IceDiff: High Resolution and High-Quality Arctic Sea Ice Forecasting with Generative Diffusion Prior

Jingyi Xu[1,*], Siwei Tu[1,*], Weidong Yang[1,†], Ben Fei[3,†],
Shuhao Li[1], Keyi Liu[1], Yeqi Luo[1], Lipeng Ma[1], Lei Bai[2]
[1]Fudan University, [2]Shanghai Artificial Intelligence Laboratory, [3]Chinese University of Hong Kong
jyxu22@m.fudan.edu.cn, wdyang@fudan.edu.cn,
benfei@cuhk.edu.hk, baisanshi@gmail.com
* Equal Contributions, † Corresponding Authors

Abstract

Variation of Arctic sea ice has significant impacts on polar ecosystems, transporting routes, coastal communities, and global climate. Tracing the change of sea ice at a finer scale is paramount for both operational applications and scientific studies. Recent pan-Arctic sea ice forecasting methods that leverage advances in artificial intelligence have made promising progress over numerical models. However, forecasting sea ice at higher resolutions is still under-explored. To bridge the gap, we propose a two-module cooperative deep learning framework, IceDiff, to forecast sea ice concentration at finer scales. IceDiff first leverages a vision transformer to generate coarse yet superior forecasting results over previous methods at a regular 25 km grid. This high-quality sea ice forecasting can be utilized as reliable guidance for the next module. Subsequently, an unconditional diffusion model pre-trained on low-resolution sea ice concentration maps is utilized for sampling down-scaled sea ice forecasting via a zero-shot guided sampling strategy and a patch-based method. For the first time, IceDiff demonstrates sea ice forecasting with a 6.25 km resolution. IceDiff extends the boundary of existing sea ice forecasting models and more importantly, its capability to generate high-resolution sea ice concentration data is vital for pragmatic usages and research. Code is available at https://github.com/EtronTech/IceDiff.

1. Introduction

Arctic sea ice is essential for global climate change and local communities. Although the exact teleconnections between mid-latitude weather and the variation of Arctic sea ice are yet to be revealed, the reduction of sea ice during the last few decades clearly has impacts that transcend the Arctic region [3, 41]. Hence, establishing skillful forecasting models at finer scales is paramount for both transportation in polar regions and scientific research in geoscience.

Existing physics-based numerical models have demonstrated the ability to accurately forecast sea ice concentration (SIC, value ranges from 0% to 100%) for short lead times of a few weeks. However, when forecasting at longer scales, their performance significantly decreases, and their forecasting skills are even worse than those of simple statistical models [3]. With the advent of deep learning methods, a series of works based on deep learning approaches have been committed to SIC forecasting with promising results. Notable studies cover forecasting leads range from daily to seasonal scale, e.g., IceNet [3] provides accurate sea ice extent (SIE, areas where the SIC value is greater than 15% and could be covered with sea ice) forecasts for 6 months lead time. SICNet [35] and SICNet$_{90}$ [34] are skillful for forecasting SIC in subsequent 7 and 90 days, respectively.

Despite the improved forecasting skills of deep learning models over previous approaches, these models are still trained on sea ice products provided by the National Sea Ice Data Center (NSIDC) or the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). These sea ice data products are primarily derived from passive microwave sensors with a resolution of 25 km, which is considered too coarse for sea ice modeling. The in situ observations are costly to acquire and extremely spare to provide sufficient spatial coverage over pan-Arctic region. This limitation hinders the representation of detailed information on sea ice concentration and the accuracy of ice edge [10, 16]. The same challenge is also presented in weather forecasts. Therefore, several endeavors are proposed to work on downscaling atmospheric states [5, 11, 43]. For instance, SwinRDM [11] combines an improved SwinRNN [21] model with a denoising diffusion probabilistic model (DDPM) [20] to achieve super-resolution (SR) in forecasting climate variables. However, forecasting sea ice at finer spatial scales, which is crucial, especially for operational applications is still under-explored.

Due to the absence of high-resolution sea ice data products and sparse in situ observations, it is not feasible to train a forecast model over pan-Arctic region at a finer scale. Two primary solutions can address this challenge: (i) Downscale sea ice maps initially and then train a forecast model using these high-resolution SIC maps. (ii) Train a forecast model at the original resolution based on the sea ice products, then downscale the forecast results to a finer scale. In the first solution, downscaling sea ice maps may introduce slight errors, which could accumulate during the forecasting stage. Additionally, training a forecast model on high-resolution SIC maps also imposes a greater computational burden. Therefore, we opt for the latter solution and introduce **IceDiff**, a two-module cooperative framework. IceDiff aims to provide skillful forecasts at three different temporal scales in the forecasting module (FM) and produce spatially downscaled high-quality SIC maps in the guided diffusion module (GDM). Specifically, in the first module, IceDiff leverages U-Net-based vision transformer and provides forecasts that are superior to existing SIC forecasting models. Then, in the second module, we utilize a well-trained DDPM on the original resolution from the NSIDC production as effective prior and downscale forecasted SIC maps guided by forecasts from the previous module at the original scale. Our DDPM is pre-trained on daily SIC maps with a resolution of 25 km, and it can also generalize well to weekly and monthly SIC maps. In every step t of the sampling process of DDPM, the intermediate variable x_0 is estimated from the denoised SIC map x_t. Then we propose a convolution kernel with optimizable parameters to simulate the up-scale degradation for the intermediate variable x_0, which results in up-scaled $\tilde{x}_0$. A distance function between the low-resolution SIC map and up-scaled $\tilde{x}_0$ is utilized to dynamically update the parameters of the convolution kernel in the reverse steps and guide the sampling of SIC map in the next time step. Furthermore, to make our IceDiff suitable for SIC maps with various resolutions, a patch-based method is also designed to downscale SIC maps at artistry resolution. We show in this work that using low-resolution SIC maps as guidance can better capture small-scale details in the generated high-resolution SIC maps. Our contributions are three folds:

- We propose a novel SIC forecasting framework IceDiff that demonstrates skills in three different temporal scales which cover subseasonal to seasonal forecasts, specifically, 7 days, weekly average of 8 weeks and monthly average of 6 months, that are commonly investigated by the community. IceDiff achieves state-of-the-art results in all three scales compared to existing deep learning models.
- IceDiff is devised with an optimizable convolutional kernel to simulate the up-scaling degradation and designs a patch-based method for downscaling SIC maps at arbitrary resolution with a single and fixed-resolution unconditional diffusion model. By utilizing the generative diffusion prior inherent in this diffusion model, the high-resolution SIC maps can be obtained through zero-shot sampling. Moreover, our IceDiff is capable of capturing small-scale details and generating high-resolution SIC maps.
- Extensive experiments show that our IceDiff could provide accurate SIC forecasts and generate high-quality sea ice samples with finer details. IceDiff not only provides a new and comprehensive baseline for data-driven sea ice research but also extends the capability of existing SIC forecasting models and pushes this area of research towards more operational usage.

