

# Prediction of Sea Ice Velocity and Concentration in the Arctic Ocean using Physics-informed Neural Network

Younghyun Koo<sup>a,b,c</sup>, Maryam Rahnemoonfar<sup>a,b</sup>

<sup>a</sup>*Department of Computer Science and Engineering, Lehigh University, 27 Memorial Dr W, Bethlehem, 18015, PA, USA*

<sup>b</sup>*Department of Civil and Environmental Engineering, Lehigh University, 27 Memorial Dr W, Bethlehem, 18015, PA, USA*

<sup>c</sup>*National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, 1540 30th St, Boulder, 80303, CO, USA*

---

## Abstract

As an increasing amount of remote sensing data becomes available in the Arctic Ocean, data-driven machine learning (ML) techniques are becoming widely used to predict sea ice velocity (SIV) and sea ice concentration (SIC). However, fully data-driven ML models have limitations in generalizability and physical consistency due to their excessive reliance on the quantity and quality of training data. In particular, as Arctic sea ice entered a new phase with thinner ice and accelerated melting, there is a possibility that an ML model trained with historical sea ice data cannot fully represent the dynamically changing sea ice conditions in the future. In this study, we develop physics-informed neural network (PINN) strategies to integrate physical knowledge of sea ice into the ML model. Based on the Hierarchical Information-sharing U-net (HIS-Unet) architecture, we incorporate the physics loss function and the activation function to produce physically plausible SIV and SIC outputs. Our PINN model outperforms the fully data-

driven model in the daily predictions of SIV and SIC, even when trained with a small number of samples. The PINN approach particularly improves SIC predictions in melting and early freezing seasons and near fast-moving ice regions.

*Keywords:* Arctic Ocean, Convolutional neural network (CNN), Hierarchical Information-sharing U-net (HIS-Unet), Physics loss function, Activation function

---

## 1. Introduction

Dramatic declines in the Arctic sea ice extent (SIE) and sea ice thickness (SIT) have been observed for the last few decades along with a warming climate resulting from the anthropogenic CO<sub>2</sub> emissions (Notz and Stroeve, 2016). Arctic SIE has been reduced by more than 50,000 km<sup>2</sup>/year, and at the same time, Arctic SIT has decreased by more than 2 m in average, with more than a 50 % SIT reduction in thick multi-year ice (Kwok, 2018). With such dramatic changes in SIE and SIT, the thermodynamic and dynamic conditions of Arctic sea ice have entered a new phase (Wang et al., 2022; Stroeve and Notz, 2018). Given that sea ice plays a critical role in the physical conditions of the atmosphere and ocean (e.g., heat exchange, salinity, etc.) (Haumann et al., 2016; Hirst, 1999; McPhee, 1992; Budikova, 2009), understanding sea ice dynamics is essential for understanding the future of the Arctic Ocean and global climate.

A traditional way to predict the sea ice dynamics in the Arctic Ocean is using numerical models based on physics laws and equations that describe sea ice in terms of mass and momentum balances, as well as its interaction

with the atmosphere and ocean (Stark et al., 2008; Fritzner et al., 2019; Hibler, 1979; Parkinson and Washington, 1979). Although physics models have successfully explained general sea ice behaviors by assuming sea ice as viscous-plastic or elastic-viscous-plastic material (Hunke and Dukowicz, 1997; Hibler, 1979), several challenges make it difficult to improve their fidelity and efficiency. First, since countless atmospheric, oceanic, and even biological variables are involved in sea ice dynamics, established physics equations have limitations in explaining every detailed mechanism of sea ice dynamics. If the sea ice dynamics are partially governed by any unknown mechanisms, physics models may not agree perfectly with the real observations. Second, uncertainties in the complicated parameterization of sea ice physical conditions, including initial and boundary conditions, can propagate through the model predictions (Blockley et al., 2020; Hunke et al., 2010). Since real observations are not always available, physics models are often based on physical assumptions. Considering that numerical models are sensitive to initial conditions and physical assumptions, they can result in significant uncertainties and discrepancies with real observations (Blanchard-Wrigglesworth et al., 2015). Third, solving coupled and non-linear partial differential equations (PDEs) is computationally intensive and expensive. Although adding more complexity to a model or increasing model resolution can enhance the model's accuracy, it exponentially increases computational costs (Hunke et al., 2010; Li et al., 2023). Hence, this computational demand inhibits improving the model performance for higher spatial and temporal resolutions.

Recently, machine learning (ML) techniques have emerged to predict sea ice behaviors with fewer complexities and a lower computational cost than

traditional numerical models. ML approaches have several benefits compared to numerical models. First, complex PDEs or parameterizations are not required for ML models because they need only observational data to be trained. ML models can learn complex non-linear patterns or relationships between multiple atmospheric and oceanic variables directly from observational data without any explicit knowledge of physical processes. Second, ML models can be easily updated and trained continuously with new observational data, and this data integration can make the model prediction more accurate and reliable. Finally, ML models require less computational cost than numerical models. While numerical models often require intensive computing resources (e.g., high-performance computing) to solve complex PDEs, ML models can be implemented relatively fast by leveraging the parallel processing ability of graph processing units (GPUs).

Based on such advantages, various machine learning models have been developed to predict sea ice concentration (SIC) and sea ice velocity (SIV), most of which rely on convolutional neural networks (CNN) (Ren and Li, 2021; Liu et al., 2021; Yan and Huang, 2018; Hoffman et al., 2023; Petrou and Tian, 2019). As a deep learning algorithm specialized for image data, CNN propagates the input data through convolutional layers with multiple filters and extracts spatial patterns and features. Based on the ability of CNN to learn complex features in image datasets, CNNs have succeeded in various applications in environmental modeling (Wang et al., 2024; Jiang et al., 2023; Sadeghi et al., 2020). Since the SIC and SIV data for the entire Arctic Ocean are often provided from remote sensing data sources as grid formats, CNN has advantages in exploiting the spatial variability in those

sea ice data.

However, the fidelity of fully data-driven ML models is highly dependent on the quantity and quality of training datasets, which leads to limitations in further improving the model performance. First and foremost, data-driven ML models require a large amount of data to ensure generalizability. If training samples are insufficient or their distribution is biased, the model can be overfitted and lack generalizability to out-of-training cases. Second, ML models are considered *black-box* models and lack inherent interpretability in how the final predictions are derived. This lack of inherent interpretability makes it difficult to understand why the model predicts physically invalid values about sea ice conditions (e.g., negative SIC or SIC > 100%).

In order to address such issues of the fully data-driven ML models and improve the model’s generalizability, it is helpful to guide the learning process to agree with fundamental physical laws and domain knowledge. By integrating physics knowledge and constraints into ML training, the ML models can yield physically consistent predictions even in the presence of imperfect data, such as missing, noise, or outliers data (Karniadakis et al., 2021; Raissi et al., 2019; Maier et al., 2023). This integration of ML and physics knowledge is often referred to as *physics-informed machine learning (PIML)*, and has been proposed for various dynamic systems (Karniadakis et al., 2021; Maier et al., 2023). Recently, many studies have attempted this PIML framework for various applications including, but not limited to, flow dynamics, material sciences, molecular simulations, chemistry (Karniadakis et al., 2021). In particular, PIML based on well-known flow laws such as Stokes equation has shown great success in modeling the dynamics of ice

sheets and glaciers (i.e., land ice) (Teisberg et al., 2021; Riel et al., 2021; Riel and Minchew, 2023; Iwasaki and Lai, 2023; Jouvet and Cordonnier, 2023; He et al., 2023). In sea ice application, Liu et al. (2024) proposed dual-task neural network architecture and incorporated a loss function based on sea ice control equation. However, it remains unclear how this PIML concept and physics-informed loss function can contribute to future sea ice predictions under rapidly changing climate conditions.

