JAMES Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2025MS005717

Climate-Adaptive and Cascade-Constrained Machine Learning Prediction for Sea Surface Height Under Greenhouse Warming

Tianmu Zheng[1], Ru Chen[1], Xin Su[1], Julian Mak[2,3], Gang Huang[4,5], and Bingzheng Yan[1]

[1]Tianjin Key Laboratory for Marine Environmental Research and Service, School of Marine Science and Technology, Tianjin University, Tianjin, China, [2]Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong, China, [3]National Oceanography Centre, Southampton, UK, [4]National Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [5]University of Chinese Academy of Sciences, Beijing, China

Key Points:

- A Dual-Attention ConvLSTM model, with a kinetic energy cascade constraint, is developed for sea surface height prediction
- The machine learning model generalizes well from the present-day climate to greenhouse-warming conditions
- Incorporating the kinetic energy cascade constraint improves machine learning performance in both current and greenhouse warming climates

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

R. Chen,
ruchen@tju.edu.cn

Citation:

Zheng, T., Chen, R., Su, X., Mak, J., Huang, G., & Yan, B. (2026). Climate-adaptive and cascade-constrained machine learning prediction for sea surface height under greenhouse warming. *Journal of Advances in Modeling Earth Systems*, *18*, e2025MS005717. https://doi.org/10.1029/2025MS005717

Received 22 DEC 2025
Accepted 1 JUL 2026

Author Contributions:

Conceptualization: Tianmu Zheng, Ru Chen
Formal analysis: Tianmu Zheng, Julian Mak, Bingzheng Yan
Funding acquisition: Ru Chen
Methodology: Tianmu Zheng, Ru Chen, Xin Su, Julian Mak, Gang Huang
Project administration: Ru Chen
Software: Tianmu Zheng, Xin Su
Supervision: Ru Chen
Validation: Xin Su, Gang Huang, Bingzheng Yan

Abstract

Machine learning (ML) has achieved remarkable success in climate and marine science. Given that greenhouse warming fundamentally reshapes ocean conditions such as stratification, circulation patterns and eddy activity, evaluating the climate adaptability of the ML models is crucial. While physical constraints have been shown to enhance the performance of ML models, kinetic energy (KE) cascade has not been used as a constraint despite its importance in regulating multi-scale ocean motions. Here we present two sea surface height (SSH) prediction models (with and without KE cascade constraint) and quantify their climate adaptability in the Kuroshio Extension. Both models exhibit only slight performance degradation under greenhouse warming conditions. Incorporating the KE cascade as a physical constraint significantly improves the model performance, reducing eddy kinetic energy errors by 19.5% in the present climate and 19.6% under greenhouse warming conditions. Additional validations using satellite observations and in the Gulf Stream region further confirm the robustness of the proposed models. Compared with the traditional KE spectrum constraint, the model with the KE cascade constraint maintains improvements in the spectrum of KE while yielding more than 14% improvements in the cross-scale transfer of KE. This work presents the first application of the KE cascade as a physical constraint for ML-based ocean state prediction and demonstrates its robust adaptability across climates, offering guidance for the further development of global ML models for both present and future conditions.

Plain Language Summary

Predicting sea surface height (SSH) is crucial for understanding ocean behavior, especially in the context of global warming and rising sea levels. In this study, we developed two machine learning (ML) models for predicting SSH in both current and future climate conditions. By incorporating a novel physical constraint, the kinetic energy (KE) cascade, into the models, we significantly enhanced prediction accuracy. The KE cascade describes how energy transfers across different spatial scales in the ocean, and integrating this concept improves the models' ability to capture the complex behavior of ocean currents and eddies. Our results show that the models remain robust under greenhouse warming, with the KE cascade constraint further improving their predictive performance. We further tested the models using satellite altimetry data and in the Gulf Stream, and compared the cascade-constrained model with the spectrum-constrained model. This work introduces a new ML approach for ocean state prediction and demonstrates how ML can complement traditional climate models in both state forecasts and dynamical analysis.

1. Introduction

Machine learning (ML) has demonstrated competitive prediction accuracy and computational efficiency in Earth system modeling (Dramsch, 2020; Eyring et al., 2024; Karpatne et al., 2018; Reichstein et al., 2019; Rolnick et al., 2022). For example, models like Pangu-Weather and FuXi have achieved remarkable prediction skill and efficiency in weather forecasting (Bi et al., 2023; L. Chen et al., 2023). However, it is essential to assess whether these ML models retain reliable prediction skill under greenhouse warming conditions (Beucler et al., 2024), a capability hereafter referred to as *climate adaptability*. Furthermore, incorporating physical constraint has been shown to enhance ML model performance (Kashinath et al., 2021; Zhu et al., 2022); however, the kinetic energy (KE) cascade has not yet been explored as a constraint. In this study, using sea surface height (SSH) prediction as

Journal of Advances in Modeling Earth Systems 10.1029/2025MS005717

Visualization: Tianmu Zheng, Xin Su, Julian Mak, Gang Huang, Bingzheng Yan
Writing – original draft: Tianmu Zheng
Writing – review & editing: Tianmu Zheng, Ru Chen, Xin Su, Julian Mak, Gang Huang, Bingzheng Yan

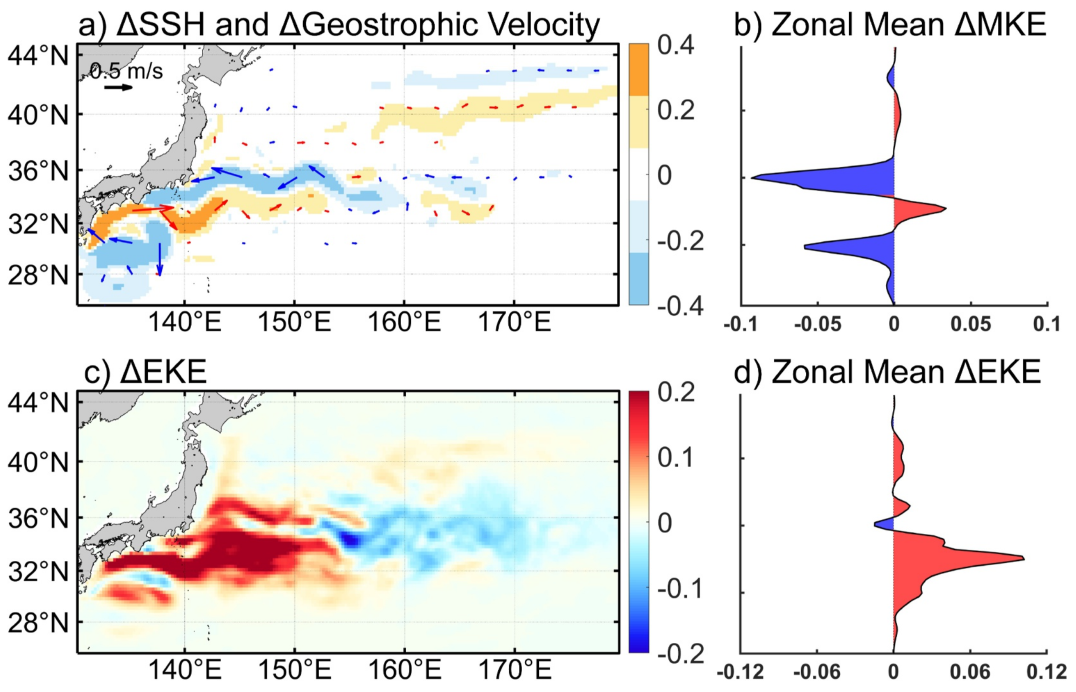

Figure 1. Oceanic responses to greenhouse warming. Δ represents the difference between the late 21st century (2092–2099) and present climate (2011–2019) from the AWI-CM-1-1-MR model under the SSP3-7.0 scenario. (a) Sea surface height changes (m, shading) and geostrophic velocity changes (arrows). (b–d) Changes in KE: (b) zonal mean kinetic energy, (c) eddy kinetic energy (EKE) spatial pattern, (d) zonal mean EKE. Percentages are computed as relative changes: (future–present)/present × 100%.

a testbed, we present a dual-attention ConvLSTM model, quantify the degree of its climate adaptability, and also assess the potential of incorporating the KE cascade to improve the prediction skill.

Climate adaptability is critical for ML-based ocean prediction, as greenhouse warming fundamentally alters ocean dynamics, including stratification, circulation patterns and eddy activity (Fox-Kemper, 2021; Li et al., 2020; Sallée et al., 2021; S. Wang et al., 2024). These changes affect both large-scale circulation and mesoscale eddies (Beech et al., 2022; Martínez-Moreno et al., 2021; Peng et al., 2022). Under greenhouse warming simulated by AWI-CM-1-1-MR under the SSP3-7.0 scenario (Beech et al., 2022; Eyring et al., 2016; O'Neill et al., 2016; Semmler et al., 2020), the area-averaged geostrophic velocity over the Kuroshio Extension region (130°–180°E, 25°–45°N) decreases by 11.5%, the mean kinetic energy drops by 46.0%, whereas eddy kinetic energy (EKE) rises by 33.1% between the present climate (2011–2019) and the late 21st century (2092–2099) (Figure 1). However, existing ocean ML models are typically trained and evaluated on temporally continuous data sets within the same climate state (Bi et al., 2023; L. Chen et al., 2023; Y. Chen et al., 2024; Ham et al., 2019; Ling et al., 2022; Rasp et al., 2018; Zhang et al., 2023). For example, Xihe was trained on GLORYS12 from 1993 to 2019 then evaluated over 2019–2020 (X. Wang et al., 2024). Although the community has started to explore the climate adaptability of ML models for atmospheric studies (Hernanz et al., 2022; Kochkov et al., 2024; O'Gorman & Dwyer, 2018) and oceanic parameterization problems (Guillaumin & Zanna, 2021; Gultekin et al., 2024; Perezhogin et al., 2025), this issue has received little attention in ocean prediction research. Therefore, our first aim is to quantify the performance degradation of ML-based ocean prediction models trained under the current climate when applied to greenhouse warming scenarios.

Our second goal is to investigate whether incorporating KE cascade as a physical constraint can enhance the ML model performance. Inspired by the demonstrated benefits of physical constraints on ML model performance (Kashinath et al., 2021; Yuval & O'Gorman, 2020; Zhu et al., 2022), we hypothesize that incorporating appropriate physical constraints could enhance ML-based ocean prediction models performance across different climate states.

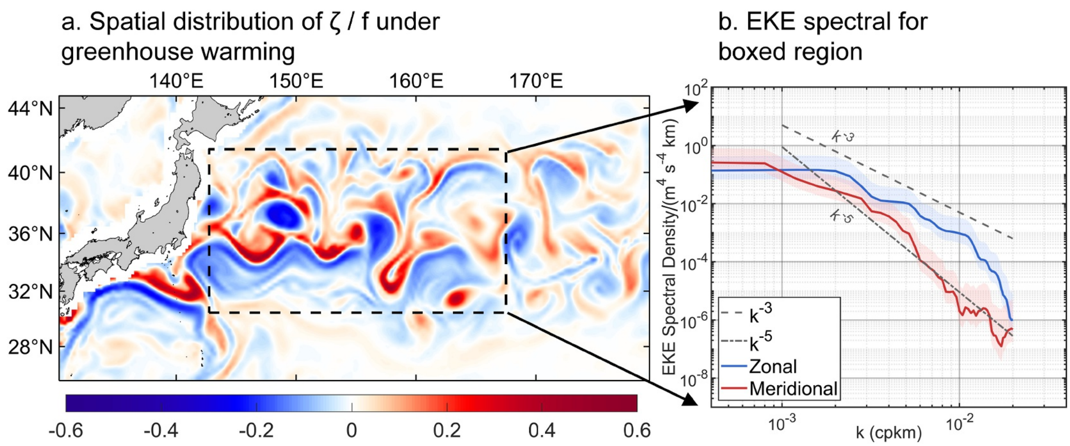

Figure 2. A snapshot of multi-scale ocean motions and energy distribution at the Kuroshio Extension on 4 February 2093. (a) Spatial distribution of normalized relative vorticity (ζ/f) under greenhouse warming. (b) Wavenumber spectra of eddy kinetic energy (EKE) for the boxed region (142.75°–167.5°E, 30.5°–41.5°N), EKE is computed from geostrophic velocity anomalies (see Table 2). Blue and red lines represent zonal and meridional spectra, with 95% confidence intervals (shaded) estimated based on the chi-square distribution (Percival & Walden, 1993). Gray dashed lines show k^{-3} and k^{-5} reference slopes.

