Machine learning for phase-resolved reconstruction of nonlinear ocean wave surface elevations from sparse remote sensing data

Ehlers, Svenja; Klein, Marco; Heinlein, Alexander; Wedler, Mathies; Desmars, Nicolas; Hoffmann, Norbert; Stender, Merten

DOI
10.1016/j.oceaneng.2023.116059

Publication date
2023

Document Version
Final published version

Published in
Ocean Engineering

Citation (APA)
Ehlers, S., Klein, M., Heinlein, A., Wedler, M., Desmars, N., Hoffmann, N., & Stender, M. (2023). Machine learning for phase-resolved reconstruction of nonlinear ocean wave surface elevations from sparse remote sensing data. *Ocean Engineering*, *288*, Article 116059. https://doi.org/10.1016/j.oceaneng.2023.116059

Ocean Engineering 288 (2023) 116059

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Machine learning for phase-resolved reconstruction of nonlinear ocean wave surface elevations from sparse remote sensing data

Svenja Ehlers [a,*], Marco Klein [b,a], Alexander Heinlein [c], Mathies Wedler [a], Nicolas Desmars [b,a], Norbert Hoffmann [a,d], Merten Stender [e]

[a] Hamburg University of Technology, Dynamics Group, Schloßmühlendamm 30, 21073 Hamburg, Germany
[b] German Aerospace Center, Institute of Maritime Energy Systems, Ship Performance Dep., 21502 Geesthacht, Germany
[c] Delft University of Technology, Delft Institute of Applied Mathematics, 2628 CD Delft, Netherlands
[d] Imperial College London, Department of Mechanical Engineering, London SW7 2AZ, United Kingdom
[e] Technische Universität Berlin, Cyber–Physical Systems in Mechanical Engineering, 10623 Berlin, Germany

ARTICLE INFO

Keywords:
Deep operator learning
Fourier neural operator
Nonlinear ocean waves
Phase-resolved surface reconstruction
X-band radar images
Radar inversion

ABSTRACT

Accurate short-term predictions of phase-resolved water wave conditions are crucial for decision-making in ocean engineering. However, the initialization of remote-sensing-based wave prediction models first requires a reconstruction of wave surfaces from sparse measurements like radar. Existing reconstruction methods either rely on computationally intensive optimization procedures or simplistic modelling assumptions that compromise the real-time capability or accuracy of the subsequent prediction process. We therefore address these issues by proposing a novel approach for phase-resolved wave surface reconstruction using neural networks based on the U-Net and Fourier neural operator (FNO) architectures. Our approach utilizes synthetic yet highly realistic training data on uniform one-dimensional grids, that is generated by the high-order spectral method for wave simulation and a geometric radar modelling approach. The investigation reveals that both models deliver accurate wave reconstruction results and show good generalization for different sea states when trained with spatio-temporal radar data containing multiple historic radar snapshots in each input. Notably, the FNO demonstrates superior performance in handling the data structure imposed by wave physics due to its global approach to learn the mapping between input and output in Fourier space.

1. Introduction

Offshore installations and vessels are strongly impacted by the dynamics of the surrounding ocean waves. Thus, accurate predictions of future wave conditions are desirable for enhancing their safe and efficient operation. For this purpose, several numerical methods have been developed, involving two fundamental steps: the assimilation and reconstruction of initial wave conditions from wave measurement data, followed by the prediction of the future wave evolution. While one line of research focuses on predicting simplified phase-averaged wave quantities based on statistical parameters, marine applications such as wind turbine installations, helicopter landings, or control of wave energy converters require phase-resolved spatio-temporal wave information $\eta(x,t)$ to identify periods of low wave conditions or enable extreme event warnings. The X-band radar is a remote sensing device that can obtain such phase-resolved wave information. However, the radar backscatter is affected by the geometrical mechanism of tilt and shadowing modulation, creating a nonlinear and sparse relationship between radar measurement intensities $\xi(x,t)$ and the actual ocean wave surface elevation $\eta(x,t)$. This makes a reconstruction of wave information from radar information necessary in the assimilation step, which is also referred to as *radar inversion* and is graphically exemplified in Fig. 1.

Contemporary phase-resolved wave reconstruction and prediction methods face a trade-off between accuracy and real-time capability. To achieve computationally efficient methods, linear wave theory (LWT) is commonly employed during the prediction step (cf. Morris et al., 1998; Naaijen and Wijaya, 2014; Hilmer and Thornhill, 2015), along with prior spectral- or texture-analysis-based reconstruction of initial wave conditions from radar data (Borge et al., 2004; Dankert and Rosenthal, 2004). However, these reconstruction methods necessitate additional calibration by wave buoys or rely on simplified assumptions concerning the radar backscatter. Furthermore, the accuracy of the linear approach decreases remarkably for larger temporal horizons of prediction and increasing wave steepness (Lünser et al., 2022), necessitating a wave

* Corresponding author.
E-mail address: svenja.ehlers@tuhh.de (S. Ehlers).

https://doi.org/10.1016/j.oceaneng.2023.116059
Received 17 May 2023; Received in revised form 21 September 2023; Accepted 11 October 2023
Available online 17 October 2023

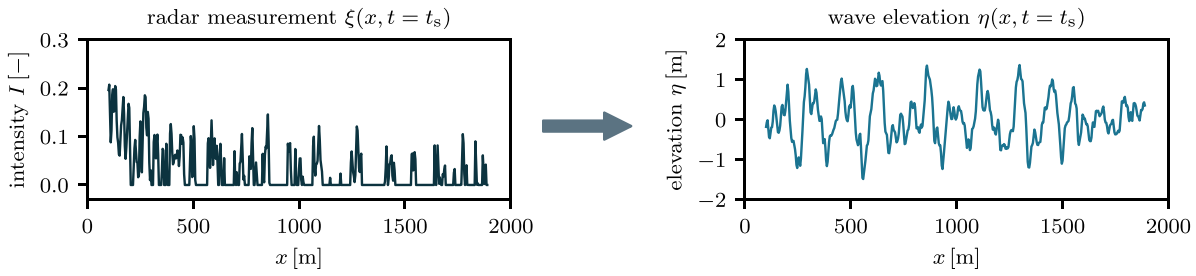

Fig. 1. Graphical illustration of the phase-resolved reconstruction task of ocean wave surfaces η from sparse radar intensity surfaces ξ for the case of waves travelling in one spatial dimension. The radar measurement (left panel) is a snapshot acquired at time instant t_s and is considered as *sparse* due to reoccurring areas with zero intensity caused by the geometrical shadowing modulation. This radar snapshot is used for reconstructing the wave surface elevation at the same time instant t_s (right panel).

prediction using nonlinear wave models, especially for capturing safety-critical events such as rogue waves (Ducrozet et al., 2007; Kharif et al., 2009). Comparative studies on phase-resolved nonlinear ocean wave prediction have demonstrated that the high-order spectral (HOS) method, introduced by West et al. (1987) and Dommermuth and Yue (1987), provides the best prediction accuracy over a wide spatio-temporal domain as well as characteristic wave steepness (Klein et al., 2020; Wu, 2004; Lünser et al., 2022; Blondel-Couprie, 2009). While the HOS prediction step itself is also numerically efficient, the reconstruction step currently represents the weakest part in the entire process (Köllisch et al., 2018): the inversion of initial conditions relies on an optimization procedure of the wave model parameters for the subsequent prediction (Wu, 2004; Blondel-Couprie, 2009), which decreases the possible horizon of prediction and hinders the real-time capability so far (Desmars, 2020). Even though the alternative for the HOS inversion proposed by Köllisch et al. (2018) is able to improve the real-time capability, this method instead assumes an unrealistic radar snapshots data rate Δt_r, making it not suitable for real-world applications (Desmars, 2020).

The aforementioned shortcomings of conventional ocean wave reconstruction and prediction methods have motivated the exploration of alternatives based on machine learning (ML) techniques. For instance, ML methods are able to predict simple phase-averaged wave quantities such as significant wave height H_s, peak period T_p or mean wave direction (cf. Deo et al., 2001; Asma et al., 2012; James et al., 2018; Wu et al., 2020; Yevnin and Toledo, 2022). Recent advancements have also allowed for the more complex task of predicting the spatio-temporal evolution of phase-resolved wave fields, achieved by training multilayer perceptrons (MLPs) (Desouky and Abdelkhalik, 2019; Law et al., 2020; Duan et al., 2020a; Zhang et al., 2022), recurrent neural networks (RNNs) (Kagemoto, 2020; Mohaghegh et al., 2021; Liu et al., 2022), or convolutional neural networks (CNNs) (Klein et al., 2022; Wedler et al., 2023) on synthetic or experimental one-dimensional elevation data. However, these studies presuppose that either temporal sequences of wave elevations can be solely measured at a single point in space by buoys $\eta(x = x_p, t)$ or snapshots of initial wave conditions are available throughout the entire space domain $\eta(x, t = t_s)$. In practice, neither of these assumptions is feasible due to the lack of directional wave information of single-point measurements and the fact that the acquisition of spatial snapshots using remote sensing systems such as radars leads to sparse and unscaled observations $\xi(x, t = t_s)$, requiring a reconstruction of wave surface elevations first.

Consequently, it would be advantageous to employ ML methods also for the phase-resolved reconstruction of wave elevations $\eta(x, t)$ from X-band radar data $\xi(x, t)$. However, as far as the authors are aware, this topic has not yet been addressed. Prior studies have solely focused on reconstructing phase-averaged statistical parameters of the prevailing sea state from radar data. For instance, Vicen-Bueno et al. (2012) and Salcedo-Sanz et al. (2015) improved the estimation of H_s by extracting scalar features from sequences of radar images $\xi(x, t)$ in a preprocessing step, which in turn were employed to train MLPs and support vector regression models. In contrast, Yang et al. (2021) extracted features from each of the consecutive radar images itself for improved H_s estimation at the current time instant. While these methods rely on handcrafted features acquired during a preprocessing step, end-to-end approaches that automatically extract important features from their input have also been proposed. For instance, Duan et al. (2020b) and Chen and Huang (2022) used CNN-based methods to estimate H_s and T_p from radar images.