2. Related Works

2.1. Deep Learning-based Sea Ice Forecasting

Deep learning models have drawn the attention of sea ice research communities and have been widely investigated for Arctic sea ice forecasting [1, 2, 22, 33]. As a pioneering work, ConvLSTM [24] predicts SIC within the region of Barents Sea and has showcased comparable results to the state-of-the-art numerical climate models. IceNet [3] employs the U-Net architecture for merging pre-selected ERA5 variables and SIC data to improve the accuracy of seasonal SIE prediction up to six months. The experimental results have shown that the forecasting skill of IceNet in pan-Arctic region is superior to the numerical sea ice simulation model SEAS5. MT-IceNet [1] adopts a similar approach to IceNet and adds bi-monthly data to further improve forecasting SIC at seasonal scale. Mu et al. [30] propose to merge atmospheric and oceanographic data in a weighted manner through a variable selection network and provide nine months of forecasting leads. Considering the complex ice-ocean-atmosphere interactions have already embedded in the spatiotemporal correlations of consecutive SIC maps [33, 35], Ren et al. [35] proposed SICNet that solely rely on SIC data to predict concentration for the next seven days. Similar to IceNet, SICNet is built upon U-Net architecture and added with channel-wise dependency via temporal-spatial attention module (TSAM), which is essentially a revised version of the convolutional block attention module. By fusing the output of two SICNet, i.e. one for modeling SIC and the other for extracting climate information, and adjusting the number of output channels, Ren et al. [34] extends the lead time to 90 days. Besides CNN and LSTM-based models, recent works leverage the latest advances in artificial intelligence to further improve forecasting performance. For instance, Ice-

Former [42] jointly models SIC and sea ice thickness (SIT) based on Transformer backbone. It first decomposes high dimensional spatio-temporal data into sequences via multivariate empirical orthogonal functions [37, 38]. Then an encoder-decoder Transformer is utilized to predict SIC and SIT for up to 45 days. Despite the performance gain of the aforementioned deep learning models, none of them have addressed forecasting sea ice at scales lower than 25 km.

2.2. Super Resolution

Super-resolution aims to reconstruct high-resolution (HR) images from low-resolution (LR) counterparts. Several works [27, 28] explore the use of Variational Autoencoders for image SR. These methods usually improve the restoration quality by learning the conditional distribution of high-resolution images induced by low-resolution images. Although these methods are able to achieve realistic perceptual quality, VAE-based methods are often not as satisfactory due to their limited generation capabilities. To this end, other deep learning methods, such as Convolutional neural networks (CNNs), have been widely used for super-resolution tasks in various domains. CNN-based methods typically use HR images corresponding to LR images to supervise training with reconstruction loss [4, 7, 36, 44]. However, these network trained solely on reconstruction loss makes it difficult to capture detailed information in images and generate diverse image results. Therefore, Generative Adversarial Networks (GANs) are proposed to solve these challenges. Stengel et al. [39] utilized GAN in a two-step SR process to achieve downscaling of climatological wind and solar irradiance. Leinonen et al. [23] introduced a recurrent, stochastic SR GAN for down-scaling time-evolving atmospheric fields, demonstrating the potential of GANs in improving the spatial resolution of LR images in atmospheric sciences applications. These models encourage the generator to generate high-quality images that the discriminator cannot distinguish from real images. Although GAN-based methods obtain satisfactory downscaled results, the diversity could be enhanced.

Currently, diffusion model-based methods have been more widely used due to their diversity in generating images and the ability to generate high-quality images comparable to GANs [17, 40]. For instance, Chen et al. [11] utilize SR based on the conditional model to recover the high spatial resolution and finer-scale atmospheric details, unfolding the ability of diffusion models to generate high-quality and detailed images in SR tasks. The objective of this study is to forecast SIC maps at a high resolution of 6.25 km. To achieve this, we have improved the SwinTransformer for predicting SIC at the standard resolution comparable to current methods. Additionally, we employ a diffusion-based SR model to generate high-resolution and high-quality results via zero-shot sampling.

3. Methodology

Our IceDiff framework comprises two modules as depicted in Figure 1. The forecasting module provides skillful forecasts at three different temporal scales. The guided diffusion module produces spatially down-scaled high-quality SIC maps with a pre-trained DDPM via zero-shot sampling and the patch-based method. We elaborate on design details in the following sections.

3.1. Forecasting Module

To achieve more accurate SIC forecasting at a finer resolution, a skillful forecasting module trained on original sea ice data products is primarily encouraged. In contrast to existing SIC forecasting methods that primarily concentrate on a single temporal scale [1, 3, 42], our objective is to enhance forecasting skill across various temporal scales and increase versatility. Encouraged by the success of Swin Transformer [25] in various data-driven weather forecasting methods [6, 11, 12] , we propose to construct forecasting module by utilizing the U-Net architecture [8] and building on top of the Swin Transformer V2 backbone [26] with ResNet$_{TSAM}$ block [35] for better capturing both spatial and temporal features of SIC maps. As illustrated in Figure 1, the FM takes N consecutive SIC maps, each of which has a shape of $[448, 304]$, to predict subsequent maps with identical temporal length as the input at 25 km spatial scale. It consists of *encoder*, *bottleneck* and *decoder*. The patch embedding layer first applies a 2×2 windows size to partition the original SIC input into patches and maps them into tokens via 2D convolution. Then, the temporal and spatial dependencies are modeled through the ResNet$_{TSAM}$ block. Both the encoder and the decoder have 3 layers of Swin Transformers that comprise a pre-configured even number of Swin Transformer Blocks for feature extraction and restoration. The calculation of two consecutive blocks can be described as follows:

$$\begin{aligned}
z_s^b &= LN(WMSA(z^{b-1})) + z^{b-1}, \\
z^b &= LN(MLP(z_s^b)) + z_s^b, \\
z_s^{b+1} &= LN(SWMSA(z^b)) + z^b, \\
z^{b+1} &= LN(MLP(z_s^{b+1})) + z_s^b,
\end{aligned} \tag{1}$$

where z_s^b and z^b represents the output feature of the (**S**hifted) **W**indow-**M**ulti-head **S**elf **A**ttention module and the MLP module for block b, respectively [26]. LN denotes the LayerNorm operation. The combination of such two consecutive blocks, i.e. the former for the employment of WMSA and the latter for SWMSA, ensures that local patterns of adjacent grids are sufficiently attended by the backbone network. We further adopt a shift window size of 28 x 19 that is consistent with the ratio of SIC input map for facilitating the SWMHA mechanism. The output of the

