

Article

STDNet: Spatio-Temporal Decompose Network for Predicting Arctic Sea Ice Concentration

Xu Zhu ^{1,2}, Jing Wang ^{1,2}, Guojun Wang ¹, Yangming Jiang ¹, Yi Sun ²  and Huihui Zhao ^{1,*} 

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; 107552203554@stu.xju.edu.cn (X.Z.); 107552203520@stu.xju.edu.cn (J.W.); wanggj@aircas.ac.cn (G.W.); jiangym@aircas.ac.cn (Y.J.)

² College of Mathematics and System Science, Xinjiang University, Urumqi 830017, China; brian@xju.edu.cn

* Correspondence: zhh@aircas.ac.cn

Abstract: In the context of global warming, the accurate prediction of Arctic Sea Ice Concentration (SIC) is crucial for the development of Arctic shipping routes. We have therefore constructed a lightweight, non-recursive spatio-temporal prediction model, the Spatio-Temporal Decomposition Network (STDNet), to predict the daily SIC in the Arctic. The model is based on the Seasonal and Trend decomposition using Loess (STL) decomposition idea to decompose the model into trend and seasonal components. In addition, we have designed the Global Sparse Attention Module (GSAM) to help the model extract global information. STDNet not only extracts seasonal signals and trend information with periodical correspondence from the data but also obtains the spatio-temporal dependence features in the data. The experimental methodology involves predicting the next 10 days based on the first 10 days of data. The prediction results provided the following metrics for the 10-day forecast of STDNet: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and coefficient of determination of 1.988%, 3.541%, 5.843%, and 0.979, respectively. The average Binary Accuracy (BACC) at the beginning of September for the period 2018–2022 reached 93.85%. The proposed STDNet model outperforms and is lighter than existing deep-learning-based SIC prediction models.

Keywords: non-recurrent spatio-temporal prediction; STL decomposition; sea ice concentration (SIC); spatio-temporal attention



Citation: Zhu, X.; Wang, J.; Wang, G.; Jiang, Y.; Sun, Y.; Zhao, H. STDNet: Spatio-Temporal Decompose Network for Predicting Arctic Sea Ice Concentration. *Remote Sens.* **2024**, *16*, 4534. <https://doi.org/10.3390/rs16234534>

Academic Editor: Yi Luo

Received: 16 October 2024

Revised: 28 November 2024

Accepted: 30 November 2024

Published: 3 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, Arctic sea ice has been melting at an unprecedented rate due to the effects of global warming and Arctic amplification. According to the latest IPCC-AR6 report, Arctic sea ice coverage shows an annual decreasing trend, and it is highly likely that the entire Arctic sea will be ice-free during summer prior to 2050 [1]. Compared to the early 1980s, both the area and thickness of sea ice have decreased by more than 50% [2,3], and the melting trend of sea ice in the Arctic is almost irreversible. Rising temperatures are melting the ice caps, and the positive feedback effect of ice cap melting is exacerbating the phenomenon of global warming [4].

As a result of this situation and its impact on the climate, the navigational potential of the Arctic channels has been continuously enhanced, gradually making navigation by ships in polar channels a reality. Arctic channels have enormous economic and social value, and they have the potential to become an important economic corridor connecting Europe, North America, and Asia. However, this cannot yet be safely actualized due to the complexity of sea ice changes and the uncertainty of sea ice forecasts, both of which cause navigational safety risks. Therefore, predicting daily changes in Arctic sea ice, and in particular predicting the Sea Ice Concentration (SIC), is of great significance, not only for providing a basis for subsequent research but also for improving the exploration and research of Arctic routes and ensuring navigational safety.

In order to better predict sea ice trends, researchers have utilized knowledge from different disciplines to study sea ice changes. For instance, numerical sea ice models have been developed based on the dynamics, thermodynamics, and thickness distribution of sea ice from known physical laws [5–13]. Common numerical models include the MIT General Circulation Model (MITgcm) [6], the Hybrid Coordinate Ocean Model (HYCOM) [7], the Modular Ocean Model (MOM6) [8], the Los Alamos sea ice model (CICE) [9], the Arctic Cap Nowcast/Forecast System (ACNFS) [12], and the Global Ice Ocean Prediction System (GIOPS) [13]. However, many of the parameters employed by such physical models are too complex [14,15] and the computational cost of running the model is also high. The rapid changes in polar sea ice and its physical properties have increased the uncertainty of the physical parameterization scheme and the prediction results of sea ice models [16–18]. In addition, certain phenomena in nature cannot be explained by clear physical laws; therefore, predicting sea ice changes using this method may lead to inconsistencies between the predicted results and actual observations [19].

Statistical models make predictions by establishing relationships between the SIC and atmospheric conditions, oceanic conditions, and sea surface variables. The Vector Autoregressive Model (VAR) [20], linear Markov model [21], and vector Markov model [22] are typical representatives. Unlike numerical modeling, statistical forecasting is data-driven; it starts from the data itself and uses statistical methods to mine and fit out changing sea ice patterns. When confronted with complex factors and unclear impact rules, statistical forecasting can, to a certain extent, compensate for the shortcomings of numerical model forecasting.

Studies predicting sea ice changes involve deterministic spatio-temporal sequence predictions. In recent years, deep-learning (DL) models have provided superior spatio-temporal predictions [23–36], and this has advanced predictions of the direction of sea ice change [37–52]. Deep learning has also become another effective method for predicting sea ice conditions. Compared with methods predicting sea ice changes from known physical laws, deep-learning algorithms can extract patterns of change from data and simulate complex dynamic systems or nonlinear changes. Deep-learning algorithms not only agree with people's knowledge of natural laws but can also achieve or even exceed the simulation effects of numerical models [37]. Many algorithms using deep-learning models have been proposed to predict sea ice changes. For example, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been proposed for predicting sea ice conditions [38–42]. The SIC is one of the most commonly predicted sea ice variables, and both short-term [43–45] and medium- to long-term predictions have been made [46,47].

However, deep-learning models have many limitations when predicting sea ice: firstly, there is a dependence on time during the model prediction process. Taking the Convolutional LSTM Network (ConvLSTM) [26] as an example, although the model integrates the advantages of CNNs and the Long Short-Term Memory (LSTM) recurrent neural network, which can simultaneously process information in various spatio-temporal dimensions, the model has certain defects. Similar to traditional temporal recurrent networks, ConvLSTM has a time-dependence problem: when the time series of the input data is too long, it forgets the historical information, leading to gradient vanishing or exploding and causing failure of model convergence [29]. In addition, short-term predictions of sea ice concentration require the input of a time series spanning 5, 10, or more days, as well as the consideration of information from multiple historical moments, which undoubtedly exacerbates the model's shortcomings. Therefore, to achieve the short-term prediction of sea ice concentration, it is necessary to consider and improve the issues of existing models to overcome time-dependent problems.

Secondly, model complexity and computational resource consumption are pressing issues that need to be addressed. Many advanced spatio-temporal prediction models have been proposed, often combining auxiliary inputs, complex neural network architectures, and sophisticated training strategies. While these methods demonstrate excellent performance, the increasing complexity of these systems undoubtedly limits their usability and scalability in practical applications [51]. For example, the Two-Stream Convolutional Long

Short-Term Memory (TS-ConvLSTM) model [51] combines two different modes of ConvLSTM to learn sea ice dynamic features, achieving significantly improved performance compared to the original ConvLSTM model. However, this improvement comes at the cost of a substantial increase in model complexity, leading to higher computational expenses. Additionally, due to the recursive nature of Recurrent Neural Networks (RNNs), data must be processed sequentially, which not only limits parallel computing capabilities but also significantly increases training and inference times.

Beyond computational complexity, existing spatio-temporal prediction models face additional challenges. For instance, complex models typically require high-performance hardware and ample storage resources, which are often difficult to provide in resource-constrained scenarios such as real-time prediction or embedded systems. Furthermore, while these models may perform well on specific datasets, they often lack robustness and are prone to overfitting when applied to different regions or climatic conditions, making it difficult to generalize to other scenarios. Additionally, the intricate architectures of complex models may lead to increased development and maintenance costs, creating technical and economic barriers for practical applications.

To address these issues, the design of lightweight networks has become essential. Lightweight networks can reduce model parameters and computational complexity while maintaining high predictive accuracy, thereby improving the usability and scalability of the models. In scenarios such as Arctic sea ice prediction, lightweight design is particularly crucial as these tasks demand extremely high real-time performance. Models need to perform inference quickly under limited computational resources while maintaining high predictive accuracy. Therefore, lightweight design is the key pathway to resolving the challenges of existing spatio-temporal prediction models, achieving a better balance between computational efficiency, predictive accuracy, and generalization capability, and providing more reliable support for practical applications.

Thirdly, traditional spatio-temporal prediction models often lack mechanisms to independently capture seasonal and trend features, making it challenging to interpret the dynamic characteristics of the data. This limitation is particularly critical for key tasks such as Arctic sea ice prediction. Currently, the utilization of seasonal and trend features in SIC data is insufficient. The mean SIC values exhibit obvious seasonal fluctuations, indicating that large-scale SIC evolution is mainly influenced by seasons [48]; in other words, SIC has obvious seasonality. In addition, the tendency of SIC to decline rapidly or rise sharply during the ice-melt and freeze-up periods and the fluctuations in environmental factors such as temperature, wind speed, ocean currents, and weather events, which can also lead to drastic changes, reflect the existence of a clear trend in SIC. In recent years, many studies have investigated how to use CNNs [39] and RNNs [42,43] to process SIC data, and some have discussed how to effectively utilize periodic information to improve the accuracy of predictions. However, deep neural networks suffer from gradient vanishing and error accumulation, which makes it difficult to capture long time dependencies in time-series data, and few studies have effectively modelled this trend or seasonality signals using deep neural networks. Therefore, when using deep learning to make short-term sea ice concentration predictions, it is necessary to consider how to effectively extract and utilize the seasonal and trend features of SIC data.

Finally, there is a need to further improve the accuracy of sea ice change predictions. For the Arctic region, spatial and temporal dependencies exist at different spatial scales. For example, the SIC changes little in the central Arctic, but there are large SIC differences at the Arctic margins. Therefore, the degree of spatio-temporal dependence also varies by region. The predefined fixed-size window makes it impossible for CNN-based models to capture multiscale dependencies. This problem limits the prediction accuracy of the current DL-based model and requires model improvement to further improve the prediction accuracy.

To address the above problems, we designed and developed a lightweight acyclic network architecture, STDNet, which comprises a basic neural network (CNN), and we

also use a Global Sparse Attention Module (GSAM), which is trained in an end-to-end manner by Mean Squared Error (MSE) loss.

The main contributions of this study are as follows:

- We proposed a lightweight non-recurrent spatio-temporal prediction model, STDNet, which innovatively decomposes data into seasonal and trend components and processes them using independent networks in parallel. Specifically, the Season network focuses on capturing the periodic seasonal signals in the data, while the Trend network targets short-term trend variations. This decomposition strategy enables STDNet to effectively extract seasonal and trend features from spatio-temporal data, significantly reducing computational costs through parallel processing while maintaining high predictive accuracy for both short-term and long-term variations. This approach overcomes the high computational complexity and scalability issues of traditional recurrent models.
- In the model design, we introduced a novel attention mechanism called the Global Sparse Attention Module. This module is designed to efficiently capture the global spatio-temporal dependencies in the data, addressing the limitations of traditional attention mechanisms that primarily focus on local information. Specifically, the module employs a sparse attention mechanism to minimize attention to irrelevant information, thereby improving computational efficiency and reducing memory usage. Meanwhile, it extracts long-range dependencies from a global perspective, enhancing the model's ability to capture complex spatio-temporal patterns. The Global Sparse Attention Module is not only simple and efficient but also highly flexible, making it applicable to a wide range of spatio-temporal prediction tasks and offering a new technical approach for future research.
- STDNet demonstrated outstanding performance in recursive prediction tasks, achieving a Mean Absolute Error (MAE) of less than 2% (1.988%) in predicting 10-day sea ice concentration (SIC) and an average Binary Accuracy (BACC) of 93.85% between 8 September and 17 September 2018–2022. To verify the model's effectiveness and generalizability, experiments were conducted on datasets for Arctic sea ice concentration, Arctic sea ice drift, and Arctic sea ice thickness. Comparative analyses with other spatio-temporal prediction models revealed that STDNet consistently outperformed existing models across multiple evaluation metrics for all datasets. Furthermore, STDNet demonstrated significantly lower model parameter counts, computational costs (FLOPs), and training time compared to existing models, achieving higher computational efficiency and resource utilization while maintaining high accuracy. These results demonstrate that STDNet not only excels in the accuracy and lightweight nature of sea ice dynamics prediction but also possesses broad applicability, making it an efficient tool for predicting various polar sea ice characteristics.

2. Materials and Methods

2.1. STDNet Overview

2.1.1. Spatio-Temporal Sequence Prediction Task

SIC prediction is a standard spatio-temporal sequence prediction task, which aims to predict the values of multiple future time steps based on historical time-series data. The definition of spatio-temporal sequence prediction is as follows: Assume a monitoring system collects data at fixed time intervals, generating c measurements at each time step. These measurements can be represented as a 2D image of size $w \times h$, where w and h denote the width and height of the image, respectively, and c represents the number of channels. Considering the temporal dimension, these measurements form a 3D tensor $X \in \mathbb{R}^{w \times h \times c}$.