Here we propose to use geostrophic KE cascade as a physical constraint. The ocean contains multi-scale motions, such as ocean circulation, jets and eddies (Figure 2). KE is redistributed across a range of spatial scales and is transferred among them through nonlinear interactions—a process known as the KE cascade (Alexakis & Biferale, 2018; Kolmogorov, 1991; Vallis, 2017). KE cascade plays a vital role in modulating multi-scale energy transfers and contributing to climate variability (Sérazin et al., 2018; Storer et al., 2023). However, to the best of our knowledge, it has not been used as a physical constraint of ML models.

To summarize, our goals are twofold. First, we evaluate the extent to which the ML models maintain adaptability across present and future climate conditions. Second, we investigate whether KE cascade constraint can enhance model performance. We choose SSH prediction at the Kuroshio Extension as the testbed for both goals. SSH is a key oceanic variable containing multi-scale dynamical signals, including large-scale circulation and meso- to submesoscale eddies (Chelton et al., 2011; Stammer, 1997). Through the geostrophic relationship, SSH directly reflects the multi-scale distribution of oceanic KE, which is shaped by cross-scale energy transfers (Scott & Wang, 2005; Storer et al., 2023). The Kuroshio Extension is an ideal region for this study due to its pronounced SSH spatiotemporal variability and strong sensitivity to climate change (Su et al., 2018; Zhou & Cheng, 2022). Our results show that the ML models we proposed for SSH prediction remain robust under greenhouse warming, and that integrating KE cascade constraint further improves their prediction skill. Robustness test using alternative data sets and regions has also been conducted. We also compare the KE cascade constraint with the KE spectrum constraint to assess their respective impacts on both the KE flux and the KE spectrum.

2. Experimental Framework

2.1. Description of Data

The model output analyzed in this study is from the AWI-CM-1-1-MR coupled climate model (Semmler et al., 2019, 2020), which contributes to the Coupled Model Intercomparison Project Phase 6 (CMIP6). The ocean component of AWI-CM-1-1-MR is based on the Finite Element Sea Ice-Ocean Model (FESOM) version 1.4 (Danilov et al., 2004), configured with an unstructured triangular mesh that allows for variable horizontal resolution (Q. Wang et al., 2014). The mesh resolution ranges from approximately 8 km in dynamically active regions to about 80 km (Semmler et al., 2020), with enhanced resolution in areas of high eddy activity (Sein et al., 2017).

The daily SSH fields analyzed here are from the r1i1p1f1 ensemble member both historical runs (1850–2014) and future projections following the SSP3-7.0 scenario (2015–2100). The SSP3-7.0 scenario is a high-emission

Table 1
Experimental Setup in This Study

Experiment name	Setup	Model	Description
DAM-C	Train and validation: 1981–2010, Test: 2012–2019	DAM	DAM trained on current climate and tested on current climate
DAM-F	Train and validation: 1981–2010, Test: 2092–2099	DAM	DAM trained on current climate and tested on future climate
CCM-C	Train and validation: 1981–2010, Test: 2012–2019	CCM	CCM trained on current climate and tested on current climate
CCM-F	Train and validation: 1981–2010, Test: 2092–2099	CCM	CCM trained on current climate and tested on future climate

pathway under which atmospheric CO_2 concentrations are projected to approximately double by the end of the century (Eyring et al., 2016; Masson-Delmotte et al., 2021; O'Neill et al., 2016).

The performance of AWI-CM-1-1-MR in simulating ocean dynamics and mesoscale eddy activity has been systematically evaluated (Beech et al., 2022). Comparison with satellite altimetry data (AVISO) shows that AWI-CM-1-1-MR captures approximately 82% of the observed EKE at the Kuroshio Extension, and the model successfully simulates the anticipated intensification of EKE in the Kuroshio Extension under greenhouse warming (Beech et al., 2022). Given its capability to capture cross-scale energy transfers from large scales to mesoscale eddies, the AWI-CM-1-1-MR data set is particularly suitable for the development and evaluation of our cascade-constrained ML algorithm.

We also use the AVISO satellite altimetry data set for independent validation. This gridded product, with a spatial resolution of $0.125° \times 0.125°$ and a daily temporal resolution, is derived from multiple satellite altimeters through unified preprocessing, inter-calibration, and optimal interpolation (Ducet et al., 2000; Taburet et al., 2019). For validation, we use the AVISO SSH data over the period 2012–2019 in the Kuroshio Extension region (130°–180°E, 25°–45°N). The data are interpolated onto a $0.25° \times 0.25°$ grid to match the resolution used in model training. Diagnostic variables are following the same equation in Table 2. The Dual-Attention ConvLSTM (DAM, Section 3.1) and Cascade-Constrained-ConvLSTM (CCM, Section 3.2) models are directly applied to the AVISO data without retraining to assess model robustness against observational data.

2.2. Experiment Setup

We designed four experiments (Figure 3a and Table 1) to investigate climate adaptability and the advantage of using the KE cascade constraint. In all experiments, the models take 21 days of SSH as input to directly predict 7 days of SSH output. Through these experiments, we can evaluate the climate adaptability of the ML models we present and assess whether incorporating the KE cascade constraint helps improve prediction accuracy and retain climate adaptability.

DAM-C versus DAM-F comparison evaluates the climate adaptability of the basic ML model by quantifying performance degradation when a model trained on current climate is applied to greenhouse warming conditions. The comparison of DAM-C versus CCM-C and DAM-F versus CCM-F demonstrates the advantage of incorporating KE cascade constraint in reducing prediction errors.

The 1981–2010 data set is partitioned into training and validation subsets in a ratio of 8:2, comprising a total of 10930 samples. Specifically, 8744 samples (1981–2004) are allocated for training and 2186 samples (2005–2010) for validation. To further verify model robustness, a 5-fold cross-validation was also conducted, with results shown in Figure S1 in Supporting Information S1.

2.3. Diagnostic Variables and Evaluation Metrics

This study evaluates the prediction skill of the ML models not only for SSH itself, but also for SSH-derived physical and higher-order statistical variables. SSH itself is evaluated using the normalized root-mean-square error (NRMSE). To further assess whether the predicted SSH fields preserve physical and statistical characteristics, we also diagnose SSH-derived variables, including information entropy, geostrophic EKE, skewness, and kurtosis—capture aspects like extreme SSH events, the intensity of geostrophic eddy activity, and SSH uncertainty (Table 2). In addition, the geostrophic KE spectral flux is employed as a complementary dynamical diagnostic, with its calculation formula shown in Figure 4. Unlike prior studies that focused on the direct

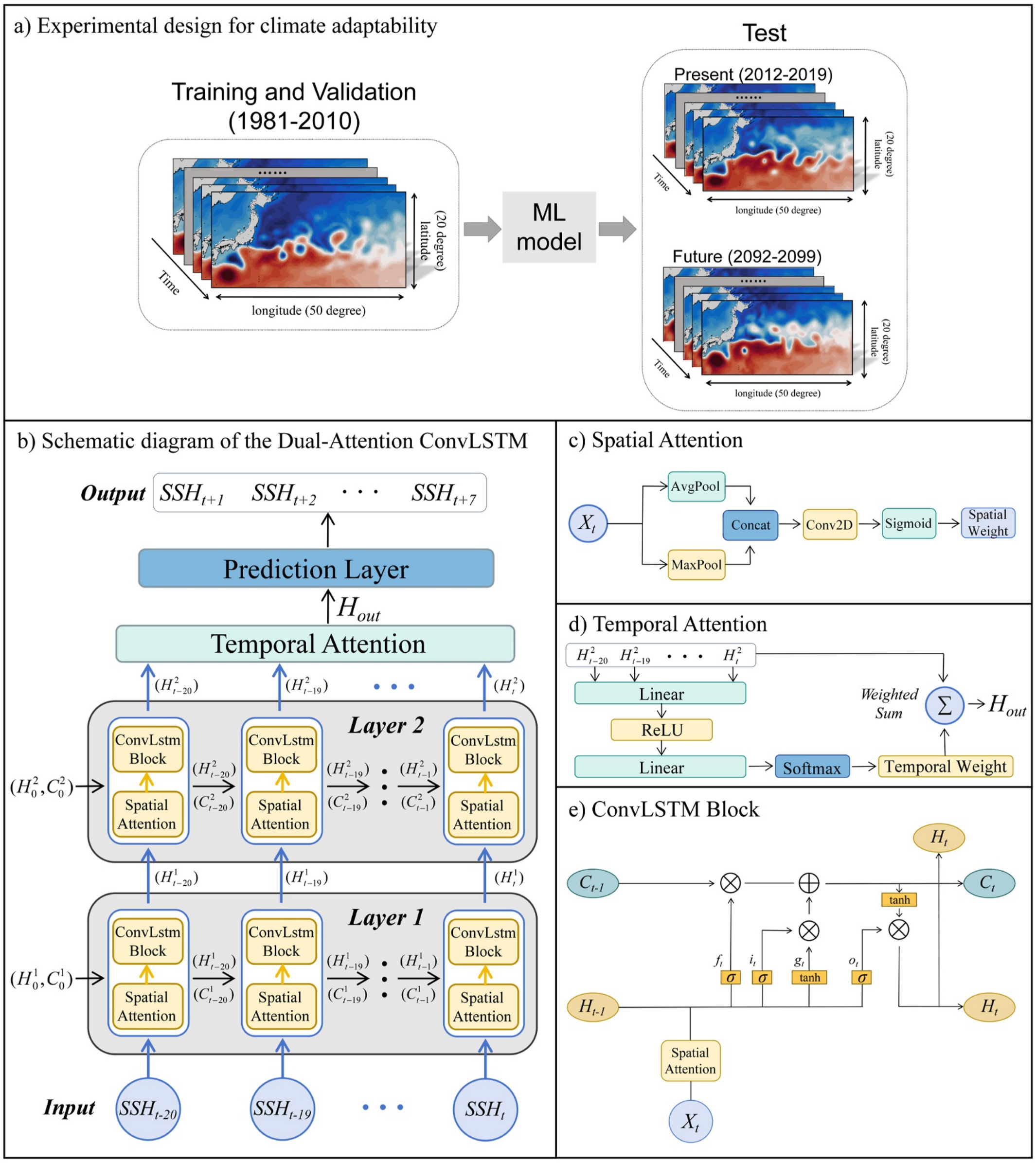

Figure 3. Climate adaptability evaluation framework and DAM model architecture. (a) Experimental design for evaluating climate adaptability. The model is trained on the 1981–2010 data and tested on both present (2012–2019) and greenhouse warming (2092–2099) climate conditions. (b) Schematic diagram of the DAM model, which consists of two ConvLSTM layers. H denotes the hidden state and C denotes the cell state, with superscripts indicating the layer index. The prediction layer uses a 3×3 convolutional layer to generate the 7-day sea surface height (SSH) forecast. (c) Spatial attention module. Average pooling and max pooling are applied in parallel across the feature maps. X denotes the input feature, including SSH and hidden-state features. The two pooled maps are then concatenated and passed through a convolutional layer and sigmoid activation to generate the spatial attention map. (d) Temporal attention module. The hidden states from all time steps are passed through two Linear layers with ReLU activation, followed by Softmax to obtain temporal weights. The weighted sum aggregates these hidden states into a single output H_{out}. (e) Structure of the ConvLSTM Block. The diagram style follows B. Shi et al. (2026).

prediction accuracy of SSH itself (Brettin et al., 2025; Sha et al., 2025; Shao et al., 2021), our expanded suite of variables provides a more complete assessment of ML model performance.

