L.H., Bethke, I., Keenlyside, N., Ogawa,
 989 F., Wang, Y.: Impact of ocean and sea ice initialisation on seasonal prediction skill
 990 in the arctic. *Journal of Advances in Modeling Earth Systems* **11**(12), 4147–4166
 991 (2019)
- 992 [55] Volodioire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier,
 993 M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., *et al.*: Evaluation of
 994 cmip6 deck experiments with cnrm-cm6-1. *Journal of Advances in Modeling Earth
 995 Systems* **11**(7), 2177–2213 (2019)
- 996 [56] Wang, W., Chen, M., Kumar, A.: Seasonal prediction of arctic sea ice extent
 997 from a coupled dynamical forecast system. *Monthly Weather Review* **141**(4),
 998 1375–1394 (2013)
- 999 [57] Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D.,
 1000 Hou, Y.-T., Chuang, H.-y., Iredell, M., *et al.*: The ncep climate forecast system
 1001 version 2. *Journal of climate* **27**(6), 2185–2208 (2014)

- 1002 [58] Liu, Y., Wang, W., Kumar, A., Collow, T.: Assessment of cpc sea ice initialization
 1003 system (cxis) and cpc weekly experimental sea ice forecasts. In: 44th NOAA
 1004 Annual Climate Diagnostics and Prediction Workshop, Durham, NC, NOAA, pp.
 1005 74–77 (2019)
- 1006 [59] Collow, T.W., Liu, Y., Wang, W., Kumar, A., DeWitt, D.: Develop improved
 1007 seasonal and week 3/4 sea ice outlook. Performance evaluations, improvement
 1008 requirements, development updates, 39 (2020)
- 1009 [60] Lin, H., Merryfield, W.J., Muncaster, R., Smith, G.C., Markovic, M., Dupont, F.,
 1010 Roy, F., Lemieux, J.-F., Dirkson, A., Kharin, V.V., *et al.*: The canadian seasonal
 1011 to interannual prediction system version 2 (cansipsv2). Weather and Forecasting
 1012 **35**(4), 1317–1343 (2020)
- 1013 [61] Hazeleger, W., Wang, X., Severijns, C., Štefănescu, S., Bintanja, R., Sterl, A.,
 1014 Wyser, K., Semmler, T., Yang, S., Hurk, B., *et al.*: Ec-earth v2. 2: description and
 1015 validation of a new seamless earth system prediction model. Climate dynamics
 1016 **39**, 2611–2629 (2012)
- 1017 [62] Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L.,
 1018 Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., *et al.*: Seas5: the new ecmwf seasonal forecast system. Geoscientific Model Development **12**(3),
 1019 1087–1117 (2019)
- 1021 [63] Li, J., Bao, Q., Liu, Y., Wu, G., Wang, L., He, B., Wang, X., Yang, J., Wu,
 1022 X., Shen, Z.: Dynamical seasonal prediction of tropical cyclone activity using the
 1023 fgoals-f2 ensemble prediction system. Weather and Forecasting **36**(5), 1759–1778
 1024 (2021)
- 1025 [64] Qiao, F., Song, Z., Bao, Y., Song, Y., Shu, Q., Huang, C., Zhao, W.: Development
 1026 and evaluation of an earth system model with surface gravity waves. Journal of
 1027 Geophysical Research: Oceans **118**(9), 4514–4524 (2013)
- 1028 [65] Chen, H., Yin, X., Bao, Y., Qiao, F.: Ocean satellite data assimilation experiments
 1029 in fio-esm using ensemble adjustment kalman filter. Science China Earth Sciences
 1030 **59**, 484–494 (2016)
- 1031 [66] Shu, Q., Qiao, F., Liu, J., Song, Z., Chen, Z., Zhao, J., Yin, X., Song, Y.: Arctic sea
 1032 ice concentration and thickness data assimilation in the fio-esm climate forecast
 1033 system. Acta Oceanologica Sinica **40**, 65–75 (2021)
- 1034 [67] Msadek, R., Vecchi, G.A., Winton, M., Gudgel, R.G.: Importance of initial conditions
 1035 in seasonal predictions of arctic sea ice extent. Geophysical Research Letters
 1036 **41**(14), 5208–5215 (2014)
- 1037 [68] Bushuk, M., Zhang, Y., Winton, M., Hurlin, B., Delworth, T., Lu, F., Jia, L.,
 1038 Zhang, L., Cooke, W., Harrison, M., *et al.*: Mechanisms of regional arctic sea

