

# Physics-Knowledge-Guided Hybrid Neural Learning for Arctic Sea Ice Concentration Evolution and Short-Range Prediction

Maqun Zhang<sup>1</sup>, Feng Gao<sup>1</sup>, Wankun Chen<sup>1</sup>, Hui Yu<sup>2</sup>, Yanhai Gan<sup>1,\*</sup>, and Junyu Dong<sup>1,\*</sup>

**Abstract**—Accurate modeling of sea ice concentration (SIC) evolution is essential for polar climate assessment and short-range sea ice prediction. Numerical and data-driven approaches constitute major foundations for SIC modeling, but the former often require complex parameterizations and substantial computation, whereas the latter rarely encode physical dependencies explicitly. This study presents the Physics-Informed Hybrid Ice Model (PIHIM), a differentiable data-driven hybrid ice model for daily SIC evolution that organizes its network structure according to the physical dependencies encoded in the sea ice continuity equation and explicitly accounts for dynamical transport, thermodynamically driven areal growth and loss, and unresolved local displacement, freeze-melt areal change, and local error closure. Two evaluation settings are adopted: reanalysis-forced simulation examines SIC evolution stability under reanalysis forcing, and forecast-forced prediction assesses short-range performance under forecast-forced conditions, with reanalysis and observational SIC serving as verification references. Results indicate enhanced ice-edge preservation and error-growth control in reanalysis-forced simulation, while PIHIM retains measurable short-range prediction skill under forecast-forced conditions. Our code will be made publicly available after the paper is accepted.

**Index Terms**—Sea ice concentration, data-driven modeling, sea ice continuity equation, forecast-forced prediction, Arctic sea ice.

## I. INTRODUCTION

POLAR sea ice constitutes a key component of the climate system, modulating air-sea heat exchange, ocean circulation, polar ecosystems, and Arctic navigation [1]–[6]. Accurate modeling of SIC evolution is therefore essential for climate monitoring and polar operations. However, SIC changes are jointly governed by atmospheric forcing, oceanic conditions, sea ice transport, and thermodynamic phase changes [7]–[10]. Their nonlinear, seasonal, and spatially heterogeneous nature complicates characterization of ice-edge displacement and low-concentration sea ice in the marginal ice zone and during freeze/melt transitions [11], [12].

Existing SIC evolution modeling approaches can be broadly grouped into physics-based numerical models and data-driven

This work was supported by the Leverhulme Trust through Project VP1-2020-044; the National Natural Science Foundation of China under Grant 42406192; the Fundamental Research Funds for the Central Universities under Grants 202413040 and 202572015; the National Science and Technology Major Project of China under Grant 2022ZD0117201; and the Postdoctoral Project of Qingdao under Grant QDBSH20240102021.

Corresponding authors: Yanhai Gan (ganyanhai@ouc.edu.cn) and Junyu Dong (dongjunyu@ouc.edu.cn).

Maqun Zhang, Feng Gao, Wankun Chen, Yanhai Gan, and Junyu Dong are with the State Key Laboratory of Physical Oceanography and Faculty of Information Science and Engineering, Ocean University of China, Qingdao 266100, China. Hui Yu is with the School of Psychology and Neuroscience, University of Glasgow, Glasgow G12 8QQ, U.K.

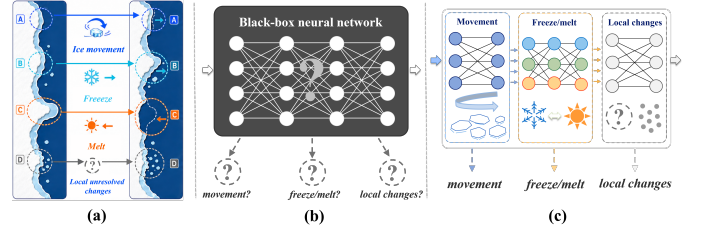


Fig. 1. Conceptual overview. (a) SIC evolution is jointly driven by ice movement, freeze/melt processes, and local unresolved changes. (b) Black-box neural networks cannot explicitly attribute SIC changes to different physical processes. (c) PIHIM decomposes neural prediction into process-guided components for movement, freeze/melt, and local changes.

models [13]–[16]. Numerical models explicitly represent sea ice dynamics and thermodynamics, but their performance depends on parameterizations of unresolved processes, and high-resolution simulations remain computationally demanding [7], [8], [13], [17]. Deep learning has also been widely used for spatiotemporal SIC prediction [14]–[16], [18]. Nevertheless, purely data-driven approaches commonly formulate SIC evolution as an end-to-end mapping, with internal states that do not readily correspond to dynamical transport, thermodynamic areal growth and decay, or residual closure. This opacity complicates error attribution and may weaken interpretability in autoregressive integration, under distribution shifts, and at high-gradient ice edges [19]–[21].

These considerations motivate a modeling strategy that combines process organization with learnable correction. The continuity equation separates horizontal transport from source-sink changes [7], [8], [22], [23], while learnable residual closure compensates for simplified physics, numerical discretization, and unresolved processes [19], [24]–[26]. Evaluation should also distinguish two settings: reanalysis-forced simulation, which examines whether the model can stably evolve SIC under prescribed reanalysis forcings, and forecast-forced prediction, which assesses short-range performance when boundary conditions are supplied by forecast products [13], [17], [18], [27]. These considerations are summarized schematically in Fig. 1.

To this end, this study proposes PIHIM, a differentiable hybrid framework for daily SIC evolution guided by the sea ice continuity equation. PIHIM decomposes one-step evolution into dynamical transport, a lightweight thermodynamics-guided source-term representation, and residual compensation. The dynamical module represents ice-motion-driven redistribution, the thermodynamic module constrains freeze/melt branches through sign constraints and SIC-based area-fraction weighting to describe areal change, and a lightweight residual-

closure network compensates for local errors arising from simplified physics, discretization, and unresolved processes.

The main contributions are as follows:

- 1) We develop a differentiable data-driven hybrid ice model for daily SIC evolution, organizing its network structure according to the physical dependencies encoded in the sea ice continuity equation.
- 2) We introduce a lightweight, data-driven thermodynamics-guided source-term representation that avoids the computational burden of a full thermodynamic sea ice model, guides learning of SIC areal change through area-fraction-weighted freeze/melt branches, and supports analysis of seasonal thermodynamic responses.
- 3) We establish a two-setting evaluation framework that distinguishes reanalysis-forced simulation from forecast-forced prediction, separately assessing evolution stability and module contributions under reanalysis forcing, and short-range prediction performance under forecast-forced conditions.

## II. RELATED WORK

### A. Physics-based Numerical Models and Machine-learning Correction

Physics-based numerical models remain a central foundation for sea ice prediction because they explicitly represent coupled atmosphere, ocean, and sea ice processes. Seasonal systems such as SEAS5 provide valuable numerical references [17], and multimodel comparisons indicate pan-Arctic skill while remaining sensitive to regional variability, initialization errors, extreme years, and model biases [13]. Machine learning has therefore been applied to correct or augment numerical systems, including seasonal sea ice prediction from Earth system models [28] and online global ice-ocean or coupled prediction systems [29], [30]. These studies demonstrate the value of combining physical models with machine learning, but most remain embedded in full numerical systems or focus on bias correction and system-level forecast performance, rather than lightweight, differentiable, process-decomposed modeling of daily SIC evolution.

### B. Data-driven SIC Prediction and Sea Ice Evolution Simulation

Deep learning is increasingly applied in sea ice remote sensing, SIC prediction, and sea ice state simulation, including automated sea ice mapping and input selection [31]. For prediction-oriented tasks, IceNet produces probabilistic seasonal forecasts [14], and IceMamba integrates attention mechanisms into a state space architecture for seasonal pan-Arctic SIC forecasting [32]. For diagnostic attribution rather than forecasting, deep neural networks have been used to reconstruct Arctic sea ice extent variability from observational daily SST anomaly fields across ocean basins and identify influential regions through explainable AI [33]. Other studies have explored lightweight U-Net models for short-range daily

prediction [34], attention-based SICNet for pan-Arctic SIC prediction [15], and IceFormer with multiscale Transformer representations for subseasonal prediction [16]. Related ocean studies have used deep learning for long-lead 3D temperature and remote-sensing reflectance prediction [35], [36], while FuXi-Ocean provides 6-hourly,  $1/12^\circ$  global ocean forecasts and addresses cumulative error amplification in autoregressive forecasting [37]. Together, these works indicate that data-driven models can efficiently capture complex spatiotemporal dependencies across remote-sensing retrieval and prediction tasks. However, most prediction and simulation models still formulate SIC or sea ice state evolution as an end-to-end mapping, leaving their internal states without direct correspondence to explicit dynamical or thermodynamic processes. This limits process diagnosis during autoregressive integration, under distributional shifts, and near high-gradient ice-edge regions.

