

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

Supplementary Material

A. Preliminary

Diffusion model is a class of generation model which consists of forward and reverse processes. The forward process is a Markov chain defined as follows:

$$q(x_1, \dots, x_T | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}). \quad (1)$$

The forward process gradually adds Gaussian noise to the training data x_0 . When the noise added at each step is small enough and the diffusion step t is large enough, pure Gaussian noise $x_t \sim \mathcal{N}(0, I)$ can be obtained after T diffusion steps. Each diffusion step is defined as follows, where β_t refers to the variance of the forward process, which can be set as a constant or as a parameter that can be learned by reparameterization.

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I). \quad (2)$$

The reverse process is the inversion of the forward process, aiming to simulate random noise from the noise distribution of each reverse step and restore data from it.

However, it is difficult to directly obtain the mean and variance of the conditional distribution $p_\theta(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta I)$ of the reverse process. According to the Bayesian formula, the conditional distribution of the reverse process can be transformed as follows:

$$q(x_{t-1} | x_t, x_0) = q(x_t | x_{t-1}, x_0) \frac{q(x_{t-1} | x_0)}{q(x_t | x_0)}. \quad (3)$$

By directly expanding the three terms at the right end of Equation (3), the mean μ_θ of the reverse process can be represented by the following equation:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, t) \right). \quad (4)$$

Among them, $\epsilon_\theta(x_t, t)$ is the noise simulation function obtained from training, which enables the model to simulate and eliminate noise in the data sampled from the reverse process.

Diffusion model uses maximum likelihood estimation to obtain the probability distribution of Markov transition in the reverse process. Specifically, the noise prediction function $\epsilon_\theta(x_t, t)$ is trained by optimizing the following denoising objectives.

$$E_{\epsilon \sim \mathcal{N}(t, \mathcal{I}), t \sim [0, T]} [\|\epsilon - \epsilon_\theta(x_t, t)\|^2]. \quad (5)$$

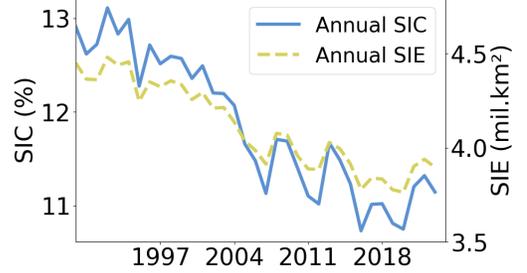


Figure 1. Annual trend of Arctic SIC and SIE of CDR dataset.

B. Annual Trend of Arctic Sea Ice

The Arctic has been declining and retreating for the last few decades. For reference, we plot the annual mean of Arctic sea ice concentration and extent in Figure 1. The years 2016 to 2023 were chosen to construct the test dataset. Specifically, we select the years 2016 and 2022 as extreme cases to further analyze the performance of IceDiff in Section G.1.

C. Guided Reverse Process

This section aims to provide a detailed derivation of the conditional reverse process formula and heuristic algorithm as well as prove that the term of N_2 in Equation (3) in the main text is a constant.

Assuming that an unconditional DDPM reverse process $p_\theta(x_{t-1} | x_t)$ is given. During the downscaling task of SIC maps, the low-resolution SIC maps can be considered as conditions. Based on this, a conditioner $p_\theta(y | x_t)$ can be employed to enhance the diffusion generator. And the gradient term $\nabla_{x_t} \log p_\theta(y | x_t)$ can be utilized to introduce guidance from the LR SIC maps y during the sampling process.

GDP[2] proposes a feasible conditioner which is formulated as follows:

$$p_\theta(y | x_t) = \frac{s}{Z} \exp(-\mathcal{L}(y, \mathcal{D}(x_t))). \quad (6)$$

In IceDiff, \mathcal{D} refers to the optimizable convolutional kernel which is utilized to simulate the upscaling process, and \mathcal{L} represents the distance function which is proposed to measure the bias between the LR SIC maps and generated HR maps after being convolved.

