https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

Deep Learning-Enhanced Background Error Covariance Estimation for Massive-Ensemble Kalman Filter in Sea Surface Temperature Forecasting

Baoxu Li[1], Hongze Leng[1], Junqiang Song[1,2], Wuxin Wang[1,2], Taikang Yuan[1], Xiang Wang[1], Jinhui Yang[1], and Hang Cao[1]

[1]College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China
[2]College of Computer Science and Technology, National University of Defense Technology, Changsha, China

Correspondence: Hongze Leng (hzleng@nudt.edu.cn)

Abstract. In recent years, deep learning-based ocean forecasting has become a prominent research focus. However, recent studies often rely on operational ocean forecast systems to provide initial conditions. Operational ocean forecast systems typically use Ensemble Data Assimilation (EDA) to generate these initial conditions. Nonetheless, the high computational cost of numerical models limits the ensemble sizes in EDA, resulting in rank deficiencies in the background error covariance matrix and introducing spurious correlations. To address these challenges and advance the operationalization of deep learning-based ocean forecasting, we propose Deepcov-EnKF, a deep learning-enhanced background error covariance method for massive Ensemble Kalman Filter (EnKF) applications in Sea Surface Temperature (SST) forecasting. The proposed method incorporates a deep learning-based SST forecasting model to generate approximately 5,000 ensemble members. It then directly maps this high-dimensional perturbation set into a background error covariance matrix using a deep neural network. This approach reduces the computational cost of covariance estimation by more than 200-fold compared to conventional techniques. Experimental results demonstrate that the forecasting model can initialize reliable 60-day SST predictions with a Root Mean Square Error (RMSE) of approximately 0.6°C. Furthermore, assimilation diagnostics reveal that Deepcov-EnKF effectively resolves the spurious correlations in the covariance matrix. The method also exhibits robust stability and outperforms advanced numerical assimilation methods during a 360-day cycling forecasting experiment. This study confirms that Deepcov-EnKF overcomes key limitations of traditional EDA frameworks, significantly enhances the accuracy of SST assimilation, and lays the foundation for high-precision marine forecasting systems.

1 Introduction

Ocean forecasting is of great significance in modern scientific research and application. High-resolution ocean forecasting models can provide accurate predictions of the state of the oceans, which is crucial for weather forecasting and climate research (Hewitt et al., 2017). The differential equation framework underlying traditional ocean numerical model predictions is difficult to comprehensively characterize the multi-scale nonlinear processes of the ocean, and there are empirical biases in parameterization schemes (such as turbulence mixing coefficients, air-sea fluxes, etc.) (Krasnopolsky et al., 2002; Köllisch et al.,

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

2018; Kim et al., 2022). The model errors show a non-linear cumulative effect in the integration process, which is prone to systematic biases in the long-term forecast (Danforth et al., 2007). At the same time, improving spatial and temporal resolution and increasing model complexity are often mutually constrained by computational costs, which limits the prediction accuracy of small and medium-scale phenomena (Xu et al., 2021; Vlachas et al., 2020).

In recent years, deep learning-based ocean forecasting models have made remarkable progress in technological breakthroughs and application practices, gradually overcoming the limitations of traditional numerical models in terms of complex nonlinear process inscription and computational efficiency (Bento et al., 2021; Abbas et al., 2024; Lu et al., 2025; Immas et al., 2021). Deep learning-based models such as Convolutional Neural Networks (CNN) (LeCun et al., 1998), Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997), 3D U-Net (Çiçek et al., 2016) have significantly improved prediction accuracy by automatically extracting spatiotemporal features from massive data. For example, compared to numerical models, the coupled CNN-LSTM model achieves more than 70% error reduction and 60% improvement in computational efficiency to predict the effective wave height (Zhang et al., 2024). In addition, large-scale deep learning-based ocean forecasting models are developing rapidly. Ocean prediction models such as XiHe (Wang et al., 2024), AI-GOMS (Xiong et al., 2023) and OceanNet (Chattopadhyay et al., 2024) can make full use of large amounts of data and establish an accurate mapping between model inputs and outputs, showing great potential in terms of prediction accuracy, timeliness and computational efficiency. In terms of accuracy, deep learning-based ocean forecasting models have been able to match or even surpass traditional numerical prediction models in terms of statistical indicators (Ali et al., 2021; Shao et al., 2022; Fei et al., 2022). Thanks to the rapid inference capability of AI models, the speed of AI ocean prediction models has been improved by orders of magnitude (Ling et al., 2024; Shao et al., 2021), breaking the time-consuming limitation of traditional numerical forecasting.

The accuracy of the initial field, as the core input to the ocean prediction system, is critical for both traditional ocean models and AI prediction models (Mu et al., 2015; Bai et al., 2022). Small initial errors can be exponentially amplified in non-linear dynamical processes, leading to prediction failures. For AI prediction models, the quality of the initial field affects the generalization ability of the data-driven model, and the difference in the distribution between the training data and the real initial state can induce model bias. Therefore, both the acquisition and optimization of high-precision initial fields are key prerequisites to improve the forecasting capability of the marine environment (Blondel et al., 2010; Mu et al., 2015). In traditional numerical models, the initial field is combined with observational data through data assimilation techniques (Dong et al., 2022; Howard et al., 2024), and its accuracy directly determines the reliability of the numerical solution of physical equations (Moore et al., 2019; Smit et al., 2021; Martin et al., 2007). The core lies in the use of mathematical optimization methods such as Ensemble Kalman Filter(EnKF) (Evensen, 1994) and variational assimilation (Courtier et al., 1994) to dynamically adjust the initial state, effectively inhibit the exponential growth of the error. Unlike traditional variational assimilation or static error statistics methods, ensemble data assimilation significantly improves the fusion of background field and observations by dynamically generating multiple ensemble members and estimating the forecast error covariance in real time (Tandeo et al., 2020; Houtekamer and Mitchell, 1998). It is widely used in meteorological and oceanographic forecasting, climate research and environmental monitoring, and is one of the main data assimilation methods in modern numerical prediction systems (Whitaker et al., 2008; Gustafsson et al., 2018). Data assimilation is the core technology for

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

improving the predictive power of traditional models and also provides more physically constrained initial conditions for AI prediction models, which is particularly important for the refined prediction of marine environments (Cummings, 2005; Smit et al., 2021; Beiser et al., 2025).

The construction of background error covariance matrices in ocean ensemble data assimilation relies on the statistical properties of the ensemble members, and the limited ensemble size can lead to biased covariance estimates (Zhang et al., 2020; Sun et al., 2024). Therefore, it is difficult to accurately represent the complex correlations of multiscale ocean dynamical processes. The uneven spatial and temporal distribution and sparseness of the observed data (e.g. currents, temperature and salinity profiles) also limit the assimilation effect (Wang et al., 2022; Alonso-González et al., 2023), especially the current data, which are much more difficult to assimilate due to their fast dynamic changes and high observation cost (Kerry et al., 2024; Toste et al., 2024). In addition, traditional ensemble assimilation methods (e.g., the EnKF) are limited in handling strongly nonlinear dynamical processes such as high-frequency turbulence or extreme events, because the underlying linear analysis step cannot fully capture the nonlinear evolution simulated by the forecast model. Moreover, these methods rely on the physical framework of the model itself, which may not be able to effectively correct forecast errors (Labahn et al., 2020; Chattopadhyay et al., 2023; Evensen et al., 2024). These limitations highlight the need for further optimization of ensemble sampling strategies, development of dynamic background error estimation methods, and integration of artificial intelligence techniques.

In recent years, there have been several studies integrating artificial intelligence into ensemble data assimilation to improve assimilation effect. Chattopadhyay et al. (2023) proposed H-EnKF, which uses pretrained deep learning data to drive a model to generate and evolve large-scale state ensembles in order to accurately compute background error covariance matrices and to reduce sampling errors. The method does not require any localization strategy and is able to significantly improve the estimation of the initial field. Hammoud et al. (2024) combined deep reinforcement learning with EnKF to propose a new assimilation strategy that uses observations of global or partial state variables for state correction. The method performs well in the Lorenz 63 and 96 systems, which is able to handle non-Gaussian observations, and overcomes the limitations of traditional EnKF. Kang et al. (2021) propose an EnKF framework based on the Convolutional Variational Auto-Encoder (CVAE), which significantly improves the accuracy of the estimation and reduces the estimation error. Chen et al. (2022) present a new method based on the gradient-free profile assimilation training framework to optimize neural network parameters using EnKF and ESMDA algorithms. The method performs well on regression problems and provides an alternative to existing neural networks for online/offline training. Mücke et al. (2024) proposed Deep Latent Space Particle Filters (D-LSPF) for real-time profile assimilation and uncertainty quantification.D-LSPF significantly speeds up complex profile assimilation tasks by combining deep learning and particle filtering without sacrificing accuracy and is capable of fusing increasingly complex models and data in real time. The above studies show that deep learning methods have demonstrated significant advantages in the field of ensemble data assimilation, with their efficiency in generating particles or ensembles and their accuracy in characterizing the background error covariance matrix exceeding that of traditional methods. However, there are still limitations to the current research results. First, the experiments are based on idealized numerical models that simplify the physical processes and have not been validated in real atmospheric or oceanic model systems. In addition, the spatial and temporal dimensions of existing

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

studies are orders of magnitude different from operational forecasting needs, and the ensemble sizes generated are far from meeting the computational requirements of global numerical prediction systems.

To facilitate the operational application of deep learning in ocean forecasting, this study proposes Deepcov-EnKF, a deep learning-enhanced background error covariance method for massive EnKF applications in Sea Surface Temperature (SST) forecasting. The proposed method leverages a deep learning-based SST prediction model to efficiently generate around 5,000 ensemble members. A deep neural network is then employed to directly map these high-dimensional perturbations into an accurate background error covariance matrix. This approach reduces computational cost by more than 200 times compared to conventional covariance estimation methods. Experimental results indicate that Deepcov-EnKF enables reliable 60-day SST forecasts. Assimilation diagnostics confirm that the method effectively suppresses spurious correlations and demonstrates strong stability throughout a 360-day cycling experiment without error accumulation. This work breaks through key bottlenecks of traditional EDA, significantly enhances SST assimilation quality. And it provides an efficient data assimilation approach that can be extended to multiple variables for the development of a new generation of high-precision marine forecasting systems.

The remainder of the paper is organized as follows. Section 2 describes the EnKF algorithm and modeling framework in detail. The data used and data processing are described in Section 3. Section 4 presents the results and Section 5 summarizes the conclusions.