Although there seems to be no relevant research on ML-based reconstruction of phase-resolved wave surfaces from sparse X-Band radar data, we hypothesize that ML offers a valuable alternative for the radar inversion task (Hypothesis 1). This hypothesis is derived from the observation that the reconstruction of zero-valued areas in the radar input, exemplified in Fig. 1, shares similarities with typical inverse problems encountered in imaging (Bertero et al., 2022; Ongie et al., 2020) such as inpainting and restoration, where ML-methods have demonstrated successful applications (Pathak et al., 2016; Zhang et al., 2017). Two neural network architectures, with network components involving either a local or global approach of data processing, are investigated in detail for their performance in our task. Specifically, we will adapt the U-Net proposed by Ronneberger et al. (2015), a fully convolutional neural network that employs a mapping approach in Euclidean space, and the Fourier neural operator (FNO) proposed by Li et al. (2020), which is designed to learn a more global mapping in Fourier space. Despite the success of CNN-based approaches in imaging problems, we hypothesize that FNO models may be better suited for handling the complex and dynamic nature of ocean waves (Hypothesis 2), since we can assume that the wave features are already explicitly encoded in the network structure, as it learns data patterns in Fourier space. In contrast, the U-Net needs to learn these wave features by aggregating information from multiple layers. Lastly, we expect that incorporating historical context via spatio-temporal radar data will enhance the reconstruction quality of both ML architectures (Hypothesis 3), which we infer from classical radar inversion methods that also rely on temporal sequences of multiple radar snapshots (cf. Dankert and Rosenthal, 2004; Borge et al., 2004).

In general, the fast inference capabilities of trained ML models, make them ideal for maintaining the real-time capability of the entire process composed of wave reconstruction and prediction (Criterion 1) due to the rapid surface reconstruction without particular data preprocessing. Besides the real-time capability, ensuring high reconstruction accuracy is crucial to prevent initial reconstruction errors that will accumulate and deteriorate the subsequent wave prediction. Hence, we strive for an empirical reference value for the surface similarity parameter (SSP) error metric (Perlin and Bustamante, 2014) of less than SSP ≤ 0.10 between ground truth and reconstructed wave surfaces (Criterion 2), which is a commonly used error threshold in ocean wave research (Klein et al., 2020; Lünser et al., 2022). In addition, the proposed ML methods must be capable of handling real-world measurement conditions of radar snapshots taken at intervals of $\Delta t_r = [1, 2]$ s (Criterion 3), a common X-band radar revolution period (Neill and Hashemi, 2018).

To summarize, the objective of this work is to develop an ML-based approach for phase-resolved radar inversion. This involves training ML models to learn mapping functions $\mathcal{M}$ that are able to reconstruct spatial wave elevation snapshots $\eta(x, t = t_s)$ from one or n_s consecutive historical radar snapshots $\xi(x, t_j)$, where $t_j = \{t_s - j\Delta t_r\}_{j=0,\dots,n_s-1}$. As obtaining ground truth wave surface elevation data for large spatial domains in real ocean conditions is almost impractical, we first generate synthetic yet highly realistic one-dimensional spatio-temporal wave surfaces $\eta(x, t)$ using the HOS method for different sea states in Section 2. The corresponding X-band radar surfaces $\xi(x, t)$ are generated using a geometric approach and incorporate tilt- and shadowing modulations. In Section 3, two neural network architectures are introduced, a U-Net-based and FNO-based network, which are investigated for their suitability for radar inversion. In Section 4, we discuss the computational results. In particular, we first compare the wave reconstruction performance of the U-Net-based and the FNO-based models, each trained using either $n_s = 1$ radar snapshot in each input or spatio-temporal input data, meaning that multiple consecutive radar snapshots n_s are provided. Afterwards, the observations are generalized for the entire data set and discussed. Finally, in Section 5, we draw conclusions based on these results and suggest future research directions.

2. Data generation and preparation

This section briefly introduces the generation of long-crested nonlinear synthetic wave data $\eta(x, t)$ using the HOS method, followed by the generation of synthetic radar data $\xi(x, t)$ that accounts for the tilt- and shadowing modulation mechanisms. The final step involves extracting a number of N input–output $(\mathbf{x}_i, \mathbf{y}_i), i = 1, \dots, N$ data samples from the synthetic radar and wave data, which we employ to train the supervised ML models. This first study on ML-based phase-resolved wave reconstruction focuses on the scenario of one-dimensional wave and radar data, driven by the advantages of easier data generation, simplified implementation, and faster neural network training with fewer computational resources.

2.1. Nonlinear synthetic wave data

To generate synthetic one-dimensional wave data, the water-wave problem can be expressed by potential flow theory. Assuming a Newtonian fluid that is incompressible, inviscid, and irrotational, the underlying wave model is described by a velocity potential $\Phi(x, z, t)$ satisfying the *Laplace equation*

$$\nabla^2 \Phi = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial z^2} = 0 \tag{1}$$

within the fluid domain, where $z = 0\,\mathrm{m}$ is the mean free surface with z pointing in upward direction. The domain is bounded by the *kinematic* and *dynamic boundary conditions* at the free surface $\eta(x, t)$ and the *bottom boundary condition* at the seabed at depth d

$$\begin{aligned} \eta_t + \eta_x \Phi_x - \Phi_z &= 0 && \text{on } z = \eta(x,t) \\ \Phi_t + g\eta + \frac{1}{2}\left(\Phi_{xx}^2 + \Phi_{zz}^2\right) &= 0 && \text{on } z = \eta(x,t) \\ \Phi_z &= 0 && \text{on } z = -d. \end{aligned} \tag{2}$$

Solving this system of equations is challenging due to the nonlinear terms in the boundary conditions, which must be satisfied additionally at the unknown free surface $\eta(x, t)$. Even though linear wave theory (Airy, 1849) provides adequate approximations for certain engineering applications, capturing realistic ocean wave effects requires modelling the nonlinear behaviour of surface gravity waves. Thus, we employ the HOS method, as formulated by West et al. (1987), which transforms the boundary conditions to the free surface and expresses them as a perturbation series of nonlinear order M around $z = 0$. In practice, an order of $M \leq 4$ is sufficient for capturing the nonlinear wave effects of interest (Desmars, 2020; Lünser et al., 2022). The HOS simulation is linearly initialized by spatial wave surface elevation snapshots $\eta(x, t_s = 0)$ sampled from the JONSWAP spectrum for finite water depth (Hasselmann et al., 1973; Bouws et al., 1985). The corresponding initial potential is linearly approximated. Subsequently, the initial elevation and potential are propagated nonlinearly in time with the chosen HOS order M. The referred JONSWAP spectrum attains its maximum at a peak frequency ω_p, whereas the peak enhancement factor γ determines the energy distribution around ω_p. The wave frequencies ω are linked to the wavenumbers k by the linear dispersion relation $\omega = \sqrt{gk \cdot \tanh(kd)}$. The relations $\omega = 2\pi/T$ and $k = 2\pi/L$ allow for substituting the peak frequency with a peak period T_p, peak wavelength L_p, or peak wavenumber k_p. Moreover, a dimensionless wave steepness parameter $\epsilon = k_p \cdot H_s/2$ is defined based on the significant wave height H_s. For more details on the HOS simulation, consider the work of Wedler et al. (2023) or Lünser et al. (2022), for example.

In this study, we select a wave domain length of 4000 m, discretized by $n_x = 1024$ grid points, resulting in $\Delta x = 3.906\,\mathrm{m}$. A peak enhancement factor of $\gamma = 3$ is employed to emulate North Sea conditions. The water depth is $d = 500\,\mathrm{m}$ and the sea state parameters peak wavelength L_p and steepness ϵ are varied systematically over $L_p \in \{80, 90, \dots, 190, 200\}\,\mathrm{m}$ and $\epsilon \in \{0.01, 0.02, \dots, 0.09, 0.10\}$, resulting in 130 possible $L_p - \epsilon$-combinations. For each $L_p - \epsilon$-combination, we generate four different initial surfaces $\eta(x, t_s = 0)$ by superimposing the wave components of the JONSWAP spectrum with random phase shifts. The subsequent wave evolution $\eta(x, t > 0)$ for $t = 0, \dots, 50\,\mathrm{s}$ with $\Delta t_{save} = 0.1\,\mathrm{s}$ is performed considering the nonlinearities imposed by HOS order $M = 4$. As a result, we generate a total of 520 unique spatio-temporal HOS wave data arrays, each of shape $E_{HOS} \in \mathbb{R}^{1024\times500}$, where $(E_{HOS})_{kj} = \eta(x_k, t_j)$ with $x_k = \cdot\Delta x$ and $t_j = j \cdot \Delta t_{save}$.

2.2. Corresponding synthetic radar data

As X-band radar systems are often pre-installed on marine structures for navigation and object detection purposes, they also gained attention for observing ocean surface elevations (Borge et al., 1999). The system antenna rotates with a device-specific revolution time Δt_r of between 1–2 s (Neill and Hashemi, 2018) while emitting radar beams along a range r. These radar beams interact with short-scale capillary waves distributed on large-scale ocean surface waves by the Bragg resonance phenomenon, resulting in backscatter to the antenna (Valenzuela, 1978). This procedure provides measurement data $\xi(r, t)$ as a proxy of wave surface elevations $\eta(r, t)$, which are not directly relatable to each other due to the influence of different modulation mechanisms. Most influential are assumed to be *tilt modulation* (Dankert and Rosenthal, 2004), *shadowing modulation* (Borge et al., 2004; Wijaya et al., 2015) or a combination of both (Salcedo-Sanz et al., 2015). In order to generate synthetic radar snapshots for this work, the modulation mechanisms are simulated according to Salcedo-Sanz et al. (2015) and Borge et al. (2004), as illustrated in Fig. 2.