Previous work [1] has derived the conditional transfor-

mation formula in the reverse process:

$$\log p_\theta(x_t|x_{t+1}, y) = \log(p_\theta(x_t|x_{t+1})p(y|x_t)) + N_1, \quad (7)$$

where N_1 refers to the conditional distribution $p_\theta(y|x_{t+1})$. Since it isn't dependent on x_t , it can be seen as a normalizing constant.

As to the reverse process of the diffusion model, the posterior $q(x_t|x_{t+1})$ used for sampling is hard to compute. Therefore, we utilize the model with parameter θ to approximate the conditional probabilities.

$$p_\theta(x_t|x_{t+1}) = \mathcal{N}(\mu_\theta, \Sigma_\theta) \quad (8)$$

$$\log p_\theta(x_t|x_{t+1}) = -\frac{1}{2}(x_t - \mu_\theta)^T \Sigma_\theta^{-1} (x_t - \mu_\theta) + C_1. \quad (9)$$

Equation (9) is a direct logarithmic expansion of Equation (8), where

$$C_1 = -\log(2\pi)^{\frac{n}{2}} (|\Sigma_\theta|^{\frac{1}{2}}) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_\theta| \quad (10)$$

Regarding the another term on the right-hand side of Equation (7) $p_\theta(y|x_t)$, it is difficult to calculate directly. Hence, Taylor expansion around $x_t = \mu_\theta$ can be used to estimate its value. By taking the first two terms of its Taylor expansion, we can obtain:

$$\log p_\theta(y|x_t) \approx \log p_\theta(y|x_t)|_{x_t=\mu_\theta} + (x_t - \mu_\theta)^T \nabla_{x_t} \log p_\theta(y|x_t)|_{x_t=\mu_\theta} \quad (11)$$

$$= C_2 + (x_t - \mu_\theta)^T g \quad (12)$$

Based on the conditioner formula Equation (6), we can approximate the value of the gradient term g :

$$\log p(y|x_t) = -\log N - s\mathcal{L}(y, \mathcal{D}(x_t)) \quad (13)$$

$$g = \nabla_{x_t} \log p(y|x_t) = -s \nabla_{x_t} \mathcal{L}(y, \mathcal{D}(x_t)). \quad (14)$$

Considering integrating Equation (9) and Equation (12), We can obtain:

$$\log p_\theta(x_t|x_{t+1})p_\theta(y|x_t) = \log p_\theta(x_t|x_{t+1}) + \log p_\theta(y|x_t) \quad (15)$$

$$\approx -\frac{1}{2}(x_t - \mu_\theta)^T \Sigma_\theta^{-1} (x_t - \mu_\theta) + (x_t - \mu_\theta)^T g + C_1 + C_2 \quad (16)$$

$$= -\frac{1}{2}(x_t - \mu_\theta - \Sigma_\theta g)^T \Sigma_\theta^{-1} (x_t - \mu_\theta - \Sigma_\theta g) + \frac{1}{2}g^T \Sigma_\theta g + C_1 + C_2 \quad (17)$$

$$= \log p(z) + N_2, \quad z \sim \mathcal{N}(\mu_\theta + \Sigma_\theta g, \Sigma_\theta), \quad (18)$$

where $N_2 = \frac{1}{2}g^T \Sigma_\theta g + C_2$ is a constant related to g .

D. Evaluation Metrics

To comprehensively evaluate our IceDiff, we select commonly used root mean square error (RMSE) and mean absolute error (MAE) for comparison of forecasting accuracy. We also leverage R^2 score to evaluate the performance:

$$R^2 = 1 - \frac{RSS}{TSS}. \quad (19)$$

where RSS represents the sum of squares of residuals and TSS denotes the total sum of squares. The Integrated Ice-Edge Error score [3] is introduced to evaluate the SIE (where the SIC value is greater than 15%) prediction:

$$IIEE = O + U, \quad (20)$$

$$O = SUM(Max(SIE_p - SIE_t, 0)), \quad (21)$$

$$U = SUM(Max(SIE_t - SIE_p, 0)), \quad (22)$$

$$SIE_p, SIE_t = \begin{cases} 1, & SIC > 15 \\ 0, & SIC \leq 15 \end{cases} \quad (23)$$

where O and U represent the overestimated and underestimated SIE between the prediction (SIE_p) and the ground truth (SIE_t), respectively. The difference between the forecasted and ground truth sea ice area (in millions of km^2) is calculated as follows:

$$SIE_{dif} = \frac{SUM(|SIE_p - SIE_t|) \times 25 \times 25}{1000000}. \quad (24)$$