Let the current time step be i , and there are J time steps of historical data from the past up to the current time. This historical time series is defined as $X_{i-J+1:i} = \{X_{i-J+1}, \dots, X_i\}$. The goal is to use this sequence to predict the data for the next K time steps, denoted as $X_{i+1:i+K} = \{X_{i+1}, \dots, X_{i+K}\}$. To formalize this prediction problem, the prediction of future

time steps can be modeled as maximizing the conditional probability $p(X_{i+1:i+K}|X_{i-J+1:i})$, as follows:

$$\hat{X}_{i+1:i+K} = \underset{X_{i+1}, \dots, X_{i+K}}{\operatorname{argmax}} p(X_{i+1:i+K}|X_{i-J+1:i}) \quad (1)$$

To simplify the prediction process, the joint conditional probability for future time steps can be decomposed using the chain rule as:

$$\hat{X}_{i+1:i+K} = \underset{X_{i+1}, \dots, X_{i+K}}{\operatorname{argmax}} \left(p(X_{i+1}|X_{i-J+1:i}) \prod_{n=2}^K p(X_{i+n}|X_{i+n-1}) \right) \quad (2)$$

The above equations indicate that spatio-temporal sequence prediction involves using historical data, $X_{i-J+1:i}$, to infer future time steps, $X_{i+1:i+K}$. This process includes both temporal dependencies and spatial pattern extraction. The complexity of this task arises from the interaction between the temporal and spatial dimensions, requiring models to effectively capture long-term dynamic dependencies as well as localized spatial features.

2.1.2. Decomposition Structure

Figure 1 illustrates the overall framework of STDNet, highlighting its core approach to solving complex long-term forecasting problems through the decomposition of time-series features. STDNet adopts a decomposition strategy [53–55], which divides the complex patterns in the input time series into trend and seasonal components, thereby simplifying the modeling of intricate temporal dynamics. Inspired by the Seasonal-Trend Decomposition using Loess (STL) method [53], STDNet captures the local temporal structures of time-series data, where the trend component reflects gradual changes over shorter time spans and the seasonal component represents periodic fluctuations or recurring patterns.

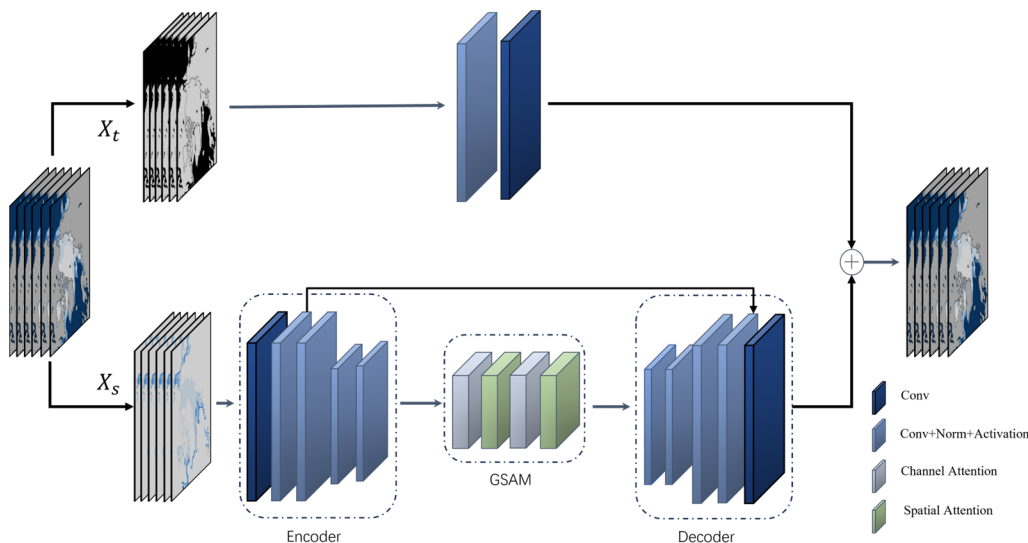


Figure 1. STD framework diagram.

For an input time series $X \in \mathbb{R}^{L \times d}$, where L is the sequence length and d is the feature dimension, the decomposition process involves two key steps. First, the trend component, $X_t \in \mathbb{R}^{L \times d}$, is extracted using a moving average method, $\text{AvgPool}(\cdot)$, which approximates the smoothing operation in STL decomposition. This can be expressed as:

$$X_t = \text{AvgPool}(X) \quad (3)$$

Next, the seasonal component $X_s \in \mathbb{R}^{L \times d}$ is obtained by subtracting the trend component from the original series:

$$X_s = X - X_t \quad (4)$$

The moving average operation, $AvgPool(\cdot)$, serves as a lightweight approximation of the smoothing process in STL decomposition. By iteratively averaging adjacent time points, it isolates the local temporal dynamics while preserving the overall structure of the time series. With a step size of 1, this process generates $L - 1$ trend values due to boundary constraints. To maintain a consistent sequence length, the final value of the trend component is extrapolated as follows:

$$X_t^{(L)} = X_t^{(L-1)} \quad (5)$$

This adjustment ensures that the trend component X_t has the same length as the original sequence X , while the seasonal component X_s captures the residual periodic variations after the trend is removed.

By employing this decomposition strategy, STDNet models trend and seasonal variations separately, embodying a “divide-and-conquer” approach to temporal feature modeling. This method not only reduces the complexity of directly modeling the entire time series but also enhances the model’s performance in capturing short-term dynamics and periodic variations. Additionally, it retains the core principles of STL decomposition, further improving the interpretability and applicability of the model. Through the specialized modeling of different temporal characteristics, STDNet effectively addresses complex spatio-temporal prediction tasks, excelling particularly in scenarios that require precise short-term forecasting and sensitivity to seasonal fluctuations.

2.2. Input

Given the input of a batch of past frames, $G \in \mathbb{R}^{B \times T \times C \times H \times W}$, with batch size B (where B is the number of batches, T is the number of frames, C is the number of channels, and $H \times W$ is the size of the image), which are decomposed to obtain the trend component and the seasonal component, respectively, the tensors with tensor reshape, $(B \times T) \times C \times H \times W$, were input into the Trend network and Season network, respectively, as shown in Figure 2a. The inputs to both networks treat each frame as an individual sample and focus on single-frame level features without considering temporal changes. In the GSAM, the hidden representations of the encoder output were reshaped into tensors of shape $B \times (T \times C) \times H \times W$, and multiple frame-level features were stacked along the time axis, as shown in Figure 2b. By forcing channel attention and spatial attention modules designed based on convolutional networks to learn from the multi-frame feature stack, the model can capture the inherent spatio-temporal evolution within the sequence data.

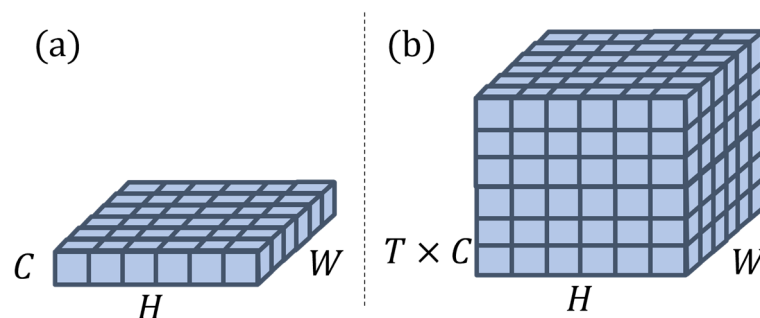


Figure 2. Input characterization: (a) Encoder/decoder input characterization, (b) GSAM input characterization.

2.3. Trend Network

Figure 3 shows the overall flowchart of STDNet. To predict the trend component, a Trend network (Figure 3a) with N_T convolutional layers with a fixed hidden size, n_T , was constructed to infer the evolutionary trend directly from the historical trend component.

The structure of the Trend network is relatively simple, and its hidden representation can be described as follows:

$$X_T^i = \sigma \left(\text{GroupNorm} \left(\text{Conv}2d_{3 \times 3} \left(X_T^{i-1} \right) \right) \right), 1 \leq i \leq N_T \quad (6)$$

$$X_T^{\text{out}} = \text{Conv}2d_{1 \times 1} \left(X_T^{N_T} \right) \quad (7)$$

where σ are the nonlinear activation functions, in which the LeakyReLU activation function was used, and GroupNorm is the normalization layer. Conv2d_{3×3} is the convolutional layer with convolutional kernel 3, Conv2d_{1×1} is the convolutional layer with convolutional kernel 1, and X_T is the trend component of the input.

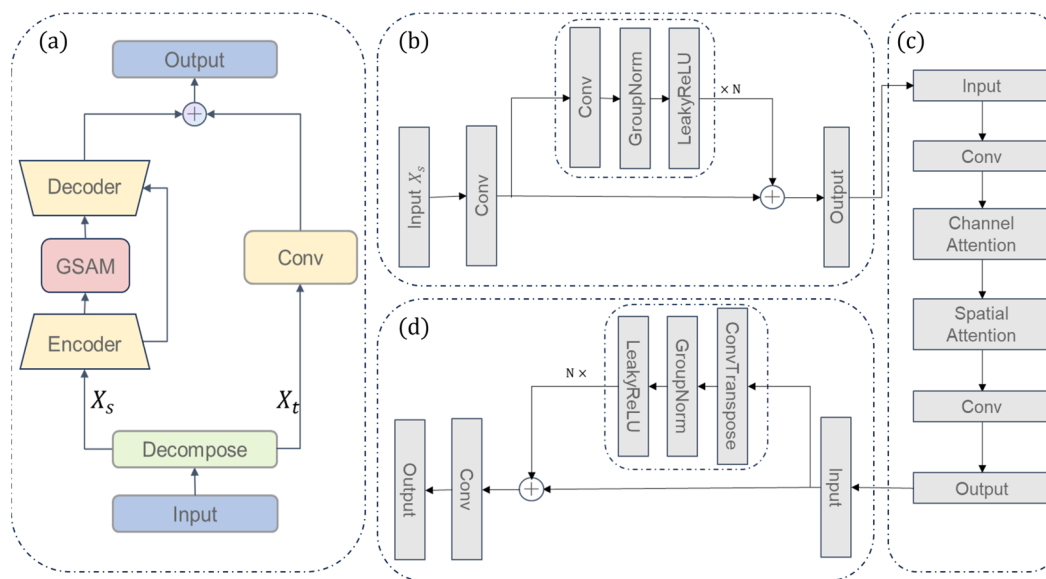


Figure 3. STDNet overall process: (a) design details, (b) Encoder, (c) GSAM, and (d) Decoder.

The design of the Trend network is intentionally simple, based on the characteristics of the trend component. The trend component, derived from the time-series decomposition, is inherently smooth and low-frequency in nature, exhibiting strong continuity and directional properties. As such, complex deep networks are unnecessary for capturing its features. By employing 3×3 convolutional kernels, the Trend network effectively captures local trend variations and progressively expands its receptive field through stacked convolutional layers, thereby accommodating long-term dependencies. The final 1×1 convolutional layer serves as a dimension reduction and feature integration step, mapping multi-channel intermediate representations to the final output.

This structural design not only meets the requirements for trend component modeling but also reduces the computational complexity of the model. Compared to using more complex network architectures, the simple Trend network can quickly capture trend features, improve training efficiency, and reduce the risk of overfitting on new data. This lightweight design of the Trend network reflects STDNet's targeted modeling of different temporal characteristics, achieving both efficiency and robust predictive performance.

2.4. Season Network

To predict the seasonal components, a Season network with an encoder–decoder structure was constructed, which includes an encoder (Figure 3b), a GSAM (Figure 3c), and a decoder (Figure 3d). The Season network utilizes an encoder to encode high-dimensional past frames into a low-dimensional potential space from which GSAM learns spatial dependencies and temporal variations. The decoder ultimately decodes the latent spatio-

temporal information into predicted future frames. The design logic of the Season network can be detailed using the following formulas:

First, the encoder is tasked with mapping high-dimensional historical seasonal data, X_S , into a low-dimensional latent space to extract its key features. The encoder consists of N_S convolutional layers with a fixed hidden size of n_S . The initial input, X_S , is first passed through a 1×1 convolutional layer for channel transformation:

$$X_S^1 = \text{Conv}2d_{1 \times 1}(X_S) \quad (8)$$

Then, it undergoes N_S convolution operations, with the output of each layer, X_S^i , expressed as:

$$X_S^i = \sigma\left(\text{GroupNorm}\left(\text{Conv}2d_{3 \times 3}\left(X_S^{i-1}\right)\right)\right), 1 \leq i \leq N_S \quad (9)$$

Here, σ is the nonlinear activation function, which is LeakyReLU in this design. To reduce the spatial dimensions of the data, downsampling is performed after every two convolutional layers. If downsampling is required, a convolutional layer with a stride of 2 is used; otherwise, a stride of 1 is maintained to preserve the size of the feature maps.

Next, the output of the encoder, $X_S^{N_S}$, enters the Global Sparse Attention Module (GSAM). The GSAM consists of N_C spatio-temporal attention modules, which are used to capture global spatial dependencies and temporal variation patterns. This can be represented as:

$$X_S^j = \text{GSAMModule}\left(X_S^{j-1}\right), N_S < j \leq N_S + N_C \quad (10)$$

Here, X_S^{j-1} is the input to the GSAM, which represents the latent representation output by the last layer of the encoder. The GSAM efficiently captures long-range spatial and temporal dependencies and significantly reduces computational complexity through sparse attention mechanisms while retaining critical data patterns.

After the GSAM processes the latent representation, $X_S^{N_S+N_C}$, the output enters the decoder for inverse restoration. The decoder comprises N_S transposed convolutional layers that gradually restore the spatial resolution of the feature maps. The output of each decoder layer can be expressed as:

$$X_S^k = \sigma\left(\text{GroupNorm}\left(\text{ConvTranspose}2d\left(X_S^{k-1}\right)\right)\right), N_S + N_C < k \leq 2N_S + N_C \quad (11)$$

Finally, the decoded seasonal output, X_S^{out} , is generated through a 1×1 convolutional layer:

$$X_S^{out} = \text{Conv}2d_{1 \times 1}\left(X_S^{2N_S+N_C}\right) \quad (12)$$

In the decoder, upsampling is performed after every two transposed convolutional layers. If upsampling is required, a transposed convolutional layer is used; otherwise, a stride of 1 is maintained to preserve the feature map size. Additionally, the Season network incorporates residual connections, passing the output of the first layer of the encoder directly to the last layer of the decoder to retain the original spatial features of the input data and enhance the spatial fidelity of the decoded output.