### C. Physics-guided and Constrained Deep Learning Models

Within the broader literature on physics-informed machine learning [38], recent sea ice studies have introduced physics-related inputs, task-specific losses, conservation constraints, and process-oriented learning to enhance the credibility of sea ice machine learning. Examples include normalized integrated ice-edge error (NIIEE) loss for ice-edge shape errors [39], thickness-informed seasonal SIC prediction [40], and physical information-guided multi-source fusion for high-resolution SIC estimation and ice-edge representation [41]. Beyond sea ice, physics-aware statistical frameworks have used meteorological variables and learned interaction factors as physical priors for multi-temporal-resolution marine heatwave forecasting [42]. In physics-guided deep learning, previous studies have explored several effective strategies for incorporating domain knowledge into neural networks, including numerical-model supervision, physics-aware architecture design, physically constrained forward propagation, and PDE-informed priors [43]–[46]. More broadly, data-driven methods can learn unresolved-process closures under physical constraints in ocean modeling [25], and mass-conserving sea ice emulators can improve interpretability and cross-climate generalization [47]. These studies indicate that physical guidance has become an active direction in sea ice and ocean machine learning. Nevertheless, existing approaches usually introduce such guidance through auxiliary variables, losses, conservation formulations, or embedded correction. Direct attempts to organize SIC evolution into differentiable structures that explicitly reflect dynamical and thermodynamic processes under a unified sea ice continuity framework remain limited.

## III. METHODOLOGY

### A. Problem Formulation and PIHM Overview

This study models the daily Arctic SIC field, denoted by  $A \in [0, 1]$ , where each grid-cell value represents the ice-covered fraction. Following common SIC and MIZ thresholds, cells with SIC not exceeding 15% are defined as open water (OW), cells with SIC greater than 15% and lower than 80% as the marginal ice zone (MIZ), and cells with SIC no lower

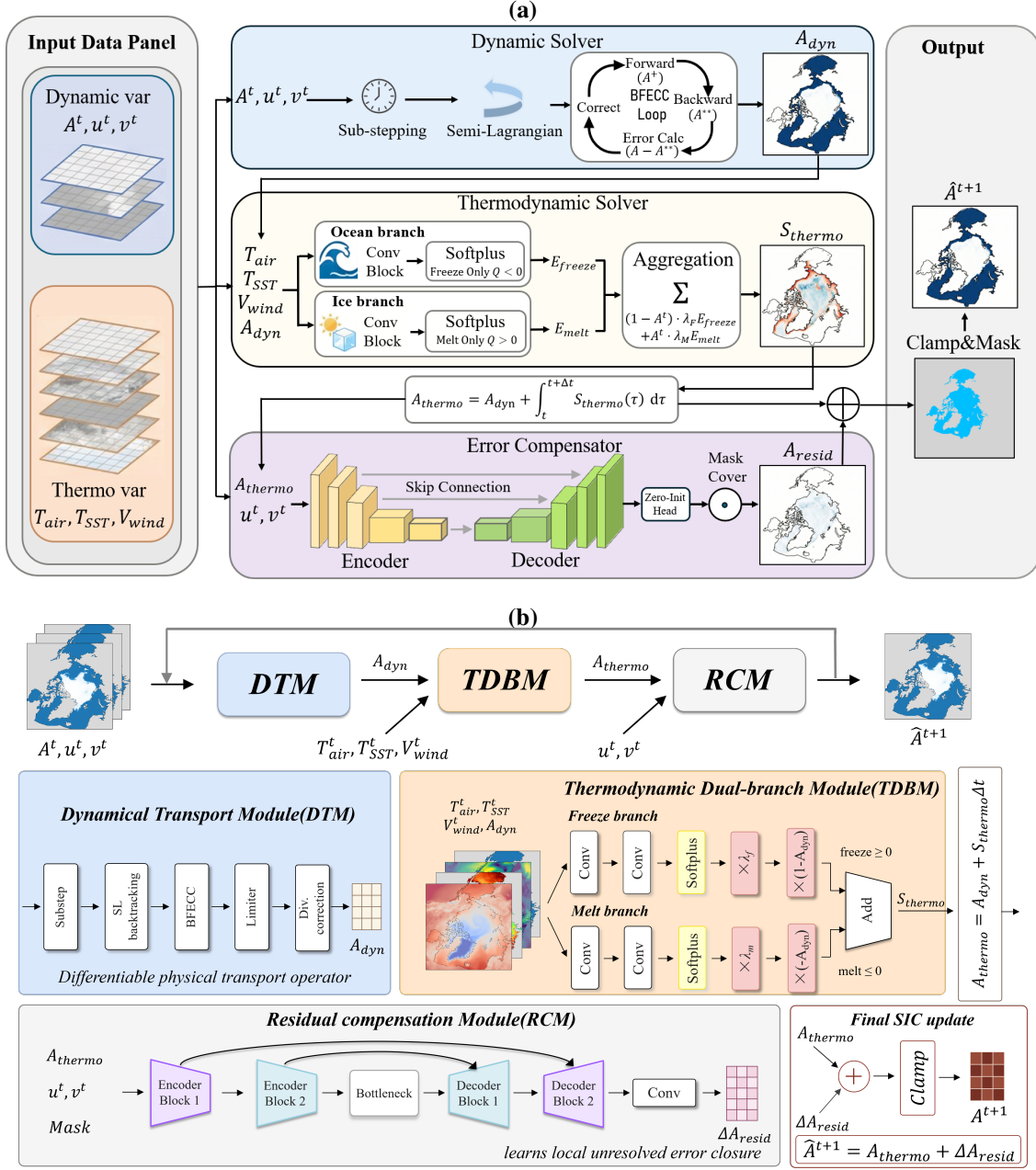


Fig. 2. PIHIM network overview. (a) The conceptual structure links the SIC continuity-equation decomposition to movement, freeze/melt processes, and local unresolved changes. (b) The implementation-level architecture presents the corresponding module inputs, outputs, and autoregressive rollout for multi-day prediction.

than 80% as pack ice (PI) [11], [12]. Given the current SIC field  $A^t$ , sea ice drift velocity  $\mathbf{v}^t = (u^t, v^t)$ , and atmospheric-oceanic forcing variables  $\mathbf{X}^t$ , the objective is to learn the one-day evolution operator in Eq. (1):

$$\hat{A}^{t+1} = \mathcal{G}_{\Theta}(A^t, \mathbf{v}^t, \mathbf{X}^t), \quad (1)$$

where  $\mathcal{G}_{\Theta}$  denotes the hybrid model. Multi-step prediction is performed autoregressively by feeding each prediction into the next step. Reanalysis and forecast forcings correspond to different evaluation settings but do not change the PIHIM model structure.

PIHIM is structured around the SIC continuity equation in

Eq. (2) [7], [8], [22], [23]:

$$\frac{\partial A}{\partial t} + \nabla \cdot (A\mathbf{v}) = S_A, \quad (2)$$

where the conservative transport term describes velocity-driven horizontal SIC redistribution, and  $S_A$  denotes the thermodynamic area source-sink term associated with freezing and melting. Based on Eq. (2), PIHIM explicitly implements dynamical transport and thermodynamic areal change, while a residual module compensates for remaining discrepancies arising from discretization approximations and unresolved local effects.

As summarized in Fig. 2, the dynamical module computes velocity-driven SIC redistribution, the thermodynamic module

estimates freeze/melt areal change from external forcings, and the residual module learns remaining local errors. This differentiable process-decomposed design provides the structural basis for module diagnosis, ablation analysis, and long-term autoregressive evaluation.

### B. Physics-guided SIC Evolution Framework

1) *Dynamical Transport Module*: The dynamical transport module represents SIC redistribution driven by horizontal sea ice drift. When thermodynamic source-sink terms are omitted, SIC follows the conservative continuity equation in Eq. (3) [7], [22], [23]:

$$\frac{\partial A}{\partial t} + \nabla \cdot (A\mathbf{v}) = 0, \quad (3)$$

where  $\mathbf{v} = (u, v)$  is the sea ice drift velocity. This process is expressed as the differentiable transport operator in Eq. (4):

$$A_{\text{dyn}} = \mathcal{D}(A^t, \mathbf{v}^t). \quad (4)$$

where  $A_{\text{dyn}}$  is the transported SIC state. In implementation,  $\mathcal{D}$  employs a semi-Lagrangian scheme with temporal substepping to enhance daily integration stability [48], [49]. Given a one-day time step  $\Delta t$  and  $N_{\text{sub}}$  transport substeps, the substep size is

$$\delta t = \frac{\Delta t}{N_{\text{sub}}}.$$

This study sets  $N_{\text{sub}} = 24$ . At each substep, departure points are backtracked along the velocity field, SIC is interpolated at these points, and a divergence correction is applied. The resulting  $A_{\text{dyn}}$  is passed to the thermodynamics-guided source-term module.

