A BATHYMETRY-CONDITIONED CNN–BIGRU FRAMEWORK FOR WAVE HINDCASTING WITH SLIDING-WINDOW EVALUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate long-horizon wave forecasting remains difficult in coastal regions where irregular bathymetry produces localized and nonlinear wave amplification. Existing approaches use physics-based models, ConvLSTM, transformers, and multiscale spatial encoders to learn ocean dynamics. However, increasing encoder complexity does not necessarily improve bathymetry-sensitive prediction. Direct multi-step models can lose spatial coherence, while recursive models can accumulate forecast drift. We propose a bathymetry-conditioned CNN-BiGRU framework with internally corrected recursive inference. The CNN jointly encodes dynamic ocean fields and an aligned bathymetric field. The BiGRU captures temporal dependencies only within the observed historical sequence. At each recursive step, the network predicts three consecutive hourly states. The first state updates the next input window. The remaining states are used to estimate the internal forecast inconsistency. The subsequent deviations between these states determine a correction factor to regulate the propagated prediction. This mechanism uses the model's own multi-step outputs to control recursive instability without future observations or external error feedback. Visual analysis indicates that the framework preserves spatial organization and follows localized amplification during extreme conditions. The experiments used 35,017 hourly CMEMS IBI samples with GEBCO-derived bathymetry. The test sequence contains a maximum crest-to-trough wave height of 17.18 m. Over a 168-hour horizon, the proposed framework achieves an MAE of 0.1375 m and an RMSE of 0.1821 m. Under the same evaluation protocol, a multi-resolution spatio-temporal baseline obtains an MAE of 0.8503 m and an RMSE of 0.9989 m, demonstrating the effectiveness of the proposed framework for extended-horizon wave forecasting.

Wave Hindcasting, Wave Forecasting, Bathymetry, CNN, Bi-GRU, Sliding window, CMEMS IBI, GEBCO

1 INTRODUCTION

Accurate ocean-wave forecasting supports coastal safety, offshore operations, and disaster preparedness. Numerical models resolve wave generation, propagation, and dissipation from physical equations. However, their cost increases at high spatial resolutions and long forecast horizons. Physics-informed learning can reduce this burden while retaining physical structure Verma et al. (2024). Machine-learning models have also reproduced high-resolution coastal wave fields with accuracy comparable to numerical simulations Shen et al. (2024). Deep learning has further shown competitive performance against operational wave models Zhang et al. (2025). Yet, long-horizon forecasting remains difficult over complex bathymetry. Abrupt depth changes induce refraction, reflection, and localized amplification. These interactions create nonuniform wave fields and increase uncertainty during extreme events.

Bathymetric information can improve high-resolution wave-height reconstruction (Kordatos et al., 2024). However, existing methods often predict a single variable and do not capture the joint long-term evolution of multiple wave fields. Recursive forecasting creates a further challenge because predicted states are repeatedly returned to the input. Local errors can therefore accumulate and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

destabilize the forecast. Structured correction has improved autoregressive regional ocean prediction (Chattopadhyay et al., 2024). Yet, bathymetry-conditioned multivariate representation and recursive error control remain disconnected. A unified framework is therefore required to learn seabed-dependent dynamics while preserving long-horizon forecast consistency.

Learning-based environmental forecasting has advanced from regional recurrent models to high-resolution, multi-scale predictors. Multivariate GRU and LSTM models capture coupled wind–wave dynamics in regional seas (Alizadeh & Nourani, 2024). CNN–BiGRU architectures additionally combine spatial encoding with bidirectional temporal learning (Thet et al., 2025). Frequency-domain representations distinguish wind-sea and swell dynamics (Lv et al., 2026), while temporal-change regularization improves recursive stability (Zhang et al., 2026). Physically consistent emulation and generative downscaling support regional prediction across spatial resolutions (Lupin-Jimenez et al., 2025). Graph models further accommodate irregular ocean geometry and atmospheric forcing (Holmberg et al., 2025). FuXi-Ocean extends learning-based forecasting to sub-daily, eddy-resolving global prediction (Niu et al., 2026). Atmospheric models have also introduced geometry-aware Transformers (Liu et al., 2025) and adaptive rollout intervals (Tian et al., 2026). STORM separates scale-specific dependencies while maintaining coherent cross-resolution evolution (Huang et al., 2026). These advances identify multi-scale representation, geometric awareness, physical consistency, and rollout stability as key forecasting requirements.

Despite this progress, key requirements remain disconnected in long-horizon coastal-wave forecasting. Recurrent and hybrid models mainly target significant wave height or short- to medium-range prediction (Alizadeh & Nourani, 2024; Thet et al., 2025; Lv et al., 2026). Their ability to preserve coupled multivariate dynamics during extended recursion is therefore unclear. Temporal regularization improves stability but relies on reference-based fine-tuning and does not infer reliability from generated states (Zhang et al., 2026). Regional emulators handle complex geometry, multiple resolutions, or atmospheric forcing (Lupin-Jimenez et al., 2025; Holmberg et al., 2025; Niu et al., 2026), but primarily predict circulation fields. Geometry-aware and adaptive-rollout methods address global structure and accumulated atmospheric error (Liu et al., 2025; Tian et al., 2026), not localized seabed transitions. STORM learns scale-specific dependencies and coherent cross-resolution evolution (Huang et al., 2026). However, it neither conditions wave dynamics on high-resolution bathymetry nor uses agreement among simultaneous predictions to regulate recursion. A unified framework is therefore needed to couple seabed geometry with multivariate wave dynamics and control long-horizon error without future observations.

To address these gaps, a bathymetry-conditioned CNN–Bi-ConvGRU is proposed for long-horizon multivariate wave forecasting. Dynamic ocean fields, high-resolution bathymetry, and an ocean mask are jointly processed. The CNN extracts local patterns associated with seabed gradients and land–ocean boundaries. The stacked Bi-ConvGRU then models their temporal evolution and generates multiple future states per forward pass. Channel-wise variation among these outputs provides an internal consistency measure. This variation is mapped to a bounded coefficient that regulates the predicted increment before recursive window updating. The procedure requires no future observations, auxiliary correction network, or inference-time retraining. The methodological novelty lies in coupling bathymetry-aware representation with prediction-derived recursive regulation. The same outputs provide both future states and the evidence used to control their propagation. Results show that the framework follows rapid transitions and high-energy events while remaining stable over extended rollouts. The principal contributions are summarized below.

- A bathymetry-conditioned CNN–Bi-ConvGRU is developed to jointly encode multivariate ocean states, high-resolution seabed geometry, and land–ocean boundaries.
- A consistency-guided recursion mechanism is introduced. It converts channel-wise variation among direct predictions into a bounded increment correction. No future observations, auxiliary network, or inference-time retraining are required.
- A leakage-free long-horizon protocol is established using only model-generated window updates. The results demonstrate lower errors, improved recovery of localized wave evolution, and greater recursive stability than the compared methods.

The remainder of this paper is organized as follows. Section 2 formulates the forecasting problem. Section 3 presents the proposed framework. Section 4 reports and discusses the results. Section 5 concludes the paper and outlines future work.

2 System Model and Problem Formulation

2.1 System Model

A coastal region is represented by an $H \times W$ grid, where H and W are its numbers of rows and columns. The indices $i \in \{1, \ldots, H\}$, $j \in \{1, \ldots, W\}$, and t denote a grid cell and an hourly time step, respectively. The physical ocean state is $\mathbf{X}_t^{\text{phy}} \in \mathbb{R}^{C_d \times H \times W}$, where C_d is the number of dynamic variables. Its c-th channel is $\mathbf{X}_t^{\text{phy},(c)}$, where $c \in \{1, \ldots, C_d\}$. VCMX is represented by channel c_v.

The bathymetric field is $\mathbf{D}^{\text{phy}} \in \mathbb{R}^{H \times W}$, where $D_{i,j}^{\text{phy}}$ is the seabed depth at cell (i, j). The ocean mask is $\mathbf{M} \in \{0, 1\}^{H \times W}$, where $M_{i,j} = 1$ denotes ocean and $M_{i,j} = 0$ denotes land. Both fields are time invariant. Training-partition statistics are used to normalize $\mathbf{X}_t^{\text{phy}}$ and $\mathbf{D}^{\text{phy}}$, producing $\mathbf{X}_t$ and $\mathbf{D}$. The complete state is

$$\mathbf{Z}_t = \text{Concat}(\mathbf{X}_t, \mathbf{D}, \mathbf{M}) \in \mathbb{R}^{(C_d+2) \times H \times W}, \tag{1}$$

where $\text{Concat}(\cdot)$ denotes channel-wise concatenation. Bathymetry represents seabed-dependent wave modulation, while $\mathbf{M}$ identifies valid ocean cells. Neither static field is predicted.