Tilt modulation refers to the variation in radar backscatter intensity depending on the local incidence angle $\tilde{\Theta}(r, t)$ between the unit normal vector $\mathbf{n}(r, t)$ perpendicular to the illuminated wave facet $\eta(r, t)$ and the unit normal vector $\mathbf{u}(r, t)$ pointing towards the antenna. As the backscatter cannot reach the antenna if the dot product $\mathbf{n}\cdot\mathbf{u}$ approaches negative values for $|\tilde{\Theta}| > \frac{\pi}{2}$, the tilt modulation $\mathcal{T}$ is simulated by

$$\mathcal{T}(r, t) = \mathbf{n}(r, t) \cdot \mathbf{u}(r, t) = \cos\tilde{\Theta}(r, t) \qquad \text{if } |\tilde{\Theta}(r, t)| \leq \frac{\pi}{2} \tag{3}$$

The *shadowing modulation* instead occurs when high waves located closer to the antenna obstruct waves at greater distances. Shadowing depends on the nominal incidence angle $\Theta(r, t)$ of a wave facet $\eta(r, t)$ with horizontal distance $R(r)$ from the antenna at height z_a above the mean sea level, geometrically expressed as

$$\Theta(r, t) = \tan^{-1}\left[\frac{R(r)}{z_a - \eta(r, t)}\right]. \tag{4}$$

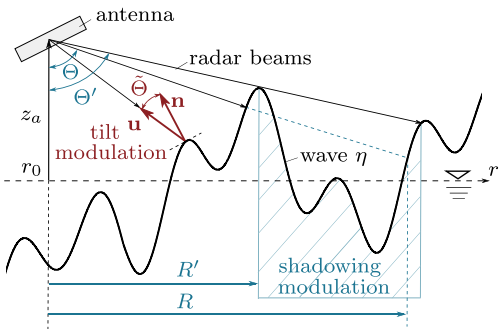

Fig. 2. Geometric display of tilt- and shadowing modulation. Tilt modulation $\mathcal{T}(r,t)$ is characterized by the local incidence angle $\tilde{\Theta}$ between surface normal vector $\mathbf{n}$ and antenna vector $\mathbf{u}$, while shadowing modulation $\mathcal{S}(r,t)$ of a wave facet occurs if another wave closer to the radar systems obstructs the radar beams.

At a specific time instance t, a wave facet $\eta(r,t)$ at point r is shadowed in case there is another facet $\eta' = \eta(r',t)$ closer to the radar $R' = R(r') < R(r)$ that satisfies the condition $\Theta' = \Theta(r',t) \geq \Theta(r,t)$. The shadowing-illumination mask $\mathcal{S}$ can be constructed from this condition as follows

$$\mathcal{S}(r,t) = \begin{cases} 0 & \text{if } R(r') < R(r) \text{ and } \Theta(r',t) \geq \Theta(r,t), \\ 1 & \text{otherwise.} \end{cases} \tag{5}$$

Assuming that tilt-and shadowing modulation contribute to the radar imaging process, the image intensity is proportional to the local radar cross-section, that is $\xi(r,t) \sim \mathcal{T}(r,t) \cdot \mathcal{S}(r,t)$. As marine radars are not calibrated, the received backscatter $\xi(r,t)$ may be normalized to a user-depended range of intensity values.

This work aims to develop a robust ML reconstruction method capable of handling even suboptimal antenna installation conditions. For this reason, we consider a X-band radar system with a comparatively low antenna installation height of $z_a = 18\,\text{m}$. This choice causes an increased amount of shadowing-affected areas in radar images, which can be inferred from Eqs. (4) and (5). Around this antenna exists a system's dead range r_{min} where the radar beams cannot reach the water surface. In this study, we estimate $r_{\text{min}} = 100\,\text{m}$, which is again a comparatively small value and results in the increased magnitude of the tilt modulation influence close to the radar. Moreover, the radar scans the wave surface with a spatial range resolution of $\Delta r = 3.5\,\text{m}$ at $n_r = 512$ grid points. Thus, the maximum observation range is computed as $r_{\text{max}} = 1892\,\text{m}$. The radar revolution period is chosen according to Criterion 3 to be a snapshot each $\Delta t_r = 1.3\,\text{s}$, i.e., $n_t = 38$ radar snapshots for 50 s of simulation time. Using these definitions, we first transform the 520 wave data arrays $E_{\text{HOS}} \in \mathbb{R}^{1024\times500}$ from their HOS grid to the radar system's grid, yielding $E_{\text{sys}} \in \mathbb{R}^{512\times38}$, where $(E_{\text{sys}})_{kj} = \eta(r_k, t_j)$ with $r_k = i \cdot \Delta r$ and $t_j = j \cdot \Delta t_r$. To obtain highly realistic corresponding radar observations, we model tilt modulation $\mathcal{T}(r,t)$ and shadowing modulation $\mathcal{S}(r,t)$, resulting in 520 radar data arrays, each denoted as $Z_{\text{sys}} \in \mathbb{R}^{512\times38}$ with $(Z_{\text{sys}})_{kj} = \xi(r_k, t_j)$.

2.3. Preparation of data for machine learning

To train a supervised learning algorithm, labelled input–output data pairs are required. As visualized in Fig. 3, from each of the 520 generated radar-wave arrays-pairs we extract six radar input snapshots $\mathbf{x}_i$ from the radar surface array Z_{sys} and wave output snapshots $\mathbf{y}_i$ from the wave surface array E_{sys} at six distinct time instances t_s with the largest possible temporal distance. Each output sample $\mathbf{y}_i \in \mathbb{R}^{512\times1}$ contains a single snapshot at time t_s, while each input sample $\mathbf{x}_i \in \mathbb{R}^{512\times n_s}$ can incorporate a number of n_s historical radar snapshots at discrete, earlier times $\{t_s - j \cdot \Delta t_r\}_{j=0,\ldots,n_s-1}$. A single snapshot ($n_s = 1$) at a time t_s can be used as input, however, as we assumed in Hypothesis 3, that larger temporal context may enhance the quality of a network's reconstruction $\hat{\mathbf{y}}_i$. Therefore, the optimal value of n_s is also a subject of investigation as discussed in Sections 4.1.2 and 4.2.2. In total, $N = 6 \cdot 520 = 3120$ input–output data pair samples are generated, each corresponding to a descriptive $L_p - \epsilon$-combination. The data set takes the of shape $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T \in \mathbb{R}^{3120\times512\times n_s}$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]^T \in \mathbb{R}^{3120\times512\times1}$ and is split into 60% training, 20% validation, and 20% test data using a stratified data split w.r.t. the sea state parameters (L_p, ϵ). This ensures an equal representation of each wave characteristic in the resulting subsets, as described in detail in A.

3. Machine learning methodology

The U-Net (Ronneberger et al., 2015) and the Fourier neural operator (FNO) (Li et al., 2020) are neural network architectures for data with grid-like structures such as our radar and wave surface elevation snapshots. Their fundamental difference is the inductive bias encoded by each architecture, which refers to prior assumptions about either the solution space or the underlying data-generating process (Mitchell, 1980; Battaglia et al., 2018). The U-Net is a special type of CNN (LeCun et al., 1989) and imposes an inductive bias by assuming that adjacent data points in Euclidean space are semantically related and learns local mappings between input patches and output features in each layer. This local information is aggregated into more global features due to the utilization of multiple downsampling and convolutional layers. In contrast, the FNO operates under the assumption that the data information can be meaningfully represented in Fourier space. It employs multiple Fourier transformations to learn a mapping between the spectral representation of the input and desired output, directly providing a global understanding of the underlying patterns in the data. This section presents the U-Net- and FNO-based architectures used in our study for radar inversion. In addition, suitable loss and metric functions are introduced for assessing the model's performance.

3.1. U-Net-based network architecture

We first adopt the U-Net concept, originally developed for medical image segmentation by Ronneberger et al. (2015), which has since been applied to a variety of image-to-image translation and surrogate modelling problems, for instance by Isola et al. (2016), Liu et al. (2018), Stoian et al. (2019), Wang et al. (2020), Eichinger et al. (2022), Niekamp et al. (2023) and Stender et al. (2023). The mirrored image dimensions in a fully convolution autoencoder network allow for the U-Net's key property, that is the use of skip-connections for concatenating the output features from the encoding path with the inputs in the decoding path. This enables the reuse of data information of different spatial scales that would otherwise be lost during downsampling and assists the optimizer to find the minimum more efficiently (Li et al., 2018).

Our proposed encoder–decoder architecture is the result of a four-fold cross-validated hyperparameter study, documented in Table A.2 in the appendix. As depicted in Fig. 4, the adapted U-Net architecture, has a depth of $n_d = 5$ consecutive encoder blocks followed by the same number of consecutive decoder blocks with skip-connections between them.