This study aims to improve the fidelity of deep learning models that predict sea ice dynamics by integrating physical knowledge into the model training. In order to enforce the deep learning model (neural network) to converge into physically valid SIV and SIC values, we explicitly adopt a physics-informed neural network (PINN) approach based on loss function. We embed this physics-informed learning strategy in the hierarchical information-sharing U-net (HIS-Unet), a CNN model for SIV and SIC predictions (Koo and Rahnemoonfar, 2024). The main contributions of this research consist of the following.

- We design physics loss functions and combine them with the data loss function to regulate physical validity of SIC and SIV.
- We modify the output layer to guarantee the physically valid SIC values.
- Our extensive experiments show that the physics-informed deep learning model can improve the performance in SIC and SIV predictions even with a small number of training samples.
- We explore the spatiotemporal variability in the improved performance

of our physics-informed learning strategies compared to fully data-driven deep learning models that do not embed physics.

The remainder of the paper is organized as follows. Section 2 provides a brief review of the physical background of sea ice modeling, machine learning for sea ice prediction, and PIML. Section 3 explains details of remote sensing and meteorological data used in this study, and section 4 presents the architecture of our CNN model and physics-informed learning strategies. The performance and implication of our PINN are discussed in Section 5.

## 2. Background

### 2.1. Physics of sea ice dynamics

In physical sea ice models, the volume and area of sea ice are determined by thermodynamic evolution and dynamic motion field. The spatiotemporal changes in SIT ( $h$ ) and SIC ( $A$ ) can be expressed by the following equations for mass conservation (Hibler, 1979; Holland and Kwok, 2012; Flato, 2004):

$$\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u}h) = S_h \quad (1)$$

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

where  $\mathbf{u}$  is the ice motion vector,  $S_h$  and  $S_A$  are the changes in SIT and SIC driven by thermodynamic sources (e.g., freezing or melting), respectively.

Additionally, the following momentum equations have been used in various numerous physical sea ice models to explain the balance of horizontal

forces on sea ice based on the most common assumption of elastic-viscous-plastic (EVP) properties of sea ice (Hibler, 1979):

$$m \frac{D\mathbf{u}}{Dt} = -mf\mathbf{k} \times \mathbf{u} + \tau_{ai} + \tau_{wi} + \mathbf{F} - mg\nabla H \quad (3)$$

where  $D/Dt = \partial/\partial t + \mathbf{u} \cdot \nabla$  is the substantial time derivative,  $m$  is the ice mass per unit area,  $\mathbf{k}$  is a unit vector normal to the surface,  $\mathbf{u}$  is the ice velocity,  $f$  is the Coriolis parameter,  $\tau_{ai}$  and  $\tau_{wi}$  are the forces due to air and water stresses,  $H$  is the elevation of the sea surface,  $g$  is the gravity acceleration, and  $\mathbf{F}$  is the force due to variations in internal ice stress. Following this equation, many previous studies described SIV as interacting with wind and ocean forcings (Rampal et al., 2016; Wang et al., 2014; Timmermann et al., 2009; Salas Mélia, 2002). Particularly, wind velocity has been treated as a major variable in SIV, contributing to up to 70 % of the sea ice velocity variances (Thorndike and Colony, 1982) depending on season or region. Nevertheless, predicting sea ice dynamics based on physical models is still challenging due to the intrinsic complexity and dependency of physical models on numerous parameterizations.

## 2.2. Neural network for sea ice prediction

Convolutional neural networks (CNNs) have been used as the most popular and efficient deep learning network for modeling SIC and SIV. First, in terms of SIC, Andersson et al. (2021) proposed a deep-learning sea ice forecasting system named IceNet to forecast monthly SIC for the next six months. Grigoryev et al. (2022) used a U-Net architecture, a type of deep CNN, for daily SIC forecasts in several subsections of the Arctic Ocean, including the Barents and Kara Seas, Labrador Sea, and Laptev Sea. Kim et al. (2020) also

used CNN to predict monthly SIC from satellite-based SIC observations, sea surface temperature, air temperature, albedo, and wind velocity data from the previous months. Similarly, Fritzner et al. (2020) proposed a CNN that predicts weekly SIC from SIC, sea surface temperature, and air temperature from the previous six days. CNN models proposed by Ren and Li (2021) and Ren et al. (2022) used the sequences of satellite-derived SIC observations to make daily SIC predictions. In some studies, long short-term memory (LSTM), an advanced recurrent neural network (RNN), has been modified and inserted into CNN architecture to obtain better performance in both spatial and temporal SIC predictions (Liu et al., 2021). CNN and LSTM have also been used for SIV prediction in several studies. Zhai and Bitz (2021) and Hoffman et al. (2023) used CNN to predict daily SIV, and their model outperformed other statistical and physical models. Petrou and Tian (2019) showed that adding LSTM units to convolutional layers improves the performance of a model that predicts SIV. In this study, we adopt a multi-task neural network framework named HIS-Unet (Koo and Rahnemoonfar, 2024) and improve the predictability of SIV and SIC of this model by incorporating physics-informed training scheme.

### *2.3. Physics-informed machine learning (PIML)*

Historically, PIML has been achieved by integrating physics knowledge into ML training in the form of: (i) data, (ii) model architecture, and (iii) optimization (Hao et al., 2023; Karniadakis et al., 2021). First, one of the straightforward methods to incorporate physics knowledge into ML is to generate training data from the desired physics knowledge. Data-driven ML models that are trained with sufficient simulation data governed by certain

underlying physics laws can accurately represent these physics laws. Given that enough high-quality and labeled data is not always available for real-world tasks, synthetic data constructed with physical simulations can provide a large amount of data with high quality. However, this approach can have limitations in perfectly reflecting real world because real-world data have different distributions from simulation data. In addition, the cost of data acquisition can be a critical issue, as simulation data is often generated via expensive experiments or large-scale simulations.

The second way to implement PIML is to design neural network architectures that implicitly embed any prior knowledge and inductive biases associated with a given task (e.g., symmetry, conservation laws, etc.) (Karniadakis et al., 2021). CNN is a popular example of this neural network architecture that achieves extensive applicability for image recognition by respecting invariance and symmetry groups in natural images (Karniadakis et al., 2021; Mallat, 2016). Another example of this category is equivariant networks, which embed the dynamic changes in spatial coordinates to preserve the equivariance of data points to rotation and translation (Satorras et al., 2022; Schütt et al., 2017). Some neural network architectures also used Lagrangian and Hamiltonian mechanics to enforce the energy conservation property of the networks (Hao et al., 2023). Furthermore, in solving PDEs, there have been several attempts to modify neural network architecture to satisfy the required initial conditions (Hao et al., 2023). Although such a model architecture approach can be effective with relatively simple and well-defined physics or symmetry groups, this approach has limitations in extending to highly complex problems.

Lastly, physics knowledge can be integrated into the optimization and learning process by imposing constraints of prior physics knowledge into the form of loss functions. As a soft manner of penalizing the loss function instead of enforcing a specific condition directly, this approach can indirectly reshape the target spaces of NN output to converge to physically plausible solutions (Karniadakis et al., 2021; Hao et al., 2023); in this study, we will refer to this approach specifically as *PINN*. This approach can be regarded as a case of multi-task learning, balancing the loss functions for two tasks of fitting both the observation data and physical constraints. Although such a soft constraint approach is the most common and flexible way for PIML, balancing these two tasks can be challenging because they can counteract the convergence to each other’s solution.