Let L be the observed sequence length. The input history is $\mathcal{H}_t = \{\mathbf{Z}_{t-L+1}, \ldots, \mathbf{Z}_t\}$. Let K be the direct prediction horizon. A forecasting model $f_{\boldsymbol{\theta}}$, parameterized by the trainable vector $\boldsymbol{\theta}$, maps this history to

$$\widehat{\mathcal{Y}}_t = f_{\boldsymbol{\theta}}(\mathcal{H}_t) = \left\{\widehat{\mathbf{X}}_{t+1}, \ldots, \widehat{\mathbf{X}}_{t+K}\right\}, \tag{2}$$

where $\widehat{\mathcal{Y}}_t$ is the predicted sequence. For $k \in \{1, \ldots, K\}$, $\widehat{\mathbf{X}}_{t+k} \in \mathbb{R}^{C_d \times H \times W}$ is the normalized prediction, $\mathbf{X}_{t+k}$ is its reference, and the hat denotes estimation. Since only states up to time t are included, the mapping is causal.

2.2 Problem Formulation

Let $H_f > K$ be the required forecast horizon and $r \in \{0, \ldots, H_f - 1\}$ the recursive index. To extend a K-step model, the first prediction is returned to the input:

$$\widehat{\mathbf{Z}}_{t+r+1} = \text{Concat}\left(\widehat{\mathbf{X}}_{t+r+1}, \mathbf{D}, \mathbf{M}\right), \qquad \mathcal{H}_{t+r+1} = \text{Shift}\left(\mathcal{H}_{t+r}, \widehat{\mathbf{Z}}_{t+r+1}\right). \tag{3}$$

Here, $\widehat{\mathbf{Z}}_{t+r+1}$ is the complete predicted state. The operator $\text{Shift}(\mathcal{H}, \mathbf{Z})$ removes the oldest state from the L-step sequence $\mathcal{H}$ and appends $\mathbf{Z}$. The sequence length therefore remains L. At $r = 0$, all inputs are observed. Later windows may contain predicted states.

Let $\mathcal{T}_{\text{tr}}$ be the set of valid training origins and $N_{\text{tr}} = |\mathcal{T}_{\text{tr}}|$ its size. Each origin provides an L-step input and K references. The number of ocean cells is $N_\Omega = \sum_{i=1}^{H} \sum_{j=1}^{W} M_{i,j}$. Training solves

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}} \frac{1}{N_{\text{tr}} K C_d N_\Omega} \sum_{t \in \mathcal{T}_{\text{tr}}} \sum_{k=1}^{K} \left\| \mathbf{M} \odot \left(\widehat{\mathbf{X}}_{t+k} - \mathbf{X}_{t+k}\right) \right\|_F^2, \tag{4}$$

where $\boldsymbol{\theta}^\star$ is the optimized parameter vector, $\odot$ denotes element-wise multiplication, and $\|\cdot\|_F$ is the Frobenius norm. The mask is broadcast across all C_d channels. The denominator averages the squared error over training origins, forecast steps, variables, and ocean cells. Land cells are excluded.

This objective controls error only within the direct horizon K. Beyond K, each prediction enters the next window through equation 3. Its error can therefore propagate and distort wave magnitude, spatial structure, and extreme VCMX evolution. The required model must capture bathymetry-dependent variation, remain causal, and limit recursive error without future observations. VCMX is evaluated separately because all dynamic channels receive equal training weight. All metrics are computed after predictions and references are restored to physical units.

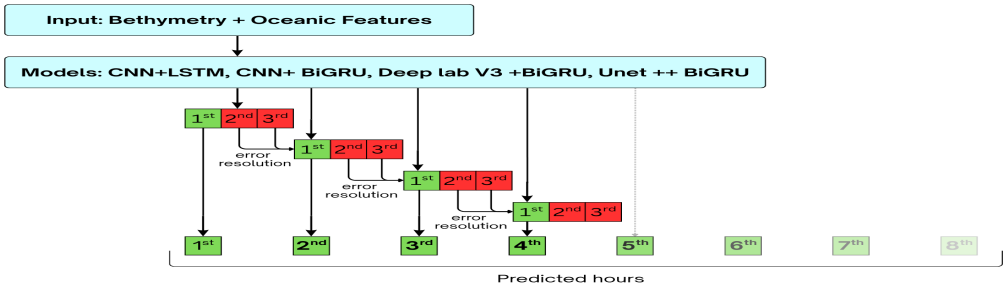

Figure 1: The proposed framework

3 Proposed Method

The proposed framework addresses bathymetry-dependent representation and recursive error propagation as shown in fig. 1. Given $\mathcal{H}_t = \{\mathbf{Z}_{t-L+1}, \ldots, \mathbf{Z}_t\}$, it first encodes each observed state. The encoded sequence is then processed temporally. A decoder generates K future states. Their internal consistency controls the state returned to the input window. This process continues until horizon H_f is reached.

3.1 Bathymetry-Conditioned Spatial Encoding

For each $\tau \in \{t - L + 1, \ldots, t\}$, where τ indexes the observation window, the encoder receives $\mathbf{Z}_\tau \in \mathbb{R}^{(C_d+2)\times H\times W}$. It contains C_d dynamic channels, bathymetry, and the ocean mask. The encoder parameters are shared across all L states.

Let $E_{\boldsymbol{\phi}}$ be the spatial encoder with trainable parameters $\boldsymbol{\phi}$. Its output is

$$\mathbf{F}_\tau = E_{\boldsymbol{\phi}}(\mathbf{Z}_\tau) = \sigma_e(\mathrm{Conv}_2[\sigma_e(\mathrm{Conv}_1[\mathbf{Z}_\tau])]), \tag{5}$$

where Conv_1 and Conv_2 are two-dimensional convolutions and $\sigma_e(\cdot)$ is the ReLU activation. Their intermediate and final outputs are $\mathbf{F}_\tau^{(1)} \in \mathbb{R}^{C_1\times H_1\times W_1}$ and $\mathbf{F}_\tau \in \mathbb{R}^{C_e\times H_e\times W_e}$, respectively. Here, C_1, H_1, W_1 and C_e, H_e, W_e denote their channel and spatial dimensions. The layers use 64 and 32 filters, respectively. Hence, $C_1 = 64$ and $C_e = 32$. Both layers use 3×3 kernels, unit stride, and same padding. Therefore, $H_1 = H_e = H$ and $W_1 = W_e = W$. Thus, $H_1 = H_e = H$ and $W_1 = W_e = W$. Joint encoding makes the spatial features dependent on the ocean state, seabed geometry, and land–ocean boundary. The resulting sequence is $\mathcal{F}_t = \{\mathbf{F}_{t-L+1}, \ldots, \mathbf{F}_t\}$.

3.2 Bidirectional Temporal Modelling

Let $R_{\boldsymbol{\gamma}}$ denote the temporal module, where $\boldsymbol{\gamma}$ is its trainable parameter set. It processes $\mathcal{F}_t$ in both directions within the observed window. The backward branch reverses only known features and accesses no state after t. Let $\ell \in \{1, \ldots, L_r\}$ index the L_r recurrent layers, and let $d \in \{\rightarrow, \leftarrow\}$ denote direction. The layer input is $\mathbf{P}_\tau^{(1)} = \mathbf{F}_\tau$. For $\ell > 1$, $\mathbf{P}_\tau^{(\ell)}$ concatenates the directional outputs of layer $\ell - 1$. Define $\pi_\rightarrow(\tau) = \tau - 1$ and $\pi_\leftarrow(\tau) = \tau + 1$. Each directional ConvGRU computes

$$\begin{aligned}
\mathbf{r}_\tau^{(\ell,d)} &= \sigma_g\Big(\mathbf{W}_{xr}^{(\ell,d)} * \mathbf{P}_\tau^{(\ell)} + \mathbf{W}_{hr}^{(\ell,d)} * \mathbf{H}_{\pi_d(\tau)}^{(\ell,d)} + \mathbf{b}_r^{(\ell,d)}\Big), \\
\mathbf{u}_\tau^{(\ell,d)} &= \sigma_g\Big(\mathbf{W}_{xu}^{(\ell,d)} * \mathbf{P}_\tau^{(\ell)} + \mathbf{W}_{hu}^{(\ell,d)} * \mathbf{H}_{\pi_d(\tau)}^{(\ell,d)} + \mathbf{b}_u^{(\ell,d)}\Big), \\
\overline{\mathbf{H}}_\tau^{(\ell,d)} &= \tanh\Big(\mathbf{W}_{xh}^{(\ell,d)} * \mathbf{P}_\tau^{(\ell)} + \mathbf{r}_\tau^{(\ell,d)} \odot \Big(\mathbf{W}_{hh}^{(\ell,d)} * \mathbf{H}_{\pi_d(\tau)}^{(\ell,d)}\Big) + \mathbf{b}_h^{(\ell,d)}\Big), \\
\mathbf{H}_\tau^{(\ell,d)} &= \Big(1 - \mathbf{u}_\tau^{(\ell,d)}\Big) \odot \overline{\mathbf{H}}_\tau^{(\ell,d)} + \mathbf{u}_\tau^{(\ell,d)} \odot \mathbf{H}_{\pi_d(\tau)}^{(\ell,d)}.
\end{aligned} \tag{6}$$

