A Neural Framework to Forecast Short-Term Waves for a Digital Twin of the Ocean

1st Bernardo M. Chagas
IDMEC, Instituto Superior Técnico
Universidade de Lisboa
Lisbon, Portugal
ORCID: 0009-0005-1048-5923

2nd Miguel S. E. Martins
IDMEC, Instituto Superior Técnico
Universidade de Lisboa
Lisbon, Portugal
ORCID: 0000-0002-6285-8737

3rd João M. C. Sousa
IDMEC, Instituto Superior Técnico
Universidade de Lisboa
Lisbon, Portugal
ORCID: 0000-0002-8030-4746

4th Susana Vieira
IDMEC, Instituto Superior Técnico
Universidade de Lisboa
Lisbon, Portugal
ORCID: 0000-0001-7961-1004

***Abstract*—Accurate short-term wave forecasting is essential for ocean monitoring and for supporting stakeholder decision-making. This paper proposes a neural framework for one-hour-ahead prediction of significant wave height, mean wave period, and mean wave direction using the ERA5 dataset. The proposed approach uses a local 3×3 spatial grid to capture the essential neighbourhood dynamics of the wave and wind fields. The information from this grid, sampled over the two previous time steps, forms a 108-dimensional input vector that represents local spatio-temporal interactions. An autoencoder is then used to compress this representation into a 32-dimensional latent space, which is subsequently mapped to the target variables using a multilayer perceptron. The performance of the model is evaluated in a spatially disjoint test region. The results show high predictive accuracy, with coefficients of determination above 0.9959 for all target variables, and low absolute errors. The proposed framework provides a computationally efficient alternative for short-term wave forecasting suitable for integration into a digital twin of the ocean.**

***Index Terms*—oceanography, forecasting methods, time series prediction, neural networks, machine learning, autoencoder, multilayer perceptron.**

I. Introduction

Weather forecasting is a computationally demanding task achieved by solving large-scale numerical models with spatial and temporal high resolution. These models, typically based on systems of partial differential equations [1], require substantial computational resources and are therefore executed on high-performance computing (HPC) infrastructures. This dependence on resources reduces overall accessibility by limiting the number of potential users and constrains the systematic generation of large ensembles, which are essential for uncertainty quantification and robust decision support.

In the context of DATAz, a Digital Twin for the Azores Free Technological Zone Project [2] [3], it is necessary to explore and develop more efficient and accessible alternatives to traditional numerical weather prediction models. The goal is to build a lightweight surrogate model based on machine learning algorithms that enables the evaluation of oceanic scenarios in an environment that does not rely solely on HPC systems. This solution is particularly relevant for the analysis of what-if scenarios, providing a powerful tool for stakeholders. Thus, the work presented here focuses on the development of a machine learning (ML) model intended to accurately predict wave parameters one hour ahead.

Machine learning-based surrogate models have emerged as computationally efficient alternatives to physics-based solvers for wave forecasting. Multilayer perceptron (MLP) and ensemble aggregation schemes [4], [5] have proven effective for predicting wave parameters, while convolutional architectures and U-Net designs, exemplified by TurboSWAN [6], better capture spatial structure. Random forests [7] have been applied to derive spatially distributed wave characteristics, while deep learning surrogates such as DELWAVE 1.0 [8] enable basin-scale wave climate representation. Recent approaches leverage spatio-temporal architectures, including ConvLSTM [9], transformers [10], and physics-informed neural networks (PINNs) [11] to enhance accuracy and physical consistency.

To develop a lightweight ML model, an innovative approach is explored in which the target variables are forecasted by using neighbourhood information rather than learning explicit spatial dependencies. This work proposes a grid-based forecasting architecture that reduces model complexity through neighbourhood aggregation. The main contributions are three-fold: (i) formulation of a one-hour-ahead wave forecasting model using a 3×3 grid centered on each target point; (ii) implementation of autoencoder-based dimensionality reduction that compresses 108 input features into a 32-dimensional latent representation; and (iii) validation of an MLP forecaster on spatially disjoint test regions, demonstrating robust generalization and low normalized prediction errors. Additionally, a baseline LSTM model is included for comparison under the same data conditions.