In the modeling of land ice, PIML has emerged as an effective way to predict ice flow, and most previous studies employed the optimization approach by adding physics loss functions or regularization terms (i.e., PINN). Teisberg et al. (2021) developed a PINN to predict ice thickness and velocity by adding a physics loss function based on the mass conservation of ice sheets. Riel et al. (2021) introduced a physics loss function derived from the governing equations of ice flow to their PINN framework to infer spatially and temporally varying basal drag in ice sheets. Riel and Minchew (2023) proposed a PINN framework to predict the distribution of ice rigidity by adding the loss functions of ice flow and Kullback-Leibler divergence. Iwasaki and Lai (2023) proposed a physics loss function to fit the model results with physics laws regarding ice flows, and this physics loss function contributed to improving the model performance when the data was contaminated by

noise. Jouvet and Cordonnier (2023) optimized their PINN by minimizing the energy associated with high-order ice-flow equations. Cheng et al. (2024) employed a PINN to infer basal sliding while filling gaps in sparsely observed ice thickness data. In the regime of sea ice, there have been several attempts to integrate physics knowledge into machine learning by training the model with the data retrieved from physical models (Palerme and Muller, 2021; Palerme et al., 2024). Liu et al. (2024) used a dual-task CNN architecture and physics-informed loss function to enforce dynamic constraints of SIC and SIV. Despite such developments of PINN for cryosphere, it is still necessary to assess how PINN and physical constraints can help future predictions of sea ice conditions. Therefore, this study explicitly applies the concept of PINN to sea ice prediction and explore the how weights to physics-informed loss function and representability of training samples contribute to the model predictability for future unseen sea ice conditions.

### 3. Data

In this study, we use daily SIV and SIC data collected from satellite observation from 2009 to 2022. We use the SIV and SIC from previous three days (inputs of the model) to predict the next day's SIV and SIC (output of the model) (Table 1). Besides satellite observations of SIV and SIC, we also use wind velocity and air temperature from reanalysis as additional input variables for the model. Additionally, we add X and Y coordinates as inputs to represent regional variability. The input datasets and their sources are summarized in Table 1. This section presents the details of these datasets.

Table 1: Input data sets for machine learning models

Dataset	Name	Spatial resolution
Sea ice velocity (u and v)	NSIDC Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors (Tschudi et al., 2020)	25 km
Sea Ice Concentration	NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration (Meier et al., 2021)	25 km
10 m wind velocity (u and v)	ECMWF Reanalysis v5 (ERA5) hourly data on single levels (Hersbach et al., 2020)	0.25°
2 m air temperature	ECMWF Reanalysis v5 (ERA5) hourly data on single levels (Hersbach et al., 2020)	0.25°

### *3.1. Sea ice velocity*

We use the NSIDC Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors version 4 (Tschudi et al., 2019, 2020) as the input and output SIV. This product derives daily SIV from three primary types of sources: (1) gridded satellite images, (2) wind reanalysis data, and (3) buoy position data from the International Arctic Buoy Program (Tschudi et al., 2019, 2020). The u component (along-x) and v component (along-y) of SIV are independently derived from each of these sources and optimally interpolated onto a 25 km Equal-Area Scalable Earth (EASE) grid. When SIV is derived from satellite images, a correlation coefficient is calculated between a small target area in a one-day image and a searching area in the next-day image. Then, the location in the next-day image with the highest correlation coefficient is determined as the displacement of ice (Tschudi et al., 2019). The mean difference between the Polar Pathfinder and buoy measurements of SIV is approximately 0.1 km/day and 0.3 km/day for u and v components, respectively (Tschudi et al., 2019). We note that the SIV of this product is valid over short distances away from the ice edge in areas where ice conditions are relatively stable, stationary, homogenous, and isotropic (Tschudi et al., 2020). In this study, we exclude SIV values close to the coastlines within 50 km (or 2 pixels) of distance.

### *3.2. Sea ice concentration*

For the SIC data, we use NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration version 4 data (Meier et al., 2021). This data set provides a Climate Data Record (CDR) of SIC (i.e., the areal fraction of ice within a grid cell) from passive microwave (PMW) data. The CDR

algorithm output is the combination of SIC estimations from two algorithms: the NASA Team algorithm (Cavalieri et al., 1984) and NASA Bootstrap algorithm (Comiso, 1986). These empirical algorithms estimate SIC from the PMW brightness temperatures at different frequencies and polarizations: vertical and horizontal polarizations at 19 GHz, 22 GHz, and 37 GHz. Several assessments showed that the error of this SIC estimation is approximately 5 % within the consolidated ice pack during cold winter conditions (Meier, 2005; Comiso et al., 1997; Ivanova et al., 2015). However, in the summer season, the error can rise to more than 20 % due to surface melt and the presence of melt ponds (Kern et al., 2020). Due to the data quality issue near coastal areas, we use the SIC data more than 50 km from the coastline. To match the coordinate systems of the SIC and SIV products, we reproject the NSIDC Sea Ice Polar Stereographic grid of the SIC product into the EASE grid of the SIV product using bilinear interpolation.

### *3.3. ERA5 climate reanalysis*

As shown in Eq. 3, sea ice dynamics are largely associated with atmospheric and oceanic circulation. Thus, we use the wind speed and air temperature as the input variables of the ML model. We use the fifth-generation ECMWF (European Centre for Medium-Range Weather Forecasts) atmospheric reanalysis (ERA5) as the data sources for wind velocity and air temperature. ERA5 provides hourly estimates of atmospheric, land, and oceanic climate variables, covering the period from January 1940 to the present (Hersbach et al., 2020). We obtain the daily average wind velocity (zonal and meridional components) at 10 m height and 2 m air temperature from this hourly data. To co-locate this data with the SIV and SIC data, we

reproject the raw ERA5 latitude-longitude grid onto the 25 km EASE grid using bilinear interpolation.

## 4. Methods

In this study, we use a hierarchical information-sharing U-net (HIS-Unet) based on Koo and Rahnemoonfar (2024) as a backbone deep learning model (Fig. 1) to make daily predictions of SIC and SIV from inputs of SIC, SIV, wind, and atmospheric temperature from the previous three days. As a fully convolutional network, the HIS-Unet is designed to predict SIC and SIV simultaneously. By sharing SIC and SIV information during the propagation processes, the HIS-Unet achieves better fidelity than other neural networks and statistical approaches in predicting both SIC and SIV. In particular, this information sharing is proven to be useful for predicting sea ice conditions where and when SIV has impacts on SIC changes (Koo and Rahnemoonfar, 2024). Since understanding sea ice dynamics requires both SIV and SIC information (e.g., Eq. 2), we use the multi-task prediction of SIV and SIC by the HIS-Unet to facilitate integrating the physics knowledge of sea ice dynamics. To embed fundamental knowledge of sea ice dynamics into the HIS-Unet architecture, we (1) introduce physics loss functions in the training phase and (2) add an additional activation function to the output layer. This section presents the details of the HIS-Unet architecture and physics-informed training strategies used in this study.

### 4.1. Hierarchical information-sharing U-net (HIS-Unet)

The HIS-Unet architecture consists of two separate task branches, each for predicting SIV and SIC (Fig. 1). The task branches are connected with

weighting attention modules (WAMS) that allow information sharing between the SIV and SIC branches. SIC and SIV branches have U-net structures (Ronneberger et al., 2015) with a series of encoders (contracting path) and decoders (expansive path). Each encoder path consists of repeated two  $3 \times 3$  convolutions and a  $2 \times 2$  max pooling operation with stride 2 for downsampling and doubling the number of features. The decoders conduct  $2 \times 2$  up-convolution and half the number of feature channels, followed by a concatenation with the cropped feature map and two  $3 \times 3$  convolutions (Ronneberger et al., 2015). The hyperbolic tangent (tanh) activation function is applied after each convolutional layer:

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4)$$

The separated U-net structures of SIV and SIC branches output SIV or SIC, respectively, but they share and transfer their information through six weighting attention modules (WAMs) during the propagation process. Six WAMs are inserted into 3 encoder steps and 3 decoder steps between SIC and SIV branches (Fig. 1).

