JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2025MS005457

Arctic High Resolution Sea-Ice Forecast System on a Heterogeneous Many-Core Computing Platform

Longjiang Mu[1], Yue Liu[2], Yuhu Chen[1], Hong Wang[1], Xianyao Chen[3], Ruizhe Song[3], Lin Zheng[2], Junting Chen[2], Lars Nerger[4], and Lixin Wu[1,3]

[1]Laoshan Laboratory, Qingdao, China, [2]Qingdao Gosci Technology Group, Qingdao, China, [3]Ocean University of China, Qingdao, China, [4]Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar-und Meeresforschung, Bremerhaven, Germany

Key Points:

- A high resolution sea-ice forecast system is developed to forecast the linear kinematic features in Arctic sea ice
- The assimilation system is refactored and optimized on a heterogeneous many-core computing platform with a speedup of 3.66×
- The linear kinematic features predictability is shown via distance, cross-correlation, orientation, and length from a hindcast experiment

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Chen and L. Wu,
chenxy@ouc.edu.cn;
lxwu@ouc.edu.cn

Citation:

Mu, L., Liu, Y., Chen, Y., Wang, H., Chen, X., Song, R., et al. (2026). Arctic high resolution sea-ice forecast system on a heterogeneous many-core computing platform. *Journal of Advances in Modeling Earth Systems*, *18*, e2025MS005457. https://doi.org/10.1029/2025MS005457

Received 16 SEP 2025
Accepted 10 MAR 2026

Author Contributions:

Conceptualization: Longjiang Mu
Data curation: Longjiang Mu
Formal analysis: Longjiang Mu
Funding acquisition: Longjiang Mu, Xianyao Chen, Lixin Wu
Investigation: Longjiang Mu
Methodology: Longjiang Mu, Lars Nerger
Project administration: Lixin Wu
Software: Longjiang Mu, Yue Liu, Yuhu Chen, Hong Wang, Lin Zheng, Junting Chen, Lars Nerger
Validation: Longjiang Mu

Abstract This study develops an ensemble forecast system for Arctic sea ice with a horizontal resolution of ∼2 km to resolve linear kinematic features (LKFs). The system is designed to run operationally on the Sunway computing architecture, which employs many-core processors with a distinct programming logic. The Parallel Data Assimilation Framework has been optimized for Sunway's architecture by leveraging fine-grained thread-level parallelism and restructuring memory hierarchies. Results from multiple process configurations show that the optimized analysis step achieves an average speedup of 9.19×, while overall performance improves by an average of 3.66× compared to CPU-based architectures. To address the strong nonlinearity of LKFs, we employ ice strength parameter perturbations and a novel localized observation error function that preserves valuable LKF information from both the model and sea ice concentration observations. Compared to Synthetic Aperture Radar observations at 1–3-day lead times, the hindcast experiment yields a bidirectionally averaged minimum Hausdorff distance of 39.7–41.9 km for large-scale LKFs, and outperforms persistence forecasts in spatial maximum cross-correlation error within the 48 hr lead time. The probability density functions of LKF orientation and length largely follow the observed distributions. Results indicate that LKF predictability can be enhanced through model dynamics inherently from well-initialized sea ice states like concentration or thickness, despite only limited details of LKFs in the observations used during assimilation. With higher-frequency data assimilation, this system shows potential for operational LKF forecasting in support of practical Arctic navigation.

Plain Language Summary Sea ice leads and ridges are common features in the Arctic. These common features form complex patterns that are crucial for modeling. Researchers are particularly keen to understand whether these features can be predicted and how predictable they are from a near real-time approach. We developed a high-resolution sea ice forecasting system with a resolution of ∼2 km and investigated the predictability of sea ice leads and ridges on a powerful computer. The results show that even with the assimilation of sea ice concentration alone, the system can predict large-scale sea ice leads and ridges within the 48 hr lead time.

1. Introduction

The decline of Arctic sea ice extent and thickness in recent decades marks a fundamental transformation of the polar environment, heralding the emergence of a "New Arctic" (Landrum & Holland, 2020). This rapid change overflows scientific interests, capturing attention also from policymakers, industries, and environmental advocates alike. Synoptic sea ice forecasts are particularly essential for navigating the Arctic Passage, particularly in forecasting linear kinematic features (LKFs), which are narrow regions of enhanced sea ice deformation where leads and ridges form. Leads form interconnected waterways within the ice pack, resembling braided river networks that offer potential navigation routes. In contrast, ridges function as natural barricades, often requiring icebreaker support or detours. This duality underscores the importance of real-time LKF nowcasting, which distinguishes navigable "waterways" from impassable "walls."

Operationally, such short-term sea ice forecasting is conducted at specialized centers using either standalone regional models or sea ice components nested within global ocean forecast systems (Blockley et al., 2020). These systems typically operate at the resolution of ∼10 km-scale, where sea ice concentration and thickness serve as primary parameters—providing coarse estimates of sea ice state. Studies (Bouchat et al., 2022; Q. Wang et al., 2016) demonstrate that increased resolution is the most direct approach for simulating the LKFs—unless using specialized fracture rheologies like elasto-brittle, Maxwell-elasto-brittle, or brittle Bingham-Maxwell

Writing – review & editing: Longjiang Mu, Xianyao Chen, Ruizhe Song, Lars Nerger

(Dansereau et al., 2016; Girard et al., 2011; Olason et al., 2022). It is clear that the inherent resolution limitation fundamentally constrains LKF forecasting capability (Hutter et al., 2022).

LKFs exhibit complex dynamics with multifractal scaling properties (Bouchat et al., 2022; Hutter & Losch, 2020; Spreen et al., 2016), and their high nonlinearity challenges numerical models attempting to simulate their evolution, let alone predict them. Forecast skill demonstrably depends on accurate initialization of sea ice concentration (e.g., J. Liu et al., 2019; Tietscheet al., 2013; Y. F. Zhang et al., 2022) and thickness (e.g., Blockley & Peterson, 2018; Mu et al., 2018; Xie et al., 2016; Yang et al., 2015), with thickness assimilation proving particularly crucial for longer-term forecasts—a finding robustly documented through idealized and hindcast experiments (e.g., Mu et al., 2022; Song et al., 2024). Nevertheless, assimilation methodologies for the nonlinear features are rather limited, and the potential predictability of LKFs through improved initialization needs more extensive investigation. Korosov et al. (2023) assimilated observational sea ice deformation data to optimize the short-term predictability of LKFs in the neXt-generation Sea Ice Model (neXtSIM), and found forecast skill to improve if sea ice concentration is constrained in regions of increased deformation. A previous study also suggested a long-term impact from the representation of sea ice leads in numerical models which could improve the following summer Arctic sea ice extent prediction (Y. Zhang et al., 2018).

In the present study, we develop an Arctic high resolution sea ice forecast system based on the model configuration from the "Estimating the Circulation and Climate of the Ocean" (ECCO2, https://ecco-group.org/) project, which is also used in the Arctic Ice Ocean Prediction System initially developed by Mu et al. (2019) and updated in Liang et al. (2020). Our new implementation introduces four key advancements: (a) increased horizontal resolution (∼2 km); (b) redesigned data assimilation methodology; (c) data assimilation code refactoring optimized for heterogeneous many-core architectures; and (d) investigation of LKF predictability. The model configuration, data assimilation methodology, forecast system design, and evaluation approach are provided in Section 2. Section 3 details the optimization of the ensemble forecast system on the Sunway architecture. Assimilation results are presented in Section 4, while Section 5 evaluates LKF forecast performance. Discussion and summary are provided in Section 6.

2. Methods and Data

2.1. The High Resolution Sea Ice Forecast System (HRSFS)

2.1.1. The High Resolution Arctic Regional Sea Ice-Ocean Coupled Model

The regional sea ice–ocean coupled model is configured based on the Massachusetts Institute of Technology general circulation model (MITgcm, Marshall et al., 1997) with a horizontal resolution of ∼2 km, which has been shown to successfully reproduce LKF statistics (Hutter & Losch, 2020). This configuration extracts one face from the global cubed-sphere grid, and discretizes the ocean and sea ice governing equations using the Arakawa C grid over the same regions. Similar configurations with lower resolution were described in Losch et al. (2010) and Nguyen et al. (2011). Vertically, the ocean model has 16 unevenly spaced layers considering the limited computation resources. The ocean state in the uppermost 120 m is represented within the first 5 layers to better simulate the ice-ocean interactions. The bathymetry is derived from "The Refined Topography data set 2" (RTopo 2) (Schaffer & Timmermann, 2016). The initial ocean hydrography is interpolated from the World Ocean Atlas 2005 (Antonov et al., 2006; Locarnini et al., 2006). Lateral boundary conditions are taken from monthly-mean ocean states from an optimized global ocean model also based on MITgcm (Menemenlis et al., 2008). Free slip conditions are used for the velocities on lateral boundaries. The diapycnal process is parameterized by the K-Profile Parameterization scheme (Large et al., 1994). Parameterizations used for the horizontal viscosity and in tracer equations take the same configuration as in Hutter and Losch (2020).

The sea ice model we use solves the momentum equation

$$m\partial_t \mathbf{v} + mf\mathbf{k}\times\mathbf{v} = -mg\nabla\phi + \boldsymbol{\tau}_\mathbf{a} + \boldsymbol{\tau}_\mathbf{o} + \nabla\cdot\boldsymbol{\sigma}, \tag{1}$$

where m is the ice load mass per unit area, f is the Coriolis parameter with vertical unit vector $\mathbf{k}$, $\nabla\phi$ is the reference density-weighted pressure gradient, $\boldsymbol{\tau}_\mathbf{a}$ and $\boldsymbol{\tau}_\mathbf{o}$ represent the atmosphere-ice and ice-ocean interfacial stresses, respectively, and σ is the internal ice stress tensor. Upon the assumption of normal flow rule and an elliptical yield curve, the closure of Equation 1 can be achieved via relating the internal ice stress σ to the sea ice strain rate $\dot{\epsilon}_{ij}$ by the

viscous-plastic (VP) rheology (Hibler, 1979) in a tensor form as $\sigma_{ij} = 2\eta\dot{\epsilon}_{ij} + \left[(\zeta - \eta)\dot{\epsilon}_{kk} - P/2\right]\delta_{ij}$, with strain rate $\dot{\epsilon}_{ij} = \frac{1}{2}(\partial_i v_j + \partial_j v_i)$. The ice compressive strength P is parameterized as $P = P^* h e^{-C^*(1-c)}$, where h is thickness, c concentration, C^* a dimensionless tuning parameter (typically 20), and P^* the ice strength constant. The bulk viscosity $\zeta = P/(2\Delta)$ and shear viscosity $\eta = \zeta/e^2$ depend on the aspect ratio e and deformation invariant $\Delta = \left(\dot{\epsilon}_I^2 + \dot{\epsilon}_{II}^2/e^2\right)^{1/2}$ calculated using two strain rate tensor invariants $\dot{\epsilon}_I$ and $\dot{\epsilon}_{II}$, where $\dot{\epsilon}_I = \dot{\epsilon}_{11} + \dot{\epsilon}_{22}$ (the divergence) and $\dot{\epsilon}_{II} = \sqrt{\left(\dot{\epsilon}_{11} - \dot{\epsilon}_{22}\right)^2 + 4\dot{\epsilon}_{12}^2}$ (the maximum shear strain). Sea ice deforms plastically when the internal stress reaches the elliptical yield curve, potentially forming either leads or ridges depending on the stress regime. Changes for sea ice thickness and sea ice concentration at local space arise from the divergence of their horizontal fluxes and the source terms from thermodynamics and ridge processes. No-slip lateral conditions are applied for sea ice velocities.

The sea ice momentum equation (Equation 1) is highly nonlinear due partly to the existence of both viscous and plastic movements in the system, which is difficult to solve resulting in high computation cost (Ip et al., 1991). At high resolutions, it is usually a bottleneck for the computation efficiency and can sometimes generate numerical noise along domain edges. Here, an iterative method with line successive relaxation (LSR) of the linearized momentum equations (J. Zhang & Hibler, 1997) is employed. The LSR method is rather a trade-off between efficiency and accuracy. The nonlinear steps are 10 times in each model step and during each nonlinear step the linear problem is solved with a tolerance of 10^{-4}, which is sufficient for data assimilation in our case.