The remainder of this paper is organized as follows. Section II introduces the dataset and describes the variables considered. Section III details the modeling methodology and general architecture. Section IV presents the network design and training procedure. Section V reports the experimental results, which are analyzed and discussed in Section VI. Finally, Section VII concludes the paper and outlines the directions for future work.

To support reproducibility and facilitate further research, we will make the full source code used in this work publicly available upon acceptance of the paper. The repository will provide access to the raw and preprocessed datasets, as well as the scripts required for data transfer and processing, model training, and result evaluation [15].

II. Dataset and Data Preparation

This work uses the ERA5 hourly reanalysis dataset on single levels [12], obtained from the Copernicus Climate Data Store. ERA5 is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and provides global atmospheric and oceanographic estimates by assimilating observations into a numerical weather prediction framework.

The data is provided on a regular latitude-longitude grid with a spatial resolution of $0.25°$ in both directions and hourly temporal resolution. This resolution offers a suitable trade-off between spatial coverage and computational tractability for regional-scale wave forecasting in the Azores area.

A. Variables chosen

The selection of input variables was guided by the physical processes governing wind-wave generation and propagation, as well as by the variables commonly employed in third-generation spectral wave models such as SWAN [13]. These models solve the wave action balance equation, accounting for wind input, non-linear wave-wave interactions, dissipation processes and bathymetric effects through depth-dependent propagation mechanisms.

In order to approximate these dynamics in a data-driven model, the ERA5 dataset used in this work includes a total of eight variables, of which three (latitude, longitude and time) serve as coordinate indices. The remaining five are physical predictors and target variables relevant to wave evolution. The atmospheric forcing variables include the 10 m wind in both u and v components (*u10* and *v10*). The wave parameters of interest are mean wave direction (*mwd*), mean wave period (*mwp*) and significant wave height (*swh*).

Because *mwd* was given in degrees, it was verified that a major prediction error was occurring for values of *mwd* closer to 0 and closer to 360, given the discontinuity of this type of data. Thus, at the cost of increasing in 1 the number of wave parameters through a sin and cos decomposition (*mwd_sin* and *mwd_cos*), the results would improve considerably [14]. After predicting each component, an aggregation was calculated using the following.

$$mwd_{pred} = \arctan\left(\frac{mwd\_sin_{pred}}{mwd\_cos_{pred}}\right)$$

All variables were downloaded using a custom Python workflow, stored in point-wise NetCDF files and subsequently merged into consolidated datasets for training and testing. This preprocessing pipeline ensured consistent spatial alignment and temporal synchronisation across all variables.

Table I summarizes the variables used by the model and their roles.

TABLE I
VARIABLES USED IN THIS WORK

Variable	Description	Units	Role
$u10$	Zonal wind component at 10 m	m s^{-1}	input
$v10$	Meridional wind component at 10 m	m s^{-1}	input
mwp	Mean wave period	s	input/target
swh	Significant wave height	m	input/target
mwd_{deg}	Mean wave direction	–	feature generation
$mwd_{\sin}$	Sine of mean wave direction	–	input/target
$mwd_{\cos}$	Cosine of mean wave direction	–	input/target

III. Proposed Framework

A. Definition of the General Approach

A square-shaped region covering the Azores Archipelago and its surrounding oceanic area was selected to define the geographical scope of the study. The respective boundaries were set from $35°$ to $41°$ North in latitude and from $23°$ to $33°$ West in longitude. The temporal window was defined from 1 January 2025 to 3 October 2025, resulting in a total of 6624 hourly observations for each spatial point.

B. Spatio-Temporal Formulation

The adopted modelling strategy follows a discrete state-space formulation, in which the future system state depends on both past system states and external forcing variables. For each spatial point, the wave properties at time $t+1$ are predicted using information from the two previous time steps (t and $t-1$).