2) *Thermodynamic Dual-branch Module (TDBM)*: Dynamical transport cannot represent SIC areal growth or decay caused by freezing and melting. Rather than solving a full sea ice thermodynamic model, which would require additional variables such as thickness, radiation, albedo, snow cover, and melt ponds, PIHIM adopts a lightweight thermodynamics-guided source-term representation for horizontal SIC areal change.

Within a grid cell, SIC change is approximated as the combination of new ice formation over available open-water area and melting over existing ice-covered area. Two branches estimate potential freezing and melting tendencies from thermodynamic forcings as shown in Eq. (5):

$$\begin{aligned} Q_f &= \mathcal{F}_{\theta_f}(\mathbf{X}_{\text{thermo}}^t), \\ Q_m &= \mathcal{F}_{\theta_m}(\mathbf{X}_{\text{thermo}}^t), \end{aligned} \quad (5)$$

where  $\mathbf{X}_{\text{thermo}}^t$  comprises near-surface air temperature, sea surface temperature, and scalar wind speed, and  $\mathcal{F}_{\theta_f}$  and  $\mathcal{F}_{\theta_m}$  are the freezing and melting branches, respectively. Their structures are provided in Supplementary Section S7.2. Non-negative transformations enforce the signs of the two tendencies in Eq. (6):

$$\begin{aligned} E_f &= \text{softplus}(Q_f), \\ E_m &= \text{softplus}(Q_m). \end{aligned} \quad (6)$$

The transported SIC state weights the available area for each process: freezing scales with  $1 - A_{\text{dyn}}$ , and melting scales with

$A_{\text{dyn}}$ . The thermodynamic source term is therefore defined by Eq. (7):

$$S_{\text{thermo}} = (1 - A_{\text{dyn}})\lambda_f E_f - A_{\text{dyn}}\lambda_m E_m, \quad (7)$$

where  $\lambda_f$  and  $\lambda_m$  are learnable scale factors. In implementation, the thermodynamic source term is evaluated only over the valid thermodynamic active region and is set to zero outside this region; the corresponding masking settings are provided in Supplementary Section S7.2. Thus, the module provides a sign-constrained and area-weighted decomposition of thermodynamic areal change without resolving full thermodynamic physics.

3) *Residual Compensation and State Update*: After dynamical transport and the thermodynamic source-term update, PIHIM forms a physics-guided intermediate state through the explicit time integration in Eq. (8):

$$A_{\text{thermo}} = A_{\text{dyn}} + S_{\text{thermo}}\Delta t. \quad (8)$$

Here,  $S_{\text{thermo}}\Delta t$  is the one-day SIC increment induced by thermodynamic areal change. Because this intermediate state cannot fully represent reference SIC evolution, owing to unresolved deformation, ice-edge fragmentation, forcing-related mismatch, interpolation error, and discretization error, a residual network estimates the remaining increment in Eq. (9):

$$\Delta A_{\text{resid}} = \mathcal{R}_{\theta}(A_{\text{thermo}}, \mathbf{v}^t). \quad (9)$$

The valid ocean mask is applied in implementation. The velocity input allows the residual module to compensate for numerical errors associated with the discretized dynamical-transport step. The residual module is not supplied with raw thermodynamic forcing variables; thermodynamic information enters the residual pathway only indirectly through  $A_{\text{thermo}}$ .

The final one-step prediction in Eq. (10) combines the physics-guided state and residual correction:

$$\hat{A}^{t+1} = A_{\text{thermo}} + \Delta A_{\text{resid}}. \quad (10)$$

In implementation,  $A_{\text{dyn}}$ ,  $A_{\text{thermo}}$ , and  $\hat{A}^{t+1}$  are clipped to  $[0, 1]$  and multiplied by the valid ocean mask.

The last layer of the residual network is zero-initialized so that the initial residual is close to zero. This initialization reduces the tendency for the residual module to override the dynamical and thermodynamic baseline at the beginning of training, while allowing it to learn remaining local biases during joint optimization.

Overall, the dynamical module accounts for horizontal transport and ice-edge displacement, the thermodynamic module accounts for freeze/melt areal change, and the residual module compensates for remaining errors. Detailed module architectures, transport settings, and the structural comparison with IceNet are provided in Supplementary Section S7.

### C. Training Strategy and Loss Function

PIHIM combines a parameter-free dynamical transport operator with trainable thermodynamic and residual modules. Because the residual network has substantial representational capacity, direct end-to-end optimization may cause the residual pathway to dominate error gradients before

the thermodynamics-guided source-term module learns stable freeze/melt tendencies. PIHIM is therefore trained with staged optimization, autoregressive rollout, and a regional area-bias constraint.

1) *Training Strategy*: Training contains a warm-up stage and a joint fine-tuning stage. During warm-up, the residual module is frozen and only the thermodynamics-guided source-term module is optimized, encouraging the model to learn forcing-driven freezing and melting tendencies from the dynamical-thermodynamic baseline. During joint fine-tuning, the residual module is unfrozen and optimized together with the thermodynamic module, so that it mainly learns remaining errors rather than replacing the process-decomposed components.

To enhance multi-step stability, training employs autoregressive rollout. Starting from  $A^t$ , the model is integrated for  $T$  consecutive steps, and each predicted SIC field is reused as the next input according to Eq. (11):

$$\hat{A}^{t+\tau} = \mathcal{G}_{\Theta}(\hat{A}^{t+\tau-1}, \mathbf{v}^{t+\tau-1}, \mathbf{X}^{t+\tau-1}), \quad \tau = 1, \dots, T. \quad (11)$$

where  $\hat{A}^t = A^t$ . This study sets  $T = 14$ , exposing the model to accumulated rollout errors during training. Training objectives, auxiliary constraints, hyperparameters, hardware configuration, and convergence diagnostics are provided in Supplementary Sections S6.2–S6.3.

2) *Loss Function*: SIC prediction requires both pixel-scale accuracy and control of regional areal drift. Pixel-wise mean squared error constrains local errors but may not sufficiently penalize systematic overestimation or underestimation of regional SIC during multi-step integration. PIHIM therefore adopts a composite loss with a pixel-wise term and a regional area-bias term.

For rollout step  $\tau$ , Eq. (12) defines the single-step loss:

$$\mathcal{L}_{\text{step}}^{(\tau)} = \mathcal{L}_{\text{pix}}^{(\tau)} + \lambda \mathcal{L}_{\text{bias}}^{(\tau)}, \quad (12)$$

where  $\lambda$  balances the two terms. The pixel-wise loss in Eq. (13) is

$$\mathcal{L}_{\text{pix}}^{(\tau)} = \frac{1}{\sum M} \sum \left[ (\hat{A}^{t+\tau} - A^{t+\tau}) \odot M \right]^2, \quad (13)$$

where  $M$  excludes land and invalid grid cells, and  $\odot$  denotes element-wise multiplication. The regional area-bias loss in Eq. (14) is

$$\mathcal{L}_{\text{bias}}^{(\tau)} = \left( \frac{\sum (\hat{A}^{t+\tau} \odot M)}{\sum M} - \frac{\sum (A^{t+\tau} \odot M)}{\sum M} \right)^2. \quad (14)$$

This term penalizes the squared difference between predicted and reference mean SIC over the valid ocean region. The main rollout loss in Eq. (15) is averaged over all rollout steps:

$$\mathcal{L} = \frac{1}{T} \sum_{\tau=1}^T \mathcal{L}_{\text{step}}^{(\tau)}. \quad (15)$$

This main rollout loss constrains step-wise spatial accuracy and multi-step regional stability; auxiliary training terms are described in Supplementary Section S6.2.

## IV. EXPERIMENTS

This section evaluates PIHIM in terms of short-range accuracy, autoregressive stability, process interpretation, and forecast-forced prediction. Subsequent experiments follow the two evaluation settings defined in Section 4.1: reanalysis-forced simulation examines SIC evolution under reanalysis forcing, and forecast-forced prediction assesses D1-D9 short-range performance under forecast-forced conditions [50]–[54]. PIHIM is compared with a persistence baseline and representative data-driven spatiotemporal prediction models, including IceNet [14], SimVP [55], FCNet [56], PredRNNv2 [57], VMRNN [58], SwinLSTM [59], and PredRNN++ [60].

### A. Experimental Design

The experiments distinguish reanalysis-forced simulation from forecast-forced prediction to avoid interpreting prescribed-forcing simulation as independent prediction performance. In reanalysis-forced simulation, ERA5/GLORYS reanalysis forcings, including sea ice drift, are supplied at each valid time. In forecast-forced prediction, each case starts from the initial SIC and analyzed forcing state available at initialization; subsequent D1-D9 rollout is driven by AIFS/RTOFS forecast forcings (RTOFS product information: <https://polar.ncep.noaa.gov/global/about/>), while future ERA5, GLORYS, and OSI SAF fields are reserved only for verification.