We also adopt the Nash-Sutcliffe Efficiency [5] to further evaluate the predicted quality:

$$NSE = \frac{1 - SUM((SIC_t - SIC_p)^2)}{SUM((SIC_t - Mean(SIC_t))^2)} \quad (25)$$

FID and Consistency are utilized to evaluate the performance of the guided diffusion module of IceDiff.

1. **FID** [4] is an objective indicator used to evaluate the quality of synthesized images. The definition of FID is as follows:

$$FID = \|\mu - \mu_w\|^2 + tr(\Sigma + \Sigma_w - 2(\Sigma \Sigma_w)^{\frac{1}{2}}), \quad (26)$$

where $\mathcal{N}(\mu, \Sigma)$ is the multivariate normal distribution estimated from Inception v3 [6] features calculated from original SIC map and the $\mathcal{N}(\mu_w, \Sigma_w)$ is estimated based on the generated down-scaled SIC map.

2. **Consistency** is adopted to measure faithfulness to the low-resolution SIC map, which refers to mean squared error (MSE) between the low-resolution SIC map and the down-scaled SIC map undergoing the convolution function.

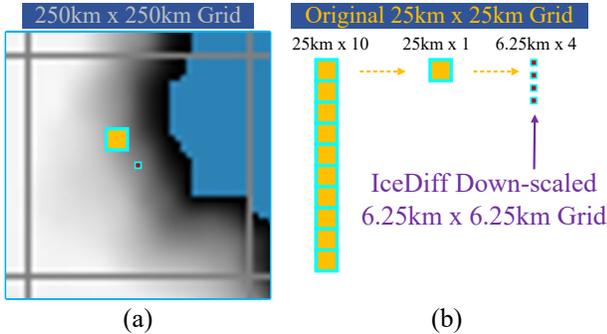


Figure 2. **Illustration of down-scaling grid.** (a) Visualization of the original and down-scaled grid on $250 \text{ km} \times 250 \text{ km}$ scale. (b) Comparison between the original $25 \text{ km} \times 25 \text{ km}$ grid and down-scaled $6.25 \text{ km} \times 6.25 \text{ km}$ grid. Down-scaling by the guided diffusion module in our IceDiff could generate fine-grained SIC maps that are consistent with their coarse counterparts.

Table 1. **Quantitative comparison between guided diffusion module in IceDiff for SIC down-scaling and other methods.**

Metrics	Interpolation-based			Diffusion-based	
	Nearest	Bilinear	Bicubic	GDP	IceDiff
Fid↓	82.07	78.52	56.70	41.39	34.77
Const.↓	21.73	19.02	14.17	8.75	7.42

E. Comparison with Original Scale SIC Map

To further verify the superiority of our method, we attempt to down-scale the original SIC map and compare IceDiff with interpolation-based methods and GDP. FID and Consistency metrics are adopted to measure the faithfulness between the down-scaled model output and the original SIC map.

As shown in Figure 2, IceDiff successfully captures small-scale structures from the original map. The generated map not only achieve a better quality but also down-scale with more clear details. By contrast, both interpolation-based methods and GDP fail to generate rich details and lack fidelity with the original SIC map on the land margin.

As demonstrated in Table 1, IceDiff outperforms interpolation-based methods and GDP in terms of both FID and Consistency metrics. A lower FID and Consistency metrics reflect a more reasonable down-scaling capability, which validates IceDiff as a promising method for generating down-scaled, detailed, and faithful SIC maps.