2 Methods

2.1 EnKF

The basic idea of EnKF is to use points in a probability space to represent one possibility of a state variable and to use the ensemble to estimate the forecast error covariance matrix (Evensen, 1994). First, N samples are generated in the initial distribution $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$ and used to construct the initial ensemble:

$$x_0^{(i)} = \bar{x}_0 + \epsilon_0^{(i)}, \quad \epsilon_0^{(i)} \sim \mathcal{N}(0, P_0), \quad i = 1, \ldots, N \tag{1}$$

where $x_0^{(i)}$ denotes the initial ensemble members. Then each ensemble member is integrated using the nonlinear model $\mathcal{M}$:

$$x_{k|k-1}^{(i)} = \mathcal{M}_{k-1}(x_{k-1|k-1}^{(i)}) + w_{k-1}^{(i)}, \quad w_{k-1}^{(i)} \sim \mathcal{N}(0, Q_{k-1}) \tag{2}$$

w denotes the mode integration error with covariance Q. The prediction mean $\bar{x}_{k|k-1}$ and the prediction perturbation matrix $X'_{k|k-1}$ are calculated from the predicted values of each ensemble member:

$$\bar{x}_{k|k-1} = \frac{1}{N} \sum_{i=1}^{N} x_{k|k-1}^{(i)} \tag{3}$$

$$X'_{k|k-1} = \left[x_{k|k-1}^{(1)} - \bar{x}_{k|k-1}, \ldots, x_{k|k-1}^{(N)} - \bar{x}_{k|k-1} \right] \tag{4}$$

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

Perturbations need to be added to the observations in the ensemble Kalman filter to avoid underestimating the posterior covariance matrix. Compute the ensemble mean model equivalent of the observations $\bar{y}_k$ and the observation perturbation matrix Y_k' accordingly:

$$y_k^{(i)} = y_k + v_k^{(i)}, \quad v_k^{(i)} \sim \mathcal{N}(0, R_k), \quad i = 1, \dots, N \tag{5}$$

$$\bar{y}_k = \frac{1}{N} \sum_{i=1}^{N} \mathcal{H}\left(x_{k|k-1}^{(i)}\right) \tag{6}$$

$$Y_k' = \left[\mathcal{H}\left(x_{k|k-1}^{(1)}\right) - \bar{y}_k, \dots, \mathcal{H}\left(x_{k|k-1}^{(N)}\right) - \bar{y}_k\right] \tag{7}$$

where $\mathcal{H}$ stands for the nonlinear observation operator, y_k represents the input observation data, $y_k^{(i)}$ represents the ith observation data point with observation error, and R_k represents the covariance matrix of observation errors. The state-observation covariance matrix P_{xy} as well as the observation-observation covariance matrix P_{yy} can be calculated based on the prediction perturbation matrix and the observation perturbation matrix:

$$P_{xy} = \frac{1}{N-1} X_{k|k-1}' Y_k'^{\top} \tag{8}$$

$$P_{yy} = \frac{1}{N-1} Y_k' Y_k'^{\top} + R_k \tag{9}$$

The Kalman gain matrix corresponding to the ensemble Kalman filtering can be expressed as:

$$K_k = P_{xy} P_{yy}^{-1} \tag{10}$$

Finally, each ensemble member is updated and the posterior statistic is calculated. The updated mean $barx_{k|k}$ and covariance $P_{k|k}$ are calculated:

$$x_{k|k}^{(i)} = x_{k|k-1}^{(i)} + K_k\left(y_k^{(i)} - \mathcal{H}\left(x_{k|k-1}^{(i)}\right)\right), \quad i = 1, \dots, N \tag{11}$$

$$\bar{x}_{k|k} = \frac{1}{N} \sum_{i=1}^{N} x_{k|k}^{(i)} \tag{12}$$

$$P_{k|k} \approx \frac{1}{N-1} X_{k|k}' X_{k|k}'^{\top}, \quad X_{k|k}' = \left[x_{k|k}^{(1)} - \bar{x}_{k|k}, \dots, x_{k|k}^{(N)} - \bar{x}_{k|k}\right] \tag{13}$$

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

2.2 Architecture of Forecasting Model

The SST prediction model constructed in this study adopts a hierarchical network architecture (shown in Fig. 1), with a Swin Transformer network as the feature extraction backbone at its core. The model input adopts a multi-source heterogeneous data fusion strategy, which mainly contains two types of key data sources: atmospheric forcing field data and SST data. The architecture adopts an encoder-decoder design, where the encoder part contains a 5-stage Swin Transformer module with 2×2 spatial downsampling at each stage, and the decoder gradually recovers the spatial resolution through transpose convolution. This design achieves full fusion of multiscale features while maintaining computational efficiency.

This study integrates sea surface temperature (SST) and forcing field data using a three-dimensional tensor structure $C \times H \times W$, where C represents the number of channels, and H and W denote the two-dimensional grid. In this case, the single channel SST 2D field, the atmospheric boundary forcing and the high-altitude dynamical field are stacked through the channel dimensions to form a $7 \times 120 \times 240$ composite input tensor$X \in R^{C_{\text{in}} \times H \times W}$. The data preprocessing core of the model uses Patchify technique to transform the raw grid data into serialized embedding vectors suitable for processing by the Transformer architecture. Each of the input data fields performs nonoverlapping local block partitioning, cutting the raw grid into 60 × 120 (i.e., 120/2 × 240/2) spatial blocks, each covering 2 × 2 neighbouring grid points. The grid point values within each block are mapped into fixed-dimension Patch Embedding through a linear projection layer to achieve the conversion from spatial grids to feature sequences. Different meteorological variables are independently blocked and embedded to preserve their physical characteristics, and then spliced into a unified multimodal input sequence in the embedding space. This process not only compresses the high-dimensional mesh data into a sequence structure, which reduces the computational complexity of the subsequent attention, but also retains the spatial correlation of the neighbouring meshes through the design of local blocks, which lays the foundation for Swin Transformer to capture the multi-modal weather patterns.

The Swin Transformer architecture implements a hierarchical feature mapping through five levels of cascade processing modules. The Swin Attention module accurately models the multi-scale spatial dependencies of meteorological fields through a hierarchical sliding-window mechanism, while significantly reducing computational complexity. The high-resolution grid is divided into local windows to focus on capturing the short-range interactions of forecast elements, avoiding the computational bottleneck of the traditional global attention at high resolution, and reducing the computational volume from $O(n^2)$ to $O(n)$. Remote spatial correlations are gradually established through a hierarchical design of merging windows layer by layer, and local boundaries are broken using window offset techniques. The attention weights of heterogeneous data such as SST, wind field and potential field are jointly calculated within the same window to implicitly learn the physical mechanism of sea-air interaction. At the same time, the relative position encoding enhances the robustness of the model in predicting ocean phenomena, and ultimately realizes efficient and high-precision spatial and temporal evolution prediction of complex ocean processes. Adaptive LayerNorm significantly improves the robustness and accuracy of the forecast by dynamically generating the Scale and Shift terms. Adaptive LayerNorm dynamically adjusts the normalization parameters based on temporal embedding and spatial context to adapt to the strong non-stationarity of the data. Time Embedding enhances the forecast accuracy in this model through a two-way encoding mechanism (initial moment and forecast time Δt). It enables the model to satisfy the time

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

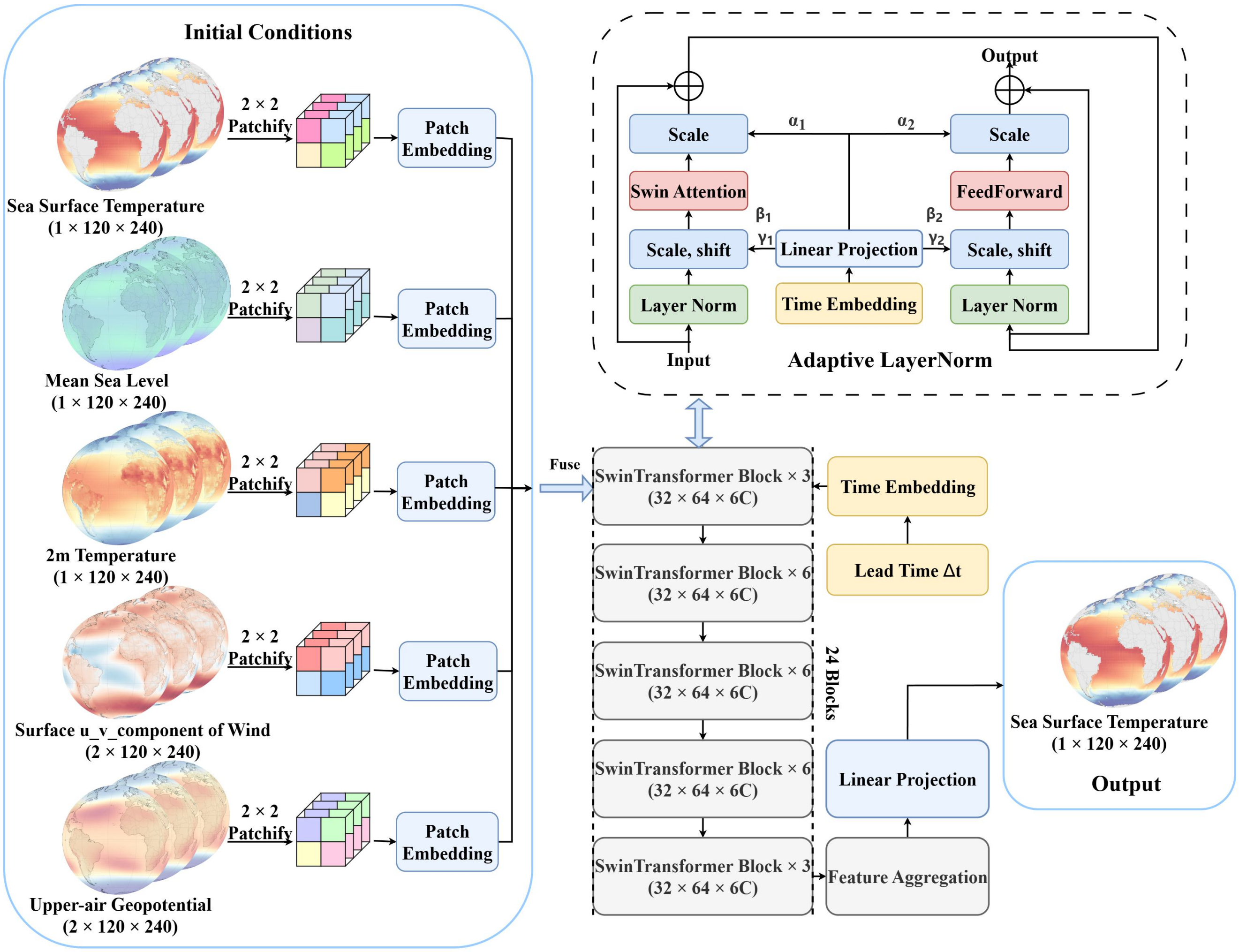

Figure 1. Overall architecture of the SST prediction model. The backbone of the prediction model is included, and the input variables include one fundamental ocean variable (sea surface temperature (SST)) and six atmospheric forcing variables (2m Temperature (t2m), Mean Sea Level (MSL), 10m Latitudinal Velocity (U), 10m Longitudinal Velocity (V), and 850hPa/500hPa Upper-air Geopotential). The detailed modelling framework of the Swin Transformer Block is shown in the upper right corner.

symmetry of physical laws, and adaptively regulate the prediction paths within a unified architecture, ultimately achieving accurate predictions across scales.