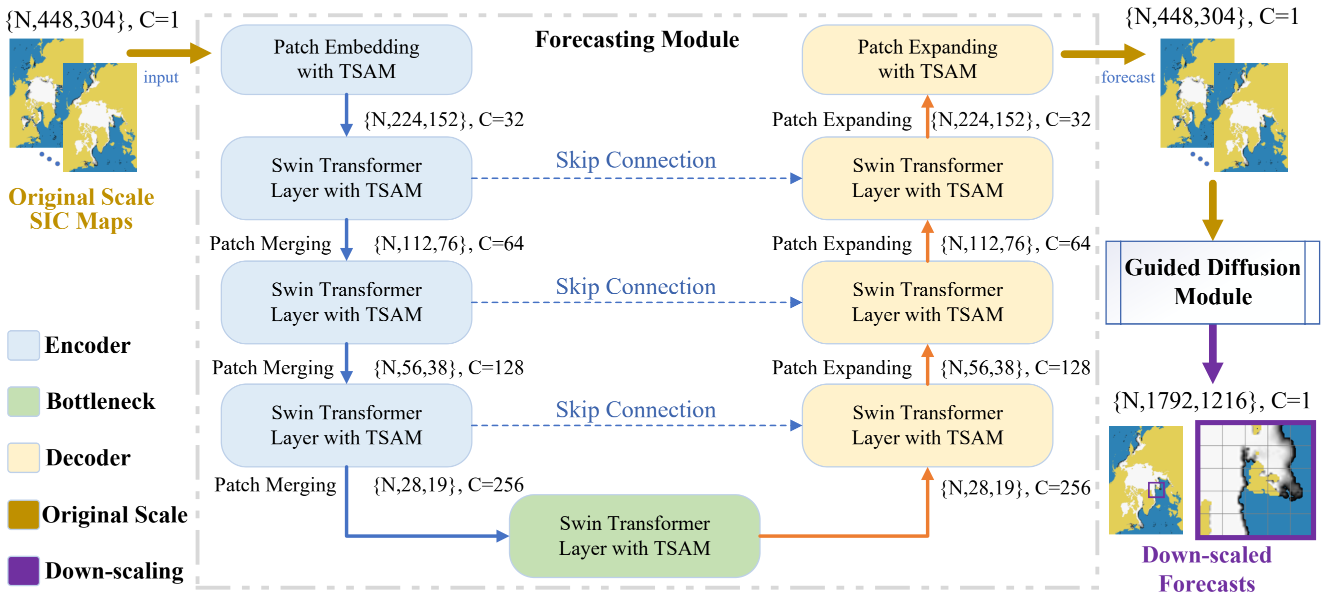


Figure 1. **Overview of IceDiff, a two-module cooperative framework to forecast sea ice concentration at a finer scale.** IceDiff first employs the forecasting module (FM) that takes history SIC data as input to generate subsequent predictions at 25 km resolution. N denotes the number of sequential time stamps, for instance, set N to 7 for seven days of forecasting leads, FM takes a tensor with the shape of $\{7, 448, 304\}$ as input and forecasts the SIC for the next 7 days. C represents the quantity of embedded channels within each layer of FM, the original SIC map data has a channel of 1 rather than 3 for regular images. The guided diffusion module (GDM) then leverages accurate forecasts as guidance to generate downscaled samples at 6.25 km resolution with high quality, see details in Figure 2.

last block within each layer is delivered to a ResNet$_{TSAM}$ block for establishing channel-wise dependencies. Before the subsequent encoder and decoder layers, patch merging and patch expanding [8] are respectively performed to reshape the feature and fit it to the desired size. Following prior works [3, 8, 35], the skip connection is added between the corresponding encoder and decoder layers to improve the performance. The final prediction is generated via a patch expanding layer with ResNet$_{TSAM}$ that up-samples the decoded feature to the size of the original SIC input. Our meticulous designs enable the forecasting module in IceDiff to be applied across various temporal scales, enhancing forecasting skills, which serves as a more effective initiation for the guided diffusion module of IceDiff.

3.2. Guided Diffusion Module

Off-the-shelf methods are limited to the forecasting of SIC maps at a scale of 25 km from sea ice data products [1, 3, 35]. However, finer-scale SIC maps remain unexplored which is of significant value to ecosystems, transporting routes, coastal communities, and global climate. Therefore, after obtaining skillful forecasts, we exploit a downscaling module based on an unconditional diffusion model pre-trained on daily SIC maps to generate realistic high-resolution SIC maps with rich details. This guided diffusion module pre-trained on daily SIC maps could generalize well to SIC maps at various scales, such as weekly or monthly SIC maps. Moreover, our guided diffusion module of IceDiff downscales the SIC forecasts from a forecasting module to a scale of 6.25 km, further expanding the accuracy and practicality of sea ice prediction.

Specifically, the reverse process of the diffusion model is conditioned on the low-resolution sea ice map y, which transforms distribution $p_\theta(x_{t-1}|x_t)$ into conditional distribution $p_\theta(x_{t-1}|x_t, y)$. Previous work [15] has derived the conditional transformation formula in the reverse process:

$$\log p_\theta(x_t|x_{t+1}, y) = \log\left(p_\theta(x_t|x_{t+1})p(y|x_t)\right) + N_1 \quad (2)$$
$$\approx \log p_\theta(z) + N_2 \quad , z \sim \mathcal{N}(z; \mu_\theta(x_t, t) + \Sigma g, \Sigma I), \quad (3)$$

where g refers to $\nabla_{x_t} \log p(y|x_t)$. $N_1 = \log p_\theta(y|x_{t+1})$, N_2 is a constant related to g (proof in Appendix). The variance of the reverse process $\Sigma = \Sigma_\theta(x_t)$ is constant.