Specifically, the state vector is defined by the historical wave properties (*mwp*, *swh*, *mwd_sin*, *mwd_cos*) while the input vector contains historical meteorological variables (*u10*, *v10*) influencing the wave evolution. This formulation allows the model to capture short-term temporal dependencies while maintaining a compact input representation suitable for efficient learning.

This relationship is expressed as follows.

$$\left(Y_{1..n}^{t+1}\right)_{\text{center}} = F\left(Y_{1..n}^{(t,t-1)},\ X_{1..m}^{(t,t-1)}\right), \tag{1}$$

where:

- $\left(Y_{1..n}^{t+1}\right)$ denotes the forecasted n wave properties for time $t+1$;
- $Y_{1..n}^{(t,t-1)}$ represents the wave properties at times t and $t-1$;
- $X_{1..m}^{(t,t-1)}$ denotes the meteorological (forcing) input variables.

The adopted temporal window uses time_horizon $= 2$ as a minimal configuration that provided strong performance for one-hour-ahead prediction. Larger temporal windows would increase input dimensionality and computational cost. The exploration of longer sequences was left for future work.

C. Explanation of the N x M Grid Used as Input Structure

Wave dynamics exhibits strong spatial dependence, as wave generation and propagation are influenced by neighbouring atmospheric conditions and fetch. To account for these effects while preserving computational efficiency, a local grid-based methodology was adopted.

As part of this exploratory study, a compact input grid based on the closest neighbourhood was developed and evaluated. For each central spatial point, information from its immediate closest neighbours was considered, forming a $N \times M$ grid. Wave properties were then forecasted one hour ahead using the combined spatio-temporal information from the central point and its surrounding neighbours. In the current implementation, this local neighbourhood is fixed to a 3×3 grid. Larger configurations (e.g., 5×5) are possible but were not explored because the present setup already achieves strong performance.

This design choice aligns with the requirements for future integration into a Digital Twin of the Ocean (DTO) framework. A computationally efficient solution is essential, as multiple models are expected to operate simultaneously within the same environment. Under such constraints, model simplicity and fast inference are critical. Rather than relying on more computationally demanding architectures such as CNN [9] or transformer-based networks [10], this work explores whether a compact neighbourhood-based representation can capture sufficient spatial context to achieve accurate short-term wave forecasts.

In addition, the use of overlapping grids increases the effective number of training samples without increasing model complexity. For instance, a 4×4 spatial region yields four distinct 3×3 grid configurations, while a 5×5 region yields nine such configurations. This strategy improves data efficiency while maintaining a lightweight model structure.

D. Selecting the Study Spacial Region

Before downloading the extended spatial dataset for the full predefined region, we first estimated that 24 latitude rows and 41 longitude columns, with 6624 time samples per point, would result in 6.5 million records in total. Thus, five smaller grids within that area were considered instead. To reduce the bias associated with grid location, four isolated grids were used for training and validation and one for testing, as shown in Figure 1.

E. Dataset generation

The training and validation data were generated by subsampling the original spatio-temporal dataset. Each sample was defined by a central point (latitude, longitude, time), from which a spatial grid and temporal window of the selected variables was extracted, excluding boundary cases. To avoid duplication and potential leakage, each sample is uniquely identified by (latitude, longitude, start_time), and repeated selections are prevented.

From a total of 245,014 valid candidates, 10,000 samples were randomly selected for training and validation. Training and validation samples are drawn from the same four training