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

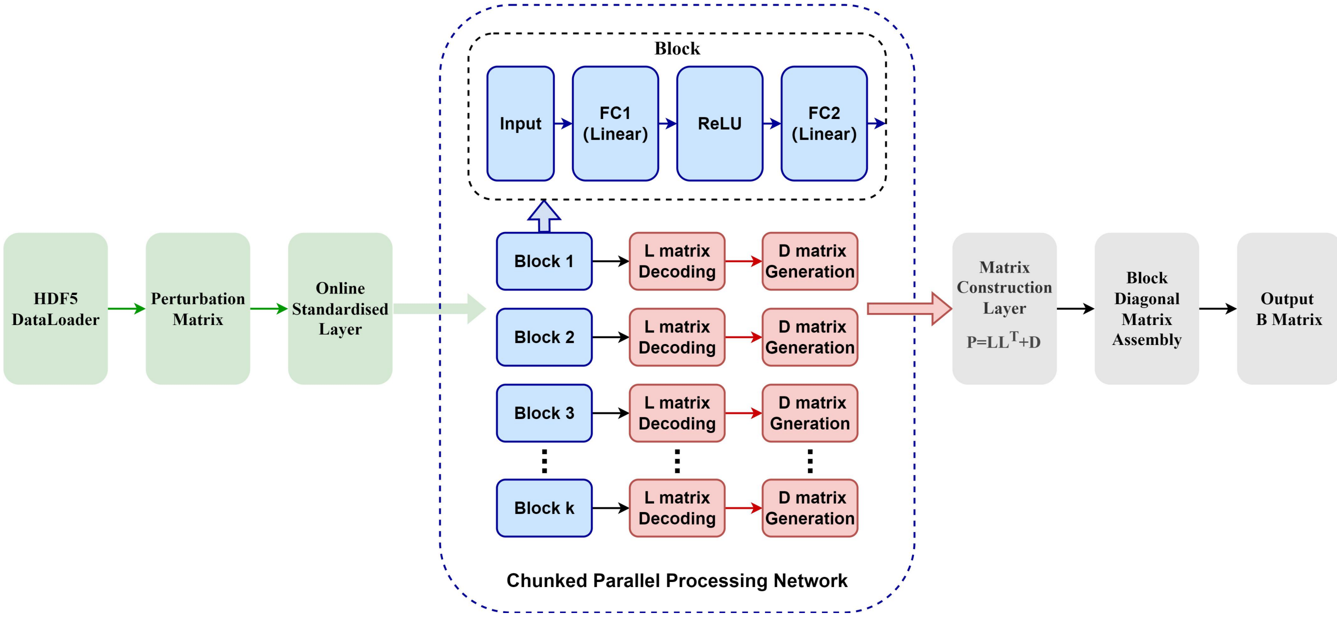

Figure 2. Overall framework of DL-BEC and network structure.

2.3 B Matrix Computing Acceleration Model

In the above prediction model, the horizontal grid is set to 120×240, in which case the computation of the background error covariance matrix B in the EnKF algorithm becomes extremely complex and time-consuming. The univariate state vector has a dimension of $N = 28,800$, and explicit storage of B requires about 10^8 elements, and the cumulative computational cost for each element leads to a very slow assimilation process. Traversing the spatial relations among all grid points to calculate the combinations of grid point pairs alone has nearly 800 million combinations, and each step of the operation consumes a large amount of computational resources. Therefore, the construction and optimization of B in the whole EnKF process is the bottleneck of real-time data assimilation. Especially when the model resolution or the number of variables continues to increase, the problem becomes more difficult.

We innovatively propose a deep learning-based generation model for background error covariance matrix (DL-BEC), aiming to break through the high-dimensional bottleneck of B matrix computation. The model takes the ensemble perturbation matrix as input, implicitly models the high-dimensional covariance structure via deep neural networks, and directly outputs B matrix estimates. The end-to-end architecture avoids the operation of explicitly constructing $N \times N$ matrices and greatly reduces computational complexity. The model inference process is fully parallelisable and real-time covariance update can be achieved under GPU acceleration. The specific model framework is shown in Fig. 2.

The core of the model lies in implicit modelling. Instead of computing or storing the complete B matrix directly, it learns a mapping from the input data to an efficient representation of the B matrix. The model learns how to extract features from the input perturbations through a neural network and gradually builds the low-rank part (L) and diagonally corrected part (D),

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

which together approximate the B matrix as

$$B \approx LL^{\top} + D. \tag{14}$$

Bannister (2008) points out that combining a low-rank flow-dependent set covariance with a diagonal variance term (to compensate for underestimation of set variance and errors arising from unresolved scales) is an effective and commonly used method for constructing a reliable background error covariance matrix. Our method is a direct application of this principle. Here, $L \in \mathbb{R}^{n \times r}$ (with $r \ll n$) captures the dominant multivariate error correlations. Its columns can be interpreted as the leading error modes. $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ is a diagonal correction that accounts for the grid-point variances and small-scale error variance not represented by the low-rank part $LL^{\top}$. The final target B matrix is then assembled from L and D. The model receives an ensemble perturbation matrix as input. Each column of this matrix represents the perturbation of the state vector of an ensemble member with respect to the ensemble average. The DataLoader module is responsible for efficiently loading the input data from storage and performing the necessary preprocessing. The Online Standardised Layer acts on each input sample to compute the mean and standard deviation of the input data for the current batch. The data is standardised using these statistics. This allows the data to be scaled consistently across batches and variables, accelerating network convergence and improving stability. The standardisation process implicitly includes first-order statistical information about the variance of the variables themselves, which is itself an important part of the diagonal elements of the B matrix.

The Feature Extraction and Matrix Construction Block (Block 1 to Block k) is the core part of the model and consists of a cascade of k structurally similar modules. Each Block is responsible for extracting deeper features from the input and progressively constructing and refining the representation of the matrix structure of B. L matrix Decoding / D matrix Generation is at the heart of each Block and receives high-level features from the output of FG2 and decodes / maps them into matrices of a particular form. The L matrix decoding generates a lower triangular matrix that captures the main, structured covariance information. The D matrix is responsible for capturing local variances not adequately represented by the low-rank component and small-scale, rapidly changing error components. L decoding and D generation are alternated to progressively refine the representation of the covariance structure.Matrix Construction Laye combines the final L and D matrices after k Block processing to form the core representation P of the model's predicted covariance matrix. P is a compact, efficient representation of the model's complete, high-dimensional background error covariance matrix B. Block Diagonal Matrix Assembly converts or interprets the single, large matrix P output by the model into the final desired B matrix.

The model is compressed from traditional $O(m^2)$ to $O(b^2)$ memory consumption by chunked low-rank decomposition technique (block size = 600, rank = 10), and achieves 28,800 dimensional real-time processing capability on a single GPU. The combination of Cholesky decomposition parameterization and eigenvalue regularization ensures that the output matrix is strictly symmetric and positive definite, with the smallest eigenvalue stabilized above 1e-6. At the same time, online normalization and mixed-accuracy training are used to achieve dynamic data adaptation. Compared with traditional sample covariance calculation methods, the entire forecast assimilation cycle experiment achieved a 7.2% acceleration on the A100 GPU. Its implicit matrix interface is seamlessly compatible with existing EnKF systems and is particularly suitable for large-scale data

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

assimilation scenarios such as meteorology and oceanography. At the same time, it compresses the single covariance modelling time to seconds, providing a solution for real-time forecasting and assimilation systems.

The training data are generated through two complementary strategies that together ensure both generic representativeness and physical relevance.First, we draw random samples over a wide range of idealized error patterns which allows the network to learn the intrinsic mapping from state perturbations to covariance matrices, without being overly tied to any particular flow realization.To guarantee that the learned mapping captures physically meaningful, flow-dependent error covariances, we augment the training dataset with perturbations extracted from ensemble forecasts. These real perturbations expose the model to the authentic variability and complex correlation structures present in operational forecasting systems.The network is trained for 100 epochs on this mixed dataset, which combines random and real perturbations.

2.4 Experimental Setup

In this study, a complete SST forecasting and assimilation cycle experimental system is constructed (the architecture is shown in Fig. 3). Its core process integrates the overall architecture of ensemble forecasting and data assimilation. In the initial field construction stage, the Breeding of Growing Modes (BGM) (Toth and Kalnay, 1993) method is used to perturb the initial field. Three time steps ($t_0 - 3\Delta t$ to t_0) before the initial moment are selected as the perturbation incubation period for specific implementation. The reason for selecting this incubation period is that our assimilation window is 24 hours, and the background field is derived from a 24-hour forecast. The growth cycle of BGM perturbations requires the capture of rapidly growing error modes that match this forecast duration. For 24-hour forecast errors, the timescale of growth is typically on the order of 1–3 days. Therefore, a 72-hour incubation period is sufficient to allow the dominant instability modes in the forecast to develop fully.The initial perturbation field is ensured to have statistical significance through successive mode integration and perturbation scale normalization.

During the forecast assimilation cycle, a pre-trained deep learning-based forecast model performs 24-hour scale forecasts for ensemble members. The generated forecast ensemble is then fused with the observations within the assimilation window. The trained DL-BEC is used to replace the explicit B matrix computation module of the traditional EnKF algorithm. The updated analytical ensemble is fed into the forecast model by serving as the initial field for the next assimilation cycle. To fully validate the system performance, the experiment is continuously executed for 360 full assimilation cycles. In this way, it is shown that the ensemble data assimilation system, which is a deep fusion of machine learning methods, significantly improves the accuracy of marine environment forecasts while maintaining long-term numerical stability.

3 Data Preprocess

The data used to train the SST daily prediction model integrate two authoritative global reanalysis datasets: Global Ocean Physics Reanalysis (GLORYS) (Jean-Michel et al., 2021) led by The Copernicus Marine Environment Monitoring Service (CMEMS), and ERA5 atmospheric reanalysis data (Hersbach et al., 2020) released by the European Center for Medium-Range Weather Forecasts (ECMWF). GLORYS is constructed based on the NEMO model, which integrates satellite observations with

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

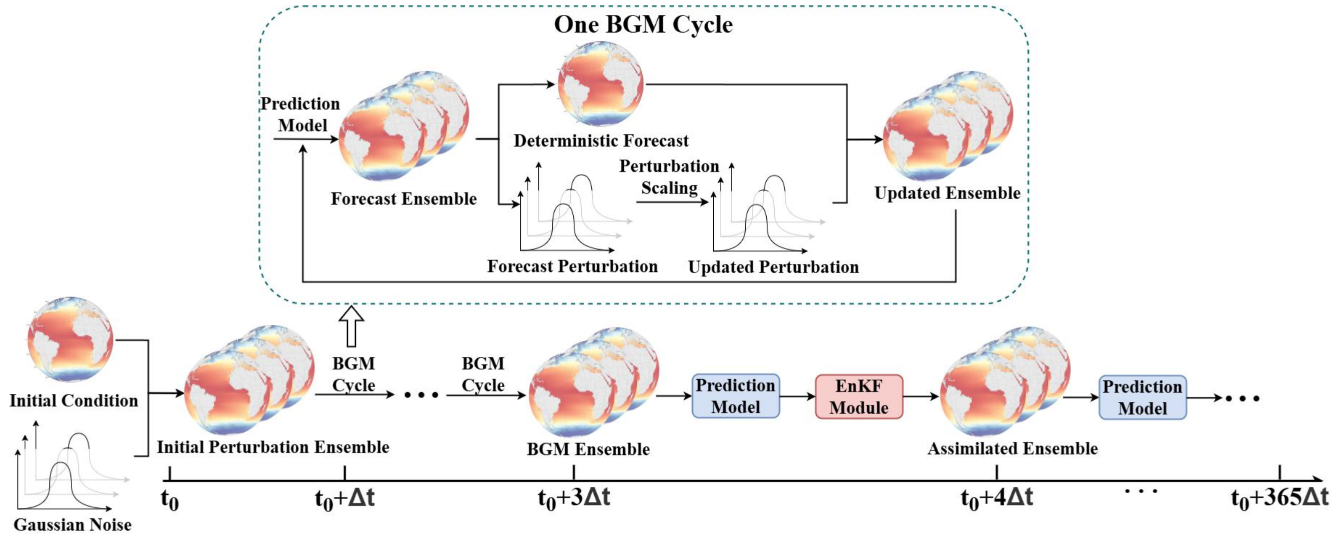

Figure 3. Overall framework of the experimental process.

in situ data such as Argo buoys, and reconstructs the high-resolution ocean dynamics through Singular Evolutive Extended Kalman (SEEK) filter (Pham et al., 1998). ERA5 relies on the IFS Cy41r2 model and 4D variational assimilation to integrate multisource observations. To keep the model training computationally feasible under our available resources, we screened and preprocessed the raw data and aggregated all fields to a uniform 1.5° × 1.5° grid. This resolution is specifically chosen to retain the large-scale thermal features that constitute the primary drivers of SST evolution on the daily timescale. At the same time, the moderate grid size substantially reduces computational complexity and memory demands, enabling thorough hyperparameter optimization and rapid model iteratio. That strikes a practical balance between accuracy and efficiency.