- ice predictability in two dynamical seasonal forecast systems. *Journal of Climate* **35**(13), 4207–4231 (2022)
- [69] Zhang, Y.-F., Bushuk, M., Winton, M., Hurlin, B., Delworth, T., Harrison, M., Jia, L., Lu, F., Rosati, A., Yang, X.: Subseasonal-to-seasonal arctic sea ice forecast skill improvement from sea ice concentration assimilation. *Journal of Climate* **35**(13), 4233–4252 (2022)
- [70] Molod, A.: The gmao high resolution coupled model and assimilation system for seasonal prediction. In: EGU General Assembly Conference Abstracts, pp. 21–12759 (2021)
- [71] Zhang, J., Rothrock, D.A.: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Monthly Weather Review* **131**(5), 845–861 (2003)
- [72] Zhang, J., Steele, M., Lindsay, R., Schweiger, A., Morison, J.: Ensemble 1-year predictions of arctic sea ice for the spring and summer of 2008. *Geophysical Research Letters* **35**(8) (2008)
- [73] Cassano, J.J., DuVivier, A., Roberts, A., Hughes, M., Seefeldt, M., Brunke, M., Craig, A., Fisel, B., Gutowski, W., Hamman, J., *et al.*: Development of the regional arctic system model (rasm): Near-surface atmospheric climate sensitivity. *Journal of Climate* **30**(15), 5729–5753 (2017)
- [74] Barthélémy, A., Goosse, H., Fichefet, T., Lecomte, O.: On the sensitivity of antarctic sea ice model biases to atmospheric forcing uncertainties. *Climate Dynamics* **51**, 1585–1603 (2018)
- [75] Horvath, S., Stroeve, J., Rajagopalan, B.: A linear mixed effects model for seasonal forecasts of arctic sea ice retreat. *Polar Geography* **44**(4), 297–314 (2021)
- [76] Lin, Y., Yang, Q., Li, X., Yang, C.-Y., Wang, Y., Wang, J., Liu, J., Chen, S., Liu, J.: Optimization of the k-nearest-neighbors model for summer arctic sea ice prediction. *Frontiers in Marine Science* **10**, 1260047 (2023)
- [77] Zeng, J., Yang, Q., Li, X., Yuan, X., Bushuk, M., Chen, D.: Reducing the spring barrier in predicting summer arctic sea ice concentration. *Geophysical Research Letters* **50**(8), 2022–102115 (2023)
- [78] Kim, E., Kruse, P., Lama, S., Bourne, J., Hu, M., Ali, S., Huang, Y., Wang, J.: Multi-task deep learning based spatiotemporal arctic sea ice forecasting. In: 2021 IEEE International Conference on Big Data (Big Data), pp. 1847–1857 (2021). IEEE
- [79] Dool, H.: Empirical Methods in Short-Term Climate Prediction. Oxford University Press, ??? (2006). <https://doi.org/10.1093/oso/9780199202782.001.0001>.