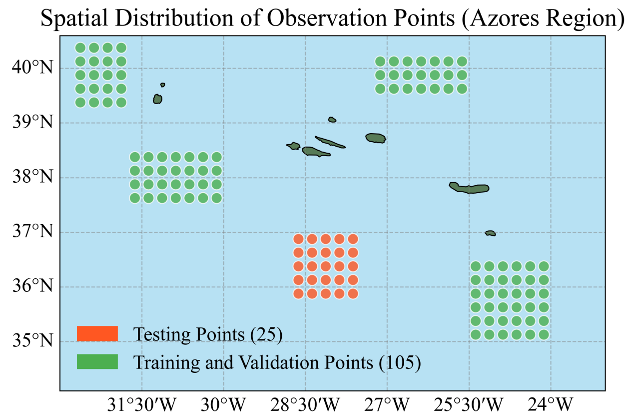

Fig. 1. Representation of the location of the 4 training and validation grids (green) and the 1 test grid (orange).

grids and then randomly split in an $80/20$ ratio. The test set consists of 2,000 samples drawn from a spatially disjoint grid from the same time period, ensuring no overlap with training and validation data.

IV. Architecture Design and Training

This section describes the overall forecasting architecture and the training procedure adopted in this study. Given the relatively high dimensionality of the spatio-temporal input representation, a two-stage learning strategy was employed. First, an autoencoder was used to learn a compact latent representation of the input features. Subsequently, a MLP was trained to forecast wave properties based on this encoded representation. The complete workflow is illustrated in Fig. 2.

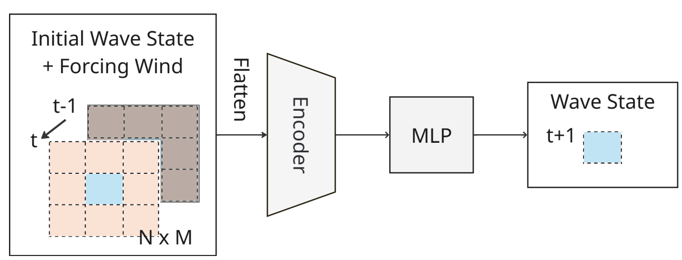

Fig. 2. Overview of the forecasting pipeline, consisting of spatio-temporal input aggregation, autoencoder-based dimensionality reduction and MLP-based wave forecasting.

A. Spatio-Temporal Sample Generation

Training samples were generated by aggregating the predefined training grids. For each sample, a central spatial point was randomly selected at a random time step t. Its eight neighbouring points were then identified, and the corresponding feature values from the two previous time steps (t and $t-1$) were collected. The resulting input vector consists of the concatenated spatio-temporal information from the 3×3 grid over two time steps. The target vector corresponds to the wave properties at the central point at time $t+1$. A memory mechanism was used to avoid duplicate samples during data generation. An analogous procedure was applied to generate the test dataset, using only the spatially disjoint test grid to ensure unbiased evaluation. For a given central point, forecasting the wave properties one time step (1 hour) ahead yields an input layer of $(8+1 \text{ points}) \times (6 \text{ features}) \times (2 \text{ time steps}) = 108$ features.

B. Data Normalization

Prior to training, and to avoid data leakage, the training set was normalized per variable type using a standard scaler, setting the mean to 0 and the standard deviation to 1. This fitted scaler was then applied to the test set, ensuring that normalization parameters were derived exclusively from the training data.

The mean and standard deviation values computed from the normalized test set tend to be close to 0 and 1, respectively, supporting the statistical consistency between the training and test distributions. The remaining differences are expected because the test set is spatially disjoint, while both sets come from the same temporal window, reflecting realistic spatial variability in local wave conditions.

C. Autoencoder-Based Dimensionality Reduction

To address the high dimensionality of the spatio-temporal input representation, a fully connected autoencoder was selected as an unsupervised feature extraction mechanism. The objective of the autoencoder is to learn, through a sequence of non-linear transformations, a compact latent representation that preserves the most relevant information required for subsequent wave forecasting while significantly reducing input size and computational cost.

The autoencoder follows a symmetric encoder-decoder architecture. The input layer maps the original feature space to a hidden layer of 128 neurons, followed by a second hidden layer with 64 neurons and finally to a latent representation of 32 dimensions. Rectified Linear Unit (ReLU) activation functions are applied after each linear transformation to introduce nonlinearity and improve representational capacity. The presented architecture was supported by the evaluation of different hyperparameter sets.

