Incomplete Observations Boost Evolutionary Performance in Ocean Modeling

Yangyang Kong[*,1,2], Yutong Jiang[*,1,2], Yanhai Gan[1,2], Junyu Dong[1,2], Feng Gao[1,2], Xiaopei Lin[1]

[1]State Key Laboratory of Physical Oceanography, Ocean University of China, Qingdao, China
[2]School of Computer Science and Technology, Ocean University of China, Qingdao, China

Key Points:

- A generative model integrating physical states and observations is proposed, following the paradigm of conventional numerical models.
- An optimization framework is formulated, enabling the generative model to directly learn ocean dynamics from sparse observations.
- Leveraging sparse observations, the model realizes self-evolution and yields improved reconstruction and prediction performance.

arXiv:2607.19147v1 [cs.LG] 21 Jul 2026

[*] These authors contributed equally to this work.

Corresponding author: Yanhai Gan, ganyanhai@ouc.edu.cn

Corresponding author: Junyu Dong, dongjunyu@ouc.edu.cn

Abstract

Data-driven methods have revolutionized ocean modeling, yet current approaches rely heavily on complete reanalysis datasets, imposing computational constraints and limiting model performance to that of the training data. Here, we present a generative state-space model and an optimization framework that enable learning directly from sparse and noisy observations. The model is essentially a hidden Markov model with a continuous state space, where oceanic physical quantities are treated as hidden states and measurements as observations, enabling a unified representation of ocean fields and observational data. Both the initial-state and state-transition modules are implemented as neural networks to capture the complexity and temporal evolution of ocean states, while the emission module is formulated as a masked Gaussian distribution. To train the model from sparse observations, we derive an optimization framework based on the expectation–maximization (EM) algorithm. The framework alternately reconstructs high-fidelity ocean fields via Langevin dynamics and optimizes deep neural networks to capture temporal evolution. Theoretical analysis shows that the framework maximizes the likelihood of observations under the generative model. For efficiency, we assume that ocean-state evolution follows a stationary, ergodic, and Markovian stochastic process and adopt only length-two state sequences during optimization. Experiments on CMIP6 simulation data and FY-3D satellite data demonstrate high-fidelity reconstruction and accurate prediction, showing that sparse observations can directly improve the model's representation of ocean-state dynamics. This work offers a scalable pathway for next-generation Earth system models to learn directly from sparse, incomplete real-world observations.

Plain Language Summary

Today's most successful artificial intelligence models for simulating the ocean and atmosphere are trained on large, carefully reconstructed global datasets that combine sparse observations with traditional computer models. While these datasets are useful, they are expensive to produce, limit the finest scales that can be resolved, and cap how accurate artificial intelligence predictions can become. In this study, we developed a new machine learning framework that can learn directly from the incomplete, sparse, and noisy measurements collected by real-world instruments such as satellites. Our approach treats the complete ocean state as a quantity that evolves over time and uses an iterative training strategy: first, it fills in missing observations to build realistic ocean fields, then it trains neural networks to predict how those fields change. Tests using climate model simulations and sea surface temperature measurements from a Chinese weather satellite show that this method improves prediction accuracy and can rebuild complete ocean maps even when much of the input data are missing. These results suggest a practical path toward artificial intelligence Earth system models that learn from real observations rather than depending solely on reconstructed datasets.

1 Introduction

In recent years, artificial intelligence (AI), spearheaded by deep learning, has driven a substantial shift in scientific discovery, with growing implications for the Earth sciences under the banner of "AI for Science" (Reichstein et al., 2019). The emergence of next-generation data-driven foundation models, exemplified by Pangu-

[*] These authors contributed equally to this work.

Weather (Bi et al., 2023), GraphCast (Lam et al., 2023), FengWu (Chen et al., 2023), and FourCastNet (Pathak et al., 2022), is reshaping Earth system modeling. These models achieve accuracies comparable to traditional numerical weather prediction (NWP) at substantially lower inference cost, representing a notable advance (Nguyen et al., 2023). These developments have made the goal of constructing a "Digital Twin" capable of simulating and predicting complex Earth systems at high resolution and fidelity more attainable than before.

However, the success of these state-of-the-art AI models rests almost invariably on large-scale, high-quality reanalysis datasets, such as ERA5 for the atmosphere (Hersbach et al., 2020) and global ocean state estimates including GLORYS and ECCO (Forget et al., 2015). Reanalysis fields are produced by fusing sparse, multimodal observations with simulations from traditional physical numerical models via data assimilation (DA). Although such products are currently the principal training data, dependence on this complete-field paradigm imposes two constraints on further progress. First, the computational cost of ocean data assimilation limits the spatiotemporal resolution of reanalysis products, thereby leaving finer, dynamically important sub-mesoscale processes under-resolved relative to the capacity of AI models (Brajard et al., 2020). Second, reanalysis accuracy is bounded by observational coverage and numerical model fidelity, creating an effective performance ceiling associated with intrinsic uncertainties. When AI models treat reanalysis data as ground truth, predictive skill is largely limited by reanalysis fidelity itself, which hinders their development into independent physical simulators that could, in principle, complement traditional numerical methods. Consequently, a central scientific question is whether dependence on reanalysis can be reduced and whether a training paradigm can be established that allows AI models to learn directly from incomplete, sparse, and noisy ocean observations—such as cloud-contaminated satellite swath measurements with large spatial gaps (Zhang et al., 2025). Meeting this challenge may move AI models beyond emulating assimilated products toward learning observation-consistent dynamics (Runge et al., 2019). Yet existing generative paradigms also face difficulties. Whether based on generative adversarial networks (GANs) (Ravuri et al., 2021), diffusion models (Mardani et al., 2023), or recent sparse reconstruction methods, these approaches encounter fundamental limitations when applied directly to such scientific tasks. Designed primarily for natural image synthesis, standard generative models are typically not constructed with explicit physical constraints. Although they can produce visually plausible static fields, they often struggle to maintain physical consistency over complex temporal evolution—a gap increasingly noted in Earth system machine learning (Sonnewald et al., 2019; Feng et al., 2025). In particular, they lack explicit mechanisms to preserve approximate geostrophic balance, mass conservation, or large-scale circulation features characteristic of ocean dynamics.

To address this gap, we propose a generative state-space modeling framework designed for incomplete observational data. By incorporating a state-space representation of temporal dynamics, the framework maintains spatiotemporal coherence within the generative process. The complete ocean field is treated as a latent state, heterogeneous measurements are linked through a masked Gaussian observation model, and deep generative networks parameterize the initial-state distribution and stochastic state transitions. Parameters are estimated using an expectation–maximization (EM) strategy (Dempster et al., 1977). In the expectation (E) step, we use Monte Carlo sampling based on Langevin dynamics; conditioned on sparse observations, this step reconstructs high-fidelity, spatiotemporally continuous oceanographic fields. In the maximization (M) step, the reconstructed fields supervise updates to the initial-state generator (a modified StyleGAN2-ADA backbone) and the state-transition model (a stochastic U-Net) (Ross et al., 2023). Iterating between the E- and M-steps allows the model to refine physical patterns from noisy and

missing data and to reconstruct and predict spatiotemporal evolution of the physical system.

2 Methodology

2.1 Probabilistic formulation as a Generative State-Space Model

To address the challenge of learning Earth system dynamics from incomplete observations (Reichstein et al., 2019; Ghil & Malanotte-Rizzoli, 1991; Brajard et al., 2020), we formulate the problem within a probabilistic generative modeling framework. Specifically, we structure the system's dynamics as a Generative State-Space Model (SSM) (Rangapuram et al., 2018), where the complete physical fields, $\mathbf{s} = \{\mathbf{s}_0, \ldots, \mathbf{s}_T\}$, are treated as state variables and the sparse satellite data, $\mathbf{o} = \{\mathbf{o}_0, \ldots, \mathbf{o}_T\}$, as observed variables. By assuming a first-order Markov process, the joint probability distribution over the entire system, $p_\theta(\mathbf{s}, \mathbf{o})$, can be factorized into three core components: an Initial State Model, a State Transition Model, and an Observation Model. Within this framework, our fundamental optimization objective is maximum likelihood estimation (MLE), which seeks parameters θ^* that maximize the marginal likelihood of the observed data $p_\theta(\mathbf{o})$. Due to the latent nature of $\mathbf{s}$, this requires an intractable high-dimensional integral over all state configurations:

$$\theta^* = \arg\max_\theta \log p_\theta(\mathbf{o}) = \arg\max_\theta \log \int p_\theta(\mathbf{s}, \mathbf{o})\, \mathrm{d}\mathbf{s}. \tag{1}$$

2.2 Learning via Monte Carlo Expectation-Maximization

Direct optimization of the MLE objective in Eq. (1) is intractable. To address this, we adopt the Expectation–Maximization (EM) framework (Dempster et al., 1977) to iteratively approach Eq. (1). Since the expectations in the M-step are computationally prohibitive, we implement a Monte Carlo EM (MCEM) approach (Wei & Tanner, 1990; Peyron et al., 2021). In the E-step, we draw one state trajectory (i.e., a realization of $\mathbf{s}$) using Langevin dynamics to approximate the posterior $p_{\theta^{(t)}}(\mathbf{s} \mid \mathbf{o})$ given observations. For a dataset comprising N samples (i.e., distinct realizations of $\mathbf{o}$), the M-step updates the parameters by solving

$$\theta^{(t+1)} = \arg\max_\theta \sum_{i=1}^{N} \log p_\theta\big(\mathbf{s}^{(i)}, \mathbf{o}^{(i)}\big). \tag{2}$$

Here, $\mathbf{s}^{(i)}$ is imputed in the E-step at iteration t and is strictly paired with $\mathbf{o}^{(i)}$. The superscript "(i)" indexes different instances in the dataset. Eq. (2) is the M-step update rule in the same arg max form as Eq. (1); the optimization objective remains maximization of $\log p_\theta(\mathbf{o})$ in Eq. (1).

Detailed mathematical derivations, including the decomposition of the evidence lower bound and the Monte Carlo rationale, are provided in Appendix A.

This formulation effectively transforms the original unsupervised learning problem into a standard supervised learning task, where the M-step updates θ by solving Eq. (2), while the overall learning goal remains Eq. (1).

2.3 Deep generative architectures

The probabilistic components of our state-space model are parameterized by advanced deep neural networks, designed to balance generation quality with computational efficiency and training stability.

2.3.1 Initial state model

We parameterize the initial state distribution $p_{\theta_{init}}(\mathbf{s}_0)$ using a modified StyleGAN2-ADA architecture (Karras, Aittala, et al., 2020). While StyleGAN2-ADA provides robust training on limited scientific datasets via adaptive discriminator augmentation, its standard convolutional blocks are computationally expensive. To achieve a lightweight design, we replace standard convolutions with Depthwise Separable Convolutions (DSC) (Chollet, 2017). However, applying the standard StyleGAN modulation-demodulation mechanism directly to depthwise convolution (DWConv) kernels leads to a mathematical cancellation effect, where the style scaling factors act as both numerator and denominator during normalization, effectively erasing style information due to the channel independence of DWConv.