In more detail, each encoder block in our U-Net-based architecture is composed of a 1D convolutional layer with $n_k = 32$ kernels of size $s_k = 5$, that are responsible for identifying specific features in the input by shifting the smaller-sized kernels, containing the networks trainable weights, across the larger input feature maps in a step-wise manner. Each convolutional layer is followed by a GeLU activation function σ (Hendrycks and Gimpel, 2016) and an average pooling downsampling layer of size 2. To summarize, in the encoding path each radar input sample $\mathbf{x}_i \in \mathbb{R}^{n_r\times n_s}$ is transformed by the first convolutional layer resulting in $v_{c1} \in \mathbb{R}^{n_r\times n_k}$, with $n_r = 512$ being the number of spatial grid-points and n_s being the historic snapshots in the radar input. Subsequently, this intermediate output is send through σ, before the

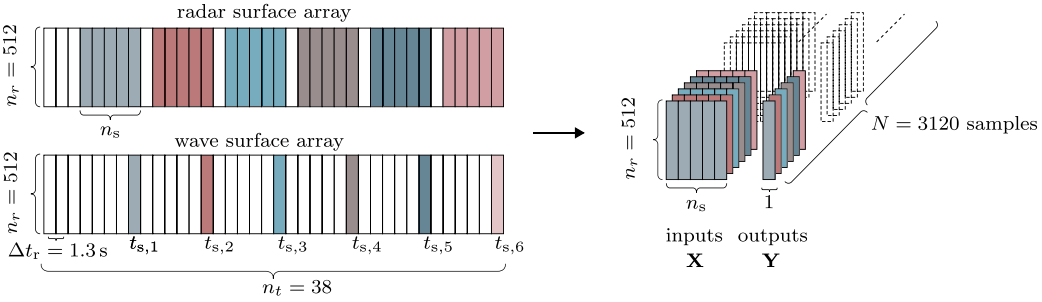

Fig. 3. Schematic representation of the ML training sample extraction process. The left-hand side illustrates one of the raw radar and wave surface simulations ($Z_{sys}, E_{sys} \in \mathbb{R}^{512\times38}$), which are utilized to extract input–output samples shown on the right-hand side. Each input $\mathbf{x}_i$ consists n_s radar snapshots acquired at intervals of $\Delta t_r = 1.3\,\mathrm{s}$, while each output $\mathbf{y}_i$ represents a single-snapshot wave surface elevation at time instant t_s. In total $N = 6 \cdot 520 = 3120$ data samples are generated.

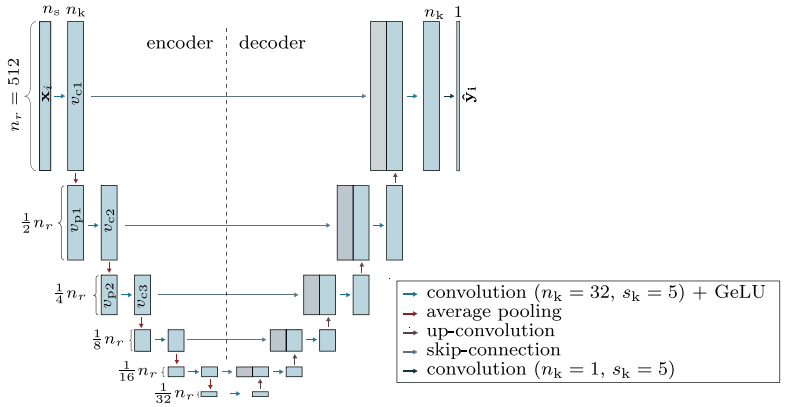

Fig. 4. Fully convolutional encoder–decoder architecture based on the U-Net (Ronneberger et al., 2015). Each input $\mathbf{x}_i$ is processed by $n_d = 5$ alternating convolutional-, activation- and average pooling layers in the encoding path. The decoding path contains convolutional-, activation- and transpose convolutional layers for a gradual upsampling to calculate the output $\hat{\mathbf{y}}_i$. Moreover, the outputs of the encoding stages are transferred to the decoding path via skip-connections.

pooling layer reduces the spatial dimension to $v_{p1} \in \mathbb{R}^{\frac{1}{2}n_r \times n_k}$. This process is repeated until the final encoding block's output is $v_{p5} \in \mathbb{R}^{\frac{1}{32}n_r \times n_k}$. Next, the decoding blocks are applied, each consisting of a convolutional layer with again $n_k = 32$ kernels of size $s_k = 5$, followed by GeLU activation. Afterwards, the feature maps' spatial dimensions are upsampled using transpose convolutional layers with linear activation. The resulting feature maps then are concatenated with the output of the corresponding stage in the encoding path via skip-connections, before the next convolution is applied. This process is repeated until the final wave output $\hat{\mathbf{y}}_i \in \mathbb{R}^{n_r \times 1}$ is calculated using a convolutional layer with a single kernel and linear activation.

As indicated above, the U-Net architecture assumes local connections between neighbouring data points, which is accomplished through two mechanisms. Firstly, the convolutional layers use kernels with a receptive field of $s_k = 5$ pixels to process different local parts of the larger input feature maps in the same manner. This is referred to as weight sharing, causing a property called *translational equivariance*: each patch of the input is processed by the same kernels. Secondly, the pooling layers induce locality by assuming that meaningful summations of information from small local regions in the intermediate feature maps can be made and creates a property referred to as *translational invariance* (Goodfellow et al., 2016).

3.2. FNO-based network architecture

In the second step, we explore a neural network based on the FNO (Li et al., 2020). While a CNN is limited to map between finite-dimensional spaces, neural operators are in addition capable to learn nonlinear mappings between a more general class of function spaces. This makes the FNO well-suited for capturing the spatio-temporal patterns that govern the dynamics of various physical problems that obey partial differential equations if the solutions are well represented in Fourier space. FNO variants have been applied to e.g., fluid dynamics (Peng et al., 2022; Li et al., 2022), simulation of multiphase flow (Yan et al., 2022; Wen et al., 2022), weather forecasting (Pathak et al., 2022), material modelling (Rashid et al., 2022; You et al., 2022), and image classification (Williamson et al., 2022).

The FNO-based iterative architecture ($\mathbf{x}_i \to v_0 \to v_1 \to \ldots \to \hat{\mathbf{y}}_i$) applied in this work is illustrated in Fig. 5, while Table A.3 in the appendix summarizes the determination of model hyperparameters by four-fold cross-validation.

The proposed FNO transforms radar input data $\mathbf{x}_i \in \mathbb{R}^{n_r \times n_s}$ into a higher-dimensional latent representation $v_0 \in \mathbb{R}^{n_r \times n_w}$ of channel width $n_w = 32$, using a linear neural network layer P with n_w nodes. Subsequently, the latent representation passes through $n_f = 3$ Fourier layers, each consisting of two paths. In the upper path, a global convolution operator defined in Fourier space is applied to each channel of v_0 separately utilizing discrete Fourier transforms F. A linear transformation R_0 is then applied to the lower-order Fourier modes after truncating the Fourier series at a maximum number of $n_m = 64$ modes. Subsequently, this scaled and filtered content is back-transformed to the spatial domain using inverse discrete Fourier transforms F^{-1}. In the lower path, a linear transformation W_0 in the spatial domain is applied to the input v_0 to account for non-periodic boundary conditions and higher-order modes that are neglected in the upper path of the Fourier layer. The outputs of the upper and lower paths are added, and the sum is passed through a nonlinear GeLU activation σ resulting in $v_1 \in \mathbb{R}^{n_r \times n_w}$, before entering the next Fourier layer. In summary, the output of the $(j+1)$-th Fourier layer is defined as

$$v_{j+1} = \sigma\left(F^{-1}\left(R_j \cdot F(v_j)\right) + W_j \cdot v_j\right). \tag{6}$$

Finally, the output v_3 of the last Fourier layer is transferred to the target wave output dimension $\hat{\mathbf{y}}_i \in \mathbb{R}^{n_r \times 1}$ using another linear layer Q. In

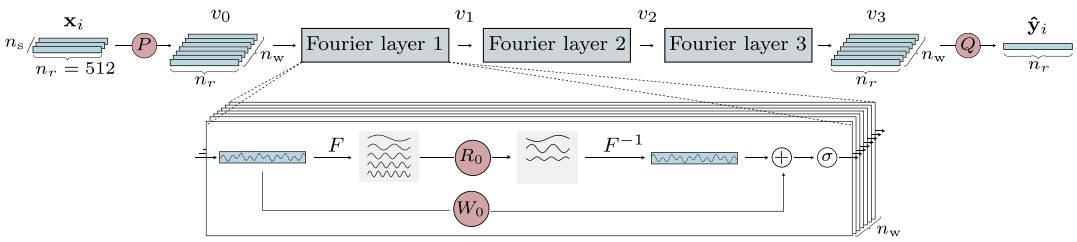

Fig. 5. Network architecture based on the Fourier neural operator (Li et al., 2020). Each input $\mathbf{x}_i$ is lifted to a higher dimensional representation v_0 of channel width d_w by a neural network P. Afterwards, $n_f = 3$ Fourier layers are applied to each channel. Finally, v_3 is transferred back to the target dimension of the output $\hat{\mathbf{y}}_i$ by another neural network Q. More specifically, each Fourier layer is composed of two paths. The upper one learns a mapping in Fourier space by adapting R_j for scaling and truncating the Fourier Series after n_m modes, while the lower one learns a local linear transform W_j.

summary, the FNOs weights correspond to $P \in \mathbb{R}^{n_s \times d_w}$, $Q \in \mathbb{R}^{n_s \times d_w}$ and all $R_j \in \mathbb{C}^{d_w \times d_w \times d_m}$ and $W_j \in \mathbb{R}^{d_w \times d_w}$. As the R_j-matrices contain the main portion of the total number of weights, most parameters are learned in the Fourier space rather than the original data space.

As previously noted, the FNO architecture incorporates a global inductive bias that assumes the input data exhibits approximately periodic properties and can be effectively represented in Fourier space. Furthermore, the FNO's design presupposes that the Fourier spectrum of the input data is smooth, enabling its frequency components to be represented by a limited number of low-wavenumber Fourier coefficients, as the R_j matrices, which are responsible for the global mapping, truncate higher-frequency modes.