The dataset was split into 80% training and 20% validation (8000 and 2000 samples, respectively) and all features were standardized prior to training. The model was trained using Adam with a learning rate of 0.0005, batch size of 64 and early stopping with patience of 10 epochs. Training converged in 213 epochs, with a best validation loss of 0.001631.

The root mean square error (RMSE) of the test set was 0.0731, indicating accurate reconstruction.

D. Multilayer Perceptron

Following dimensionality reduction using the autoencoder, the 32-dimensional latent representations were used as inputs to the MLP for wave parameter prediction at the central point. The MLP consists of two hidden layers: the first with 256 neurons and the second with 128 neurons, each followed by ReLU activations to introduce nonlinearity. Dropout regularization was applied to mitigate overfitting, with rates of 0.05 in the first layer and 0.025 in the second.

The dataset was again split into 80% training and 20% validation, resulting in 8000 and 2000 samples, respectively. This split is a random partition of samples from the same four training grids, while final testing is performed on the spatially disjoint test grid from the same temporal window. The network was trained using the Adam optimizer with a learning rate of 0.001 and a mean squared error (MSE) loss function. Training proceeded for up to 1000 epochs with early stopping based on validation loss and a batch size of 64. The learning rate was automatically reduced when improvements in validation loss plateaued, allowing the optimizer to converge more effectively and avoid getting trapped in shallow minima. The best model was achieved after 393 epochs, reaching a validation loss of 0.003837, indicating effective learning of the non-linear mapping from encoded atmospheric-oceanic features to the target wave parameters.

This configuration allows the MLP to leverage the compact, informative latent space produced by the autoencoder while maintaining sufficient capacity to model complex relationships in the data.

E. LSTM Baseline Model

As a benchmark, an LSTM baseline was implemented using the same samples and preprocessing pipeline. The input tensor is organized as $(\text{TIME\_HORIZON} = 2, \text{features} = 54)$, with each step containing a flattened 3×3 grid and 6 variables per point ($9 \times 6 = 54$). The LSTM processes the two steps sequentially, and the final hidden output is mapped to the targets by a linear layer. The model was trained for 500 epochs under the same conditions as the proposed framework.

F. Code Availability

To support reproducibility and facilitate further research, the full source code used in this work is publicly available. The implementation includes data preprocessing, model training and evaluation scripts corresponding to the experiments reported in this paper. The repository is hosted on GitHub and can be accessed at [15].

V. Results

A. Evaluation Metrics

Model performance was evaluated using three metrics: the Mean Absolute Error (MAE), RMSE, and the coefficient of determination (R^2). The proposed forecasting model and the LSTM baseline were evaluated for a one-hour prediction horizon, using the same dataset split and preprocessing pipeline, with training conducted on the shared training regions and testing performed on the spatially disjoint test region. Table II presents a side-by-side comparison using MAE, RMSE, and R^2.

To evaluate and compare both models, three consolidated visual analyses were generated: a 2-day sliding forecasting panel with inline errors (Fig. 3), parity plots (Fig. 4), and error distributions (Fig. 5).

TABLE II
SIDE-BY-SIDE TEST SET COMPARISON: PROPOSED MODEL VS. LSTM BASELINE (ONE-HOUR HORIZON)

Target	Proposed Model			LSTM Baseline		
	MAE	RMSE	R^2	MAE	RMSE	R^2
mwd (deg)	1.2765	2.1684	0.9996	3.1527	4.7307	0.9981
mwp	0.0691	0.0943	0.9959	0.2105	0.2845	0.9624
swh	0.0417	0.0591	0.9979	0.1751	0.2330	0.9669