To resolve this, we introduce a structural decoupling strategy for style injection. Specifically, we restrict the style modulation operation to the DWConv layer, allowing it to exclusively adjust the relative amplitudes of individual feature channels. Subsequently, we utilize the Pointwise Convolution (PWConv) layer to linearly mix these style-modulated features, transforming amplitude differences into structural variations. Crucially, the demodulation (normalization) operation, originally performed at the DWConv layer, is migrated to the PWConv layer. This "inject-at-DWConv, normalize-at-PWConv" design avoids mathematical cancellation, ensuring effective style propagation across scales while significantly improving parameter efficiency.

2.3.2 Stochastic state transition model

The transition probability $p_{\theta_{trans}}(\mathbf{s}_{t+1}|\mathbf{s}_t)$ is learned via a conditional adversarial framework inspired by Pix2Pix (Isola et al., 2017). The generator employs a Stochastic U-Net architecture (Ronneberger et al., 2015), which takes the concatenation of the current state $\mathbf{s}_t$ and a latent noise tensor $\mathbf{z}$ as input. The U-Net's multi-scale skip connections efficiently capture the multi-resolution spatial features inherent in ocean-meteorological systems. Unlike the local PatchGAN discriminator used in standard Pix2Pix, we employ a global conditional convolutional classifier as the discriminator to enforce global physical consistency constraints (Mirza & Osindero, 2014), a strategy increasingly emphasized in recent data-driven Earth system modeling to prevent unphysical predictions.

To enable probabilistic forecasting and enhance optimization flexibility, we incorporate an explicit noise injection mechanism. Gaussian noise sampled from $\mathcal{N}(0, \mathbf{I})$ is injected into each layer of the decoder after passing through learned affine transformations (Karras et al., 2019). This design provides dual benefits: it transforms the deterministic mapping into a stochastic process to capture the intrinsic uncertainty of physical evolution, and it expands the solution space flexibility, providing critical manifold support for the gradient-based posterior sampling in the E-step.

2.3.3 Observation model

We establish an explicit probabilistic observation model $p(\mathbf{o}_t|\mathbf{s}_t)$ to mathematically describe the link between the hidden system state $\mathbf{s}_t$ and the incomplete observation $\mathbf{o}_t$. We assume that observations are measurements taken from a sparse spatial subset of the true physical field, corrupted by Gaussian noise. This is formalized as a conditional Gaussian distribution, consistent with variational data assimilation principles in oceanography (Kalnay, 2003; Evensen, 2009; Bennett, 2002):

$$p(\mathbf{o}_t|\mathbf{s}_t) = \mathcal{N}(\mathbf{o}_t \mid \mathbf{M}_t \odot \mathbf{s}_t, \sigma^2\mathbf{I}). \tag{3}$$

where $\mathbf{M}_t$ is a binary mask operator representing the spatial location of observations (e.g., determined by satellite orbits or cloud cover in real-world datasets, or generated artificially in simulated datasets), and $\sigma^2\mathbf{I}$ represents the measurement uncertainty.

2.4 Iterative Optimization via Expectation-Maximization

In practice, directly optimizing the log-likelihood in Eq. (1) is extremely challenging. On one hand, explicitly modeling the joint probability of state and observation variables is difficult. On the other hand, computing the marginal likelihood requires integrating over the latent state space, which is intractable even if the joint probability is known. Intuitively, constructing a model from incomplete observations faces a dilemma of circular dependency between "state reconstruction" and "model learning" (Ghahramani & Roweis, 1998): the precise inference of the complete physical field $\mathbf{s}$ requires an accurate model as a prior, yet training an accurate model relies on a comprehensive representation of the complete field $\mathbf{s}$.

To address this circular dependency, we employ the Expectation-Maximization (EM) algorithm (Dempster et al., 1977), transforming the problem into two alternating core steps. The fundamental objective remains to maximize the observation likelihood $p_\theta(\mathbf{o})$. In the E-step, we obtain samples from the posterior distribution $p_\theta(\mathbf{s}|\mathbf{o})$ via Langevin dynamics to achieve an efficient reconstruction of the system state, a strategy increasingly adopted in data-driven oceanography (Fablet et al., 2021; Bocquet et al., 2019). In the subsequent M-step, these reconstructed state samples are used as high-quality training data to update the parameters of the joint probability model. The E-step and M-step alternate, driving the iterative evolution of the entire framework.

Our proposed framework is theoretically applicable to sequential problems of arbitrary length T, offering a generalized solution for learning spatiotemporal dynamics from incomplete data. The detailed algorithmic flow of a single EM iteration is summarized in Algorithm 1.

2.5 E-step: Efficient state reconstruction via latent space sampling

This step addresses the core challenge of state reconstruction within our framework, specifically aiming to "achieve efficient reconstruction of spatiotemporal system states conditional on observational data." Within the EM framework, the mathematical objective of the E-step is to sample from the posterior distribution $p_\theta(\mathbf{s}|\mathbf{o})$, given the current model parameters θ and the incomplete observations $\mathbf{o}$.

Directly performing Langevin dynamics sampling on the posterior in the pixel space ($\mathbf{s}$-space) faces significant challenges: the dimensionality of $\mathbf{s}$ is extremely high, and its probability distribution can be exceedingly complex and rugged. To address this, we adopt a more efficient and stable latent space sampling strategy (Pang et al., 2021). The core idea is to shift from directly sampling the high-dimensional state $\mathbf{s}$ to sampling the lower-dimensional stochastic latent vector $\mathbf{z}$ that governs the generation of $\mathbf{s}$, which has shown promising results in inverting complex geophysical systems (Peyron et al., 2021).

Specifically, we view the system state as being generated by our pre-trained state models from a sequence of latent noise vectors. Here, the latent noise vectors $\mathbf{z}$ encompass both the input noise and the layer-wise injected noise within the Initial State Model and the State Transition Model. Consequently, our objective transforms from sampling $\mathbf{s}$ to sampling $\mathbf{z}$:

$$\mathbf{z} \sim p_\theta(\mathbf{z}|\mathbf{o}). \tag{4}$$

Algorithm 1 One Iteration of the EM-based Learning Framework

Require: Models: Initial state generator $G_{\theta_{\text{init}}}$, Transition model $F_{\theta_{\text{trans}}}$.
Require: Data: Sparse observations $\{\mathbf{o}_t\}_{t=0}^{T}$, Mask operators $\{\mathbf{M}_t\}_{t=0}^{T}$.
Require: Hyperparameters: Langevin steps K, Step size η, Prior weights $\Lambda = \{\lambda_{\mathbf{z}}\}$.
Ensure: Updated parameters $\theta^{\text{new}} = \{\theta_{\text{init}}, \theta_{\text{trans}}\}$.

E-Step: State Reconstruction via Latent Space Sampling

Initialize latent variables $\mathcal{Z} = \{\mathbf{z}_0, \ldots, \mathbf{z}_T\} \sim \mathcal{N}(0, \mathbf{I})$.
for $k = 1$ to K **do**
  Forward simulation to generate physical states:
  $\mathbf{s}_0 \leftarrow G_{\theta_{\text{init}}}(\mathbf{z}_0)$
  for $t = 1$ to T **do**
    $\mathbf{s}_t \leftarrow F_{\theta_{\text{trans}}}(\mathbf{s}_{t-1}, \mathbf{z}_t)$
  end for
  Compute energy (negative log-posterior):
  $\mathcal{E}(\mathcal{Z}) \leftarrow \sum_{t=0}^{T} \frac{\|(\mathbf{o}_t - \mathbf{s}_t) \odot \mathbf{M}_t\|_2^2}{2\sigma_t^2} + \sum_{\mathbf{z} \in \mathcal{Z}} \lambda_{\mathbf{z}} \|\mathbf{z}\|_2^2$
  Update latents via Langevin Dynamics step:
  $\mathcal{Z} \leftarrow \mathcal{Z} - \frac{\eta}{2} \nabla_{\mathcal{Z}} \mathcal{E}(\mathcal{Z}) + \sqrt{\eta} \cdot \mathcal{N}(0, \mathbf{I})$
end for
Output of E-Step: Reconstructed full states $\hat{\mathcal{S}}$ derived from the final $\mathcal{Z}$.

M-Step: Model Parameter Updates

Optimize Initial-State Model:
$\mathcal{L}_{\text{init}} \leftarrow \sum_{t=0}^{T} \mathcal{L}_G(\theta_{\text{init}}; \hat{\mathbf{s}}_t)$
$\theta_{\text{init}} \leftarrow \text{AdamUpdate}(\theta_{\text{init}}, \nabla_{\theta_{\text{init}}} \mathcal{L}_{\text{init}})$

Optimize Transition Model:
$\mathcal{L}_{\text{trans}} \leftarrow \sum_{t=1}^{T} \mathcal{L}_F(\theta_{\text{trans}}; \hat{\mathbf{s}}_{t-1}, \hat{\mathbf{s}}_t)$
$\theta_{\text{trans}} \leftarrow \text{AdamUpdate}(\theta_{\text{trans}}, \nabla_{\theta_{\text{trans}}} \mathcal{L}_{\text{trans}})$

return Updated parameters θ^{new}.

Since $p_\theta(\mathbf{z}|\mathbf{o})$ is intractable to compute, we utilize Langevin dynamics for sampling (Welling & Teh, 2011). To do this, we need to compute the score of the noise posterior distribution:

$$\begin{aligned} \nabla_{\mathbf{z}} \log p_\theta(\mathbf{z}|\mathbf{o}) &= \nabla_{\mathbf{z}} \log p_\theta(\mathbf{z}, \mathbf{o}) - \nabla_{\mathbf{z}} \log p_\theta(\mathbf{o}) \\ &= \nabla_{\mathbf{z}} \log p(\mathbf{z}) + \nabla_{\mathbf{z}} \log p_\theta(\mathbf{o}|\mathbf{z}) \\ &= \nabla_{\mathbf{z}} \log p(\mathbf{z}) + \nabla_{\mathbf{z}} \log p_\theta(\mathbf{o}|\mathbf{s}). \end{aligned} \tag{5}$$

Since the prior $p(\mathbf{z})$ is a pre-defined Gaussian distribution and the likelihood $p_\theta(\mathbf{o}|\mathbf{s})$ is also Gaussian, the score $\nabla_{\mathbf{z}} \log p_\theta(\mathbf{z}|\mathbf{o})$ can be easily computed. With the score of the noise posterior distribution, we can employ Langevin dynamics to sample from this posterior:

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \frac{\eta}{2} \nabla_{\mathbf{z}} \log p_\theta(\mathbf{z}_k|\mathbf{o}) + \sqrt{\eta} \cdot \mathbf{w}_k, \quad \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{6}$$

where k differentiates the iterations of the Langevin dynamics. To enhance the practical performance of this latent space Langevin sampling, we identified three key effective strategies through a series of comparative experiments.