The one-layer, zero heat capacity formulation (Parkinson & Washington, 1979; Semtner, 1976) is used for the sea ice thermodynamics. Sea ice is divided into seven thickness categories as in Hibler (1984) to allow proper ice growth also possible over thick ice regions. The ice thickness distribution (ITD) is not used though it is already provided in the MITgcm.

Sea ice ridge is generated when sea ice concentration in the cell exceeds 1.0 due to convergence in the advection step (Hibler, 1979). The sea ice albedos in dry or wet conditions are set to 0.69 and 0.66, respectively, and snow albedos are 0.84 and 0.79 correspondingly. The lead closing parameter H_0 in the model is set to 0.6074. The model is integrated from 1 January 2006 to 31 December 2019, which serves as baseline simulations for parameter tuning. Sea ice concentration and sea ice thickness from the Pan-Arctic Ice-Ocean Modeling and Assimilation System (J. Zhang & Rothrock, 2003) are used to initialize this run with a sea ice strength parameter $P^* = 2.264 \times 10^4$ Nm^{-2}. During the simulation period, the "JRA-55 based surface data set for driving ocean–sea-ice models (JRA55-do)" is used. The model is further integrated to 2021 with restart files output for hindcast experiments. An evaluation against satellite observations for simulated sea ice extent (Figure S1 in Supporting Information S1) and thickness (Figure S2 in Supporting Information S1) is provided in the supplementary.

2.1.2. Data Assimilation

Data assimilation in HRSFS employs an ensemble-based approach using the Parallel Data Assimilation Framework (PDAF, Nerger & Hiller, 2013, https://pdaf.awi.de). The analysis step is conducted offline and accelerated with highly parallelized functionality taking into account the large degrees of freedom inherent in high-resolution modeling. Each ensemble member is then initialized from the analyzed fields and continues the integration. During this process, the local Error Subspace Transform Kalman Filter (LESTKF), a variant of the classical ensemble Kalman Filter, is applied. The LESTKF is reported in Nerger et al. (2012) to preserve more ensemble information and therefore outperforms the Local Singular Evolutive Interpolated Kalman filter that was used in Mu et al. (2018). In addition, the Local Nonlinear Ensemble Transform Filter (LNETF, Tödter & Ahrens, 2015) that is able to capture nonlinearity in high-dimensional systems is tested for high resolution sea ice concentration assimilation in the study.

The state vector in the HRSFS assimilation system consists of sea ice concentration, sea ice thickness, and ocean temperature and salinity above 2,000.0 m. Ocean states below this depth undergo only minimal changes during the analysis and are therefore excluded to save memory and computational resources. Sea ice velocity is not included in the state vector due to its small inertia that usually has a rather short memory (Mu et al., 2022). The assimilated observation types are sea ice concentration, sea ice thickness, and sea surface temperature.

High-resolution modeling needs correspondingly high-resolution observations for effective data assimilation. Without such observations, the model state becomes overly smoothed and fails to constrain fine-scale spatio-temporal features. For sea ice concentration, our operational forecast system employs 3.25 km-resolution remote sensing data retrieved from the Advanced Microwave Scanning Radiometer 2 (AMSR2) in the University of Bremen (Melsheimer & Spreen, 2019). This product is selected because its spatial resolution is close to our model grid and it is operationally available with only a one-day latency. The ARTIST Sea Ice (Kaleschke et al., 2001) algorithm with tie points is used to convert the brightness temperatures in 89 GHz channels to sea ice concentrations. Although full calibration of this product remains incomplete, its depiction of sea ice leads—manifested as linear corridors of reduced concentration within consolidated ice—provides particularly valuable constraints for assimilation. An alternative 1 km-resolution product (Ludwig et al., 2020) merging MODIS thermal infrared data with AMSR2 observations offers higher spatial detail but suffers from longer latency and is only available in winter seasons due to cloud contamination. However, inspection of this higher resolution AMSR2 data reveals that only large-scale sea ice leads are captured. Nevertheless, it is currently the only available SIC containing sea ice lead information suitable for operational forecasting.

The CS2SMOS sea ice thickness operationally maintained in the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research is used to constrain the sea ice thickness state. Taking advantage of the complementary features between the retrieved thickness from CryoSat-2 and the Soil Moisture and Ocean Salinity (SMOS) (Ricker et al., 2017), CS2SMOS thickness is a merged product employing the optimal interpolation approach. The spatial resolution for CS2SMOS is 25 km, which is much lower than the model resolution. We adopt version 203 as it provides daily analysis. Despite its 2-day latency and resolution too coarse to resolve small-scale processes, this product remains effective for constraining the large-scale sea ice thickness in the model, which evolves slowly over short timescales.

The daily global sea surface temperature L4 product with 0.05° resolution provided by the Group for High Resolution Sea Surface Temperature (GHRSST, Good et al., 2020, http://www.ghrsst.org) is assimilated. This product merges satellite observations for sea surface temperature from multiple sensors on different platforms using the optimal interpolation method, and can be obtained operationally with one-day latency. In the Arctic Ocean, the resolution is around 1.8 km, which is close to the model's resolution.

2.1.3. Observation Errors

Observation errors normally used for data assimilation are combinations of the representation errors together with instrumental errors (Janjić et al., 2018). Careful treatment of these errors is critical as they dramatically affect the analyzed state vector. SIC exhibits the well-known non-Gaussian distribution. However, previous studies (e.g., Mu et al., 2019; Tietsche et al., 2013; Xie & Zhu, 2010; Yang et al., 2016) for SIC assimilation generally assumed Gaussian error distributions when applying the ensemble-based filters. Nonetheless, it turns out to work well for models with low horizontal resolution, where the nonlinearity in sea ice models such as LKFs is less prevalent and the smeared SIC by the filter still retains sufficient information.

For high-resolution forecasting where preserving small-scale features is essential, we propose a novel two-component observation error scheme that explicitly accounts for (a) the global non-Gaussian SIC distribution via an inverted Beta shape function and (b) local nonlinear structures through adaptive spatial scaling.

1. Shape function from inverted Beta distribution: Ice-free regions (SIC = 0) provide deterministic information in areas distant from the marginal ice zone, enabling ensemble filters to correct overestimated sea ice. Accounting for ice-free conditions reveals a bimodal SIC distribution with peaks at 0 and approaching 1 (Figure 1c, green line). This pattern is well-approximated by the Beta distribution (a continuous probability distribution). A uniform sea ice observation error for such distribution will surely degrade the performance of the filter in the area where LKFs are generated or around the sea ice edge (See Section 4.2). Intuitively, the SIC observation error should satisfy three requirements: (a) Approach zero as SIC $\rightarrow$ 0 (ice-free conditions); (b) Remain near-constant within the marginal ice zone, consistent with Gaussian error characteristics; (c) Be smaller in compact ice regions than in transition zones, yet larger than in ice-free areas. A function fulfilling these criteria can be derived from an inverted Beta distribution. We define this as a shape function $\psi(C)$ for SIC C, where $\psi : [0,1] \rightarrow \mathbb{R}^{+}$ represents observation error magnitude. The specific functional form is not unique; our system employs the empirical fit $\psi(C) = 0.01 \sec(2.95C + 1.63) + 0.17$ as shown in Figure 1c (blue curve), with $\psi(C)$ bounded between 0.01 and 0.16.

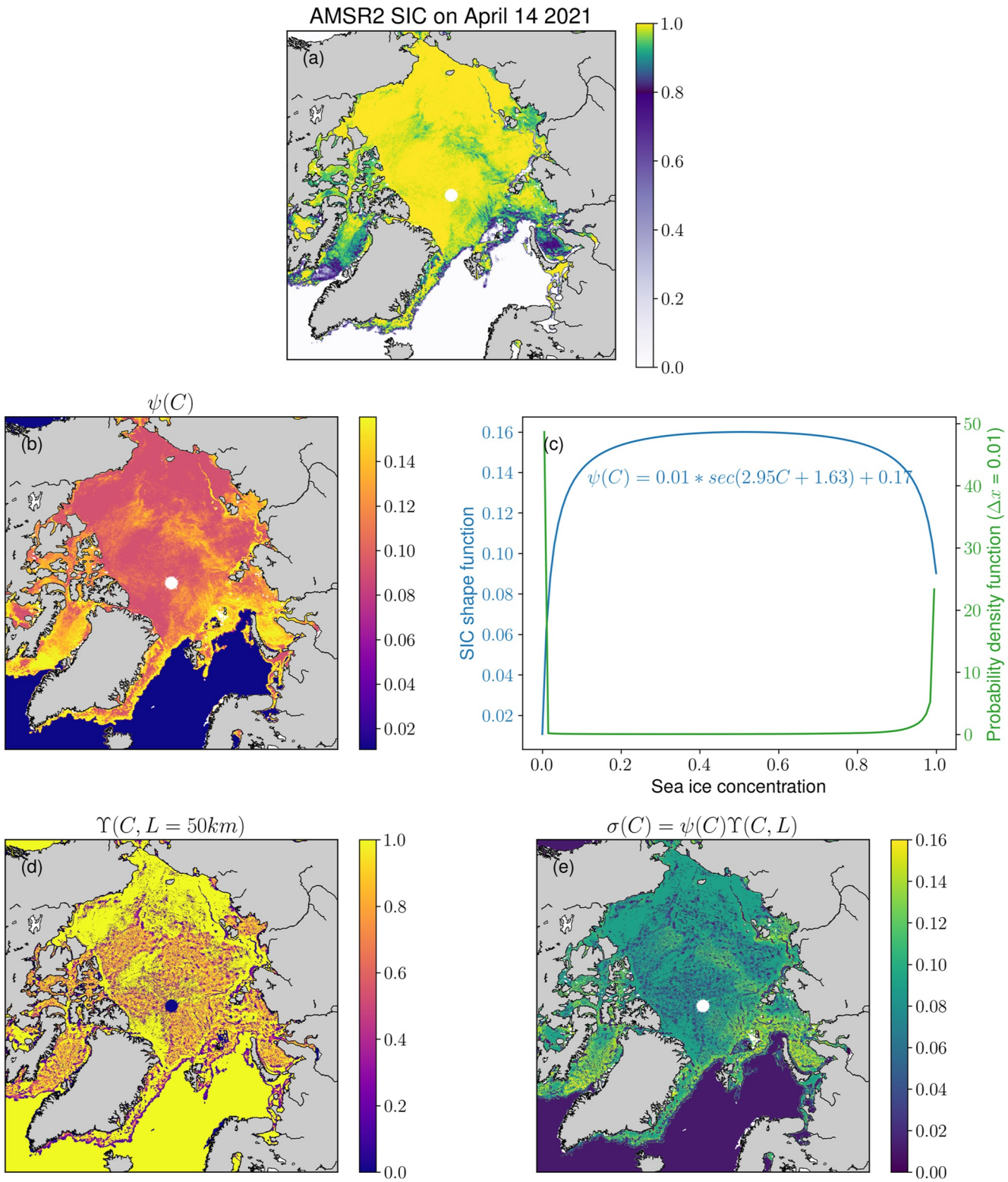

Figure 1. The SIC error on 14 April 2021. The Advanced Microwave Scanning Radiometer 2 SIC data used for assimilation are shown in (a). The shape function related to SIC (C) is shown in (b). The SIC distribution and the shape function used in our system are shown in (c) with green and blue lines, respectively. The nonlinearity function calculated with a local distance of $L = 50$km is shown in (d). The constructed SIC error $\sigma(C)$ used in the data assimilation system on that day is given in (e).