3.3. Training and evaluation

Both the U-Net- and FNO-based architecture are implemented using the PyTorch library (Paszke et al., 2019). To enable a fair comparison and account for wave training data of varying spatial scales, the mean of the relative L2-norm of the error is employed as loss function $\mathcal{L}$ for both architectures. The relative L2-norm error for one sample i is defined as follows, where $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i \in \mathbb{R}^{512\times 1}$ represent the true and reconstructed wave surface

$$\mathrm{nL2}(\mathbf{y}_i, \hat{\mathbf{y}}_i) = \mathrm{nL2}_i = \frac{\|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2}{\|\mathbf{y}_i\|_2}. \tag{7}$$

While we use the subscript i to represents a sample-specific error $\mathrm{nL2}_i$, the value nL2 without a subscript denotes the mean value across a number of samples N, for example the mean error across the training set $\mathcal{L} := \mathrm{nL2} = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \mathrm{nL2}(\mathbf{y}_i, \hat{\mathbf{y}}_i)$. To minimize the loss, we use the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.001. The training is executed for 800 epochs on an NVIDIA GeForce RTX 3050 Ti Laptop GPU. For both, the U-Net-based models $\mathcal{M}_{\mathrm{U},n_s}$ and FNO-based models $\mathcal{M}_{\mathrm{F},n_s}$, only the models with the lowest test loss within the 800 epochs is stored for performance evaluation and visualization.

Established machine learning metrics based on Euclidean distances treat the deviation of two surfaces in frequency or phase as amplitude errors (Wedler et al., 2022). Therefore, we introduce the surface similarity parameter (SSP) proposed by Perlin and Bustamante (2014) as an additional performance metric

$$\mathrm{SSP}(\mathbf{y}_i, \hat{\mathbf{y}}_i) = \mathrm{SSP}_i = \frac{\sqrt{\int |F_{\mathbf{y}_i}(k) - F_{\hat{\mathbf{y}}_i}(k)|^2 dk}}{\sqrt{\int |F_{\mathbf{y}_i}(k)|^2 dk} + \sqrt{\int |F_{\hat{\mathbf{y}}_i}(k)|^2 dk}} \in [0, 1], \tag{8}$$

where k denotes the wavenumber vector and $F_{\mathbf{y}_i}$ denotes the discrete Fourier transform of a surface $\mathbf{y}_i$. The SSP is a normalized error metric, with $\mathrm{SSP}_i = 0$ indicating perfect agreement and $\mathrm{SSP}_i = 1$ a comparison against zero or of phase-inverted surfaces. As the SSP combines phase-, amplitude-, and frequency errors in a single quantity, it is used in recent ocean wave prediction and reconstruction studies by Klein et al. (2020, 2022), Wedler et al. (2022, 2023), Desmars et al. (2021, 2022) and Lünser et al. (2022).

While metrics such as the $\mathrm{nL2}_i$ or SSP_i evaluate the average reconstruction quality of each $\hat{\mathbf{y}}_i \in \mathbb{R}^{n_r \times 1}$ across the entire spatial domain r with $n_r = 512$ grid points, it is important to consider the potential imbalance in reconstruction error between those areas where the radar input $\mathbf{x}_i$ was either shadowed or visible. This imbalance ratio can be quantified by $\frac{\mathrm{nL2}_{\mathrm{shad}_i}}{\mathrm{nL2}_{\mathrm{vis}_i}}$. Here, $\mathrm{nL2}_{\mathrm{shad}_i} = \mathrm{nL2}(\mathbf{y}_{\mathrm{shad}_i}, \hat{\mathbf{y}}_{\mathrm{shad}_i})$ and $\mathrm{nL2}_{\mathrm{vis}_i} = \mathrm{nL2}(\mathbf{y}_{\mathrm{vis}_i}, \hat{\mathbf{y}}_{\mathrm{vis}_i})$ are the errors of the output wave elevations in the shadowed or visible areas, respectively. We separate the visible and shadowed parts using the shadowing mask S introduced in Eq. (5), where $\mathbf{y}_{\mathrm{vis}_i} = S \cdot \mathbf{y}_i$ and $\mathbf{y}_{\mathrm{shad}_i} = (1 - S) \cdot \mathbf{y}_i$. Afterwards, all cells with zero entries are removed from the output arrays, such that the number of visible or invisible data points is n_{vis_i} or n_{shad_i}, respectively, and $\mathbf{y}_{\mathrm{vis}_i}, \hat{\mathbf{y}}_{\mathrm{vis}_i} \in \mathbb{R}^{n_{\mathrm{vis}_i} \times 1}$ and $\mathbf{y}_{\mathrm{shad}_i}, \hat{\mathbf{y}}_{\mathrm{shad}_i} \in \mathbb{R}^{n_{\mathrm{shad}_i} \times 1}$ satisfy $n_{\mathrm{vis}_i} + n_{\mathrm{shad}_i} = n_r = 512$. To conclude, a high value of the ratio indicates that the reconstruction in areas that were shadowed in the input is much worse than in the visible areas. We thus not only strive for low $\mathrm{nL2}_i$ values, but also for low $\frac{\mathrm{nL2}_{\mathrm{shad}_i}}{\mathrm{nL2}_{\mathrm{vis}_i}}$ ratios to achieve uniform reconstructions. We use a ratio metric only based on the Euclidean distance based $\mathrm{nL2}_i$ and not for the SSP_i, as small sections of $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i$ cannot be meaningfully considered in Fourier space.

Table 1
Reconstruction results averaged across the entire test set evaluated with different metrics for the U-Net-based models $\mathcal{M}_{\mathrm{U},n_s}$ and FNO-based models $\mathcal{M}_{\mathrm{F},n_s}$ trained with either one or multiple radar snapshots n_s in each sample's input.

Model					Mean errors across $N_{\text{test}} = 624$ test set samples		
Name	Architecture	n_s	Epochs	Investigated in	nL2	$\frac{\mathrm{nL2}_{\mathrm{shad}}}{\mathrm{nL2}_{\mathrm{vis}}}$	SSP
$\mathcal{M}_{\mathrm{U},1}$	U-Net-based	1	150	Section 4.1.1	0.329	2.679	0.171
$\mathcal{M}_{\mathrm{U},10}$	U-Net-based	10	592	Section 4.1.2	0.123	1.755	0.061
$\mathcal{M}_{\mathrm{F},1}$	FNO-based	1	721	Section 4.2.1	0.242	1.886	0.123
$\mathcal{M}_{\mathrm{F},9}$	FNO-based	9	776	Section 4.2.2	0.153	1.381	0.077

4. Results

This work explores the potential of utilizing machine learning for the reconstruction of one-dimensional ocean wave surfaces η from radar measurement surfaces ξ at a time instance t_s. Therefore, each radar input sample $\mathbf{x}_i \in \mathbb{R}^{n_r \times n_s}$, with $n_r = 512$ being the number of spatial grid points in range direction and n_s being the number of radar snapshots, is acquired according to Section 2. Each input $\mathbf{x}_i$ is to be mapped to the desired wave surface output $\mathbf{y}_i \in \mathbb{R}^{n_r \times 1}$ via a ML model $\mathcal{M}$. We examine the impact of the inductive bias of the U-Net-based models $\mathcal{M}_{\mathrm{U},n_s}$ and the FNO-based models $\mathcal{M}_{\mathrm{F},n_s}$ proposed in Section 3, as well as the impact of the number of historical radar snapshots n_s included in each input $\mathbf{x}_i$. We train the models using a total data set of $N_{\text{train}} = 2496$ samples and thus to learn the mapping $\mathcal{M} : \mathbf{X} \to \mathbf{Y}$ with $\mathbf{X} \in \mathbb{R}^{2496 \times n_r \times n_s}$, $\mathbf{Y} \in \mathbb{R}^{2496 \times n_r \times 1}$. Afterwards, we evaluate their performance using the previously excluded test set of $N_{\text{test}} = 624$ samples. The results are summarized in Table 1, and are discussed regarding the pre-stated Hypothesis 1–Hypothesis 3 and Criterion 1–Criterion 3 in detail in the subsequent subsections.