2.5.1 Increasing latent degrees of freedom

We identified that the composition of the optimizable latent variables is critical for sampling fidelity. In our framework, the latent variable $\mathcal{Z}$ is defined as a composite high-dimensional set encompassing all sources of stochasticity: the initial input vectors for both the initial state and transition models, as well as the noise vectors injected into every layer of both networks. Crucially, to further enhance optimization flexibility, we expanded the channel dimension of these layer-wise noise injections from single-channel to multi-channel. This design significantly multiplies the optimizable parameter space, enabling fine-grained tuning at specific feature levels.

Optimization with low degrees of freedom (e.g., optimizing only initial noise) faces a dual challenge: a constrained search space that imposes a theoretical ceiling on solution quality, and a rugged energy landscape populated with local minima. In contrast, our joint optimization strategy mitigates the first challenge by expanding the search space and addresses the second by decomposing the highly coupled global optimization into multi-level cooperative tasks. This structural flexibility allows the optimizer to escape global stagnation by adjusting local noise variables, thereby accessing higher-quality solution regions inaccessible to global-only adjustments.

2.5.2 Choice of activation function

We further observed that the choice of non-linear activation functions also contributes to sampling fidelity. Empirical comparisons indicated that employing Tanh as the primary activation yields improved fitting accuracy compared to standard ReLU (Nair & Hinton, 2010) or LeakyReLU (Maas et al., 2013) configurations. Interestingly, we initially hypothesized that this benefit stemmed from avoiding the non-smooth "kink" of ReLUs at zero; however, preliminary exploratory experiments did not support this. Consequently, we revised our hypothesis: Tanh's benefit likely arises from its property of providing a maximal gradient at the origin. Unlike ReLU or LeakyReLU, where gradients near zero can be small or abrupt, Tanh appears to offer a more favorable driving force for our gradient-based latent optimization process. This finding aligns with recent advances in Physics-Informed Neural Networks (PINNs), where smooth activation functions (e.g., Tanh) are preferred for modeling continuous physical quantities (Raissi et al., 2019). Given these empirical benefits, we uniformly adopted Tanh throughout our model architecture.

2.5.3 Choice of noise prior distribution

We further investigated the impact of the prior distribution type for the injected noise vectors $\boldsymbol{\varepsilon}$. We compared two standard choices: a Uniform distribution $\mathcal{U}(-1, 1)$ and a Standard Gaussian distribution $\mathcal{N}(0, \mathbf{I})$. As visualized in Fig. 1, a striking contrast was observed. Employing a Uniform distribution (Fig. 1a) induces unnatural spatial discontinuities, particularly visible as vertical seams, thereby compromising the physical coherence of the reconstructed field. Conversely, the Gaussian noise prior (Fig. 1b) yields spatially smooth and physically consistent reconstructions, effectively eliminating these artifacts.

We attribute this phenomenon to the optimization dynamics in the latent space. The log-prior of a Uniform distribution has zero gradient within its support, failing to provide effective regularization guidance for unobserved regions. In contrast, the Gaussian prior imposes a smooth quadratic penalty ($\propto -\|\boldsymbol{\varepsilon}\|^2$), providing continuous gradient feedback that regularizes the latent variables. This property is highly compatible with our gradient-based Langevin sampling framework and is consistent with the standard assumption of Gaussian background error covariance

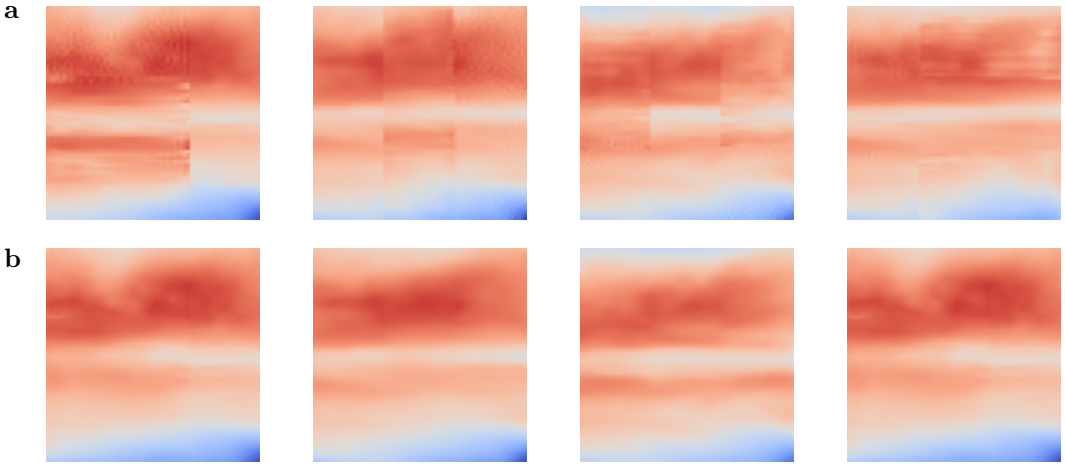

Figure 1: Comparison of spatial continuity under different noise priors. a, Reconstructions using a uniform prior $\mathcal{U}(-1, 1)$ exhibit fracture-like artifacts. b, Reconstructions using a Gaussian prior $\mathcal{N}(0, \mathbf{I})$ demonstrate smooth structures.

in optimal interpolation and variational data assimilation (Lorenc, 1986; Bennett, 2002). Consequently, we utilize Gaussian noise for all injected noise layers.

2.6 M-step: Model evolution via multi-objective learning

In the M-step, we update the model parameters $\theta = \{\theta_{\text{init}}, \theta_{\text{trans}}\}$ by maximizing the expected log-likelihood of the complete data (Dempster et al., 1977). We utilize the high-fidelity state trajectories $\hat{\mathbf{s}} = \{\hat{\mathbf{s}}_0, \ldots, \hat{\mathbf{s}}_T\}$ reconstructed in the E-step as pseudo-ground-truth training data. Given the architectural orthogonality between the initial state distribution and the temporal dynamics, we decouple the global optimization into two parallel sub-tasks:

It is worth noting that we employ different adversarial loss formulations for the two sub-tasks. For the Initial State Model, we adopt the non-saturating loss with R1 and PL regularization as recommended by StyleGAN2-ADA (Karras, Aittala, et al., 2020). This choice is motivated by its proven stability and effectiveness in training high-fidelity unconditional generative models. For the State Transition Model, we utilize the original min-max adversarial loss from the conditional GAN framework (Goodfellow et al., 2014; Mirza & Osindero, 2014), consistent with the foundational Pix2Pix (Isola et al., 2017) approach for image-to-image translation tasks. This deliberate choice allows each component to leverage the most established and robust training strategy for its specific task.

2.6.1 Initial state model update

We optimize the parameters θ_{init} (encompassing both the generator $G_{\theta_{\text{init}}}$ and discriminator $D_{\theta_{\text{init}}}$) to match the distribution of the reconstructed initial states $\hat{\mathbf{s}}_0$.

The generator objective minimizes the adversarial loss combined with path length regularization:

$$\mathcal{L}_G = \mathbb{E}_{\mathbf{z}} \left[\text{softplus} \left(-D_{\theta_{\text{init}}}(G_{\theta_{\text{init}}}(\mathbf{z})) \right) \right] + \lambda_{pl} \mathcal{L}_{pl}. \tag{7}$$

where $\mathcal{L}_{pl}$ denotes the Path Length (PL) regularization (Karras, Laine, et al., 2020). This term encourages a smooth latent space, essential for the stability of the gradient-based Langevin sampling in the E-step.

The discriminator objective maximizes the distinction between real and generated images, stabilized by the R1 gradient penalty:

$$\mathcal{L}_D = \mathbb{E}_{\hat{\mathbf{s}}_0}\left[\text{softplus}\left(-D_{\theta_{\text{init}}}(\hat{\mathbf{s}}_0)\right)\right] + \mathbb{E}_{\mathbf{z}}\left[\text{softplus}\left(D_{\theta_{\text{init}}}(G_{\theta_{\text{init}}}(\mathbf{z}))\right)\right] + \lambda_{R1}\mathcal{L}_{R1}. \tag{8}$$

where $\mathcal{L}_{R1} = \mathbb{E}_{\hat{\mathbf{s}}_0}[\|\nabla D_{\theta_{\text{init}}}(\hat{\mathbf{s}}_0)\|^2]$ penalizes the gradient norm on real data to prevent mode collapse.

2.6.2 State Transition Model Update

The transition parameters θ_{trans} (parameterizing the Stochastic U-Net) are optimized to capture the stochastic physical dynamics across the entire temporal sequence. We construct the training set by extracting all valid transition pairs $\{(\hat{\mathbf{s}}_t, \hat{\mathbf{s}}_{t+1})\}_{t=0}^{T-1}$ from the high-fidelity trajectories reconstructed in the E-step. The optimization is formulated as a min-max game over the cumulative loss across all time steps:

$$\min_{G_{\theta_{\text{trans}}}} \max_{D_{\theta_{\text{trans}}}} \mathcal{L}_{\text{trans}} = \sum_{t=0}^{T-1} \left(\mathcal{L}_{\text{cGAN}}^{(t)} + \lambda \mathcal{L}_{L1}^{(t)}\right). \tag{9}$$

where the conditional adversarial term $\mathcal{L}_{\text{cGAN}}^{(t)}$ (Mirza & Osindero, 2014) ensures that the predicted transition from $\hat{\mathbf{s}}_t$ to $\hat{\mathbf{s}}_{t+1}$ is distributionally indistinguishable from real physical evolution:

$$\mathcal{L}_{\text{cGAN}}^{(t)} = \mathbb{E}_{(\hat{\mathbf{s}}_t, \hat{\mathbf{s}}_{t+1})}[\log D_{\theta_{\text{trans}}}(\hat{\mathbf{s}}_t, \hat{\mathbf{s}}_{t+1})] + \mathbb{E}_{\hat{\mathbf{s}}_t, \mathbf{z}}[\log(1 - D_{\theta_{\text{trans}}}(\hat{\mathbf{s}}_t, G_{\theta_{\text{trans}}}(\hat{\mathbf{s}}_t, \mathbf{z})))]. \tag{10}$$

To further enforce pixel-wise fidelity of the prediction, we incorporate the L1 reconstruction loss:

$$\mathcal{L}_{L1}^{(t)} = \mathbb{E}_{(\hat{\mathbf{s}}_t, \hat{\mathbf{s}}_{t+1}), \mathbf{z}}[\|\hat{\mathbf{s}}_{t+1} - G_{\theta_{\text{trans}}}(\hat{\mathbf{s}}_t, \mathbf{z})\|_1]. \tag{11}$$

Here, $\mathbf{z}$ denotes the injected noise tensor modeling prediction uncertainty. In Eq. (9), λ balances perceptual realism with structural accuracy.