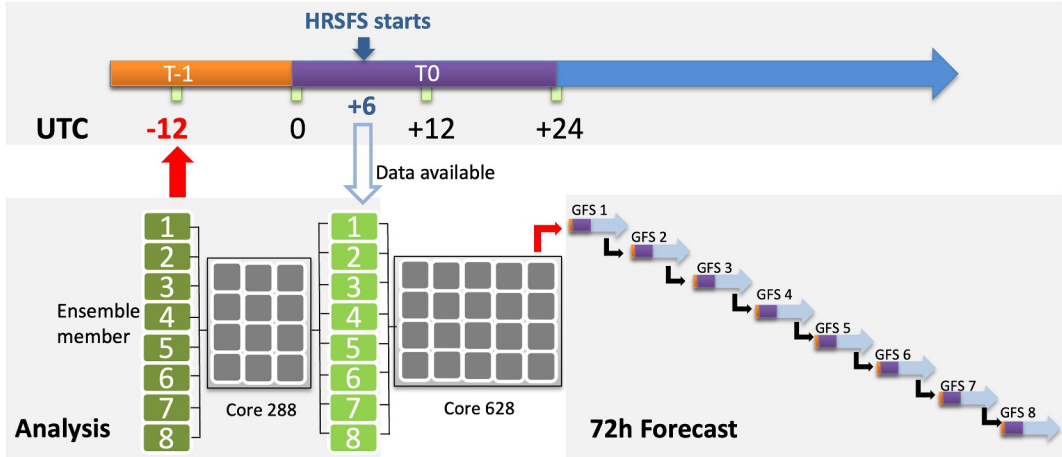

Figure 2. Schematic of the high-resolution sea ice forecast system (HRSFS) workflow. Atmospheric forcing data and observation data are preprocessed before 06:00 UTC daily. The analysis step (T-12hr) assimilates observations using eight ensemble members with 288 cores to generate the initial conditions. Subsequently, a 72-hr sequential ensemble forecast is initiated, with each of the eight members utilizing 628 cores.

2. Nonlinearity scaling for local features: The lead area in satellite derived products could have non-zero concentration such as the lead connecting to the New Siberian Islands in Figure 1a because the algorithm used for retrieval inevitably introduces smearing and averaging (Andersen et al., 2007). Given the current scarcity of observations resolving the fractal features in sea ice, such SIC information remains crucial for LKF analysis and forecasting. To explicitly incorporate these locally nonlinear effects, we introduce a scaling term $\Upsilon(C,L) = \mathbf{P}(C,L)/\mathbf{P}(C_{\text{mode}},L)$, where $\mathbf{P}(C,L)$ is the probability density function (PDF) of observed SIC values C within a local spatial window of scale L, and $C_{\text{mode}} = argmax\mathbf{P}(C,L)$ is the modal SIC value in that window. This term (Figure 1d) thus quantifies the deviation apart from the background value, with $\Upsilon \in (0,1]$. The complete SIC observation error σ (Figure 1e) is now the combination of the shape function (Figure 1b) with the nonlinearity function (Figure 1d), that is, $\sigma(C) = \psi(C)\Upsilon(C,L)$, where L is an empirical parameter that is tested to be around 50 km in our system.

For the ice thickness, the uncertainties provided in the CS2SMOS data product are directly applied for assimilation. The minimum error for sea ice thickness is set to 0.01 m, while a value of 0.5 m is assigned for sea ice thicker than 5.0 m. For sea surface temperature error, a global uniform value of 0.5°C is prescribed. This value is estimated by averaging the temperature errors provided in the data set over the model domain excluding sea ice covered regions.

2.2. Real-Time Forecast

We use the real-time atmospheric weather forecasts from the Global Ensemble Forecast System (GEFS) operated by National Centers for Environmental Prediction (NCEP) for the experiment. The downward short-wave and long-wave radiation fluxes, 10-m U and V wind components, 2-m specific humidity and temperature, and the total precipitation are used to generate the forcing fields. These variables are provided at 3-hourly intervals with 0.5° horizontal resolution.

The data-downloading and preprocessing for observations and atmospheric forecasts are fully prepared before 6:00 UTC each day. The HRSFS then executes its daily data assimilation cycle from that time (Figure 2). HRSFS has an ensemble size of eight for both analysis and forecast, correspondingly 8 members from the GEFS are selected. The ensemble size is not that large considering the computation cost is high for the 2 km resolution model. During the analysis step, the LESTKF filter corrects the background state vector (propagated from the forecast of the previous cycle) considering both simulation errors approximated by the covariances in the ensemble states and the observation errors. The weights for observations within the local range around the analysis grid point are calculated via a quasi-Gaussian weight function (Gaspari & Cohn, 1999). During this process, 288 cores are used for the parallel data assimilation.

Following the analysis step 12 hr prior, HRSFS initiates a 3-day ensemble sea ice forecast driven by atmospheric predictions. To estimate the probability space of the initial ensemble, we construct a matrix of state vectors from high-frequency daily model output in April 2021. After removing its temporal mean, this matrix is decomposed into empirical orthogonal functions (EOFs). Initial perturbations are then generated via second-order exact sampling and reconstructed using the leading eight EOFs and their corresponding singular values. To further enhance ensemble spread, random perturbations preserving the residual mean and covariance structure are added. Each ensemble member is then initialized by applying a corresponding perturbation to the state on the starting date. The simulation for each ensemble member uses 628 cores and is conducted in sequence, spanning from $T - 12$hr to $T + 60$hr. Since SIC follows a binomial distribution with more values at 0.9, inadequate forecast spread would cause rapid ensemble collapse. GEFS ensemble forecasts provide additional system uncertainties, while sea ice strength parameters (P^*) are perturbed across members to generate diverse lead characteristics (length, curvature, orientation). Specifically, P^* is varied linearly across the eight ensemble members from 1.464×10^4 to 2.864×10^4 with a constant increment of 0.175×10^4. Each member is assigned a fixed, globally uniform value of P^*. In this study, the real-time HRSFS hindcast is running from 25 April to 27 May 2021.

2.3. Evaluation Approach

2.3.1. The Evaluation Data Set

To circumvent cloud obscuration and shadow artifacts inherent in optical satellite sensors (e.g., MODIS, VIIRS), we evaluate Linear Kinematic Feature (LKF) forecasts in the HRSFS using sea ice deformation derived from a mosaic Level-4 sea ice drift product (Global Ocean - High Resolution SAR Sea Ice Drift Time Series, 2025). This product, generated by the Technical University of Denmark (DTU) Space Department from SAR satellite data, provides all-weather and high-quality data unaffected by clouds or darkness. The DTU Space produces gridded ice displacement fields with a nominal temporal span between processed swaths of 24 hr and a nominal grid resolution of 10 km in the Arctic. This data set is obtained by applying the Matthews Correlation Coefficient (Matthews, 1975) processing on Sentinel-1 SAR, Envisat's Advanced SAR swath data or Radar Satellite Scan-SAR Wide mode data. Sea ice deformation rate is calculated as $\sqrt{\dot{\varepsilon}_I^2 + \dot{\varepsilon}_{II}^2}$, where $\dot{\varepsilon}_I$ and $\dot{\varepsilon}_{II}$ are the invariants of the strain rate as aforementioned. The two invariants are in fact the divergence and shear of the sea ice velocity field, respectively. The DTU SAR data directly provide the divergence and shear calculated from the retrieved sea ice displacements. Note that deformation fields derived from sea-ice motion data inherently carry significant uncertainty, with divergence estimates being particularly susceptible (Itkin, 2025; Plante et al., 2025).

2.3.2. The LKF Detection Method

We use the method described in Hutter et al. (2019) to detect the LKFs based on the deformation rates. This algorithm classifies pixels with higher deformation rates compared to the surroundings as LKF pixels and connects smaller LKF segments based on their distance, orientation and deformation. The same method is applied to both SAR observations and model simulations. The resulting LKFs are used for comparison.

2.3.3. Metrics

Due to the fractal nature of LKFs, it is normally difficult to directly compare the forecasted LKF and the SAR-observed LKF, that is, we don't know which LKF in observations should be used to compare against the modeled one. Therefore, for each SAR-observed LKF (denoted S), we identify the nearest forecasted LKF (F) using a bidirectional distance $D(F,S)$ defined as: $D(F,S) = (H_{F\mapsto S} + H_{S\mapsto F})/2$, where $H_{F\mapsto S}$, a variant of the Hausdorff distance, measures the minimum distance from F pointing to S. They are hereby calculated as

$$H_{F\mapsto S} = \frac{1}{N}\sum_{i=1}^{N} \inf_{f\in F}^{i} \inf_{s\in S} d(f,s), \tag{2}$$

$$H_{S\mapsto F} = \frac{1}{N}\sum_{i=1}^{N} \inf_{s\in S}^{i} \inf_{f\in F} d(f,s), \tag{3}$$

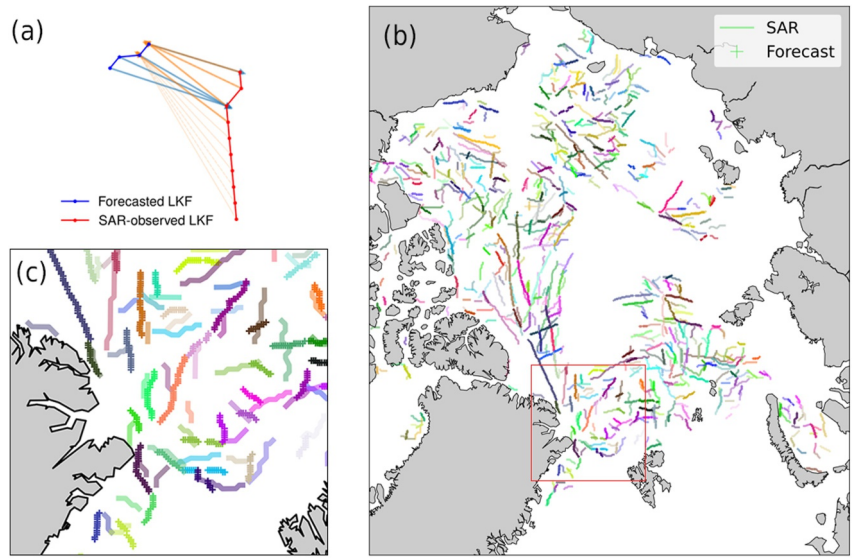

Figure 3. (a) The bi-directionally averaged minimum Hausdorff distance. The arrows between the linear kinematic features (LKFs) connect points to the closest point on the other LKF. In this plot, the bolder arrows indicate the distances involved in the bidirectionally averaged minimum Hausdorff distance (bAMHD) calculation, where $H_{F \mapsto S}$ is calculated by the blue arrows pointing to the red LKF, and vice versa for $H_{S \mapsto F}$. (b) LKFs from Synthetic Aperture Radar observations (bold-light) and their nearest forecasted LKFs used for evaluation (thin-deep) in terms of bAMHD distance on 2 May 2021. Note that the same color is used for each LKF pair. (c) The zoomed-in view of the red box in panel (b).

where f and s are the points along the LKFs F and S, $\inf^i$ the ith smallest value in the set, $d(f, s)$ the Haversine distance on the Earth sphere between locations f and s, N the minimum dimension of F or S. We designate D as the bidirectionally averaged minimum Hausdorff distance (bAMHD), conceptually illustrated by the averaged bold distances shown in Figure 3a.

Exact one-to-one correspondence between simulated and SAR-observed LKFs is not possible due to multifractal properties, temporal variability, orientation differences, and different resolution. The spatial resolution of SAR observations inherently also limits the precision of LKF detection. In this study, SAR-derived LKFs serve as the reference data set. For each observed LKF, we identify its nearest simulated counterpart based on the bAMHD distance, focusing exclusively on true-positive matches. To exclude effects from small-scale features with strong nonlinearities, only forecasted LKFs exceeding 50 km in length are retained since the nominal resolution of SAR data is 10 km, where we expect large scale LKF may be predictable. Figures 3b and 3c illustrates this comparative pairing approach.

The LKF forecast skill is also assessed using the spatial Maximum Cross-Correlation algorithm, following Korosov et al. (2023). This method compares modeled sea ice deformation fields against SAR-derived counterparts. To reconcile their differing spatial resolutions, the model outputs (~2 km) are upscaled to the SAR grid (~10 km) via a k-dimensional tree search with distance-weighted averaging within a 15-km radius. The metric A_{MCC} is computed by sliding a 30 × 30 pixel template within a 36 × 36 pixel search window, where the fraction of successful matches exceeding a correlation threshold of 0.35 defines the forecast skill. The resulting ensemble mean and spread of the $1 - A_{MCC}$ represent the forecast error.