NRMSE is used to evaluate the SSH itself (Equation 1), while the physical and statistical variables are assessed using the Mean Absolute Error (MAE) (Equation 2) and correlation R (Equation 3). The NRMSE, MAE and correlation R are defined as follows:

Table 2
Diagnostic Variables for Sea Surface Height Evaluation

Variable	Physical Means
Information entropy $(i,j) = -\sum_{k=1}^{K} p_k(i,j)\log_2 p_k(i,j)$	Quantifies SSH uncertainty by measuring probability distribution uniformity; higher values indicate greater uncertainty and variability (Shannon, 1948)
$\mathrm{EKE}_g\ (i,j) = \frac{1}{2}\left[\left(u_g'(i,j)\right)^2 + \left(v_g'(i,j)\right)^2\right]$	Quantifies the intensity of geostrophic eddy motions (Chelton et al., 2011; Stammer, 1997)
Skewness $(i,j) = \frac{\overline{(\mathrm{SSH}'(i,j))^3}}{\sigma^3(i,j)}$	Asymmetry of SSH distribution: positive values indicate right-tail, and vice versa (Sura & Gille, 2010; Von Storch & Zwiers, 2002)
Kurtosis $(i,j) = \frac{\overline{(\mathrm{SSH}'(i,j))^4}}{\sigma^4(i,j)} - 3$	Frequency of extreme SSH events: high values indicate more frequent extremes, and vice versa (Sura & Gille, 2010; Von Storch & Zwiers, 2002)

Definitions

- (i,j): Represent spatial coordinate indices
- N: Total number of data points
- K: Total number of bins ($K = 10$, equally-spaced bins spanning SSH range $[-0.3, 1.7]$ m)
- k: Bin index ($k = 1, 2, \ldots, K$)
- P_k: Probability of SSH values in the k-th bin
- $\overline{(\cdot)}$: Time-averaged
- SSH': SSH deviation from the time average
- σ: Standard deviation of SSH'
- u_g, v_g: Eastward and northward geostrophic velocities:

$u_g = -\frac{g}{f}\frac{\partial \mathrm{SSH}}{\partial y}$

$v_g = \frac{g}{f}\frac{\partial \mathrm{SSH}}{\partial x}$

Where g is gravitational acceleration, f is the Coriolis parameter, x and y are the longitudinal and latitudinal coordinates

$$\mathrm{NRMSE_{SSH}} = \frac{\sqrt{\frac{1}{N}\sum_{(i,j)}\left(\mathrm{SSH^{pred}}(i,j) - \mathrm{SSH^{real}}(i,j)\right)^2}}{\max \mathrm{SSH^{real}}(i,j) - \min \mathrm{SSH^{real}}(i,j)} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{(i,j)}\left|X^{\mathrm{real}}(i,j) - X^{\mathrm{pred}}(i,j)\right| \tag{2}$$

$$R = \frac{\sum_{(i,j)}\left(X^{\mathrm{real}}(i,j) - \overline{X^{\mathrm{real}}}\right)\left(X^{\mathrm{pred}}(i,j) - \overline{X^{\mathrm{pred}}}\right)}{\sqrt{\sum_{(i,j)}\left(X^{\mathrm{real}}(i,j) - \overline{X^{\mathrm{real}}}\right)^2\sum_{(i,j)}\left(X^{\mathrm{pred}}(i,j) - \overline{X^{\mathrm{pred}}}\right)^2}} \tag{3}$$

where X represents the diagnostic variables listed in Table 2, (i,j) represent spatial coordinate indices, and N denotes the total number of grid points, $*^{\mathrm{pred}}$ represents the predicted data from the model and $*^{\mathrm{real}}$ represents the true data.

Uncertainties are estimated using a bootstrapping technique (R. Chen et al., 2014; Chernick, 2011; Guan et al., 2022). We randomly select a subset of size N_{subset} to estimate the metrics and repeat this process N_{boot} times, yielding a 95% confidence interval half-width of $2\sigma_{\mathrm{boot}}/\sqrt{N_{\mathrm{boot}}}$. σ_{boot} is the standard deviation of the N_{boot} subsample estimates. The value $N_{\mathrm{subset}} \cdot N_{\mathrm{boot}}$ is chosen to be close to, but not greater than, the total number of available data points, so that each subsampled data set is approximately independent.

3. Machine Learning Method

To evaluate the climate adaptability and the advantage of KE cascade constraints, a suitable ML architecture must first be selected for SSH prediction. Our purpose is neither to identify the optimal architecture among existing ML models, nor develop an entirely new architecture. Instead, we aim to use a reasonable and competitive ML-based prediction model as a testbed for examining climate adaptability and the effect of the KE cascade constraint. Among various ML approaches for geoscientific variable prediction, ConvLSTM has proven effective in various spatiotemporal prediction tasks, including precipitation nowcasting, ocean wave forecasting and wind speed prediction (X. Shi et al., 2015; Song et al., 2022; L. Zheng et al., 2023), and attention mechanisms have demonstrated remarkable capability in capturing important features and dependencies in sequential data

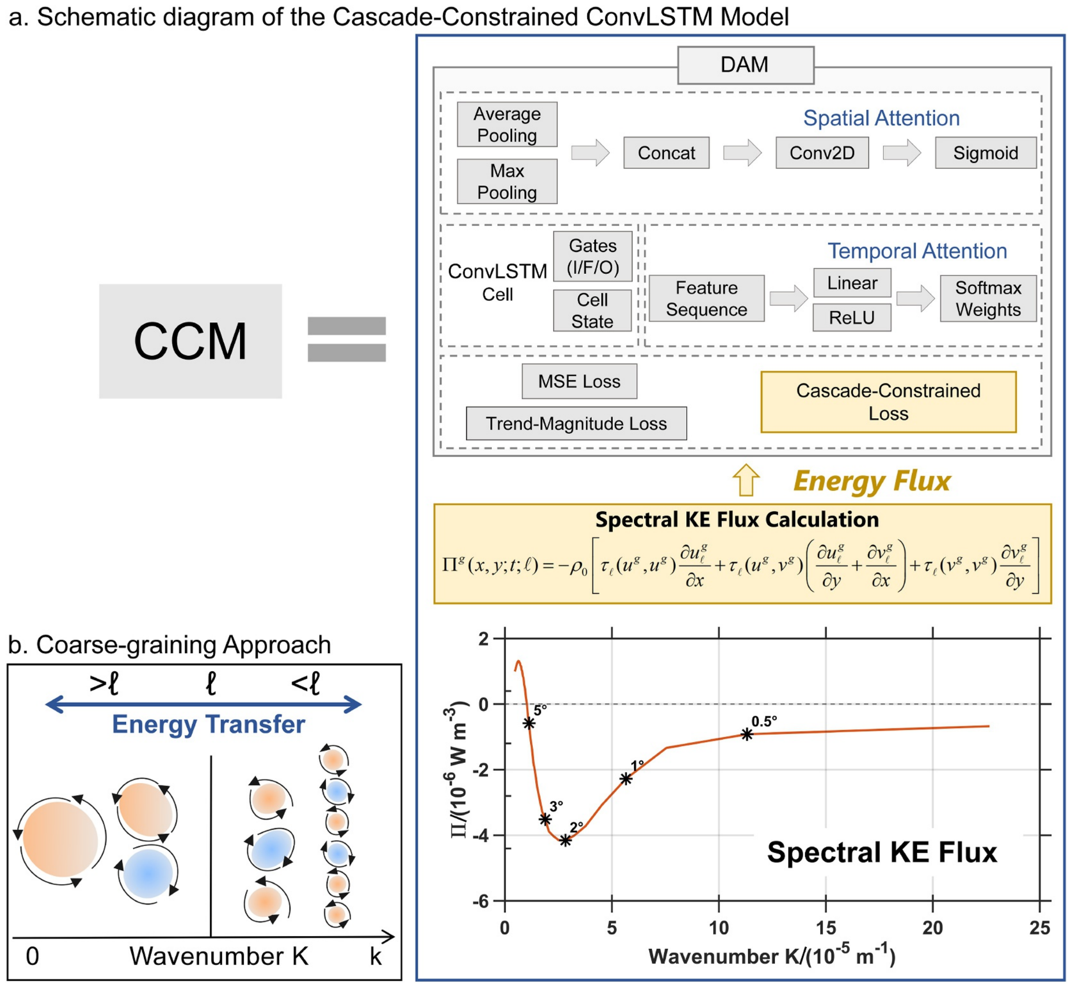

Figure 4. Architecture of the Cascade-Constrained-ConvLSTM (CCM) model. (a) Schematic diagram showing the integration of DAM components (spatial attention, temporal attention, and trend-aware loss) with KE cascade constraint. The spectral KE flux $\Pi(k)$ displays the inverse cascade with our selected filter scales (5°, 3°, 2°, 1°, 0.5°) marked by stars, which are used for computing the physics-informed loss. (b) The spectral KE flux denotes the KE transfer rate between larger-scale (larger than the separation scale ℓ) and smaller-scale motions. Here we apply the coarse-graining approach to estimate the spectral KE flux (Aluie et al., 2018).

(Bahdanau et al., 2014; Lin et al., 2020; Vaswani et al., 2017; Woo et al., 2018). Recently, Shi et al. demonstrated that integrating attention mechanisms into ConvLSTM improves the prediction of sea surface temperature and sea level anomaly, with robust performance in both the Bohai Sea and the South China Sea (B. Shi et al., 2024, 2026). Beyond these variables, this approach has also been successfully applied to ocean remote sensing reflectance prediction (Zhou et al., 2024).

These successes suggest that combining ConvLSTM with attention mechanisms provides a well-suited architecture for our research goals. Building upon this foundation, we further extend this approach by developing two models: (a) the Dual-Attention-ConvLSTM Model (DAM, Section 3.1), which incorporates spatiotemporal attention and Trend-Aware Loss function, and (b) the Cascade-Constrained-ConvLSTM Model (CCM, Section 3.2), which further integrates the KE cascade as a physical constraint. The experimental setup is described in Section 2.2.

3.1. Architecture of DAM Model

DAM integrates both spatial and temporal attention mechanisms within the ConvLSTM framework to enhance SSH prediction (Figure 3b; comparisons with other methods are provided in Table S1 in Supporting

Information S1; see Text S1 in Supporting Information S1 for the description of other SSH prediction methods). The spatial attention, inspired by Convolutional Block Attention Module (Woo et al., 2018), is embedded within each ConvLSTM cell (Figure 3b). It applies average and max pooling along the channel dimension, followed by a convolutional layer and sigmoid activation to generate a spatial attention map. This map is pixel-wise multiplied with the input, enabling the model to focus on regions with strong SSH variability at each time step (Figure 3c). The temporal attention module operates after each ConvLSTM layer processes the entire input sequence. It computes learned weights over the hidden states across all time steps through a two-layer neural network with softmax normalization (Qin et al., 2017), producing a weighted combination that highlights the most informative time steps for prediction (Figure 3d).