3 Results

3.1 A Generative State-Space Modeling Framework for Incomplete Observations

To address the challenge of capturing complex ocean physical dynamics from sparse satellite swath data, we constructed a data-driven generative state-space modeling framework (Fig. 2a). Targeting the rigorous demands of high-resolution ocean simulation, we implemented tailored designs within the model architecture. First, the Initial State Model (Fig. 2b) adopts a lightweight modified StyleGAN2-ADA (Karras, Aittala, et al., 2020), incorporating a Depthwise Separable Convolution (DWConv) strategy (Chollet, 2017) to significantly reduce computational complexity while maintaining generation quality. Second, the State Transition Model (Fig. 2c) utilizes a Stochastic U-Net (Ronneberger et al., 2015; Kohl et al., 2018); by injecting random noise at the decoder levels, it explicitly captures the inherent uncertainties within the physical evolution process. Complementing these generative components, we defined an explicit probabilistic Observation Model to bridge the generated states with real-world data, formulating observations as sparse measurements with Gaussian noise.

Furthermore, to effectively handle the intrinsic cyclic dependency between "model learning" (which requires complete data) and "state reconstruction" (which requires a trained model), we formalized the training process as an iterative optimization framework based on the Expectation-Maximization (EM) algorithm (Brajard et al., 2020) (Fig. 2d). Unlike traditional incremental learning, this system achieves iterative improvement directly from spatiotemporally varying sparse observational data by alternating between the E-step and the M-step.

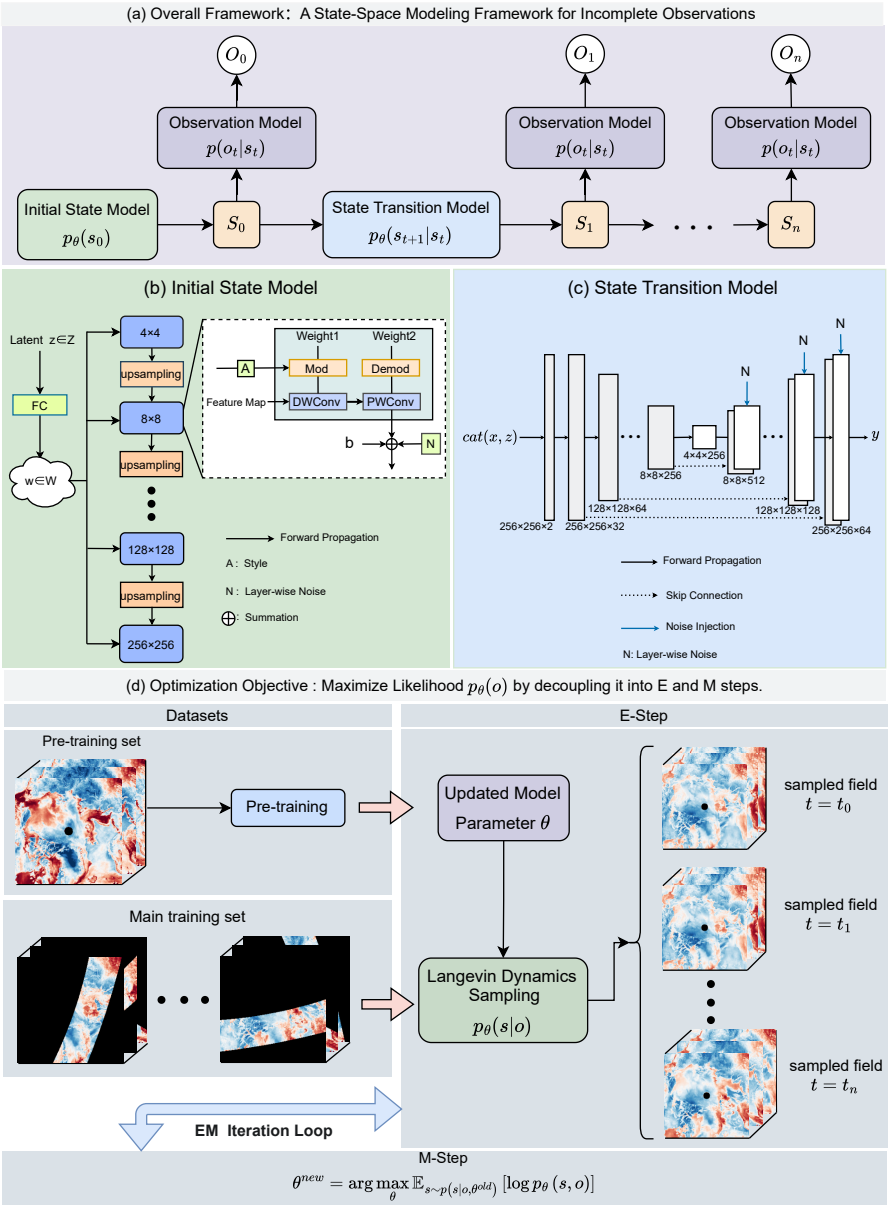

Figure 2: Schematic of the generative state-space modeling framework. **a**, Probabilistic graphical model illustrating the hidden Markov process. Latent complete physical states (S) evolve temporally and emit partial observations (O) via the observation model. **b**, **c**, Neural architectures for the generator components. **b**, The Initial State Model ($p_\theta(\mathbf{s}_0)$) utilizes a StyleGAN2-ADA backbone with modulated depthwise separable convolutions (DWConv + PWConv) to map latent code $\mathbf{w}$ to spatial fields. **c**, The State Transition Model ($p_\theta(\mathbf{s}_{t+1}|\mathbf{s}_t)$) employs a stochastic U-Net that processes concatenated inputs ($\text{cat}(\mathbf{x}, \mathbf{z})$) with multi-scale skip connections and noise injection (N). **d**, Optimization workflow based on the Expectation-Maximization (EM) algorithm. The pipeline begins with model initialization using a pre-training dataset. In the main loop, raw satellite swath data serve as the primary training dataset. The E-step infers the posterior of hidden states (S) via Langevin dynamics sampling (Sampled Field), which are subsequently used in the M-step to update model parameters θ according to the maximization objective shown in the bottom panel.

3.2 Experimental design and dual-track validation strategy

3.2.1 Experimental setup and datasets

To comprehensively evaluate performance across controlled and operational scenarios, we implemented a dual-track validation strategy. For rigorous quantitative assessment, we established an Observing System Simulation Experiment (OSSE) framework derived from the CMIP6 BCC-CSM2-MR climate model (Wu et al., 2019), treating daily sea surface temperature (tos) fields as the ground truth (Griffies et al., 2016). Synthetic incomplete observations were generated by applying structural masks—occluding approximately 67% of the spatial domain—superimposed with Gaussian noise to mimic sensor imperfections.

To assess operational generalization, we employed real-world Level-1 swath data from the FY-3D Microwave Radiation Imager (MWRI), focusing on the Arctic region ($> 66°$N). In contrast to the standardized simulation grid, this dataset is characterized by high-frequency yet non-uniform temporal sampling ($\sim$1.72 h intervals) driven by polar orbital dynamics, presenting naturally occurring along-track data voids without ground-truth references.

3.2.2 Progressive training protocol

To bridge the gap between limited data availability and high-fidelity modeling requirements, we implemented a rigorous two-stage strategy analogous to curriculum learning. The process commences with a warm-up phase, where the generative models are initialized via supervised pre-training on a limited subset of complete physical fields (approx. 800 samples), thereby establishing a robust prior over spatial statistics. Subsequently, the framework transitions to the core EM-based iterative learning phase utilizing a larger corpus of exclusively incomplete observations (approx. 1,500 samples). This transition enables the model to progressively internalize stochastic physical dynamics while simultaneously enhancing reconstruction fidelity through the alternating Expectation and Maximization steps, effectively unlocking the capability to learn directly from sparse data.

3.2.3 Task formulation

While our proposed framework is theoretically applicable to sequential problems of arbitrary length T, in this study, we focus on the fundamental unit of temporal evolution: the $T = 1$ scenario, comprising two consecutive frames ($t = 0, 1$). This setting allows us to balance computational costs while capturing core thermodynamic dynamics. Consequently, the specific experimental objective across both simulated and real-world scenarios is to reconstruct the complete system state $\mathbf{s} = \{\mathbf{s}_0, \mathbf{s}_1\}$ from the sparse observation frames $\mathbf{o} = \{\mathbf{o}_0, \mathbf{o}_1\}$.

3.3 High-fidelity reconstruction and forecasting in simulation

3.3.1 Evaluation metrics

To comprehensively assess model performance, we employed Root Mean Square Error (RMSE) (Wilks, 2011), Mean Absolute Error (MAE) (Wilks, 2011), and Structural Similarity Index (SSIM) (Wang et al., 2004) as evaluation metrics, capturing both pixel-level accuracy and structural fidelity. Crucially, to ensure physical relevance, all quantitative errors were calculated in degrees Celsius (°C) within the valid physical range of 15–35 °C.

3.3.2 Baseline model for comparison

To quantitatively demonstrate the efficacy of our iterative EM-based learning framework, we establish a baseline model, hereafter referred to as the Pre-trained Model. This model is derived from the initial warm-up phase of our training protocol. Specifically, it is trained via standard supervised learning on a limited set of complete ground-truth fields. The Final Model, in contrast, is the result of the full iterative training process, which refines the pre-trained parameters by learning directly from incomplete observations. This comparison is designed to isolate and quantify the performance gains attributable to the EM-driven learning from sparse data.

3.3.3 High-fidelity reconstruction under structural masks

Our central hypothesis is that the proposed Expectation–Maximization (EM) framework enables the model to bootstrap accurate physical representations from severely incomplete observations. Within an Observing System Simulation Experiment (OSSE), we take daily sea surface temperature (SST) fields from CMIP6 simulations as ground truth. Imperfect inputs are generated by applying structural masks together with additive Gaussian noise, emulating the irregular coverage characteristic of satellite retrievals.

Training follows a sequential split with two evaluated checkpoints. First, a **Pre-trained** baseline is obtained from a short warm-up stage on an early subset of complete, unmasked SST fields, which initializes the network and establishes basic spatiotemporal dynamics. The pipeline then proceeds to the main stage on the remaining data, where inputs are *exclusively* masked and noise-corrupted to mimic observation-like incompleteness. Starting from this baseline, we run 17 EM iterations: in each E-step, latent variables are refined via Langevin dynamics to reconstruct full spatiotemporal states; in each M-step, model parameters are updated with Adam using these reconstructions as targets. The **Final model** denotes the parameters after completing these EM iterations.