LKF orientation and length serve as additional evaluation metrics. The LKF orientation is calculated on the projected coordinate using the "North Polar Stereographic Projection" in the Basemap package, where the central meridian is set to 0° and the bounding latitude is set to 65°N. The Arctic domain is divided into 50 km × 50 km grid cells. For each cell, the mean orientation of the contained LKF segments is computed. The LKF length is calculated as the cumulative geodesic distance between consecutive points along each LKF, determined via the Haversine formula on the WGS84 ellipsoid.

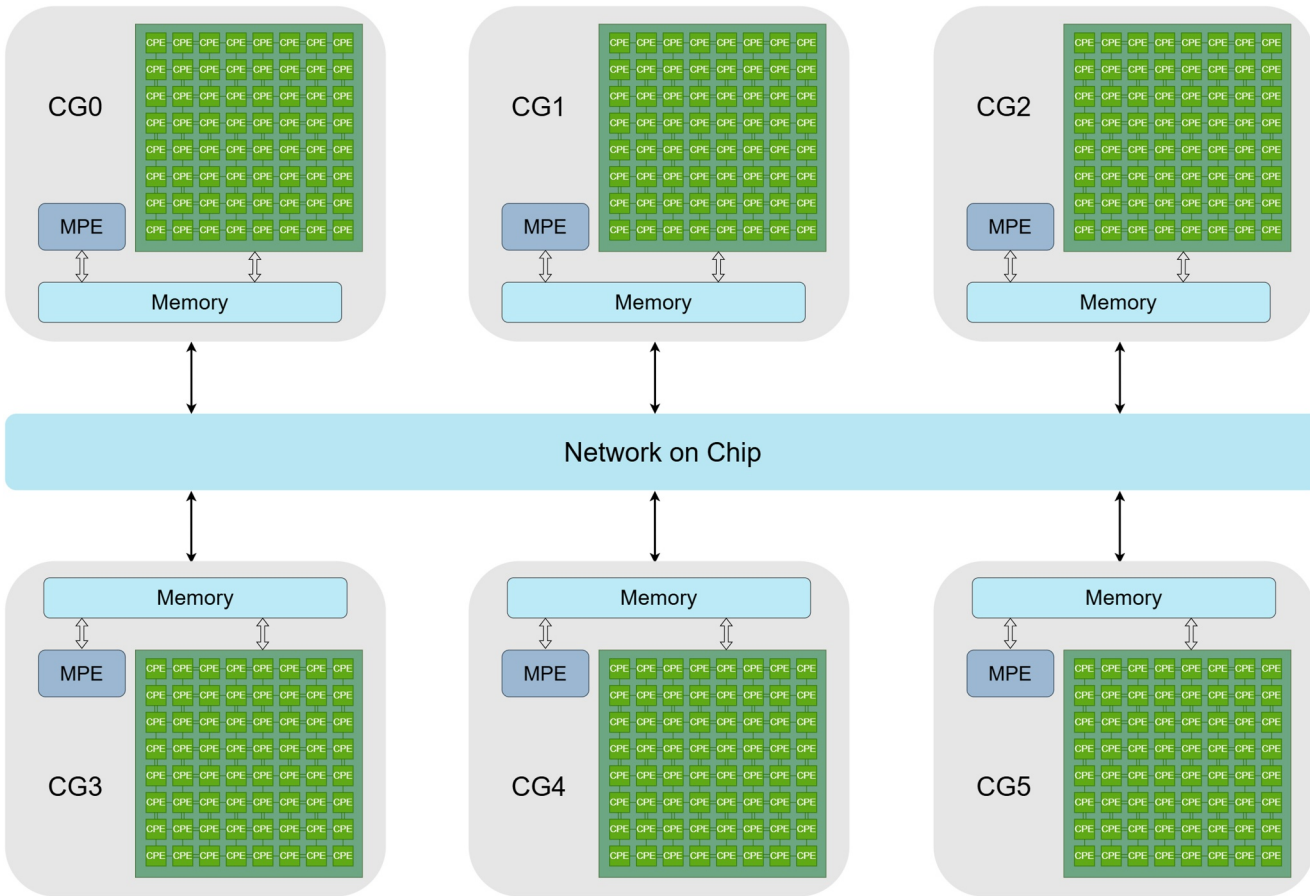

Figure 4. Sunway architecture.

3. HRSFS on a Heterogeneous Many-Core Computing System

3.1. System Architecture

The new generation Sunway supercomputer is equipped with the SW26010Pro many-core processor, featuring an innovative architectural design. As illustrated in Figure 4, each SW26010Pro processor contains six core groups (CGs), totaling 390 cores, which are interconnected via a network on chip. Each CG is composed of one management processing element (MPE), 64 computing processing elements (CPEs), and one memory controller. Both MPE and CPE adopt reduced instruction set computer architecture (Xu et al., 2022) but serve distinct purposes: the MPE is designed for computation scheduling, process management, task allocation, and data communication, while the CPEs are optimized for aggregating computing capacity and accelerating core functional modules, arranged in an 8×8 grid structure.

Each CG is equipped with 16 GB DDR4 main memory, accessible to both MPE and 64 CPEs via the memory controller. Additionally, each CPE features a 256 KB fast local device memory (LDM), part of which can be dynamically configured as a 32 KB or 128 KB hardware-managed data cache. The processor supports two high-speed memory access modes: direct memory access and remote memory access. Direct memory access enables continuous data transfer between LDM and main memory. RMA facilitates efficient communication among CPEs within the same CG.

The software environment of the Sunway platform has been continuously enhanced alongside architectural upgrades. The Sunway Accelerated Computing Architecture, introduced in 2021, integrates a basic compiler and core libraries to support development. Its parallel programming model employs "MPI + Athread": MPI manages inter-CG tasks and communication, while Athread optimizes intra-CG parallelism (Li et al., 2021). This programming model enables full utilization of the computational capabilities of the SW26010Pro processor.

3.2. Optimization Strategy

3.2.1. Filter Optimization

The LESTKF partitions the global state space into independent local subspaces, offering distinct advantages over global counterparts (Vetra-Carvalho et al., 2018). This filter is already coded in PDAF. This decomposition minimizes data contention and lock overhead by enabling concurrent, thread-safe processing of subspaces without global synchronization. By confining computations to localized regions, LESTKF improves cache locality, reducing memory latency and maximizing data retrieval efficiency. These architectural benefits align seamlessly with the Sunway multi-core architecture, providing a solid foundation for thread-level optimizations. Workload distribution across cores remains balanced through dynamic adjustment of subspace granularity, where the number of local domains assigned per CPE adapts to the total domain count and available CPE resources. Concurrently, cache-friendly access patterns maximize shared memory utilization. This synergy between LESTKF's design and hardware capabilities accelerates data assimilation workflows.

PDAF timing metrics show that the analysis phase dominates computational costs as it performs the core routine of combining model forecasts with observations. Employing Bayesian inference and ensemble-based filters, this phase computes an optimal posterior state estimate by quantifying uncertainties through ensemble statistics while efficiently handling high-dimensional spaces, adjusting prior states through observational constraints, and updating error covariance matrices. The resulting analysis state provides initial conditions for subsequent forecasts.

3.2.2. Parallelization Method

Following data partitioning through MPI, the analysis loop on Sunway triggers CPE processing while the MPE enters a waiting state. Each CPE cluster processes a subset of the partitioned data, and upon completion of the computations, the MPE resumes operation to handle the few dependent variables. This design ensures that the MPE focuses on management tasks, while CPEs handle compute-intensive analysis steps, thereby achieving a balanced distribution of workload.

The Linear Algebra PACKage (LAPACK) and the Basic Linear Algebra Subprograms that are needed during analysis are adapted to align with the interface requirements. These adaptations enable direct execution on CPEs by resolving architecture-specific constraints (e.g., instruction set compatibility and memory access patterns), thereby ensuring efficient processing of linear algebra operations with minimized computational overhead.

The implementation fully exploits LESTKF's localization properties and ensemble operation independence by minimizing the number of variables that exhibit dependencies within the analysis loop—such as those involved in global reductions of local observation statistics (e.g., summed observation counts and maximum local observation size). Computational tasks are predominantly offloaded to CPEs, while MPE involvement is restricted to resolving these few synchronization points, which drastically reduces cross-memory data transfers and enhances parallel efficiency. In this strategy, coarse-grained MPI parallelism orchestrates data partitioning, and fine-grained Athread computation optimizes CPE throughput, ultimately maximizing both task-level parallelism and overall computational performance.

3.2.3. Memory Optimization

In the LESTKF analysis, a large number of local variables are utilized exclusively during the computation and become obsolete upon its completion. To optimize this, a hierarchical memory management strategy was implemented: variables that are non-iterative in the analysis and not reused after processing are allocated directly in the LDM. Readers are referred to Vetra-Carvalho et al. (2018) for more details. This approach exploits the low-latency access characteristics of LDM to minimize redundant data transfers. Additionally, sequential computational steps leading to the final results were consolidated, with intermediate operations fused to reduce memory access overhead and improve data reuse efficiency.

For intermediate variables involved in computations outside the analysis loop, dedicated arrays were allocated in main memory. These variables are directly written into the main memory arrays by the CPEs using load/store instructions, thereby avoiding unnecessary data movement between LDM and main memory.

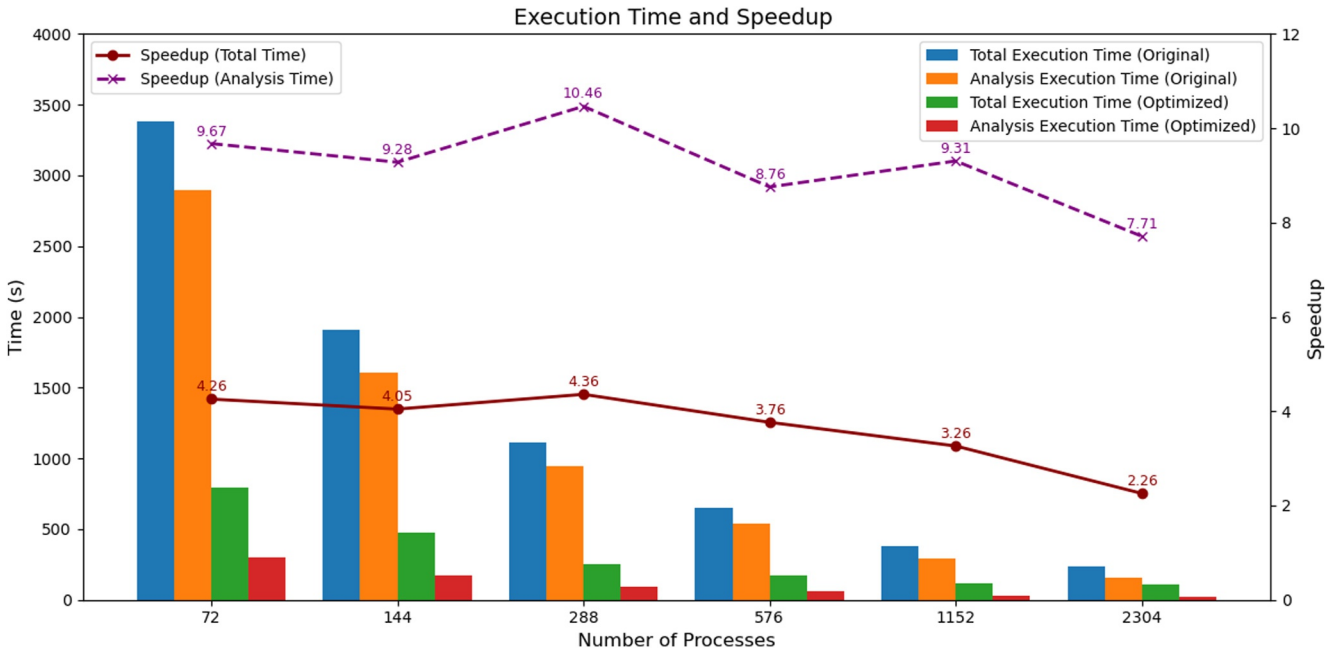

Figure 5. Execution time comparison.