This structure of the Season network fully considers the essential characteristics of seasonal components, namely periodicity and complex spatio-temporal interactions. It combines the efficient representation capability of the encoder–decoder structure with the powerful spatio-temporal dependency modeling capability of the GSAM. The encoder extracts multi-level features and reduces spatial dimensions through convolutional layers, the GSAM captures complex spatio-temporal dependencies in the latent space, and the decoder gradually restores the resolution of the feature maps to generate high-quality predictions. This design effectively captures the periodic and spatial characteristics of seasonal components while optimizing computational resources. This targeted design

enables the Season network to perform exceptionally well in complex spatio-temporal prediction tasks, achieving a balance between high accuracy and efficiency in seasonal component modeling.

2.5. GSAM

The attention mechanism is an important technological innovation in the field of deep learning that aims to improve the efficiency and performance of a model in processing input data. Inspired by the human brain's selective attention to external information, this mechanism enables models to concentrate more on the most significant parts of the input data. In the attention mechanism, the model dynamically adjusts its attention to different input elements by learning a set of weights, thus enabling it to focus on information relevant to the current task. Unlike traditional neural networks, the attention mechanism allows the model to selectively focus on different parts of the input data to capture key information more efficiently. This flexibility allows the model to be adapted to the processing of different tasks and variable-length sequences. In addition, by selectively focusing on certain parts of the input sequence, the attention mechanism reduces the global processing of the entire input, thus enhancing computational efficiency and reducing computational costs. This is crucial for training large-scale data and complex models. Overall, the attention mechanism has driven the development and application of deep-learning models in several domains through its advantages in flexibility, performance enhancement, and adaptation to sequence processing. Its successful application in tasks such as natural language processing and computer vision has made the attention mechanism one of the core technologies in deep-learning research and applications.

The Convolutional Block Attention Module (CBAM) is a simple and effective feed-forward convolutional neural network attention module [56]. The CBAM sequentially infers the attention map along the channel and the spatial dimensions of given intermediate feature maps. The attention map is then multiplied with the input feature map for adaptive feature refinement. As CBAM is a lightweight and universal module, it can be seamlessly integrated into any CNN architecture with negligible overheads and trained end-to-end with the underlying CNN. However, due to reduced information and dimensional separation, the CBAM approach utilizes only a limited visual representation of the receptive field, ignoring the importance of preserving both channel and spatial aspects to enhance cross-dimensional interactions.

Based on this, the Global Attention Mechanism (GAM) enhances the performance of deep neural networks by reducing information approximation and amplifying global interaction representation [57]. Based on the convolutional spatial attention submodule, a 3D multilayer perceptron-based alignment channel attention submodule was proposed. The results showed that although the GAM improves the prediction accuracy, due to the removal of the pooling layer that originally reduced the dimension, this costless operation leads to a surge in computational demand and a significant increase in computational costs.

To amplify the cross-dimensional interactions while reducing the computational cost, this study improves on original work and proposes the GSAM, which is divided into two parts: the channel attention module (Figure 4a) and the spatial attention module (Figure 4b).

The channel attention module compresses the feature map in the spatial dimension to produce a channel attention map. For a single image, channel attention focuses on which elements of the image are most important. Initially, 3D substitution is used to preserve the 3D information. Subsequently, to effectively calculate channel attention, a Sparse Multilayer Perceptron (sMLP) [58] is utilized to compress the spatial dimensions of the input feature map and amplify the cross-dimensional channel-space dependencies. Finally, the output and input feature maps of this module are multiplied and replaced as input dimensions into the spatial attention module.

Similarly, the Spatial Attention Module (SAM) compresses channels and focuses on the location of most of the informative parts of an image, thus complementing channel attention. The feature maps output from the channel attention module are used as the input

feature maps for this module, and the spatial information is fused using two convolutional layers. A reduction ratio is set to r to compress the channels, reducing the dimension to C/r channels. Finally, this feature map is multiplied with the input feature map of this module to obtain the final generated feature map.

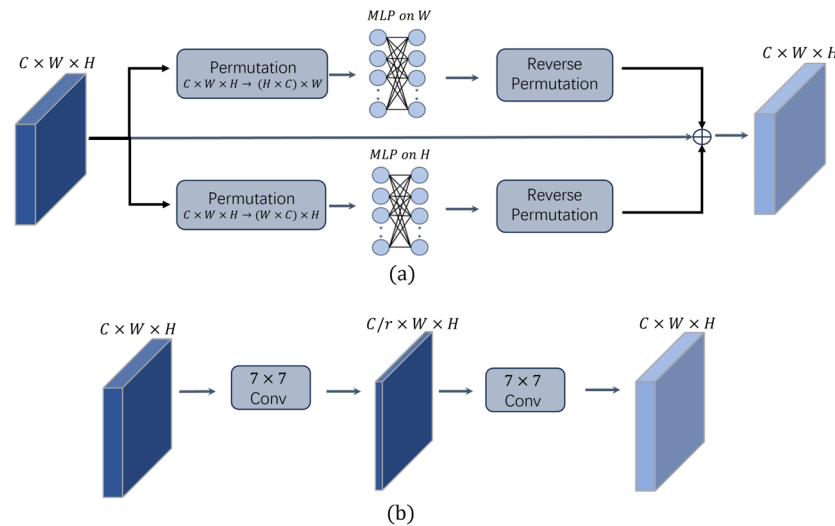


Figure 4. GSAM: (a) channel attention module, (b) spatial attention module.

Unlike previous attention mechanisms, the traditional MLP (Figure 5a) is replaced by sMLP in this study, as this addresses two major drawbacks of the traditional MLP. In this respect, we aimed to reduce the number of parameters to avoid overfitting and sought to reduce computational complexity (especially with a large number of input cells) to achieve multi-stage processing in a pyramid structure.

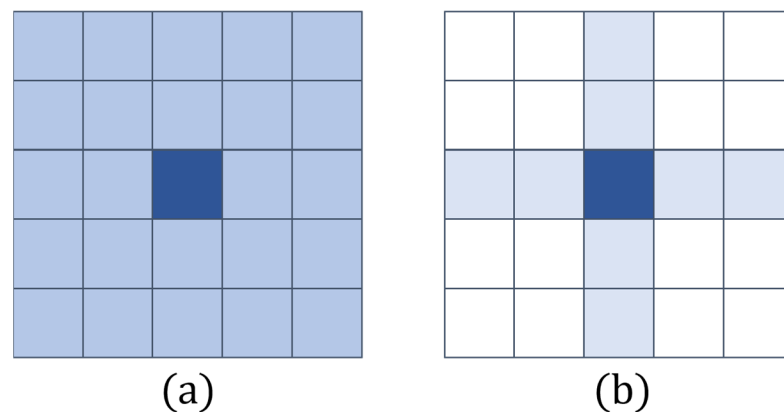


Figure 5. The sparse MLP reduces the computational complexity of the MLP through sparse connections and weight sharing. In (a) MLP, the dark blue token interacts with all the light blue tokens in a single MLP layer. In contrast, in one (b) sMLP layer, the dark blue token only interacts with the light blue tokens in the horizontal and vertical directions.

In the sMLP, sparse connections and weight sharing are employed to achieve the design objectives. As shown in Figure 5b, cells in the sMLP interact directly only with cells on the same row or column, rather than with all other cells. Furthermore, all rows and columns share the same projection weights, respectively. Figure 5b demonstrates the implementation of the sMLP module, which consists of three pathways. In addition to the identity map shown in the middle, the other two paths are responsible for mixing cells along the horizontal and vertical directions, respectively. Using the sMLP not only efficiently aggregates spatial dimensions but also reduces the amount of computation.

3. Experiments

In this section, we report a series of experiments conducted to evaluate the performance of STDNet. Section 3.1. describes the dataset used for the experiments, along with preprocessing details. Details about implementing the experiment and evaluation criteria are then described in Section 3.2. Subsequently, to validate the effectiveness of STDNet, the evaluation performance is demonstrated in Section 3.3., which includes quantitative and qualitative comparisons with other existing methods, while exploring the issue of computational efficiency. Finally, in Section 3.4., ablation studies are provided to evaluate the key components of STDNet and the combination of each loss function.

3.1. Dataset

Currently, the large-scale Arctic observation network is still in its developmental stage, with the spatiotemporal continuity of sea ice observation data needing improvement, particularly due to the scarcity and uneven quality of in situ observations. High-quality datasets are a prerequisite for data-driven models to achieve reliable predictions, as the quality of the data directly determines the upper limit of intelligent model performance.

Sea Ice Concentration (SIC), as a key indicator for polar environmental monitoring and sea ice change prediction, is critical for understanding the impacts of climate change and plays a vital role in supporting Arctic navigation, resource exploration, and climate adaptation strategies. Existing prediction models typically integrate SIC with other oceanic and atmospheric parameters, such as sea surface temperature, air temperature, and albedo, to capture the dynamics influencing SIC. These factors are primarily embedded in the spatiotemporal dependencies of SIC time series. However, significant differences in resolution and uncertainty exist among sea ice data products from various sources (e.g., satellite observations, reanalysis, and numerical simulations), resulting in SIC predictions being strongly associated with the uncertainties of these data.

With over 40 years of accumulated satellite-based SIC data, rich spatiotemporal dependencies related to SIC variability have been captured. Therefore, we propose using a single SIC dataset to accurately predict SIC over the next 10 days. This approach reduces uncertainties introduced by multiple data sources and enhances the practical applicability of the prediction model.

The daily SIC data used in this study were obtained from the Nimbus-7 SSMR and DMSP SSM/I-SSMIS passive microwave data versions provided by the NSIDC [59]. These data have a spatial resolution of 25 km and are based on an Arctic stereographic projection. The region of interest covers 448×304 grids. In each 25 km grid cell, SIC values below 15% detected by passive microwave instruments were set to 0% due to high uncertainty, with the remaining data ranging from 15 to 100%. A threshold of 15% was used as the critical minimum SIC. The dataset spans 35 years from 1988 to 2022, including a total of 12,772 days, avoiding potential errors due to inconsistent observation intervals in the data from 1978 to 1988. The parameters of the SIC data used in this study include a spatial resolution of 25 km, daily temporal resolution, and Arctic stereographic projection. Figure 6 shows two examples of SIC in March and September 2022, respectively. The average SIC in the study area exhibits significant seasonal fluctuations [48], indicating that the large-scale evolution of SIC is primarily influenced by seasonal changes. Overall, SIC reaches its maximum in March and its minimum in September.

The 35-year SIC dataset was divided into a training set, a validation set, and a test set. The first 29 years (1988–2016) were used for training, the following year (2017) was used for validation, and the last five years (2018–2022) were used for testing. The goal of the model is to predict the SIC for the subsequent 10 days based on the SIC data from the previous 10 days. A 20-day sliding window was used to generate input/output sequence samples; however, the model was trained and predicted using only the first 10 days as input and the subsequent 10 days as ground truth for supervised training. It is important to note that the 20-day sliding window was used solely to generate complete sequence data containing both input and output, not for direct prediction.

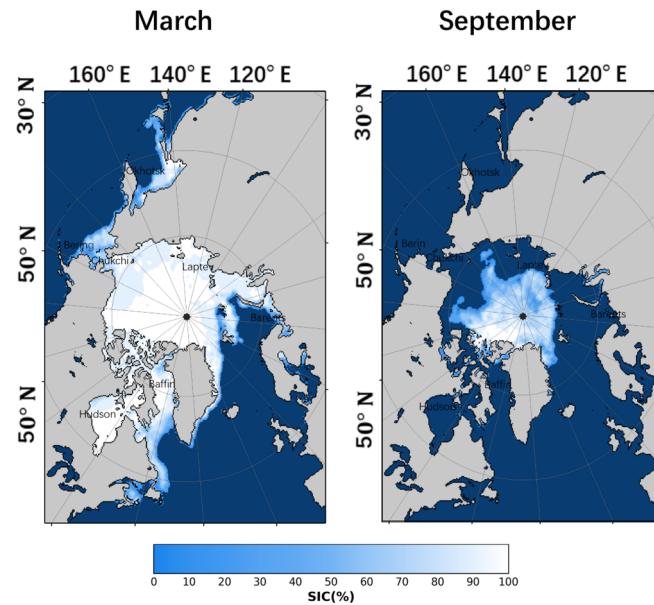


Figure 6. SIC charts for March and September.

During the testing phase, to objectively evaluate the model's prediction capability for 10 consecutive days, the sliding step was set to 10 days. This ensured that the test samples were completely independent and non-overlapping in time, providing more reliable evaluation results. Each image had a spatial resolution of $448 \times 304 \times 1$ (H $\times$ W $\times$ C). The training set consisted of 10,564 samples, the validation set included 348 samples, and the test set comprised 35 samples, with each sample corresponding to 350 consecutive days of SIC data for a given year.

3.2. Implementation Details and Evaluation Criteria

The STDNet model was trained end-to-end without any pre-training. The number of layers and hidden sizes in the Trend module were set to 1 and 128, respectively, the number of layers and hidden sizes in the Season module for encoders and decoders were set to 4 and 16, respectively, and the number of layers and hidden sizes in the GSAM were set to 2 and 128, respectively. STDNet was optimized using the MSE loss function, which is similar to the traditional spatio-temporal prediction model.

The Adam optimizer was used to train STDNet, with an initial learning rate set to 0.001; the learning rate was adjusted to 90% of the original after every 2641 iterations. The batch size of the model was set to 4 and training was stopped after 52,820 iterations.