Each WAM first receives information from the SIV and SIC branches and calculates the weighted sum of them (Fig. 1b). Letting a WAM receive the SIV feature map ( $\xi_{in,SIV}$ ; height  $H$ , width  $W$ , channels  $C$ ) and SIC feature map ( $\xi_{in,SIC}$ ;  $H \times W \times C$ ), the input shared information ( $\xi_{in,share}$ ) is determined by multiplying linear weights,  $A_{in,SIV}$  and  $A_{in,SIC}$ , to  $\xi_{in,SIV}$  and  $\xi_{in,SIC}$ , respectively (Fig. 1b). Then, this shared information  $\xi_{in,share}$  passes through sequentially arranged channel attention (Fig. 1c) and spatial attention modules (Fig. 1d). The channel attention highlights what channel is meaningful, and the spatial attention highlights where an informative part

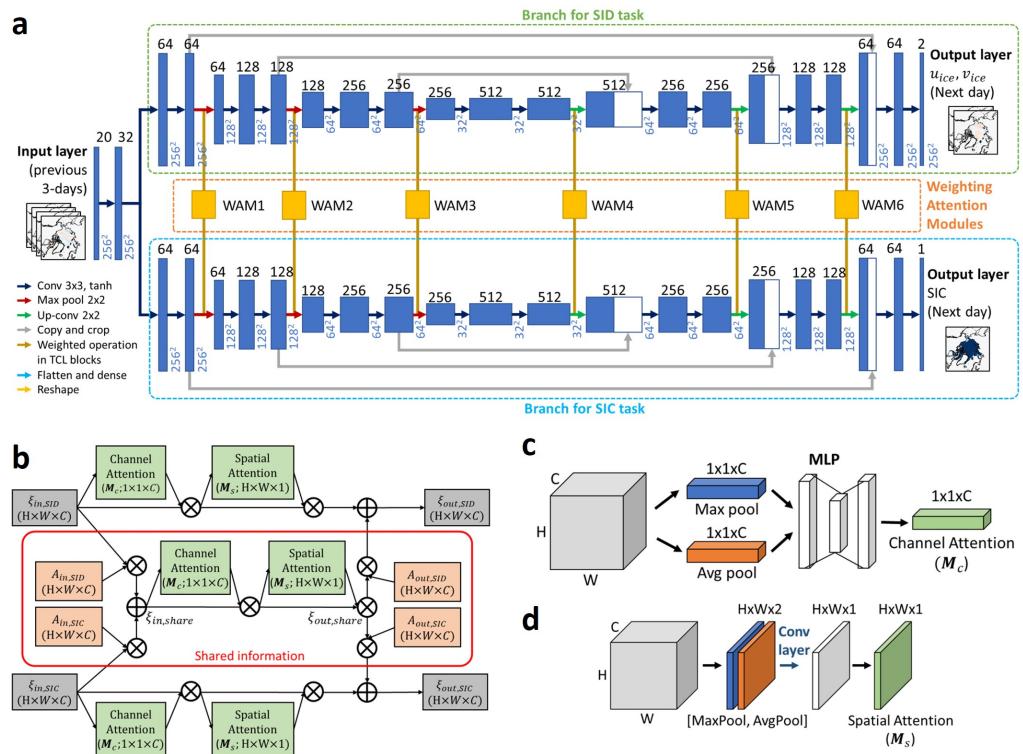


Figure 1: (a) Architecture of Hierarchical information-sharing U-net (HIS-Unet); (b) Weighting attention module (WAM) in HIS-Unet; (c) Channel attention module and (d) Spatial attention module in a WAM (Koo and Rahmoomfar, 2024).

is spatially located (Woo et al., 2018). These channel and spatial attention modules are also applied to the input SIV and SIC feature maps ( $\xi_{in,SIV}$  and  $\xi_{in,SIC}$ ) (Fig. 1b). Then, the attention shared information is sent to the SIV and SIC branches after multiplying output weights ( $A_{out,SIV}$  and  $A_{out,SIC}$ ) and adding attention SIV and SIC information, respectively. More details of how the WAMs enable sharing and highlighting SIC and SIV information are described in Koo and Rahnemoonfar (2024).

#### *4.2. Physics-informed training*

In order to make the HIS-Unet incorporate physics knowledge of sea ice, we embed two physics-informed regularizations: (1) include physics loss functions along with the data loss function and (2) insert the sigmoid activation function to the SIC branch to guarantee valid SIC values.

The original HIS-Unet is optimized by the mean square error (MSE) objective loss function ( $L_{data}$ ). The MSE data loss term ( $L_{data}$ ) is calculated by the following equation:

$$L_{data} = \sum(|u_p - u_o|^2 + |v_p - v_o|^2 + |A_p - A_o|^2) \quad (5)$$

where  $u$  and  $v$  denote x-component and y-component of SIV, respectively,  $A$  denotes SIC, and the subscript  $o$  means observation and  $p$  means prediction by HIS-Unet. In addition to this MSE data loss function, we design physics loss functions that apply physical constraints: (i)  $L_{sat}$  constraining valid SIV values and (ii)  $L_{therm}$  constraining thermodynamically valid ice growth.

First, we regularize the valid SIV values associated with SIC values. Since the PMW-derived SIV can be only defined where sea ice presents with  $> 15$

% of SIC (Tschudi et al., 2019), SIV should be zero where SIC is less than 15 %. Therefore, we design the first physics loss term as follows:

$$L_{sat} = \begin{cases} |u_p^2 + v_p^2|, & \text{if } A_p < 0.15 \\ 0, & \text{if } A_p \geq 0.15 \end{cases} \quad (6)$$

The next physics loss term ( $L_{therm}$ ) is based on Eq.2, which explains the thermodynamic and dynamic SIC changes. The thermodynamic SIC changes ( $S_A$ ), the right term in Eq. 2, can be calculated by temporal changes of SIC and the combination of advection and divergence of SIC. In the case of daily SIC prediction, the daily  $S_A$  can be assumed not to exceed (-1, 1). That is, we assume that the thermodynamic freezing and melting of sea ice are unlikely to saturate SIC from 0 % to 100 % or remove the entire sea ice from 100 % to 0 % within a day in most cases. Thus, based on this assumption, we design the second physics loss term as follows:

$$L_{therm} = \text{ReLU}\left(\left|\frac{\partial A_p}{\partial t} + \nabla \cdot (\mathbf{u}_p A_p)\right| - 1\right) \quad (7)$$

where the time derivative term ( $\frac{\partial A_p}{\partial t}$ ) is derived by subtracting the previous-day SIC from the output SIC, and the spatial derivative term ( $\nabla \cdot (\mathbf{u}_p A_p)$ ) is derived by calculating the spatial gradients in output SIV and SIC grids. By using  $L_{therm}$ , we can assign a linearly increasing penalty as the thermodynamic ice growth prediction exceeds 1.