F. Ablation Study on variants of the Forecasting Module

To further explore the proposed forecasting module, we also construct two variants based on the hyper-parameters: the

Table 2. **Ablation study on the variants of forecasting module in IceDiff.** We compare 2 variants of FM, i.e. Base and Small, to evaluate the impact of different configurations

L.T.	Variant	RMSE↓	MAE↓	R^2 ↑	NSE↑	IIEE↓	SIE_{diff} ↓
7.D.	Small	0.0403	0.0083	0.988	0.986	867	0.0538
	Base	0.0396	0.0080	0.989	0.987	835	0.0315
8.W.A.	Small	0.0579	0.0116	0.971	0.966	1383	0.1239
	Base	0.0553	0.0112	0.973	0.969	1353	0.0871
6.M.A.	Small	0.0696	0.0195	0.901	0.902	2607	0.4744
	Base	0.0648	0.0168	0.919	0.913	2016	0.2657

number of blocks and multi-heads in each layer:

$$FM_{Small} : Blocks[2, 2, 6, 2], Heads[2, 4, 4, 8]$$

$$FM_{Base} : Blocks[2, 2, 18, 2], Heads[4, 8, 16, 32],$$

where the last number is the setup in the bottleneck. The performance improves when more blocks and heads are added to the forecasting module (as in Table 2). Hence we choose to use FM_{Base} model throughout this work to forecast SIC for the guided diffusion module in IceDiff to perform down-scaling.

G. Additional Visualization Results

G.1. Extreme Case Analysis

For the year 2016, historically low annual mean SIC was recorded. In 2022, an abrupt increase of annual average than previous yeas was observed and it reaches the highest value within the test time period. We use our IceDiff at daily scale for the forecasting of September 1st (during the end of melting season when SIC typically reaches its lowest value) in 2016 (Figure 3(a)) and 2022 (Figure 3(b)). Our IceDiff could produce accurate forecasts of SIC and SIE in those two extreme cases. Considering the historically low annual mean in 2016 and the abnormal SIC fluctuation in 2022 when compared to the annual pattern during test time period, our IceDiff are capable of providing reliable forecasting on 25 km grid.

G.2. Down-scaling Quality

Figure 2 presents the difference between the original SIC map and the output of IceDiff. The down-scaled SIC map has a minimum of 6.25 km grid length, which is a quarter of the original 25 km, and provides more detail at finer scales.

Figure 4 demonstrates that our IceDiff is more consistent with the original SIC map than the other two baselines.

Figure 5 provides additional visualization results of three different down-scaling methods. Note that both IceDiff and GDP are based on diffusion models, a few dark pixels (SIC value close to zero) in the open water area are generated during the denoising process.