4. Assimilation Results

4.1. Analysis Performance of HRSFS on the Sunway Architecture

The performance of the optimized PDAF on the Sunway architecture is evaluated through execution time analysis. The results focus on three key metrics: total execution time, analysis phase efficiency, and parallel scalability across varying process counts (72–2,304 processes), which is shown in Figure 5.

The total execution time of the analysis demonstrates substantial improvements after optimization. At 72 processes, the original implementation required 56 min for a full data assimilation cycle, while the optimized version completed the same task in just 13.3 min. As the number of processes scales to 2,304, the original runtime drops to 4 min, whereas the optimized framework achieves a reduction to 1.77 min. The speedup curve (red in Figure 5) shows that the total runtime achieves a 4.26× speedup at 72 processes and retains a 2.26× speedup at 2,304 processes, with the largest speedup found using 288 processes.

The analysis hotspot, a critical time-consuming computational bottleneck in the original implementation, shows even more performance gains. A speedup of 9.68× is observed with 72 processes, and 7.71× with 2,304 processes, indicating that the optimization preserves high efficiency even at large parallel scales. The optimized version reduces analysis duration from 48.3 to 5 min at 72 processes and from 2.57 to 0.33 min at 2,304 processes. It is noted that this statistic is only for the actual analysis step computation without including IO.

For the total run time of the offline assimilation process, scaling from 72 to 2,304 processes, which represents a 32-fold increase in computational resources, the original implementation reduces analysis time by 93%, while the optimized version achieves a comparable 87% reduction. This indicates that the optimized version maintains good parallel scalability. Throughout the scaling curve, the total speedup consistently outperforms the original version by approximately 2 ×, confirming that our architectural refinements specifically target the most computationally intensive components.

The optimization of the data assimilation system paves the way for the high resolution forecast system. The 72 hr forecasts with 8 ensemble members now can be finished in 1.5 hr using 1,617 cores. Analysis within 2 min potentially also enables high frequency data assimilation, which has been proved to dramatically improve the forecast results (K. Wang et al., 2023).

Table 1
The Configuration of SIC Assimilation Experiments

Experiment	Filter	Observation error	Ensemble size
GS	LESTKF	0.15 uniformly	8
GSLsigma	LESTKF	$\sigma(C) = \psi(C)\Upsilon(C, L)$	8
NL	LNETF	0.15 uniformly	8

4.2. Assimilation of High Resolution SIC

The operational SIC observations currently in use are derived from AMSR2 passive microwave radiometer data at 3.25 km resolution. However, assimilating such high-resolution data within an ensemble framework presents challenges, as open water occurrence represents an intermittent phenomenon with skewed distributions that conventional filters struggle to handle. We demonstrate this through three assimilation experiments (Table 1). The experiment *GS* is configured with the LESTKF and a globally uniform SIC observation error of 0.15, inherently assuming Gaussian distributions. Experiment *GSLsigma* uses the same LESTKF filter but with SIC error estimated by the inversed Beta distribution function together with the nonlinearity function aforementioned in Section 2.1.3, while still maintaining Gaussian distribution assumptions. Experiment *NL* employs the LNETF with the same SIC error configuration as *GS*.

We first evaluate the analyzed SIC at the basin scale. Compared to the SIC observations (Figure 6d), the analyzed SIC distributions from all experiments generally capture the observed patterns, with extensive areas of high concentration (above 0.8) around the North Pole and north of the Severnaya Zemlya Islands. However, in these regions, *GS* produces SIC values that are higher than observed, albeit smoothly distributed. Conversely, *NL* yields lower SIC but exhibits discontinuities during transitions from high to low concentrations, characterized by blotchy patterns. *GSLsigma* generates the most realistic analysis, with smooth transitions and a more accurate sea ice edge. Both *GS* and *NL* fail to adequately constrain low concentrations at the sea ice edge, where excess sea ice is present north of the Barents Sea. These results suggest that the Gaussian assumption tends to produce overly smooth fields, while the nonlinear filter creates more abrupt, patchy structures.

The performance in capturing LKFs, specifically sea ice leads here, varies considerably among the three experiments. For instance, the prominent lead in the East Siberian Sea (Figures 6d and 6e) is absent in *GS* but successfully captured by both *GSLsigma* (Figure 6b) and *NL* (Figure 6c). While NL's ability to resolve such

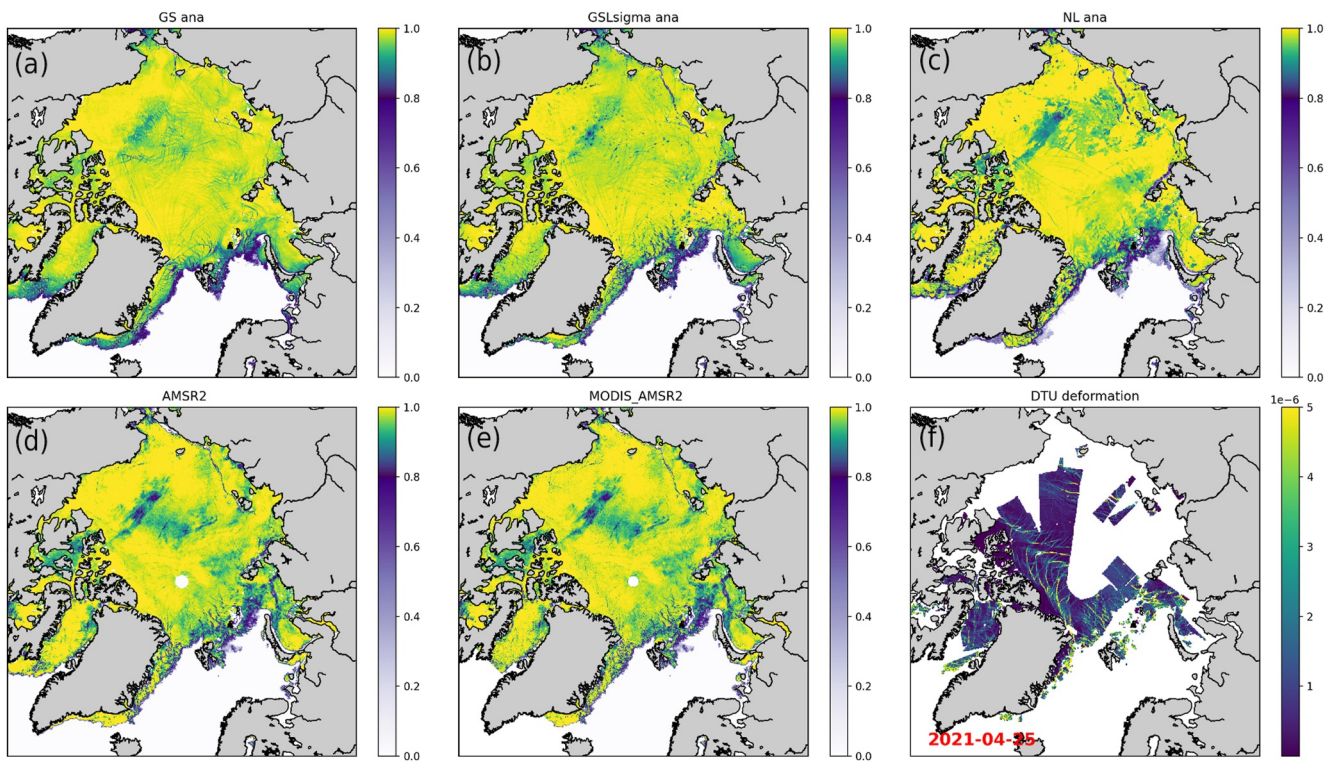

Figure 6. Analyzed SIC on 25 April 2021 using a different assimilation method. Sea ice analysis from *GS* is denoted as GS ana (a), from *GSLsigma* as GSLsigma ana (b), and from *NL* as NL ana (c). SIC with a resolution of 3.25 km retrieved from Advanced Microwave Scanning Radiometer 2, and of 1 km from the MODIS-AMSR2 merged product is shown in (d) and (e). Sea ice deformation (s^{-1}) denoting potential linear kinematic features from DTU Synthetic Aperture Radar data is shown in (f).

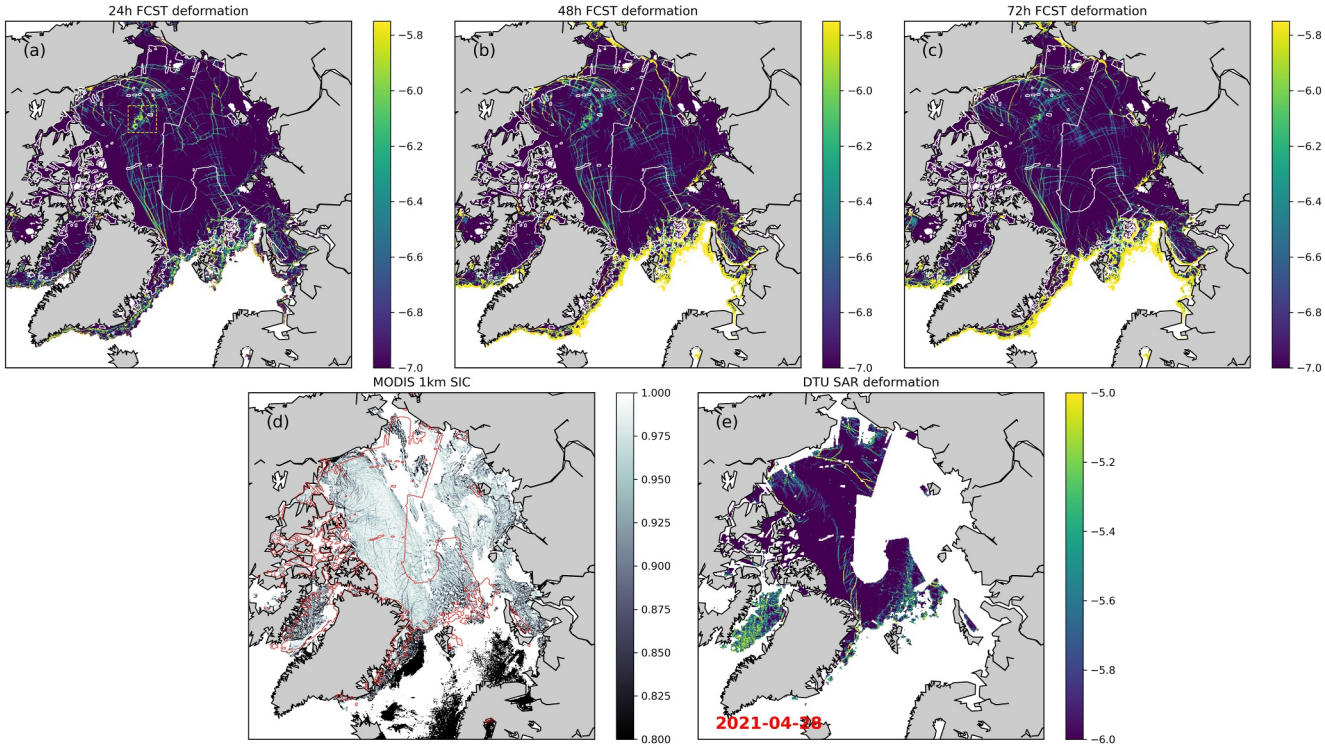

Figure 7. Linear kinematic feature (LKF) forecasts represented by sea ice deformation rate (log10 scale, s^{-1}) on 28 April 2021. Forecasts on lead times of 24, 48, and 72 hr are shown in (a), (b), and (c), respectively. SIC observation from MODIS is shown in (d). Sea ice deformation from DTU Synthetic Aperture Radar (SAR) data is shown in (e). White (a,b,c) and red (d) boundaries mark SAR satellite data coverage areas. The forecasts use different colormap ranges than the SAR data, as the ensemble mean reduces the magnitude of deformation signals from shifting LKFs.

nonlinear features is expected, only GSLsigma realistically represents smaller leads north of Svalbard. Furthermore, GSLsigma preserves simulated LKFs in consolidated ice regions ($SIC \rightarrow 1.0$), which is crucial for the forecast system as the SIC observation cannot resolve enough LKFs, whereas NL retains only large-scale features. North of Greenland and the Canadian Arctic Archipelago (CAA)—regions with dense LKFs evident in both MODIS-AMSR2 (Figure 6e) and SAR data (Figure 6f)—the oversmoothing in *GS* and scaling limitations in *NL* result in poor performance. These results demonstrate that *GSLsigma* reconciles large-scale extremes with small-scale features, combining advantages of both nonlinear and Gaussian-filter approaches, which is adopted in the hindcast. Despite data assimilation, a notable discrepancy persists in the spatial patterns of leads (represented by sea ice concentration) and LKFs (represented by deformation rate) between the analysis and observations (Figure 6). This indicates that developing an accurate LKF forecasting system is challenging when the initial conditions, derived from the assimilated analysis, are already significantly biased.