Consequently, the total physics loss term ( $L_{phy}$ ) and the final objective loss functions ( $L$ ) are defined as follows:

$$L_{phy} = \lambda_{sat} L_{sat} + \lambda_{therm} L_{therm} \quad (8)$$

$$L = L_{data} + L_{phy} \quad (9)$$

where  $\lambda_{sat}$  and  $\lambda_{therm}$  are the relative weights of  $L_{sat}$  and  $L_{therm}$ , respectively, to the data loss term. In this study, we conduct experiments with different  $\lambda_{sat}$  and  $\lambda_{therm}$  of 0, 0.2, 1.0, and 5.0 and examine how this weight changes the model performance.

As the second regularization, we enforce the output SIC to lie within the value range of 0 to 1 (0 to 100 %) by adding the sigmoid activation function to the last output layer of the SIC branch:

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (10)$$

This can constrain the valid range of SIC outputs, even for out-of-training samples.

#### *4.3. Training strategy*

In our HIS-Unet, we use the previous 3 days of SIV (x- and y-components), SIC, air temperature, and wind velocity (x- and y- components) as the inputs to predict the next day's SIV and SIC. Consequently, the input layer has 18 channels of  $256 \times 256$  grid size. All input values are normalized to -1 to 1 based on the nominal maximum and minimum values that each variable can have. All the data is collected for 14 years from 2009 to 2022; the first seven years of data (2009-2015) are utilized as training data, and the remaining seven years of data (2016-2022) are utilized as test data. To examine how the model performance is changed by the number of training data, we train the model with three different training sample sizes: (1) using all 2009-2015 data as training samples (i.e., 100 % sampling), (2) randomly selecting 50 % of 2009-2015 data as training samples (i.e., 50 % sampling), and (3) randomly selecting 20 % of 2009-2015 data as training samples (i.e., 20 % sampling).

The total number of training samples is 2,177 for the 100 % sampling, 1,088 for 50 % sampling, and 436 for the 20 % sampling. The HIS-Unet models trained with three different sample sizes are applied to the 2016-2022 test data. All models are optimized by the Adam stochastic gradient descent algorithm (Kingma and Ba, 2017) with 100 epochs and a 0.001 learning rate. All scripts are executed on eight NVIDIA RTX A5000 GPUs with 24 GB of memory.

#### 4.4. Model performance

The model performance is assessed by the root mean square error (RMSE):

$$\text{RMSE}(\hat{y}, y) = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (11)$$

where  $\hat{y}$  denotes predicted values,  $y$  denotes true values,  $N$  is the number of data points. In the case of SIV, we calculate RMSE for both x- and y-component SIV, and the average of x- and y-component RMSEs is determined as the SIV RMSE to embrace the magnitude and angle error of SIV. We assess and compare the RMSEs from the HIS-Unet with three different training sample cases and four different  $\lambda_{sat}$  and  $\lambda_{therm}$  (i.e., 0, 0.2, 1.0, and 5.0). We use the HIS-Unet without any physics-informed regularization (i.e., trained only with the data loss  $L_{data}$  and without the sigmoid activation function to the SIC branch) as the baseline model, and this purely data-driven model is notated as *No-Phy*. Meanwhile, we call the physics-informed HIS-Unet simply PINN. We compare the RMSE differences between PINN and *No-Phy* for different training sample sizes (20 %, 50 %, and 100 %). Additionally, we examine how these RMSE differences vary by month and region in seven test years (2019-2022). To assess the statistical significance

of RMSE improvement by PINN, we conduct paired t-tests between RMSEs of PINN and RMSEs of *No-Phy*. If the p-value from the t-test is below 0.05, we determine that the RMSE difference between PINN and *No-Phy* is significant.

In addition to RMSE, we use the mean absolute error (MAE) and anomaly correlation coefficient (ACC) to evaluate the predictive performance of PINN and *No-Phy*.

$$\text{MAE}(\hat{y}, y) = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N} \quad (12)$$

$$\text{ACC}(\hat{y}, y) = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^N (y_i - \bar{y}_i)^2}} \quad (13)$$

Lower RMSE and MAE values indicate higher prediction accuracy, while a higher ACC (i.e., approaching 1) reflects stronger agreement between predicted and observed anomalies, indicating better predictive skill. We note that the performance of *No-Phy* model trained with full training samples can be found in Koo and Rahnemoonfar (2024) in detail. The *No-Phy* HIS-Unet showed better performance than another simple statistical model (e.g., linear regression), physical model (Hybrid Coordinate Ocean Model; HyCOM), simple convolutional network, and independent-branch U-nets. Since the HIS-Unet performance has already been assessed in the previous study, this study is focused on assessing the effect of using a physics-informed learning strategy rather than assessing the performance of the HIS-Unet architecture itself.

## 5. Results and discussion

### 5.1. Performance of PINN

We evaluate the predictive performance of *No-Phy* and PINNs with different  $\lambda_{sat}$  and  $\lambda_{therm}$  settings, and check whether PINNs can improve the model performance compared to *No-Phy*. Additionally, we investigate how this improvement pattern varies by different training sample sizes. Figure 2 shows the SIV and SIC RMSE changes by PINNs as functions of different training sample sizes,  $\lambda_{sat}$ , and  $\lambda_{therm}$ . Table 2 shows the RMSE, MAE, and ACC of *No-Phy* baseline and PINN with 0.2  $\lambda_{therm}$  and 0.2  $\lambda_{sat}$ . According to Table 2, the errors of both SIV and SIC predictions decrease with more training samples.

Regarding SIV, the integration of physics loss functions, including both  $\lambda_{therm}$  and  $\lambda_{sat}$ , improves the model fidelity relative to the *No-Phy* baseline model (Fig. 2a). The reduction in SIV RMSE by the physics loss term is observed in all sample sizes, but this effect is more significant in the 20 % sample size: SIV RMSE decreases by up to 0.10 km/day when  $\lambda_{sat}$  is set to 0.2. When 50 % and 100 % of training samples are used, the SIV RMSE decreases by up to 0.02 km/day and 0.05 km/day, respectively. We also find improvements of SIV MAE and ACC by PINN in all training sample cases (Table 2).

On the other hand, regarding SIC, the PINNs show lower SIC RMSEs than the *No-Phy* baseline model (Fig. 2), and the improvement of the PINN is more significant for smaller sample sizes. When 20 % of training samples are used, the PINNs show approximately 1.1 % lower SIC RMSE than the *No-Phy* model. When the training samples are set to 50 % and 100 %, the