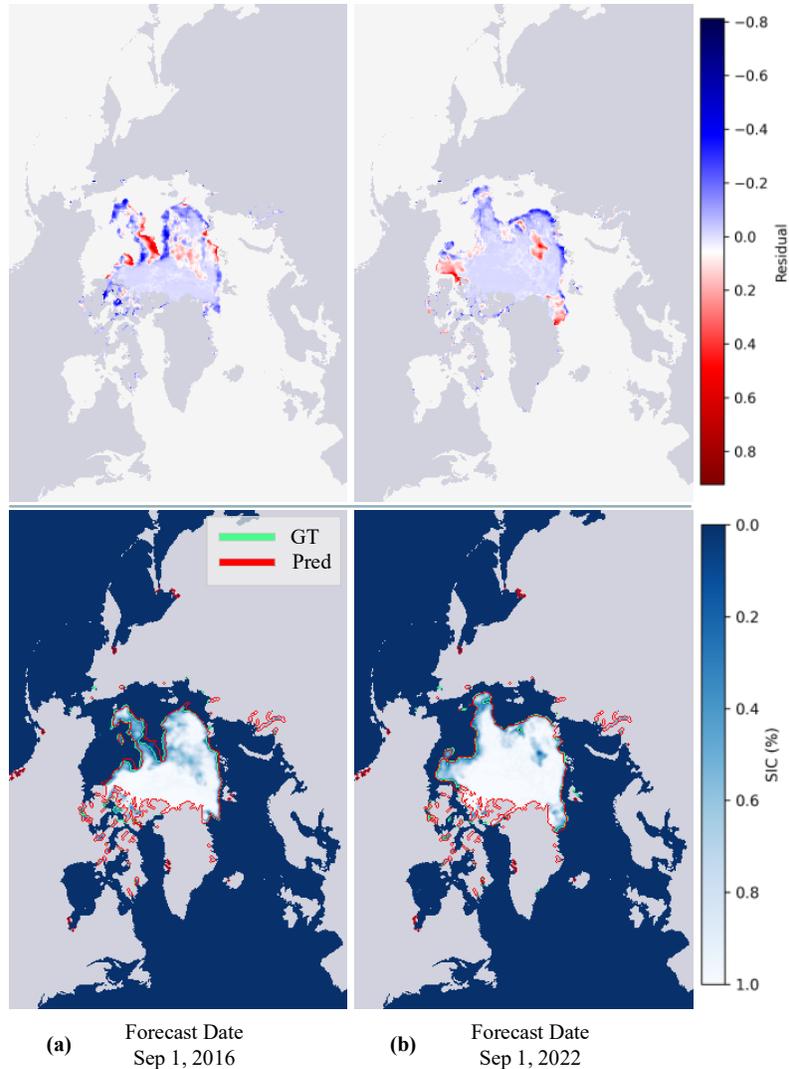


Figure 3. **Performance of IceDiff in extreme cases.** The upper two figures represents the forecasting residuals of SIC while the lower two illustrates the predicted SIE by IceDiff in (a) September 1st, 2016 and (b) September 1st, 2022.

References

- [1] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 1
- [2] 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. 1
- [3] 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. 2
- [4] 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. 2
- [5] 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. 2
- [6] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 2

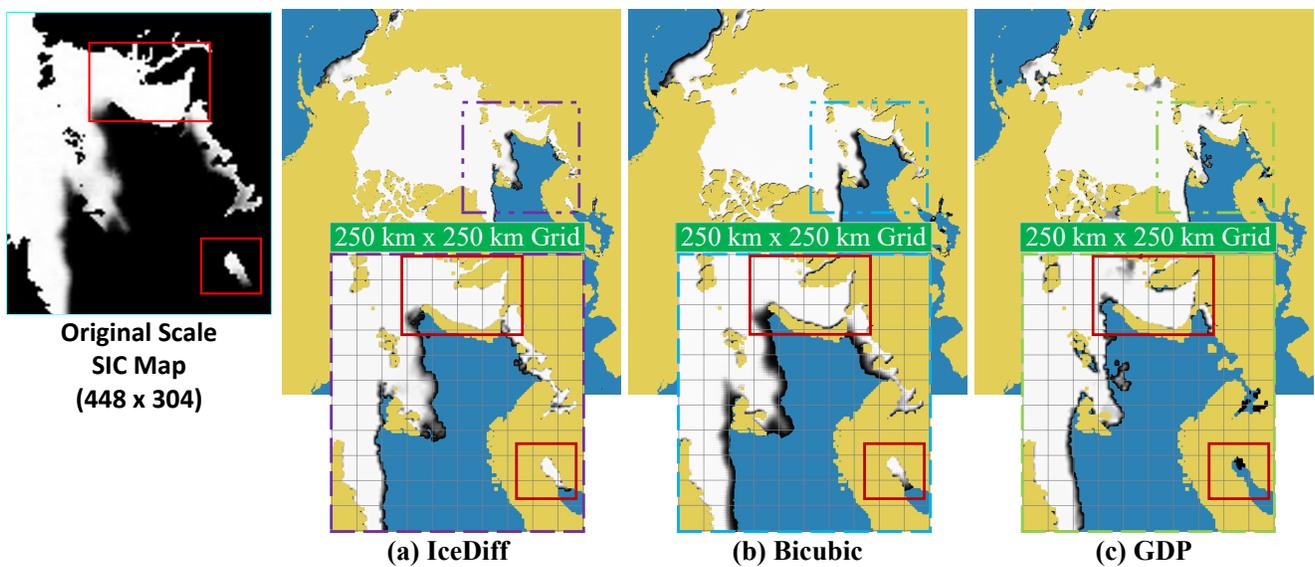


Figure 4. **Down-scaling Quality.** Comparison of SIC map at original scale and down-scaled by three different methods (March 27th, 2016).