5. Performance of LKF Forecast

5.1. Spatial Distribution of the LKF Forecast

LKF forecasts across different lead times (i.e., forecast horizons) (Figure 7) exhibit consistent spatial distributions basin-wide. Significant LKF density occurs along the Beaufort Gyre periphery, with two major bands extending: one toward Fram Strait north of the Canadian Arctic Archipelago (hereafter FS band), and another across the Transpolar Drift toward the Barents Sea (hereafter BS band). Both MODIS SIC (Figure 7d) and SAR (Figure 7e) observations confirm the entire FS band, and also reveal parts of the BS band. Although SAR data gaps exist north of the Laptev and East Siberian Seas (Figure 7d), available MODIS SIC data still show LKFs whose orientations align with forecast patterns. Forecasts north of the Barents Sea capture numerous small-scale LKFs resulting from ice-island interactions, consistent with both MODIS and SAR observations.

Narrow, short LKFs along the FS band north of CAA appear less pronounced in SAR data than in MODIS. While forecasts generate plausible LKFs in this region, their excessive deformation suggests longer features than SAR

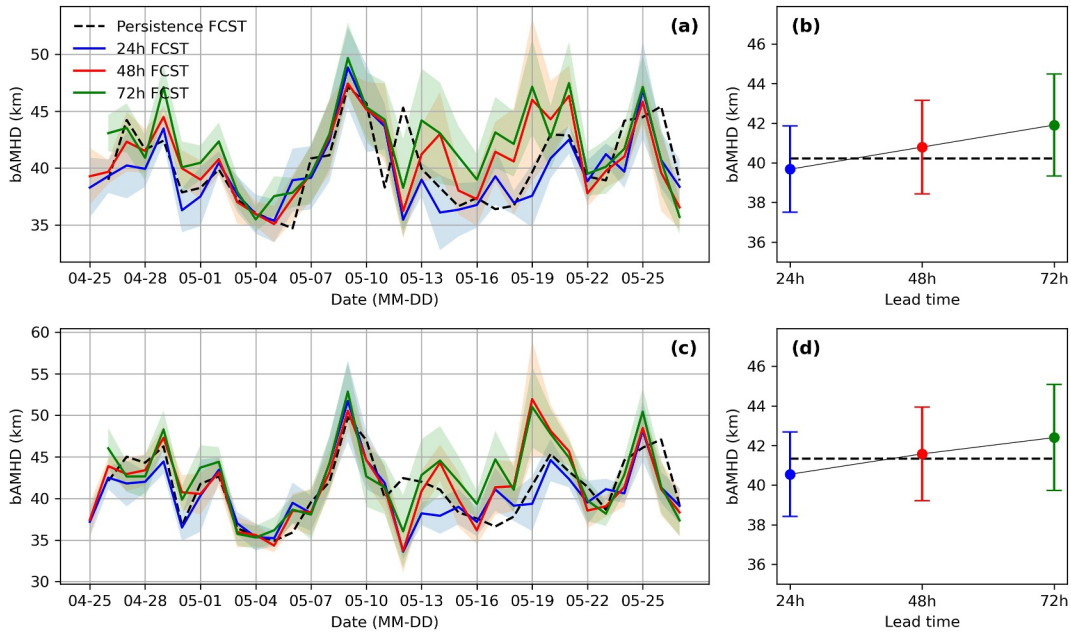

Figure 8. The bidirectionally averaged minimum Hausdorff distance (bAMHD) distance between forecasted and SAR-observed linear kinematic features. Smaller bAMHD indicates better skill. (a) Shows the bAMHD calculated against the Synthetic Aperture Radar (SAR) observation on 0h and (b) shows the distance with respect to the lead time. Correspondingly, (c) and (d) show the distance calculated against the SAR observation on 12 hr. The shading area and error bars denote the ensemble spread. Note that currently the daily mean forecast is used for the calculation. Persistence forecasts are shown in black dashed lines.

observed. For the BS band covered by observations in the Beaufort Sea, forecasts and SAR agree on the LKF orientation along the Transpolar Drift direction. However, SAR reveals a continuous, dominant LKF with high-intensity deformation (Figure 7e), while forecasts produce dispersed, fragmented LKFs lacking this coherence. A spurious forecast feature appears in the Beaufort Gyre between the two major bands (yellow dashed region in Figure 7a), but this region shows neither high deformation in SAR nor distinct leads in MODIS SIC. While 3-day forecasts differ primarily in LKF development and dissipation, the consistency of the large-scale LKF pattern indicates that skillful LKF forecasting is feasible if initial conditions are perfectly constrained.

5.2. LKF Predictability

LKF forecast skill over time is evaluated using bAMHD distance. Note that bAMHD only measures large scale LKFs. Distances are computed relative to observed deformation products derived from 24-hr drift trajectories. To be robust, two overlapping windows are used daily: 00:00 to 00:00 + 24 hr UTC and 12:00 to 12:00 + 24 hr UTC. As shown in Figures 8a and 8c, 24 hr forecasts generally outperform 48 and 72 hr forecasts. Exceptions occur between May 3 and 10, when the SAR data coverage is limited to north of Fram Strait. Similar performance is found when validating against 12 hr SAR data (Figure 8c). Notably, 48 hr forecasts show significantly better performance relative to 72 hr forecasts when compared against 0h SAR data versus 12 hr data.

Figures 8b and 8d present forecast skill as a function of lead time. Against 0 hr data, bAMHD values are 39.7, 40.8, and 41.9 km for 24, 48, and 72 hr lead times, respectively. The forecast outperforms persistence (the analyzed state) at a 24 hr lead time, but not at 48 hr. Ensemble spreads vary minimally across lead times, ranging from 2.18 to 2.58 km. When evaluated against 12 hr data, bAMHD values are approximately 1 km larger for all lead times, while ensemble spreads remain comparable. These statistics further confirm the findings in Section 5.1.

To assess the spatial performance of forecasted LKFs, the $1 - A_{MCC}$ error metric was evaluated against SAR-derived deformation over the study period. Time-series analysis (Figures 9a and 9c) indicates that while all forecast lead times generally track the observed deformation events, the 24 hr forecast consistently yields the highest spatial fidelity, often outperforming the persistence baseline. A systematic increase in error with lead time

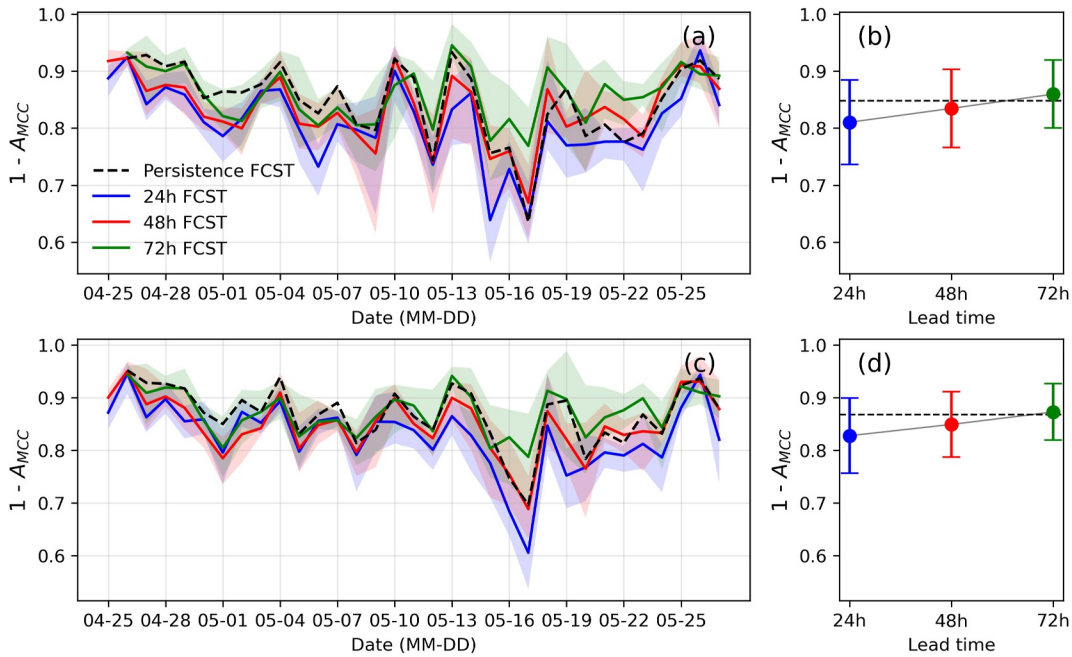

Figure 9. The forecast error $1 - A_{MCC}$ against Synthetic Aperture Radar observations. Smaller $1 - A_{MCC}$ indicates higher forecast skill. (a) Shows the forecast error calculated against the observation on 0h and (b) shows the error with respect to the lead time. Correspondingly, (c) and (d) show the error calculated against the observation on 12 hr. The shading area and error bars denote the ensemble spread. Note that currently the daily mean forecast is used for the calculation. Persistence forecasts are shown in black dashed lines.

is evident in the statistical summaries (Figures 9b and 9d), where the ensemble mean error rises from approximately 0.82 at 24 hr to 0.87 at 72 hr. In contrast to the bAMHD metric, the ensemble spreads of $1 - A_{MCC}$ are narrowing with lead time. Comparison between the 0 and 12 hr observation windows shows high consistency in error patterns, suggesting that the model's spatial skill is robust across different intra-day satellite overpasses. In terms of spatial correlation, the results demonstrate that HRSFS shows significant forecast skill over persistence for lead times up to 48 hr.

Across all lead times, the PDFs for the ensemble mean LKF orientation and length are statistically indistinguishable (Figures 10a and 10b). Forecasted LKF orientations (Figure 10a) show PDF peaks at $\pm[49°, 44°]$ and $[-2.6°, 2.6°]$, which corresponds to the peaks in SAR observations. The relatively low occurrence of LKF orientation in the range of $\pm[20°, 40°]$ and $\pm[75°, 80°]$ is also found both in forecasts and observations. All forecasts successfully capture the observed large orientations north of CAA extending toward the Fram Strait, as well as those in the East Siberian Sea (Figure 10c). The forecasts also indicate negative orientations north of the Chukchi Sea and in the Beaufort Sea. While negative orientations north of Greenland are present in the forecasts, they are not as clearly evident in the observations. Due to resolution limitations, the SAR observations cannot resolve the larger LKFs along the East Greenland Current and north of the Svalbard Islands. For LKFs shorter than 100 km, the PDF spreads across different forecast lead times are consistent; however, the spread increases for longer LKFs (Figure 10b). Between 100 and 1,000 km, all forecasted lengths are underestimated compared to observations. This systematic underestimation may be mitigated by an ITD parameterization, which is not employed in the current model configuration (Hutter & Losch, 2020). This interpretation is further supported by PDFs from individual ensemble members (not shown) that use different internal stress (P^*); their differences are most pronounced within the aforementioned underestimated length range.