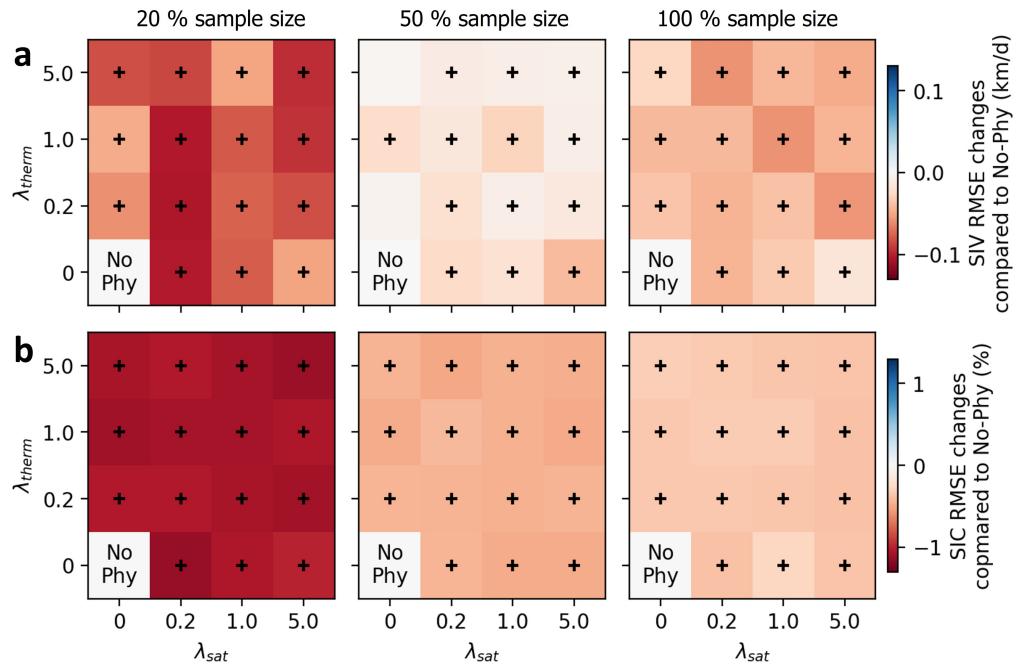


Figure 2: (a) The reduction of SIV RMSE by PINNs with different training sample sizes (20, 50, and 100 %) and combinations of  $\lambda_{sat}$  and  $\lambda_{therm}$  (0, 0.2, 1.0, and 5.0), relative to the SIV RMSE of the *No-Phy* model. (b) The reduction of SIC RMSE by PINNs with different training sample sizes and combinations of  $\lambda_{sat}$  and  $\lambda_{therm}$ , relative to the SIC RMSE of the *No-Phy* model. Cross markers indicate statistically significant reduction.

Table 2: Assessment of SIV and SIC predictions for PINN ( $\lambda_{therm} = 0.2$  and  $\lambda_{sat} = 0.2$ ) and *No-Phy* with different training sample sizes. The best accuracy is highlighted in bold.

Sample size		20 %			50 %			100 %		
SIV	Model	RMSE (km/d)	MAE (km/d)	ACC	RMSE (km/d)	MAE (km/d)	ACC	RMSE (km/d)	MAE (km/d)	ACC
	PINN	2.778	1.954	0.853	2.732	1.915	0.859	<b>2.629</b>	<b>1.844</b>	<b>0.867</b>
	No-Phy	2.873	2.053	0.843	2.759	1.936	0.856	2.684	1.880	0.863
SIC	Model	RMSE (%)	MAE (%)	ACC	RMSE (%)	MAE (%)	ACC	RMSE (%)	MAE (%)	ACC
	PINN	6.687	3.441	0.975	6.222	3.118	0.978	<b>5.854</b>	<b>2.917</b>	<b>0.981</b>
	No-Phy	7.393	4.335	0.969	6.611	3.533	0.975	6.197	3.274	0.978

SIC RMSEs are approximately 0.5 % and 0.3 % lower for the PINNs than the *No-Phy* models, respectively. Besides the RMSE comparison, the MAE and ACC are also improved by PINN (Table 2). These results suggest that the incorporation of physics-informed architecture and optimization is beneficial for SIC predictability under limited training sample scenarios. However, the weights to the physics loss function are not so significant for the SIC accuracy: no consistent trend in SIC RMSE is observed with increasing  $\lambda_{therm}$  or  $\lambda_{sat}$  values.

### 5.2. Temporal characteristics of model performance

Figure 3 shows the monthly RMSE of SIV prediction for seven test years (2016-2022) with different sample sizes. For the PINN model, we set  $\lambda_{sat} = 0.2$  and  $\lambda_{therm} = 0.2$  based on the optimal configuration identified in Figure 2. In general, SIV prediction shows relatively lower errors in summer (June-August) due to the slow drift speed during these months (Fig. A.7c). The improvement of SIV RMSE by PINN (i.e., RMSE difference between PINN and *No-Phy* model) is evident in most months when the sample size is 20

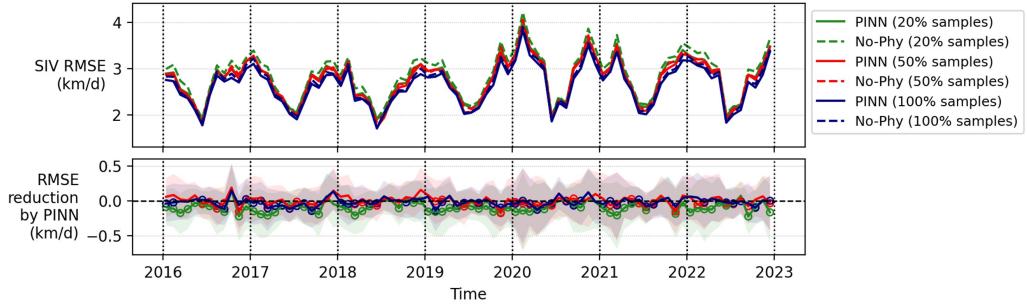


Figure 3: Monthly SIV RMSE of *No-Phy* model and PINN ( $\lambda_{sat} = 0.2$  and  $\lambda_{therm} = 0.2$ ) for seven test years (2016-2022). On the top panel, the solid lines indicate monthly SIV RMSEs of PINN models, and the dashed lines indicate *No-Phy* models. The bottom panel shows the differences between each PINN model and the *No-Phy* model. Statistically significant decreases in RMSE by PINN are highlighted as circles in the bottom panel. Statistical significance of each month is determined by conducting a t-test on all daily RMSE results from PINN and *No-Phy*. The shaded areas in the bottom panels show the 25-75 % quantile of RMSE difference.

%, and the RMSE reduction is relatively greater during the winter months (January to April) (Fig. 3). For the 50 % and 100 % sample sizes, no significant temporal trend in the SIV performance improvement is found.

Monthly RMSEs of SIC for test years are depicted in Figure 4. In contrast to the SIV results, SIC exhibits relatively higher errors during the summer months from June to July. While SIC remains consistently high from January to May, it decreases rapidly in these months (Fig. A.7d); the model could have difficulty capturing this trend. The most interesting finding is that SIC RMSE is highly dependent on training sample size: SIC RMSE decreases with a greater sample size. As a result, a more remarkable improvement in model accuracy by PINN is observed with fewer sample sizes. When the models are trained with 20 % of training samples, PINN reduces the SIC RMSE by >

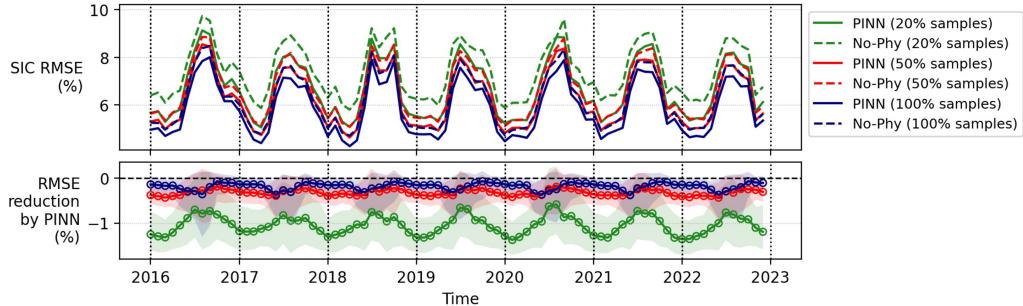


Figure 4: Monthly SIC RMSE of *No-Phy* model and PINN ( $\lambda_{sat} = 0.2$  and  $\lambda_{therm} = 0.2$ ) for seven test years (2016-2022). On the top panel, the solid lines indicate monthly SIC RMSEs of PINN models, and the dashed lines indicate *No-Phy* models. The bottom panel shows the differences between each PINN model and the *No-Phy* model. Statistically significant decreases in RMSE by PINN are highlighted as circles in the bottom panel. Statistical significance of each month is determined by conducting a t-test on all daily RMSE results from PINN and *No-Phy*. The shaded areas in the bottom panels show the 25-75 % quantile of RMSE difference.