For evaluation, both checkpoints are tested under five structural occlusion patterns that remove the top, bottom, center, left, or right two-thirds of the spatial domain; reported metrics are averages across these scenarios. Quantitative results in Table 1 support the hypothesis: relative to the pre-trained baseline, the final model yields marked gains across deterministic scores for reconstructing the initial state $\mathbf{s}_0$ and the one-step-ahead forecast $\mathbf{s}_1$.

Table 1: Quantitative evaluation on the simulated dataset. The table demonstrates the performance improvement of state reconstruction ($\mathbf{s}_0$) and forecasting ($\mathbf{s}_1$) via EM iteration. RMSE and MAE are reported in degrees Celsius (°C).

Model	**States**	**RMSE (°C)** ↓	**MAE (°C)** ↓	**SSIM** ↑
Pre-trained Model	$\mathbf{s}_0$	0.9696	0.7034	0.9014
	$\mathbf{s}_1$	0.7587	0.4804	0.9138
Final Model	$\mathbf{s}_0$	**0.6969**	**0.4621**	**0.9355**
	$\mathbf{s}_1$	**0.5532**	**0.3197**	**0.9496**

Specifically, for the reconstruction of the initial state $\mathbf{s}_0$, the Final Model reduced the RMSE by 28.13% and the MAE by 34.30%, alongside a significant improvement in SSIM. This indicates that the EM iteration enables the model to

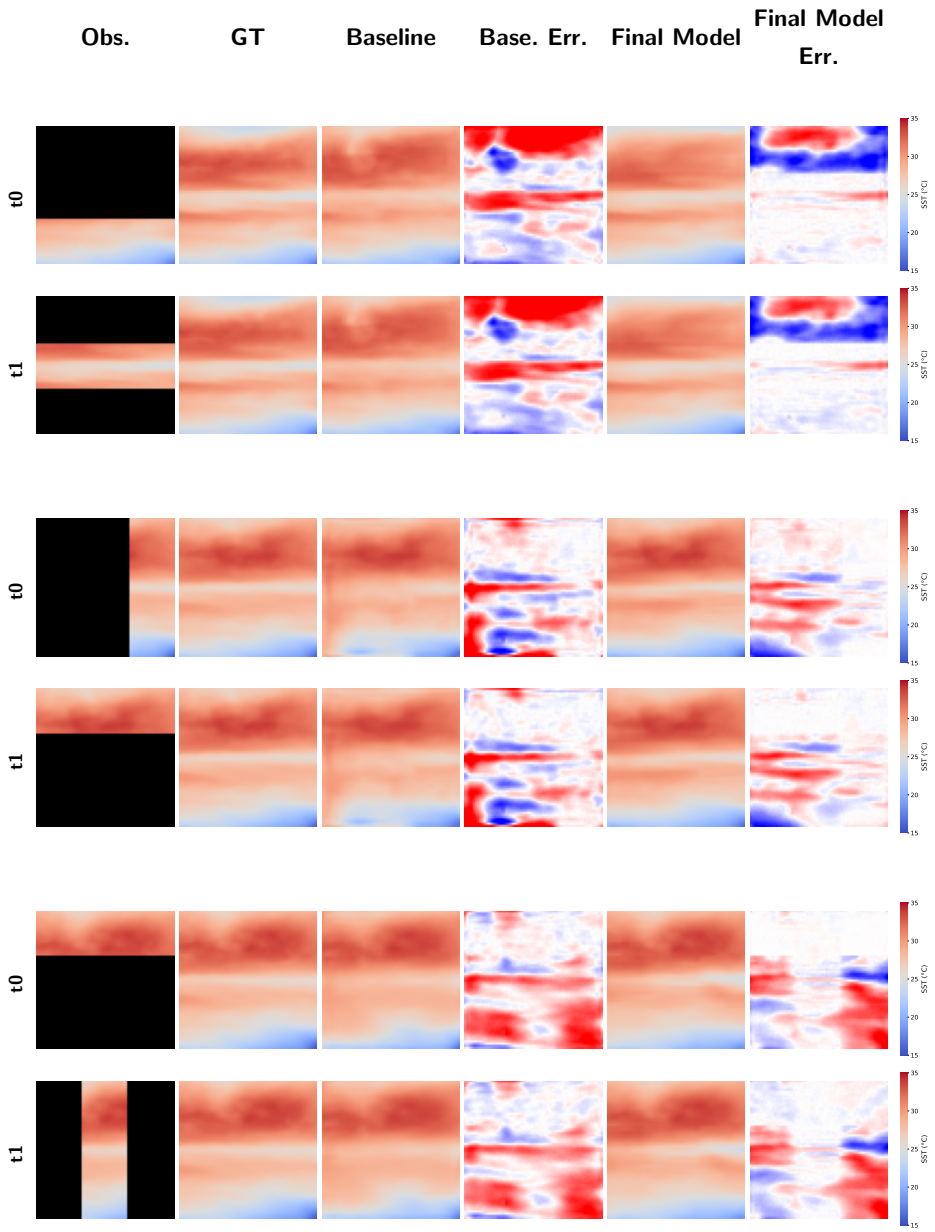

Figure 3: Qualitative comparison on representative simulation samples. Columns from left to right: Sparse Observation, Ground Truth, Baseline prediction, Baseline Error, Final Model output, Final Model Error, and Scale. The reduction in residual intensity in the final column highlights the model's ability to recover fine-scale structures.

learn a more realistic prior distribution of physical fields, thereby enhancing its data assimilation capability. The performance improvement for $\mathbf{s}_1$ is equally significant, with an RMSE reduction of 27.09% and an MAE reduction of 33.45%. Since the reconstruction of $\mathbf{s}_1$ relies heavily on the physical evolution from $\mathbf{s}_0$ (governed by the State Transition Model), this result strongly evidences that the EM framework significantly improves the model's precision in simulating physical dynamics.

An interesting phenomenon observed is that the reconstruction errors for $\mathbf{s}_1$ are consistently lower than those for $\mathbf{s}_0$. We hypothesize that this is because the reconstruction of $\mathbf{s}_1$ is constrained by the physical evolution from $\mathbf{s}_0$, whereas $\mathbf{s}_0$ lacks constraints from past physics.

To complement quantitative metrics, we examined representative reconstruction results under structural masking conditions (**Fig. 3**). The Pre-trained Model, while capturing global spatial patterns, exhibits a tendency towards over-smoothing in unobserved regions. This results in a loss of high-frequency physical details, as reflected by higher residual magnitudes in the error maps. Conversely, the Final Model recovers sharper and physically more coherent structures, including distinct sub-mesoscale eddies (McWilliams, 2016) and temperature fronts. The visible reduction in error magnitude across the masked regions suggests that the EM-driven iterative process effectively corrects the inference bias present in the initialization phase, leading to representations that are closer to the ground truth (Ross et al., 2023).

3.3.4 Forecasting capability of the State Transition Model

Beyond assessing the coupled reconstruction performance, we further isolated and evaluated the pure forecasting capability of the State Transition Model ($G_{\theta_{\text{trans}}}$) itself. To this end, we designed a standard one-step forecasting experiment.

In this setup, we directly used the complete, noise-free ground-truth states $\mathbf{s}_0$ from the simulated dataset as input. We then fed these perfect initial conditions into the state transition model to generate a one-step prediction, $\hat{\mathbf{s}}_1 = G_{\theta_{\text{trans}}}(\mathbf{s}_0)$. Comparing $\hat{\mathbf{s}}_1$ with the true subsequent state $\mathbf{s}_1$ isolates the transition model's skill in simulating physical dynamics. Table 2 summarizes one-step forecasting performance when $\mathbf{s}_0$ is taken from the ground truth.

Table 2: Performance evaluation of the one-step forecasting task. The input is the ground-truth state $\mathbf{s}_0$. Best results are in bold.

Model	**RMSE (°C)** ↓	**MAE (°C)** ↓	**SSIM** ↑
Pre-trained model	0.9263	0.7288	0.9306
Final model	**0.7431**	**0.5479**	**0.9408**

The results in Table 2 show that, even when provided with identical perfect initial conditions, the final model still outperforms the pre-trained model in forecasting accuracy. Specifically, the final model achieves noticeable reductions in both RMSE and MAE, together with a modest improvement in SSIM.

This substantial performance gain provides strong evidence that our EM framework does more than just end-to-end data fitting; it genuinely enables the State Transition Model itself to distill more accurate and authentic physical evolution laws from the iterative process of reconstruction and refinement. This result confirms

that the model, trained via our proposed method, evolves into a more powerful and physically consistent standalone predictor.

3.4 Generalization to real-world satellite observations

3.4.1 Pre-training set construction for real-world data

Beyond the quantitative benchmarks on simulated datasets, validating robustness against intrinsic data voids is crucial for operational viability. We applied the framework to real-world Level-1 swath data from the FY-3D satellite, where data incompleteness stems from orbital scanning geometries rather than artificial masks. In the absence of complete ground truth, we prioritized a qualitative assessment. To establish the Pre-trained Model (Baseline) for this dataset, we constructed a pre-training set via a multi-step pipeline: raw satellite swaths were first georeferenced and projected onto a regular Arctic grid (north of 66°N); then, the approximately 14 daily orbital passes were sequentially aggregated to form a complete daily composite; finally, to recover the high-frequency temporal dynamics consistent with the satellite's orbital cycle ($\sim$1.72 h), these daily composites were temporally interpolated to generate a sequence of approximately 14 spatially complete fields per day (Griffies et al., 2016).

3.4.2 Structural fidelity and texture recovery

Fig. 4 consistently reveals a marked enhancement in structural fidelity across all three distinct observation scenarios. The Pre-trained Model (middle column) exhibits a characteristic "over-smoothing" bias, producing fields that are spatially continuous but lack high-frequency definition. This is particularly evident in the blurred ice edges and the homogenized texture within the main ice pack. In contrast, the Final Model (right column) successfully recovers sharp thermal gradients and intricate structural patterns, such as fine ice filaments and fragmented floes, within the unobserved regions. This suggests that the model has not merely learned to replicate smoothed, interpolated states, but has learned to sample from the complex manifold of valid physical states.

3.4.3 Spatiotemporal coherence and physical plausibility

Furthermore, a crucial test lies in the model's ability to maintain dynamic consistency over time. A detailed inspection of the temporal evolution from T_0 to T_1 highlights the superior performance of the Final Model. While the Baseline's evolution of fine-scale features appears less coherent across the transition in the middle column, the Final Model renders a more physically plausible evolution across all scenarios. In the first case (right column), the ice formations exhibit a clear, coherent rotational motion from T_0 to T_1, accompanied by a slight downward drift. This high-fidelity dynamic reconstruction is not an isolated success, as similar physically consistent evolutions are observed in the other scenarios in the right column as well, aligned with known Arctic sea-ice rheology.