The experimental setup included a system equipped with an Intel Xeon Platinum 8352 V 2.10 GHz twelve-core processor and an NVIDIA GeForce RTX 4090 with 24 GB of GPU memory, using Python 3.10.0 and PyTorch 2.1.0.

To comprehensively evaluate the prediction performance of STDNet, the following metrics were used: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), and Binary Accuracy (BACC), and the formulas are as follows, respectively,

$$RMSE = \sqrt{\sum_{i=0}^N \sum_{j=0}^N (X_{ij} - \hat{X}_{ij})^2} \quad (13)$$

$$MAE = \sum_{i=0}^N \sum_{j=0}^N |X_{ij} - \hat{X}_{ij}| \quad (14)$$

$$MAPE = \sum_{i=0}^N \sum_{j=0}^N \left| \frac{X_{ij} - \hat{X}_{ij}}{X_{ij}} \right| \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=0}^N \sum_{j=0}^N (X_{ij} - \hat{X}_{ij})^2}{\sum_{i=0}^N \sum_{j=0}^N (X_{ij} - \bar{X})^2} \quad (16)$$

$$BACC = \left(1 - \frac{IIEE}{\text{area of active grid cell region}} \right) \times 100\% \quad (17)$$

where N is the total number of pixels in a row or column of the input image, X_{ij} is the value at the position of the pixel in the i row and j column of the non-terrestrial grid of the real image (NSIDC), $\hat{X}_{ij}$ is the pixel value corresponding to the non-terrestrial grid of the predicted image, and $\bar{X}$ is the average value of the non-terrestrial grid of the real image. RMSE, MAPE, and MAE are common indicators used in error evaluation, and they can directly reflect the relative or absolute errors between the predicted and observed results. For MAPE, the denominator was replaced by 0.1 when X_{ij} was zero. The coefficient of determination, R^2 , measured in terms of correlation, provides a clearer picture of the differences and similarities in evolutionary trends.

BACC [38] was mainly used to measure the proximity of the predicted image sea ice edge to the real image. Grids with $SIC \leq 15\%$ were defined as open water and grids with $SIC > 15\%$ as sea ice. The Sea Ice Extent (SIE) was defined as the total area covered by grid cells with $SIC > 15\%$. The Integrated Ice Edge Error (IIEE) was the sum of all regions where the projected SIE is overestimated or underestimated, and the area of the active grid cell region was the maximum daily SIE observed from 1988 to 2022.

3.3. Results

The MAE, MAPE, RMSE, and R^2 of STDNet test data are shown in Table 1.

Table 1. STDNet evaluation indicators.

Metrics	2018	2019	2020	2021	2022	Mean
MAE	1.937	1.975	2.014	2.003	2.011	1.988
MAPE	3.575	3.936	3.441	3.313	3.438	3.541
RMSE	5.715	5.841	5.956	5.823	5.879	5.843
R^2	0.980	0.978	0.977	0.979	0.979	0.979

The results showed that the indicators varied slightly from year to year. The minimum MAE is 1.937% in 2018, the minimum MAPE is 3.313% in 2021, the minimum RMSE is 5.715% in 2018, and the maximum R^2 is 0.98 in 2018. The average MAE, MAPE, RMSE, and R^2 values were 1.988%, 3.541%, 5.843%, and 0.979, respectively. The STDNet model demonstrated a good prediction performance, with low error and high accuracy.

To show the prediction capability of STDNet more intuitively, the prediction results of the test data are visualized.

Figure 7a–e show the single-day, non-terrestrial, grid-averaged SIC true values (from NSIDC); STDNet predicted values; and MAE for the years 2018 to 2022. An analysis of the images revealed how STDNet was highly aligned with the actual observations for predicting the melting and freezing trends of sea ice and extreme points. The MAE ranged from 0.52 to 7.48%, which reflected only a small forecast difference between the five years, and a clear 10-day periodicity was evident. This periodic change primarily resulted from the increased error due to the extended prediction step. Figure 7f–j illustrate the spatial distribution of the MAE for SIC predictions from 2018 to 2022. The annual average MAE for all forecast days at each grid point was calculated in these charts. In most of the Arctic, this error was typically less than 2%. The spatial MAE rose slightly only in some sea ice margins but remained below 6%. As a result, STDNet performed well overall in the Arctic SIC forecast at 10-day intervals.

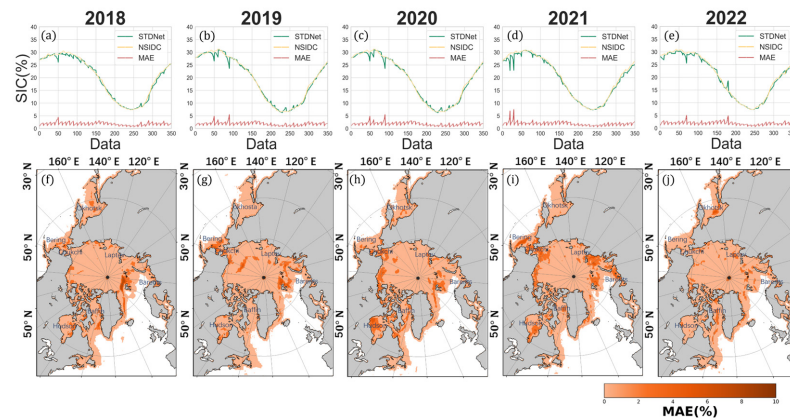


Figure 7. Performance of model test set: (a–e) 2018–2022 single-day, non-terrestrial, grid-averaged SIC true values (NSIDC); STDNet predicted values; and MAE. (f–j) 2018–2022 average spatial distribution of MAE of SIC predicted by STDNet.

To explore why the MAE increased in the sea ice edge region, the sea ice edge was calculated based on the predicted value of SIC, and the proximity was assessed using the BACC metric [38]. As shown in Figure 7a–e, the sea ice concentration reaches its minimum in September. Predictions from 8 September to 17 September, 2018–2022, were selected for evaluation, and the results, covering the period from 2020 to 2022, are shown in Figure 8, where the observed ice edge is the black contour and the STDNet predicted ice edge is the green contour.

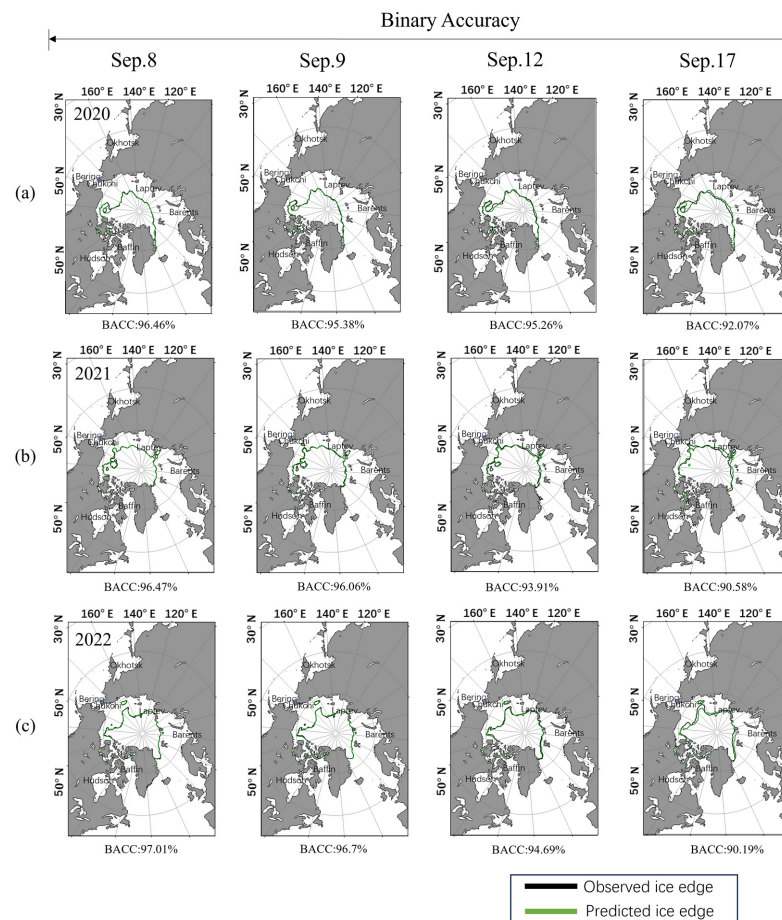


Figure 8. The BACC of STDNet during the melting season (September 8 to September 17): (a–c) 2020–2022.

STDNet predicted ice edges close to the real ice edge with an average BACC value of 93.85%, but the predicted edges tended to expand outward as the prediction step increased. This indicated that the SIE predicted by STDNet was slightly larger, showing some lag effects. This was primarily due to the strong nonlinear character of the high-frequency, small-scale evolution during the melt period and the presence of interpolation errors in the data. Overall, the prediction results of STDNet were satisfactory and reliable.

3.4. Comparative Experiments

The approach was compared with several existing models, including ConvLSTM [26], PredRNN [28], PredRNN++ [29], MIM [31], PhyDNet [32], PredNet [33], PredRNNv2 [34], MAU [35], and SimVP [36]. The average metric performances are shown in Table 2. It is evident that the MAE/MAPE/RMSE/ R^2 of our proposed STDNet were all superior to those of the existing methods. All experiments were conducted on the same dataset, and the excellent performance of each metric confirmed the predictive performance of STDNet.

Table 2. Comparison of model performance with STDNet and other models in Arctic sea ice concentration prediction. All metrics reached optimal values, demonstrating the outstanding performance of STDNet.

Models	MAE (%)	MAPE (%)	RMSE (%)	R^2
ConvLSTM (NeurIPS'2015)	2.633	7.457	6.808	0.971
PredNet (ICLR'2017)	2.045	6.222	6.621	0.972
PredRNN (NeurIPS'2017)	2.197	4.974	5.936	0.977
PredRNN++ (ICML'2018)	2.145	5.392	5.886	0.978
MIM (CVPR'2019)	2.425	8.490	6.114	0.976
PhyDNet (CVPR'2020)	4.162	10.875	10.895	0.925
MAU (NeurIPS'2021)	2.411	6.594	6.290	0.975
PredRNNv2 (TPAMI'2022)	2.526	7.917	6.407	0.974
SimVP (CVPR'2022)	2.376	6.506	6.397	0.974
STDNet	1.988	3.541	5.843	0.979

The best results are in bold.

To conduct investigations in greater detail, a set of prediction values for March and September were selected, when the SIC reached its maximum and minimum values, respectively. The dates from 12 March to 21 March and 8 September to 17 September of 2022 were specifically chosen.

Figures 9–11 illustrate the detailed conditions from 12 March to 21 March. Comparisons of the daily quantitative performances of MAE, MAPE, and R^2 for each model are shown in Figure 9. The data here represent the number of days from 12 March to 21 March. The prediction performance of our proposed STDNet was significantly better than that of the other existing methods. In particular, our model consistently outperformed other methods in long-term forecasting, and the gap between the metrics was more pronounced as the number of forecast days increased.

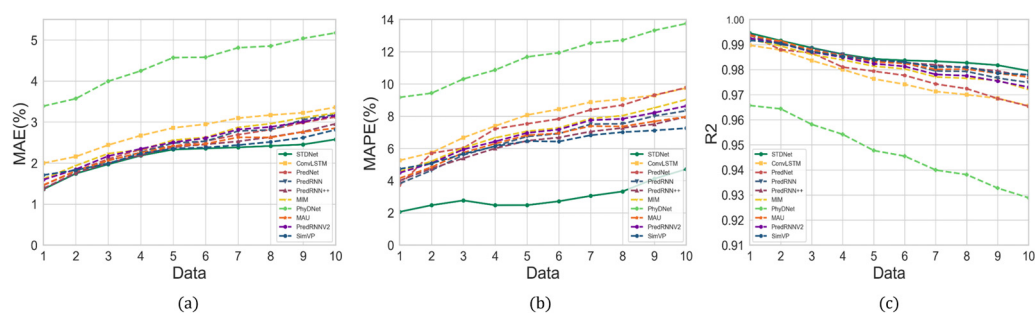


Figure 9. Comparison of quantitative time-steps between STDNet and other methods for three metrics from March 12 to March 21: (a) MAE, (b) MAPE, and (c) R^2 .

To qualitatively evaluate the predictive performance of the proposed STDNet, comparative experiments on daily predicted SIC were conducted by comparing with observed images to assess how well the prediction model predicted future SIC. Figure 10 shows the observed images and some of the images predicted by the better performing models. Figure 11 shows the spatial distribution of the daily residuals for these models, highlighting the differences between observed and predicted images. In Figure 11, the yellow areas indicate positive residuals, meaning the predictions are lower than the actual values, and the blue areas indicate negative residuals, where predictions are higher than actual values. As shown in Figures 10 and 11, the SIC predicted by the STDNet model is basically consistent with the observed image with small residuals.