Therefore, the sampling process of $p_\theta(x_{t-1}|x_t, y)$ integrates the gradient term g with the diffusion model $(\mu, \Sigma) = (\mu_\theta(x_t), \Sigma_\theta(x_t))$, which plays a role in controlling the direction of map generation. We use a heuristic algorithm (proof in Appendix) to approximate the value of g:

$$\log p(y \mid x_t) = -\log N - s\mathcal{L}(\mathcal{D}(\tilde{x}_0), y) \quad (4)$$
$$g = \nabla_{x_t} \log p(y|x_t) = -s\nabla_{x_t}\mathcal{L}(\mathcal{D}(x_t), y). \quad (5)$$

Among them, s is the scaling factor used to control the degree of guidance. And $\mathcal{L}$ is a distance function used to measure the distance deviation between two SIC maps. $\mathcal{D}$ is an up-scale convolution function with optimizable parameters, where the parameters are randomly initialized and dynamically optimized with the gradient of distance function at

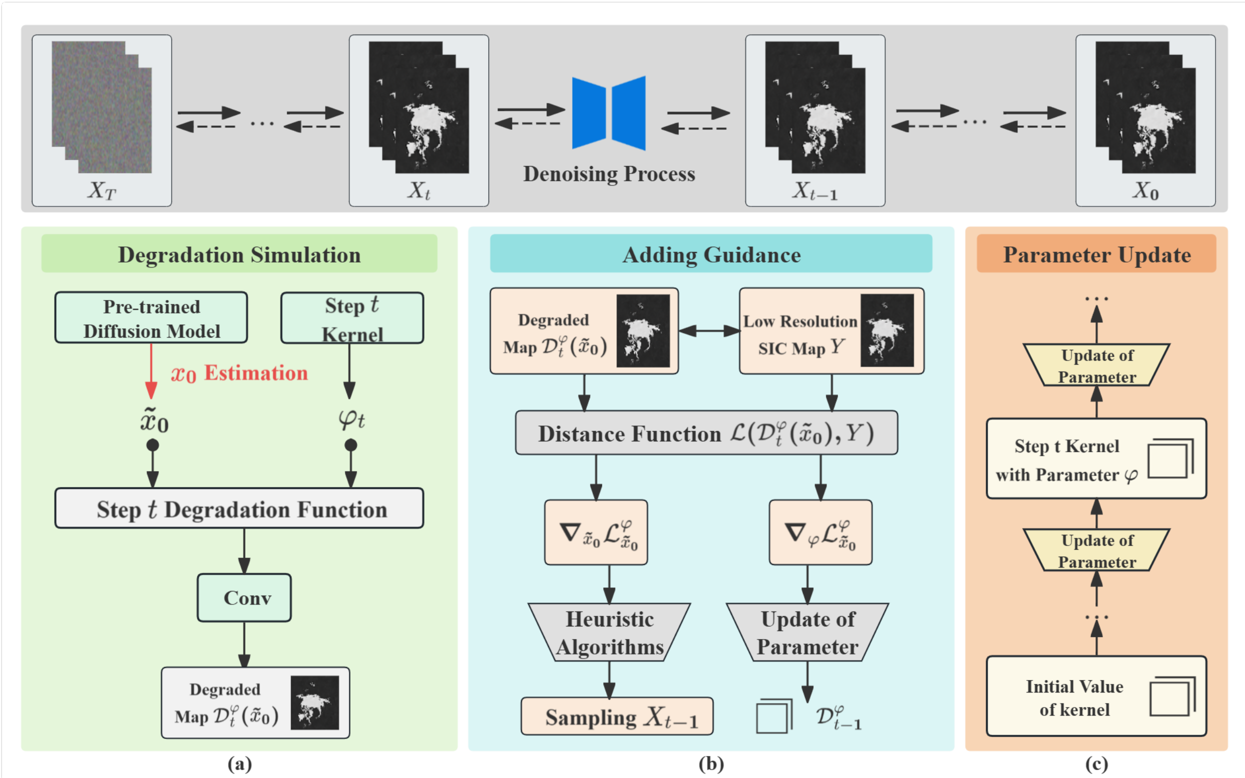


Figure 2. **Overview of guided diffusion module in IceDiff for SIC map downscaling. (a)** Pre-trained diffusion model is used to estimate $\tilde{x}_0$ at reverse step t. The optimizable convolutional kernel of step t reduces the resolution of X_t so that guidance from low-resolution SIC maps can be incorporated to sample X_{t-1}. **(b)** Low-resolution SIC map Y is integrated during the sampling process. A distance function describes the distance between a low-resolution SIC map and the generated high-resolution SIC map x after convolution. The gradient of distance function could be used to guide the sampling of x_{t-1} and **(c)** dynamically update the parameters of the convolutional kernel.

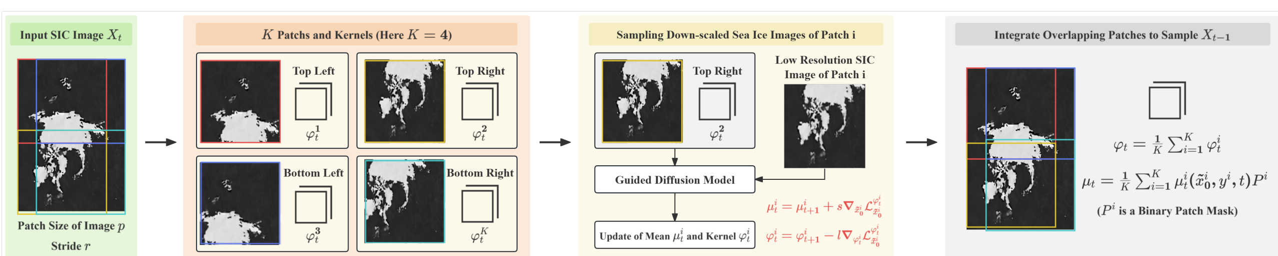


Figure 3. **Illustration of the patch-based method.** We set patch size $p = 256$ and stride $r = 128$ to fit the size of pre-trained models. This setting will divide the example map into $K = 4$ sub-maps boxed with different colors. For each sub-map, an individual optimizable convolution kernel with adaptive parameters is set to add guidance from the corresponding low-resolution SIC sub-map. At each sampling step t, the average estimated noise on four overlapping blocks will be utilized to update the mean μ used to sample x_{t-1}.

every reverse step. $\mathcal{D}$ connects two different resolution SIC maps, allowing low-resolution SIC maps to provide more effective guidance.