For observational data, we used the OISST (Optimum Interpolation Sea Surface Temperature) dataset (Reynolds, 1993; Huang et al., 2021). This is a global SST product released by the National Oceanic and Atmospheric Administration (NOAA) of the United States. It combines multisource satellite remote sensing infrared observations and actual measurement data from field buoys, ships, etc., covering a time span from 1981 to the present. We interpolated the raw OISST data onto a global 1.5° grid with a temporal resolution of one day. The OISST accounts for different uncertainties and provides an internal analysis error estimate. Therefore we construct R as a diagonal matrix, with the diagonal entries set to the square of the corresponding OISST error standard deviation.

To effectively capture and learn the complex dynamic interactions between the ocean and the atmosphere, we carefully selected key boundary conditions as model input, including sea-land masks and six key atmospheric variables. The input data covers SST and six atmospheric forcing variables (2m Temperature (t2m), Mean Sea Level (MSL), 10m Latitudinal Velocity (U), 10m Longitudinal Velocity (V), and 850hPa/500hPa Upper-air Geopotential). Their spatial dimensions are uniformly represented as a grid of $C \times 120 \times 240$. Two-dimensional variables are regarded as a single layer in the C dimension ($C = 1$), while three-dimensional variables contain multiple layers ($C > 1$, representing different vertical heights). We integrate all the

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

variables along their respective C dimensions to form a unified tensor structure and input it simultaneously. We divide the dataset into training, validation and test datasets. The training dataset includes data from 1993 to 2021, the validation dataset includes data from 2022, and the test dataset includes data from 2023.

4 Results

4.1 Prediction Results of Forecast Model

We conducted a single deterministic prediction and a ensemble prediction with a ensemble quantity of 5,000 using the same initial field. The global SST forecast biases distribution is shown in Figure 4. To quantitatively evaluate the performance and error level of the prediction model, this study uses RMSE (Root Mean Square Error) and ACC (Anomaly Correlation Coefficient) as evaluation indicators. The calculation formulas are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{15}$$

$$\mathrm{ACC} = \frac{\sum_{i=1}^{N}\hat{a}_i a_i}{\sqrt{\sum_{i=1}^{N}\hat{a}_i^2}\cdot\sqrt{\sum_{i=1}^{N}a_i^2}} \tag{16}$$

Where $a_i = y_i - \bar{y}$, $\hat{a}_i = \hat{y}_i - \bar{y}$. N represents the total number of samples (number of spatial grid points or time steps), $\hat{y}_i$ represents the predicted SST value for the i sample, y_i represents the reanalysis SST value for the i sample, and $\bar{y}$ represents historical daily climatological SST.

From the comparison of SST distribution maps, it can be seen that the AI prediction model trained in this study maintains good spatial consistency with actual reanalysis data within a 30-day forecast period. The model successfully captures the main distribution characteristics and trends of global SST. In terms of characterizing SST anomalies (such as significant warm/cold anomaly zones), the model's predictive performance remains at a high level overall. Although, there is still room for improvement in some details. A typical example is the cold water mass in the Antarctic Ocean: the model's predictions of its spatial extent and intensity extremes show some deviation compared to reanalysis data. It is worth noting that ensemble forecasts and deterministic forecasts exhibit complementary performance. Ensemble forecasts exhibit superior spatial continuity (smoother representation), with their predicted global SST distribution patterns more closely aligning with reality . On the other hand, deterministic forecasts perform better in depicting fine-scale structural details. For example, in capturing the intricate features of SST fluctuations in the equatorial Pacific region, deterministic forecasts demonstrate relatively superior capability.

By analyzing the trends in RMSE and ACC as forecast lead time changes, we can conduct a more precise quantitative assessment of the error evolution in both deterministic and ensemble forecasts. During the experiment, we defined periods of 15 days or less as short-term, 15 to 45 days as medium-term, and periods exceeding 45 days as long-term. The results show that the RMSE of deterministic forecasts remains stable at approximately 0.6°C even in forecasts as long as 60 days, which

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

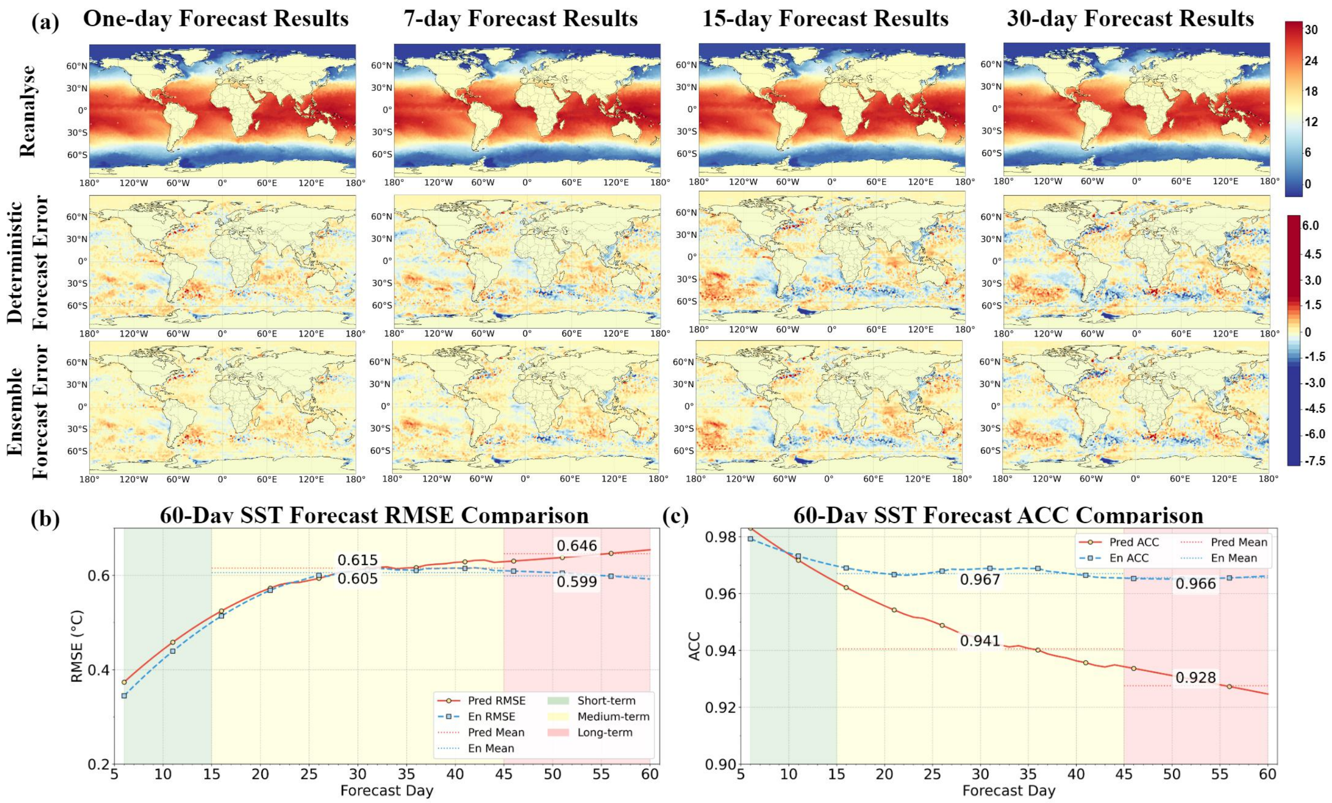

Figure 4. Forecast results of the prediction model. a. Comparison of global SST distribution maps for deterministic forecasts and ensemble forecasts error within 30 days. b. Comparison of RMSE for global SST for deterministic forecasts and ensemble forecasts within 60 days without spin-up period. Pred means deterministc forecast and En means ensemble mean. c. Comparison of ACC for global SST for deterministic forecasts and ensemble forecasts within 60 days without spin-up period.

fully demonstrates the model's exceptional long-term forecast performance. Ensemble forecasts exhibit relatively high RMSE in short-term forecasts, primarily due to disturbances introduced by initial conditions. As the forecast lead time increases, the average RMSE of ensemble forecasts gradually decreases below that of deterministic forecasts. This highlights the core advantage of ensemble averaging methods in suppressing error growth and enhancing the accuracy of medium- to long-term forecasts. Whether deterministic or ensemble forecasts, the ACC values remain above 0.9 at the 60-day ultra-long forecast lead time. This indicates that the model effectively captures the primary spatio-temporal evolution characteristics of the SST anomaly field. Within the long-term forecast range, the ensemble forecast's ACC values not only significantly exceed those of deterministic forecasts but also exhibit much smaller temporal fluctuations, demonstrating higher stability. This indicates that the long-term forecast results generated by ensemble forecasting possess stronger robustness and reference value.