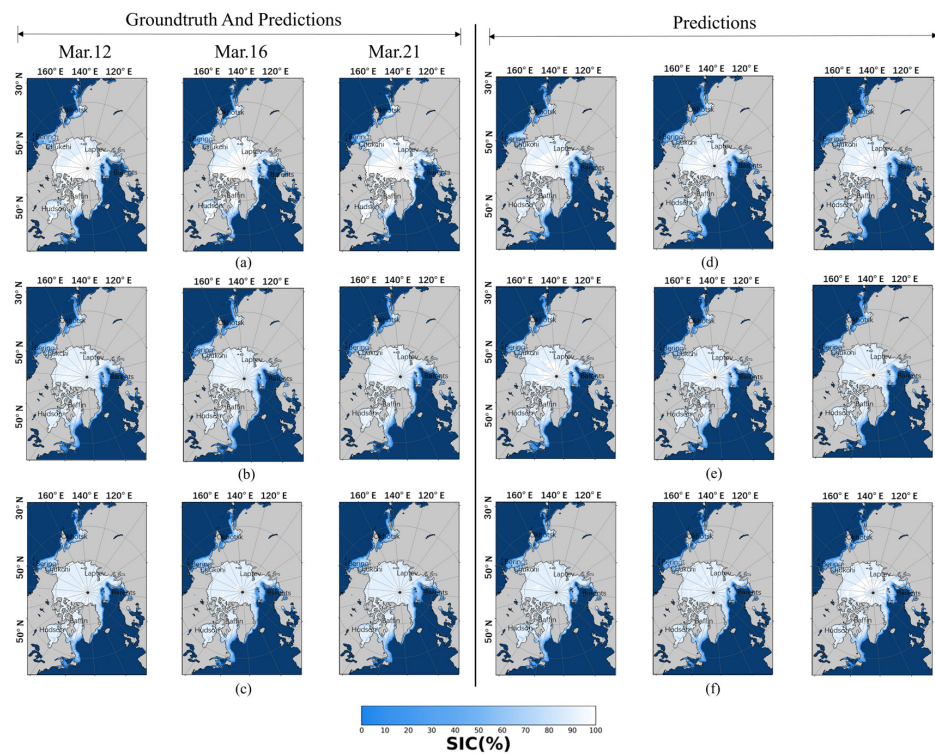


Figure 10. Comparison of the prediction performance of STDNet with some of the methods from March 12 to March 21: (a) NSIDC, (b) ConvLSTM, (c) PredRNN++, (d) PredRNNV2, (e) SimVP, (f) STDNet. All experiments were conducted on the same dataset, using the previous 10 days of data to predict the following 10 days.

Overall, the prediction errors for the central Arctic ice cap across all models were relatively small, generally within 5%, and mostly tended to be overestimated. In the projected SIC, most of the anomalies occurred in the marginal ice areas of the North Atlantic and North Pacific due to fluctuations at the ice edge. Although all models generated similar outputs in terms of quantity and quality, the consistency of the SIC images predicted by the STDNet model with the observed images surpassed that of the other comparative models. In this respect, both the overestimated and underestimated areas were geographically similar, but the number of STDNet errors was smaller than for the comparative models. The following is a brief qualitative description of the predictions for each day: Images from the first two days show that the residuals for most regions of the STDNet prediction images were in the range of -5% to 5% , except near the Barents Sea where the residuals reached -10% . In contrast, the residuals of the predicted images of other comparison algorithms were generally in the regions of -10% to 10% in the edge ice region. From day 5 to day 10, the residuals in the edge ice areas of STDNet's predicted images gradually increased, but the residuals in most central areas remained within the range of $(-5\%, 5\%)$. The consistency in the ice edge regions was continually superior to that of the other comparative models.

Figures 12–14 illustrate the detailed conditions from 8 September to 17 September. Figure 12 shows a comparison between the quantitative performance of the models in terms of daily MAE, MAPE, and R^2 . The data in the figure represent the number of days from 8 September to 17 September. Similar to March, the predictive performance of our proposed STDNet was significantly better than that of other existing methods.

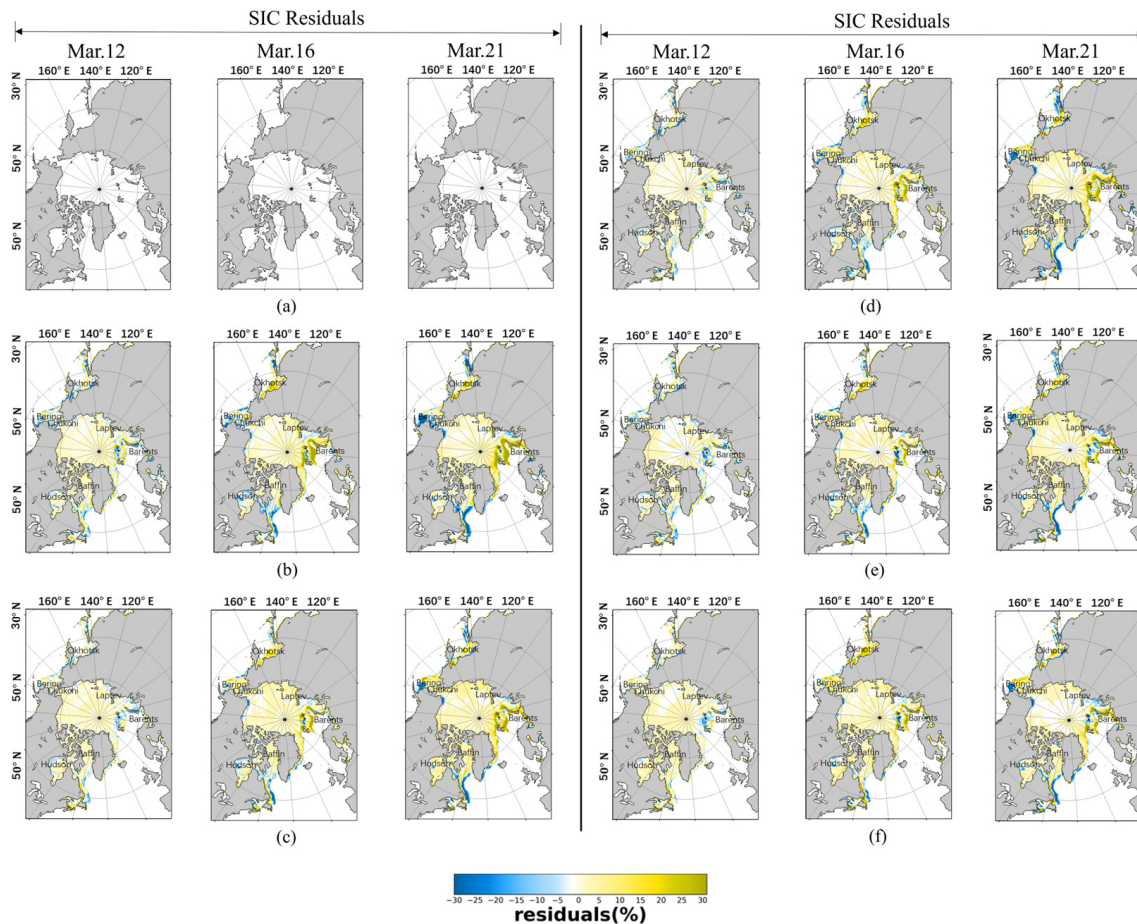


Figure 11. Residuals distribution of predicted SIC by selected models and NSIDC SIC from March 12 to March 21: (a) NSIDC, (b) ConvLSTM, (c) PredRNN++, (d) PredRNNV2, (e) SimVP, (f) STDNet.

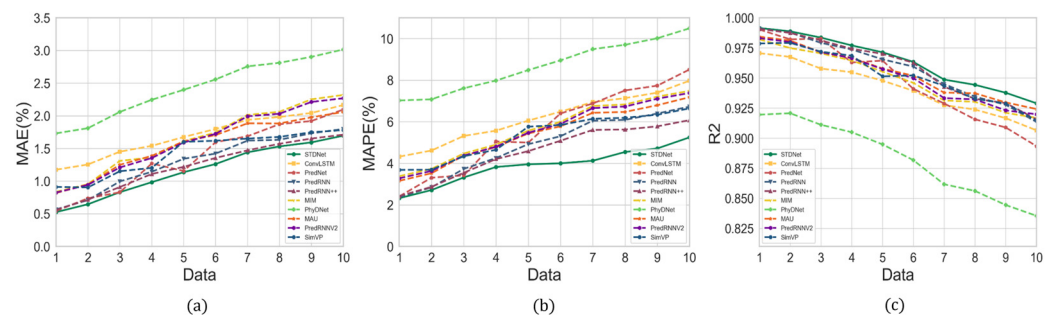


Figure 12. Quantitative time-step comparison between STDNet and other methods for three metrics: (a) MAE, (b) MAPE, and (c) R^2 from 8 September to 17 September.

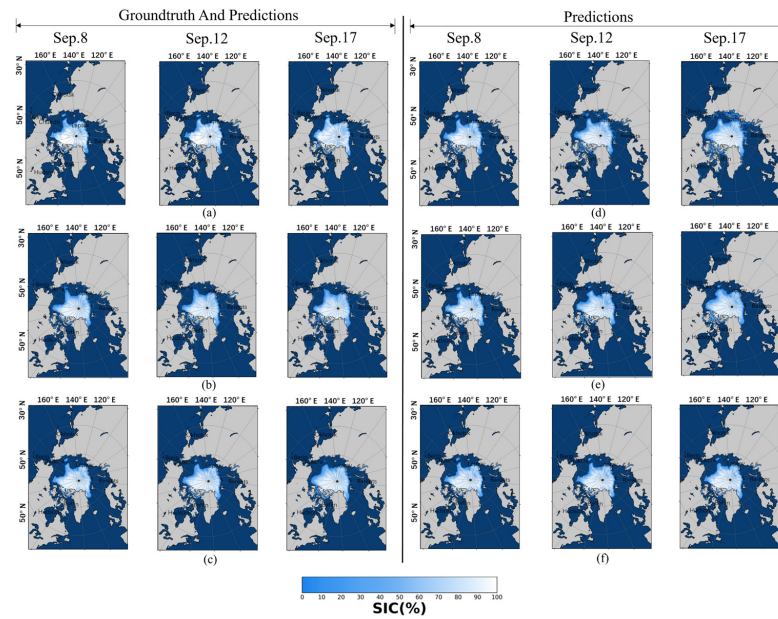


Figure 13. Comparison between the prediction performance of STDNet and some of the methods from 8 September to 17 September: (a) NSIDC, (b) ConvLSTM, (c) PredRNN++, (d) PredRNNV2, (e) SimVP, (f) STDNet. All experiments were conducted on the same dataset, using the previous 10 days of data to predict the following 10 days.

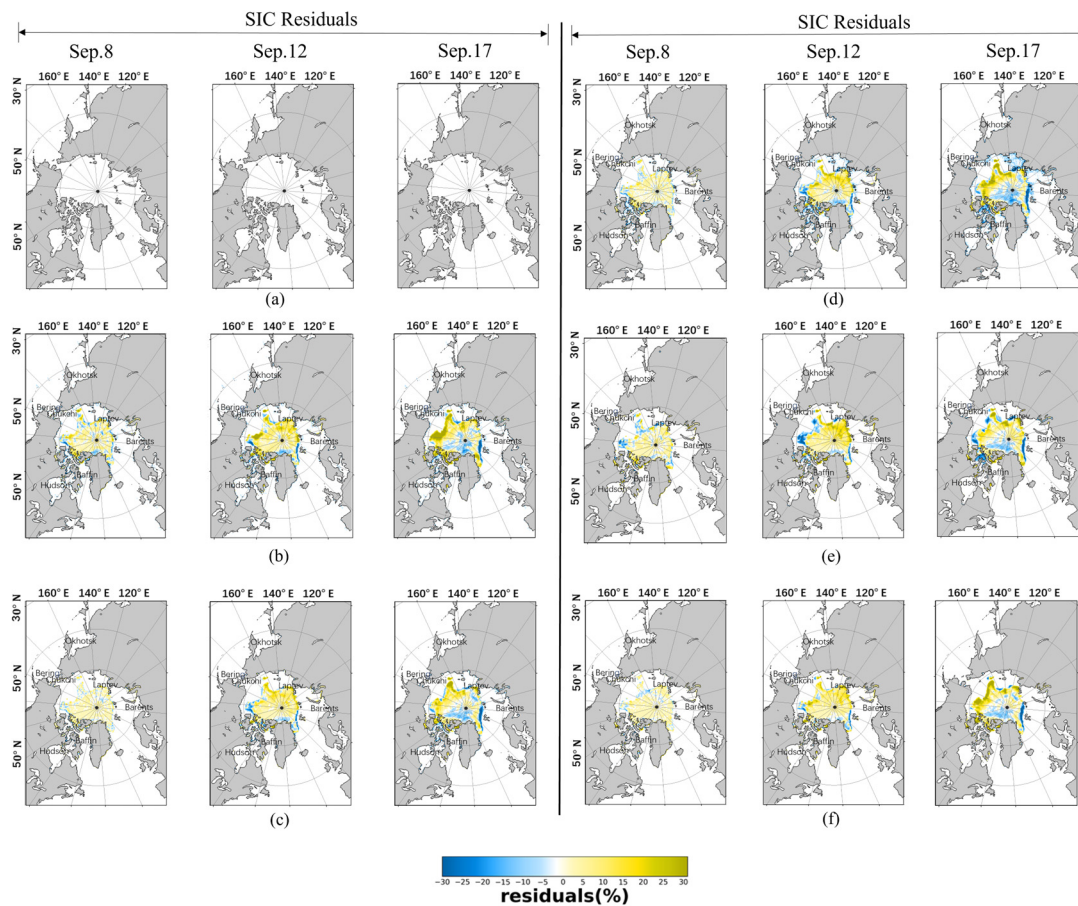


Figure 14. Residual distribution of SIC predicted by selected models and NSIDC SIC from September 8 to September 17: (a) NSIDC, (b) ConvLSTM, (c) PredRNN++, (d) PredRNNV2, (e) SimVP, (f) STDNet.

Figure 13 shows the observed images for September and some of the images predicted by the better performing models. Figure 14 shows the spatial distribution of the daily residuals for these models. In September, the sea ice extent was significantly reduced but, similar to March, most anomalies in the predicted images occurred at the ice edges, with errors in the central Arctic ice cap remaining small and generally within 5%. STDNet's predicted images show better consistency than those of other comparative models, and the residuals in the ice edge areas are relatively smaller.