As shown in Algorithm 1, GDM undergoes T reverse steps to gradually restore pure Gaussian noise $x_T \sim \mathcal{N}(0, I)$ to high-resolution SIC maps. In each reverse step t, the diffusion model calculates the estimated variable $\tilde{x}_0$, which then undergoes the up-scale convolution function $\mathcal{D}$ at time t to establish a distance metric $\mathcal{L}$ with the low-resolution SIC map, where $\mathcal{D}_t^{\varphi}$ signifies that its convolutional kernel parameters are $\varphi = \varphi_t$ at the reverse steps t. The $\mathcal{L}$ subscript $\varphi = \varphi_t$ indicates that it is a function dependent on the convolutional kernel parameter φ at reverse step t. The gradients of distance function $\mathcal{L}$ with respect to $\tilde{x}_0$ and the convolution kernel parameters of the up-scale convolution function are used to update $\tilde{x}_0$ and the convolution kernel parameters, respectively. The update of convolution kernel parameters ensures the accuracy of the model's

Table 1. **Quantitative results of forecasts**. IceDiff is evaluated on three different temporal scales. The output length of IceDiff equals a pre-defined forecasting lead as depicted in Figure 1. The validated accuracy of IceDiff could provide reliable guidance for the guided diffusion module to generate high-quality downscaled SIC. BT represents that the SIC data [14] are generated by the Bootstrap retrieval algorithm [13] and the data distribution of CDR is considered to be closer to BT rather than NT [29].

Temporal Scale	Lead Time	Dataset	Methods	RMSE↓	MAE↓	R^2↑	NSE↑	IIEE↓	SIE_{dif}↓
Subseasonal	7 Days (Daily)	NT	SICNet [35]	0.0430	0.0091	0.978	0.975	1175	0.1127
		NT	**IceDiff**	**0.0324**	**0.0077**	**0.987**	**0.986**	**899**	**0.0186**
		CDR	SICNet [35]	0.0490	0.0100	0.982	0.979	976	0.0718
		CDR	**IceDiff**	**0.0396**	**0.0080**	**0.989**	**0.987**	**835**	**0.0315**
	45 Days (Daily)	BT	IceFormer [42]	0.0660	0.0201	0.960	-	-	-
	8 Weeks Average (Weekly)	NT	**IceDiff**	0.0403	0.0108	0.975	0.972	1170	0.1051
		CDR	**IceDiff**	0.0553	0.0112	0.973	0.969	1353	0.0871
	90 Days (Daily)	NT	SICNet_{90} [34]	-	0.0512	-	-	-	-
Seasonal	6 Months Average (Monthly)	NT	IceNet [3]	0.1820	0.0916	0.567	-	-	-
		NT	MT-IceNet [1]	0.0777	0.0197	0.915	-	-	-
		NT	**IceDiff**	**0.0573**	**0.0188**	**0.958**	**0.952**	**2014**	**0.2419**
		CDR	**IceDiff**	0.0648	0.0168	0.919	0.913	2016	0.2657

Algorithm 1 Guided diffusion module in IceDiff with the guidance of low-resolution SIC map y. An unconditional diffusion model pre-trained on SIC maps $\epsilon_\theta(x_t, t)$ is given.

Input: Low-resolution SIC map y, downscale convolution function $\mathcal{D}$ composed of optimized convolutional kernels $\mathcal{K}$ with parameters φ, learning rate l, guidance scale s, distant function $\mathcal{L}$.

Output: Output high-resolution SIC map x_0 conditioned on y.

Sample x_T from $\mathcal{N}(0, I)$

for all t from T to 1 **do**

$\tilde{x}_0 = \frac{x_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t,t)}{\sqrt{\bar{\alpha}_t}}$

$\mathcal{L}_{\varphi,\tilde{x}_0} = \mathcal{L}(y, \mathcal{D}^\varphi(\tilde{x}_0))$

$\tilde{x}_0 \leftarrow \tilde{x}_0 - \frac{s(1-\bar{\alpha}_t)}{\sqrt{\bar{\alpha}_{t-1}\beta_t}} \nabla_{\tilde{x}_0} \mathcal{L}_{\varphi,\tilde{x}_0}$

$\tilde{\mu}_t = \frac{\sqrt{\bar{\alpha}_{t-1}\beta_t}}{1-\bar{\alpha}_t}\tilde{x}_0 + \frac{\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t$

$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$

Sample x_{t-1} from $\mathcal{N}(\tilde{\mu}_t, \tilde{\beta}_t I)$

$\varphi \leftarrow \varphi - l\nabla_\varphi \mathcal{L}_{\varphi,\tilde{x}_0}$

end for

return x_0

up-scaling process, making guidance of low-resolution SIC maps more precise and effective.

Patch-based Strategy. Furthermore, pre-trained models can only generate fixed-size SIC maps. To tackle the problem of sampling downscaled SIC forecasts at any resolution, we employ a patch-based strategy. As shown in Figure 3, the patch-based strategy partitions SIC maps into several sub-regions based on a fixed stride. For each individual sub-region, the gradient terms of distance loss between SIC maps upscaled by the convolutional kernel and the corresponding low-resolution SIC sub-region maps are computed individually. Subsequently, the respective mean of Gaussian distribution in sub-regions (Eq. 3) and convolution kernel parameters are updated based on the gradient terms. The entire map is updated by taking the mean of the updated values across all sub-regions, each weighted by a binary patch mask that serves as a quantifier of the regional scope, thereby refining the overall mean and convolution kernel parameters of the sampled high-resolution SIC maps. By leveraging this strategy, our model is capable of downscaling sea ice maps to the desired resolution and the practicality of the model has been further enhanced.

Table 2. **Quantitative comparison between IceDiff for SIC downscaling and other methods**. Const. and FID represents the Consistency and Fréchet Inception Distance metrics, respectively.

Method	7 Days		8 Week Avg.		6 Month Avg.	
	Const.↓	FID↓	Const.↓	FID↓	Const.↓	FID↓
Nearest	20.01	78.45	21.98	80.17	23.07	88.36
Bilinear	17.88	70.75	20.02	77.29	18.31	80.73
BiCubic	13.13	54.71	13.40	68.85	10.09	70.18
DGP [32]	23.18	75.42	28.62	81.42	29.74	85.47
GDP [17]	8.26	40.15	10.54	52.40	10.96	55.72
IceDiff	**6.84**	**29.34**	**7.25**	**31.29**	**6.62**	**39.79**

4. Experiments

4.1. Dataset and Baseline Setup

We evaluate our proposed IceDiff framework on the Climate Data Record of Passive Microwave Sea Ice Concentration Version 4 dataset from NSIDC [29] (CDR in Table 1). It records daily SIC data starting from October 25^{th} 1978 and provides the coverage of the pan-Arctic region (N:-39.36°, S:-89.84°, E:180°, W:-180°). The SIC map is formed of 448 × 304 pixels and each of which represents an area of 25 km × 25 km grid and has a value ranging from 0% to 100%. We choose data from October 25^{th} 1978 to 2013 as the training dataset, the years 2014 and 2015 are selected as validation set, and data from 2016 to 2023 are used for testing. To comprehensively validate and verify the performance of our forecasting module, we utilize additional data [16] that are generated through the retrieval

Table 3. **Ablation study on TSAM.** We compare the forecasting results between IceDiff with or without the TSAM.