Here, $\mathbf{r}_\tau^{(\ell,d)}$ and $\mathbf{u}_\tau^{(\ell,d)}$ are the reset and update gates. The tensors $\overline{\mathbf{H}}_\tau^{(\ell,d)}$ and $\mathbf{H}_\tau^{(\ell,d)}$ are the candidate and updated hidden states. The operators $*$ and $\odot$ denote two-dimensional convolution and element-

wise multiplication. The functions $\sigma_g(\cdot)$ and $\tanh(\cdot)$ are the sigmoid and hyperbolic-tangent activations. For $q \in \{r, u, h\}$, $\mathbf{W}_{xq}^{(\ell,d)}$, $\mathbf{W}_{hq}^{(\ell,d)}$, and $\mathbf{b}_q^{(\ell,d)}$ denote the input kernels, recurrent kernels, and biases. Boundary hidden states are initialized to zero.

The module contains $L_r = 2$ bidirectional ConvGRU layers with 64 and 32 hidden channels. All recurrent convolutions use 3×3 kernels, unit stride, and same padding. The terminal directional states are fused as

$$\mathbf{V}_t = R_{\boldsymbol{\gamma}}(\mathcal{F}_t) = \mathrm{Concat}\Big(\mathbf{H}_t^{(L_r,\rightarrow)}, \mathbf{H}_{t-L+1}^{(L_r,\leftarrow)}\Big), \tag{7}$$

where $\mathbf{V}_t \in \mathbb{R}^{64\times H\times W}$ is the final bidirectional representation. Its 64 channels are obtained by concatenating the 32 terminal channels from each direction.

3.3 Multi-Step Wave-Field Prediction

Let $G_{\boldsymbol{\psi}}$ denote the decoder, where $\boldsymbol{\psi}$ is its trainable parameter set. It maps the bidirectional representation $\mathbf{V}_t \in \mathbb{R}^{64\times H\times W}$ to

$$\widehat{\mathcal{Y}}_t = G_{\boldsymbol{\psi}}(\mathbf{V}_t) = \mathrm{Reshape}[\mathrm{Conv}_{1\times1}(\mathbf{V}_t)] \in \mathbb{R}^{K\times C_d\times H\times W}. \tag{8}$$

Here, $\mathrm{Conv}_{1\times1}$ denotes a pointwise convolution that produces KC_d channels. The operator $\mathrm{Reshape}(\cdot)$ separates these channels into K forecast steps and C_d dynamic variables. Thus, $\widehat{\mathcal{Y}}_t = \{\widehat{\mathbf{X}}_{t+1}, \ldots, \widehat{\mathbf{X}}_{t+K}\}$, where $\widehat{\mathbf{X}}_{t+k} \in \mathbb{R}^{C_d\times H\times W}$ is the prediction at direct lead $k \in \{1, \ldots, K\}$. All predictions share $\mathbf{V}_t$. The first output advances the recursive window. The remaining outputs characterize immediate evolution and support consistency estimation.

3.4 Consistency-Guided Recursive Forecasting

At recursive step $r \in \{0, \ldots, H_f - 1\}$, the current window $\mathcal{H}_{t+r}$ produces K direct predictions. For variable $c \in \{1, \ldots, C_d\}$, the relative change between direct leads k and $k+1$ is

$$\delta_{t+r,k}^{(c)} = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} M_{i,j}\left|\widehat{X}_{t+r+k+1,i,j}^{(c)} - \widehat{X}_{t+r+k,i,j}^{(c)}\right|}{\sum_{i=1}^{H}\sum_{j=1}^{W} M_{i,j}\left|\widehat{X}_{t+r+k,i,j}^{(c)}\right| + \epsilon}, \qquad k = 1, \ldots, K-1. \tag{9}$$

Here, $\widehat{X}_{t+r+k,i,j}^{(c)}$ is the predicted value of variable c at cell (i, j) and direct lead k. The mask $M_{i,j}$ excludes land cells. The constant $\epsilon = 10^{-8}$ prevents division by zero. Hence, $\delta_{t+r,k}^{(c)}$ is dimensionless.

For $K \geq 3$, the channel-wise inconsistency score is

$$d_{t+r}^{(c)} = \frac{1}{K-2}\sum_{k=1}^{K-2}\left|\delta_{t+r,k+1}^{(c)} - \delta_{t+r,k}^{(c)}\right|. \tag{10}$$

For $K = 3$, this reduces to $d_{t+r}^{(c)} = \left|\delta_{t+r,2}^{(c)} - \delta_{t+r,1}^{(c)}\right|$. A smaller score indicates more uniform short-term evolution. The score is mapped to

$$\alpha_{t+r}^{(c)} = \mathrm{clip}\Big(1 - \lambda d_{t+r}^{(c)}, \alpha_{\min}, 1\Big), \tag{11}$$

where $\lambda = 0.5$ controls the correction strength, $\alpha_{\min} = 0.1$ is its lower bound, and $\mathrm{clip}(x, a, b) = \min\{\max\{x, a\}, b\}$.

Let $\mathbf{X}_{t+r}^{\mathrm{in}}$ denote the latest dynamic state in $\mathcal{H}_{t+r}$. At $r = 0$, $\mathbf{X}_t^{\mathrm{in}} = \mathbf{X}_t$ is observed. For $r \geq 1$, $\mathbf{X}_{t+r}^{\mathrm{in}} = \widetilde{\mathbf{X}}_{t+r}$ is the preceding corrected state. Each channel is propagated as

$$\widetilde{\mathbf{X}}_{t+r+1}^{(c)} = \mathbf{X}_{t+r}^{\mathrm{in},(c)} + \alpha_{t+r}^{(c)}\left(\widehat{\mathbf{X}}_{t+r+1}^{(c)} - \mathbf{X}_{t+r}^{\mathrm{in},(c)}\right), \tag{12}$$

where the tilde denotes correction. Equation equation 12 regulates only the predicted increment. The corrected channels form $\widetilde{\mathbf{X}}_{t+r+1}$. The next complete state and input window are

$$\widetilde{\mathbf{Z}}_{t+r+1} = \mathrm{Concat}\Big(\widetilde{\mathbf{X}}_{t+r+1}, \mathbf{D}, \mathbf{M}\Big), \qquad \mathcal{H}_{t+r+1} = \mathrm{Shift}\Big(\mathcal{H}_{t+r}, \widetilde{\mathbf{Z}}_{t+r+1}\Big). \tag{13}$$

Only $\widetilde{\mathbf{X}}_{t+r+1}$ advances the window. The remaining $K-1$ predictions are used only for consistency estimation. No future reference is required.

3.5 TRAINING PROCEDURE

The trainable parameters are $\boldsymbol{\theta} = \{\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\psi}\}$, corresponding to the encoder, temporal module, and decoder. The data are divided chronologically into training from January 2020 to December 2022, validation from January to June 2023, and testing from July to December 2023. Their valid forecast-origin sets are denoted by $\mathcal{T}_{\mathrm{tr}}$, $\mathcal{T}_{\mathrm{val}}$, and $\mathcal{T}_{\mathrm{ev}}$. Each dynamic variable is scaled to $[0, 1]$ using training-set extrema. Bathymetry is scaled separately, while the ocean mask remains binary. The same transformations are applied to all partitions. No input–target window crosses a partition boundary.

Each sample contains $L = 48$ historical states and $K = 3$ future targets. The grid dimensions are $H = 72$ and $W = 90$, with $C_d = 8$ dynamic and two static channels. For a mini-batch $\mathcal{B} \subset \mathcal{T}_{\mathrm{tr}}$, the input and target dimensions are $|\mathcal{B}| \times L \times (C_d + 2) \times H \times W$ and $|\mathcal{B}| \times K \times C_d \times H \times W$, respectively. For forecast origin t, the model predicts $\widehat{\mathcal{Y}}_t = f_{\boldsymbol{\theta}}(\mathcal{H}_t)$, where $f_{\boldsymbol{\theta}} = G_{\boldsymbol{\psi}} \circ R_{\boldsymbol{\gamma}} \circ E_{\boldsymbol{\phi}}$. Its masked multi-step objective is

$$\mathcal{L}_{\mathrm{tr}}(\boldsymbol{\theta}) = \frac{\sum_{t\in\mathcal{B}} \sum_{k=1}^{K} \sum_{c=1}^{C_d} \sum_{i=1}^{H} \sum_{j=1}^{W} M_{i,j} \left(\widehat{X}^{(c)}_{t+k,i,j} - X^{(c)}_{t+k,i,j} \right)^2}{|\mathcal{B}| K C_d N_\Omega}, \tag{14}$$

where $|\mathcal{B}|$ is the batch size and $N_\Omega = \sum_{i=1}^{H} \sum_{j=1}^{W} M_{i,j}$ is the number of ocean cells. All variables and forecast steps are weighted equally. Land cells are excluded. Optimization, checkpoint selection, implementation settings, and consistency-parameter selection are detailed in Appendix A.3.