The evaluation of the predictive model's performance indicates that the STDNet model demonstrates significant advantages in Sea Ice Concentration (SIC) prediction. First, its prediction results show higher consistency with observed images compared to other benchmark models. Second, STDNet is more accurate in capturing the detailed changes along the sea ice edges, which is a critical aspect of SIC prediction, as the prediction of sea ice edge areas is crucial for assessing ice extent and understanding the dynamic changes in sea ice. Finally, STDNet maintains high predictive accuracy across different seasons and varying ranges of sea ice concentration, indicating the model's robustness and adaptability, making it better suited for SIC prediction under extreme weather conditions.

The definition of a lightweight model lies in achieving efficient predictive performance with minimal resource consumption, including parameters, computational cost, and time. This design philosophy focuses on optimizing model structure, reducing redundant parameters, and lowering computational complexity, enabling the model to perform exceptionally well, even under hardware resource constraints. Table 3 summarizes the comparison between the proposed STDNet and other existing models in terms of lightweight characteristics. The training times reported in the table were measured on a single NVIDIA GeForce RTX 4090 GPU, providing a comprehensive perspective for evaluating the performance of STDNet. Compared to other methods, STDNet features the fewest parameters (Params), the fastest training time, and lower Floating-point Operations per second (FLOPs). This demonstrates that STDNet's lightweight design is not only reflected in the reduction of parameters and computational costs but also in significantly shorter training times and outstanding predictive accuracy, thereby achieving a balance between computational efficiency and predictive performance. This makes STDNet a highly efficient model choice that maintains high accuracy while being suitable for practical application scenarios.

Table 3. Results of experiment conducted to compare computational costs with other methods. All experiments were performed on the same dataset.

Models	FLOPs (G)	Params (M)	Training time (h)
ConvLSTM (NeurIPS'2015)	133.000	4.133	1.767
PredNet (ICLR'2017)	7.041	3.129	5.300
PredRNN (NeurIPS'2017)	252.000	1.558	3.600
PredRNN++ (ICML'2018)	368.000	2.479	4.683
MIM (CVPR'2019)	398.000	12.446	5.383
PhyDNet (CVPR'2020)	510.000	3.092	5.400
MAU (NeurIPS'2021)	37.345	8.931	5.583
PredRNNV2 (TPAMI'2022)	253.000	1.559	4.650
SimVP (CVPR'2022)	44.714	3.488	1.800
STDNet	23.878	1.490	1.567

The best results are in bold.

Overall, STDNet achieves high-precision predictions while maintaining a lightweight model structure, offering significant practical advantages. By minimizing the model's parameter count and FLOPs, it effectively reduces computational resource consumption, making it highly adaptable for resource-constrained environments. Its streamlined architecture significantly shortens training time and enhances inference speed, ensuring faster model updates and real-time predictions. Despite its lightweight design, STDNet consistently delivers superior accuracy, demonstrating its robustness and effectiveness across various tasks. This combination of efficiency and high performance ensures that STDNet

is well-suited for real-world applications, striking an ideal balance between operational feasibility and predictive capability.

3.5. Ablation Studies

This section presents the results of ablation studies conducted to evaluate the contribution of each component in the model, including the network architecture design and spatio-temporal attention mechanism construction. To evaluate the contribution of the decomposition idea and GSAM to SIC prediction, ablation experiments were performed on the dataset without decomposition or without the GSAM. To ensure the fairness of the comparison, all parameter settings and experimental configurations were identical. The results are summarized in Table 4.

Table 4. Ablation studies conducted to assess the role of decomposition and GSAM. All experiments were conducted on the same dataset, using data from the previous 10 days to predict the subsequent 10 days.

Models	MAE (%)	MAPE (%)	RMSE (%)	R ²
STDNet no Decompose	2.097	4.593	5.882	0.978
STDNet no GSAM	2.106	4.865	5.935	0.977
STDNet	1.988	3.541	5.843	0.979

The results showed that models without decomposition performed significantly worse across all metrics, demonstrating that STL decomposition effectively extracts seasonal signals and trend information from the data. Additionally, the absence of the GSAM led to a notable deterioration in all metrics, confirming that GSAM was beneficial for capturing the spatio-temporal dependencies in the data. In summary, the ablation study validated that both decomposition and the GSAM contributed positively to achieving satisfactory model performance.

4. Discussion

Comparing prediction results across different regions or data types is valuable, as model performance can vary significantly across different application scenarios. In the following Sections 4.1–4.3, STDNet evaluates the prediction results of various models on Antarctic sea ice concentration, Arctic sea ice drift, and Arctic sea ice thickness. These comparisons highlight STDNet’s excellent generalization and broad applicability.

4.1. Antarctic Sea Ice Concentration

The Antarctic daily SIC data we selected also come from the Nimbus-7 SSMR and DMSP SSM/I-SSMIS passive microwave data provided by NSIDC [59]. The dataset has a spatial resolution of 25 km and is based on a polar stereographic projection of Antarctica. Similar to the Arctic, data values from passive microwave instruments below 15% in a given 25 km grid cell are considered too uncertain to use. Therefore, values below 15% were set to 0% ice, while the remaining data represent concentrations ranging from 15 to 100%. The 15% threshold is used as the critical threshold for the lowest SIC, ensuring data comparability across seasons and years.

The data from 1988 to 2016 were used for training, 2017 for validation, and 2018 to 2022 for testing. The spatial resolution of the images is $332 \times 316 \times 1$. To ensure that the convolutional layer in the model can perform multiple downsamplings, the images were cropped to $328 \times 312 \times 1$. This adjustment ensures data comparability across seasons and annual intervals.

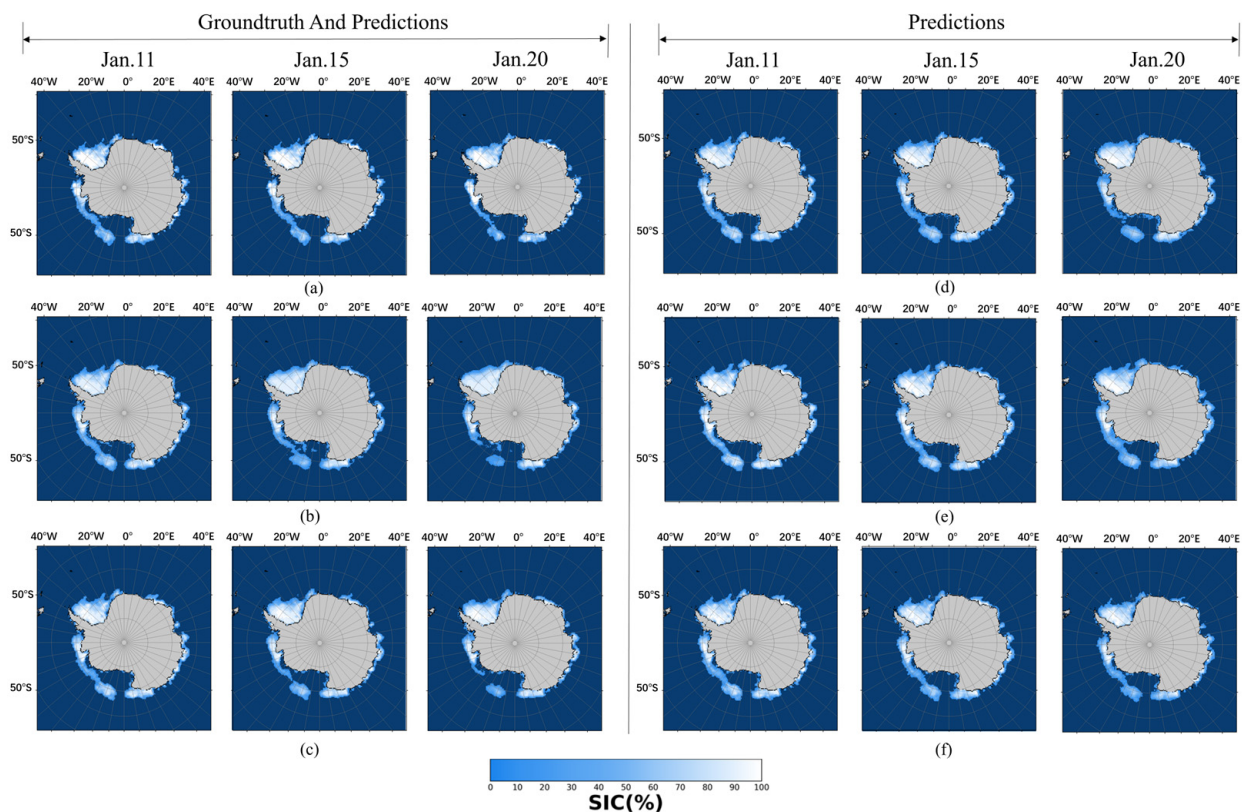
To quantitatively assess STDNet’s prediction performance on Antarctic SIC data, several existing models were compared, and the average metric performance of all methods is presented in Table 5. Compared to other methods, STDNet demonstrates optimized MAE, MAPE, RMSE, and R² for the 10-day Antarctic SIC forecast.

Table 5. Comparison of model performance with STDNet and other models in Antarctic sea ice concentration prediction.

Models	MAE (%)	MAPE (%)	RMSE (%)	R ²
ConvLSTM (NeurIPS'2015)	2.496	5.665	6.549	0.963
PredNet (ICLR'2017)	2.224	5.554	7.333	0.954
PredRNN (NeurIPS'2017)	2.310	5.980	6.106	0.968
PredRNN++ (ICML'2018)	2.271	5.875	6.021	0.970
MIM (CVPR'2019)	2.289	5.465	6.131	0.968
PhyDNet (CVPR'2020)	4.590	6.500	10.597	0.904
MAU (NeurIPS'2021)	2.382	5.906	6.254	0.966
PredRNNV2 (TPAMI'2022)	2.359	6.001	6.282	0.966
SimVP (CVPR'2022)	2.494	6.091	6.465	0.964
STDNet	2.165	5.335	5.841	0.971

The best results are in bold.

As shown in Figure 15, to qualitatively evaluate STDNet's prediction performance, an image comparison is provided with several models that exhibit better prediction accuracy. The melting period of Antarctic sea ice, which typically occurs during the southern hemisphere's summer (around November to February), is opposite to that of the Arctic. Therefore, a prediction image from three days in January 2022, specifically from 11th January to 20th January, was selected. STDNet outperforms other models, producing images that are not only closer to real observations in terms of location but also more detailed in capturing edge contours and morphology. This results in higher clarity and consistency in critical areas, reducing misclassification and omission compared to other models.

**Figure 15.** Comparison of the prediction performance of STDNet with selected methods on Antarctic sea ice concentration between 11 January and 20 January: (a) NSIDC, (b) ConvLSTM, (c) PredRNN++, (d) PredRNNV2, (e) SimVP, and (f) STDNet.

There are significant differences between Antarctic and Arctic Sea Ice Concentration (SIC) data in terms of geographic location, seasonal variability, dynamic properties, and influencing factors. Compared to the Arctic, Antarctic sea ice exhibits more drastic seasonal variations and a less consistent trend, reflecting its more complex characteristics. STDNet demonstrates excellent performance in predicting Antarctic SIC, highlighting the model's robustness and accuracy in handling complex and variable sea ice dynamics. Additionally, the model effectively captures the unique seasonality and trends of both Arctic and Antarctic sea ice, showcasing strong generalization capabilities.

4.2. Arctic Sea Ice Drift

For Sea Ice Drift (SID) data, the NSIDC Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vector Version 4 [60] was used. This high-resolution dataset is employed to study Arctic and Antarctic sea ice dynamics, providing daily global sea ice motion vectors from 1988 to the present. The dataset integrates multiple sources, including satellite imagery (e.g., AVHRR, SMMR, SSMI, SSMI/S, AMSR), NCEP/NCAR wind reanalysis data, and International Arctic Buoy Program (IABP) buoy-position data. By combining these sources, the u-component (x-direction) and v-component (y-direction) of sea ice motion are generated. To minimize polar projection distortion and ensure data accuracy, the u- and v-components are projected onto a 25 km grid (EASE-Grid) using optimal interpolation.

The data from 1988 to 2016 were divided into a training set, 2017 into a validation set, and 2018 to 2022 into a test set. The spatial resolution of the u-component and v-component images is $361 \times 361 \times 1$. To unify the inputs, these components are combined in the channel dimension, resulting in an input image resolution of $361 \times 361 \times 2$. To ensure that the model's convolutional layers can perform multiple downsamplings, the images are cropped to $360 \times 360 \times 2$.

When comparing the prediction results of STDNet with those of selected models, the average metric performance is presented in Table 6. For the Arctic SID 10-day prediction, STDNet is optimal in RMSE and R^2 , while slightly trailing PredRNN++ in MAE and MAPE. However, considering the differences in model complexity, this slight accuracy difference is deemed acceptable.

Table 6. Comparison of model performance with STDNet and other models in Arctic sea ice drift prediction. Lower RMSE and higher R^2 indicate satisfactory model performance.

Models	MAE (cm/s)	MAPE (cm/s)	RMSE (cm/s)	R^2
ConvLSTM (NeurIPS'2015)	1.469	0.506	7.165	0.941
PredRNN++ (ICML'2018)	1.382	0.443	6.446	0.952
PredRNNV2 (TPAMI'2022)	1.490	0.495	6.831	0.945
SimVP (CVPR'2022)	1.560	0.478	6.831	0.947
STDNet	1.450	0.473	6.384	0.953

The best results are in bold.

Figure 16 shows the residual distribution images of several models over three days from 11 March to 20 March 2022. Comparing the residual images, the STDNet model demonstrates superior performance in residual distribution, especially in long-term predictions. The residual images of the STDNet model exhibit a more uniform distribution and smaller residual values. Due to the drastic changes in sea ice drift data, the residuals at the sea ice edges are generally higher. However, compared to other models, the STDNet model shows fewer and smaller residuals, highlighting its excellent accuracy and stability in handling complex sea ice edge features.