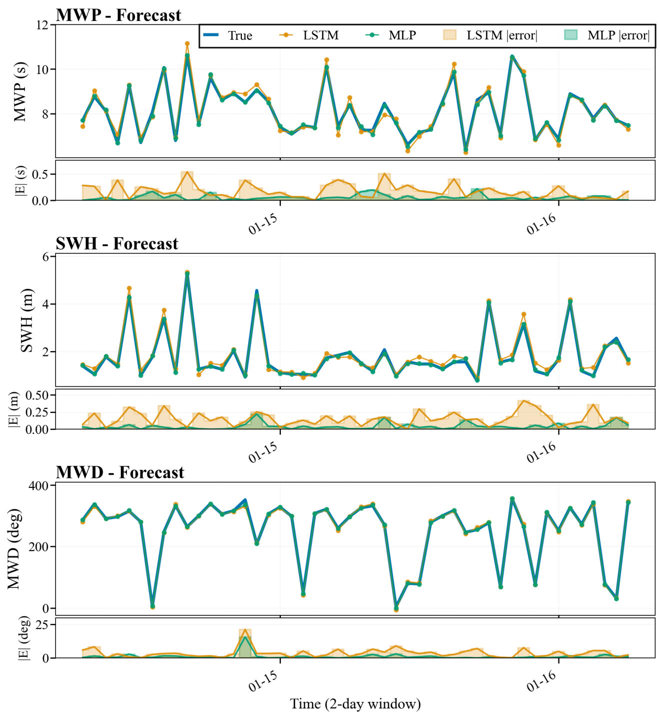

Fig. 3. 2-Day Sliding Forecasting Analysis: Predictions with Inline Error Panels.

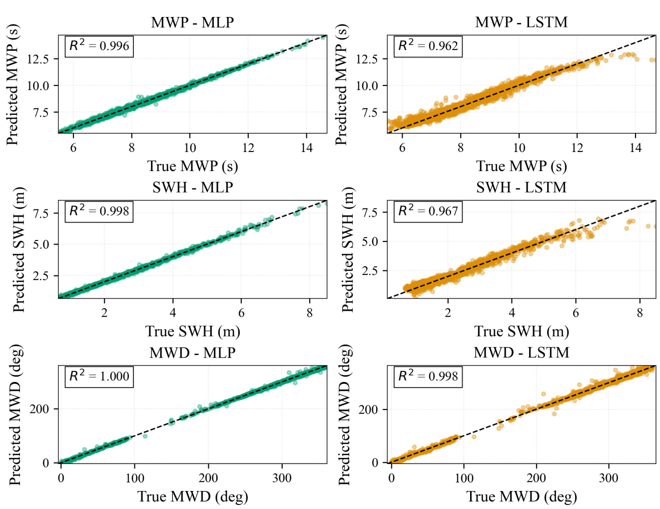

Fig. 4. Parity plots comparing MLP and LSTM predictions for the three target variables.

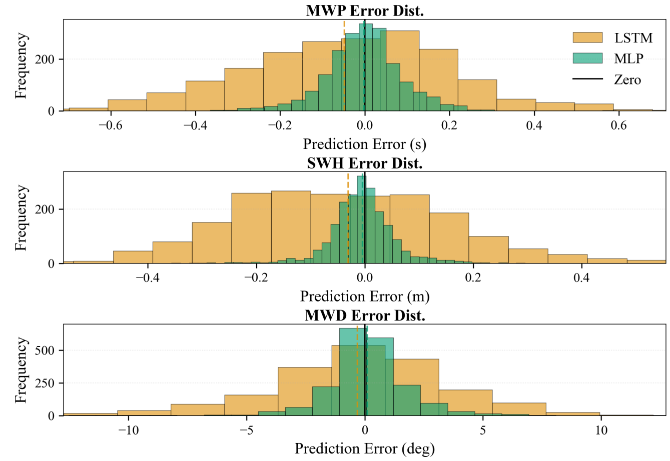

Fig. 5. Error distribution comparison between MLP and LSTM for all variables.

VI. DISCUSSION

The results in Table II, together with Figs. 3–5, indicate that the proposed model delivers high predictive accuracy for one-hour wave forecasting across all target variables, achieving better performance than the LSTM alternative. The MAE and the RMSE are low and the coefficient of determination (R^2) exceeds 0.9959 for all variables, indicating excellent agreement between predicted and true values.