1075 <https://doi.org/10.1093/oso/9780199202782.001.0001>

- 1076 [80] Finn, T.S., Durand, C., Farchi, A., Bocquet, M., Chen, Y., Carrassi, A.,
1077 Dansereau, V.: Deep learning subgrid-scale parametrisations for short-term fore-
1078 casting of sea-ice dynamics with a maxwell elasto-brittle rheology. *The Cryosphere*
1079 **17**(7), 2965–2991 (2023)
- 1080 [81] Finn, T.S., Durand, C., Farchi, A., Bocquet, M., Rampal, P., Carrassi, A.: Gen-
1081 erative diffusion for regional surrogate models from sea-ice simulations. *Journal*
1082 of Advances in Modeling Earth Systems **16**(10), 2024–004395 (2024)
- 1083 [82] Gregory, W., Bushuk, M., Adcroft, A., Zhang, Y., Zanna, L.: Deep learning of
1084 systematic sea ice model errors from data assimilation increments. *Journal of*
1085 Advances in Modeling Earth Systems **15**(10), 2023–003757 (2023)
- 1086 [83] Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., Zanna, L.: Machine learning
1087 for online sea ice bias correction within global ice-ocean simulations. *Geophysical*
1088 *Research Letters* **51**(3), 2023–106776 (2024)
- 1089 [84] Cavalieri, D.J., Gloersen, P., Campbell, W.J.: Determination of sea ice parameters
1090 with the nimbus 7 smmr. *Journal of Geophysical Research: Atmospheres* **89**(D4),
1091 5355–5369 (1984)
- 1092 [85] Comiso, J.C.: Characteristics of arctic winter sea ice from satellite multispectral
1093 microwave observations. *Journal of Geophysical Research: Oceans* **91**(C1), 975–
1094 994 (1986)
- 1095 [86] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater,
1096 J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., et al.: Era5 monthly averaged
1097 data on single levels from 1979 to present. Copernicus Climate Change Service
1098 (C3S) Climate Data Store (CDS) **10**, 252–266 (2019)
- 1099 [87] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater,
1100 J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., et al.: Era5 monthly averaged
1101 data on pressure levels from 1979 to present. Copernicus Climate Change Service
1102 (C3S) Climate Data Store (CDS) (2019)
- 1103 [88] DiGirolamo, N., Parkinson, C., Cavalieri, D., Gloersen, P., Zwally, H.: Sea ice
1104 concentrations from nimbus-7 smmr and dmsp ssm/i-ssmis passive microwave
1105 data, version 2. NASA National Snow and Ice Data Center Distributed Active
1106 Archive Center (2022)
- 1107 [89] Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Liu, Y.: Vmamba:
1108 Visual state space model. arXiv preprint arXiv:2401.10166 (2024)
- 1109 [90] Chen, H., Song, J., Han, C., Xia, J., Yokoya, N.: Changemamba: Remote sens-
1110 ing change detection with spatio-temporal state space model. arXiv preprint

- 1111 arXiv:2404.03425 (2024)
- 1112 [91] Chen, T., Tan, Z., Gong, T., Chu, Q., Wu, Y., Liu, B., Ye, J., Yu, N.: Mim-
1113 istd: Mamba-in-mamba for efficient infrared small target detection. arXiv preprint
1114 arXiv:2403.02148 (2024)
- 1115 [92] Ma, X., Zhang, X., Pun, M.-O.: Rs 3 mamba: Visual state space model for remote
1116 sensing image semantic segmentation. IEEE Geoscience and Remote Sensing
1117 Letters (2024)
- 1118 [93] Liao, W., Zhu, Y., Wang, X., Pan, C., Wang, Y., Ma, L.: Lightm-unet: Mamba
1119 assists in lightweight unet for medical image segmentation. arXiv preprint
1120 arXiv:2403.05246 (2024)
- 1121 [94] Chen, K., Chen, B., Liu, C., Li, W., Zou, Z., Shi, Z.: Rsmamba: Remote sensing
1122 image classification with state space model. IEEE Geoscience and Remote Sensing
1123 Letters (2024)
- 1124 [95] Yue, Y., Li, Z.: Medmamba: Vision mamba for medical image classification. arXiv
1125 preprint arXiv:2403.03849 (2024)
- 1126 [96] Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. arXiv preprint
1127 arXiv:1607.06450 (2016)
- 1128 [97] Chollet, F.: Xception: Deep learning with depthwise separable convolutions.
1129 In: Proceedings of the IEEE Conference on Computer Vision and Pattern
1130 Recognition, pp. 1251–1258 (2017)
- 1131 [98] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recogni-
1132 tion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern
1133 Recognition, pp. 770–778 (2016)
- 1134 [99] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint
1135 arXiv:1412.6980 (2014)