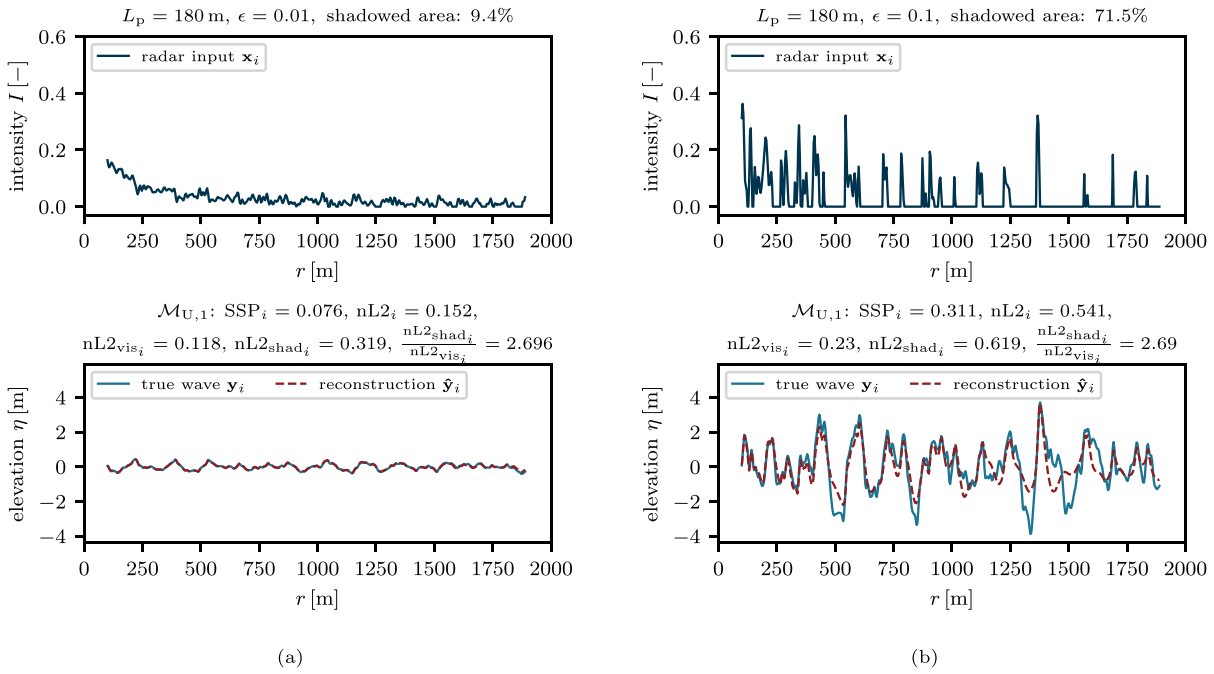

Fig. 6. Two samples from the test set described by the same wavelength $L_p = 180\,m$, but different steepness ϵ, reconstructed by the U-Net-based architecture $\mathcal{M}_{U,1}$. (a) Small ϵ cause minor impact of the shadowing modulation and allow accurate reconstructions. (b) Larger ϵ create more extensive shadowed areas and cause higher reconstruction errors.

4.1. Performance of the U-Net-based model

In the first step of our investigation, we examine the ability of U-Net-based models $\mathcal{M}_{U,n_s}$ to reconstruct wave surfaces along the full spatial dimension, which covers $r_{max} - r_{min} = 1792\,m$ on $n_r = 512$ grid points. We use the $N_{train} = 2496$ samples of single snapshot ($n_s = 1$) radar input data for training. Afterwards, we utilize the same architecture to determine the best number of historical snapshots $n_s > 1$ required in the radar inputs to achieve the best reconstruction performance. We also visually compare reconstructed wave elevations $\hat{\mathbf{y}}_i$ of two test set samples with their corresponding true elevations $\mathbf{y}_i$.

4.1.1. U-Net using single-snapshot radar data

Mapping of single snapshot radar data ($n_s = 1$) refers to mapping a radar snapshot $\mathbf{x}_i \in \mathbb{R}^{n_r \times 1}$ to a wave snapshot $\mathbf{y}_i \in \mathbb{R}^{n_r \times 1}$, with $n_r = 512$ spatial grid points, that are recorded at the same time instant t_s. According to Table 1 the U-Net-based model $\mathcal{M}_{U,1}$ trained with the available $N_{train} = 2496$ samples achieves a reconstruction performance given by a mean loss value of nL2 = 0.329 across all $N_{test} = 624$ test set samples after 150 epochs of training. Afterwards, the model tends to overfit the training data, as shown in the loss curve in Fig. B.15(a). The observed error corresponds to a mean value of SSP = 0.171 across all test set samples, which fails to satisfy the Criterion 2 of reconstruction errors below SSP ≤ 0.1.

To identify the origin of reconstruction errors, we employed model $\mathcal{M}_{U,1}$ to generate reconstructions $\hat{\mathbf{y}}_i$ for two exemplary radar input samples $\mathbf{x}_i$ from the test set. Despite the stratified data split ensuring an equal distribution of sea state parameter combinations (L_p, ϵ) in the training and test set, the errors are unevenly distributed across individual samples i, as exemplarily illustrated in Fig. 6: The sample in Fig. 6(a) corresponds to a peak wavelength $L_p = 180\,m$ and small amplitudes caused by a small steepness of $\epsilon = 0.01$. It exhibits a minor impact from the shadowing modulation mechanism only affecting 9.4% of the total radar-illuminated surface $\mathbf{x}_i$ in the top panel. The corresponding surface reconstruction $\hat{\mathbf{y}}_i$ generated by $\mathcal{M}_{U,1}$ in the bottom panel closely approximates the true wave elevation $\mathbf{y}_i$, as evidenced by the sample-specific error of $nL2_i = 0.152$ or $SSP_i = 0.076$. In contrast, the second sample in Fig. 6(b) with the same $L_p = 180\,m$ but increased $\epsilon = 0.10$ shows 71.5% of the spatial r-domain being affected from shadowing modulation causing zero-valued intensities. This results in a high reconstruction error of $nL2_i = 0.541$ or $SSP_i = 0.311$. Particularly the shadowed areas seem to contribute to the poor reconstruction, as their error is $\frac{nL2_{shad_i}}{nL2_{vis_i}} = 2.69$ times higher than in the visible areas.

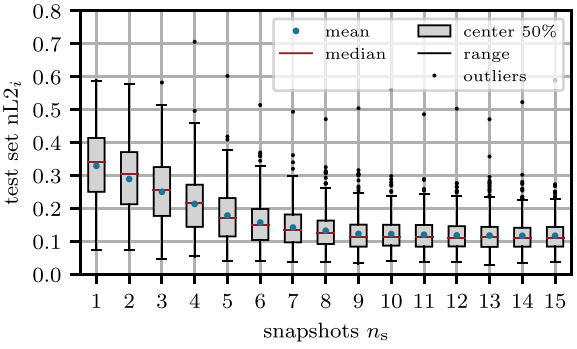

Fig. 7. Boxplot depicting the error distribution on test set, depending on the number of historical radar snapshots n_s provided to train U-Net-based architectures $\mathcal{M}_{U,n_s}$. The best model performance is achieved for $n_s = 10$.

4.1.2. U-Net using spatio-temporal radar data

To improve the reconstruction quality of the U-Net-based architecture, especially for high wave steepness, we took inspiration from classical spectral-analysis- and optimization-based reconstruction approaches (cf. Borge et al., 2004; Wu, 2004). These approaches use spatio-temporal radar data by considering temporal sequences of n_s historical radar snapshots for reconstruction. Thus, we use multiple historical radar snapshots n_s that satisfy Criterion 3 with $\Delta t_r = 1.3\,s$ for each input sample $\mathbf{x}_i \in \mathbb{R}^{512 \times n_s}$, while the outputs remain single snapshots $\mathbf{y}_i \in \mathbb{R}^{n_r \times 1}$ at the respective last time instant t_s. We conducted 14 additional training runs of the same architecture using all $N_{train} = 2496$ samples of the training set with increasing n_s in the inputs $\mathbf{x}_i$, to determine the best number of snapshots n_s in the boxplot in Fig. 7.

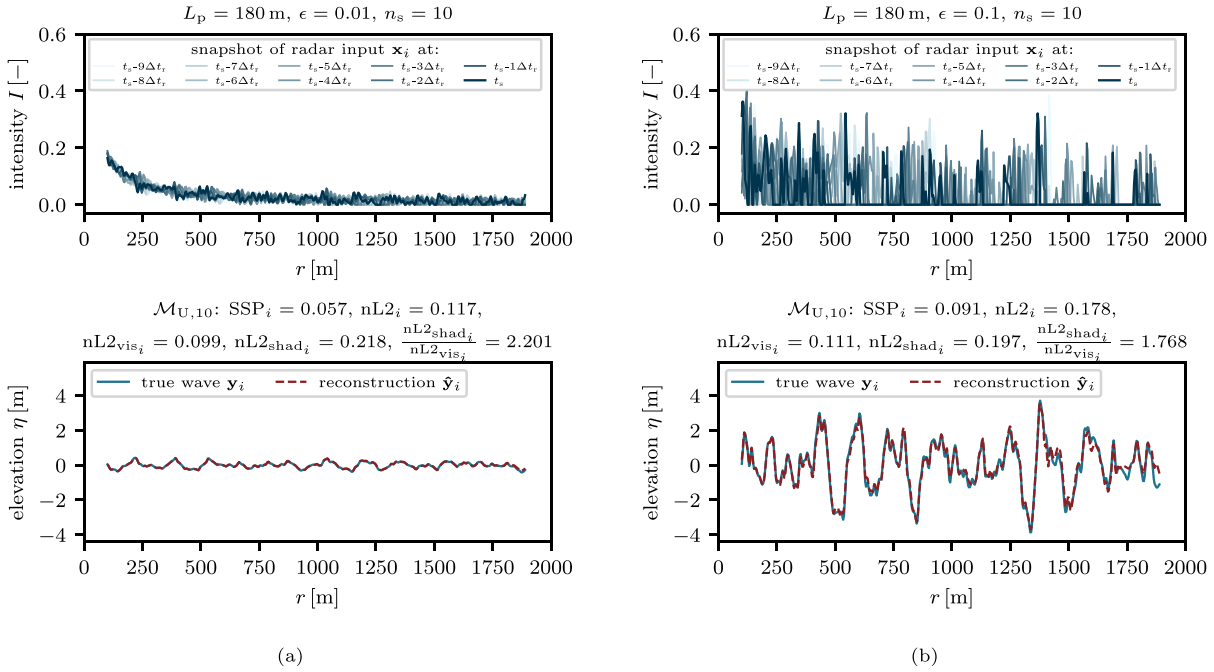

Fig. 8. Two samples from the test set described by the same wavelength $L_p = 180\,\mathrm{m}$, but different wave steepness ϵ, reconstructed by the U-Net-based architecture trained with $n_s = 10$ historical snapshots in the radar input $\mathcal{M}_{U,10}$. Compared to $\mathcal{M}_{U,1}$, a strong reconstruction improvement is observed, especially for the sample with high $\epsilon = 0.10$ in (b).