Within each setting, all trainable baselines are evaluated using the same data splits, input variables, evaluation windows, and verification references as PIHIM. Table 1 summarizes the main settings, and detailed protocols, data roles, metric definitions, information-availability constraints, preprocessing, and vector-rotation details are provided in Supplementary Sections S1 and S6.1.

The metrics include mean absolute error (MAE), root-mean-square error (RMSE), anomaly correlation coefficient (ACC), Integrated Ice Edge Error (IIEE), sea ice area (SIA), and sea ice extent (SIE); area and extent diagnostics are reported in  $10^6 \text{ km}^2$ .

### B. Reanalysis-forced Simulation

This section adopts the reanalysis-forced simulation setting defined in Section 4.1 to evaluate PIHIM. The 14-day benchmark assesses short-range SIC evolution accuracy, while the 90-day, calendar-year, and multi-year integrations examine autoregressive rollout stability. Supplementary initial-condition perturbation and extreme-year withholding diagnostics are provided in Supplementary Sections S2.4 and S2.6.

1) *Fourteen-day Benchmark*: The 14-day benchmark compares PIHIM with Persistence and selected spatiotemporal prediction baselines under reanalysis-forced simulation, with the results reported in Table 2.

Among the evaluated models, PIHIM attains the lowest MAE, RMSE, and IIEE and the highest ACC. Relative to Persistence, it reduces MAE and IIEE by about 41% and 61%, respectively, indicating that it does not simply preserve the initial SIC field. IceNet is the closest baseline, but PIHIM

TABLE I  
SUMMARY OF EXPERIMENTAL SETTINGS, DATA, METRICS, AND BASELINES

| Category                   | Setting                                                                  | Role                                                                                                              |
|----------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| Evaluation setting         | reanalysis-forced simulation                                             | Setting defined in Section 4.1 for assessing SIC evolution under reanalysis forcing                               |
| Evaluation setting         | forecast-forced prediction                                               | Setting defined in Section 4.1 for assessing D1-D9 short-range prediction under forecast-forced conditions        |
| Reanalysis/reference       | ERA5 [50], GLORYS [51]                                                   | ERA5 provides atmospheric forcings; GLORYS provides SST, SIC, and sea ice velocity                                |
| Forecast forcings          | AIFS [52], RTOFS [53]                                                    | AIFS provides air temperature and wind speed, RTOFS provides SST, and the sea ice velocity channel is set to zero |
| Observational verification | OSI SAF [54]                                                             | Satellite SIC for assessing whether the results depend on the GLORYS reference                                    |
| Numerical reference        | SEAS5 [17]                                                               | Serves as a contextual numerical reference rather than a task-equivalent baseline                                 |
| Metrics                    | MAE, RMSE, Bias, ACC, IIEE, SIA/SIE bias, trend error                    | Evaluate grid errors, spatial structure, ice-edge error, area/extent drift, and long-term trend                   |
| Model baselines            | Persistence, IceNet, SimVP, FCNet, PredRNNv2, VMRNN, SwinLSTM, PredRNN++ | Cover persistence and data-driven spatiotemporal prediction baselines                                             |

TABLE II  
FOURTEEN-DAY REANALYSIS-FORCED SIMULATION BENCHMARK

| Model       | MAE           | RMSE          | Bias          | ACC           | IIEE          |
|-------------|---------------|---------------|---------------|---------------|---------------|
| Persistence | 0.0389        | 0.1094        | 0.0009        | 0.6999        | 0.6202        |
| PredRNN++   | 0.0297        | 0.0799        | 0.0028        | 0.8256        | 0.4227        |
| VMRNN       | 0.0294        | 0.0714        | 0.0024        | 0.8594        | 0.3536        |
| SwinLSTM    | 0.0294        | 0.0712        | 0.0023        | 0.8604        | 0.3476        |
| PredRNNv2   | 0.0277        | 0.0703        | 0.0008        | 0.8627        | 0.3673        |
| FCNet       | 0.0265        | 0.0714        | 0.0005        | 0.8600        | 0.3431        |
| SimVP       | 0.0247        | 0.0655        | 0.0009        | 0.8809        | 0.2795        |
| IceNet      | 0.0238        | 0.0621        | 0.0020        | 0.8916        | 0.2680        |
| PIHIM       | <b>0.0229</b> | <b>0.0592</b> | <b>0.0004</b> | <b>0.9057</b> | <b>0.2437</b> |

still attains lower MAE and IIEE and higher ACC. Calendar-aligned lead-day heatmaps for the five metrics are provided in Supplementary Section S2.5 to indicate that these averaged results are not dominated by isolated seasonal periods.

2) *Seasonal-scale and Long-term Stability*: Longer integrations assess the autoregressive stability of PIHIM beyond the 14-day benchmark. SIA is computed as the SIC-weighted area over the valid ocean region, whereas SIE is computed as the area of grid cells with SIC at least 15% (NSIDC terminology: <https://nsidc.org/learn/ask-scientist/what-difference-between-sea-ice-area-and-extent>); both are reported in  $10^6 \text{ km}^2$ . The 90-day rollout evaluates intraseasonal stability, calendar-year integration evaluates full annual freeze-melt cycles, and multi-year continuous integration diagnoses long-term areal drift without annual reset. Extended lead-time curves, annual statistics, and multi-year trend diagnostics are provided in Supplementary Sections S2.1–S2.3.

a) *Ninety-day Autoregressive Rollout*: The 90-day rollout comprises 12 seasonal cases initialized on January 1, April 1, July 1, and October 1 in each year from 2022 to 2024. Table 3 reports endpoint-averaged metrics at Day 30, Day 60, and Day 90.

PIHIM exhibits no apparent error divergence through Day 90. MAE and RMSE remain near 0.026–0.027 and 0.068–0.071, ACC remains above 0.87, and IIEE remains near  $0.27 - 0.30 \times 10^6 \text{ km}^2$ . The Day-90 metrics should be

TABLE III  
ENDPOINT DIAGNOSTICS FOR 90-DAY REANALYSIS-FORCED SIMULATION

| Lead time | MAE    | RMSE   | ACC    | IIEE   | SIA bias |
|-----------|--------|--------|--------|--------|----------|
| Day 30    | 0.0270 | 0.0679 | 0.8710 | 0.2692 | 0.0157   |
| Day 60    | 0.0272 | 0.0710 | 0.8742 | 0.3038 | 0.0332   |
| Day 90    | 0.0257 | 0.0677 | 0.8808 | 0.2665 | 0.0057   |

TABLE IV  
CALENDAR-YEAR INTEGRATION DIAGNOSTICS FOR 2022–2024

| Year | MAE    | ACC    | IIEE   | End-of-year SIA bias |
|------|--------|--------|--------|----------------------|
| 2022 | 0.0276 | 0.8478 | 0.2980 | 0.0252               |
| 2023 | 0.0260 | 0.8869 | 0.2709 | 0.0086               |
| 2024 | 0.0262 | 0.8917 | 0.2700 | 0.0313               |

interpreted as endpoint averages over seasonal cases, not as cumulative averages along a single error sequence.

b) *Calendar-year Integration*: Calendar-year integration starts at the beginning of each test year and continues to year end, evaluating whether PIHIM captures summer melt, autumn–winter freeze-up, and intra-annual SIA/SIE variability. Table 4 reports the corresponding year-wise diagnostic metrics.

As shown in Fig. 3, the SIA/SIE curves reproduce the main annual phases, indicating that PIHIM maintains stable areal evolution over complete annual cycles. Across 2022–2024, MAE remains around 0.026–0.028, and IIEE remains around  $0.27 - 0.30 \times 10^6 \text{ km}^2$ . The lower ACC and higher IIEE in 2022 also indicate that long-term performance remains affected by interannual ice-edge conditions and external forcing states.

c) *Multi-year Continuous Integration*: Multi-year continuous integration evaluates stability without annual reset over the 2022–2024 test period. Table 5 reports linear SIA trends for the reference and PIHIM simulation.