The DAM model uses two ConvLSTM layers with hidden dimension 64 and 3×3 kernels, totaling 451,534 trainable parameters. It is trained using the AdamW optimizer with a learning rate of 2.5×10^{-4} and weight decay of 0.01, OneCycleLR scheduling, batch size 8, for up to 100 epochs. Early stopping is applied with a patience of 10 epochs based on the validation loss.

Besides the dual-attention mechanisms, DAM incorporates a loss function that captures SSH evolution patterns beyond traditional numerical accuracy metrics (Figure 3b). $\mathcal{L}_{\mathrm{DAM}}$ is formulated as:

$$\mathcal{L}_{\mathrm{DAM}} = \alpha\mathcal{L}_{\mathrm{MSE}} + \beta\mathcal{L}_{\mathrm{trend-magnitude}} \tag{4}$$

where:

$$\mathcal{L}_{\mathrm{MSE}} = \mathrm{MSE}\left(\mathrm{SSH}^{\mathrm{pred}}, \mathrm{SSH}^{\mathrm{real}}\right) \tag{5}$$

$$\mathcal{L}_{\mathrm{trend-magnitude}} = \mathrm{MSE}\left(\mathrm{SSH}^{\mathrm{pred}}_{\mathrm{diff}}, \mathrm{SSH}^{\mathrm{real}}_{\mathrm{diff}}\right) \times (1 + \mathrm{weight}) \tag{6}$$

MSE represent mean squared error, $\mathrm{SSH}^{*}_{\mathrm{diff}} = \mathrm{SSH}^{*}_{t+1} - \mathrm{SSH}^{*}_{t}$ represent temporal differences between consecutive time steps. $*^{\mathrm{pred}}$ represents the predicted data from the model and $*^{\mathrm{real}}$ represents the true data. The weight term assigns higher importance to large variations:

$$\mathrm{weight} = c \times \mathbb{I}\left(\left|\mathrm{SSH}^{\mathrm{real}}_{\mathrm{diff}}\right| > \tau\right) \tag{7}$$

where $\mathbb{I}$ is the indicator function that equals 1 when the condition is true and 0 otherwise, τ is the threshold above which temporal changes receive additional emphasis during training, and c is a weight multiplier controlling the additional emphasis on large-change regions. The hyperparameters α, β, τ, and c are optimized using Bayesian hyperparameter optimization (Shahriari et al., 2015; Snoek et al., 2012), with $\beta = 1 - \alpha$. The search spaces are $\alpha \in [0.2, 0.8]$, $\tau \in [0.01, 0.5]$, and $c \in [0.5, 5.0]$, with search intervals of 0.05, 0.01, and 0.25, respectively. The optimal values obtained are $\alpha = 0.4$, $\beta = 0.6$, $\tau = 0.04$, and $c = 2.75$.

3.2. Architecture of CCM Model

While physics-informed approaches have incorporated various oceanic constraints (Meng et al., 2021, 2023; Zhu et al., 2022), the KE cascade representing cross-scale energy transfers has not been utilized. Here we incorporate the spectral KE flux as a constraint into our model.

Building upon DAM, we develop CCM that incorporates KE cascade into the neural network architecture (Figure 4). The cascade-constraint loss measures the spectral KE flux error:

$$\mathcal{L}_{\mathrm{cascade}} = \mathrm{MSE}\left(\Pi^{\mathrm{pred}}, \Pi^{\mathrm{true}}\right) \tag{8}$$

where the spectral KE fluxes (Π) using the coarse-graining approach (Aluie et al., 2018):

$$\Pi^{g}(x, y; t; \ell) = -\rho_0\left[\tau_\ell(u^g, u^g)\frac{\partial u^g_\ell}{\partial x} + \tau_\ell(u^g, v^g)\left(\frac{\partial u^g_\ell}{\partial y} + \frac{\partial v^g_\ell}{\partial x}\right) + \tau_\ell(v^g, v^g)\frac{\partial v^g_\ell}{\partial y}\right] \tag{9}$$

where $\rho_0 = 1027.4\ \mathrm{kg/m^3}$, $\tau_\ell(u,u)$ represents the Reynolds stress tensor, u^g and v^g represent the zonal and meridional geostrophic velocity, and u^g_ℓ, v^g_ℓ represent the geostrophic velocity components at scales larger than ℓ. The Reynolds stress tensor is defined as:

$$\tau_\ell(\mathbf{u},\mathbf{u}) = (\mathbf{u}\mathbf{u})_\ell - \mathbf{u}_\ell\mathbf{u}_\ell \tag{10}$$

where this general form is applied to the geostrophic velocity components in Equation 9. This tensor quantifies the force exerted by small-scale motions (scales $< \ell$) on large-scale motions (scales $> \ell$). We evaluate Π^g at five representative scales (5°, 3°, 2°, 1°, 0.5°) by coarse-graining approach (Aluie et al., 2018). The first two loss components (MSE, trend-magnitude loss) are identical to those defined in the DAM model. The total CCM loss becomes:

$$\mathcal{L}_{\mathrm{CCM}} = \alpha\mathcal{L}_{\mathrm{MSE}} + \beta\mathcal{L}_{\mathrm{trend-magnitude}} + \delta\mathcal{L}_{\mathrm{cascade}} \tag{11}$$

The hyperparameters α, β, δ, τ, and c are jointly optimized using Bayesian hyperparameter optimization (Shahriari et al., 2015; Snoek et al., 2012), with $\beta = 1 - \alpha - \delta$. The search spaces are $\alpha \in [0.2, 0.8]$, $\delta \in [0.05, 0.5]$, $\tau \in [0.01, 0.5]$, and $c \in [0.5, 5.0]$, with search intervals of 0.05, 0.05, 0.01, and 0.25, respectively. The optimal values obtained are $\alpha = 0.35$, $\beta = 0.2$, $\delta = 0.45$, $\tau = 0.35$, and $c = 2.25$. With these settings, the model not only minimizes pixel-wise prediction error but also constrains the cross-scale KE transfer during training. This addresses one of the limitations of data-driven ML models, which often produce physically inconsistent outputs and erroneous energy distribution across scales despite achieving high forecast skill (Beucler et al., 2021; Bonavita, 2024; Irrgang et al., 2021; Pasquini et al., 2026).

In addition, we develop a Spectrum-Constrained model (SCM) that shares the same architecture as CCM but replaces the cascade loss with a KE spectrum loss, allowing us to evaluate the effects of constraining energy distribution rather than energy transfer. The spectrum-constraint loss is defined as:

$$L_{\mathrm{spectrum}} = \mathrm{MSE}\left(\log_{10}\left(E^{\mathrm{pred}}\right), \log_{10}\left(E^{\mathrm{true}}\right)\right) \tag{12}$$

where E^{pred} and E^{true} denote the predicted and true KE spectrum, respectively. The logarithmic transform places the KE spectrum at a comparable order of magnitude across wavenumbers. E^{pred} and E^{true} are computed from geostrophic velocities as:

$$E(k) = c_k\frac{1}{2}\left(\frac{|\hat{u}^g(k)|^2}{N_x} + \frac{|\hat{v}^g(k)|^2}{N_x}\right) \tag{13}$$

where k is the zonal wavenumber, and $\hat{u}^g$ and $\hat{v}^g$ are the Fourier coefficients of the zonal and meridional geostrophic velocities, respectively. N_x is the number of zonal grid points, and c_k is the one-sided spectrum correction factor, with $c_k = 1$ for the zero and Nyquist components and $c_k = 2$ otherwise. Following the same Bayesian hyperparameter optimization procedure as CCM, the optimal hyperparameters are $\alpha = 0.35$, $\beta = 0.35$, $\delta = 0.3$, $\tau = 0.22$, and $c = 1.75$.

All models (DAM, CCM, and SCM) are optimized within their respective hyperparameter search spaces using the same Bayesian optimization framework. For each model, 30 hyperparameter configurations are explored, and each configuration is evaluated by training the model for 30 epochs. This helps ensure that performance comparisons among models primarily reflect the impact of the physical constraints applied.

4. Result: Climate Adaptability and Advantage of Cascade Constraints

4.1. Climate Adaptability of DAM Model

The Kuroshio Extension undergoes intense changes under greenhouse warming, making it an idealized region for climate adaptability evaluation. In the context of greenhouse warming, the EKE band intensifies and shifts westward, getting broader and less coherent. Similarly, the high information entropy region expands both meridionally and zonally, reflecting the broadening of eddy-active zones into previously quiescent regions

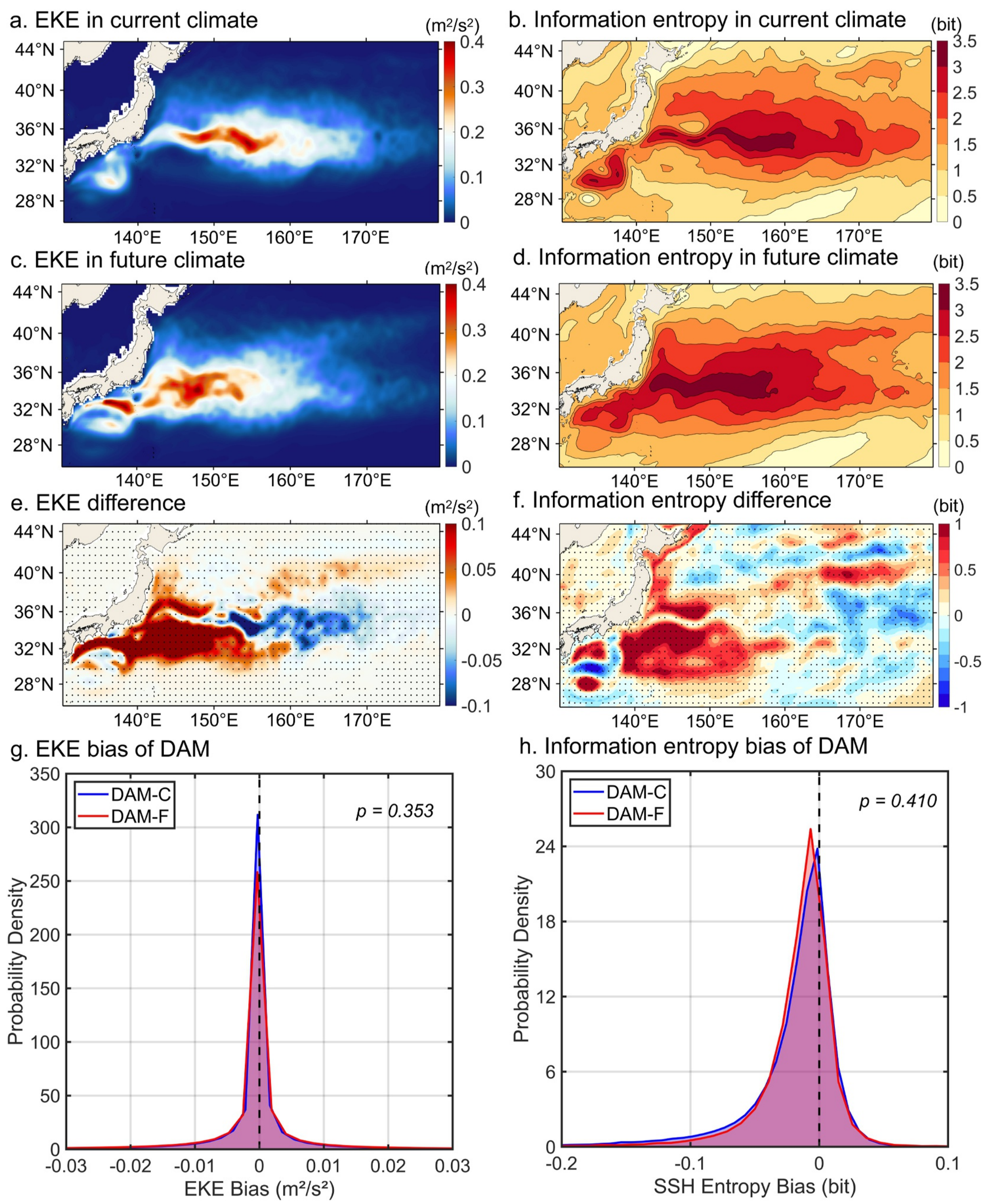

Figure 5. Spatial patterns and biases of the time-mean eddy kinetic energy (EKE) and information entropy for the DAM model. (a) EKE (m^2/s^2) in current climate (2012–2019). (b) Information entropy of sea surface height (SSH) (bit) in current climate. (c) EKE in future climate (2092–2099) under greenhouse warming. (d) Information entropy of SSH in future climate. (e) Difference in EKE between future and current climate, calculated as future minus current. This panel contains the same information as Figure 1, but additionally indicates statistical significance. Stippling denotes statistically significant differences at the 95% confidence level. (f) Difference in information entropy between future and current climate. (g) Probability Density Functions (PDFs) of the EKE bias for current climate (blue) and future climate (red). (h) PDFs of the information entropy bias for current and future climate. The dashed vertical lines in panels (g and h) indicate the line of zero bias. The p-values from two-sample Kolmogorov–Smirnov (KS) tests indicate no statistically significant difference between the DAM-C and DAM-F bias distributions (EKE: $p = 0.353$; information entropy: $p = 0.410$), confirming no significant change in the bias distributions across climate states (Lanzante, 2021; Massey, 1951).