0.5 % in all months, and such RMSE reduction is more significant in winter months from January to March.

### 5.3. Spatial characteristics of model performance

In this section, we examine where the PINN approach is more or less effective and discuss the implications of such spatial patterns. Herein, we discuss only the PINN with  $\lambda_{sat} = 0.2$  and  $\lambda_{therm} = 0.2$  configuration because (1) this PINN setting shows overall best RMSE in all sampling sizes and (2) the spatial patterns of PINN improvement appear similar for all different  $\lambda_{sat}$  and  $\lambda_{therm}$  values. The spatial distributions of SIV RMSE for PINN and *No-Phy* models with different sample sizes are shown in Figure 5. As previously discussed in Figures 2 and 3, the most substantial RMSE improvement occurs

with 20 % of training samples (Fig. 5h). When 20 % of training samples are used, the SIV RMSE improvement by PINN reaches up to 10 % near the central Arctic and northern Canadian Archipelago (Fig. 5g). When 50 % of training samples are used, RMSE reductions are observed across the central Arctic, but it is not statistically significant, with less than 5 % SIV improvement in most regions (Fig. 5h). Similarly, when 100 % samples are used to train the model, <5 % of SIV RMSE reduction is observed across the Arctic Ocean except central Arctic (Fig. 5i).

The spatial distribution of SIC RMSE improvement (Fig. 6) exhibits a different pattern from SIV. Above all, SIC RMSE improvement is observed across the Arctic Ocean. The PINN approach plays a particularly important role in improving SIC RMSE when only 20 % of training samples are used. When only 20 % of training samples are used to train the model, the baseline SIC RMSE of the *No-Phy* model exceeds 4 % across the Arctic Ocean (Fig. 6a). However, by adopting PINN, SIC RMSE decreases significantly by up to 50 % in the central Arctic (Fig. 6g). The improvement of SIC RMSE by PINN is attenuated when more training samples are used. Although SIC RMSE improvement is not significant with 50 % and 100 % sample experiments, the effectiveness of PINN still appears widespread over the central Arctic with up to ~20 % (Fig. 6h and i).

#### 5.4. Implications to future Arctic sea ice prediction

Given that many studies have shown the Arctic sea ice entered a new state in the 2010s with a dramatic loss of sea ice cover (Stroeve and Notz, 2018; Meier et al., 2014; Guemas et al., 2016; Döscher et al., 2014), historical satellite-based sea ice observations spanning over more than 40 years since

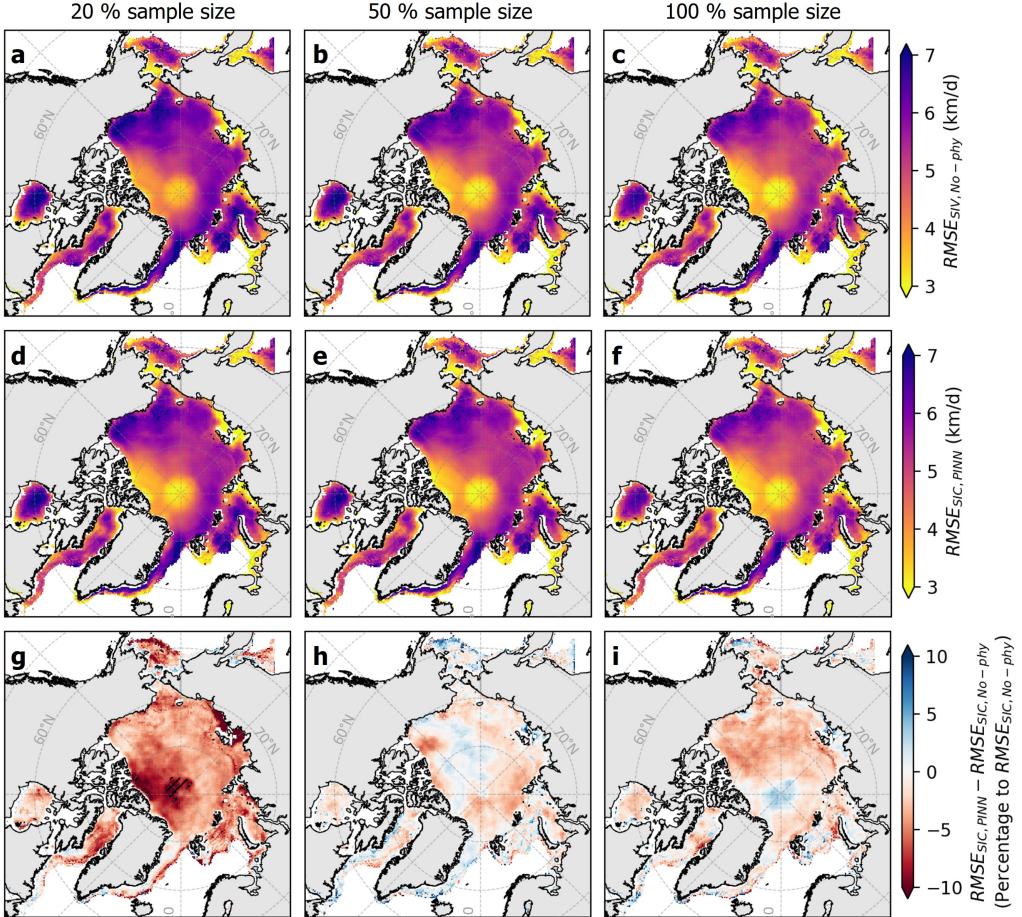


Figure 5: The SIV RMSE map of the No-Phy model for (a) 20 % training samples, (b) 50 % training samples, and (c) 100 % training samples. The SIV RMSE map of the PINN model ( $\lambda_{sat} = 0.2$  and  $\lambda_{therm} = 0.2$ ) for (d) 20 % training samples, (e) 50 % training samples, and (f) 100 % training samples. The relative SIV RMSE difference map between PINN and No-Phy model for (g) 20 % training samples, (h) 50 % training samples, and (i) 100 % training samples. On the bottom panels of the RMSE differences, the reddish color indicates the improvement in SIV RMSE by PINN over the No-Phy model (i.e., reduction in RMSE).