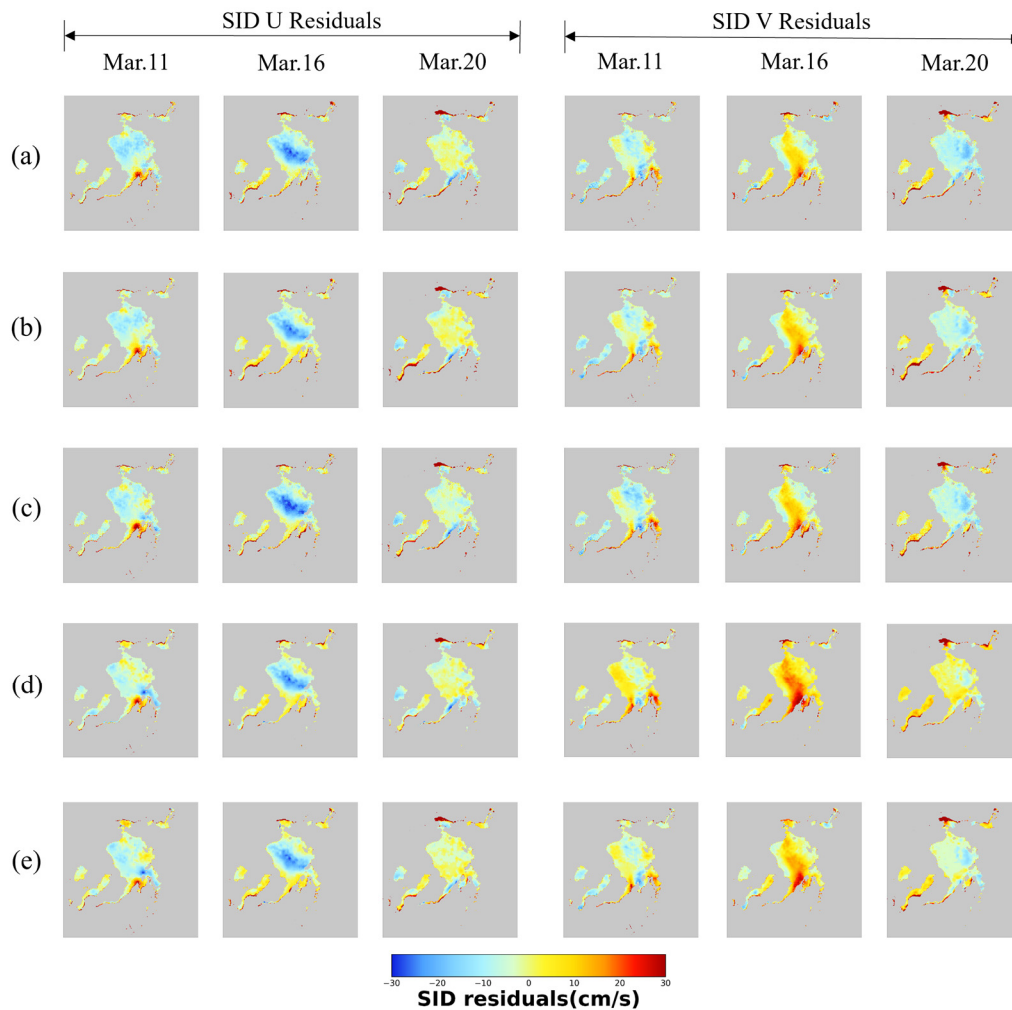


Figure 16. Residual distributions of sea ice drift u- and v-components predicted by selected models and real images from March 11 to March 20: (a) ConvLSTM, (b) PredRNN++, (c) PredRNNV2, (d) SimVP, and (e) STDNet.

4.3. Arctic Sea Ice Thickness

For Sea Ice Thickness (SIT) data, the SMOS thin sea ice thickness dataset version 300, provided by the University of Bremen, was selected [61]. This dataset is derived from the inversion of sea ice thickness using L-band brightness temperature observations at high incidence angles ($50\text{--}70^\circ$) from the SMOS satellite. Version 300 data have been finely calibrated and optimized, making it particularly suitable for studies in thin ice regions, especially during the freeze–thaw periods of sea ice thickness changes.

To generate sea ice thickness data, brightness temperature data from the SMOS satellite is first acquired and corrected to eliminate noise from atmospheric conditions, ionospheric effects, and other interferences. Physical and empirical models are then established to relate brightness temperature to sea ice thickness, considering the impact of high incidence angles on brightness temperature measurements. Finally, an optimization algorithm is employed to invert sea ice thickness from the brightness temperature data, with the results stabilized and enhanced through the combination of multiple observations.

The Arctic SIT data span from 2010 to the present. Data from 2010 to 2018 were used for training, 2019 for validation, and 2020 to 2022 for testing. The spatial resolution of the images is $896 \times 608 \times 1$.

Table 7 compares the evaluation metrics of STDNet with other models, showing that STDNet achieves optimal MAE, MAPE, RMSE, and R^2 for the Arctic SIT 10-day forecast. To qualitatively assess STDNet’s prediction performance, images predicted by each model

for three days from 17 September to 26 September 2022, were selected for comparison. As shown in Figure 17, the SIT images predicted by STDNet exhibit significant advantages in image quality and detail capture. Compared to other models, STDNet's generated images more clearly depict the hierarchical structure of sea ice thickness, are more consistent with real observations, and perform better in edge processing.

Table 7. Comparison of model performance with STDNet and other models in Arctic sea ice thickness prediction. All metrics also reached optimal values, proving that STDNet maintains its performance across different datasets.

Method	MAE (cm)	MAPE (cm)	RMSE (cm)	R ²
ConvLSTM	1.373	0.240	3.957	0.944
PredRNN++	1.115	0.193	3.310	0.960
PredRNNV2	1.470	0.292	4.140	0.938
SimVP	1.487	0.316	3.518	0.955
STDNet	1.082	0.173	3.250	0.962

STDNet excels not only in predicting SIC data but also in predicting SID and SIT data, reflecting its strong generalization ability. Regardless of data type or diverse sea ice environmental conditions, STDNet maintains consistently robust performance and demonstrates exceptional adaptability, ensuring high-quality prediction results across varying sea ice dynamics.

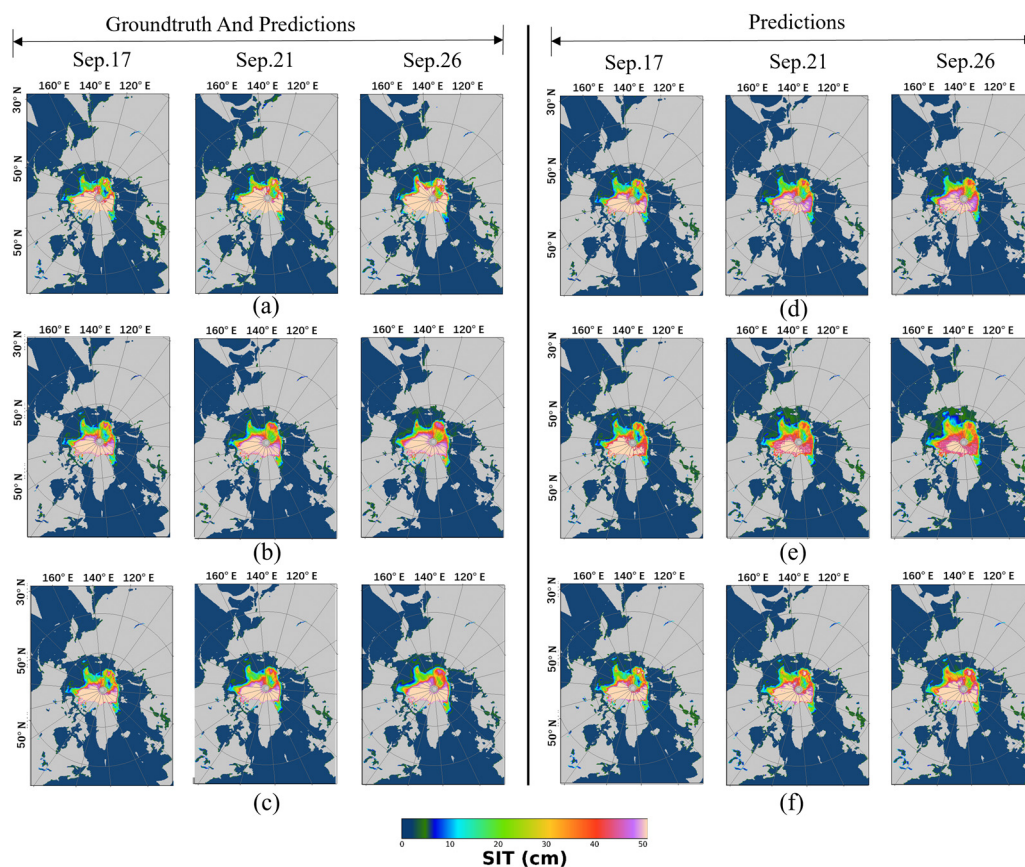


Figure 17. Comparison of the prediction performance of STDNet with selected methods on Arctic sea ice thickness between 17 September and 26 September: (a) NSIDC, (b) ConvLSTM, (c) PredRNN++, (d) PredRNNV2, (e) SimVP, and (f) STDNet.

4.4. Navigational Risk Prediction

STDNet demonstrates significant potential in the field of navigational risk prediction, capable of integrating dynamic multi-source data such as meteorological, oceanic, and environmental factors to provide precise predictions for route safety. For example, in polar routes, sea ice dynamics and meteorological conditions (such as wind speed and air pressure) are critical to navigation safety. By modeling the long-term trends of sea ice coverage and short-term meteorological fluctuations, STDNet can accurately predict future navigational conditions, offering scientific support for polar route planning and real-time navigation. Similarly, for coastal routes, in scenarios such as typhoon path prediction and storm surge simulation, STDNet can combine oceanographic data (e.g., current speed and wave height) with meteorological factors to assess the potential impact of severe weather on routes, reducing shipping risks caused by natural disasters.

The lightweight design and efficient inference capabilities of STDNet make it particularly well-suited for real-time applications in navigational scenarios. It can be deployed in ship navigation systems, dynamically predicting route risks by integrating real-time environmental data and providing optimized route suggestions, thereby enhancing the safety and economic efficiency of maritime transport. Additionally, STDNet's fast training capability allows it to quickly adapt to different maritime regions and environmental conditions, offering customized prediction support for global navigation routes.

In the future, STDNet can be further integrated with satellite remote sensing and Automatic Identification System (AIS) data to enhance its capabilities in the dynamic monitoring of complex sea areas, real-time sea ice change alerts, and collision risk prediction in densely navigated regions. By integrating multi-source data and optimizing model deployment, STDNet will provide efficient, accurate, and scalable solutions for navigational risk prediction, delivering both safety and efficiency improvements to the global shipping industry.

5. Conclusions

We propose a lightweight, non-recurrent, spatio-temporal prediction model, STDNet, for predicting daily SIC in the Arctic. Specifically, the model is divided into seasonal and trend parts based on the STL decomposition concept. In addition, we introduce the GSAM, which enables STDNet to not only extract seasonal signals and trend information in the data but also capture the spatio-temporal dependencies in the data, thus providing end-to-end predictions of daily SIC for the following 10 days. STDNet was tested using SIC data from 2018 to 2022. Comprehensive experimental results showed the following: (1) using only satellite remote sensing SIC data as input, STDNet predicted 10-day SICs with an overall MAE of 1.988%, MAPE of 3.541%, RMSE of 5.843%, and R^2 of 0.979; (2) the prediction performance of STDNet outperformed existing advanced methods both quantitatively and qualitatively; (3) our model requires fewer parameters but provides a higher fitting efficiency due to the densely connected structure and the sparse MLP. (4) In the 10-day predictions for Antarctic SIC, Arctic SID, and Arctic SIT, STDNet has demonstrated exceptional performance; it not only maintains high accuracy but also exhibits strong generalization capabilities.

Despite these advances, a number of issues remain: (1) Uncertainty in sea ice data. As a data-driven approach, deep learning does not require prior physical principles, but it has high requirements on the training data, and the noise in the training data can affect the model's prediction. Currently, due to the limitations of satellite observation technologies, the accuracy of SIC observations is poor in the summer and fall. (2) Complexity of sea ice change patterns. Due to the great differences in atmospheric circulation, ocean current patterns, and hydrological characteristics in different regions, the changes in sea ice show strong variability and spatio-temporal heterogeneity. Although deep learning excels at fitting complex nonlinear patterns, it is primarily used in fields such as image processing. The current mainstream loss functions, such as MAE and SSIM, including the MSE used in this paper, consider the magnitude of the error or the difference in the visual contrast.

There is no loss function for model training for the spatial heterogeneity or local variability of sea ice.

To make the deep-learning model more suitable for the short-term forecasting of sea ice, we aim in future research to introduce elements in the atmosphere and the ocean that are related to sea ice changes, to enable the patterns of change of sea ice to be characterized, to reduce the impact caused by uncertainties in sea ice data, and to improve the loss function by adding constraints on physical mechanisms.

Author Contributions: Conceptualization, X.Z. and H.Z.; methodology, X.Z. and H.Z.; software, X.Z.; validation, X.Z., J.W. and H.Z.; formal analysis, X.Z.; investigation, H.Z.; resources, H.Z.; data curation, X.Z. and H.Z.; writing—original draft preparation, X.Z. and J.W.; writing—review and editing, X.Z. and H.Z.; visualization, J.W.; supervision, H.Z., G.W., Y.J. and Y.S.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 52101405).

Data Availability Statement: The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author. The data used in this study were provided by the National Snow and Ice Data Center (NSIDC) and were critical to this research. These data are available upon request, subject to the data-sharing policy of the center.