ConvLSTM, CNN–Bi-ConvGRU, ASPP–Bi-ConvGRU, and depthwise-separable U-Net++–Bi-ConvGRU are trained independently under identical data splits, preprocessing, loss, optimizer, training budget, and checkpoint selection. Each model receives the same ten input channels and predicts the same eight variables. Thus, only the spatial–temporal architecture differs. Further implementation details and the complete training and inference algorithm are provided in Appendix A.3.

3.6 LONG-HORIZON EVALUATION PROTOCOL

Let $\mathcal{T}_{\mathrm{ev}}$ be the set of valid test origins and $N_{\mathrm{ev}} = |\mathcal{T}_{\mathrm{ev}}|$ its size. An origin $t \in \mathcal{T}_{\mathrm{ev}}$ is valid when its L observed inputs and H_f future references lie within the test partition. Starting from observed states, H_f recursive updates are performed. Each update generates K outputs, applies equation 9– equation 12, and updates the window using equation 13. This produces

$$\widetilde{\mathcal{Y}}_t^{(H_f)} = \left\{ \widetilde{\mathbf{X}}_{t+1}, \ldots, \widetilde{\mathbf{X}}_{t+H_f} \right\}. \tag{15}$$

The sequence $\widetilde{\mathcal{Y}}_t^{(H_f)}$ contains the H_f corrected forecasts. References are accessed only after the complete rollout. No future input or ground-truth refresh is used. Bathymetry and the ocean mask remain fixed.

Predictions and references are inverse-normalized before scoring. For variable c, the physical-unit errors are

$$\mathrm{MAE}_c = \frac{\sum_{t\in\mathcal{T}_{\mathrm{ev}}} \sum_{h=1}^{H_f} \sum_{i=1}^{H} \sum_{j=1}^{W} M_{i,j} \left| \widetilde{X}^{\mathrm{phy},(c)}_{t+h,i,j} - X^{\mathrm{phy},(c)}_{t+h,i,j} \right|}{N_{\mathrm{ev}} H_f N_\Omega}, \tag{16}$$

$$\mathrm{RMSE}_c = \sqrt{\frac{\sum_{t\in\mathcal{T}_{\mathrm{ev}}} \sum_{h=1}^{H_f} \sum_{i=1}^{H} \sum_{j=1}^{W} M_{i,j} \left(\widetilde{X}^{\mathrm{phy},(c)}_{t+h,i,j} - X^{\mathrm{phy},(c)}_{t+h,i,j} \right)^2}{N_{\mathrm{ev}} H_f N_\Omega}}. \tag{17}$$

Here, $h \in \{1, \ldots, H_f\}$ is the forecast lead. The quantities $\widetilde{X}^{\mathrm{phy},(c)}_{t+h,i,j}$ and $X^{\mathrm{phy},(c)}_{t+h,i,j}$ are the corrected prediction and reference in physical units. VCMX performance is obtained by setting $c = c_v$. Variables with different units are reported separately. Any multivariate aggregate is computed in the normalized domain and labelled accordingly.

All retrained methods use identical forecast origins, input windows, masks, transformations, horizons, and metrics. Each origin uses the same observed initialization. A separate 696-hour interval from November 2023 evaluates rapid amplification around the 17.18-m VCMX event of 4–5 November. It supplements the complete held-out evaluation. The correction requires $\mathcal{O}(KC_dN_\Omega)$ operations per recursive step. Since K and C_d are fixed, its overhead is linear in N_Ω.

Table 1: Experimental setup and implementation details.

Parameter	Value	Parameter	Value
Data source	CMEMS IBI–GEBCO	Region	North-East Atlantic Ocean
Study period	2020–2023	Hourly states	35,017
Grid size	72×90	Input channels	$8 + 2$
Training data	Jan. 2020–Dec. 2022	Validation data	Jan.–Jun. 2023
Test data	Jul.–Dec. 2023	Input window	48 h
Direct horizon	3 h	Recursive horizon	168 h
Loss	Masked MSE	Optimizer	Adam
Learning rate	10^{-4}	Batch size	4
Maximum epochs	50	Patience	5 epochs
Framework	PyTorch Lightning	Precision	16-bit AMP
Hardware	$2\times$ GTX 1080 Ti	Target	VCMX
Metrics	MAE and RMSE	Domain	Ocean cells

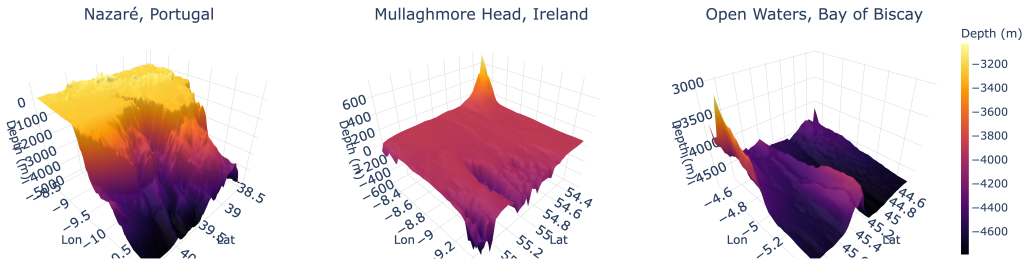

Figure 2: Bathymetry of Nazaré Canyon, Mullaghmore Head, and the open-water Bay of Biscay.

4 Results and Discussion

The proposed framework is evaluated in terms of predictive accuracy, spatial fidelity, and stability over extended recursive horizons. All methods use identical data partitions, preprocessing operations, forecast origins, and evaluation metrics. Quantitative comparisons are complemented by spatial analyses of high-energy wave events.

Experiments are conducted over Nazaré Canyon, within $38.5°$–$40.5°$N and $11.0°$–$8.5°$W. This region is selected because its abrupt seabed transitions produce nonuniform wave propagation and localized amplification. It therefore provides a challenging setting for evaluating bathymetry-aware representation and long-horizon forecast stability. Hourly fields are obtained from the CMEMS Atlantic–Iberian Biscay Irish Ocean Wave Reanalysis (Toledano et al., 2022) and mapped onto a 72×90 grid. High-resolution GEBCO bathymetry (GEBCO Bathymetric Compilation Group 2020, 2020) is interpolated onto the same grid. Eight dynamic variables, bathymetry, and a binary ocean mask form the ten input channels. VCMX denotes maximum crest-to-trough wave height and serves as the primary evaluation variable.

The dataset contains 35,017 hourly states from January 2020 to December 2023. The data are divided chronologically into training (January 2020–December 2022), validation (January–June 2023), and testing (July–December 2023) partitions. Variable-wise scaling is fitted only to the training data. Each sample uses 48 observed hours to predict the next three hourly states. No sample crosses a partition boundary. All errors are computed in physical units after inverse transformation.

The experimental configuration is summarized in Table 1. All architectures are trained using identical data partitions, preprocessing operations, loss functions, and computational budgets. Model selection and correction-parameter selection are performed only on the validation partition. The test partition is reserved for final evaluation.

During evaluation, the complete 168-hour forecast is generated before the corresponding reference fields are accessed. Ground-truth refresh is therefore excluded. The same ocean mask, forecast origins, inverse transformation, and metric definitions are used for every compared method. This controlled protocol ensures that the reported differences arise from the forecasting architectures rather than unequal data exposure or evaluation conditions.

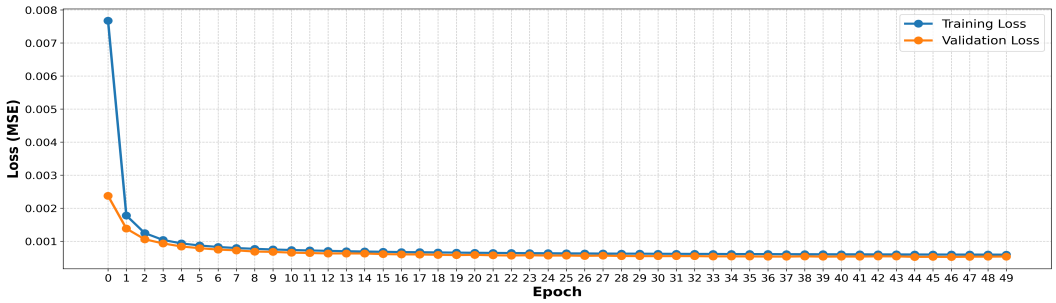

Figure 3: Training and validation MSE of the proposed CNN–Bi-ConvGRU over 50 epochs.