Fig. 4 shows that PIHIM reproduces the reference trend direction, with a SIA trend error of  $0.0125 \times 10^6 \text{ km}^2 \text{ yr}^{-1}$ . This indicates limited areal drift under prescribed reanalysis forcings. Because the diagnostic covers only the 2022–2024 test period and remains forcing-dependent, it should be inter-

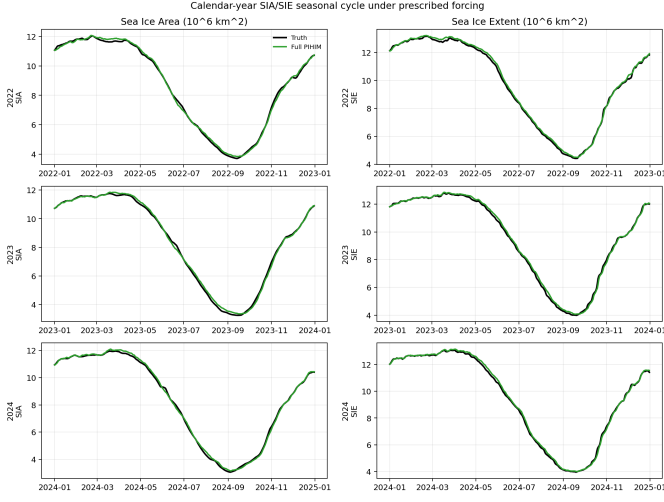


Fig. 3. Seasonal cycles of SIA/SIE in calendar-year integrations from 2022 to 2024

TABLE V  
TREND-STABILITY DIAGNOSTICS FOR MULTI-YEAR CONTINUOUS INTEGRATION

| Continuous integration period | Truth SIA trend | PIHIM SIA trend | SIA trend error |
|-------------------------------|-----------------|-----------------|-----------------|
| 2022-2024                     | -0.1318         | -0.1193         | 0.0125          |

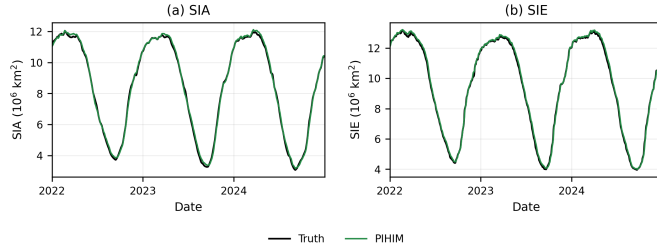


Fig. 4. SIA/SIE trajectories in multi-year continuous integration over the 2022-2024 test period

interpreted as a stability diagnostic rather than evidence of climate-scale trend prediction capability.

### C. Mechanistic Interpretation

This section interprets PIHIM behavior through component ablation, seasonal-regional error attribution, and thermodynamic source-term diagnostics. The central questions are whether the model gain arises from coupled process modules, where the main errors occur, and whether the thermodynamic outputs exhibit reasonable seasonal phase and perturbation responses. Complete diagnostics are provided in Supplementary Sections S3 and S8.

1) *Component Contribution*: Table 6 reports inference-time diagnostic ablations on the same trained PIHIM. Specific modules are disabled or retained to examine their immediate contributions; the variants are not treated as independently retrained architectures for model selection.

The full model attains the lowest errors and the highest ACC among these diagnostic variants. Dynamic + Thermo has higher errors and lower ACC than the full model, indicating

TABLE VI  
INFERENCE-TIME DIAGNOSTIC ABLATION RESULTS

| Variant                | MAE           | RMSE          | ACC           | IIEE          |
|------------------------|---------------|---------------|---------------|---------------|
| Dynamic + Thermo       | 0.0580        | 0.1398        | 0.5725        | 0.6355        |
| Thermo + compensation  | 0.0340        | 0.0832        | 0.8275        | 0.2875        |
| Dynamic + compensation | 0.0290        | 0.0708        | 0.8586        | 0.2750        |
| Full PIHIM             | <b>0.0229</b> | <b>0.0592</b> | <b>0.9057</b> | <b>0.2437</b> |

TABLE VII  
SEASONAL-REGIONAL ERROR ATTRIBUTION

| Region     | Season     | MAE    | RMSE   | Bias    |
|------------|------------|--------|--------|---------|
| MIZ        | transition | 0.1384 | 0.1820 | 0.0167  |
| MIZ        | freezing   | 0.1302 | 0.1685 | -0.0011 |
| MIZ        | melting    | 0.1192 | 0.1612 | 0.0113  |
| Pack ice   | transition | 0.0265 | 0.0474 | 0.0002  |
| Pack ice   | freezing   | 0.0227 | 0.0426 | -0.0045 |
| Pack ice   | melting    | 0.0347 | 0.0615 | -0.0138 |
| Open water | transition | 0.0087 | 0.0500 | 0.0076  |
| Open water | freezing   | 0.0062 | 0.0328 | 0.0016  |
| Open water | melting    | 0.0056 | 0.0372 | 0.0050  |

TABLE VIII  
SEASONALITY OF THERMODYNAMIC SOURCE TERMS

| Condition         | Mean freeze source | Mean melt source | Mean combined source |
|-------------------|--------------------|------------------|----------------------|
| Melting season    | 0.00265            | -0.00320         | -0.00050             |
| Freezing season   | 0.01828            | -0.00125         | 0.01673              |
| Transition season | 0.00905            | -0.00113         | 0.00786              |

that explicit process terms alone do not close all SIC evolution errors. Dynamic + compensation and Thermo + compensation also remain below full PIHIM, indicating that dynamical transport, thermodynamic source terms, and residual compensation are complementary. Retrained structural ablations are provided in Supplementary Section S3.1.

2) *Seasonal and Regional Error Attribution*: Table 7 decomposes errors by ice regime (OW, MIZ, and pack ice) and season (melting, freezing, and transition).

Errors are concentrated in the MIZ, where MAE reaches 0.1192-0.1384 across seasons. Open-water errors are substantially lower, with MAE near 0.006, and pack ice errors remain relatively low. Thus, the primary modeling challenge does not lie in stable open water or the pack ice interior, but in the rapidly evolving ice-edge zone, where low-concentration ice, contour displacement, and freeze-melt phase changes interact.

3) *Thermodynamic Source-term Diagnostics*: The thermodynamic module is a sign-constrained dual-branch source term with SIC area-fraction weighting: the freezing branch contributes non-negative increments weighted by open-water fraction, and the melting branch contributes non-positive increments weighted by ice-covered fraction. The Table 8 reports seasonal mean source terms, and Table 9 reports responses to thermodynamic forcing perturbations.

The net source is negative in the melting season, positive in the freezing season, and intermediate during the transition season. Warming perturbations in  $T_{air}$  and SST both shift the net source toward melting, with a stronger response to SST. These results indicate that the thermodynamic source term

TABLE IX  
RESPONSE OF THERMODYNAMIC SOURCE TERMS TO EXTERNAL  
THERMODYNAMIC FORCING PERTURBATIONS

| Perturbation experiment | Mean combined source response |
|-------------------------|-------------------------------|
| Tair +1 °C              | -0.00091                      |
| SST +0.5 °C             | -0.00490                      |

TABLE X  
COMPARISON OF 99-CASE FORECAST-FORCED PREDICTION UNDER  
GLORYS VERIFICATION

| Model     | MAE           | RMSE          | Bias          | ACC           | IIEE          |
|-----------|---------------|---------------|---------------|---------------|---------------|
| IceNet    | 0.0477        | 0.1112        | -0.0106       | 0.7065        | 0.5967        |
| SimVP     | 0.0457        | 0.1077        | 0.0197        | 0.7283        | 0.5799        |
| VMRNN     | 0.0399        | 0.0988        | 0.0100        | 0.7576        | 0.5329        |
| SwinLSTM  | 0.0386        | 0.0972        | 0.0080        | 0.7633        | 0.5196        |
| FCNet     | 0.0385        | 0.0980        | <b>0.0030</b> | 0.7372        | 0.5336        |
| PredRNNv2 | 0.0367        | 0.0942        | 0.0098        | 0.7778        | 0.5354        |
| PredRNN++ | 0.0356        | 0.0958        | 0.0046        | 0.7557        | 0.5288        |
| PIHIM     | <b>0.0321</b> | <b>0.0912</b> | 0.0082        | <b>0.7839</b> | <b>0.4711</b> |

has the expected seasonal phase and perturbation direction. Supplementary diagnostics further show that the weak global-mean melt signal mainly reflects post-transport correction demand and dynamical-state effects, rather than insufficient melting-response capacity of the dual-branch structure.

#### D. Forecast-forced Prediction

Following the forecast-forced prediction setting in Section 4.1 and the information-availability constraints in Supplementary Section S1.3, forecast sea ice velocity is excluded because short-range drift is sensitive to wind, ocean currents, internal ice mechanics, and data assimilation differences, with reported error growth in drift/deformation forecasts [61]–[63]. Additional diagnostics based on RTOFS forecast sea ice velocity forcing are provided in Supplementary Section S9. Incorporating reliable forecast sea ice velocity and recalibrating the dynamical module are left for future work.

Two test sets are reported: a 99-case set initialized from 2024-04-01 to 2026-04-15 for overall forecast-forced evaluation, and a SEAS5-intersection 25-case set for same-date contextual comparison. GLORYS and OSI SAF serve as verification references; complete test-set definitions and supplementary results are provided in Supplementary Section S4.