Acknowledgments: The authors would like to thank the editors and reviewers for their insightful suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Li, X.; Stephenson, S.R.; Lynch, A.H.; Goldstein, M.A.; Bailey, D.A.; Veland, S. Arctic shipping guidance from the CMIP6 ensemble on operational and infrastructural timescales. *Clim. Chang.* **2021**, *167*, 23. [\[CrossRef\]](#)
- Shi, J.; Lee, W.-J.; Liu, Y.; Yang, Y.; Wang, P. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Trans. Ind. Appl.* **2012**, *48*, 1064–1069. [\[CrossRef\]](#)
- Peng, Z.; Yu, D.; Huang, D.; Heiser, J.; Yoo, S.; Kalb, P. 3D cloud detection and tracking system for solar forecast using multiple sky imagers. *Sol. Energy* **2015**, *118*, 496–519. [\[CrossRef\]](#)
- Screen, J.A.; Simmonds, I. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature* **2010**, *464*, 1334–1337. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kaminski, T.; Kauker, F.; Toudal Pedersen, L.; Voßbeck, M.; Haak, H.; Niederdrenk, L.; Hendricks, S.; Ricker, R.; Karcher, M.; Eicken, H. Arctic mission benefit analysis: Impact of sea ice thickness, freeboard, and snow depth products on sea ice forecast performance. *Cryosphere* **2018**, *12*, 2569–2594. [\[CrossRef\]](#)
- Yang, Q.; Losa, S.N.; Losch, M.; Tian-Kunze, X.; Nerger, L.; Liu, J.; Kaleschke, L.; Zhang, Z. Assimilating SMOS sea ice thickness into a coupled ice-ocean model using a local SEIK filter. *J. Geophys. Res. Ocean.* **2014**, *119*, 6680–6692. [\[CrossRef\]](#)
- Zhang, J. Sea ice properties in high-resolution sea ice models. *J. Geophys. Res. Ocean.* **2021**, *126*, e2020JC016686. [\[CrossRef\]](#)
- Adcroft, A.; Anderson, W.; Balaji, V.; Blanton, C.; Bushuk, M.; Dufour, C.O.; Dunne, J.P.; Griffies, S.M.; Hallberg, R.; Harrison, M.J. The GFDL global ocean and sea ice model OM4. 0: Model description and simulation features. *J. Adv. Model. Earth Syst.* **2019**, *11*, 3167–3211. [\[CrossRef\]](#)
- Horvat, C.; Jones, D.R.; Iams, S.; Schroeder, D.; Flocco, D.; Feltham, D. The frequency and extent of sub-ice phytoplankton blooms in the Arctic Ocean. *Sci. Adv.* **2017**, *3*, e1601191. [\[CrossRef\]](#)
- Fu, H.; Wu, X.; Li, W.; Zhang, L.; Liu, K.; Dan, B. Improving the accuracy of barotropic and internal tides embedded in a high-resolution global ocean circulation model of MITgcm. *Ocean Model.* **2021**, *162*, 101809. [\[CrossRef\]](#)
- Hunke, E.C.; Lipscomb, W.H.; Turner, A.K.; Jeffery, N.; Elliott, S. CICE: The Los Alamos Sea ice model documentation and software user's manual version 5.1 LA-CC-06-012. *T-3 Fluid Dyn. Group Los Alamos Natl. Lab.* **2015**, 675, 15.
- Hebert, D.A.; Allard, R.A.; Metzger, E.J.; Posey, P.G.; Preller, R.H.; Wallcraft, A.J.; Phelps, M.W.; Smedstad, O.M. Short-term sea ice forecasting: An assessment of ice concentration and ice drift forecasts using the US Navy's Arctic Cap Nowcast/Forecast System. *J. Geophys. Res. Ocean.* **2015**, *120*, 8327–8345. [\[CrossRef\]](#)
- Smith, G.C.; Roy, F.; Reszka, M.; Surcel Colan, D.; He, Z.; Deacu, D.; Belanger, J.M.; Skachko, S.; Liu, Y.; Dupont, F. Sea ice forecast verification in the Canadian global ice ocean prediction system. *Q. J. R. Meteorol. Soc.* **2016**, *142*, 659–671. [\[CrossRef\]](#)
- Blockley, E.; Vancoppenolle, M.; Hunke, E.; Bitz, C.; Feltham, D.; Lemieux, J.-F.; Losch, M.; Maisonnave, E.; Notz, D.; Rampal, P. The future of sea ice modeling: Where do we go from here? *Bull. Am. Meteorol. Soc.* **2020**, *101*, E1304–E1311. [\[CrossRef\]](#)

15. Hunke, E.C.; Lipscomb, W.H.; Turner, A.K. Sea-ice models for climate study: Retrospective and new directions. *J. Glaciol.* **2010**, *56*, 1162–1172. [\[CrossRef\]](#)
16. Girard, L.; Weiss, J.; Molines, J.-M.; Barnier, B.; Bouillon, S. Evaluation of high-resolution sea ice models on the basis of statistical and scaling properties of Arctic sea ice drift and deformation. *J. Geophys. Res. Ocean.* **2009**, *114*. [\[CrossRef\]](#)
17. Hutchings, J.K.; Roberts, A.; Geiger, C.A.; Richter-Menge, J. Spatial and temporal characterization of sea-ice deformation. *Ann. Glaciol.* **2011**, *52*, 360–368. [\[CrossRef\]](#)
18. Kim, R.; Tremblay, L.B.; Brunette, C.; Newton, R. A regional seasonal forecast model of Arctic minimum sea ice extent: Reflected solar radiation versus late winter coastal divergence. *J. Clim.* **2021**, *34*, 6097–6113. [\[CrossRef\]](#)
19. Blanchard-Wrigglesworth, E.; Cullather, R.; Wang, W.; Zhang, J.; Bitz, C. Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook. *Geophys. Res. Lett.* **2015**, *42*, 8042–8048. [\[CrossRef\]](#)
20. Wang, L.; Yuan, X.; Ting, M.; Li, C. Predicting summer Arctic sea ice concentration intraseasonal variability using a vector autoregressive model. *J. Clim.* **2016**, *29*, 1529–1543. [\[CrossRef\]](#)
21. Yuan, X.; Chen, D.; Li, C.; Wang, L.; Wang, W. Arctic sea ice seasonal prediction by a linear Markov model. *J. Clim.* **2016**, *29*, 8151–8173. [\[CrossRef\]](#)
22. Wang, L.; Yuan, X.; Li, C. Subseasonal forecast of Arctic sea ice concentration via statistical approaches. *Clim. Dyn.* **2019**, *52*, 4953–4971. [\[CrossRef\]](#)
23. Andrade, J.; Katoch, S.; Turaga, P.; Spanias, A.; Tepedelenlioglu, C.; Jaskie, K. Formation-aware cloud segmentation of ground-based images with applications to PV systems. In Proceedings of the 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), Patras, Greece, 5–17 July 2019; pp. 1–7.
24. Ranzato, M.; Szlam, A.; Bruna, J.; Mathieu, M.; Collobert, R.; Chopra, S. Video (language) modeling: A baseline for generative models of natural videos. *arXiv* **2014**, arXiv:1412.6604.
25. Srivastava, N.; Mansimov, E.; Salakhudinov, R. Unsupervised learning of video representations using lstms. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 843–852.
26. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-c. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 802–810.
27. Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Deep learning for precipitation nowcasting: A benchmark and a new model. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5622–5632.
28. Wang, Y.; Long, M.; Wang, J.; Gao, Z.; Yu, P.S. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 879–888.
29. Wang, Y.; Gao, Z.; Long, M.; Wang, J.; Philip, S.Y. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 5123–5132.
30. Wang, Y.; Jiang, L.; Yang, M.-H.; Li, L.-J.; Long, M.; Fei-Fei, L. Eidetic 3D LSTM: A model for video prediction and beyond. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
31. Wang, Y.; Zhang, J.; Zhu, H.; Long, M.; Wang, J.; Yu, P.S. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9154–9162.
32. Guen, V.L.; Thome, N. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11474–11484.
33. Lotter, W.; Kreiman, G.; Cox, D. Deep predictive coding networks for video prediction and unsupervised learning. *arXiv* **2016**, arXiv:1605.08104.
34. Wang, Y.; Wu, H.; Zhang, J.; Gao, Z.; Wang, J.; Philip, S.Y.; Long, M. Predrnn: A recurrent neural network for spatiotemporal predictive learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 2208–2225. [\[CrossRef\]](#)
35. Chang, Z.; Zhang, X.; Wang, S.; Ma, S.; Ye, Y.; Xinguang, X.; Gao, W. Mau: A motion-aware unit for video prediction and beyond. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 26950–26962.
36. Gao, Z.; Tan, C.; Wu, L.; Li, S.Z. Simvp: Simpler yet better video prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 3170–3180.
37. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#)
38. Andersson, T.R.; Hosking, J.S.; Pérez-Ortiz, M.; Paige, B.; Elliott, A.; Russell, C.; Law, S.; Jones, D.C.; Wilkinson, J.; Phillips, T. Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nat. Commun.* **2021**, *12*, 5124. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Kim, Y.J.; Kim, H.-C.; Han, D.; Lee, S.; Im, J. Prediction of monthly Arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. *Cryosphere* **2020**, *14*, 1083–1104. [\[CrossRef\]](#)
40. Ren, Y.; Li, X. Predicting daily arctic sea ice concentration in the melt season based on a deep fully convolution network model. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 1–16 July 2021; pp. 5540–5543.
41. Ren, Y.; Li, X.; Zhang, W. A data-driven deep learning model for weekly sea ice concentration prediction of the pan-arctic during the melting season. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4304819. [\[CrossRef\]](#)

42. Wei, J.; Hang, R.; Luo, J.-J. Prediction of pan-arctic sea ice using attention-based LSTM neural networks. *Front. Mar. Sci.* **2022**, *9*, 860403. [\[CrossRef\]](#)
43. Liu, Q.; Zhang, R.; Wang, Y.; Yan, H.; Hong, M. Daily prediction of the arctic sea ice concentration using reanalysis data based on a convolutional lstm network. *J. Mar. Sci. Eng.* **2021**, *9*, 330. [\[CrossRef\]](#)
44. Zheng, Q.; Li, W.; Shao, Q.; Han, G.; Wang, X. A mid-and long-term arctic sea ice concentration prediction model based on deep learning technology. *Remote Sens.* **2022**, *14*, 2889. [\[CrossRef\]](#)
45. Liu, Y.; Bogaardt, L.; Attema, J.; Hazeleger, W. Extended-range arctic sea ice forecast with convolutional long short-term memory networks. *Mon. Weather Rev.* **2021**, *149*, 1673–1693. [\[CrossRef\]](#)
46. Chi, J.; Kim, H.-c. Prediction of arctic sea ice concentration using a fully data driven deep neural network. *Remote Sens.* **2017**, *9*, 1305. [\[CrossRef\]](#)
47. Mu, B.; Luo, X.; Yuan, S.; Liang, X. IceTFT v 1.0. 0: Interpretable Long-Term Prediction of Arctic Sea Ice Extent with Deep Learning. *Geosci. Model Dev. Discuss.* **2023**, *2023*, 1–28.
48. Serreze, M.C.; Barrett, A.P.; Crawford, A.D.; Woodgate, R.A. Monthly variability in Bering Strait oceanic volume and heat transports, links to atmospheric circulation and ocean temperature, and implications for sea ice conditions. *J. Geophys. Res. Ocean.* **2019**, *124*, 9317–9337. [\[CrossRef\]](#)
49. Choi, M.; De Silva, L.W.A.; Yamaguchi, H. Artificial neural network for the short-term prediction of arctic sea ice concentration. *Remote Sens.* **2019**, *11*, 1071. [\[CrossRef\]](#)
50. Wang, L.; Scott, K.A.; Clausi, D.A. Sea ice concentration estimation during freeze-up from SAR imagery using a convolutional neural network. *Remote Sens.* **2017**, *9*, 408. [\[CrossRef\]](#)
51. Chi, J.; Bae, J.; Kwon, Y.-J. Two-stream convolutional long-and short-term memory model using perceptual loss for sequence-to-sequence Arctic sea ice prediction. *Remote Sens.* **2021**, *13*, 3413. [\[CrossRef\]](#)
52. Zheng, Q.; Wang, R.; Han, G.; Li, W.; Wang, X.; Shao, Q.; Wu, X.; Cao, L.; Zhou, G.; Hu, S. A Spatio-Temporal Multiscale Deep Learning Model for Subseasonal Prediction of Arctic Sea Ice. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4300522.
53. Anderson, O.D.; Kendall, M.G. Time-Series, 2nd edn. *Statistician* **1976**, *25*, 308. [\[CrossRef\]](#)
54. Cleveland, R.; William, C.; Terpenning, I. STL: A seasonal-trend decomposition procedure based on loess. *J. Off. Stat.* **1990**, *6*, 3–73.
55. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
56. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
57. Liu, Y.; Shao, Z.; Hoffmann, N. Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv* **2021**, arXiv:2112.05561.
58. Tang, C.; Zhao, Y.; Wang, G.; Luo, C.; Xie, W.; Zeng, W. Sparse MLP for image recognition: Is self-attention really necessary? In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022; pp. 2344–2351.
59. DiGirolamo, N.E.; Parkinson, D.J.; Gloersen, C.P.; Zwally, H.J. *Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 2*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2022.
60. Tschudi, M.A.; Meier, W.N.; Stewart, J.S. An enhancement to sea ice motion and age products at the National Snow and Ice Data Center (NSIDC). *Cryosphere* **2020**, *14*, 1519–1536. [\[CrossRef\]](#)
61. Huntemann, M.; Heygster, G.; Kaleschke, L.; Krumpen, T.; Mäkynen, M.; Drusch, M. Empirical sea ice thickness retrieval during the freeze-up period from SMOS high incident angle observations. *Cryosphere* **2014**, *8*, 439–451. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.