This robust spatiotemporal consistency across diverse observation geometries strongly evidences that through the cyclic interplay of E-step and M-step, the framework has successfully internalized the governing physical laws. The model has learned not merely to "inpaint" static images, but to simulate a dynamic trajectory, effectively avoiding the temporal flickering often associated with simple interpolation methods (Brajard et al., 2020). The ability to generalize these learned laws to novel, real-world scenarios highlights its immense potential for data-sparse operational oceanography (Bell et al., 2015).

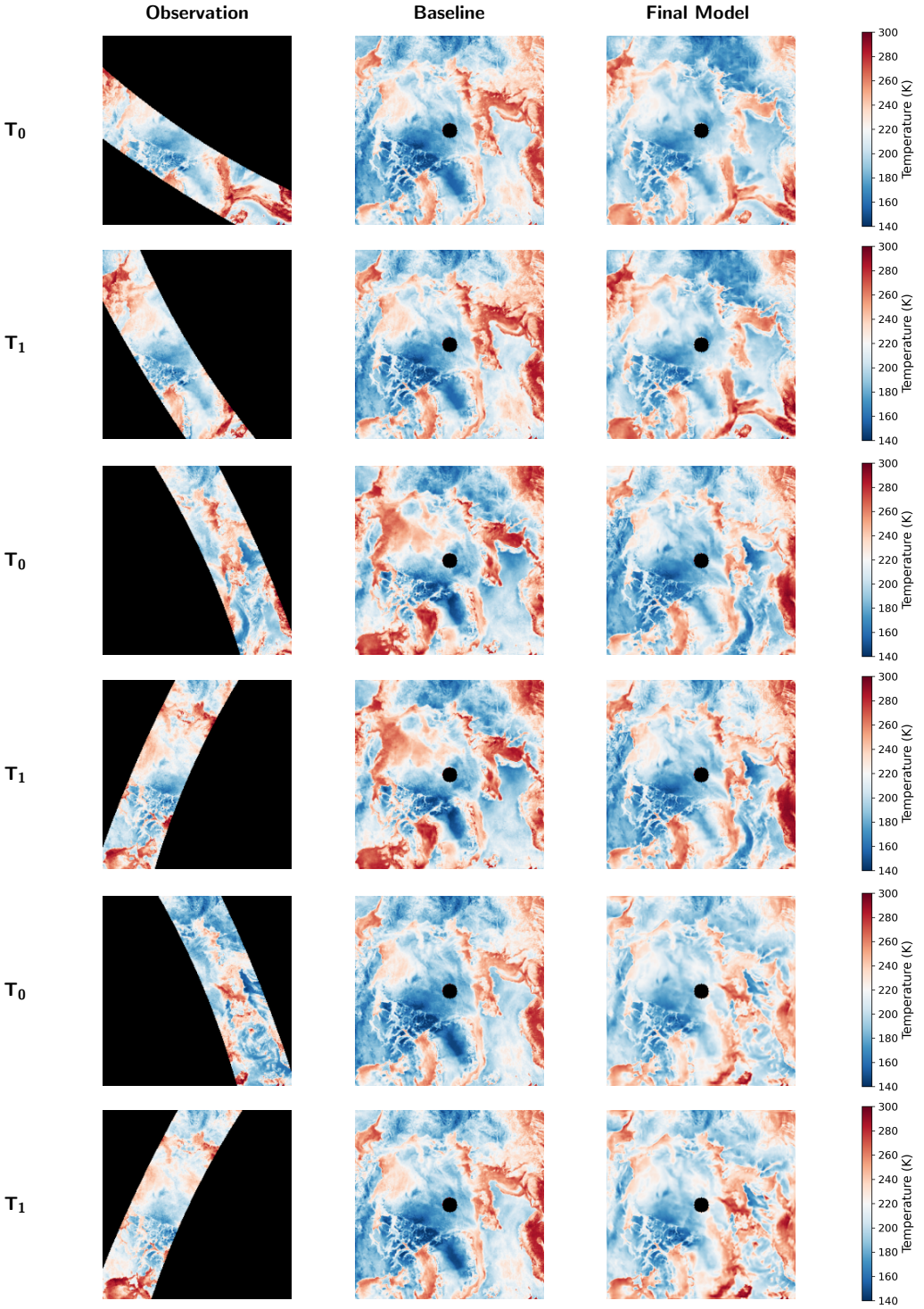

Figure 4: Robust spatiotemporal reconstruction from sparse satellite swaths. The figure compares raw observations, pre-trained baseline, and final model results across consecutive time steps for three samples. The framework ensures temporal continuity and recovers fine-scale structures.

4 Discussion and Conclusion

4.1 Breaking the data dependency barrier

The central achievement of this study is the establishment and validation of a generative state-space modeling framework capable of iterative learning directly from sparse, incomplete observations. In the rapidly evolving landscape of AI for Earth System Science (AI4ESS), the reliance on pristine, complete reanalysis datasets (e.g., ERA5) has long been a bottleneck, limiting models to the role of "emulators" rather than independent "simulators." Our work fundamentally challenges this status quo. By demonstrating that an AI model can be trained end-to-end using only partial observations, we provide a tangible pathway toward autonomous AI Earth system modeling, liberating the field from the computational and resolution constraints imposed by traditional data assimilation systems.

4.2 The efficacy of the EM paradigm

Our experimental results provide compelling evidence that the Expectation-Maximization (EM) algorithm serves as a robust paradigm for addressing the inherent circular dependency between "state reconstruction" and "model learning." As evidenced by the significant performance leap from the "Pre-trained Model" to the "Final Model" (Table 1), the framework successfully instigates a "bootstrapping" learning process. This is a critical finding: it implies that deep generative models possess the capacity to refine their understanding of physical dynamics by leveraging their own imperfect reconstructions. Through the iterative interplay of the E-step (probabilistic state inference) and the M-step (parameter optimization), the model converges from a coarse approximation to a high-performance state that captures intricate physical dynamics, effectively distilling coherent physical laws from fragmented data.

4.3 Implications for future Earth system modeling

This core discovery paves a new, viable path for the development of next-generation AI Earth system models. It demonstrates that perfect supervision is not a prerequisite for learning complex physics. Instead, by designing sophisticated iterative learning frameworks, we can extract knowledge directly from the noisy, imperfect observations of the real world. This capability is particularly transformative for regions or variables where high-quality reanalysis data is scarce or nonexistent (e.g., deep ocean variables or polar regions). By enabling models to learn from raw satellite or in-situ observations, we move closer to constructing independent, powerful physical process simulators that can complement, or potentially supersede, traditional numerical methods in specific domains.

4.4 Limitations and future outlook

While our framework shows immense promise, we dge certain limitations. First, the computational cost of the Langevin dynamics sampling in the E-step is non-trivial, potentially limiting scalability for high-frequency, global-scale operational forecasting. Future work will explore accelerating this process via variational inference (Kingma & Welling, 2013) or distilled diffusion solvers (Salimans & Ho, 2022). Second, while we achieve high fidelity, ensuring strict adherence to conservation laws (e.g., mass and energy conservation) remains a challenge for purely data-driven generative models. Integrating physical constraints (Physics-Informed Machine Learning) into our probabilistic framework represents a promising direction (Raissi et al., 2019). Despite these challenges, this work stands as a foundational step to-

wards a more data-efficient and physically grounded era of AI meteorology and oceanography.

Appendix A Derivation of the Generative State-Space Model

A1 Maximum Likelihood Estimation Objective

Consider a dataset of partial observations $\mathbf{o} = (\mathbf{o}^{(1)}, \mathbf{o}^{(2)}, \ldots, \mathbf{o}^{(N)})$. Here, the superscript within the parentheses denotes different samples in the dataset, rather than time points. It should be noted that each sample in the dataset may contain observations from multiple different time points. We treat different samples in the dataset as independent and identically distributed. Afterwards, the goal of Maximum Likelihood Estimation (MLE) is to find the optimal parameters θ^* that maximize the probability of the data:

$$\theta^* = \arg\max_\theta p_\theta(\mathbf{o}) = \arg\max_\theta \prod_{i=1}^{N} p_\theta(\mathbf{o}^{(i)}). \tag{A1}$$

For simplicity, we maximize the log-likelihood to transform the product into a sum:

$$\theta^* = \arg\max_\theta \log p_\theta(\mathbf{o}) = \arg\max_\theta \sum_{i=1}^{N} \log p_\theta(\mathbf{o}^{(i)}). \tag{A2}$$

In our framework, we model the joint probability of the system state and the observation. Therefore, the probability of the observation is actually a marginal probability. Hence, to get $p_\theta(o^{(i)})$ in Eq. (A2), we need to integrate over the state space:

$$p_\theta(\mathbf{o}^{(i)}) = \int p_\theta(\mathbf{s}^{(i)}, \mathbf{o}^{(i)})\, \mathrm{d}\mathbf{s}^{(i)} = \int p_\theta(\mathbf{s}^{(i)}) p_\theta(\mathbf{o}^{(i)} \mid \mathbf{s}^{(i)})\, \mathrm{d}\mathbf{s}^{(i)}. \tag{A3}$$

In the framework, $p_\theta(\mathbf{s}^{(i)})$ is modeled through sophisticated neural networks (i.e. initial state model and state transition model), making the integral difficult to calculate. Consequently, the objective defined in Eq. (A2) can not be directly optimized.

A2 Decomposition of the Log-Likelihood

For the sake of discussion, let's single out and analyze a specific term $\log p_\theta(\mathbf{o}^{(i)})$ from Eq. (A2). In fact, $\log p_\theta(\mathbf{o}^{(i)})$ can be decomposed as follows:

$$\log p_\theta(\mathbf{o}^{(i)}) = \log p_\theta(\mathbf{s}^{(i)}, \mathbf{o}^{(i)}) - \log p_\theta(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)}). \tag{A4}$$

Furthermore, we introduce an auxiliary distribution $q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})$ satisfying:

$$q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)}) \geq 0, \quad \int q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})\, \mathrm{d}\mathbf{s}^{(i)} = 1. \tag{A5}$$

Taking the expectation of the log likelihood in Eq. (A4) with respect to q, we obtain:

$$\mathbb{E}_q[\log p_\theta(\mathbf{o}^{(i)})] = \mathbb{E}_q[\log p_\theta(\mathbf{s}^{(i)}, \mathbf{o}^{(i)})] - \mathbb{E}_q[\log p_\theta(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})]. \tag{A6}$$

The term on the left-hand side of Eq. (A6) can be expanded as follows:

$$\begin{aligned} \mathbb{E}_q[\log p_\theta(\mathbf{o}^{(i)})] &= \int q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)}) \log p_\theta(\mathbf{o}^{(i)})\, \mathrm{d}\mathbf{s}^{(i)} \\ &= \log p_\theta(\mathbf{o}^{(i)}) \int q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})\, \mathrm{d}\mathbf{s}^{(i)} \\ &= \log p_\theta(\mathbf{o}^{(i)}). \end{aligned} \tag{A7}$$

Here, we utilized the definition of q in Eq. (A5). Combining the results from Eq. (A6) and Eq. (A7), we obtain the decomposition of the marginal log-likelihood:

$$\log p_\theta(\mathbf{o}^{(i)}) = \mathbb{E}_q\left[\log p_\theta(\mathbf{s}^{(i)}, \mathbf{o}^{(i)})\right] - \mathbb{E}_q\left[\log p_\theta(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})\right]. \tag{A8}$$

The second term on the right-hand side of Eq. (A8) relates to the cross-entropy between the auxiliary distribution $q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})$ and the posterior $p_\theta(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})$.