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

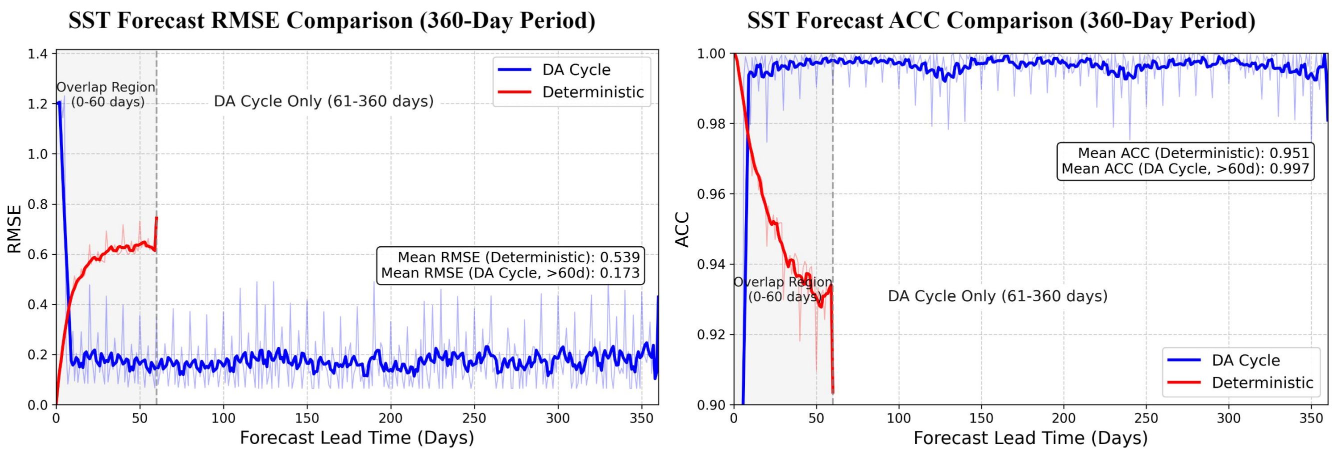

Figure 5. RMSE and ACC of SST in the 360-day forecast assimilation cycle experiment as a function of forecast days.

4.2 Results of Data Assimilation

This study conducted a 360-day forecast-assimilation cycle experiment to assess the stability and assimilation effectiveness of the system during long-term continuous operation. Figure 5 shows the changes in RMSE and ACC of SST predictions during the cycle experiment. Throughout the 360-day cycle experiment, the RMSE of SST predictions was stably maintained at a low level (approximately 0.2°C). This strongly demonstrates that the effective assimilation of real observational data continuously provides high-precision initial fields, laying a solid foundation for subsequent forecasts. More importantly, during the long-term continuous cycle process, neither the forecast model nor the assimilation system exhibited significant performance degradation, and the forecast results maintained a high degree of stability. This fully validates the robustness of the model and its potential reliability in operational environments. The SST forecast ACC value in the cycle experiment remained consistently stable at a high level above 0.95, significantly outperforming single independent forecasts. This indicates that the assimilation process, which continuously incorporates real observational data, not only optimizes initial conditions but also significantly enhances the model's effective forecast lead time.

To accurately assess the contribution of the assimilation process to forecast performance, this study systematically compared the results of single-day forecasts driven directly by observational data with those obtained after a 360-day forecast-assimilation cycle. The results are shown in Figure 6. The RMSE of forecasts generated after a 360-day continuous assimilation cycle showed a significant decrease compared to those of direct observational forecasts. It is worth noting that fluctuations occurred during the middle and latter stages of the 360-day cycle experiment. These quantitative results strongly demonstrate that the assimilation algorithm successfully integrates prior dynamical information from the background field with constraints from real-time observational data. Assimilation generates significantly optimized initial fields, which are the fundamental guarantee for subsequent high-precision forecasts. Continuous cycle assimilation not only optimizes the initial field in a single step but also effectively suppresses error accumulation by continuously utilizing new observations to correct the system state. The

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

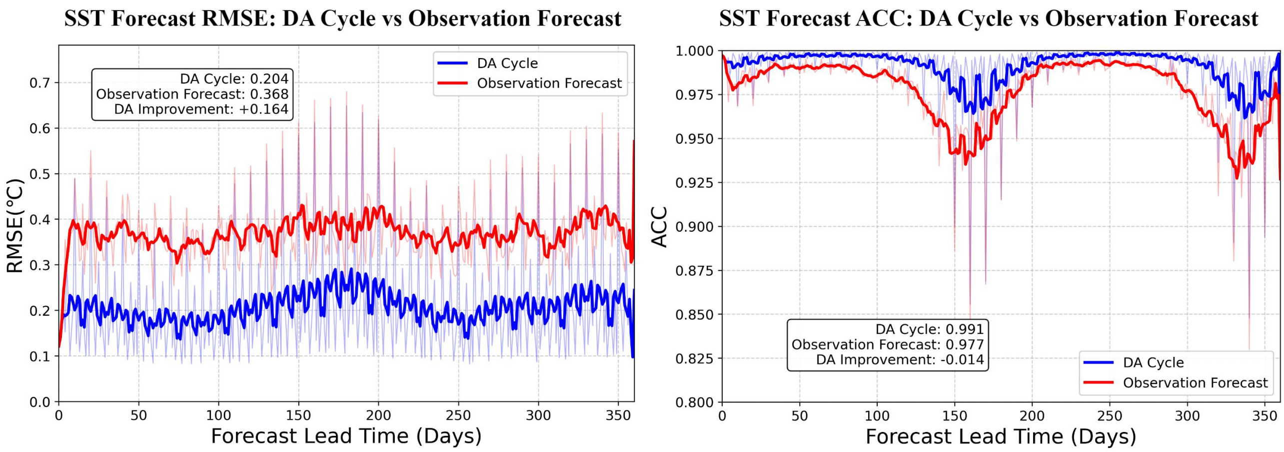

Figure 6. Comparison of RMSE and ACC for forecast assimilation cycles and RMSE and ACC for one-day forecasts using direct observations.

introduction and continuous operation of a data assimilation system can significantly improve SST prediction accuracy, extend the effective forecast lead time, and enhance the robustness of the forecast system.

In addition, to thoroughly evaluate the effectiveness of the data assimilation framework, we conducted two separate 60-day forecast experiments. One using raw observational data directly input into the forecast model, and the other using initial fields generated through data assimilation as input into the forecast model. OISST data covers the entire globe and can be used to initialise forecasts. We then calculated and compared the RMSE and ACC of the two forecast results, with detailed results shown in Figure 7. As clearly shown in Figure 7, using the initial fields generated through data assimilation as model inputs significantly reduced the forecast RMSE compared to directly using raw observational data. On average, the RMSE for the 60-day forecast decreased by 10.4%. In particular, the reduction in RMSE was particularly pronounced in the first 15 days of the forecast. Meanwhile, the forecast ACC also showed a significant improvement trend. These results collectively and strongly indicate that the assimilation process effectively integrates background field information with observational data, not only making the initial field more reasonable but also significantly reducing forecast errors. The direct benefit is a notable extension of the forecast lead time. Even within the 60-day forecast period, the improvement in forecast performance remains evident. In summary, this comparative experiment fully demonstrates that the data assimilation framework we propose can significantly enhance the performance of the SST forecast model. The integration of this framework with the forecast model forms a stable forecast assimilation system, providing a reliable technical foundation for maintaining long-term high-stability SST forecasts.

4.3 DL-BEC Effectiveness Verification

The core objective of designing the DL-BEC is to efficiently complete the calculation of the B matrix, significantly reducing its generation time. At the same time, it strictly ensures that the generated B matrix complies with physical constraints and

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

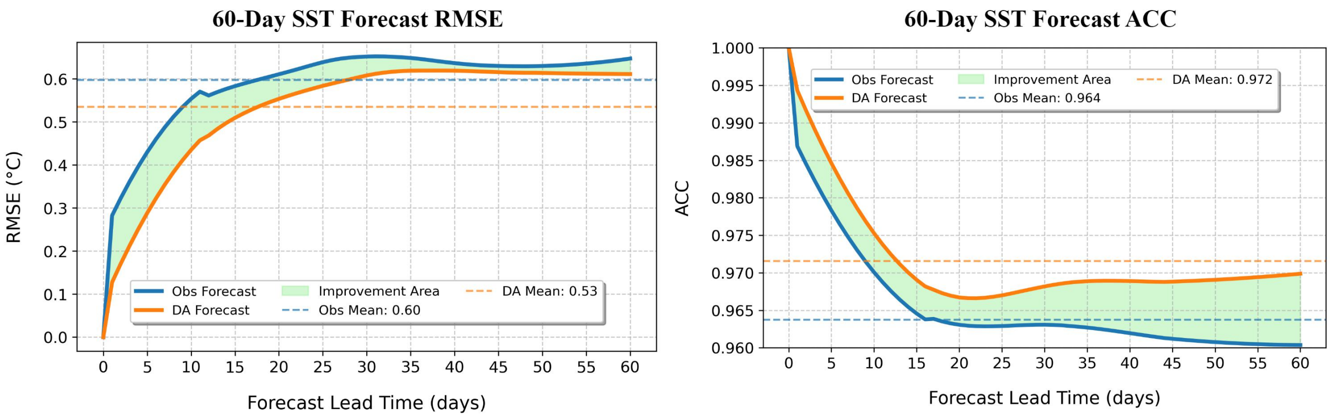

Figure 7. Comparison of RMSE and ACC for 60-day forecasts using observation data and assimilated background fields as inputs.

maintains its mathematical positive definiteness. To quantitatively evaluate the operational efficiency of the DL-BEC in practical applications, we have designed a rigorous comparative experimental plan. In the experiment, the initial ensemble size is set to 5,000 members. We conducted 360 complete forecast-assimilation cycle experiments using both the traditional EnKF assimilation scheme and the assimilation scheme integrated with the DL-BEC. Under the strict condition that computational resource configurations (CPU/GPU type, quantity, etc.) were held entirely consistent, precise measurements were taken of the time required for both approaches to generate the B matrix, the duration of a single assimilation run, and the total cumulative runtime. The comparison results of the runtime between the two experiments are detailed in Figure 8.

As shown in Figure 8, the introduction of the DL-BEC significantly reduces the computational time required for the entire assimilation process. Specifically, the computation time for B matrix was reduced by 99.6%, the time required for a single assimilation decreased by 14.2%, and the total time to complete all cyclic experiments was reduced by 7.2%. The assimilation process also involves other computational steps, such as the calculation of the Kalman gain matrix. These steps also incur a corresponding computational cost, which reduces the overall acceleration ratio. This result strongly demonstrates that the DL-BEC can significantly improve the computational efficiency of assimilation. The improved efficiency makes it possible to obtain high-precision initial field information in a shorter computational time. This not only ensures the timeliness of assimilation results but also directly enhances the real-time forecasting capability of the forecast model.

To validate the physical plausibility of the B matrix generated by the DL-BEC, we plotted its spatial distribution map for visual analysis. The B matrix directly calculated from 5,000 ensemble members was used as a high-precision benchmark. Based on this, we plotted and compared the distribution maps of the B matrix generated by the DL-BEC and the traditional direct calculation method for different set sizes (see Figure 9). By visually comparing the morphological and structural characteristics of these distribution maps, we aim to assess the DL-BEC's ability to maintain the physical properties of the B matrix under different ensemble sizes.