By being trained on data from multiple spatial regions, the model learns wave dynamics in a manner that is not tied to a specific location, resulting in a location-agnostic grid representation. Consequently, the model can be applied to previously unseen regions, as demonstrated by its performance on the test region (different from the four grids used for training), thereby validating its capability to generalize across distinct spatial domains.

In Fig. 3, which presents the 2-day sliding forecasting analysis with predictions and inline error panels, the proposed methodology demonstrates superior performance compared to the LSTM approach. The error panels indicate that the proposed method maintains a lower and more consistent error amplitude, whereas the LSTM exhibits more irregular error patterns. This variability undermines the reliability of a stable confidence interval, as the LSTM alternates between periods of low error and instances of significantly larger deviations.

Using the parity plot in Fig. 4, it is verified that the LSTM performs more poorly during storm events. Due to the unbalanced data distribution (not specifically rebalanced in preprocessing), LSTM was not able to forecast correctly larger waves and the respective larger periods, which is a relevant limitation for storm-oriented applications.

Regarding the error distributions in Fig. 5, both models present approximately normal-shaped distributions, supporting the quality of the data and preprocessing methodology. However, the MLP error is visibly narrower, and its mean is closer to zero, while the LSTM error mean is slightly biased across variables, reinforcing a tendency to underpredict the wave parameters.

The computational efficiency of the proposed MLP-based framework further supports its suitability for deployment in a Digital Twin of the Ocean (DTO) environment. The model comprises only 41,860 trainable parameters, corresponding to a compact memory footprint of approximately 0.16 MB. Inference benchmarking on a standard CPU shows an average latency of 0.22 ms per sample, enabling a throughput of over 4,500 predictions per second. These results indicate that the model can deliver near-instantaneous predictions while requiring minimal computational resources.These characteristics are advantageous in DTO systems, where scalability is essential. The low latency and small memory footprint support efficient integration into real-time or large-scale simulation pipelines.

VII. Conclusions and Future Work

This paper proposes a compact, data-driven framework for short-term wave forecasting that explicitly encodes local spatial neighbourhoods and short-term temporal dynamics through a structured input grid. By combining a 3×3 spatial representation with autoencoder-based dimensionality reduction and a multilayer perceptron, the proposed model achieves high one-hour-ahead predictive accuracy for significant wave height, mean wave period, and mean wave direction.

In addition to the comparison with the LSTM architecture, the inclusion of a benchmark model provides further context for the results, showing that the proposed methodology consistently outperforms standard approaches across all wave parameters. This reinforces that the observed performance gains are not solely due to model complexity, but rather to the effectiveness of the structured spatial-temporal representation.

The results demonstrate that accurate short-horizon wave forecasts can be obtained without relying on large spatial domains or convolutional architectures. Instead, explicitly modelling immediate neighbourhood interactions proves sufficient to capture the dominant local wave-wind interactions. The model achieves coefficients of determination that exceed 0.9959 for all target variables while maintaining a lightweight architecture with fewer than 50,000 trainable parameters, highlighting its suitability for computationally efficient deployment. These findings support the feasibility of structured local grids as an effective alternative to more complex spatial learning strategies for short-term wave prediction, particularly in applications where efficiency, interpretability and scalability are critical.

Several directions for future research arise from this work. First, although the current implementation adopts a fixed 3×3 spatial neighbourhood, larger configurations (e.g., 5×5) are possible and may further improve performance. These alternatives were not explored in this study because the present configuration already achieves strong results, but their evaluation remains an important direction to better understand the trade-off between spatial coverage and model complexity. Second, the adopted temporal window uses a horizon of 2 as a minimal configuration that provided strong performance for one-hour-ahead prediction. Extending this temporal context to longer input sequences may enhance the model's ability to capture more complex temporal dependencies, although this comes at the cost of increased input dimensionality and computational requirements. The systematic exploration of longer temporal windows is therefore left for future work.