A3 Iterative Optimization via the EM Algorithm

We employ an iterative optimization strategy. Let $\theta^{(n)}$ denote the model parameters at the n-th iteration. Since q in Eq. (A8) can be any distribution, we can readily set it to be equal to the exact posterior under the current parameters:

$$q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)}) = p_{\theta^{(n)}}(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)}). \tag{A9}$$

In this situation, the second term on the right-hand side of Eq. (A8) becomes the entropy of q. According to Gibbs' Inequality, any update to θ will result in an increase in the second term on the right-hand side of Eq. (A8). Therefore, if we can increase the expected value of the complete-data log-likelihood (the first term on the right-hand side of Eq. (A8)) by updating θ, we will also be able to increase the log-likelihood of the observed data (the left-hand side of Eq. (A8)). Incorporating Eq. (A2), the M-step at iteration $n+1$ solves

$$\theta^{(n+1)} = \arg\max_\theta \sum_{i=1}^{N} \mathbb{E}_{\mathbf{s}^{(i)} \sim q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})}\left[\log p_\theta(\mathbf{s}^{(i)}, \mathbf{o}^{(i)})\right]. \tag{A10}$$

A30.1 Monte Carlo Approximation. Since the analytic formula of the auxiliary distribution q is unknown, exact computation of the expectations in Eq. (A10) is infeasible, necessitating the employment of Monte Carlo integration. Specifically, for an observation sample $\mathbf{o}^{(i)}$, we can approximate the expectation by drawing samples from the auxiliary distribution:

$$\mathbb{E}_{\mathbf{s}^{(i)} \sim q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})}[\log p_\theta(\mathbf{s}^{(i)}, \mathbf{o}^{(i)})] \approx \frac{1}{M} \sum_{j=1}^{M} \log p_\theta(\mathbf{s}^{(i,j)}, \mathbf{o}^{(i)}). \tag{A11}$$

In Eq. (A11), $\mathbf{s}^{(i,j)}$ denotes a sample drawn from $q(\mathbf{s}^{(i)} \mid \mathbf{o}^{(i)})$, and M signifies the number of samples for approximation. Theoretically, the larger the value of M, the more accurate the approximation of the expectation. However, in our experiments, we found that satisfactory results can be achieved even when $M = 1$. Therefore, to balance computational efficiency with accuracy, we set $M = 1$ in the implementation. By substituting Eq. (A11) back into Eq. (A10), we can get

$$\theta^{(n+1)} = \arg\max_\theta \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{M} \log p_\theta(\mathbf{s}^{(i,j)}, \mathbf{o}^{(i)}). \tag{A12}$$

Since M is always equal to 1 in the implementation, we can omit it for simplicity. In this case, Eq. (A12) further simplifies to

$$\theta^{(n+1)} = \arg\max_\theta \sum_{i=1}^{N} \log p_\theta(\mathbf{s}^{(i)}, \mathbf{o}^{(i)}). \tag{A13}$$

Eq. (A13) matches Eq. (2) in the main text; the optimization objective remains Eq. (A2). Here, $\mathbf{s}^{(i)}$ represents the high-fidelity physical fields sampled during the E-step. The superscript "(i)" corresponds to different instances in the dataset, ensuring that each system state trajectory is strictly paired with its specific observation sequence.

A4 E-Step and M-Step Decoupling

A41 E-Step: Efficient State Sampling

Consider the states and observations over a sequence of time steps $t = 0, 1, \ldots, T$:

$$\mathbf{s} = (\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_T), \quad \mathbf{o} = (\mathbf{o}_0, \mathbf{o}_1, \ldots, \mathbf{o}_T). \tag{A14}$$

The posterior distribution of $\mathbf{s}$ given $\mathbf{o}$ is given by Bayes' rule:

$$p_\theta(\mathbf{s} \mid \mathbf{o}) \propto p_\theta(\mathbf{o} \mid \mathbf{s})\, p_\theta(\mathbf{s}). \tag{A15}$$

Assuming conditional independence of observations, the likelihood factorizes as the product over all time steps:

$$p_\theta(\mathbf{o} \mid \mathbf{s}) = \prod_{t=0}^{T} p_\theta(\mathbf{o}_t \mid \mathbf{s}_t). \tag{A16}$$

Combining Eqs. (A15) and (A16), the gradient of the log-posterior with respect to the state $\mathbf{s}$ is:

$$\nabla_{\mathbf{s}} \log p_\theta(\mathbf{s} \mid \mathbf{o}) = \sum_{t=0}^{T} \nabla_{\mathbf{s}} \log p_\theta(\mathbf{o}_t \mid \mathbf{s}_t) + \nabla_{\mathbf{s}} \log p_\theta(\mathbf{s}). \tag{A17}$$

Directly utilizing Eq. (A17) to perform Langevin dynamics sampling in the $\mathbf{s}$-space is intractable due to the high dimensionality of the physical state $\mathbf{s}$ and the complexity of its prior distribution, which renders the computation of the posterior score function prohibitively difficult. Instead, we leverage the generative mapping $\mathbf{s} = \mathcal{G}(\mathbf{z})$ (where $\mathbf{z} = (\mathbf{z}_0, \ldots, \mathbf{z}_T)$ is a lower-dimensional latent variable) and perform Langevin dynamics sampling in the latent space $\mathbf{z}$ to approximate the posterior $p_\theta(\mathbf{z} \mid \mathbf{o})$. The gradient of the log-posterior with respect to $\mathbf{z}$ is decomposed as:

$$\nabla_{\mathbf{z}} \log p_\theta(\mathbf{z} \mid \mathbf{o}) = \nabla_{\mathbf{z}} \log p_\theta(\mathbf{o} \mid \mathbf{z}) + \nabla_{\mathbf{z}} \log p_\theta(\mathbf{z}). \tag{A18}$$

Expanding the observation likelihood term (where $\mathbf{s}_t$ is a function of $\mathbf{z}$):

$$\nabla_{\mathbf{z}} \log p_\theta(\mathbf{o} \mid \mathbf{z}) = \sum_{t=0}^{T} \nabla_{\mathbf{z}} \log p_\theta(\mathbf{o}_t \mid \mathbf{s}_t). \tag{A19}$$

Thus, the score function with respect to the latent variable $\mathbf{z}$ is given by:

$$\nabla_{\mathbf{z}} \log p_\theta(\mathbf{z} \mid \mathbf{o}) = \sum_{t=0}^{T} \nabla_{\mathbf{z}} \log p_\theta(\mathbf{o}_t \mid \mathbf{s}_t) + \nabla_{\mathbf{z}} \log p_\theta(\mathbf{z}). \tag{A20}$$

A42 M-Step: Decoupling the Parameter Updates

The M-step in Eq. (A13) maximizes the complete-data log-likelihood over the N sampled trajectories $\{(\mathbf{s}^{(i)}, \mathbf{o}^{(i)})\}_{i=1}^{N}$. We now factorize the joint probability $p_\theta(\mathbf{s}, \mathbf{o})$ based on our generative state-space model structure. For a single trajectory $(\mathbf{s}^{(i)}, \mathbf{o}^{(i)}) = \{(\mathbf{s}_0^{(i)}, \mathbf{o}_0^{(i)}), (\mathbf{s}_1^{(i)}, \mathbf{o}_1^{(i)}) \ldots, (\mathbf{s}_T^{(i)}, \mathbf{o}_T^{(i)})\}$, the log-likelihood decomposes as:

$$\log p_\theta(\mathbf{s}^{(i)}, \mathbf{o}^{(i)}) = \log p_{\theta_{\text{init}}}(\mathbf{s}_0^{(i)}) + \sum_{t=1}^{T} \log p_{\theta_{\text{trans}}}(\mathbf{s}_t^{(i)} \mid \mathbf{s}_{t-1}^{(i)}) + \sum_{t=0}^{T} \log p(\mathbf{o}_t^{(i)} \mid \mathbf{s}_t^{(i)}). \tag{A21}$$

Since the observation likelihood terms $\log p(\mathbf{o}_t^{(i)} \mid \mathbf{s}_t^{(i)})$ are explicitly modeled by the Gaussian distribution and possess no learnable parameters, they are constant with respect to θ. Therefore, these terms can be omitted in the M-step update.

Substituting the decomposition in Eq. (A21) into Eq. (A13), and noting that θ_{init} and θ_{trans} are disjoint parameter sets (θ_{init} and θ_{trans} are encapsulated within

the initial state model and the state transition model, respectively), the M-step update decouples into two independent sub-problems:

1. **Initial State Model Update:**

$$\theta_{\text{init}}^{\text{new}} = \arg\max_{\theta_{\text{init}}} \sum_{i=1}^{N} \log p_{\theta_{\text{init}}}(\mathbf{s}_0^{(i)}) \tag{A22}$$

   This is equivalent to minimizing the negative log-likelihood, corresponding to the loss $\mathcal{L}_{\text{initial}}$ defined in the main text.

2. **State Transition Model Update:**

$$\theta_{\text{trans}}^{\text{new}} = \arg\max_{\theta_{\text{trans}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \log p_{\theta_{\text{trans}}}(\mathbf{s}_t^{(i)} \mid \mathbf{s}_{t-1}^{(i)}) \tag{A23}$$

   This corresponds to maximizing the conditional log-likelihood of the transitions, which maps to the loss $\mathcal{L}_{\text{transition}}$.

Thus, the M-step effectively minimizes the combined loss $\mathcal{L}_M = \mathcal{L}_{\text{initial}} + \mathcal{L}_{\text{transition}}$.