1) *99-case Forecast-forced Prediction Results:* Table 10 and Table 11 report the D1-D9 mean results for the 99-case set under GLORYS and OSI SAF verification, respectively.

PIHIM obtains the lowest MAE, RMSE, and IIEE and the highest ACC under both verification references. Although all models show larger errors under OSI SAF than under GLORYS, PIHIM retains a consistent advantage, suggesting that its performance is not specific to the GLORYS reference.

2) *SEAS5-intersection 25-case Contextual Comparison:* The 25-case comparison uses initialization dates common to SEAS5 and PIHIM. SEAS5-EnsMean is the 51-member mean, while SEAS5-BestMember and SEAS5-WorstMember characterize the ensemble range rather than task-equivalent baselines.

TABLE XI  
COMPARISON OF 99-CASE FORECAST-FORCED PREDICTION UNDER OSI  
SAF VERIFICATION

| Model     | MAE           | RMSE          | Bias           | ACC           | IIEE          |
|-----------|---------------|---------------|----------------|---------------|---------------|
| SimVP     | 0.0613        | 0.1437        | 0.0227         | 0.7076        | 0.9656        |
| VMRNN     | 0.0594        | 0.1487        | 0.0115         | 0.6843        | 1.0365        |
| SwinLSTM  | 0.0591        | 0.1475        | 0.0106         | 0.6887        | 1.0267        |
| FCNet     | 0.0586        | 0.1441        | 0.0080         | 0.6879        | 0.9774        |
| PredRNN++ | 0.0572        | 0.1489        | 0.0094         | 0.6707        | 1.0054        |
| IceNet    | 0.0554        | 0.1249        | -0.0167        | 0.7939        | 0.9106        |
| PredRNNv2 | 0.0545        | 0.1386        | 0.0113         | 0.7212        | 0.9579        |
| PIHIM     | <b>0.0448</b> | <b>0.1157</b> | <b>-0.0004</b> | <b>0.8071</b> | <b>0.8473</b> |

TABLE XII  
SEAS5-INTERSECTION 25-CASE RESULTS UNDER GLORYS  
VERIFICATION

| Model             | MAE           | RMSE          | Bias          | ACC           | IIEE          |
|-------------------|---------------|---------------|---------------|---------------|---------------|
| SEAS5-WorstMember | 0.0614        | 0.1345        | -0.0163       | 0.6035        | 0.8566        |
| SEAS5-EnsMean     | 0.0597        | 0.1310        | -0.0162       | 0.6148        | 0.8253        |
| SEAS5-BestMember  | 0.0597        | 0.1315        | -0.0159       | 0.6169        | 0.8208        |
| IceNet            | 0.0479        | 0.1110        | -0.0107       | 0.7073        | 0.5851        |
| SimVP             | 0.0463        | 0.1088        | 0.0207        | 0.7286        | 0.5756        |
| VMRNN             | 0.0399        | 0.0984        | 0.0106        | 0.7619        | 0.5236        |
| FCNet             | 0.0394        | 0.0990        | <b>0.0039</b> | 0.7357        | 0.5303        |
| SwinLSTM          | 0.0385        | 0.0966        | 0.0087        | 0.7700        | 0.5075        |
| PredRNNv2         | 0.0367        | 0.0940        | 0.0107        | 0.7824        | 0.5232        |
| PredRNN++         | 0.0356        | 0.0954        | 0.0046        | 0.7629        | 0.5201        |
| PIHIM             | <b>0.0319</b> | <b>0.0903</b> | 0.0088        | <b>0.7897</b> | <b>0.4561</b> |

TABLE XIII  
SEAS5-INTERSECTION 25-CASE RESULTS UNDER OSI SAF  
VERIFICATION

| Model             | MAE           | RMSE          | Bias           | ACC           | IIEE          |
|-------------------|---------------|---------------|----------------|---------------|---------------|
| SEAS5-WorstMember | 0.0675        | 0.1473        | -0.0029        | 0.6810        | 1.2147        |
| SEAS5-EnsMean     | 0.0657        | 0.1441        | -0.0036        | 0.6942        | 1.1638        |
| SEAS5-BestMember  | 0.0656        | 0.1448        | -0.0034        | 0.6929        | 1.1644        |
| SimVP             | 0.0624        | 0.1457        | 0.0242         | 0.7107        | 0.9996        |
| FCNet             | 0.0603        | 0.1463        | 0.0097         | 0.6900        | 1.0144        |
| VMRNN             | 0.0600        | 0.1499        | 0.0119         | 0.6910        | 1.0779        |
| SwinLSTM          | 0.0598        | 0.1486        | 0.0114         | 0.6955        | 1.0668        |
| PredRNN++         | 0.0582        | 0.1501        | 0.0100         | 0.6783        | 1.0393        |
| IceNet            | 0.0563        | 0.1264        | -0.0166        | 0.7965        | 0.9390        |
| PredRNNv2         | 0.0551        | 0.1398        | 0.0123         | 0.7277        | 0.9968        |
| PIHIM             | <b>0.0453</b> | <b>0.1164</b> | <b>-0.0005</b> | <b>0.8129</b> | <b>0.8745</b> |

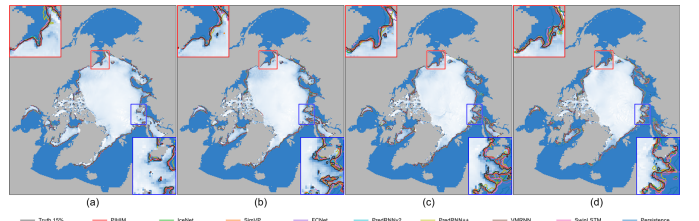


Fig. 5. Rapid ice-edge retreat case in reanalysis-forced simulation. Panels (a)-(d) correspond to D2, D6, D10, and D14 from the July 2, 2024 initialization.

The 25-case results are consistent with the 99-case evaluation. Under both verification references, PIHIM outperforms the selected learning baselines and SEAS5 ensemble summaries, while SEAS5 remains a contextual numerical reference rather than part of the full 99-case comparison.

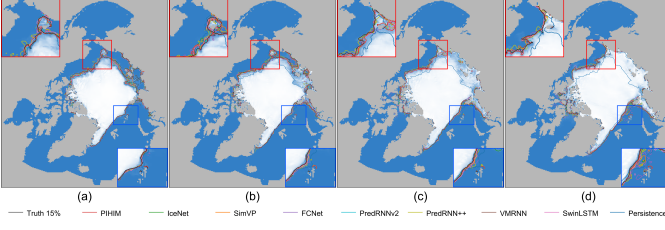


Fig. 6. Rapid ice-edge expansion case in reanalysis-forced simulation. Panels (a)-(d) correspond to D2, D6, D10, and D14 from the October 16, 2024 initialization.

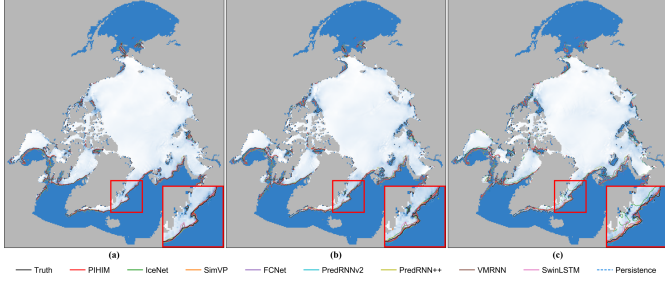


Fig. 7. GLORYS-verified ice-edge case in 99-case forecast-forced prediction. Panels (a)-(c) correspond to D3, D6, and D9 from the May 22, 2024 initialization.

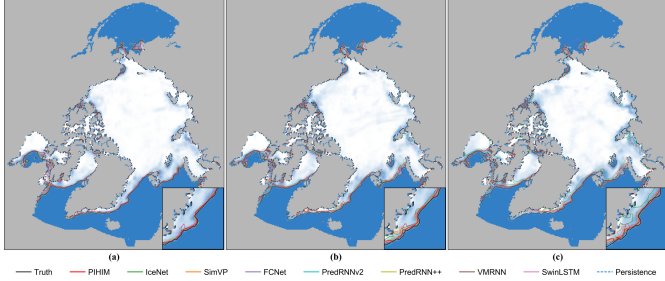


Fig. 8. OSI SAF-verified ice-edge case in 99-case forecast-forced prediction. Panels (a)-(c) correspond to D3, D6, and D9 from the May 22, 2024 initialization.

### E. Ice-edge Spatial Case Analysis

Spatial cases supplement the global statistics by comparing 15% SIC ice-edge contours and local morphology under rapid retreat/expansion and forecast-forced conditions.