Lead Time	TSAM	RMSE↓	MAE↓	R^2↑	NSE↑	IIEE↓	SIE_{dif}↓
7 Days	✗	0.0403	0.0082	0.9881	0.9860	874	0.0336
	✓	**0.0396**	**0.0080**	**0.9885**	**0.9865**	**835**	**0.0315**
8 Week Avg.	✗	**0.0535**	0.0116	**0.975**	**0.971**	1386	0.1420
	✓	0.0553	**0.0112**	0.973	0.969	**1353**	**0.0871**
6 Month Avg.	✗	0.0763	0.0194	0.883	0.875	2610	0.4919
	✓	**0.0648**	**0.0168**	**0.919**	**0.913**	**2016**	**0.2657**

algorithm (NASA Team [9], NT in Table 1). The forecasting module is trained on three different temporal scales, i.e. the input of SIC maps are 7 days, 8 weeks average, and 6 months average, covering forecasting leads from subseasonal to seasonal scale. Corresponding deep learning-based models are selected for comparison, as illustrated in Table 1. Note that we follow the split of datasets in respective baselines to preprocess the additional data. We re-implemented SICNet (the source code is not available) based on the original paper and trained it under an environment identical to our forecasting module. To evaluate the performance of our downscaling, five methods are selected to set up baselines for downscaling quality comparison.

4.2. Implementation Details

We independently trained the forecasting module of IceDiff on SIC dataset of three aforementioned temporal scales for 60 epochs. For each temporal scale, the forecasting module outputs the same length of input data as forecasting leads, i.e. 7 days, 8 weeks average, and 6 months averages. The dimensions of embedding features of encoder, decoder, and bottleneck are all set to 32. We utilize the same hyperparameter of TSAM as in [35]. The forecasting module is optimized by AdamW under the mean square error (MSE) objective using Pytorch on one NVIDIA A100 80GB GPU for all experiments with a learning rate of 0.0005. All daily SIC maps are resized and cropped to 256 $\times$ 256 to train our GDM of IceDiff. using AdamW with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We train in 16-bit precision using loss-scaling, but maintain 32-bit weights, EMA, and optimizer state. We use an EMA rate of 0.9999. When sampling with 1000 timesteps, we use the same noise schedule as for training.

4.3. Evaluation Metrics

To quantitatively evaluate our IceDff, we select commonly used root mean square error (RMSE) and mean absolute error (MAE) for the comparison of forecasting accuracy. R^2 score and Nash-Sutcliffe Efficiency [31] are utilized to analyze the similarity of spatial pattern between the ground truth and predicted SIC map. We adopt Integrated Ice-Edge Error score [18] and calculate the sea ice area differences (SIE_{dif}), in millions of km^2) to further evaluate the prediction of SIE (where the SIC value is greater than 15%).

As to the evaluation of downscaling quality, following [11] we utilize the Fréchet Inception Distance (FID) [19] for measuring the downscaling fidelity. In addition, we adopt Consistency to quantify the faithfulness between the downscaled map and the low-resolution map.

The detailed equations for calculating the abovementioned metrics can be found in Appendix.

Table 4. **Ablation study on the ratio of Input/Forecast number of SIC maps**. Hf., Db. and Eq. represent that the forecasting module takes input data which has *half*, *double*, and *equal* length of forecasting leads, respectively. For the Hf. of 7 Days, we set the input length to 3.

Lead Time	I./F.	RMSE↓	MAE↓	R^2↑	NSE↑	IIEE↓	SIE_{dif}↓
7 Days	Hf.	0.0416	0.0084	0.987	0.985	905	0.0538
	Db.	0.0401	0.0084	0.988	0.986	889	0.0599
	Eq.	**0.0396**	**0.0080**	**0.989**	**0.987**	**835**	**0.0315**
8 Week Avg.	Hf.	0.0580	0.0118	0.970	0.966	1405	0.1176
	Db.	0.0594	0.0121	0.969	0.964	1435	0.1221
	Eq.	**0.0553**	**0.0112**	**0.973**	**0.969**	**1353**	**0.0871**
6 Month Avg.	Hf.	0.0904	0.0232	0.842	0.830	3192	0.6289
	Db.	0.0722	0.0179	0.897	0.889	2027	0.2795
	Eq.	**0.0648**	**0.0168**	**0.919**	**0.913**	**2016**	**0.2657**

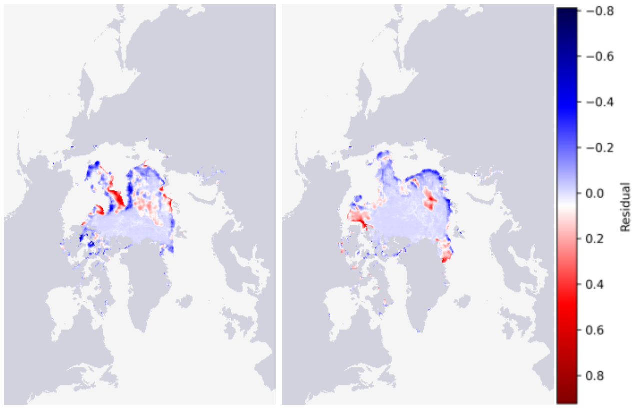
4.4. Main Results

In this subsection, we evaluate the proposed IceDiff with baseline methods for validation of the effectiveness.

Set up new baselines for SIC prediction. The forecasting module of IceDiff is trained independently to predict SIC at 7 days, 8 weeks average, and 6 months average. As shown in Table 1, the performance of our proposed forecasting module of IceDiff is superior to SICNet in all evaluation metrics. The forecasting skill of FM across different SIC datasets is consistent in three temporal scales. Specifically, it outperforms SOTA seasonal forecasting baselines by a noticeable margin. Considering the overall performance of our FM, it could provide reliable SIC forecasts and sufficiently support the subsequent downscaling module.