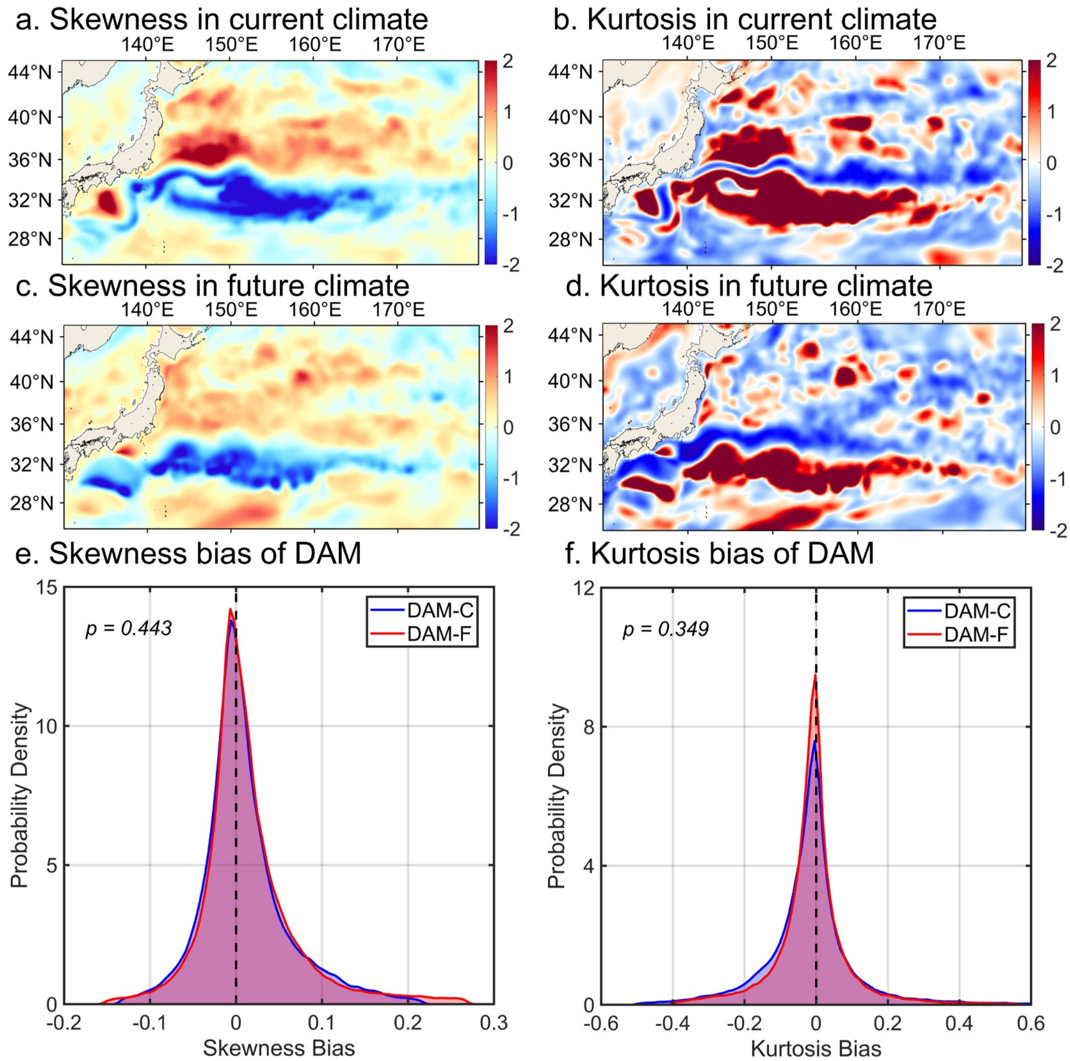

Figure 6. Spatial patterns and biases of the time-mean skewness and kurtosis for the DAM model. (a) Skewness and (b) kurtosis in current climate (2012–2019). (c) Skewness and (d) kurtosis in future climate (2092–2099) under greenhouse warming. (e) Probability Density Functions (PDFs) of the skewness bias for current (blue) and future climate (red). (f) PDFs of the kurtosis bias for current and future climate conditions. The dashed vertical lines in panels (e and f) indicate the line of zero bias. KS tests following the same procedure as in Figure 5 show no statistically significant difference between the DAM-C and DAM-F bias distributions (skewness: $p = 0.443$; kurtosis: $p = 0.349$).

(Figures 5a–5f). The change of skewness and kurtosis reflects shifts in the distribution asymmetry and extreme event statistics of SSH (Figures 6a–6d). The spatial patterns of these metrics under current climate are consistent with previous studies (Beech et al., 2022; Sura & Gille, 2010; Von Storch & Zwiers, 2002), the EKE response to greenhouse warming also aligns with trends in literature (Barceló-Llull et al., 2025; Beech et al., 2022).

We evaluated the climate adaptability of the DAM model by comparing current climate case (DAM-C) and future climate case (DAM-F) experiments (Table 1). Across all evaluated variables, the Probability Density Functions of biases show no statistically significant changes between DAM-C and DAM-F (Figures 5g, 5h, 6e, and 6f). For EKE, biases are within ±0.01 m^2/s^2 for 85.90% in DAM-C and 83.93% in DAM-F. Information entropy biases fall within ±0.1 bit for 95.33% (DAM-C) and 96.99% (DAM-F). The corresponding proportions for skewness within ±0.1 are 90.75% in DAM-C and 89.97% in DAM-F, while those for kurtosis within ±0.2 are 87.50% in DAM-C and 90.42% in DAM-F. These results indicate that the DAM model maintains robust performance across different climate states. Performance comparisons with other baseline methods under different climate states are provided in Table S1 in Supporting Information S1. These comparisons are used to place our models in the context of

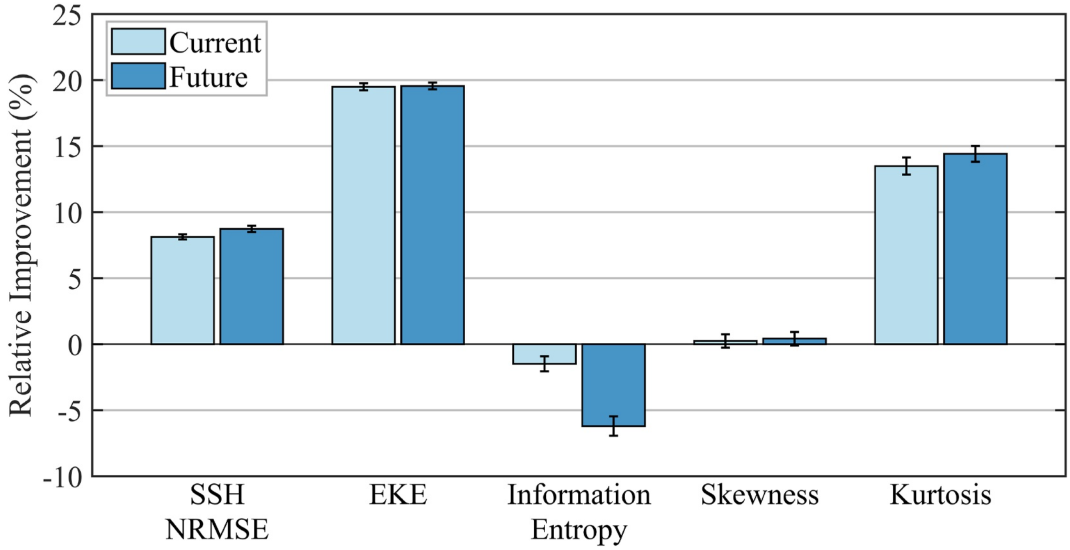

Figure 7. Performance comparison between DAM and CCM models showing the time-mean relative improvement by incorporating KE cascade constraint. Bar charts display the relative improvement (%) of CCM over DAM for five evaluation metrics: NRMSE of sea surface height, eddy kinetic energy Mean Absolute Error (MAE), information entropy MAE, skewness MAE, and kurtosis MAE. Light blue bars represent current climate conditions (2012–2019), while dark blue bars represent future climate conditions (2092–2099). The relative improvement is calculated as: (DAM-C − CCM-C)/DAM-C × 100% for current climate and (DAM-F − CCM-F)/DAM-F × 100% for future climate. Error bars indicate 95% confidence intervals.

common SSH prediction methods, rather than to claim that the selected architecture outperforms all possible alternatives.

To quantitatively evaluate the climate adaptability of the DAM model, we compared the metrics (Table 2) described between the DAM-C and DAM-F experiments. The relative changes in the MAE for information entropy and kurtosis are negative, indicating improved prediction skill under greenhouse warming. Similarly, DAM-F (future-climate experiment) shows slightly better skill in representing the spatial structure of EKE. Other metrics or variables in Table 3 exhibit modest degradation under greenhouse warming, with SSH NRMSE

Table 3
Performance Comparison Between the DAM-C (Current Climate) and DAM-F (Future Climate) Experiments

	DAM-C	DAM-F	Relative change (%)
Normalized Root-Mean-Square Error (Equation 1)			
SSH (m)	0.0116 ± 0.0001	0.0123 ± 0.0001	**+6.0**
Mean Absolute Error (Equation 2)			
Information Entropy (bit)	0.0263 ± 0.0006	0.0234 ± 0.0005	**−11.0**
EKE (m^2/s^2)	0.0090 ± 0.0006	0.0107 ± 0.0008	**+18.6**
Skewness	0.0395 ± 0.0008	0.0423 ± 0.0011	**+7.1**
Kurtosis	0.1008 ± 0.0033	0.0830 ± 0.0030	**−17.7**
Spatial Correlation Between True and Predicted Values (Equation 3)			
Information Entropy	0.9983 ± 0.0001	0.9986 ± 0.0002	+0.03
EKE	0.9414 ± 0.0007	0.9471 ± 0.0010	**+0.61**
Skewness	0.9967 ± 0.0002	0.9941 ± 0.0003	**−0.26**
Kurtosis	0.9945 ± 0.0005	0.9902 ± 0.0010	**−0.43**

Note. Relative change = (DAM-F − DAM-C)/DAM-C × 100%. Bold values indicate statistically significant difference between DAM-C and DAM-F.