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

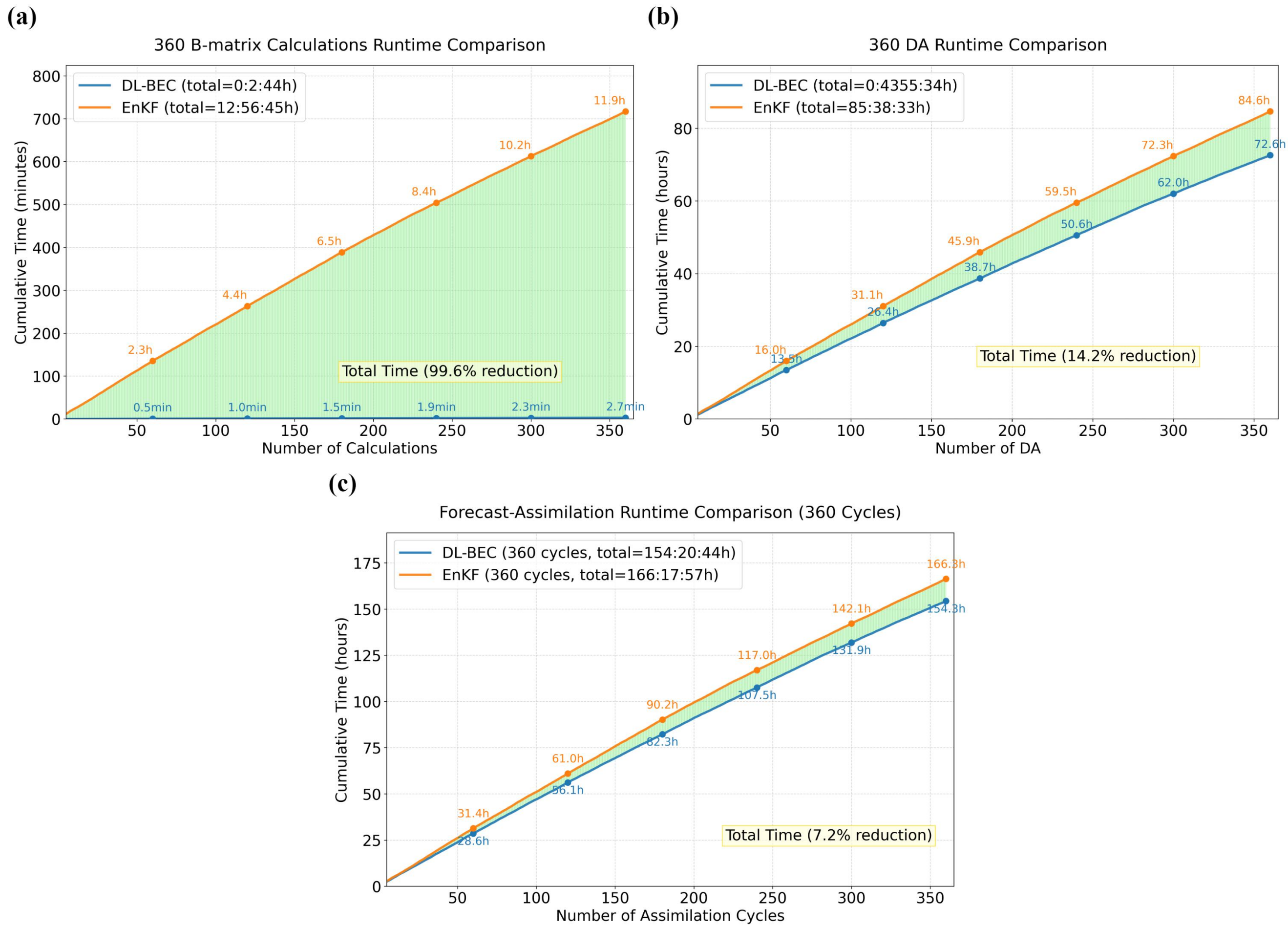

Figure 8. Comparison of experimental times before and after using the DL-BEC. Figure (a) compares the computational time for the two schemes' B matrix calculations; Figure (b) compares the computational time for the two schemes' assimilation processes; Figure (c) compares the total experimental time for the two schemes.

Under 28,800-dimensional data conditions, as the ensemble size continues to increase, the B matrix obtained using traditional computational methods gradually approaches the true state. When the ensemble size is small, the B matrix obtained using traditional computational methods contains significant spurious cross-correlations. As the ensemble size continues to increase, these spurious correlations gradually weaken, and the rank of the matrix also approaches the full-rank state. This phenomenon is fully consistent with the expectations of traditional EnKF theory. Therefore, when computational cost constraints typically only allow for smaller ensemble sizes, techniques such as covariance expansion and localization must be employed to enforce reasonable physical constraints on the B matrix. In contrast, the DL-BEC demonstrates significant advantages. When the ensemble size reaches 1,000, the B matrix simulated by the DL-BEC is already very close to the true state. As the ensemble members increases, the B matrix computed by DL-BEC increasingly approximates the true situation, thereby enhancing its

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

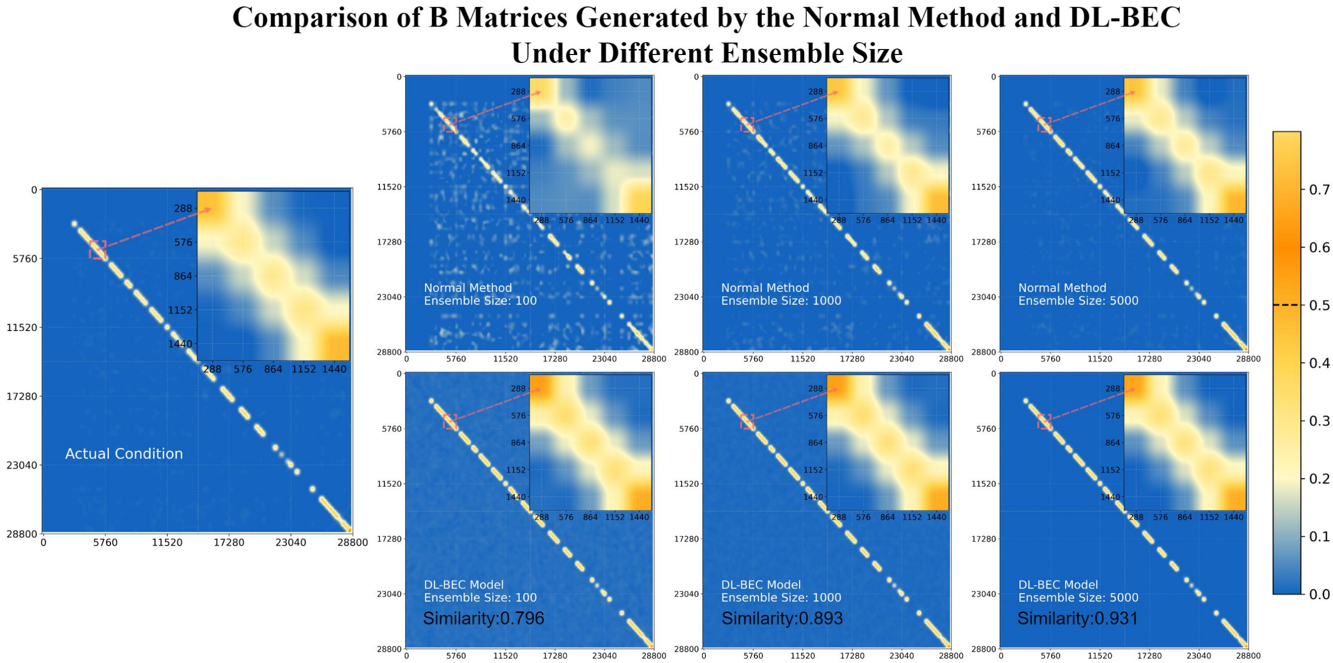

Figure 9. Comparison of B matrices generated by the normal direct calculation method and DL-BEC under different ensemble sizes. The left side shows the B matrix under ideal real condition, the top shows the B matrix obtained by traditional calculation, and the bottom shows the B matrix generated by DL-BEC, along with its similarity to the actual condition. From left to right, the ensemble sizes are 100, 1000, and 5000, respectively.

accuracy. This result indicates that under smaller ensemble size, the DL-BEC can more accurately characterize the background error covariance structure. This has important application value for improving the efficiency and accuracy of traditional numerical weather prediction systems. It provides new possibilities for more effectively assimilating high-dimensional observational data under computationally constrained scenarios.

Additionally, to validate the actual performance of DL-BEC, we designed a comparative experiment: we input the initial fields obtained through traditional computational methods and those generated by DL-BEC into the forecast model and conducted a 60-day forecast. In the experiment, the ensemble size for both methods was set to 5,000. The experimental results (see Figure 10) clearly show that, under the same ensemble size, the initial fields obtained using DL-BEC exhibit superior forecast performance. Specifically, since DL-BEC can effectively suppress the spurious correlations introduced by traditional methods, its forecast results have a significantly lower RMSE than the traditional EnKF during the initial stage of the forecast. In addition, ACC has also been significantly improved. After adopting DL-BEC, the RMSE was significantly reduced by 13.9% within the 15-day short-term forecast period, while the ACC was also improved by 1%. The experimental results strongly confirm that the DL-BEC significantly enhances the effectiveness of data assimilation by more accurately characterizing the covariance structure of background field errors. This improvement is directly reflected in subsequent forecast stages, particularly in the

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

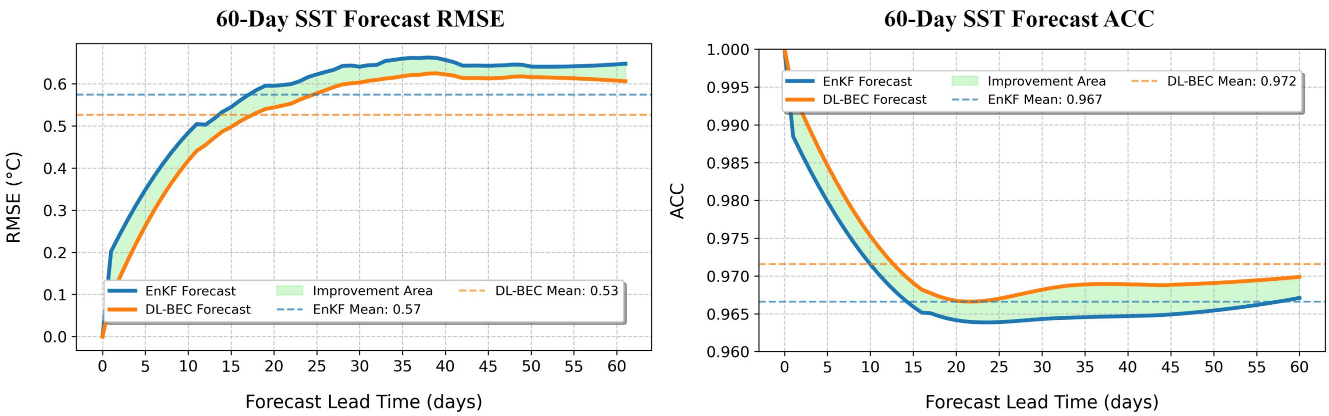

Figure 10. Comparison of RMSE and ACC for 60-day forecasts using the assimilation background fields obtained using DL-BEC and those calculated using traditional methods as inputs.

short-term forecast period, where it effectively reduces forecast errors. The DL-BEC plays a crucial role in optimizing the overall performance of the entire assimilation-forecasting system.

4.4 Comparison with Common Numerical Model Assimilation Methods

To validate the accuracy of our proposed assimilation methodology, this study conducted a series of comparative experiments by integrating our forecasting model with assimilation schemes. Critically, adopting the same forecast model ensures that any difference in performance can be attributed to the assimilation method itself, rather than to discrepancies in model physics, resolution, or surface forcing. We don't use freely available operational SST forecasts (e.g., from CMEMS), where the underlying forecasting systems differ substantially and would confound the assessment of the assimilation technique.