The boxplot shows that the model's mean performance across the entire test set significantly improves up to a value of $n_s = 10$, confirming Hypothesis 3 as the reconstruction quality improves by incorporating multiple radar snapshots in the input. Moreover, the sample-specific error values $nL2_i$ become less scattered around the mean value. The model $\mathcal{M}_{U,10}$, determined by the boxplot analysis, achieves a final mean reconstruction performance of $nL2 = 0.123$ or $SSP = 0.061$ across the $N_{test} = 624$ test set samples, as shown in Table 1, now satisfying Criterion 2 of $SSP \leq 0.10$ and thus confirms Hypothesis 1. In addition, it yields a lower ratio of $\frac{nL2_{shad}}{nL2_{vis}} = 1.755$ compared to 2.679 for model $\mathcal{M}_{U,1}$, indicating a more balanced reconstruction between shadowed and visible areas on average. Moreover, the model does not exhibit early overfitting anymore, achieving the best performance after 592 epochs, shown in Fig. B.15(b).

Fig. 8 further confirms the improvement of the reconstruction, using $n_s = 10$ radar snapshots in each input to train $\mathcal{M}_{U,10}$, by depicting the same two exemplary test set samples reconstructed by $\mathcal{M}_{U,1}$ in Fig. 6 before. The top panels display the most recent (t_s) radar snapshot present in $\mathbf{x}_i \in \mathbb{R}^{n_r \times n_s}$ in the darkest shading and preceding snapshots at $t_j = \{t_s - j\Delta t_r\}_{j=0,\dots,n_s-1}$ in increasingly lighter shades. Compared to Fig. 6, the sample with small $\epsilon = 0.01$ in Fig. 8(a) experiences only a slight reduction in reconstruction error, while the sample with $\epsilon = 0.10$ in Fig. 8(b) exhibits a substantial reduction around one-third of the previous sample-specific $nL2_i$ or SSP_i value. The improved performance seems mainly attributable to the enhanced reconstruction of shadowed areas.

4.2. Performance of the FNO-based model

The U-Net-based model $\mathcal{M}_{U,10}$ already supported Hypothesis 1 and Hypothesis 3 by demonstrating the potential to reconstruct wave surface elevations from radar data in general and improving the reconstruction quality by including additional historical radar data in the input. However, we also hypothesized that the FNO-based architecture may outperform CNN-based methods, such as the U-Net, due to its global inductive bias (Hypothesis 2), which may be beneficial for the wave data structure. To investigate this, we again use the entire set of $N_{train} = 2496$ samples to train FNO-based models $\mathcal{M}_{F,n_s}$ with $n_s = 1$ radar snapshot in each of the inputs $\mathbf{x}_i \in \mathbb{R}^{n_r \times n_s}$ first. Subsequently, we determine the number $n_s > 1$ to achieve the best reconstruction performance. Both investigations again are conducted on the entire domain of 1792 m ($n_r = 512$) and we compare true and reconstructed elevations $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i \in \mathbb{R}^{n_r \times 1}$ of two exemplary samples.

4.2.1. FNO using single-snapshot radar data

The FNO-based model $\mathcal{M}_{F,1}$ trained with $n_s = 1$ snapshot in each input, attains its best performance $nL2 = 0.240$ after 721 training epochs, as shown Table 1 and demonstrated in the loss curve in Fig. B.16(a). Although the corresponding mean $SSP = 0.123$ across all $N_{test} = 624$ samples in the test set does not attain the Criterion 2, the error still presents a notable improvement compared to the SSP value of 0.171 previously obtained by the U-Net-based model $\mathcal{M}_{U,1}$. Moreover, $\mathcal{M}_{F,1}$ not only reduces the mean nL2 or SSP error but also reconstructs the waves more uniformly between shadowed and visible areas compared to $\mathcal{M}_{U,1}$. This is evident by the decrease in the mean $\frac{nL2shad}{nL2vis}$-ratio from 2.679 to 1.886.

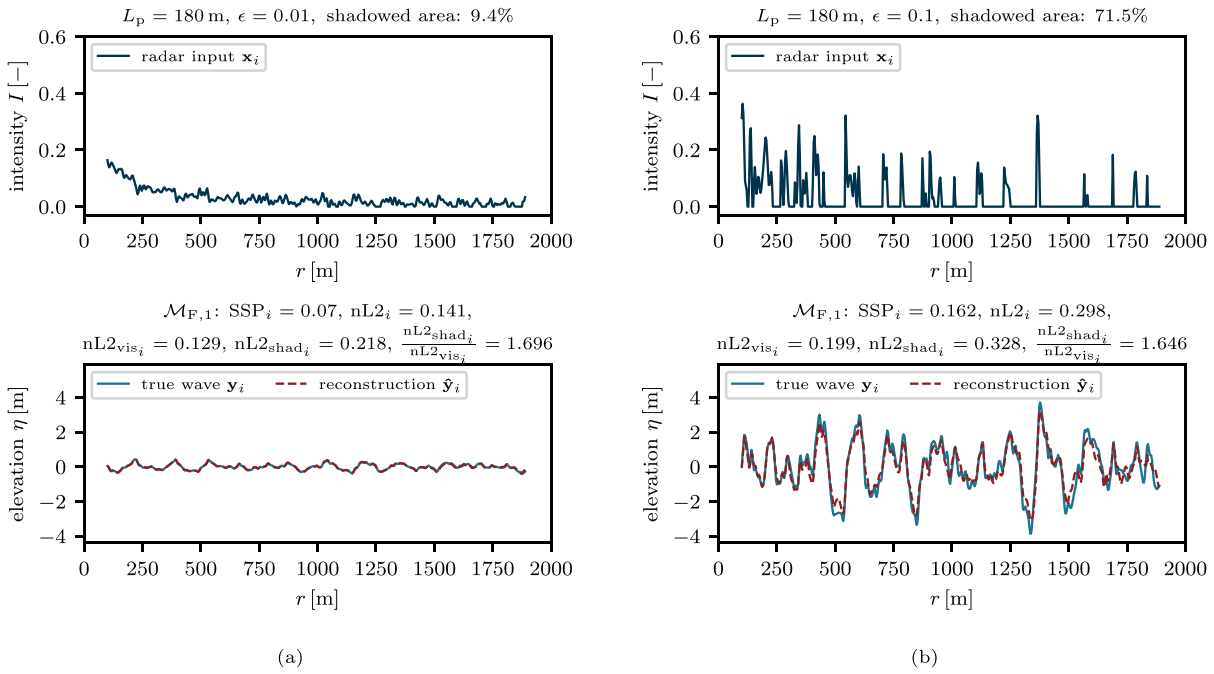

Fig. 9. Two samples from the test set described by the same wavelength $L_p = 180\,m$, but different wave steepness ϵ reconstructed by the FNO-based architecture $\mathcal{M}_{F,1}$. The $\mathcal{M}_{F,1}$ outperforms the $\mathcal{M}_{U,1}$ in reconstructing the shadowed areas, especially noticeable for the sample with large $\epsilon = 0.10$ in (b).

This improved wave reconstruction can be illustrated by comparing the reconstructions of the same two exemplary test set samples generated by $\mathcal{M}_{F,1}$ in Fig. 9 to $\mathcal{M}_{U,1}$ in Fig. 6. As depicted in Fig. 9(a), the sample-specific $nL2_i$ or SSP_i metrics are only slightly improved, but the ratio $\frac{nL2_{shad_i}}{nL2_{vis_i}}$ is substantially smaller than observed using $\mathcal{M}_{U,1}$ before. These observations are even more pronounced for the sample with high ϵ in Fig. 9(b). The $\mathcal{M}_{F,1}$ reduces the error in terms of $nL2_i$ or SSP_i by almost half and also produces a more uniform reconstruction between shadowed and visible areas.

4.2.2. FNO using spatio-temporal radar data

Although the FNO-based model $\mathcal{M}_{F,1}$ outperforms the U-Net-based model $\mathcal{M}_{U,1}$, it does not achieve the desired reconstruction quality of $SSP \leq 0.10$ (Criterion 2). To enhance the model performance we analyse the effect of including multiple historical snapshots in each input $\mathbf{x}_i \in \mathbb{R}^{512 \times n_s}$ for the training of this architecture. Again, 14 additional training runs were conducted, each with an increasing number of n_s. The results, depicted in Fig. 10, demonstrate an initial improvement in performance for the models $\mathcal{M}_{F,n_s}$ with increasing n_s which is slightly less notable than that observed for the U-Net-based models $\mathcal{M}_{U,n_s}$ in Fig. 7 before. The FNO-based models achieve the best performance for $n_s = 9$ input snapshots, beyond which the mean error slightly increases.

According to Table 1, the model $\mathcal{M}_{F,9}$ attains a mean performance of $nL2 = 0.153$ on the test set, after 776 training epochs, as depicted by the loss curve in Fig. B.16(b). This error value corresponds to a mean $SSP = 0.076$, fulfilling the Criterion 2 of a $SSP \leq 0.10$. However, in comparison to the U-Net-based model $\mathcal{M}_{U,10}$, which achieved a final mean value of $SSP = 0.061$, the performance of $\mathcal{M}_{F,9}$ measured in terms of nL2 or SSP is slightly inferior, even though in the single-snapshot case $\mathcal{M}_{F,1}$ outperformed $\mathcal{M}_{U,1}$. Nevertheless, compared to all investigated models, $\mathcal{M}_{F,9}$ on average achieves the best reconstruction uniformity between shadowed and visible areas indicated by a mean $\frac{nL2_{shad}}{nL2_{vis}} = 1.381$ on test

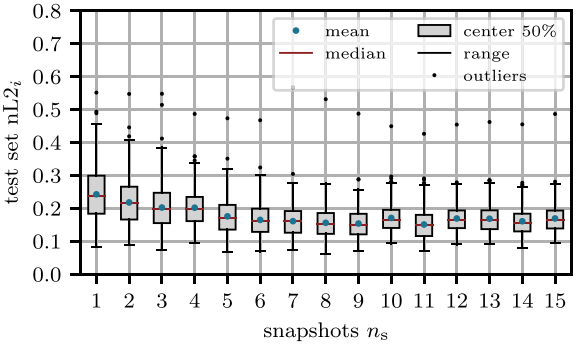

Fig. 10. Boxplot depicting the error distribution on the test set, depending on the number of historical radar snapshots n_s provided to train FNO-based architectures $\mathcal{M}_{F,n_s}$. The best model performance is achieved for $n_s = 9$. Afterwards, the errors slightly increase again.