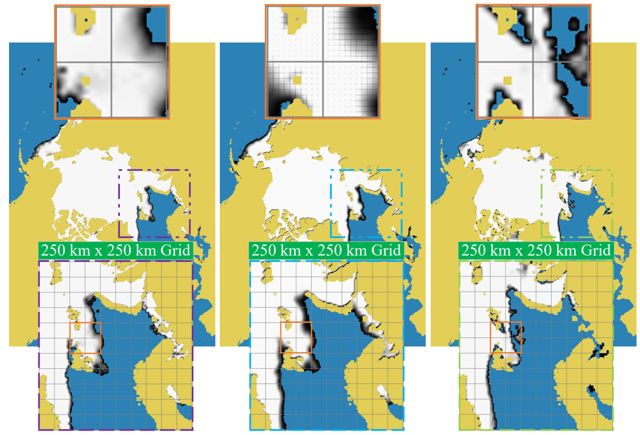
Analysis of the end of the melting season. To further verify the performance of IceDiff in extreme cases, we select two years of the end of the melting season (early September) when the mean Arctic SIC typically reaches its minimum for validation: (i) In 2016, it recorded historically low annual SIC; (ii) In 2022, the annual SIC has an abrupt increase compared to the previous year (annual trends are plotted in the Appendix). We present the SIC residuals (predicted SIC minus the ground truth) of September 1^{st} 2016 and 2022 in Figure 4(a) and Figure 4(b), respectively. The overall residuals are relatively low and the prediction errors mainly appear in part of the sea ice edge regions. The accurate forecasting results in temporal periods with highly varying SIC validate the effectiveness of IceDiff.

High downscaling quality. We systematically compare the results of IceDiff with other downscaling methods. In terms of interpolation-based methods, we select nearest, bilinear, and bicubic interpolation to compare with our diffusion-based GDM. Besides, we select DGP [32] as a GAN-based method and GDP [17] as a DDPM-based method for fur-



(a) Forecast Date Sep 1, 2016 (b) Forecast Date Sep 1, 2022

Figure 4. **Qualitative analysis of the end of the melting season.** The annual mean Arctic SIC has recorded a historical low in 2016 and it drastically increased in 2022 when compared to the previous annual mean. The residual plots of early September in **(a)**2016 and **(b)**2022 demonstrate where the FM overestimates (positive figures) or underestimates (negative figures) SIC.



(a) IceDiff **(b) Bicubic** **(c) GDP**

Figure 5. **Qualitative analysis of downscaled SIC map.** We visualize the SIC map by coloring the land and the ocean in yellow and blue, respectively. Since SIC has a range of $[0, 1]$, we map it to $[0, 255]$ and replicate it for all 3 RGB channels. The darker the pixels appear, the SIC values are more closer to 0%. For better identification of the difference, we crop out a region and plot a grid size of 250 km and further display details on 4×4 grid.

ther comparison. Interpolation-based methods upsample the low-resolution sea ice map, which uses existing pixel information to estimate missing pixel values.

As shown in Figure 5, our GDM produces high-quality results and small-scale structures are successfully captured and rich details are generated through the guidance from the input map during each sampling step. By contrast, interpolation-based methods do not integrate details information of the low-resolution map, resulting in blurry edges and contours in the generated results. More quality analysis can be found in the appendix.

To further validate the superiority of our method, we follow SwinRDM [11] and GDP [17] to utilize FID and Consistency to evaluate our method, respectively. As shown in Table 2, compared to interpolation-based methods, our IceDiff improves FID score by 31.11 on average, reflecting it has the capacity to help generate a more realistic high-resolution map with a satisfactory texture restoration effect. A lower Consistency demonstrates the structural difference between the downscaled map and the original SIC input is relatively imperceptible, which better preserves the spatial distribution of sea ice forecast.

4.5. Ablation Studies on Design of FM

As shown in Table 3, by adopting TSAM to FM, the performance has improved for all metrics at 7 days and 6 months average lead times while the effects for 8 weeks average lead time are not all positive. Overall, incorporating TSAM to establish channel-wise dependency is favorable for SIC forecast models. To examine the impacts of input SIC data length, we trained FM for input lengths of half, double, and equal to the forecasting leads. The results in Table 4 show the peak of the performance at the ratio where the input length equals the forecasting leads.

5. Conclusion and Limitations

We propose a novel pan-Arctic high-resolution SIC forecasting framework, IceDiff, which comprises a forecasting module and a guided diffusion module, these two modules closely cooperate with each other to fulfill accurate forecasting of SIC at a finer and more practical spatial scale. The forecasting module of IceDiff is capable of producing accurate forecasts at three different temporal scales and the guided diffusion module leverages reliable forecasts as guidance to generate high-quality down-scaled SIC maps. The forecasting skills of IceDiff at 7 days lead are superior to SOTA deep learning-based models. IceDiff also sets up a new baseline for weekly average forecasting of 8 weeks and provides competitive results in seasonal SIC forecasts. To the best of our knowledge, IceDiff is the first attempt to forecast SIC at 6.25 km spatial scale which is a quarter of the grid size (25 km) that current methods perform on. The experimental results show that our framework can generate down-scaled SIC superior to off-the-shelf methods and it is vital for pushing forecasting models towards operational usage and further promoting sea ice research.

Existing sea ice concentration data products are all retrieved from satellite observations with spatial resolution limited to 25 km. A finer-scale ground truth is missing compared with observations from stations in weather forecasting. Therefore, our limitation lies in the evaluations of IceDiff on these in situ observations. We will focus on evaluating our IceDiff on higher-resolution SIC observations.

References

[1] Sahara Ali and Jianwu Wang. Mt-icenet-a spatial and multi-temporal deep learning model for arctic sea ice forecasting. In *2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*, pages 1–10. IEEE, 2022. 2, 3, 4, 6

[2] Sahara Ali, Yiyi Huang, Xin Huang, and Jianwu Wang. Sea ice forecasting using attention-based ensemble lstm. *arXiv preprint arXiv:2108.00853*, 2021. 2

[3] Tom R Andersson, J Scott Hosking, María Pérez-Ortiz, Brooks Paige, Andrew Elliott, Chris Russell, Stephen Law, Daniel C Jones, Jeremy Wilkinson, Tony Phillips, et al. Seasonal arctic sea ice forecasting with probabilistic deep learning. *Nature communications*, 12(1):5124, 2021. 1, 2, 3, 4, 6

[4] Saeed Anwar, Salman Khan, and Nick Barnes. A deep journey into super-resolution: A survey. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020. 3

[5] Seth Bassetti, Brian Hutchinson, Claudia Tebaldi, and Ben Kravitz. Diffesm: Conditional emulation of temperature and precipitation in earth system models with 3d diffusion models. *Journal of Advances in Modeling Earth Systems*, 16(10): e2023MS004194, 2024. 1

[6] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970): 533–538, 2023. 3

[7] Bruno Buongiorno Nardelli, Davide Cavaliere, Elodie Charles, and Daniele Ciani. Super-resolving ocean dynamics from space with computer vision algorithms. *Remote Sensing*, 14(5):1159, 2022. 3

[8] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022. 3, 4

[9] D.J. Cavalieri, P. Gloersen, and W.J. Campbell. Determination of sea ice parameters with the nimbus-7 smmr. *Journal of Geophysical Research*, 89:5355–5369, 1984. 7