1) *Rapid Retreat and Expansion Cases in Reanalysis-forced Simulation:* The July 2 and October 16, 2024 cases correspond to the fastest 14-day retreat and expansion diagnostics. Fig. 5 and Fig. 6 show that PIHIM keeps the 15% contour closer to GLORYS than the baselines under both conditions; complete D1-D14 sequences are provided in Supplementary Section S5.1.

2) *99-case Forecast-forced Prediction Spatial Case:* Fig. 7 and Fig. 8 present the May 22, 2024 forecast-forced case under GLORYS and OSI SAF verification. PIHIM remains comparatively stable from D3 to D9, consistent with the 99-case statistics; complete sequences are provided in Supplementary Section S5.2.

3) *Contextual Spatial Comparison on the SEAS5-intersection Samples:* Fig. 9 and Fig. 10 provide contextual spatial comparisons with SEAS5 for the April 1, 2026

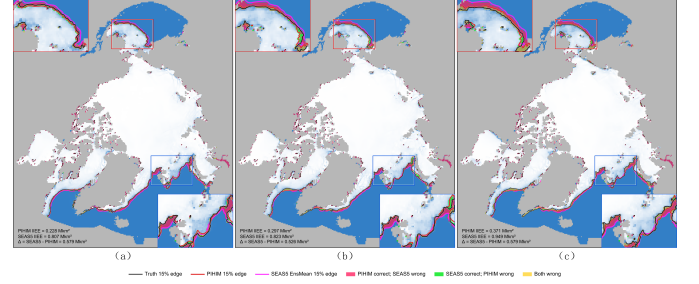


Fig. 9. GLORYS-verified ice-edge comparison in SEAS5-intersection 25-case forecast-forced prediction. Panels (a)-(c) correspond to D3, D6, and D9 from the April 1, 2026 initialization.

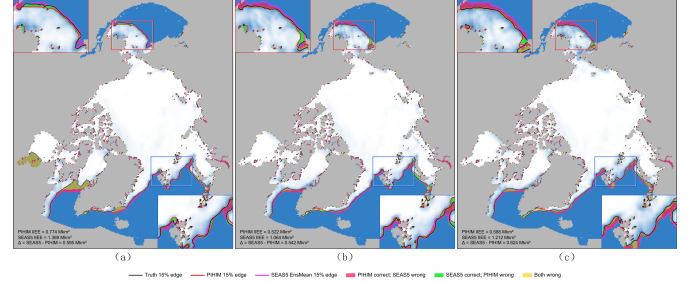


Fig. 10. OSI SAF-verified ice-edge comparison in SEAS5-intersection 25-case forecast-forced prediction. Panels (a)-(c) correspond to D3, D6, and D9 from the April 1, 2026 initialization.

initialization on the 25-case set. Under both verification references, PIHIM shows smaller ice-edge displacement than SEAS5 EnsMean, consistent with Table 12 and Table 13; complete D1-D9 sequences are provided in Supplementary Section S5.3.

## V. CONCLUSION

This study proposed PIHIM for daily Arctic SIC evolution modeling and short-range prediction. Guided by the sea ice continuity equation, PIHIM decomposes SIC evolution into dynamical transport, thermodynamic areal change, and residual compensation, providing a process-decomposed alternative to purely end-to-end SIC sequence models. Experiments show that PIHIM improves ice-edge preservation, error-growth control, and areal stability in reanalysis-forced simulation, while retaining consistent short-range advantages under GLORYS and OSI SAF verification in forecast-forced prediction. Module diagnostics further indicate complementary roles of dynamical transport, thermodynamic source-term representation, and residual compensation in spatial redistribution, freeze/melt areal change, and local error closure.

PIHIM remains a lightweight data-driven model rather than a full numerical sea ice model. Its thermodynamic module does not explicitly resolve sea ice thickness, snow cover, melt ponds, albedo, or fine-scale energy-budget processes, and forecast-forced prediction still depends on external atmospheric and oceanic forecast quality. Future work will incorporate more reliable thickness, energy-budget, and sea ice motion information, refine thermodynamic source-term representation and forecast-forcing error attribution, and explore integration with data assimilation and broader sea ice applications.

## CONFLICT OF INTEREST

None of the authors have a conflict of interest to disclose.