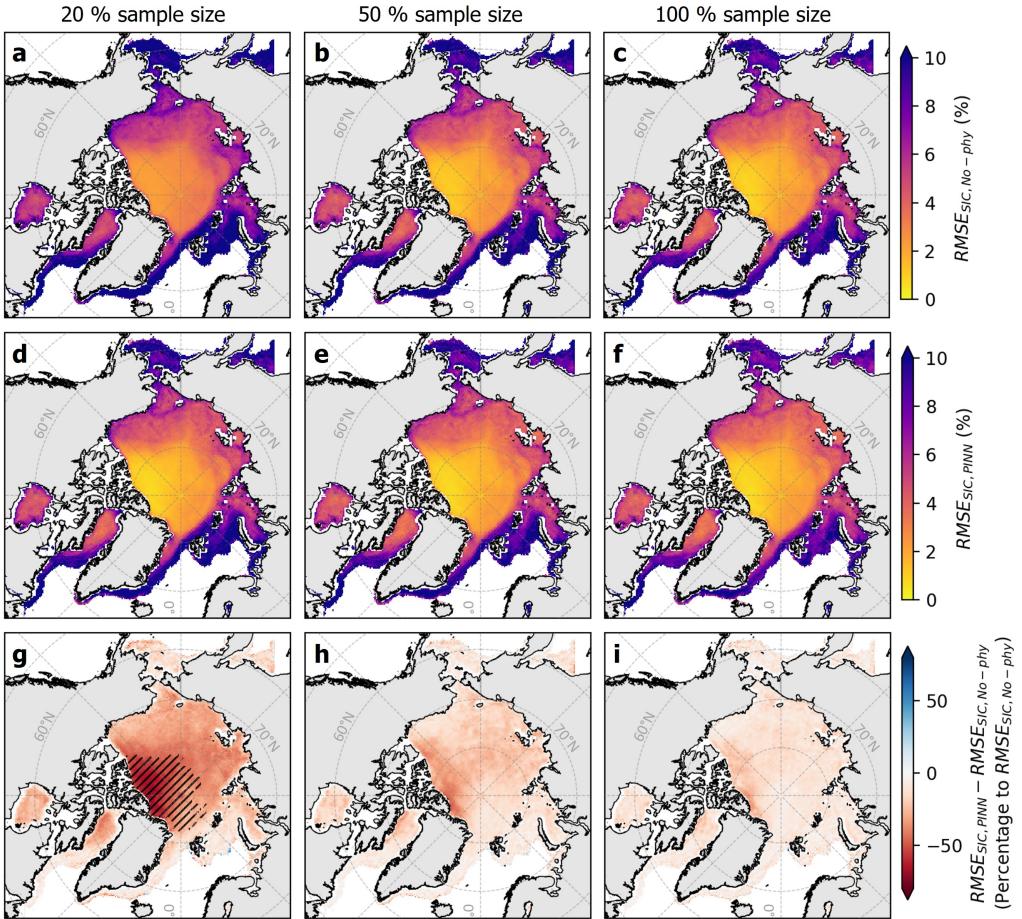


Figure 6: The SIC RMSE map of the No-Phy model for (a) 20 % training samples, (b) 50 % training samples, and (c) 100 % training samples. The SIC RMSE map of the PINN model ( $\lambda_{sat} = 0.2$  and  $\lambda_{therm} = 0.2$ ) for (d) 20 % training samples, (e) 50 % training samples, and (f) 100 % training samples. The relative SIC RMSE difference map between PINN and No-Phy model for (g) 20 % training samples, (h) 50 % training samples, and (i) 100 % training samples. On the bottom panels of the RMSE differences, the reddish color indicates the improvement in SIC RMSE by PINN over the No-Phy model (i.e., reduction in RMSE).

the 1980s may no longer fully represent the recent and future non-stationary sea ice state in the Arctic. Moreover, considering that the Arctic Ocean is projected to be nearly ice-free during summer as early as 2030-2040 (Massonnet et al., 2012; Overland and Wang, 2013; Árthun et al., 2021), the future sea ice dynamics in the Arctic Ocean will likely be different from those in historical observations. Hence, this raises concerns about the reliability of purely data-driven deep learning models trained solely on historical records, as such models may struggle to generalize to recent and future non-stationary sea ice conditions under rapid climate change. However, our PINN strategy provides a more robust alternative to fully data-driven learning by integrating physical principles directly into the learning process, thereby reducing reliance on historical data alone and enhancing model generalizability in a non-stationary climate regime.

In this study, we train our PINN models with 2009-2015 data and evaluate them with 2016-2022 data. Compared to the training seven years, the test seven years are characterized as fast-moving sea ice and lower SIC (Fig. A.7). Despite these differences, our PINN indeed shows significant improvement in both SIV and SIC RMSEs compared to the fully data-driven *No-Phy* model. We highlight that our PINN models maintain consistent performance regardless of the number of training samples, whereas the performance of *No-Phy* is dependent on the number of training samples (Figs. 5 and 6). This suggests the potential of the PINN framework to guarantee significant predictive performance and generalizability even when the training data do not fully represent unseen sea ice conditions. Moreover, considering that the significant improvements by PINN are observed near multi-year ice (MYI)

regions in the central Arctic and Canadian Archipelago (Figs. 5 and 6), our findings suggest that PINNs can contribute to obtaining consistent model performance despite the ongoing reduction of MYI coverage.

## 6. Conclusion

In this study, we propose a physics-informed learning strategy to improve the fidelity of the existing deep learning model for the prediction of daily sea ice velocity (SIV) and sea ice concentration (SIC) retrieved from spaceborne remote sensing data. Our physics-informed learning strategy is achieved through two pathways: (1) design additional physics-informed loss functions that regularize SIV values based on SIC values and regularize SIC values based on the assumption of daily thermodynamic ice growth and melting; (2) insert sigmoid activation function to restrict the output SIC values into the range of 0 to 1. We implement extensive experiments by employing the Hierarchical information-sharing U-net (HIS-Unet), which predicts SIV and SIC through a series of information sharing between SIV and SIC branches. In order to investigate the impact of training samples on the model robustness, we train the physics-informed neural network (PINN) and normal data-driven neural network (*No-Phy*) separately with three different training sample ratios (100 %, 50 %, and 20 %) and four different weights to the physics-informed loss terms (0.0, 0.2, 1.0, and 0.5).

The results exhibit that the physics-informed learning strategy improves both SIC and SIV predictions. In particular, the improvement of SIC prediction by PINN is more obvious than SIV with a small number of samples. Physics-informed learning strategy helps achieve the benefits of HIS-Unet

even in the case of a lack of sufficient training datasets. Therefore, this physics-informed learning strategy can contribute to improving the performance of deep learning models for sea ice prediction, even if the past sea ice data cannot fully represent the future non-stationary sea ice conditions affected by rapid climate change. Additionally, this strategy can also be easily applied for a multi-day forecast of sea ice conditions combined with recurrent network architectures, such as LSTM.

### **Software and data availability**

The codes that were used for the prediction of sea ice concentration and velocity using Python language (version 3.11) based on PyTorch library can be found in Github: [https://github.com/BinaLab/PINN\\_seaice](https://github.com/BinaLab/PINN_seaice). This repository was created by Younghyun Koo (e-mail: kooala317@gmail.com) in 2023 and contains program codes (111 MB) and training data (3.41 GB). The authors' experimental environment was as follows:

- OS: Windows 11 Pro
- CPU: Intel(R) Core(TM) i7-11700F
- RAM: 16.0 GB
- GPU: NVIDIA GeForce RTX 3070

The sea ice velocity and concentration data can be downloaded free of charge from NSIDC (Tschudi et al., 2019; Meier et al., 2021), and the ERA5 weather data can be downloaded free of charge from the Copernicus Climate Data Store (Hersbach et al., 2020).

## CRediT authorship contribution statement

**Younghyun Koo:** Software, Validation, Formal analysis, Data curation, Visualization, Writing - Original Draft. **Maryam Rahnemoonfar:** Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding acquisition.

## Acknowledgments

This work is supported by the U.S. National Science Foundation (NSF) BIGDATA (IIS-1838230, IIS-2308649) and NSF Leadership Class Computing (OAC-2139536) awards.

## Appendix A. Arctic SIC and SIV conditions from 2016 to 2022

Figures A.7a and b show the average SIV and SIC, respectively, over the Arctic Ocean in 2022. The Fram Strait in eastern Greenland shows the fastest ice movement. Figures A.7c and d show the monthly mean SIV and SIC, respectively, in 2022. SIV decreased from January to June and increased again thereafter. SIC was consistent to 90 % from January to May, decreased to ~70 % until September, and increased again from October.

## References

Andersson, T.R., Hosking, J.S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D.C., Wilkinson, J., Phillips, T., Byrne, J., Tietsche, S., Sarojini, B.B., Blanchard-Wrigglesworth, E., Aksenov, Y., Downie, R., Shuckburgh, E., 2021. Seasonal arctic sea ice forecasting with probabilistic deep learning. *Nature Communications* 12.