[10] D.J. Cavalieri, P. Gloersen, and W.J. Campbell. Deriving long-term time series of sea ice cover from satellite passive-microwave multisensor data sets. *Journal of Geophysical Research*, 104:15803–15814, 1999. 1

[11] Lei Chen, Fei Du, Yuan Hu, Zhibin Wang, and Fan Wang. Swinrdm: integrate swinrnn with diffusion model towards high-resolution and high-quality weather forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 322–330, 2023. 1, 3, 7, 8

[12] Lei Chen, Xiaohui Zhong, Feng Zhang, Yuan Cheng, Yinghui Xu, Yuan Qi, and Hao Li. Fuxi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, 6(1):190, 2023. 3

[13] J.C. Comiso. Characteristics of arctic winter sea ice from satellite multispectral microwave observations. *Journal of Geophysical Research*, 91:975–994, 1986. 6

[14] Josefino C. Comiso. Bootstrap sea ice concentrations from nimbus-7 smmr and dmsp ssm/i-ssmis, version 3. 2017. 6

[15] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 4

[16] N. DiGirolamo, C.L. Parkinson, D.J. Cavalieri, P. Gloersen, and H.J. Zwally. Sea ice concentrations from nimbus-7 smmr //and dmsp ssm/i-ssmis passive microwave data. (nsidc-0051, version 1). *National Snow and Ice Data Center Distributed Active Archive Center.*, 1996. 1, 6

[17] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9935–9946, 2023. 3, 6, 7, 8

[18] Helge F Goessling, Steffen Tietsche, Jonathan J Day, Ed Hawkins, and Thomas Jung. Predictability of the arctic sea ice edge. *Geophysical Research Letters*, 43(4):1642–1650, 2016. 7

[19] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. 7

[20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1

[21] Yuan Hu, Lei Chen, Zhibin Wang, and Hao Li. Swinvrnn: A data-driven ensemble forecasting model via learned distribution perturbation. *Journal of Advances in Modeling Earth Systems*, 15(2):e2022MS003211, 2023. 1

[22] Young Jun Kim, Hyun-Cheol Kim, Daehyeon Han, Sanggyun Lee, and Jungho Im. Prediction of monthly arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. *The Cryosphere*, 14(3): 1083–1104, 2020. 2

[23] Jussi Leinonen, Daniele Nerini, and Alexis Berne. Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7211–7223, 2020. 3

[24] Yang Liu, Laurens Bogaardt, Jisk Attema, and Wilco Hazeleger. Extended-range arctic sea ice forecast with convolutional long short-term memory networks. *Monthly Weather Review*, 149(6):1673–1693, 2021. 2

[25] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 3

[26] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022. 3

[27] Zhi-Song Liu, Wan-Chi Siu, and Yui-Lam Chan. Photo-realistic image super-resolution via variational autoencoders.

IEEE Transactions on Circuits and Systems for video Technology, 31(4):1351–1365, 2020. 3

[28] Zhi-Song Liu, Wan-Chi Siu, Li-Wen Wang, Chu-Tak Li, and Marie-Paule Cani. Unsupervised real image super-resolution via generative variational autoencoder. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 442–443, 2020. 3

[29] W.N. Meier, F. Fetterer, A.K. Windnagel, and J.S. Stewart. Noaa/nsidc climate data record of passive microwave sea ice concentration. (g02202, version 4). *National Snow and Ice Data Center*, 2021. 6

[30] Bin Mu, Xiaodan Luo, Shijin Yuan, and Xi Liang. Icetft v 1.0. 0: Interpretable long-term prediction of arctic sea ice extent with deep learning. *Geoscientific Model Development Discussions*, 2023:1–28, 2023. 2

[31] J.E. Nash and J.V. Sutcliffe. River flow forecasting through conceptual models part i — a discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970. 7

[32] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7474–7489, 2021. 6, 7

[33] Zisis I Petrou and Yingli Tian. Prediction of sea ice motion with convolutional long short-term memory networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57 (9):6865–6876, 2019. 2

[34] Yibin Ren and Xiaofeng Li. Predicting the daily sea ice concentration on a subseasonal scale of the pan-arctic during the melting season by a deep learning model. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023. 1, 2, 6

[35] Y Ren, X Li, and W Zhang. A data-driven deep learning model for weekly sea ice concentration prediction of the pan-arctic during the melting season, ieee t. geosci. remote, 60, 4304819, 2022. 1, 2, 3, 4, 6, 7

[36] Eduardo Rocha Rodrigues, Igor Oliveira, Renato Cunha, and Marco Netto. Deepdownscale: A deep learning strategy for high-resolution weather forecast. In *2018 IEEE 14th International Conference on e-Science (e-Science)*, pages 415–422. IEEE, 2018. 3

[37] Qi Shao, Guangchao Hou, Wei Li, Guijun Han, Kangzhuang Liang, and Yang Bai. Ocean reanalysis data-driven deep learning forecast for sea surface multivariate in the south china sea. *Earth and Space Science*, 8(7):e2020EA001558, 2021. 3

[38] S Sparnocchia, N Pinardi, and E Demirov. Multivariate empirical orthogonal function analysis of the upper thermocline structure of the mediterranean sea from observations and model simulations. In *Annales Geophysicae*, pages 167–187. Copernicus GmbH, 2003. 3

[39] Karen Stengel, Andrew Glaws, Dylan Hettinger, and Ryan N King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815, 2020. 3

[40] Siwei Tu, Weidong Yang, and Ben Fei. Taming generative diffusion for universal blind image restoration. *arXiv preprint arXiv:2408.11287*, 2024. 3

[41] Lei Wang, Xiaojun Yuan, and Cuihua Li. Subseasonal forecast of arctic sea ice concentration via statistical approaches. *Climate Dynamics*, 52:4953–4971, 2019. 1

[42] Qingyu Zheng, Ru Wang, Guijun Han, Wei Li, Xuan Wang, Qi Shao, Xiaobo Wu, Lige Cao, Gongfu Zhou, and Song Hu. A spatio-temporal multiscale deep learning model for subseasonal prediction of arctic sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 3, 6

[43] Xiaohui Zhong, Lei Chen, Jun Liu, Chensen Lin, Yuan Qi, and Hao Li. Fuxi-extreme: Improving extreme rainfall and wind forecasts with diffusion model. *Science China Earth Sciences*, pages 1–13, 2024. 1

[44] Ruofan Zhou and Sabine Susstrunk. Kernel modeling super-resolution on real low-resolution images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2433–2443, 2019. 3