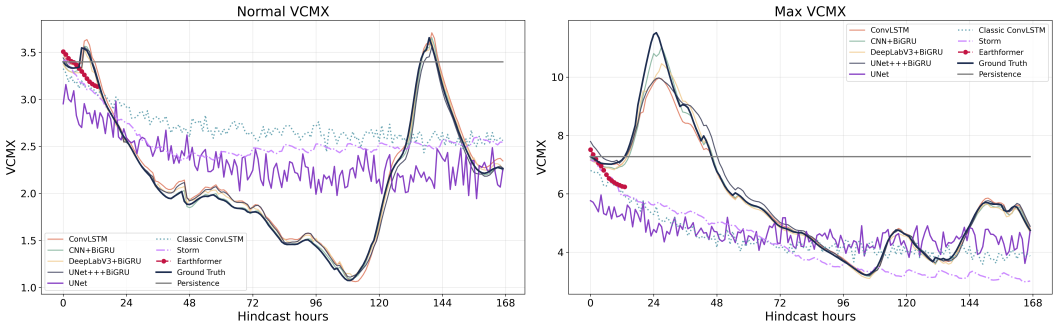

Figure 4: VCMX forecasts at Nazaré under normal and high-wave conditions.

Figure 2 compares the seabed geometry of the three evaluation domains. Nazaré Canyon exhibits the strongest depth variation, with an abrupt transition from the continental shelf to the deep canyon. Mullaghmore Head represents a shallower coastal domain with localized seabed relief. In contrast, the Bay of Biscay region is dominated by deep open-water bathymetry with smoother large-scale variation. These distinct geometries allow the model to be evaluated under canyon-dominated, coastal-shelf, and deep-ocean conditions. The comparison therefore examines whether bathymetry-conditioned features remain informative across heterogeneous seabed structures.

Figure 3 presents the training and validation losses of the proposed CNN–Bi-ConvGRU. Both losses decrease sharply during the initial epochs and subsequently approach stable values. The validation loss remains close to the training loss and does not exhibit a late-stage increase. This behavior indicates stable optimization under the chronological data split. It also suggests that the selected checkpoint does not result from an isolated fluctuation in the validation objective.

Figure 4 compares 168-hour VCMX forecasts at Nazaré Canyon. The proposed CNN–Bi-ConvGRU closely follows the reference under both normal and high-wave conditions. It captures the dominant peaks, rapid decay, and secondary amplification with limited long-horizon drift. The other trained spatio-temporal models recover the main trend but show larger deviations near extreme peaks. U-Net, classical ConvLSTM, the cross-scale model, and persistence fail to preserve the complete temporal evolution.

Figures 5 and 6 compare 168-hour VCMX forecasts under coastal and open-water bathymetric conditions. Both representative normal-energy sequences and high-VCMX sequences are examined. The proposed model follows the phase and amplitude variations of the reference series throughout the recursive horizon. In contrast, STORM and U-Net tend toward weakly varying trajectories and fail to reproduce several rapid transitions. This behavior is most evident around the dominant peaks, where localized wave amplification must be recovered from the preceding multivariate states.

At Mullaghmore Head, the proposed model captures the timing of the two major high-VCMX events and follows the intervening decay. A moderate reduction in peak magnitude remains visible, but

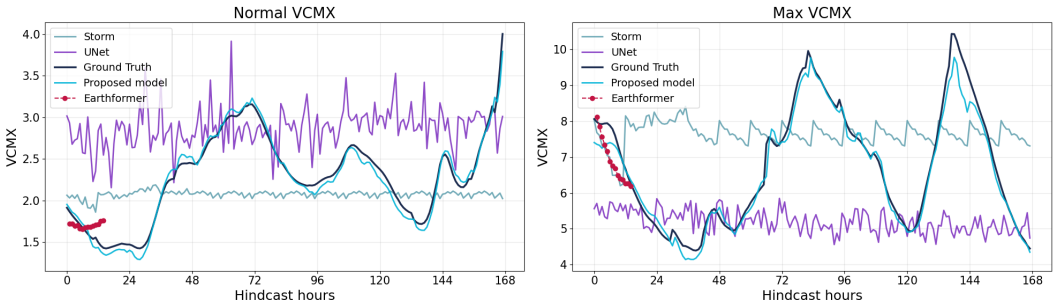

Figure 5: VCMX forecasts at Mullaghmore Head under normal and extreme wave conditions.

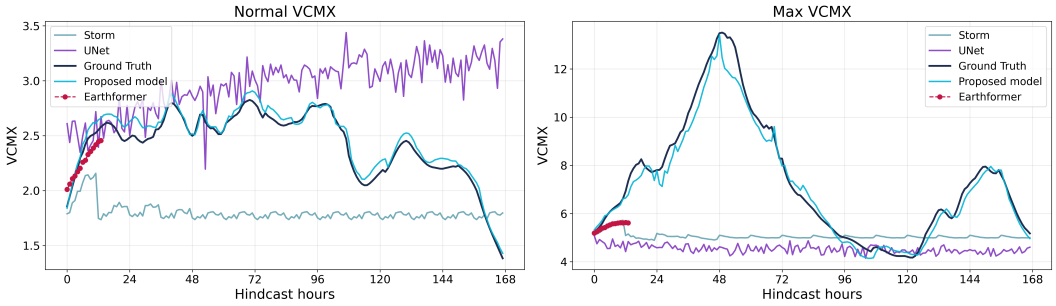

Figure 6: VCMX forecasts over the Bay of Biscay under normal and extreme conditions.

the temporal evolution is preserved. The normal-energy sequence is also reconstructed without the persistent offset shown by the compared models. These observations indicate that the learned representation remains effective under a coastal bathymetric structure that differs from the canyon-dominated Nazaré domain.

A similar pattern is observed over the open-water Bay of Biscay domain. The proposed model reproduces the principal rise, peak, and decay of the extreme sequence. It also follows the lower-amplitude variations of the normal-energy sequence. STORM remains close to a nearly constant level after the initial hours, while U-Net exhibits a persistent bias and misses the dominant extreme peak. The proposed model slightly underestimates the largest crest-to-trough height, but retains its occurrence time and overall temporal structure. The result indicates that the consistency-guided rollout controls drift without forcing the forecast toward a constant climatological state.

Table 2 reports VCMX errors across the three bathymetric regions. Within each region, all methods use the same test sequence, ocean mask, inverse transformation, and 168-hour recursive horizon. The proposed CNN–Bi-ConvGRU achieves the lowest reported MAE and RMSE under both normal and high-VCMX conditions. Over the complete test set, it obtains an MAE of 0.1512 m and an RMSE of 0.1991 m. The cross-scale spatio-temporal model yields 0.8501 m and 1.2447 m, respectively.

The regional results show consistent behaviour across distinct seabed conditions. At Nazaré, the proposed model attains MAE values of 0.0499 m and 0.1375 m for normal and high-VCMX cases. It also maintains low errors at Mullaghmore Head and in the open-water Bay of Biscay. These results support the trajectories shown in Figures 5 and 6. The model therefore preserves both average evolution and high-energy wave variations during extended recursion.

A progressive ablation is performed to examine the effects of recursive evaluation, bathymetric conditioning, and bidirectional temporal modelling. Four configurations are considered. The first is a vanilla ConvLSTM without bathymetry or consistency-guided recursion. The second applies recursive regulation without bathymetric input. The third adds bathymetry while retaining the ConvLSTM

Table 2: Regional VCMX forecasting performance over a 168-hour horizon.

		Normal VCMX		High VCMX	
Region	**Method**	**MAE**	**RMSE**	**MAE**	**RMSE**
Bay of Biscay	Cross-scale model	1.1564	1.6075	1.8148	2.3950
	U-Net	0.5424	0.5986	2.9044	2.9513
	Proposed	**0.0715**	**0.0838**	**0.3482**	**0.3721**
Mullaghmore Head	Cross-scale model	1.7390	2.2543	2.3418	2.9245
	U-Net	0.8360	0.9999	1.6880	1.9697
	Proposed	**0.0754**	**0.0925**	**0.2374**	**0.2843**
Nazaré Canyon	Cross-scale model	0.5352	0.6352	0.8503	0.9989
	U-Net	0.5890	0.7179	1.6255	1.9283
	Proposed	**0.0499**	**0.0697**	**0.1375**	**0.1821**

Table 3: Ablation of bathymetry, correction, and temporal modeling.