## REFERENCES

- [1] Serreze, M. C. & Meier, W. N. The Arctic's sea ice cover: trends, variability, predictability, and comparisons to the Antarctic. *Ann. N. Y. Acad. Sci.* **1436**, 36–53 (2019).
- [2] Screen, J. A. & Simmonds, I. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature* **464**, 1334–1337 (2010).
- [3] Post, E. et al. Ecological Consequences of Sea-Ice Decline. *Science* **341**, 519–524 (2013).
- [4] Stroeve, J. & Notz, D. Changing state of Arctic sea ice across all seasons. *Environ. Res. Lett.* **13**, 103001; <https://doi.org/10.1088/1748-9326/aade56> (2018).
- [5] Kwok, R. Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018). *Environ. Res. Lett.* **13**, 105005; <https://doi.org/10.1088/1748-9326/aac3ec> (2018).
- [6] Melia, N., Haines, K. & Hawkins, E. Sea ice decline and 21st century trans-Arctic shipping routes. *Geophys. Res. Lett.* **43**, 9720–9728 (2016).
- [7] Hibler, W. D. III. A Dynamic Thermodynamic Sea Ice Model. *J. Phys. Oceanogr.* **9**, 815–846 (1979).
- [8] Notz, D. Challenges in simulating sea ice in Earth System Models. *WIREs Clim. Change* **3**, 509–526 (2012).
- [9] Olonscheck, D., Mauritsen, T. & Notz, D. Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations. *Nat. Geosci.* **12**, 430–434 (2019).
- [10] Massonnet, F. et al. Arctic sea-ice change tied to its mean state through thermodynamic processes. *Nat. Clim. Change* **8**, 599–603 (2018).
- [11] Rolph, R. J., Feltham, D. L. & Schroder, D. Changes of the Arctic marginal ice zone during the satellite era. *Cryosphere* **14**, 1971–1984 (2020).
- [12] Horvat, C. Marginal ice zone fraction benchmarks sea ice and climate model skill. *Nat. Commun.* **12**, 2221; <https://doi.org/10.1038/s41467-021-22004-7> (2021).
- [13] Bushuk, M. et al. Predicting September Arctic Sea Ice: A Multimodel Seasonal Skill Comparison. *Bull. Am. Meteorol. Soc.* **105**, E1170–E1203 (2024).
- [14] Andersson, T. R. et al. Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nat. Commun.* **12**, 5124; <https://doi.org/10.1038/s41467-021-25257-4> (2021).
- [15] Ren, Y., Li, X. & Zhang, W. A Data-Driven Deep Learning Model for Weekly Sea Ice Concentration Prediction of the Pan-Arctic During the Melting Season. *IEEE Trans. Geosci. Remote Sens.* **60**, 4304819; <https://doi.org/10.1109/TGRS.2022.3177600> (2022).
- [16] Zheng, Q. et al. A Spatiotemporal Multiscale Deep Learning Model for Subseasonal Prediction of Arctic Sea Ice. *IEEE Trans. Geosci. Remote Sens.* **62**, 4300522; <https://doi.org/10.1109/TGRS.2024.3355238> (2024).
- [17] Johnson, S. J. et al. SEAS5: the new ECMWF seasonal forecast system. *Geosci. Model Dev.* **12**, 1087–1117 (2019).
- [18] Palermé, C. et al. Improving short-term sea ice concentration forecasts using deep learning. *Cryosphere* **18**, 2161–2176 (2024).
- [19] Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
- [20] Li, Y. et al. An Explainable Deep Learning Model for Daily Sea Ice Concentration Forecast. *IEEE Trans. Geosci. Remote Sens.* **62**, 4301217; <https://doi.org/10.1109/TGRS.2024.3386930> (2024).
- [21] Liu, Q. et al. Physics-informed deep convolutional network for combined sea ice concentration and velocity prediction. *Ocean Eng.* **313**, 119440; <https://doi.org/10.1016/j.oceaneng.2024.119440> (2024).
- [22] Thorndike, A. S., Rothrock, D. A., Maykut, G. A. & Colony, R. The thickness distribution of sea ice. *J. Geophys. Res.* **80**, 4501–4513 (1975).
- [23] Lipscomb, W. H. & Hunke, E. C. Modeling Sea Ice Transport Using Incremental Remapping. *Mon. Weather Rev.* **132**, 1341–1354 (2004).
- [24] Willard, J., Jia, X., Xu, S., Steinbach, M. & Kumar, V. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. *ACM Comput. Surv.* **55**, 1–37 (2022).
- [25] Zanna, L. & Bolton, T. Data-Driven Equation Discovery of Ocean Mesoscale Closures. *Geophys. Res. Lett.* **47**, e2020GL088376; <https://doi.org/10.1029/2020GL088376> (2020).
- [26] Rasp, S., Pritchard, M. S. & Gentile, P. Deep learning to represent subgrid processes in climate models. *Proc. Natl. Acad. Sci. USA* **115**, 9684–9689 (2018).
- [27] Murphy, A. H. & Winkler, R. L. A General Framework for Forecast Verification. *Mon. Weather Rev.* **115**, 1330–1338 (1987).
- [28] He, Z., Wang, Y., Brajard, J., Wang, X. & Shen, Z. Correcting errors in seasonal Arctic sea ice prediction of Earth system models with machine learning. *Cryosphere* **19**, 3279–3293 (2025).
- [29] Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A. & Zanna, L. Machine Learning for Online Sea Ice Bias Correction Within Global Ice-Ocean Simulations. *Geophys. Res. Lett.* **51**, e2023GL106776; <https://doi.org/10.1029/2023GL106776> (2024).
- [30] Gregory, W. et al. Advancing global sea ice prediction capabilities using a fully coupled climate model with integrated machine learning. *Sci. Adv.* **12**, eady8957; <https://doi.org/10.1126/sciadv.ady8957> (2026).
- [31] Chen, X. et al. A comparative study of data input selection for deep learning-based automated sea ice mapping. *Int. J. Appl. Earth Obs. Geoinf.* **131**, 103920; <https://doi.org/10.1016/j.jag.2024.103920> (2024).
- [32] Wang, W., Yang, W., Wang, L., Wang, G. & Lei, R. Seasonal forecasting of Pan-Arctic sea ice with state space model. *npj Clim. Atmos. Sci.* **8**, 172; <https://doi.org/10.1038/s41612-025-01058-0> (2025).
- [33] Li, Y. et al. Significance of Atlantic sea surface temperature anomalies to Arctic sea ice variability revealed by deep learning. *npj Clim. Atmos. Sci.* **9**, 70; <https://doi.org/10.1038/s41612-026-01347-2> (2026).
- [34] Grigoryev, T. et al. Data-Driven Short-Term Daily Operational Sea Ice Regional Forecasting. *Remote Sens.* **14**, 5837; <https://doi.org/10.3390/rs14225837> (2022).
- [35] Xiao, C. et al. Prediction of long lead monthly three-dimensional ocean temperature using time series gridded Argo data and a deep learning method. *Int. J. Appl. Earth Obs. Geoinf.* **112**, 102971; <https://doi.org/10.1016/j.jag.2022.102971> (2022).
- [36] Zhou, G., Chen, J., Liu, M. & Ma, L. A spatiotemporal attention-augmented ConvLSTM model for ocean remote sensing reflectance prediction. *Int. J. Appl. Earth Obs. Geoinf.* **129**, 103815; <https://doi.org/10.1016/j.jag.2024.103815> (2024).
- [37] Niu, Y. et al. A deep learning global ocean forecasting model with sub-daily and eddy-resolving resolution. *npj Clim. Atmos. Sci.*; <https://doi.org/10.1038/s41612-026-01444-2> (2026).
- [38] Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
- [39] Ren, Y. & Li, X. Predicting the Daily Sea Ice Concentration on a Subseasonal Scale of the Pan-Arctic During the Melting Season by a Deep Learning Model. *IEEE Trans. Geosci. Remote Sens.* **61**, 4301315; <https://doi.org/10.1109/TGRS.2023.3279089> (2023).
- [40] Ren, Y., Li, X. & Wang, Y. SICNet<sub>season</sub> V1.0: a transformer-based deep learning model for seasonal Arctic sea ice prediction by incorporating sea ice thickness data. *Geosci. Model Dev.* **18**, 2665–2678 (2025).
- [41] Liu, X., Deng, W., Liu, L., Ding, J. & Cheng, X. Sea ice concentration estimation via physical information-guided multi-source data fusion and spatial continuity preservation. *Int. J. Appl. Earth Obs. Geoinf.* **150**, 105339; <https://doi.org/10.1016/j.jag.2026.105339> (2026).
- [42] Su, X. et al. A lightweight physics-aware framework for multi-scale marine heatwaves forecasting. *npj Clim. Atmos. Sci.* **9**, 104; <https://doi.org/10.1038/s41612-026-01367-y> (2026).
- [43] Meng, Y., Rigall, E., Chen, X., Gao, F., Dong, J. & Chen, S. Physics-Guided Generative Adversarial Networks for Sea Subsurface Temperature Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **34**, 3357–3370; <https://doi.org/10.1109/TNNLS.2021.3123968> (2023).
- [44] Mao, Y., Gu, Y., Sha, L., Shao, H., Wang, Q. & Abdelzaher, T. Phys-Taylor: Partially Physics-Knowledge-Enhanced Deep Neural Networks via NN Editing. *IEEE Trans. Neural Netw. Learn. Syst.* **36**, 447–461; <https://doi.org/10.1109/TNNLS.2023.3325432> (2025).
- [45] Wang, F. et al. Inherently Interpretable Physics-Informed Neural Network for Battery Modeling and Prognosis. *IEEE Trans. Neural Netw. Learn. Syst.* **36**, 1145–1159; <https://doi.org/10.1109/TNNLS.2023.3329368> (2025).
- [46] Menicali, L., Richter, D. H. & Castruccio, S. Bayesian Neural Networks With Physics-Informed Priors With Application to Boundary Layer Velocity. *IEEE Trans. Neural Netw. Learn. Syst.* **36**, 18048–18061; <https://doi.org/10.1109/TNNLS.2025.3577508> (2025).
- [47] Gregory, W. et al. FloeNet: A mass-conserving global sea ice emulator that generalizes across climates. *arXiv* 2603.12449; <https://doi.org/10.48550/arXiv.2603.12449> (2026).
- [48] Dupont, T. F. & Liu, Y. Back and forth error compensation and correction methods for semi-Lagrangian schemes with application to level set interface computations. *Math. Comput.* **76**, 647–668 (2007).
- [49] Selle, A., Fedkiw, R., Kim, B.-M., Liu, Y. & Rossignac, J. An Unconditionally Stable MacCormack Method. *J. Sci. Comput.* **35**, 350–371 (2008).
- [50] Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

- [51] Lellouche, J.-M. et al. The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. *Front. Earth Sci.* **9**, 698876; <https://doi.org/10.3389/feart.2021.698876> (2021).
- [52] Lang, S. et al. AIFS – ECMWF’s data-driven forecasting system. *arXiv* 2406.01465; <https://doi.org/10.48550/arXiv.2406.01465> (2024).
- [53] Chassignet, E. P. et al. US GODAE: Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM). *Oceanography* **22**, 64–75 (2009).
- [54] Lavergne, T. et al. Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records. *Cryosphere* **13**, 49–78 (2019).
- [55] Gao, Z., Tan, C., Wu, L. & Li, S. Z. SimVP: Simpler yet Better Video Prediction. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 3170–3180 (IEEE, 2022).
- [56] Zhang, J., Gao, F., Gan, Y., Dong, J. & Du, Q. Frequency-Compensated Network for Daily Arctic Sea Ice Concentration Prediction. *IEEE Trans. Geosci. Remote Sens.* **63**, 5104315; <https://doi.org/10.1109/TGRS.2025.3564457> (2025).
- [57] Wang, Y. et al. PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 2208–2225 (2023).
- [58] Tang, Y., Dong, P., Tang, Z., Chu, X. & Liang, J. VMRNN: Integrating Vision Mamba and LSTM for Efficient and Accurate Spatiotemporal Forecasting. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* 5663–5673; <https://doi.org/10.1109/CVPRW63382.2024.00575> (IEEE, 2024).
- [59] Tang, S., Li, C., Zhang, P. & Tang, R. SwinLSTM: Improving Spatiotemporal Prediction Accuracy using Swin Transformer and LSTM. In *Proc. IEEE/CVF International Conference on Computer Vision* 13424–13433; <https://doi.org/10.1109/ICCV51070.2023.01239> (IEEE, 2023).
- [60] Wang, Y., Gao, Z., Long, M., Wang, J. & Yu, P. S. PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in Spatiotemporal Predictive Learning. In *Proc. International Conference on Machine Learning* 5123–5132 (PMLR, 2018).
- [61] Mohammadi-Aragh, M., Goessling, H. F., Losch, M., Hutter, N. & Jung, T. Predictability of Arctic sea ice on weather time scales. *Sci. Rep.* **8**, 6514; <https://doi.org/10.1038/s41598-018-24660-0> (2018).
- [62] Palerme, C. & Müller, M. Calibration of sea ice drift forecasts using random forest algorithms. *Cryosphere* **15**, 3989–4004 (2021).
- [63] Korosov, A., Rampal, P., Ying, Y., Ólason, E. & Williams, T. Towards improving short-term sea ice predictability using deformation observations. *Cryosphere* **17**, 4223–4240 (2023).