6. Discussions and Summary

This study developed a high-resolution sea ice forecast system at ~2 km scale and presented ensemble-based 72-hr forecasts of Arctic LKFs. We optimized the PDAF for Sunway's heterogeneous many-core architecture, reducing the analysis time by nearly an order of magnitude by employing the fine-grained thread-level

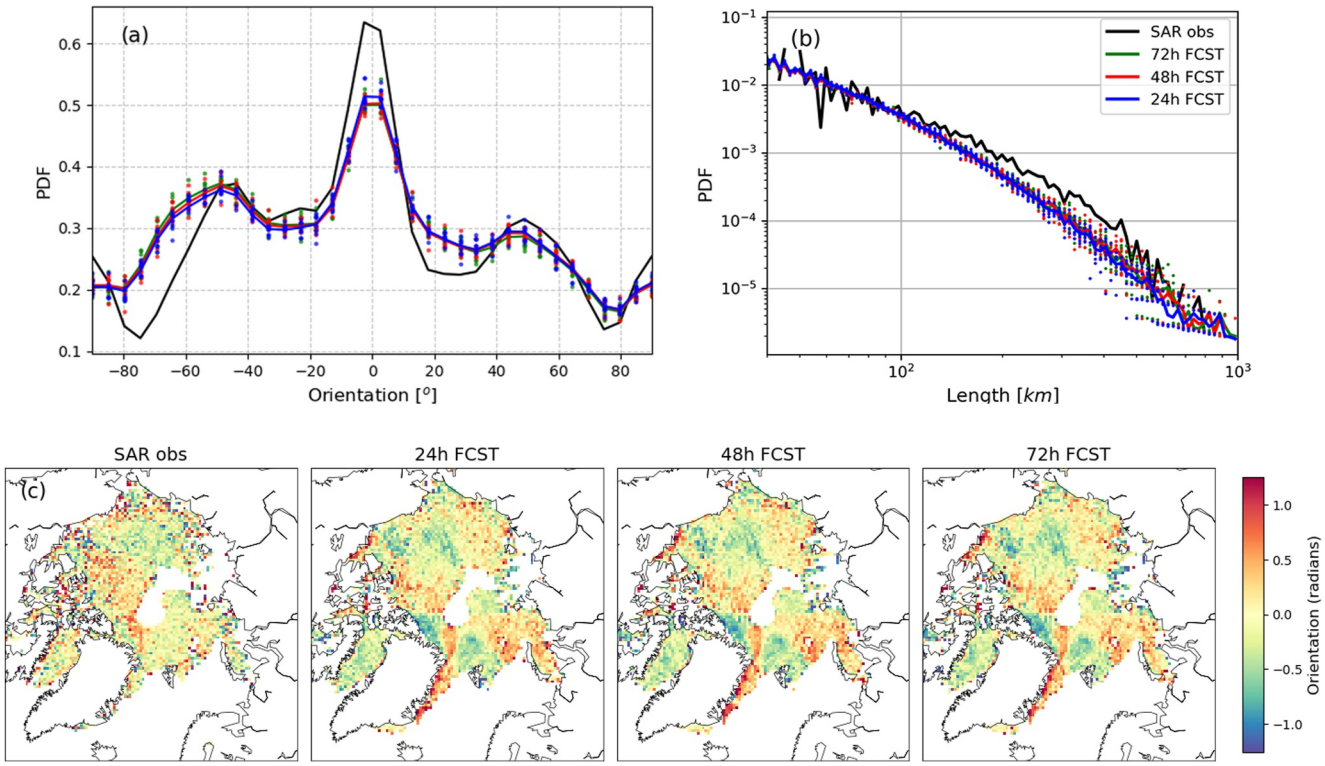

Figure 10. The probability density functions (PDFs) for the orientation (a) and length (b) of forecasted and SAR-observed linear kinematic features (LKFs). The PDFs were computed using Gaussian kernel density estimation (gaussian_kde from SciPy) to eliminate biases introduced by arbitrary histogram bin edges. Each dot in (a) and (b) represents the PDF from an individual ensemble member. The solid lines, colored correspondingly, show the ensemble mean PDF for each group. (c) Displays the spatial distribution of LKF orientations. Note that only forecasted LKFs within the SAR-observed region are included in this statistical analysis.

parallelism, which previously was the primary bottleneck. This advancement dramatically improved computational feasibility and demonstrates potential for future high-frequency analysis.

Forecasting nonlinear LKFs is difficult due to their high intermittency and anisotropy, a problem analogous to forecasting precipitation in atmospheric systems and eddies in oceanic systems. Therefore, we employed an ensemble approach by perturbing the ice strength parameter to account for uncertainties in the system. To assimilate SIC observations that contain signatures of LKFs, which often manifest as linear features of reduced sea ice concentration within the pack ice, we designed a localized observation error function. During analysis, this function preserves most of the simulated LKFs in the forecasts that would otherwise be filtered out when they are not resolved in observations, while still assimilating valuable LKF information from SIC observations. This process generates SIC fields that are smooth and consistent with observations at both basin and local scales.

The LKF hindcast experiment addresses two key questions: to what spatiotemporal scale LKFs can be forecast, and to what degree of accuracy. Using the bAMHD metric for large-scale LKFs, the experiment demonstrates forecast skill with bAMHD values of 39.7–41.9 km across 24- to 72-hr lead times. Another metric that measures the maximum cross-coefficients over space shows that the forecast skill can outperform the persistence with lead time more than 48 hr. For small-scale LKFs (<100 km), however, results reveal a higher spatial occurrence in forecasts than in observations. While these small-scale features cannot be perfectly initialized, the performance can be improved through proper sea ice strength parameter perturbations. Additional findings show that forecasted LKF distributions statistically match SAR observations in orientation and length, with minimal differences across lead times. Critically, this demonstrates the achievable skill level for 3-day LKF forecasts—even with assimilation limited to SIC observations. The model's physical representation of sea ice dynamics, derived from the SIC state rather than resolution alone, emerges as the primary determinant on predictability. Specifically, continuous data assimilation maintains the realistic heterogeneity of sea ice state, which defines the internal ice strength distribution. This structural foundation enables the model's rheology to effectively translate atmospheric forcing into dynamic deformation features, allowing the forecast to outperform persistence.

It is also worth noting that the predicted LKFs include not only leads but also ridges. The data assimilation step in this study utilizes SIC data that explicitly resolve a few leads. While ridges are difficult to observe directly, incorporating ridge information would likely improve forecasts of navigational barriers, which represents an important direction for future research. Despite the relatively coarse resolution of current observations, the dynamic signatures of leads and ridges are inherently manifested in SIC or deformation fields, allowing this information to be utilized indirectly. This is consistent with the methodology of Korosov et al. (2023), in which deformation information is implicitly used through SIC, which also shows a positive effect on LKF forecasts.

Advancing LKF forecasting requires improvements both in data assimilation and modeling frameworks. Several foundational research directions warrant further investigation. Data assimilation methodologies demand refinement, specifically optimizing ensemble design through different perturbation techniques and various atmospheric forcing. Key priorities include tuning localization radii, addressing representative errors, and integrating multisource observational data sets (e.g., sea ice concentration, thickness, drift, and deformation). For instance, LKFs could be directly extracted from high-resolution MODIS SIC data and then assimilated as localized observational features. Their detailed structure would help correct the model's representation of ice deformation, thereby enhancing its predictive skill in regions where MODIS provides coverage. Quantitative assessment of forecast skill, particularly regarding LKF persistence and spatial density, is also essential to evaluate the forecast performance.

Operational implementation must balance physical accuracy with computational efficiency. Critical aspects like the convergence criteria and rheology related parameters, such as yield curve eccentricity and lead closing parameters, require systematic investigation on their forecast effects. Additionally, given the chaotic nature of LKFs, the predictability is fundamentally scale-dependent. While forecasting the precise location and timing of individual LKF remains inherently challenging, a practical solution is to provide probabilistic forecasts. These forecasts map the likelihood of LKF occurrence within specific domains across an ensemble, thereby directly quantifying the uncertainty and translating it into actionable risk information. This could be further explored to develop a high-resolution, decision-relevant sea ice forecasting.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

MITgcm is open-source software available at https://github.com/MITgcm/MITgcm. The Parallel Data Assimilation Framework (PDAF) is accessible via http://pdaf.awi.de. The linear kinematic feature (LKF) detection code is publicly available at https://github.com/nhutter/lkf_tools. Modified PDAF core routines optimized for many-core architectures, along with supporting data are available from Mu (2025).

Acknowledgments
Computing resources are financially supported by Laoshan Laboratory (No. LSKJ202300303, No. LSKJ202202301). The research is supported by the Taishan Scholars Program (TSPD20240807) and the National Natural Science Foundation of China (42176235). LM developed the assimilation and forecast system, refactored the assimilation code, conducted the experiments and drafted the paper. YL, YC, and HW optimized the assimilation code on Sunway computing platform. XC contributed to the discussion of the work and gave financial support on the project. RS, LZ, JC helped on data analysis. LN is the main developer of PDAF and edited the manuscript. LW gave supports on the funding and computation resources. All authors contributed to the system development. The authors would like to thank Damien Ringeisen and two other anonymous reviewers for their constructive comments on the paper.

References

Andersen, S., Tonboe, R., Kaleschke, L., Heygster, G., & Pedersen, L. T. (2007). Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration arctic sea ice. *Journal of Geophysical Research*, *112*(C8). https://doi.org/10.1029/2006jc003543

Antonov, J. I., Locarnini, R., Boyer, T., Mishonov, A., & Garcia, H. (2006). *World ocean atlas 2005, volume 1: Salinity*. Government Printing Office.

Blockley, E., & Peterson, K. (2018). Improving Met Office seasonal predictions of Arctic sea ice using assimilation of CryoSat-2 thickness. *The Cryosphere*, *12*, 3419–3438, https://doi.org/10.5194/tc-12-3419-2018

Blockley, E., Vancoppenolle, M., Hunke, E., Bitz, C., Feltham, D., Lemieux, J.-F., et al. (2020). The future of sea ice modeling: Where do we go from here? *Bulletin of the American Meteorological Society*, *101*(8), E1304–E1311. https://doi.org/10.1175/bams-d-20-0073.1

Bouchat, A., Hutter, N., Chanut, J., Dupont, F., Dukhovskoy, D., Garric, G., et al. (2022). Sea ice rheology experiment (SIREx): 1. scaling and statistical properties of sea-ice deformation fields. *Journal of Geophysical Research: Oceans*, *127*, e2021JC017667. https://doi.org/10.1029/2021JC017667

Dansereau, V., Weiss, J., Saramito, P., & Lattes, P. (2016). A Maxwell elasto-brittle rheology for sea ice modelling. *The Cryosphere*, *10*(3), 1339–1359. https://doi.org/10.5194/tc-10-1339-2016

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Girard, L., Bouillon, S., Weiss, J., Amitrano, D., Fichefet, T., & Legat, V. (2011). A new modeling framework for sea-ice mechanics based on elasto-brittle rheology. *Annals of Glaciology*, *52*(57), 123–132. https://doi.org/10.3189/172756411795931499

Global Ocean - High Resolution SAR Sea Ice Drift Time Series. (2025). E.U. copernicus marine service information (CMEMS). *Marine Data Store (MDS)*. https://doi.org/10.48670/mds-00328

Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., et al. (2020). The current configuration of the ostia system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sensing*, *12*(4), 720. https://doi.org/10.3390/rs12040720

Hibler, W., III. (1979). A dynamic thermodynamic sea ice model. *Journal of Physical Oceanography*, *9*(4), 815–846. https://doi.org/10.1175/1520-0485(1979)009<0815:adtsim>2.0.co;2

Hibler, W., III. (1984). The role of sea ice dynamics in modeling co2 increases. *Climate processes and climate sensitivity*, *29*, 238–253.