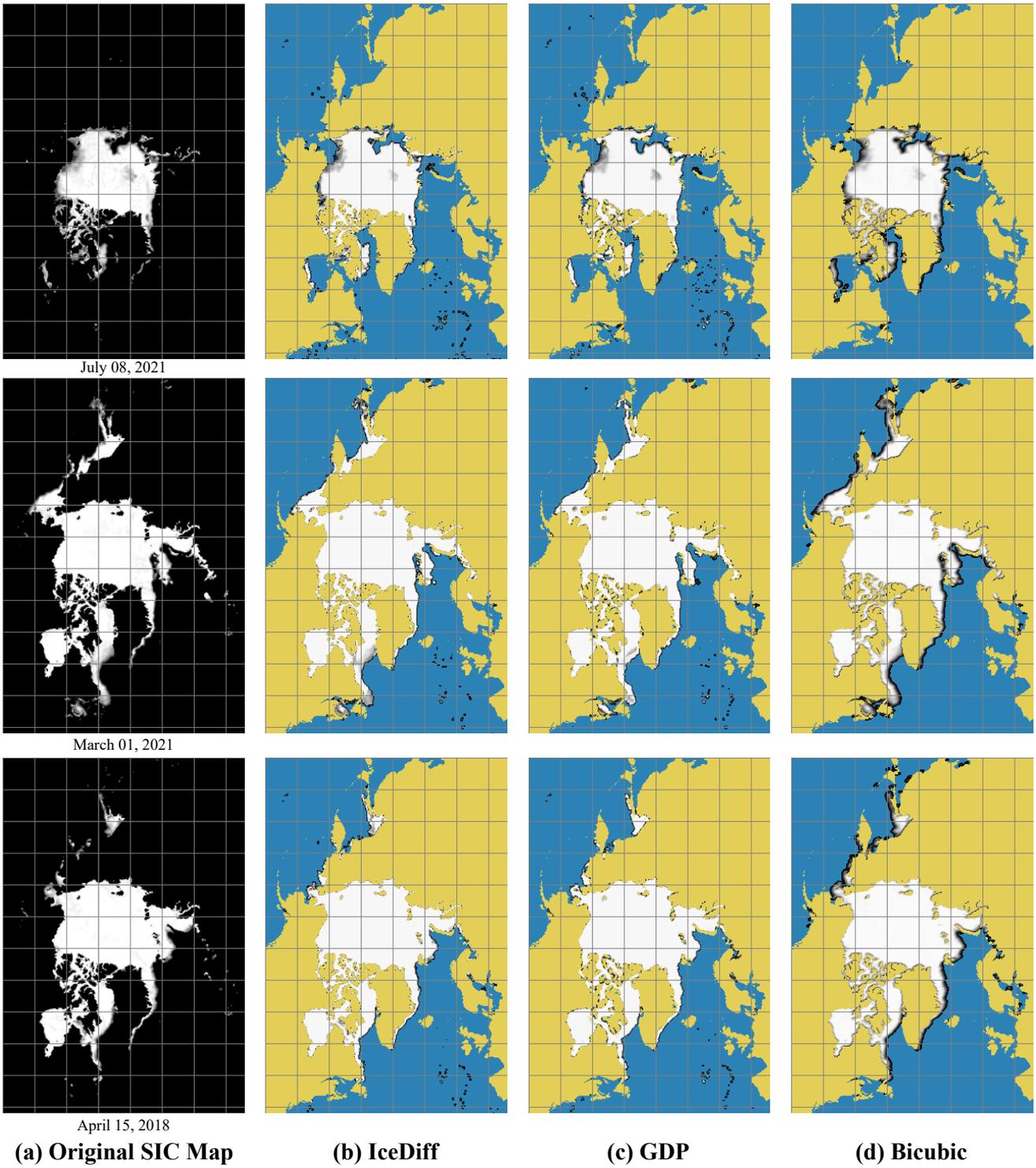


Figure 5. **Down-scaling Quality.** Comparison of SIC map at original scale and down-scaled by three different methods.