	Bathymetry	**Correction**	**Normal VCMX**		**High VCMX**	
Configuration			**MAE**	**RMSE**	**MAE**	**RMSE**
Vanilla ConvLSTM	No	No	0.1384	0.8629	1.5289	1.7974
ConvLSTM + recursion	No	Yes	3.6417	4.2885	2.5484	3.2350
ConvLSTM + bathymetry	Yes	Yes	0.0956	0.1230	0.2199	0.2821
Proposed CNN–Bi-ConvGRU	Yes	Yes	**0.0499**	**0.0697**	**0.1375**	**0.1821**

backbone. The final configuration replaces the temporal backbone with the proposed Bi-ConvGRU. All configurations use a 48-hour observation window and are evaluated over the same VCMX sequences.

Recursive deployment without bathymetric input produces substantial error growth. The correction mechanism alone cannot recover spatial forcing that is absent from the input representation. Adding bathymetry sharply reduces both MAE and RMSE. This result shows that recursive regulation depends on an informative seabed-conditioned state. Replacing the ConvLSTM with Bi-ConvGRU provides a further reduction under both wave regimes. The complete framework therefore benefits from the complementary roles of bathymetric encoding, recursive regulation, and bidirectional temporal modelling.

The correction adds no trainable parameters or auxiliary network. For K direct predictions, C_d predicted variables, and N_Ω ocean cells, its cost is $\mathcal{O}(KC_dN_\Omega)$ per recursive update. It operates directly on the predicted tensors and requires no additional model inference. Since K and C_d are fixed, its overhead scales linearly with N_Ω. The CNN–Bi-ConvGRU therefore remains the dominant computational component.

5 Conclusion and Future Scope

A bathymetry-conditioned CNN–Bi-ConvGRU framework has been developed for long-horizon multivariate wave-field forecasting. Dynamic ocean variables, seabed depth, and land–ocean boundaries are jointly encoded. Forecast inconsistency is then estimated from multiple direct outputs and used to regulate recursive state propagation. No future observations, auxiliary correction network, or inference-time retraining are required. Over the 168-hour Nazaré high-VCMX sequence, the proposed framework achieves an MAE of 0.1375 m and an RMSE of 0.1821 m. It also maintains low errors across the coastal-shelf and open-water evaluations. The results further show that rapid transitions and high-VCMX events are tracked across coastal and open-water conditions. However, the present evaluation is limited to three North Atlantic regions and reanalysis-derived inputs. Future work will examine additional ocean basins, observation-driven forcing, probabilistic uncertainty estimation, and explicit physical constraints for extreme-wave prediction.

References

Mohamad Javad Alizadeh and Vahid Nourani. Multivariate GRU and LSTM Models for Wave Forecasting and Hindcasting in the Southern Caspian Sea. *Ocean Engineering*, 298:117193,

2024. doi: 10.1016/j.oceaneng.2024.117193.

Ashesh Chattopadhyay, Michael Gray, Tianning Wu, Anna B. Lowe, and Ruoying He. Oceannet: A principled neural operator-based digital twin for regional oceans. *Scientific Reports*, 14:21181, 2024. doi: 10.1038/s41598-024-72145-0.

GEBCO Bathymetric Compilation Group 2020. The gebco_2020 grid - a continuous terrain model of the global oceans and land., 2020. URL https://doi.org/10.5285/a29c5465-b138-234d-e053-6c86abc040b9.

Daniel Holmberg, Emanuela Clementi, Italo Epicoco, and Teemu Roos. Accurate Mediterranean Sea Forecasting via Graph-Based Deep Learning. *Scientific Reports*, 15:45051, 2025. doi: 10.1038/s41598-025-31177-w.

Qihe Huang, Zhengyang Zhou, Yangze Li, Jiaming Ma, Kuo Yang, Binwu Wang, Xu Wang, and Yang Wang. STORM: Synergistic Cross-Scale Spatio-Temporal Modeling for Weather Forecasting. In *International Conference on Learning Representations*, 2026.

Ioannis Kordatos, Athanasios Donas, George Galanis, Ioannis Famelis, and Alex Alexandridis. Significant wave height prediction in nested domains using radial basis function neural networks. *Ocean Engineering*, 305:117865, 2024. doi: 10.1016/j.oceaneng.2024.117865.

Yang Liu, Zinan Zheng, Jiashun Cheng, Fugee Tsung, Deli Zhao, Yu Rong, and Jia Li. CirT: Global Subseasonal-to-Seasonal Forecasting With Geometry-Inspired Transformer. In *International Conference on Learning Representations*, 2025.

Leonard Lupin-Jimenez, Moein Darman, Subhashis Hazarika, Tianning Wu, Michael Gray, Ruoying He, Anthony Wong, and Ashesh Chattopadhyay. Simultaneous Emulation and Downscaling With Physically Consistent Deep Learning-Based Regional Ocean Emulators. *Journal of Geophysical Research: Machine Learning and Computation*, 2(3):e2025JH000851, 2025. doi: 10.1029/2025JH000851.

Tao Lv, Aifeng Tao, Peng Tian, Jun Fan, Wei Xu, and Jinhai Zheng. A Frequency-Domain and Spatiotemporal Global Joint Learning Method for Wind–Swell Separation and Short-Term Wave Field Modeling. *Ocean Engineering*, 343:123392, 2026. doi: 10.1016/j.oceaneng.2025.123392.

Yuan Niu, Qiusheng Huang, Xiaohui Zhong, Anboyu Guo, Lei Chen, Dianjun Zhang, Zhen-Qiang Zhou, Xiaoyan Jia, Libo Wu, Renhe Zhang, Jiawei Qi, Xuefeng Zhang, and Hao Li. A Deep Learning Global Ocean Forecasting Model With Sub-Daily and Eddy-Resolving Resolution. *npj Climate and Atmospheric Science*, 2026. doi: 10.1038/s41612-026-01444-2.

Jian Shen, Zhengui Wang, Jiabi Du, Yinglong J Zhang, and Qubin Qin. Machine learning-based wave model with high spatial resolution in chesapeake bay. *Earth and Space Science*, 11(3): e2023EA003303, 2024.

Phyusin Thet, Aifeng Tao, Tao Lv, and Jinhai Zheng. Wave Height Forecasting in the Bay of Bengal Using Multivariate Hybrid Deep Learning Models. *Journal of Marine Science and Engineering*, 13(8):1412, 2025. doi: 10.3390/jmse13081412.

Jindong Tian, Yifei Ding, Ronghui Xu, Hao Miao, Chenjuan Guo, and Bin Yang. ARROW: An Adaptive Rollout and Routing Method for Global Weather Forecasting. In *International Conference on Learning Representations*, 2026.

Cristina Toledano, Malek Ghantous, Pablo Lorente, Alice Dalphinet, Lotfi Aouf, and Marcos G. Sotillo. Impacts of an altimetric wave data assimilation scheme and currents-wave coupling in an operational wave system: The new copernicus marine ibi wave forecast service. *Journal of Marine Science and Engineering*, 10(4), 2022. ISSN 2077-1312. doi: https://doi.org/10.3390/jmse10040457. URL https://www.mdpi.com/2077-1312/10/4/457.

Yogesh Verma, Markus Heinonen, and Vikas Garg. Climode: Climate and weather forecasting with physics-informed neural odes. In *International Conference on Learning Representations*, volume 2024, pp. 8408–8430, 2024.

Ziliang Zhang, Huaming Yu, Danqin Ren, Chenyu Zhang, Minghua Sun, and Xin Qi. Ocean wave forecasting with deep learning as alternative to conventional models. *Journal of Advances in Modeling Earth Systems*, 17(11):e2025MS005285, 2025.

Ziliang Zhang, Huaming Yu, Xin Qi, Qian Li, Jiacheng Ye, Shouwen Qiao, Xiaoping Zhao, and Ge Li. Improving Long-Term Forecast Stability of OceanCastNet: Efficient Fine-Tuning via a Temporal Change Loss Function. *Ocean Engineering*, 354:124869, 2026. doi: 10.1016/j.oceaneng.2026.124869.

A Appendix

This appendix provides the training algorithm and supplementary results omitted from the main paper due to space constraints. The reported settings, symbols, and model names follow the definitions used in the main text.

A.1 Bathymetric Data Consistency

Figure 7 summarizes the annual elevation coverage of the Nazaré domain. The minimum elevation ranges from -5001 to -4989 m. The maximum remains close to 662 m after 2019. The depth-covered area varies from $30{,}240.31$ to $30{,}361.04$ km^2. These small variations arise from annual data coverage and preprocessing. They are not interpreted as physical seabed changes. A fixed bathymetric field is therefore used throughout model training and evaluation.