Specifically, we implemented a three-dimensional variational (3DVar) scheme akin to that employed by the NCEP operational Global Real-Time Ocean Forecasting System (RTOFS v2.0) (Mehra and Rivin, 2010) and the U.S. Navy's Global Ocean Forecasting System (GOFS 3.1) (Metzger et al., 2017). Following Parrish and Derber (1992), the 3DVar uses a static background-error covariance matrix B. In parallel, we implemented the SEEK filter, which serves as the assimilation engine for the GLORYS. Both assimilation schemes were coupled with the identical SST forecasting model proposed herein, and a 360-day assimilation cycle experiment was conducted. The relevant results are displayed in Figure 11.

Experimental results demonstrate that our proposed Deepcov-EnKF assimilation framework exhibits superior performance compared to current advanced numerical assimilation methods. During a 360-day cyclical forecasting experiment, this framework achieved significantly lower RMSE and markedly higher ACC. This advantage is primarily attributed to our use of a large number of ensemble members and a rapid yet accurate B-matrix estimation method. Further experiments demonstrate that DL-BEC possesses strong operational application potential, with prospects for effectively enhancing the assimilation efficiency and simulation performance of existing ocean numerical forecast models in the future.

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

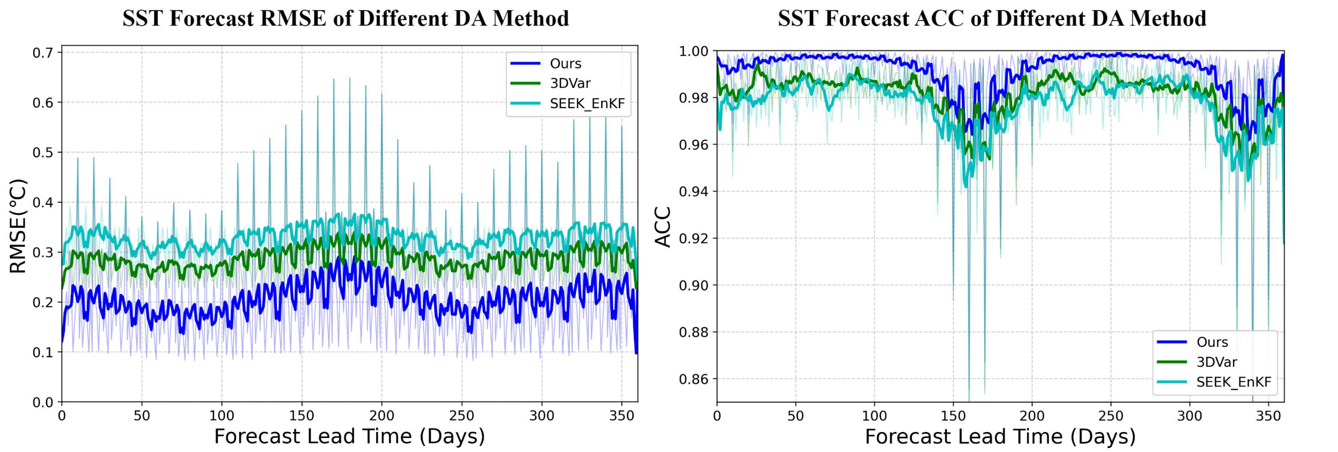

Figure 11. Comparison of DL-BEC with Current Operational Ocean Assimilation Methods

5 Conclusions

The integration of deep learning models into ocean variable forecasting, combined with data assimilation techniques, represents a prominent research frontier in marine science. In this study, we explore the potential of combining deep learning-based SST forecasting with the EnKF assimilation algorithm. To this end, we developed Deepcov-EnKF method that couples a rapid SST prediction model with an efficient assimilation system. At the core of this method are two main components: an advanced SST forecasting model and DL-BEC. The forecasting model employs an encoder-decoder structure to effectively learn complex features and achieve accurate SST estimation through mult-isource heterogeneous data fusion. On the assimilation side, the method integrates the traditional EnKF with DL-BEC, which uses block low-rank decomposition to handle ultra-high-dimensional data efficiently. Through Cholesky decomposition parameterization and eigenvalue regularization, the DL-BEC ensures that the resulting covariance matrix remains symmetric and positive definite, thus preserving physical validity. Using a 360-day of observational data, we comprehensively evaluated the model's forecasting skill and assimilation performance. Results indicate that Deepcov-EnKF maintains a low RMSE throughout a 60-day prediction horizon, demonstrating robust long-term stability. The forecast-assimilation cycle operates reliably, confirming the method's practicality. Moreover, Deepcov-EnKF performs consistently well across different ensemble sizes and marine regions, highlighting its strong generalization capability and robustness. Compared to current advanced numerical assimilation methods, Deepcov-EnKF demonstrates superior performance. This study provides a valuable theoretical and practical foundation for deeper integration of deep learning into marine forecasting and data assimilation, demonstrating significant potential to enhance both prediction accuracy and computational efficiency.

Our core approach involves generating ensemble forecasts from perturbed initial conditions and incorporating data assimilation to enhance predictive accuracy. The forecasting model was trained using high-quality reanalysis data to better represent ocean-atmosphere dynamics. During evaluation, daily model forecasts were systematically validated against corresponding

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

reanalysis fields. Results demonstrate that the forecasting model maintains robust performance over a 60-day forecast horizon, with the RMSE remaining as low as approximately 0.6°C and the ACC sustaining a high value above 0.9 even at day 60. To assess the assimilation contribution, forecast comparisons were conducted using initial conditions derived from both assimilated and non-assimilated data. The findings reveal that the assimilated initial fields substantially enhance forecast performance across both short and long time scales, reducing the overall RMSE by 10.4%. A key innovation introduced in this study is the DL-BEC, which enables efficient real-time computation of the error covariance matrix in massive ensemble members. The DL-BEC achieved a 99.6% reduction in computation time compared to conventional methods. Moreover, under identical ensemble sizes, it effectively suppresses spurious correlations, thereby significantly improving assimilation stability and accuracy. The results not only offer a practical tool for advancing ocean phenomenon studies but also strengthen the foundation for developing higher-resolution ocean prediction systems.

Under real-time observational conditions, the method proposed in this study is capable of providing long-term and stable predictions of SST distribution. This capability offers substantial scientific and practical value for applications such as maritime navigation safety planning, marine ecosystem monitoring and conservation, as well as climate change research. Furthermore, the global SST outputs generated by the method can serve as a key for predicting and analyzing major marine phenomena(such as global marine heatwaves), thereby offering reliable data support for oceanographic studies. The developed DL-BEC also exhibits strong transfer potential and can be adapted to EDA systems in other domains, such as meteorology. Additionally, this model establishes an important technical foundation for accelerating B matrix calculations and maintaining B matrix stability under conditions of limited ensemble members.

The proposed method holds considerable promise for broad application. The present spatial resolution of 1.5°×1.5° was adopted as a pragmatic choice under current computational resources. As computing capacity grows, the resolution can be raised within the same model framework to better identify and accurately depict mesoscale and small-scale ocean phenomena such as eddies and fronts, particularly in critical sea areas. Likewise, the current focus on a single variable (SST) serves to validate the feasibility of the approach. In principle, the model is inherently parallelizable and can be extended to assimilate multiple variables simultaneously. And its scale can be expanded to realize multivariate coupled assimilation, thereby enabling a full description of the three-dimensional, multi-variable ocean state. The memory consumption challenges posed by the block-storage strategy of the B matrix can be addressed through optimized storage schemes. The systematic overestimation of diagonal elements by the DL-BEC, which may unduly amplify correlations between adjacent grid points, can be corrected by further investigation into network architecture design and training strategies, leading to enhanced data assimilation quality. With these developments, the method will advance toward a more mature and practical stage, offering extensive application prospects.

The top priority is to significantly enhance the spatial resolution of the model and incorporate key ocean variables to achieve a more comprehensive and detailed characterization of ocean conditions. It's necessary to deeply optimize the core architecture of the forecast model and fine-tune the training hyperparameters to maximize model performance improvements. And a variable resolution strategy can be adopted, maintaining a global coarse-resolution grid while implementing high-resolution grid refinement only in key sea areas, forming a nested grid system. This approach can effectively capture mesoscale ocean

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

phenomena in key regions while significantly controlling overall computational complexity and training costs. Additionally, the training strategy of the DL-BEC can be optimized, exploring more effective loss functions to suppress issues such as overestimation of diagonal elements. A more reasonable network architecture and training hyperparameters can be carefully designed. A more optimal block strategy can be adopted for the storage and computation of the B matrix, balancing memory usage and computational efficiency. Through model architecture and training optimization, innovative variable-resolution nested grid technology, and DL-BEC algorithm improvements, we will strive to overcome current limitations, driving the model toward higher resolution, multi-variable, efficient, and stable directions, ultimately building a more powerful and practical ocean forecasting and assimilation system.

Code and data availability. The atmospheric reanalysis data set is available at https://doi.org/10.24381/cds.adbb2d47 (Hersbach et al., 2020) and the ocean reanalysis data set is available at https://doi.org/10.48670/moi-00021 (Jean-Michel et al., 2021). The ocean observation data set is available at https://www.ncei.noaa.gov/products/optimum-interpolation-sst (Huang et al., 2021). The code for this study is at https://doi.org/10.5281/zenodo.20993446.

Author contributions. Baoxu Li developed the DL-BEC code, performed the experiments, and wrote the paper. Hongze Leng and Junqiang Song supervised the theoretical aspects of data assimilation and deep learning. Wuxin Wang and Taikang Yuan carried out the data processing. Jinhui Yang and Xiang Wang provided the computational platforms and resources. Hang Cao created and optimised the figures. All authors contributed to the final version of the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors would like to thank Lilan Huang, Xingyu Chai, and Ruisheng Hu for providing valuable suggestions for this article. We thank the National Supercomputer Center in Tianjin and Beijing PARATERA Tech Corp for providing computing resources.

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

References

Abbas, M., Min, Z., Liu, Z., and Zhang, D.: Unravelling oceanic wave patterns: A comparative study of machine learning approaches for predicting significant wave height, Applied Ocean Research, 145, 103 919, 2024.

Ali, A., Fathalla, A., Salah, A., Bekhit, M., and Eldesouky, E.: Marine data prediction: an evaluation of machine learning, deep learning, and statistical predictive models, Computational Intelligence and Neuroscience, 2021, 8551 167, 2021.

Alonso-González, E., Aalstad, K., Pirk, N., Mazzolini, M., Treichler, D., Leclercq, P., Westermann, S., López-Moreno, J. I., and Gascoin, S.: Spatio-temporal information propagation using sparse observations in hyper-resolution ensemble-based snow data assimilation, Hydrology and Earth System Sciences, 27, 4637–4659, 2023.

Bai, G., Wang, Z., Zhu, X., and Feng, Y.: Development of a 2-D deep learning regional wave field forecast model based on convolutional neural network and the application in South China Sea, Applied Ocean Research, 118, 103 012, 2022.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, Quarterly Journal of the Royal Meteorological Society, 134, 1951–1970, 2008.

Beiser, F., Holm, H. H., Lye, K. O., and Eidsvik, J.: Multi-level data assimilation for ocean forecasting using the shallow-water equations, Journal of Computational Physics, 524, 113 722, 2025.