Appendix B Open Research

Data Availability Statement

CMIP6 model output was obtained from the Earth System Grid Federation (ESGF) CMIP6 interface hosted at the Lawrence Livermore National Laboratory node (`https://esgf-node.llnl.gov/projects/cmip6/`). This study uses daily sea surface temperature (`tos`; CMIP6 table `Oday`) from the Beijing Climate Center Climate System Model version 2 with medium resolution (`BCC-CSM2-MR`) `historical` experiment, variant label `r1i1p1f1`, on the native ocean grid (`gn`).

Satellite data are FengYun-3D (`FY-3D`) Microwave Radiation Imager (`MWRI`) Level-1 (`L1`) products distributed by the National Satellite Meteorological Center (NSMC), China Meteorological Administration. Data access, product metadata, and distribution policies are available through the NSMC data portal: `https://data.nsmc.org.cn/DataPortal/cn/data/dataset.html?dataTypeCode=L1&satelliteCode=FY3D&instrumentTypeCode=MWRI`. Users must comply with NSMC registration requirements, data-use terms, and official citation guidance for the specific product version and observation period used in this study.

Software and Code Availability Statement

The code used to implement the methods of this study is publicly available in the GitHub repository `https://github.com/kyy-logs/Incomplete-Observations-Boost-Evo--lutionary-Performance-in-Ocean-Modeling`. The software is distributed under the license terms provided in the repository file `LICENSE.txt`. Installation, dependencies, and containerized execution are described in the repository `README.md` and `Dockerfile`.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Natural Science Foundation of China (Grant No. 42406192), the Fundamental Research Funds for the Central Universities (Grant No. 202413040), the National Science and Technology Major Project of China (Grant No. 2022ZD0117201), the Key R&D Program of Shandong Province (Grant No. 2025CXPT185), and the Postdoctoral Project of Qingdao (Grant No. QDBSH20240102021).

We dge the World Climate Research Programme (WCRP) and the Climate Model Intercomparison Project (CMIP6) for providing the climate model outputs. We thank the Beijing Climate Center (BCC) for making the BCC-CSM2-MR historical simulations available through the Earth System Grid Federation (ESGF, `https://esgf-node.llnl.gov/projects/cmip6/`). We also thank the National Satellite Meteorological Center (NSMC), China Meteorological Administration, for providing the FengYun-3D (FY-3D) Microwave Radiation Imager (MWRI) Level-1 data (`https://data.nsmc.org.cn`).

References

Bell, M. J., Schiller, A., Le Traon, P.-Y., Smith, N. R., Dombrowsky, E., & Wilmer-Becker, K. (2015). An introduction to GODAE OceanView. *Journal of Operational Oceanography*, *8*(S1), s2–s11. Retrieved from `https://doi.org/10.1080/1755876X.2015.1022041` doi: 10.1080/1755876X.2015.1022041

Bennett, A. F. (2002). *Inverse modeling of the ocean and atmosphere*. Cambridge University Press. doi: 10.1017/CBO9780511535895

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, *619*, 533–538. doi: 10.1038/s41586-023-06185-3

Bocquet, M., Brajard, J., Carrassi, A., & Bertino, L. (2019). Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. *Nonlinear Processes in Geophysics*, *26*(3), 143–162. doi: 10.5194/npg-26-143-2019

Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. *Journal of Computational Science*, *44*, 101171. Retrieved from `https://www.sciencedirect.com/science/article/abs/pii/S1877750320304725` doi: 10.1016/j.jocs.2020.101171

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., et al. (2023). Fengwu: Pushing the skillful global medium-range weather forecast out to 10 days. *arXiv preprint arXiv:2304.02948*. Retrieved from `https://arxiv.org/abs/2304.02948`

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition (cvpr)* (pp. 1251–1258). doi: 10.1109/CVPR.2017.195

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x

Evensen, G. (2009). *Data assimilation: The ensemble kalman filter*. Springer Science & Business Media. doi: 10.1007/978-3-642-03711-5

Fablet, R., Beauchamp, L., Drumetz, L., & Rousseau, F. (2021). Joint interpolation and representation learning for irregular sampled satellite-derived geophysical fields. *Frontiers in Applied Mathematics and Statistics*, *7*, 655224. doi: 10.3389/fams.2021.655224

Feng, D., Tan, Z., Lin, Z., Xu, D., Yu, C.-W., He, Q., et al. (2025). A comparative study of physics-informed and data-driven neural networks for compound

flood simulation at river-ocean interfaces: A case study of hurricane irene. *Journal of Geophysical Research: Machine Learning and Computation*, *2*(4). doi: 10.1029/2025jh000758

Forget, G., Campin, J.-M., Heimbach, P., Hill, C. N., Ponte, R. M., & Wunsch, C. (2015). Ecco version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*, *8*(10), 3071–3104. doi: 10.5194/gmd-8-3071-2015

Ghahramani, Z., & Roweis, S. T. (1998). Learning nonlinear dynamical systems using an em algorithm. In *Advances in neural information processing systems (neurips)* (Vol. 11). Retrieved from `https://proceedings.neurips.cc/paper/1998/hash/34ed066df378efacc9b924ec161e7639-Abstract.html`

Ghil, M., & Malanotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. *Advances in Geophysics*, *33*, 141–266. doi: 10.1016/S0065-2687(08)60442-2

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems (neurips)* (Vol. 27). Retrieved from `https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html`

Griffies, S. M., Danabasoglu, G., Durack, P. J., et al. (2016). Omip contribution to cmip6: experimental and diagnostic protocol for the physical component of the ocean model intercomparison project. *Geoscientific Model Development*, *9*, 3231–3291. doi: 10.5194/gmd-9-3231-2016

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi: 10.1002/qj.3803

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the ieee conference on computer vision and pattern recognition (cvpr)* (pp. 1125–1134). doi: 10.1109/CVPR.2017.632

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press. doi: 10.1017/CBO9780511802270

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. In *Advances in neural information processing systems (neurips)* (Vol. 33, pp. 12104–12114). Retrieved from `https://proceedings.neurips.cc/paper/2020/hash/8d30aa96e72440759f74bd2306c1fa3d-Abstract.html`

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (cvpr)* (pp. 4401–4410). doi: 10.1109/CVPR.2019.00453

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of StyleGAN. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (cvpr)* (pp. 8110–8119). doi: 10.1109/CVPR42600.2020.00813

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*. Retrieved from `https://arxiv.org/abs/1312.6114`

Kohl, S. A. A., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J. R., Maier-Hein, K. H., et al. (2018). A probabilistic U-Net for segmentation of ambiguous images. In *Advances in neural information processing systems (neurips)* (Vol. 31). Retrieved from `https://arxiv.org/abs/1806.05034`

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416–1421. doi: 10.1126/science.adi2336

Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *112*(474), 1177–1194. doi: 10.1002/qj.49711247414

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (Vol. 30, p. 3). Retrieved from https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf

Mardani, M., Brennan, N., Pinheiro, P., et al. (2023). Generative residual diffusion modeling for super-resolution and forecasting of high-resolution weather and climate. *arXiv preprint arXiv:2309.15214*. Retrieved from https://arxiv.org/abs/2309.15214

McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *472*(2191), 20160117. doi: 10.1098/rspa.2016.0117

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*. Retrieved from https://arxiv.org/abs/1411.1784

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (icml)* (pp. 807–814). Retrieved from https://mlanthology.org/icml/2010/nair2010icml-rectified/

Nguyen, T., Brandstetter, J., Kapoor, A., Campbell, J. K., & Gupta, A. (2023). Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*. Retrieved from https://arxiv.org/abs/2301.10343

Pang, B., Han, T., Nijkamp, E., Zhu, S.-C., & Wu, Y. N. (2021). Learning latent space energy-based models via MCMC. In *Proceedings of the 38th international conference on machine learning (icml)* (Vol. 139, pp. 8394–8404). Retrieved from http://proceedings.mlr.press/v139/pang21a.html

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). Fourcastnet: A global data-driven high-resolution forecasting model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*. Retrieved from https://arxiv.org/abs/2202.11214

Peyron, M., Fillion, A., Gürol, S., Marchais, V., Gratton, S., Boudier, P., & Goret, G. (2021, September). Latent space data assimilation by using deep learning. *Quarterly Journal of the Royal Meteorological Society*, *147*(740), 3759–3777. Retrieved from https://doi.org/10.1002/qj.4153 doi: 10.1002/qj.4153

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. doi: 10.1016/j.jcp.2018.10.045

Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., & Januschowski, T. (2018). Deep state space models for time series forecasting. In *Advances in neural information processing systems (neurips)* (Vol. 31). Retrieved from https://proceedings.neurips.cc/paper/2018/hash/5cf68969faa67deb9e8358e2bf9f8545-Abstract.html

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, *597*, 672–677. doi: 10.1038/s41586-021-03854-z

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*, 195–204. doi: 10.1038/s41586-019-0912-1

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the international conference on medical image computing and computer-assisted intervention (miccai)* (pp. 234–241). doi: 10.1007/978-3-319-24574-4_28

Ross, A., Li, Z., Perezhogin, P., Fernandez-Granda, C., & Zanna, L. (2023). Bench-

marking of machine learning ocean subgrid parameterizations in an idealized model. *Journal of Advances in Modeling Earth Systems*, *15*(1), e2022MS003258. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003258 doi: 10.1029/2022MS003258

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., et al. (2019). Inferring causation from time series in earth system sciences. *Nature Communications*, *10*(1), 2553. doi: 10.1038/s41467-019-10105-3

Salimans, T., & Ho, J. (2022). Progressive distillation for fast sampling of diffusion models. In *International conference on learning representations (iclr)*. Retrieved from https://openreview.net/forum?id=TIdIXIpzhoI

Sonnewald, M., Wunsch, C., & Heimbach, P. (2019). Unsupervised learning reveals geography of global ocean dynamical regions. *Earth and Space Science*, *6*(5), 784–794. doi: 10.1029/2018EA000519

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. doi: 10.1109/TIP.2003.819861

Wei, G. C. G., & Tanner, M. A. (1990). A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, *85*(411), 699–704. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10474930 doi: 10.1080/01621459.1990.10474930

Welling, M., & Teh, Y. W. (2011). Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (icml)* (pp. 681–688). Retrieved from https://icml.cc/2011/papers/398_icmlpaper.pdf

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (3rd ed., Vol. 100). Amsterdam; Boston: Academic Press. Retrieved from https://linkinghub.elsevier.com/retrieve/isbn/9780123850225

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., ... others (2019). The beijing climate center climate system model (bcc-csm): the main progress from cmip5 to cmip6. *Geoscientific Model Development*, *12*(4), 1573–1600. doi: 10.5194/gmd-12-1573-2019

Zhang, X., Zhou, Y., Ning, C., & Li, X. (2025). Physically guided gan-based super resolution of buoy observations for internal solitary waves. *Journal of Geophysical Research: Machine Learning and Computation*, *2*(4). doi: 10.1029/2025jh000951