A.2 Feature-Correlation Analysis

The raw dataset provides the 22 physical variables listed in Table 4. From these variables, 14 additional features are constructed to represent directional interactions, spectral concentration, Stokes drift, and energy dissipation. The final input contains 18 predictor channels. These comprise four directly measured variables, namely VCMX, VHM0, VTPK, and VTM02, and the 14 engineered features in Table 5. The model predicts eight physical wave variables, with VCMX treated as the primary evaluation target. Thus, the number of input features and predicted variables is $C_f = 18$ and $C_d = 8$, respectively. Only historical feature values are supplied to the model.

Table 4: Physical variables available in the raw dataset.

Variable	Description
VCMX	Maximum crest-to-trough wave height, $H_{c,\max}$
VHM0	Spectral significant wave height, H_{m0}
VHM0_SW1	Spectral significant primary-swell wave height
VHM0_SW2	Spectral significant secondary-swell wave height
VHM0_WW	Spectral significant wind-wave height
VMDR	Mean wave direction, M_{dir}
VMDR_SW1	Mean primary-swell direction
VMDR_SW2	Mean secondary-swell direction
VMDR_WW	Mean wind-wave direction
VMXL	Maximum crest height
VPED	Principal wave direction at the spectral peak
VSDX	Eastward Stokes-drift component
VSDY	Northward Stokes-drift component
VTM01_SW1	Primary-swell period from spectral moments $(0, 1)$
VTM01_SW2	Secondary-swell period from spectral moments $(0, 1)$
VTM01_WW	Wind-wave period from spectral moments $(0, 1)$
VTM02	Mean wave period from spectral moments $(0, 2)$, T_{m02}
VTM10	Mean wave period from spectral moments $(-1, 0)$, T_{m-10}
VTPK	Wave period at the spectral peak, T_p
TAUX	Eastward surface stress associated with ocean viscous dissipation
TAUY	Northward surface stress associated with ocean viscous dissipation
TLA	Wave-induced mixing energy flux into seawater

Let $\epsilon = 10^{-6}$ be a numerical-stability constant. For two angular variables θ_1 and θ_2, their circular difference is

$$\Delta_{\mathrm{circ}}(\theta_1, \theta_2) = \min(|\theta_1 - \theta_2| \bmod 360^\circ,\ 360^\circ - (|\theta_1 - \theta_2| \bmod 360^\circ)) \,. \tag{18}$$

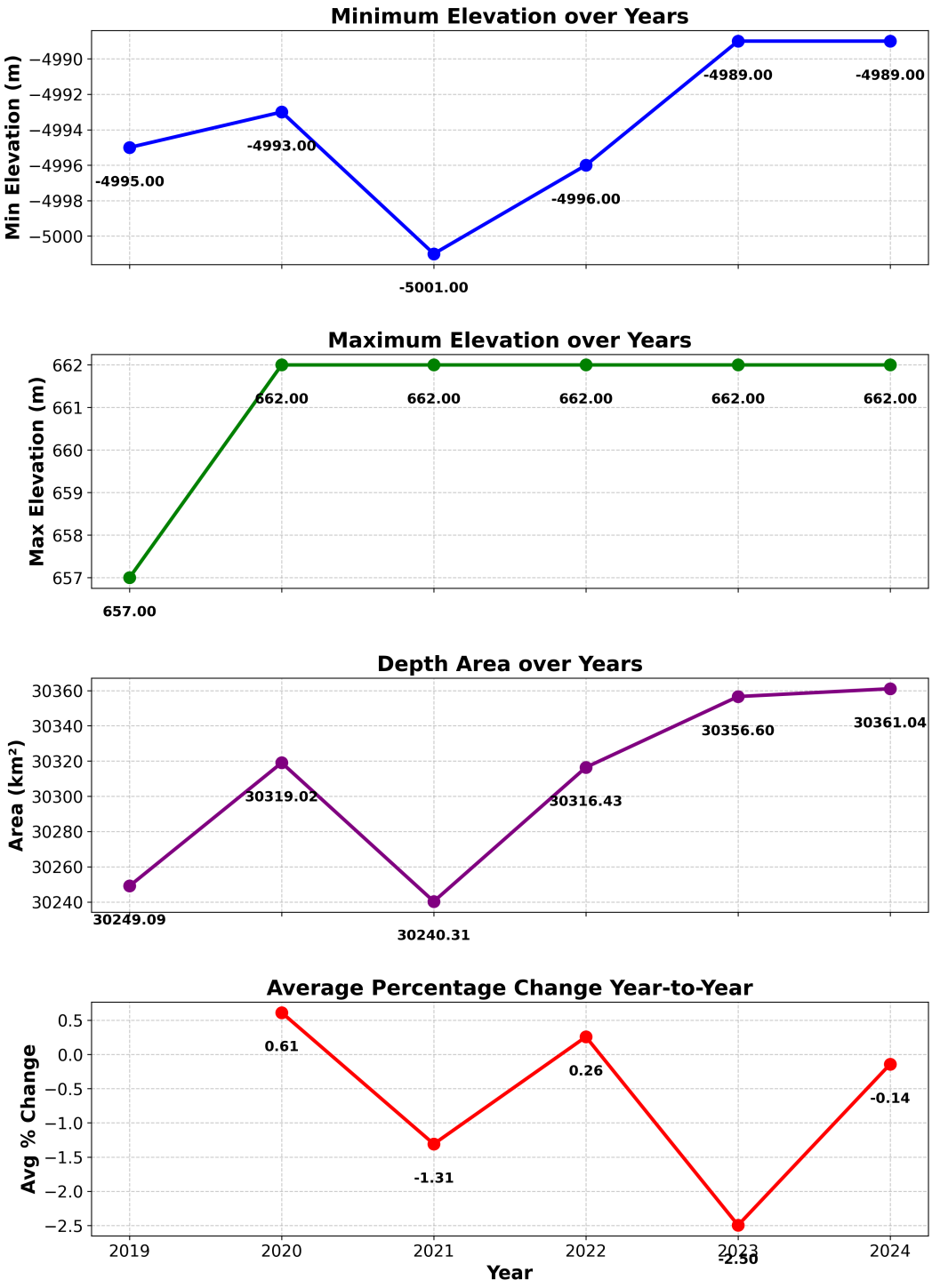

Figure 7: Annual elevation and depth-coverage statistics for the Nazaré domain.

Figure 8 reports the Pearson correlations among the source variables and the 18 retained predictors. VCMX is strongly associated with the wave-height variables VHM0 and VMXL. Several wave-period variables also exhibit substantial mutual dependence. The engineered features supplement these variables with directional, spectral, drift, stress, and interaction information. High correlations

Table 5: Engineered predictor features.

Feature	Definition	Physical interpretation
VSDmag	$\sqrt{\mathrm{VSDX}^2 + \mathrm{VSDY}^2}$	Stokes-drift magnitude
VMDR_sin	$\sin(\pi\,\mathrm{VMDR}/180)$	Sine encoding of mean wave direction
VMDR_cos	$\cos(\pi\,\mathrm{VMDR}/180)$	Cosine encoding of mean wave direction
delta_theta_SW	$\Delta_{\mathrm{circ}}(\mathrm{VMDR_SW1}, \mathrm{VMDR_SW2})$	Directional separation between swell systems
swell_energy_ratio	$\mathrm{VHM0_SW2}/(\mathrm{VHM0_SW1} + \epsilon)$	Relative secondary-to-primary swell height
delta_theta_SW1_WW	$\Delta_{\mathrm{circ}}(\mathrm{VMDR_SW1}, \mathrm{VMDR_WW})$	Separation between primary swell and wind waves
inv_VSDmag	$1/(\mathrm{VSDmag} + \epsilon)$	Inverse Stokes-drift magnitude
spectral_sharpness	$\mathrm{VTPK}/(\mathrm{VTM02} + \epsilon)$	Proxy for spectral concentration
log_TLA	$\ln(\mathrm{TLA} + \epsilon)$	Compressed mixing-energy flux
TLA_x_swell_ratio	$\mathrm{log_TLA} \times \mathrm{swell_energy_ratio}$	Dissipation–swell interaction
TLA_x_delta_theta	$\mathrm{log_TLA} \times \mathrm{delta_theta_SW}$	Dissipation–direction interaction
TAU_mag	$\sqrt{\mathrm{TAUX}^2 + \mathrm{TAUY}^2}$	Resultant surface-stress magnitude
resultant_angle_TAU	$\mathrm{atan2}(\mathrm{TAUY}, \mathrm{TAUX})$	Resultant surface-stress direction
resultant_angle_VSD	$\mathrm{atan2}(\mathrm{VSDY}, \mathrm{VSDX})$	Resultant Stokes-drift direction