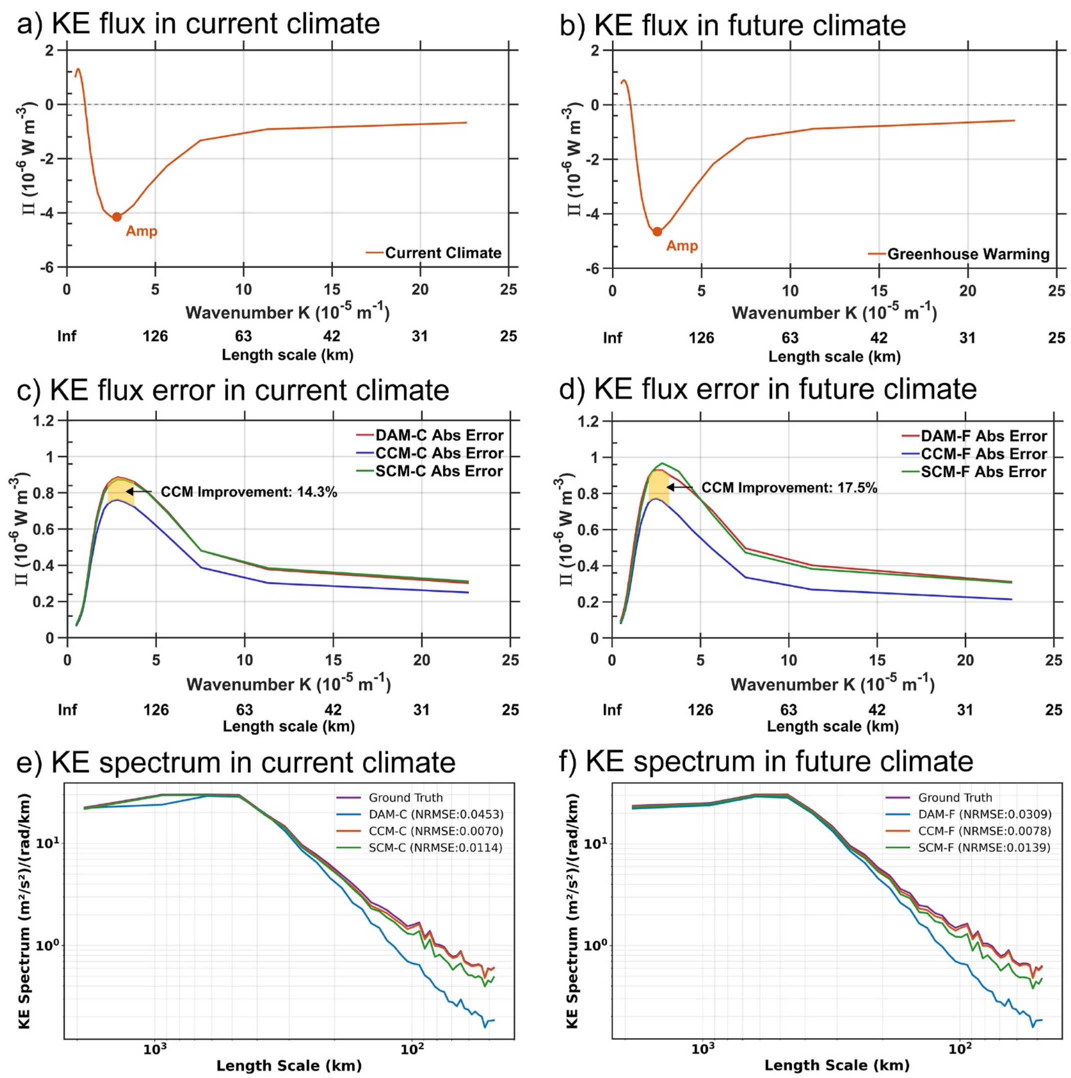

Figure 8. KE cascade spectra and model performance comparison. (a) Spectral KE flux Π(k) in the current climate, showing the characteristic amplitude minimum (Amp, red). (b) The same as (a), but for greenhouse warming scenario. (c) Mean Absolute Error between predicted and true spectral KE flux for the DAM-C (red), CCM-C (blue) and SCM-C (green) experiments. Yellow shading highlights the wavenumber range encompassing the five points with largest DAM-C absolute errors, where CCM shows 14.3% average improvement. (d) The same as (c), but for experiments in the future climate. (e) KE spectrum comparison in the current climate between the truth and the three experiments (DAM-C, CCM-C, SCM-C). (f) As in (e), but for the future-climate experiments (DAM-F, CCM-F, SCM-F).

increasing from 0.0116 to 0.0123 (+6.0%) and EKE exhibiting the largest relative degradation at 18.6%. However, the spatial pattern correlations remain consistently above 0.94 across all variables, indicating that the DAM model generalizes well to the greenhouse warming climate.

4.2. Advantage of Incorporating KE Cascade Constraint

To evaluate the impact of incorporating KE cascade constraint, we compared experiments with and without the constraint (see Table 1 for experiment descriptions). The key message is that adding this constraint consistently reduces the prediction errors across the evaluation metrics.

For both current and future climate scenarios, incorporating KE cascade constraint (CCM) notably improves prediction skill compared to experiments without these constraint (DAM). As shown in Figure 7, under the current climate, CCM-C exhibits 8.13% lower SSH NRMSE values than DAM-C (0.0107 vs. 0.0116). Under greenhouse warming, this advantage becomes even more pronounced, with CCM-F exhibiting 8.73% lower errors

Table 4
Spatial Correlations Between True and Predicted Metrics in Each Experiment

		DAM-C	CCM-C	DAM-F	CCM-F
Variables	Information Entropy	0.9983 ± 0.0001	0.9985 ± 0.0001	0.9986 ± 0.0002	0.9986 ± 0.0001
	EKE	0.9414 ± 0.0007	**0.9603 ± 0.0005**	0.9471 ± 0.0010	**0.9638 ± 0.0006**
	Skewness	0.9967 ± 0.0002	0.9969 ± 0.0002	0.9941 ± 0.0003	0.9943 ± 0.0003
	Kurtosis	0.9945 ± 0.0005	0.9952 ± 0.0004	0.9902 ± 0.0010	0.9921 ± 0.0010

Note. Bold values indicate statistically significant improvements of CCM over DAM.

than DAM-F (0.0112 vs. 0.0123). Incorporating this cascade constraint boosts the EKE prediction skill by 19.49% under the current climate and 19.55% under greenhouse warming. Kurtosis also shows substantial improvements with over 13% error reduction in both climate conditions.

The KE cascade depicts cross-scale transfer among multi-scale motions (Alexakis & Biferale, 2018; Kolmogorov, 1991; Vallis, 2017). The performance of this cascade constraint, especially in improving EKE, can be attributed to their ability to enforce physically consistent energy transfers, which greatly modulate mesoscale eddy energy (Geng et al., 2025). By constraining cross-scale KE fluxes during training, CCM also better captures the KE cascade characteristics.

The spatial correlations between true and predicted values further confirm the advantage of the cascade-constrained algorithm (CCM) in capturing EKE patterns (Table 4). All experiments show high spatial pattern correlations (>0.94). In particular, EKE, as an energy-related variable, exhibits statistically significant improvements, with correlations rising from 0.9414 to 0.9603 under the current climate and from 0.9471 to 0.9638 under greenhouse warming, highlighting the importance of preserving physically consistent energy transfers in ML ocean models.

4.3. KE Cascade Versus Spectrum Constraints

The spectral KE flux $\Pi(k)$ depicts the KE transfer rate across selected spatial scales (Figures 8a and 8b). The negative value of $\Pi(k)$ indicates inverse cascade, where energy flows from small to large scales. The spectral flux diagnosed here (Figures 8a and 8b) closely resembles the AVISO satellite observations (Geng et al., 2025; Scott & Wang, 2005; S. Wang et al., 2015). Comparison between Figures 8a and 8b suggests that the inverse cascade amplitude intensifies under greenhouse warming, consistent with findings in literature (Geng et al., 2025). These results suggest that the AWI-CM-1-1-MR model realistically captures cross-scale energy transfer at the Kuroshio Extension.

A useful way to think about the KE spectrum and $\Pi(k)$ is that the spectrum tells us how much energy resides at each scale, whereas the flux $\Pi(k)$ quantifies the direction and strength of KE transfer across scales. Matching the spectral distribution does not automatically guarantee that the cross-scale transfer is correct, and the reverse is also possible. This motivates an explicit comparison between KE cascade-based and spectrum-based constraints. We therefore consider three configurations: the unconstrained DAM, the Cascade-Constrained model (CCM), and a Spectrum-Constrained model (SCM) that replaces the cascade loss used in CCM with a KE spectrum loss while keeping the remaining architecture and loss components unchanged. For the flux $\Pi(k)$, CCM produces the largest reduction in absolute error near the wavenumber band where DAM exhibits its maximum flux errors, with 14.3% improvement in the current climate and 17.5% under greenhouse warming (Figures 8c and 8d). In contrast, SCM does not consistently reduce the KE-flux error relative to DAM, indicating that directly constraining the KE spectrum does not necessarily improve the representation of cross-scale energy transfer.

The benefits of including physical constraints in ML models are also evident in the predicted KE spectrum (Figures 8e and 8f). DAM underestimates KE at eddy scales in both climates, whereas CCM and SCM reproduce the reference spectrum more closely. Notably, CCM improves both KE spectrum and KE spectral flux, whereas SCM primarily targets the spectrum and yields a more limited improvement in the spectral flux (Figures 8c–8f). These results suggest that, compared to constraining the KE spectrum, constraining the cascade provides a more comprehensive improvement in energetic consistency than constraining the spectrum alone.

Table 5
Model Performance of DAM and CCM Across AVISO and Gulf Stream Validations

	Performance of DAM			Performance of CCM		
	DAM-C	DAM-C (AVISO)	DAM-C (Gulf)	CCM-C	CCM-C (AVISO)	CCM-C (Gulf)
Normalized Root-Mean-Square Error (Equation 1)						
SSH (m)	0.0116	0.0148	0.0109	**0.0107**	**0.0133**	**0.0102**
Mean Absolute Error (Equation 2)						
Information Entropy (bit)	0.0263	0.0367	0.0223	0.0267	**0.0242**	0.0236
EKE (m^2/s^2)	0.0090	0.0116	0.0054	**0.0072**	**0.0103**	**0.0044**
Skewness	0.0395	0.0588	0.0441	0.0394	**0.0430**	0.0433
Kurtosis	0.1008	0.1176	0.1539	**0.0872**	**0.0942**	0.1500

Note. Bold values indicate statistically significant improvements of CCM over DAM. DAM-C and CCM-C are defined in Table 1. DAM-C (AVISO) denotes validation on AVISO altimetry without retraining, that is, the DAM-C models trained on the AWI-CM-1-1-MR data set are directly evaluated on AVISO; CCM-C (AVISO) is defined in the same way. DAM-C (Gulf) denotes evaluation in the Gulf Stream region using models retrained under the DAM-C experiment for the Gulf Stream configuration; CCM-C (Gulf) is defined in the same way. For the information-entropy metric in the Gulf Stream experiments, the SSH binning range is adjusted to [−1.3, 0.7] m to better match the SSH range in the Gulf Stream.

4.4. Validation in AVISO and Gulf Stream Regions

To further assess the robustness of the DAM and CCM models, we evaluate their performance against AVISO satellite altimetry data (Section 2.1) in the Kuroshio Extension. The models presented in Sections 4.1 and 4.2 are applied directly to AVISO without additional retraining. Both models retain high skill in the AVISO evaluation, and CCM shows improvements over DAM. Therefore, analyses based on both CMIP6 and AVISO show that adding the cascade constraint improves the prediction skill of SSH and related dynamical metrics, particularly reducing errors in geostrophic EKE.