Bento, P., Pombo, J., Mendes, R., Calado, M., and Mariano, S.: Ocean wave energy forecasting using optimised deep learning neural networks, Ocean Engineering, 219, 108 372, 2021.

Blondel, E., Bonnefoy, F., and Ferrant, P.: Deterministic non-linear wave prediction using probe data, Ocean Engineering, 37, 913–926, 2010.

Chattopadhyay, A., Nabizadeh, E., Bach, E., and Hassanzadeh, P.: Deep learning-enhanced ensemble-based data assimilation for high-dimensional nonlinear dynamical systems, Journal of Computational Physics, 477, 111 918, 2023.

Chattopadhyay, A., Gray, M., Wu, T., Lowe, A. B., and He, R.: OceanNet: a principled neural operator-based digital twin for regional oceans, Scientific Reports, 14, 21 181, 2024.

Chen, C., Dou, Y., Chen, J., and Xue, Y.: A novel neural network training framework with data assimilation: C. Chen et al., The Journal of Supercomputing, 78, 19 020–19 045, 2022.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International conference on medical image computing and computer-assisted intervention, pp. 424–432, Springer, 2016.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Quarterly Journal of the Royal Meteorological Society, 120, 1367–1387, 1994.

Cummings, J. A.: Operational multivariate ocean data assimilation, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131, 3583–3604, 2005.

Danforth, C. M., Kalnay, E., and Miyoshi, T.: Estimating and correcting global weather model error, Monthly weather review, 135, 281–299, 2007.

Dong, R., Leng, H., Zhao, J., Song, J., and Liang, S.: A framework for four-dimensional variational data assimilation based on machine learning, Entropy, 24, 264, 2022.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, 1994.

Evensen, G., Vossepoel, F. C., and Van Leeuwen, P. J.: Iterative ensemble smoothers for data assimilation in coupled nonlinear multiscale models, Monthly Weather Review, 152, 1277–1301, 2024.

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

Fei, T., Huang, B., Wang, X., Zhu, J., Chen, Y., Wang, H., and Zhang, W.: A hybrid deep learning model for the bias correction of sst numerical forecast products using satellite data, Remote Sensing, 14, 1339, 2022.

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., et al.: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres, Quarterly Journal of the Royal Meteorological Society, 144, 1218–1256, 2018.

Hammoud, M. A. E. R., Raboudi, N., Titi, E. S., Knio, O., and Hoteit, I.: Data assimilation in chaotic systems using deep reinforcement learning, Journal of Advances in Modeling Earth Systems, 16, e2023MS004 178, 2024.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly journal of the royal meteorological society, 146, 1999–2049, 2020.

Hewitt, H. T., Bell, M. J., Chassignet, E. P., Czaja, A., Ferreira, D., Griffies, S. M., Hyder, P., McClean, J. L., New, A. L., and Roberts, M. J.: Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales?, Ocean Modelling, 120, 120–136, 2017.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural computation, 9, 1735–1780, 1997.

Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Monthly weather review, 126, 796–811, 1998.

Howard, L. J., Subramanian, A., and Hoteit, I.: A machine learning augmented data assimilation method for high-resolution observations, Journal of Advances in Modeling Earth Systems, 16, e2023MS003 774, 2024.

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., and Zhang, H.-M.: Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1, Journal of Climate, 34, 2923–2939, 2021.

Immas, A., Do, N., and Alam, M.-R.: Real-time in situ prediction of ocean currents, Ocean Engineering, 228, 108 922, 2021.

Jean-Michel, L., Eric, G., Romain, B.-B., Gilles, G., Angélique, M., Marie, D., Clément, B., Mathieu, H., Olivier, L. G., Charly, R., et al.: The Copernicus global 1/12 oceanic and sea ice GLORYS12 reanalysis, Frontiers in Earth Science, 9, 698 876, 2021.

Kang, X., Kokkinaki, A., Power, C., Kitanidis, P. K., Shi, X., Duan, L., Liu, T., and Wu, J.: Integrating deep learning-based data assimilation and hydrogeophysical data for improved monitoring of DNAPL source zones during remediation, Journal of Hydrology, 601, 126 655, 2021.

Kerry, C. G., Roughan, M., Keating, S., Gwyther, D., Brassington, G., Siripatana, A., and Souza, J. M. A.: Comparison of 4-dimensional variational and ensemble optimal interpolation data assimilation systems using a Regional Ocean Modeling System (v3. 4) configuration of the eddy-dominated East Australian Current system, Geoscientific Model Development, 17, 2359–2386, 2024.

Kim, J., Kim, T., Yoo, J., Ryu, J.-G., Do, K., and Kim, J.: STG-OceanWaveNet: Spatio-temporal geographic information guided ocean wave prediction network, Ocean Engineering, 257, 111 576, 2022.

Köllisch, N., Behrendt, J., Klein, M., and Hoffmann, N.: Nonlinear real time prediction of ocean surface waves, Ocean engineering, 157, 387–400, 2018.

Krasnopolsky, V. M., Chalikov, D. V., and Tolman, H. L.: A neural network technique to improve computational efficiency of numerical oceanic models, Ocean Modelling, 4, 363–383, 2002.

Labahn, J. W., Wu, H., Harris, S. R., Coriton, B., Frank, J. H., and Ihme, M.: Ensemble Kalman filter for assimilating experimental data into large-eddy simulations of turbulent flows, Flow, Turbulence and Combustion, 104, 861–893, 2020.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86, 2278–2324, 1998.

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

Ling, F., Ouyang, L., Larbi, B. R., Luo, J.-J., Han, T., Zhong, X., and Bai, L.: Improving global weather and ocean wave forecast with large artificial intelligence models, Science China Earth Sciences, 67, 3641–3654, 2024.

Lu, X., Peng, Z., Li, C., Chen, L., Qiao, G., Li, C., Yang, B., and He, Q.: An innovative deep learning model for accurate wave height predictions with enhanced performance for extreme waves, Ocean Engineering, 322, 120 502, 2025.

Martin, M., Hines, A., and Bell, M.: Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133, 981–995, 2007.

Mehra, A. and Rivin, I.: A real time ocean forecast system for the North Atlantic Ocean, TAO: Terrestrial, Atmospheric and Oceanic Sciences, 21, 2, 2010.

Metzger, E. J., Helber, R. W., Hogan, P. J., Posey, P. G., Thoppil, P. G., Townsend, T. L., Wallcraft, A. J., Smedstad, O. M., Franklin, D. S., Zamudo-Lopez, L., and Phelps, M. W.: Global Ocean Forecast System 3.1 Validation Test, https://api.semanticscholar.org/CorpusID:134962358, 2017.

Moore, A. M., Martin, M. J., Akella, S., Arango, H. G., Balmaseda, M., Bertino, L., Ciavatta, S., Cornuelle, B., Cummings, J., Frolov, S., et al.: Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems: A more complete picture of the state of the ocean, Frontiers in Marine Science, 6, 90, 2019.

Mu, M., Duan, W., Chen, D., and Yu, W.: Target observations for improving initialization of high-impact ocean-atmospheric environmental events forecasting, National Science Review, 2, 226–236, 2015.

Mücke, N. T., Bohté, S. M., and Oosterlee, C. W.: The deep latent space particle filter for real-time data assimilation with uncertainty quantification, Scientific Reports, 14, 19 447, 2024.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's spectral statistical-interpolation analysis system, Monthly Weather Review, 120, 1747–1763, 1992.

Pham, D. T., Verron, J., and Roubaud, M. C.: A singular evolutive extended Kalman filter for data assimilation in oceanography, Journal of Marine systems, 16, 323–340, 1998.

Reynolds, R. W.: Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperatures, Journal of climate, 6, 768–774, 1993.

Shao, Q., Hou, G., Li, W., Han, G., Liang, K., and Bai, Y.: Ocean reanalysis data-driven deep learning forecast for sea surface multivariate in the South China Sea, Earth and Space Science, 8, e2020EA001 558, 2021.

Shao, Q., Zhao, Y., Li, W., Han, G., Hou, G., Li, C., Liu, S., Gong, Y., Liu, H., and Qu, P.: A simple statistical intra-seasonal prediction model for sea surface variables utilizing satellite remote sensing, Remote Sensing, 14, 1162, 2022.

Smit, P., Houghton, I., Jordanova, K., Portwood, T., Shapiro, E., Clark, D., Sosa, M., and Janssen, T.: Assimilation of significant wave height from distributed ocean wave sensors, Ocean Modelling, 159, 101 738, 2021.

Sun, H.-X., Wang, S., Zheng, X., and Chen, S. X.: High-dimensional ensemble Kalman filter with localization, inflation, and iterative updates, Quarterly Journal of the Royal Meteorological Society, 150, 4870–4884, 2024.

Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M., and Zhen, Y.: A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation, Monthly Weather Review, 148, 3973–3994, 2020.

Toste, R., Böck, C. S., da Silva, M. S., Moraes, N. O., Soares, A. E., Nehme, D. M., de Freitas Assad, L. P., Landau, L., Barreto, F., and da Silva Júnior, C. L.: CODAR data assimilation into an integrated ocean forecasting system for the Brazilian Southeastern coast, Ocean Modelling, 188, 102 331, 2024.

https://doi.org/10.5194/egusphere-2026-3488
Preprint. Discussion started: 7 September 2026

Toth, Z. and Kalnay, E.: Ensemble forecasting at NMC: The generation of perturbations, Bulletin of the american meteorological society, 74, 2317–2330, 1993.

Vlachas, P.-R., Pathak, J., Hunt, B. R., Sapsis, T. P., Girvan, M., Ott, E., and Koumoutsakos, P.: Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics, Neural Networks, 126, 191–217, 2020.

Wang, X., Wang, R., Hu, N., Wang, P., Huo, P., Wang, G., Wang, H., Wang, S., Zhu, J., Xu, J., et al.: Xihe: A data-driven model for global ocean eddy-resolving forecasting, arXiv preprint arXiv:2402.02995, 2024.

Wang, Y., Yuan, Z., Xie, C., and Wang, J.: Temporally sparse data assimilation for the small-scale reconstruction of turbulence, Physics of Fluids, 34, 2022.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., and Toth, Z.: Ensemble data assimilation with the NCEP global forecast system, Monthly Weather Review, 136, 463–482, 2008.

Xiong, W., Xiang, Y., Wu, H., Zhou, S., Sun, Y., Ma, M., and Huang, X.: Ai-goms: Large ai-driven global ocean modeling system, arXiv preprint arXiv:2308.03152, 2023.

Xu, L., Chen, N., Chen, Z., Zhang, C., and Yu, H.: Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions, Earth-Science Reviews, 222, 103 828, 2021.

Zhang, J., Luo, F., Quan, X., Wang, Y., Shi, J., Shen, C., and Zhang, C.: Improving wave height prediction accuracy with deep learning, Ocean Modelling, 188, 102 312, 2024.

Zhang, L., Sidoti, D., Bienkowski, A., Pattipati, K. R., Bar-Shalom, Y., and Kleinman, D. L.: On the identification of noise covariances and adaptive Kalman filtering: A new look at a 50 year-old problem, IEEE Access, 8, 59 362–59 388, 2020.