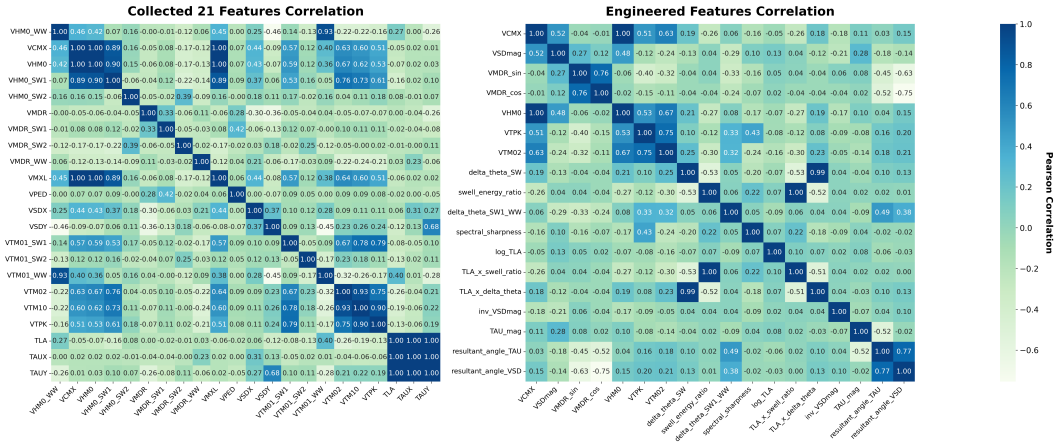

Figure 8: Pearson correlations among the source variables and retained predictor features.

among some engineered predictors are expected because they share source variables. The analysis identifies statistical dependence and does not imply causality. Every transformation is computed at the current or an earlier time step. No future target is used.

A.3 Training and Recursive Inference

Algorithm 1 summarizes model training and long-horizon inference. The maximum number of epochs is $E_{\max} = 50$. Adam is used with a learning rate of 10^{-4} and a mini-batch size of 4. The validation loss is evaluated after every epoch. The checkpoint $\boldsymbol{\theta}^{\star}$ with the lowest validation loss is retained. Training stops after five consecutive non-improving epochs. The correction parameters $\lambda = 0.5$ and $\alpha_{\min} = 0.1$ are selected using the validation set and fixed before testing.

Here, $\mathcal{T}_{\mathrm{tr}}$, $\mathcal{T}_{\mathrm{val}}$, and $\mathcal{T}_{\mathrm{ev}}$ are the training, validation, and test forecast-origin sets. The symbols K and H_f denote the direct and recursive forecast horizons. The index e denotes the training epoch, r denotes the recursive step, and $c \in \{1, \ldots, C_d\}$ denotes a dynamic variable. The correction is applied only after training and does not modify $\boldsymbol{\theta}^{\star}$. No future reference or ground-truth refresh is used during recursion.

The implementation uses PyTorch Lightning with 16-bit automatic mixed precision. Distributed data parallelism is performed on two NVIDIA GeForce GTX 1080 Ti GPUs. These settings affect computational efficiency but do not alter the forecasting model or evaluation protocol.

Algorithm 1 Training and consistency-guided forecasting

Require: $\mathcal{T}_{\mathrm{tr}}, \mathcal{T}_{\mathrm{val}}, \mathcal{T}_{\mathrm{ev}}, E_{\max}, K, H_f$
Fit normalization using $\mathcal{T}_{\mathrm{tr}}$ and initialize $\boldsymbol{\theta}$
for $e = 1, \ldots, E_{\max}$ **do**
  for each mini-batch $\mathcal{B} \subset \mathcal{T}_{\mathrm{tr}}$ **do**
    Compute $\mathcal{L}_{\mathrm{tr}}(\boldsymbol{\theta}; \mathcal{B})$
    Update $\boldsymbol{\theta}$ using Adam
  end for
  Evaluate $\mathcal{L}_{\mathrm{val}}$ on $\mathcal{T}_{\mathrm{val}}$
  if $\mathcal{L}_{\mathrm{val}}$ improves **then**
    $\boldsymbol{\theta}^{\star} \leftarrow \boldsymbol{\theta}$
  end if
  Stop after five non-improving epochs
end for
for each $t \in \mathcal{T}_{\mathrm{ev}}$ **do**
  Initialize $\mathcal{H}_t$ using L observed states
  for $r = 0, \ldots, H_f - 1$ **do**
    $\widehat{\mathcal{Y}}_{t+r} \leftarrow f_{\boldsymbol{\theta}^{\star}}(\mathcal{H}_{t+r})$
    Compute $d_{t+r}^{(c)}$ and $\alpha_{t+r}^{(c)}$ for all c
    Obtain $\widetilde{\mathbf{X}}_{t+r+1}$ using equation 12
    Update $\mathcal{H}_{t+r+1}$ using equation 13
  end for
  Return $\{\widetilde{\mathbf{X}}_{t+1}, \ldots, \widetilde{\mathbf{X}}_{t+H_f}\}$
end for

A.4 Complete Architecture Comparison

The complete architecture comparison is reported in Table 6. Each model uses the same 48-hour observation window, three-hour direct horizon, chronological partitions, normalization parameters, ocean mask, loss function, optimizer, and checkpoint rule. The table reports VCMX errors in metres. Results for the complete test partition and the high-VCMX interval are presented separately.

Table 6: Complete VCMX architecture comparison.

	Complete test set		**High VCMX**	
Architecture	**MAE**	**RMSE**	**MAE**	**RMSE**
ConvLSTM	0.0904	0.1167	0.2552	0.3228
CNN–Bi-ConvGRU	**0.0476**	**0.0660**	**0.1512**	**0.1991**
ASPP–Bi-ConvGRU	0.0738	0.0998	0.2127	0.2965
U-Net++–Bi-ConvGRU	0.0855	0.1122	0.3258	0.4153

CNN–Bi-ConvGRU achieves the lowest MAE and RMSE in both evaluation settings. ASPP–Bi-ConvGRU improves upon ConvLSTM but remains less accurate than the compact CNN encoder. U-Net++–Bi-ConvGRU provides no additional gain despite its more complex spatial pathway. Under the present grid resolution and training conditions, bidirectional temporal representation is therefore more effective than increasing spatial-encoder complexity.

A.5 Observation-Window Sensitivity

Table 7 evaluates observation windows of 12, 24, 48, and 72 hours. All remaining settings are fixed. The 24-hour window gives the lowest error under normal conditions. The 48-hour window achieves the lowest high-VCMX MAE and RMSE. Increasing the window to 72 hours does not improve either regime. The 48-hour setting is retained because high-energy wave prediction is the primary objective.

Table 7: Sensitivity to the observation-window length.

	Normal VCMX		High VCMX	
Lookback	**MAE**	**RMSE**	**MAE**	**RMSE**
12 h	0.0540	0.0740	0.1653	0.2126
24 h	**0.0480**	**0.0674**	0.1496	0.1925
48 h	0.0499	0.0697	**0.1375**	**0.1821**
72 h	0.0548	0.0758	0.1568	0.2020

A.6 Model Size and Evaluation Scope

Table 8 summarizes the model sizes and evaluation horizons recorded for the implemented methods. Parameter counts refer to the evaluated implementations. EarthFormer is listed for completeness but is not treated as a horizon-matched baseline because its available prediction interval is 14 hours.

Table 8: Model size and evaluation configuration.

Method	**Lookback**	**Evaluation horizon**	**Parameters**
Vanilla ConvLSTM	48 h	168 h	289 K
Cross-scale spatio-temporal model	48 h	168 h	221 K
EarthFormer	12 h	14 h	166 K
U-Net	48 h	168 h	3.47 M
Proposed CNN–Bi-ConvGRU	48 h	168 h	534 K

The proposed model is smaller than the evaluated U-Net but larger than ConvLSTM and the cross-scale model. Its additional parameters arise primarily from bidirectional recurrent processing. This comparison describes implementation size only. It does not establish computational superiority because runtime and memory measurements are not reported.

A.7 Additional Implementation Details

The spatial encoder contains two 3×3 convolutional layers with 64 and 32 filters. The temporal module contains two bidirectional ConvGRU layers with 64 and 32 hidden channels. Unit stride and same padding are used throughout. A 1×1 decoder produces KC_d output channels. Adam is used with learning rate 10^{-4} and batch size 4. Training is limited to 50 epochs and uses an early-stopping patience of five epochs. The implementation uses PyTorch Lightning, 16-bit automatic mixed precision, and two NVIDIA GeForce GTX 1080 Ti GPUs.

The direct horizon is $K = 3$, while the recursive horizon is $H_f = 168$. At each recursive step, three direct states are generated. Their channel-wise variation determines the correction coefficient. Only the corrected first state is appended to the input window. The procedure therefore advances by one hour without accessing a future observation.