This improvement by adding cascade constraints also holds in other energetic regions. For example, we examined regional transferability of the DAM and CCM models by conducting experiments in the Gulf Stream, another energetic western boundary current. Because the land–sea mask differs from that in the Kuroshio Extension, we retrain both DAM and CCM for the Gulf Stream configuration using the same training strategy described above. Both models achieve high skill in the Gulf Stream evaluation, and CCM generally improves over DAM across most metrics (Table 5).

5. Discussion and Conclusions

This study develops two ML models for SSH prediction and systematically evaluates their prediction skill and degree of climate adaptability. The novelty of DAM lies in integrating the dual attention mechanisms into ConvLSTM and developing a Trend-Magnitude Loss, which improves SSH prediction by approximately 3% in SSH NRMSE (Table S1 in Supporting Information S1). CCM additionally introduces the KE cascade constraint through the ML training process. Both models maintain robust performance under altered oceanic conditions. Incorporating the KE cascade constraint can further improve the prediction skill of SSH and the relevant variables (e.g., geostrophic EKE, kurtosis). Additional validations using AVISO observations and in the Gulf Stream region further confirm the robustness of the proposed models. A comparison with a KE-spectrum constraint further suggests that the cascade constraint improves both the KE flux and the KE spectrum, whereas the spectrum constraint primarily improves the KE spectrum. Additional extended-forecast tests further indicate that the advantage of the KE cascade constraint is maintained beyond the 7-day prediction horizon (Text S2 and Figure S2 in Supporting Information S1).

This study provides a systematic and transferable framework for evaluating the climate adaptability of ML models. This framework could be applied across different ocean regions, data sets, climate scenarios and ML methods, which will be our future work. The KE cascade constraint introduced in this study, together with the dual attention mechanism, is architecture-agnostic and modular, allowing seamless integration into diverse ML architectures, including Transformers and Graph Neural Networks (Vaswani et al., 2017; Wu et al., 2020), opening a pathway for its application to emerging ML approaches in ocean modeling.

This work has broader implications. For climate prediction, the climate adaptability demonstrated here suggests that ML models trained on present-day could complement CMIP6 in projecting future ocean states, without requiring retraining on future climate data. The appropriate merging of data-driven and physics-based models could eventually lead to an AI Model Intercomparison Project (AIMIP), which aims to predict and interpret climate change similar to CMIP. For weather prediction, the cascade constraint could be extended to other geoscience applications, such as wind speed forecasting, where atmospheric KE cascades similarly regulate the distribution of energy across scales (Schubert et al., 2023). For state estimation, future observations are unavailable in operational forecasting, the ML-predicted SSH could serve as pseudo-observations to be assimilated into numerical ocean models, providing additional constraints during the forecast period when no real-time observations exist (Cremer et al., 2025). For understanding physical processes, the improvement achieved by incorporating the cascade constraint demonstrates that cross-scale energy transfers are a key mechanism regulating SSH variability. Moreover, the ML model can represent the KE cascade process, suggesting that ML models are not only useful for prediction but also potentially valuable for dynamical diagnosis.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The daily SSH data from AWI-CM-1-1-MR used in this study are available from the World Data Center for Climate at DKRZ, including the historical simulations (Semmler et al., 2022a) and SSP3-7.0 scenario simulations (Semmler et al., 2022b). The AVISO/DUACS satellite altimetry data used in this study are available from the Copernicus Marine Service (Copernicus Marine Service, 2024). The code used for model training and evaluation is publicly available from Zenodo (T. Zheng, 2026).

Acknowledgments

We thank the anonymous reviewers for their constructive comments and insightful suggestions, which helped improve the clarity and quality of this manuscript. This study was funded by National Natural Science Foundation of China (Grants 42476007 and 42076007).

References

Alexakis, A., & Biferale, L. (2018). Cascades and transitions in turbulent flows. *Physics Reports*, *767*, 1–101. https://doi.org/10.1016/j.physrep.2018.08.001

Aluie, H., Hecht, M., & Vallis, G. K. (2018). Mapping the energy cascade in the North Atlantic Ocean: The coarse-graining approach. *Journal of Physical Oceanography*, *48*(2), 225–244. https://doi.org/10.1175/jpo-d-17-0100.1

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Barceló-Llull, B., Rosselló, P., Combes, V., Sánchez-Román, A., Pujol, M. I., & Pascual, A. (2025). Kuroshio Extension and Gulf Stream dominate the Eddy Kinetic Energy intensification observed in the global ocean. *Scientific Reports*, *15*(1), 21754. https://doi.org/10.1038/s41598-025-06149-9

Beech, N., Rackow, T., Semmler, T., Danilov, S., Wang, Q., & Jung, T. (2022). Long-term evolution of ocean eddy activity in a warming world. *Nature Climate Change*, *12*(10), 910–917. https://doi.org/10.1038/s41558-022-01478-3

Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., et al. (2024). Climate-invariant machine learning. *Science Advances*, *10*(6), eadj7250. https://doi.org/10.1126/sciadv.adj7250

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, *126*(9), 098302. https://doi.org/10.1103/physrevlett.126.098302

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*(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Bonavita, M. (2024). On some limitations of current machine learning weather prediction models. *Geophysical Research Letters*, *51*(12), e2023GL107377. https://doi.org/10.1029/2023gl107377

Brettin, A. E., Zanna, L., & Barnes, E. A. (2025). Learning propagators for sea surface height forecasts using Koopman Autoencoders. *Geophysical Research Letters*, *52*(4), e2024GL112835. https://doi.org/10.1029/2024gl112835

Chelton, D. B., Schlax, M. G., & Samelson, R. M. (2011). Global observations of nonlinear mesoscale eddies. *Progress in Oceanography*, *91*(2), 167–216. https://doi.org/10.1016/j.pocean.2011.01.002

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023). FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, *6*(1), 190. https://doi.org/10.1038/s41612-023-00512-1

Chen, R., McClean, J. L., Gille, S. T., & Griesel, A. (2014). Isopycnal eddy diffusivities and critical layers in the Kuroshio Extension from an eddying ocean model. *Journal of Physical Oceanography*, *44*(8), 2191–2211. https://doi.org/10.1175/jpo-d-13-0258.1

Chen, Y., Wang, Y., Huang, G., & Tian, Q. (2024). Coupling physical factors for precipitation forecast in China with graph neural network. *Geophysical Research Letters*, *51*(2), e2023GL106676. https://doi.org/10.1029/2023gl106676

Chernick, M. R. (2011). *Bootstrap methods: A guide for practitioners and researchers*. John Wiley & Sons.

Copernicus Marine Service. (2024). Global ocean gridded L4 sea surface heights and derived variables reprocessed 1993 ongoing (SEALEVEL_GLO_PHY_L4_MY_008_047) [Dataset]. *E.U. Copernicus Marine Service Information (CMEMS), Marine Data Store (MDS)*. https://doi.org/10.48670/moi-00148

Cremer, C. J. M., Mariegaard, J. S., & Andersson, H. J. (2025). A hybrid data assimilation and machine learning approach for enhancing operational forecasting in 2D hydrodynamic models. *Journal of Hydroinformatics*, *27*(3), 493–507. https://doi.org/10.2166/hydro.2025.289

Danilov, S., Kivman, G., & Schröter, J. (2004). A finite-element ocean model: Principles and evaluation. *Ocean Modelling*, *6*(2), 125–150. https://doi.org/10.1016/s1463-5003(02)00063-x

Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in Geophysics*, *61*, 1–55.

Ducet, N., Le Traon, P.-Y., & Reverdin, G. (2000). Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and-2. *Journal of Geophysical Research*, *105*(C8), 19477–19498. https://doi.org/10.1029/2000jc900063

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Collins, W. D., Gentine, P., Barnes, E. A., Barreiro, M., Beucler, T., et al. (2024). Pushing the frontiers in climate modelling and analysis with machine learning. *Nature Climate Change*, *14*(9), 916–928. https://doi.org/10.1038/s41558-024-02095-y

Fox-Kemper, B. (2021). Ocean, cryosphere and sea level change. In *AGU Fall Meeting Abstracts*.

Geng, Q., Chen, R., Qiu, B., Jing, Z., Su, X., Huang, G., et al. (2025). Intensification of oceanic inverse energy cascade under global warming. *arXiv preprint arXiv:2505.06327*.

Guan, W., Chen, R., Zhang, H., Yang, Y., & Wei, H. (2022). Seasonal surface eddy mixing in the Kuroshio Extension: Estimation and machine learning prediction. *Journal of Geophysical Research: Oceans*, *127*(3), e2021JC017967. https://doi.org/10.1029/2021jc017967

Guillaumin, A. P., & Zanna, L. (2021). Stochastic-deep learning parameterization of ocean momentum forcing. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002534. https://doi.org/10.1029/2021ms002534

Gultekin, C., Subel, A., Zhang, C., Leibovich, M., Perezhogin, P., Adcroft, A., et al. (2024). An analysis of deep learning parameterizations for ocean subgrid eddy forcing. *arXiv preprint arXiv:2411.06604*.

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7

Hernanz, A., García-Valero, J. A., Domínguez, M., & Rodríguez-Camino, E. (2022). A critical view on the suitability of machine learning techniques to downscale climate change projections: Illustration for temperature with a toy experiment. *Atmospheric Science Letters*, *23*(6), e1087. https://doi.org/10.1002/asl.1087

Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., & Saynisch-Wagner, J. (2021). Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nature Machine Intelligence*, *3*(8), 667–674. https://doi.org/10.1038/s42256-021-00374-3

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2018). Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, *31*(8), 1544–1554. https://doi.org/10.1109/tkde.2018.2861006

Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmaeilzadeh, S., et al. (2021). Physics-informed machine learning: Case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200093. https://doi.org/10.1098/rsta.2020.0093

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., et al. (2024). Neural general circulation models for weather and climate. *Nature*, *632*(8027), 1060–1066. https://doi.org/10.1038/s41586-024-07744-y

Kolmogorov, A. N. (1991). The local structure of turbulence in incompressible viscous fluid for very large Reynolds Numbers. *Doklady Akademii Nauk SSSR*, *30*, 301.

Lanzante, J. R. (2021). Testing for differences between two distributions in the presence of serial correlation using the Kolmogorov–Smirnov and Kuiper's tests.

Li, G., Cheng, L., Zhu, J., Trenberth, K. E., Mann, M. E., & Abraham, J. P. (2020). Increasing ocean stratification over the past half-century. *Nature Climate Change*, *10*(12), 1116–1123. https://doi.org/10.1038/s41558-020-00918-2

Lin, Z., Li, M., Zheng, Z., Cheng, Y., & Yuan, C. (2020). Self-attention ConvLSTM for spatiotemporal prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Ling, F., Luo, J.-J., Li, Y., Tang, T., Bai, L., Ouyang, W., & Yamagata, T. (2022). Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean Dipole. *Nature Communications*, *13*(1), 7681. https://doi.org/10.1038/s41467-022-35412-0

Martínez-Moreno, J., Hogg, A. M., England, M. H., Constantinou, N. C., Kiss, A. E., & Morrison, A. K. (2021). Global changes in oceanic mesoscale currents over the satellite altimetry record. *Nature Climate Change*, *11*(5), 397–403. https://doi.org/10.1038/s41558-021-01006-9

Massey, F. J., Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68–78. https://doi.org/10.1080/01621459.1951.10500769

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., et al. (2021). Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change (Vol. *2*, No. 1, p. 2391).