Hutter, N., Bouchat, A., Dupont, F., Dukhovskoy, D., Koldunov, N., Lee, Y. J., et al. (2022). Sea ice rheology experiment (sirex): 2. Evaluating linear kinematic features in high-resolution sea ice simulations. *Journal of Geophysical Research: Oceans*, *127*(4), e2021JC017666. https://doi.org/10.1029/2021jc017666

Hutter, N., & Losch, M. (2020). Feature-based comparison of sea ice deformation in lead-permitting sea ice simulations. *The Cryosphere*, *14*(1), 93–113. https://doi.org/10.5194/tc-14-93-2020

Hutter, N., Zampieri, L., & Losch, M. (2019). Leads and ridges in arctic sea ice from rgps data and a new tracking algorithm. *The Cryosphere*, *13*(2), 627–645. https://doi.org/10.5194/tc-13-627-2019

Ip, C. F., Hibler, W. D., III, & Flato, G. M. (1991). On the effect of rheology on seasonal sea-ice simulations. *Annals of Glaciology*, *15*, 17–25. https://doi.org/10.3189/1991aog15-1-17-25

Itkin, P. (2025). Novel methods to study sea ice deformation, linear kinematic features and coherent dynamic clusters from imaging remote sensing data. *The Cryosphere*, *19*(3), 1135–1151. https://doi.org/10.5194/tc-19-1135-2025

Janjić, T., Bormann, N., Bocquet, M., Carton, J., Cohn, S., Dance, S. L., et al. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278. https://doi.org/10.1002/qj.3130

Kaleschke, L., Lüpkes, C., Vihma, T., Haarpaintner, J., Bochert, A., Hartmann, J., & Heygster, G. (2001). SSM/I sea ice remote sensing for mesoscale ocean-atmosphere interaction analysis. *Canadian Journal of Remote Sensing*, *27*(5), 526–537. https://doi.org/10.1080/07038992.2001.10854892

Korosov, A., Rampal, P., Ying, Y., Ólason, E., & Williams, T. (2023). Towards improving short-term sea ice predictability using deformation observations. *The Cryosphere*, *17*(10), 4223–4240. https://doi.org/10.5194/tc-17-4223-2023

Landrum, L., & Holland, M. M. (2020). Extremes become routine in an emerging new arctic. *Nature Climate Change*, *10*(12), 1108–1156. https://doi.org/10.1038/s41558-020-0892-z

Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, *32*(4), 363–403. https://doi.org/10.1029/94rg01872

Li, Y., Duan, X., Gan, L., Wan, W., & Yang, G. (2021). Enabling large-scale simulation of cam on the sunway taihulight supercomputer. *IEEE Transactions on Computers*(99), 1-1.

Liang, X., Zhao, F., Li, C., Zhang, L., & Li, B. (2020). Evaluation of arciops sea ice forecasting products during the ninth chinare-arctic in summer 2018. *Adv. Polar Sci*, *31*(1), 14–25.

Liu, J., Chen, Z., Hu, Y., Zhang, Y., Ding, Y., Cheng, X., et al. (2019). Towards reliable arctic sea ice prediction using multivariate data assimilation. *Science Bulletin*, *64*(1), 63–72. https://doi.org/10.1016/j.scib.2018.11.018

Locarnini, R., Mishonov, A., Ji, A., Boyer, T., & Garcia, H. (2006). *World ocean atlas 2005, volume 1: Temperature*. U.S. Government Printing Office.

Losch, M., Menemenlis, D., Campin, J.-M., Heimbach, P., & Hill, C. (2010). On the formulation of sea-ice models. Part 1: Effects of different solver implementations and parameterizations. *Ocean Modelling*, *33*(1–2), 129–144. https://doi.org/10.1016/j.ocemod.2009.12.008

Ludwig, V., Spreen, G., & Pedersen, L. T. (2020). Evaluation of a new merged sea-ice concentration dataset at 1 km resolution from thermal infrared and passive microwave satellite data in the arctic. *Remote Sensing*, *12*(19), 3183. https://doi.org/10.3390/rs12193183

Marshall, J., Adcroft, A., Hill, C., Perelman, L., & Heisey, C. (1997). A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research*, *102*(C3), 5753–5766. https://doi.org/10.1029/96JC02775

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta*, *405*(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Melsheimer, C., & Spreen, G. (2019). AMSR2 ASI sea ice concentration data, Arctic, version 5.4 (NetCDF) (July 2012 - December 2019) [Dataset]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.898399

Menemenlis, D., Campin, J.-M., Heimbach, P., Hill, C., Lee, T., Nguyen, A., et al. (2008). Ecco2: High resolution global ocean and sea ice data synthesis. *Mercator Ocean Quarterly Newsletter*, *31*, 13–21.

Mu, L. (2025). Forecast results from an Arctic high resolution sea ice forecast system [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.17131688

Mu, L., Liang, X., Yang, Q., Liu, J., & Zheng, F. (2019). Arctic ice ocean prediction system: Evaluating sea-ice forecasts during xuelong's first trans-arctic passage in summer 2017. *Journal of Glaciology*, *65*(253), 813–821. https://doi.org/10.1017/jog.2019.55

Mu, L., Nerger, L., Streffing, J., Tang, Q., Niraula, B., Zampieri, L., et al. (2022). Sea-ice forecasts with an upgraded awi coupled prediction system. *Journal of Advances in Modeling Earth Systems*, *14*(12), e2022MS003176. https://doi.org/10.1029/2022ms003176

Mu, L., Yang, Q., Losch, M., Losa, S. N., Ricker, R., Nerger, L., & Liang, X. (2018). Improving sea ice thickness estimates by assimilating cryosat-2 and smos sea ice thickness data simultaneously. *Quarterly Journal of the Royal Meteorological Society*, *144*(711), 529–538. https://doi.org/10.1002/qj.3225

Nerger, L., & Hiller, W. (2013). Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers & Geosciences*, *55*, 110–118. https://doi.org/10.1016/j.cageo.2012.03.026

Nerger, L., Janjić, T., Schröter, J., & Hiller, W. (2012). A unification of ensemble square root kalman filters. *Monthly Weather Review*, *140*(7), 2335–2345. https://doi.org/10.1175/mwr-d-11-00102.1

Nguyen, A. T., Menemenlis, D., & Kwok, R. (2011). Arctic ice-ocean simulation with optimized model parameters: Approach and assessment. *Journal of Geophysical Research*, *116*(C4), C04025. https://doi.org/10.1029/2010JC006573

Olason, E., Boutin, G., Korosov, A., Rampal, P., Williams, T., Kimmritz, M., et al. (2022). A new brittle rheology and numerical framework for large-scale sea-ice models. *Journal of Advances in Modeling Earth Systems*, *14*(8), e2021MS002685. https://doi.org/10.1029/2021ms002685

Parkinson, C. L., & Washington, W. M. (1979). A large-scale numerical model of sea ice. *Journal of Geophysical Research*, *84*(C1), 311–337. https://doi.org/10.1029/jc084ic01p00311

Plante, M., Lemieux, J.-F., Tremblay, L. B., Bouchat, A., Ringeisen, D., Blain, P., et al. (2025). A sea ice deformation and rotation rate dataset (2017–2023) from the environment and climate change Canada automated sea ice tracking system (eccc-asits). *Earth System Science Data*, *17*(2), 423–434. https://doi.org/10.5194/essd-17-423-2025

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., & Haas, C. (2017). A weekly Arctic sea-ice thickness data record from merged CryoSat-2 and SMOS satellite data. *The Cryosphere*, *11*(4), 1607–1623. https://doi.org/10.5194/tc-11-1607-2017

Schaffer, J., & Timmermann, R. (2016). Greenland and antarctic ice sheet topography, cavity geometry, and global bathymetry (rtopo-2), links to netcdf files. *PANGAEA*, 543–557. https://doi.org/10.1594/PANGAEA.856844

Semtner, A. J., Jr. (1976). A model for the thermodynamic growth of sea ice in numerical investigations of climate. *Journal of Physical Oceanography*, *6*(3), 379–389. https://doi.org/10.1175/1520-0485(1976)006<0379:amfttg>2.0.co;2

Song, R., Mu, L., Loza, S. N., Kauker, F., & Chen, X. (2024). Assimilating summer sea-ice thickness observations improves arctic sea-ice forecast. *Geophysical Research Letters*, *51*(13), e2024GL110405. https://doi.org/10.1029/2024gl110405

Spreen, G., Kwok, R., Menemenlis, D., & Nguyen, A. (2016). Sea-ice deformation in a coupled ocean–sea-ice model and in satellite remote sensing data. *The Cryosphere*, *11*(4), 1553–1573. https://doi.org/10.5194/tc-11-1553-2017

Tietsche, S., Notz, D., Jungclaus, J., & Marotzke, J. (2013). Assimilation of sea-ice concentration in a global climate model–physical and statistical aspects. *Ocean Science*, *9*(1), 19–36. https://doi.org/10.5194/os-9-19-2013

Tödter, J., & Ahrens, B. (2015). A second-order exact ensemble square root filter for nonlinear data assimilation. *Monthly Weather Review*, *143*(4), 1347–1367. https://doi.org/10.1175/mwr-d-14-00108.1

Vetra-Carvalho, S., Van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., et al. (2018). State-of-the-art stochastic data assimilation methods for high-dimensional non-gaussian problems. *Tellus A: Dynamic Meteorology and Oceanography*, *70*(1), 1–43. https://doi.org/10.1080/16000870.2018.1445364

Wang, K., Ali, A., & Wang, C. (2023). Local analytical optimal nudging for assimilating amsr2 sea ice concentration in a high-resolution pan-arctic coupled ocean (hycom 2.2. 98) and sea ice (cice 5.1. 2) model. *The Cryosphere*, *17*(10), 4487–4510. https://doi.org/10.5194/tc-17-4487-2023

Wang, Q., Danilov, S., Jung, T., Kaleschke, L., & Wernecke, A. (2016). Sea ice leads in the arctic ocean: Model assessment, interannual variability and trends. *Geophysical Research Letters*, *43*(13), 7019–7027. https://doi.org/10.1002/2016gl068696

Xie, J., Counillon, F., Bertino, L., Tian-Kunze, X., & Kaleschke, L. (2016). Benefits of assimilating thin sea ice thickness from smos into the topaz system. *The Cryosphere*, *10*(6), 2745–2761. https://doi.org/10.5194/tc-10-2745-2016

Xie, J., & Zhu, J. (2010). Ensemble optimal interpolation schemes for assimilating argo profiles into a hybrid coordinate ocean model. *Ocean Modelling*, *33*(3–4), 283–298. https://doi.org/10.1016/j.ocemod.2010.03.002

Xu, J., Fu, J., Gan, L., Chen, Y., Huang, Z., & Yang, G. (2022). Accelerating cryo-em reconstruction of relion on the new sunway supercomputer. In *2022 IEEE intl conf on parallel & distributed processing with applications, big data & cloud computing, sustainable computing & communications, social computing & networking (ispa/bdcloud/socialcom/sustaincom)* (pp. 129–138).

Yang, Q., Losa, S. N., Losch, M., Tian-Kunze, X., Nerger, L., Liu, J., et al. (2015). Assimilating smos sea ice thickness into a coupled ice-ocean model using a local seik filter. *Journal of Geophysical Research: Oceans*, *119*(10), 6680–6692. https://doi.org/10.1002/2014jc009963

Yang, Q., Losch, M., Losa, S. N., Jung, T., Nerger, L., & Lavergne, T. (2016). Brief communication: The challenge and benefit of using sea ice concentration satellite data products with uncertainty estimates in summer sea ice data assimilation. *The Cryosphere*, *10*(2), 761–774. https://doi.org/10.5194/tc-10-761-2016

Zhang, J., & Hibler, W., III. (1997). On an efficient numerical method for modeling sea ice dynamics. *Journal of Geophysical Research*, *102*(C4), 8691–8702. https://doi.org/10.1029/96jc03744

Zhang, J., & Rothrock, D. A. (2003). Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Monthly Weather Review*, *131*(5), 845–861. https://doi.org/10.1175/1520-0493(2003)131⟨0845:MGSIWA⟩2.0.CO;2

Zhang, Y., Cheng, X., Liu, J., & Hui, F. (2018). The potential of sea ice leads as a predictor for summer arctic sea ice extent. *The Cryosphere*, *12*(12), 3747–3757. https://doi.org/10.5194/tc-12-3747-2018

Zhang, Y. F., Bushuk, M., Winton, M., Hurlin, B., Delworth, T., Harrison, M., et al. (2022). Subseasonal-to-seasonal arctic sea ice forecast skill improvement from sea ice concentration assimilation. *Journal of Climate*, *35*(13), 4233–4252. https://doi.org/10.1175/jcli-d-21-0548.1