Third, the framework will be extended to multi-hour forecasting horizons using an autoregressive formulation, where model outputs are recursively used as inputs. Particular attention will be given to understanding error accumulation and propagation over time, as well as identifying practical limits on forecast lead time for which the methodology remains reliable.

Finally, the methodology will be applied to other datasets to assess its suitability as a surrogate modeling tool for physics-based wave models. In particular, future work will investigate the integration of the proposed framework within SWAN simulations, evaluating its potential to accelerate ensemble-based or high-resolution wave modeling workflows. This extension would further support the use of data-driven surrogates in a digital twin of the ocean.

References

[1] G. J. Komen, L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann and P. A. E. M. Janssen, *Dynamics and Modelling of Ocean Waves*. Cambridge, UK: Cambridge University Press, 1994.

[2] D. Jones, C. Snider, A. Nassehi, J. Yon and B. Hicks, "Characterising the Digital Twin: A systematic literature review," *CIRP Journal of Manufacturing Science and Technology*, vol. 29, pp. 36–52, 2020, doi:10.1016/j.cirpj.2020.02.002.

[3] A. Tzachor, O. Hendel and C. E. Richards, "Digital twins: a stepping stone to achieve ocean sustainability?," *npj Ocean Sustainability*, vol. 2, no. 1, p. 16, 2023, doi:10.1038/s44183-023-00023-9.

[4] S. C. James, Y. Zhang and F. O'Donncha, "A machine learning framework to forecast wave conditions," *Coastal Engineering*, vol. 137, pp. 1–10, 2018.

[5] F. O'Donncha, Y. Zhang, B. Chen and S. C. James, "Ensemble model aggregation for ocean-wave forecasting," *Journal of Marine Systems*, vol. 203, p. 103206, 2020.

[6] C. Gautier, E. de Korte, G. van Hemert and J. van Nieuwkoop, "Confidence intervals for SWAN wave forecasts using a surrogate North Sea wave model to apply wind ensembles," Tech. Rep., Dec. 2024.

[7] J. Chen, A. C. Pillai, L. Johanning and I. Ashton, "Using machine learning to derive spatial wave data: A case study for a marine energy site," *Environmental Modelling & Software*, vol. 142, p. 105066, 2021.

[8] P. Mlakar, A. Ricchi, S. Carniel, D. Bonaldo and M. Ličer, "DELWAVE 1.0: Deep learning surrogate model of surface wave climate in the Adriatic Basin," *Geoscientific Model Development*, vol. 17, no. 12, pp. 4705–4725, 2024.

[9] J. Yang, T. Zhang, J. Zhang, X. Lin, H. Wang and T. Feng, "A ConvLSTM nearshore water level prediction model with integrated attention mechanism," *Frontiers in Marine Science*, vol. 11, 2024.

[10] J. Ji, J. He, M. Lei, M. Wang and W. Tang, "Spatio-temporal transformer network for weather forecasting," *IEEE Transactions on Big Data*, vol. 11, no. 2, pp. 372–387, 2025.

[11] J. Donnelly, A. Daneshkhah and S. Abolfathi, "Physics-informed neural networks as surrogate models of hydrodynamic simulators," *Science of the Total Environment*, vol. 912, p. 168814, 2024.

[12] H. Hersbach *et al.*, "ERA5 hourly data on single levels from 1940 to present," Copernicus Climate Change Service (C3S), ECMWF, 2025. [Online]. Available: https://cds.climate.copernicus.eu

[13] Delft University of Technology, "SWAN scientific and technical documentation: Cycle III, version 41.51," Technical Report, 2024. [Online]. Available: https://swanmodel.sourceforge.io/download/zip/swantech.pdf

[14] S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics*. Singapore: World Scientific, 2001.

[15] B. M. Chagas, "Short-term wave forecasting," GitHub repository, 2026. [Online]. Available: https://github.com/BernardoMChagas/Short-Term_Wave_Forecasting