Meng, Y., Gao, F., Rigall, E., Dong, R., Dong, J., & Du, Q. (2023). Physical knowledge-enhanced deep neural network for sea surface temperature prediction. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–13. https://doi.org/10.1109/tgrs.2023.3257039

Meng, Y., Rigall, E., Chen, X., Gao, F., Dong, J., & Chen, S. (2021). Physics-guided generative adversarial networks for sea subsurface temperature prediction. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(7), 3357–3370. https://doi.org/10.1109/tnnls.2021.3123968

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2548–2563. https://doi.org/10.1029/2018ms001351

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Pasquini, F., Baatsen, M., François, B., Theeuwes, N., & Schmeits, M. (2026). Assessing the ability of a stretched-grid deep-learning weather prediction model to capture physical balances. *arXiv preprint arXiv:2604.01454*.

Peng, Q., Xie, S.-P., Wang, D., Huang, R. X., Chen, G., Shu, Y., et al. (2022). Surface warming–induced global acceleration of upper ocean currents. *Science Advances*, *8*(16), eabj8394. https://doi.org/10.1126/sciadv.abj8394

Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for physical applications*. Cambridge University Press.

Perezhogin, P., Adcroft, A., & Zanna, L. (2025). Generalizable neural-network parameterization of mesoscale eddies in idealized and global ocean models. *Geophysical Research Letters*, *52*(19), e2025GL117046. https://doi.org/10.1029/2025gl117046

Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al. (2022). Tackling climate change with machine learning. *ACM Computing Surveys*, *55*(2), 1–96. https://doi.org/10.1145/3485128

Sallée, J.-B., Pellichero, V., Akhoudas, C., Pauthenet, E., Vignes, L., Schmidtko, S., et al. (2021). Summertime increases in upper-ocean stratification and mixed-layer depth. *Nature*, *591*(7851), 592–598. https://doi.org/10.1038/s41586-021-03303-x

Schubert, R., Vergara, O., & Gula, J. (2023). The open ocean kinetic energy cascade is strongest in late winter and spring. *Communications Earth & Environment*, *4*(1), 450. https://doi.org/10.1038/s43247-023-01111-x

Scott, R. B., & Wang, F. (2005). Direct evidence of an oceanic inverse kinetic energy cascade from satellite altimetry. *Journal of Physical Oceanography*, *35*(9), 1650–1666. https://doi.org/10.1175/jpo2771.1

Sein, D. V., Koldunov, N. V., Danilov, S., Wang, Q., Sidorenko, D., Fast, I., et al. (2017). Ocean modeling on a mesh with resolution following the local Rossby radius. *Journal of Advances in Modeling Earth Systems*, *9*(7), 2601–2614. https://doi.org/10.1002/2017ms001099

Semmler, T., Danilov, S., Gierz, P., Goessling, H. F., Hegewald, J., Hinrichs, C., et al. (2020). Simulations for CMIP6 with the AWI climate model AWI-CM-1-1. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS002009. https://doi.org/10.1029/2019ms002009

Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Barbi, D., Hegewald, J., et al. (2019). IPCC DDC: AWI AWI-CM1. 1MR model output prepared for CMIP6 ScenarioMIP ssp585.

Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Barbi, D., Hegewald, J., et al. (2022a). CMIP6_supplemental CMIP AWI AWI-CM-1-1-MR [Dataset]. *World Data Center for Climate (WDCC) at DKRZ*. https://doi.org/10.26050/WDCC/C6sCMAWAWM

Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Barbi, D., Hegewald, J., et al. (2022b). CMIP6_supplemental ScenarioMIP AWI AWI-CM-1-1-MR [Dataset]. *World Data Center for Climate (WDCC) at DKRZ*. https://doi.org/10.26050/WDCC/C6sSPAWAWM

Sérazin, G., Penduff, T., Barnier, B., Molines, J.-M., Arbic, B. K., Müller, M., & Terray, L. (2018). Inverse cascades of kinetic energy as a source of intrinsic variability: A global OGCM study. *Journal of Physical Oceanography*, *48*(6), 1385–1408. https://doi.org/10.1175/jpo-d-17-0136.1

Sha, W., Jin, D., Liu, L., Zhang, X., Ling, F., Zhang, F., et al. (2025). A South China sea surface absolute dynamic topography prediction model based on convolutional long short-term memory network with self-attention mechanism. *Geophysical Research Letters*, *52*(17), e2025GL117019. https://doi.org/10.1029/2025gl117019

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, *104*(1), 148–175. https://doi.org/10.1109/jproc.2015.2494218

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shao, Q., Li, W., Han, G., Hou, G., Liu, S., Gong, Y., & Qu, P. (2021). A deep learning model for forecasting sea surface height anomalies and temperatures in the South China Sea. *Journal of Geophysical Research: Oceans*, *126*(7), e2021JC017515. https://doi.org/10.1029/2021jc017515

Shi, B., Ge, C., Lin, H., Xu, Y., Tan, Q., Peng, Y., & He, H. (2024). Sea surface temperature prediction using ConvLSTM-based model with deformable attention. *Remote Sensing*, *16*(22), 4126. https://doi.org/10.3390/rs16224126

Shi, B., Lin, H., Zhu, Y., He, H., Tan, Q., Peng, Y., & Liu, J. (2026). CA-ConvLSTM: A cross-attention-based deep learning framework for joint prediction of sea surface temperature and sea level anomaly. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, *28*.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, *25*.

Song, T., Han, R., Meng, F., Wang, J., Wei, W., & Peng, S. (2022). A significant wave height prediction method based on deep learning combining the correlation between wind and wind waves. *Frontiers in Marine Science*, *9*, 983007. https://doi.org/10.3389/fmars.2022.983007

Stammer, D. (1997). Global characteristics of ocean variability estimated from regional TOPEX/POSEIDON altimeter measurements. *Journal of Physical Oceanography*, *27*(8), 1743–1769. https://doi.org/10.1175/1520-0485(1997)027<1743:gcoove>2.0.co;2

Storer, B. A., Buzzicotti, M., Khatri, H., Griffies, S. M., & Aluie, H. (2023). Global cascade of kinetic energy in the ocean and the atmospheric imprint. *Science Advances*, *9*(51), eadi7420. https://doi.org/10.1126/sciadv.adi7420

Su, Z., Wang, J., Klein, P., Thompson, A. F., & Menemenlis, D. (2018). Ocean submesoscales as a key component of the global heat budget. *Nature Communications*, *9*(1), 775. https://doi.org/10.1038/s41467-018-02983-w

Sura, P., & Gille, S. T. (2010). Stochastic dynamics of sea surface height variability. *Journal of Physical Oceanography*, *40*(7), 1582–1596. https://doi.org/10.1175/2010jpo4331.1

Taburet, G., Sanchez-Roman, A., Ballarotta, M., Pujol, M.-I., Legeais, J.-F., Fournier, F., et al. (2019). DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Science*, *15*(5), 1207–1224. https://doi.org/10.5194/os-15-1207-2019

Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Von Storch, H., & Zwiers, F. W. (2002). *Statistical analysis in climate research*. Cambridge University Press.

Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., et al. (2014). The Finite Element Sea Ice-Ocean Model (FESOM) v. 1.4: Formulation of an ocean general circulation model. *Geoscientific Model Development*, *7*(2), 663–693. https://doi.org/10.5194/gmd-7-663-2014

Wang, S., Jing, Z., Wu, L., Sun, S., Chen, Z., Ma, X., & Gan, B. (2024). A more quiescent deep ocean under global warming. *Nature Climate Change*, *14*(9), 961–967. https://doi.org/10.1038/s41558-024-02075-2

Wang, S., Liu, Z., & Pang, C. (2015). Geographical distribution and anisotropy of the inverse kinetic energy cascade, and its role in the eddy equilibrium processes. *Journal of Geophysical Research: Oceans*, *120*(7), 4891–4906. https://doi.org/10.1002/2014jc010476

Wang, X., Wang, R., Hu, N., Wang, P., Huo, P., Wang, G., et al. (2024). XiHe: A data-driven model for global ocean eddy-resolving forecasting. *arXiv preprint arXiv:2402.02995*.

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 4–24. https://doi.org/10.1109/tnnls.2020.2978386

Yuval, J., & O'Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications*, *11*(1), 3295. https://doi.org/10.1038/s41467-020-17142-3

Zhang, G., Chen, R., Li, X., Li, L., Wei, H., & Guan, W. (2023). Temporal variability of global surface eddy diffusivities: Estimates and machine learning prediction. *Journal of Physical Oceanography*, *53*(7), 1711–1730. https://doi.org/10.1175/jpo-d-22-0251.1

Zheng, L., Lu, W., & Zhou, Q. (2023). Weather image-based short-term dense wind speed forecast with a ConvLSTM-LSTM deep learning model. *Building and Environment*, *239*, 110446. https://doi.org/10.1016/j.buildenv.2023.110446

Zheng, T. (2026). Climate-adaptive and cascade-constrained machine learning prediction for sea surface height [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.20196637

Zhou, G., Chen, J., Liu, M., & Ma, L. (2024). A spatiotemporal attention-augmented ConvLSTM model for ocean remote sensing reflectance prediction. *International Journal of Applied Earth Observation and Geoinformation*, *129*, 103815. https://doi.org/10.1016/j.jag.2024.103815

Zhou, G., & Cheng, X. (2022). Impacts of oceanic fronts and eddies in the Kuroshio-Oyashio Extension region on the atmospheric general circulation and storm track. *Advances in Atmospheric Sciences*, *39*(1), 22–54. https://doi.org/10.1007/s00376-021-0408-4

Zhu, Y., Zhang, R.-H., Moum, J. N., Wang, F., Li, X., & Li, D. (2022). Physics-informed deep-learning parameterization of ocean vertical mixing improves climate simulations. *National Science Review*, *9*(8), nwac044. https://doi.org/10.1093/nsr/nwac044

References From the Supporting Information

Balogun, A.-L., & Adebisi, N. (2021). Sea level prediction using ARIMA, SVR and LSTM neural network: Assessing the impact of ensemble Ocean-Atmospheric processes on models' accuracy. *Geomatics, Natural Hazards and Risk*, *12*(1), 653–674. https://doi.org/10.1080/19475705.2021.1887372

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., et al. (2007). The HYCOM (hybrid coordinate ocean model) data assimilative system. *Journal of Marine Systems*, *65*(1–4), 60–83. https://doi.org/10.1016/j.jmarsys.2005.09.016

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, J., van den Dool, H. M., & Barnston, A. G. (1996). Long-lead seasonal temperature prediction using optimal climate normals. *Journal of Climate*, *9*(4), 809–817. https://doi.org/10.1175/1520-0442(1996)009<0809:llstpu>2.0.co;2

Li, L., Vidard, A., Le Dimet, F.-X., & Ma, J. (2019). Topological data assimilation using Wasserstein distance. *Inverse Problems*, *35*(1), 015006. https://doi.org/10.1088/1361-6420/aae993

Li, W., Xie, Y., Deng, S.-M., & Wang, Q. (2010). Application of the multigrid method to the two-dimensional Doppler radar radial velocity data assimilation. *Journal of Atmospheric and Oceanic Technology*, *27*(2), 319–332. https://doi.org/10.1175/2009jtecha1271.1

Liu, J., Jin, B., Wang, L., & Xu, L. (2022). Sea surface height prediction with deep learning based on attention mechanism. *IEEE Geoscience and Remote Sensing Letters*, *19*, 3039062. https://doi.org/10.1109/lgrs.2020.3039062

Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, *9*(4), 347–404. https://doi.org/10.1016/j.ocemod.2004.08.002

Yang, Z., Yang, X., & Lin, Q. (2021). TCTN: A 3D-temporal convolutional transformer network for spatiotemporal predictive learning. *arXiv preprint arXiv:2112.01085*.