Fig. 11 shows the reconstructions $\hat{\mathbf{y}}_i$ for the same two exemplary radar inputs $\mathbf{x}_i$ from the test set used before, now generated by the trained FNO-based model $\mathcal{M}_{F,9}$. Compared to $\mathcal{M}_{F,1}$ in Fig. 9 both samples experience an almost similar increase in reconstruction quality measured in terms of the sample-specific SSP_i and $nL2_i$ errors. In addition, these values are comparable to that achieved by $\mathcal{M}_{U,10}$ in Fig. 8. However, for the sample with small $\epsilon = 0.01$ in Fig. 11(a), $\mathcal{M}_{F,9}$ generates a more balanced reconstruction than $\mathcal{M}_{U,10}$, as reflected by the reduction of $\frac{nL2_{shad_i}}{nL2_{vis_i}}$ from 2.201 to 1.665 for this individual sample. For the higher-steepness sample in Fig. 11(b), the increase of reconstruction uniformity given by $\frac{nL2_{shad_i}}{nL2_{vis_i}}$ is less significant but still present.

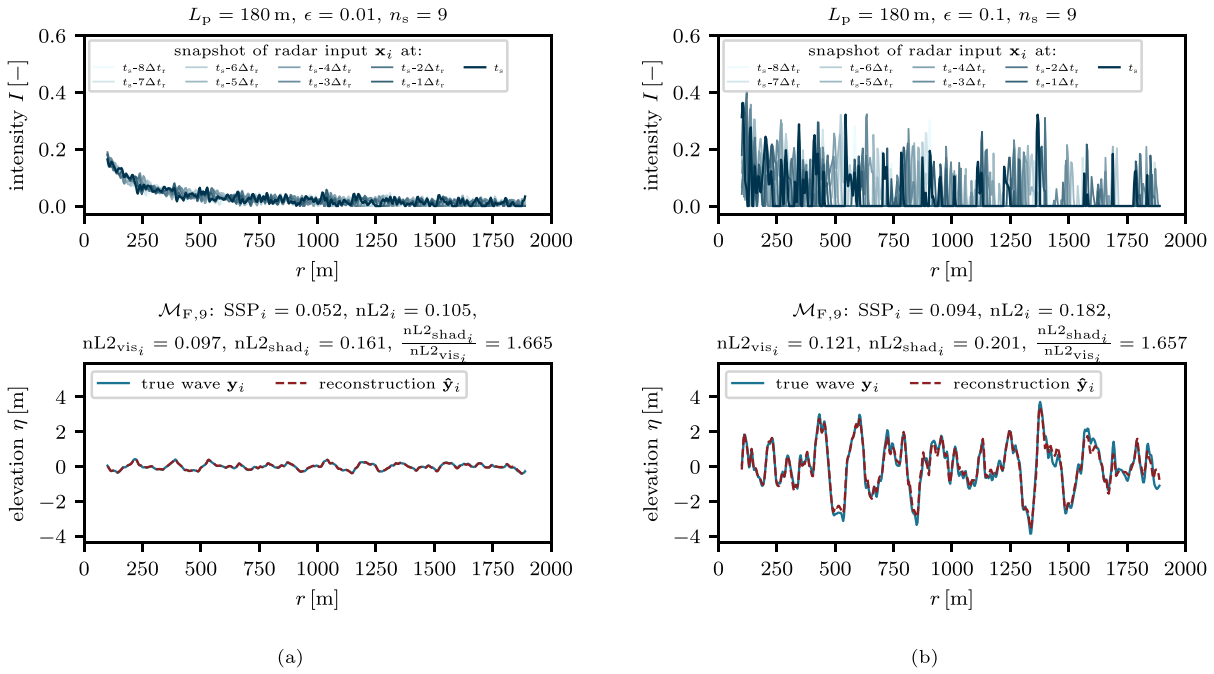

Fig. 11. Two samples from the test set described by the same wavelength $L_p = 180\,m$, but different wave steepness ϵ, reconstructed by the FNO-based architecture trained with $n_s = 9$ historical snapshots in the radar input $\mathcal{M}_{F,9}$. Compared to $\mathcal{M}_{F,1}$ a reconstruction improvement is visible for both samples. Moreover, the reconstruction quality on the entire r-domain is almost equivalent to the results of $\mathcal{M}_{U,10}$, but especially for the small steepness sample in (a) the error ratio between shadowed and visible areas is remarkably smaller using $\mathcal{M}_{F,9}$ which indicates the potential of a more uniform reconstruction.

4.3. Comparative discussion

The aforementioned visual observations described for Figs. 6, 8, 9 and 11 have been limited to the examination of only two exemplary samples from the test set, both described by peak wavelength $L_p = 180\,m$ and either steepness $\epsilon = 0.01$ or $\epsilon = 0.10$. To avoid any possible incidental observations, the generalization of the error values needs to be examined. This can be achieved by plotting sample-specific error values such as $nL2_i$ against each sample's describing combination of peak wavelength L_p and steepness ϵ for all $N_{test} = 624$ test set samples reconstructed using the U-Net-based models $\mathcal{M}_{U,1}$ and $\mathcal{M}_{U,10}$ or the FNO-based models $\mathcal{M}_{F,1}$ and $\mathcal{M}_{F,9}$.

4.3.1. Discussion of overall reconstruction quality

Fig. 12 illustrates the reconstruction error as the mean $nL2_i$ value across 4–5 samples available for each specific $L_p - \epsilon$-combination included in the test set. Additionally, red dots in the cell centres indicate the combinations that achieved a mean $SSP_i \leq 0.10$ (Criterion 2). Fig. 12(a) confirms the findings presented in Section 4.1.1 for the U-Net-based model $\mathcal{M}_{U,1}$ trained with one radar snapshot ($n_s = 1$) in each input $\mathbf{x}_i$. The errors between the true $\mathbf{y}_i$ and reconstructed wave output $\hat{\mathbf{y}}_i$ increase with increasing steepness ϵ and thus with increasing wave height. Moreover, we now observe that this effect occurs almost independent of the peak wavelength L_p of each sample. For samples described by $\epsilon > 0.02$, $\mathcal{M}_{U,1}$ fails to meet the Criterion 2 as the corresponding errors exceed SSP_i values of 0.10. This is attributable to the geometrical radar imaging problem demonstrated in Fig. 2, showing that the increase in wave height caused by increased ϵ results in more and larger shadowed areas. Fig. 13 demonstrates that the occurrence of shadowing mainly increases with increasing ϵ and is less influenced by L_p. While $\epsilon = 0.01$ on average only causes around 10%, $\epsilon = 0.10$ instead causes approximately 70–75 % of each input $\mathbf{x}_i$ being affected by shadowing modulation. This results in areas along the spatial range r containing zero-valued intensities that complicate the radar inversion task.

Understanding the challenges faced by model $\mathcal{M}_{U,1}$ in reconstructing shadowed areas, requires revisiting the U-Net's local mode of operation, outlined in Section 3.1, and the exemplary radar input depicted in the upper panel of Fig. 6(b). Due to shadowing, numerous local areas exhibit zero-intensities covering up to approximately 200 m, especially for greater distances from the radar system. However, the kernels in the first convolutional layer with a kernel size of $s_k = 5$ only cover a domain of $s_k \cdot \Delta r = 17.5\,m$ while being shifted across the input feature map in a step-wise manner. While the U-Net's translational equivariance property is useful for translating radar intensities to wave surface elevation regardless of their spatial location, it thus also causes kernels to be shifted across large areas with zero input only, which cannot be processed in a meaningful way. Although the pooling layers subsequently reduce the dimension of feature maps, resulting in an increased ratio of kernel size to feature size, the problem of radar inversion can be assumed to be based on the mapping of individual pixel values, known as low-level features. These features are learned in the early layers of a CNN-based network (Zeiler and Fergus, 2014). Accordingly, the initial stages of the U-Net-based architecture are more important for our task than for its original purpose of image segmentation (Ronneberger et al., 2015) that is based on mid- to high-level features extracted in the later layers. For this reason, we face problems applying $\mathcal{M}_{U,1}$ for reconstruction, as important kernels in the early layers receive a significant amount of sparse, not valuable content. Although increasing the kernel size s_k is a theoretically possible solution, doing so would compromise the U-Net's local key property. Moreover, when processing two-dimensional surfaces with 2D convolutional kernels in future research, it would result in a quadratic increase in the number of weights, leading to computational issues.

For this reason, the approach of providing $n_s = 10$ consecutive radar snapshots governed according to Criterion 3 for the training of U-Net-based model $\mathcal{M}_{U,10}$ in Section 4.1.2 more effectively accounts for the sparsity in the input data. The upper panel of Fig. 8(b) demonstrated the presence of input information across the majority of the r-domain. The wave surfaces undergo shape variations while travelling towards the radar due to differing phase velocities of their components caused by dispersion. This results in a different part of the radar surface being shadowed or visible at each time step and seem to allow to capture

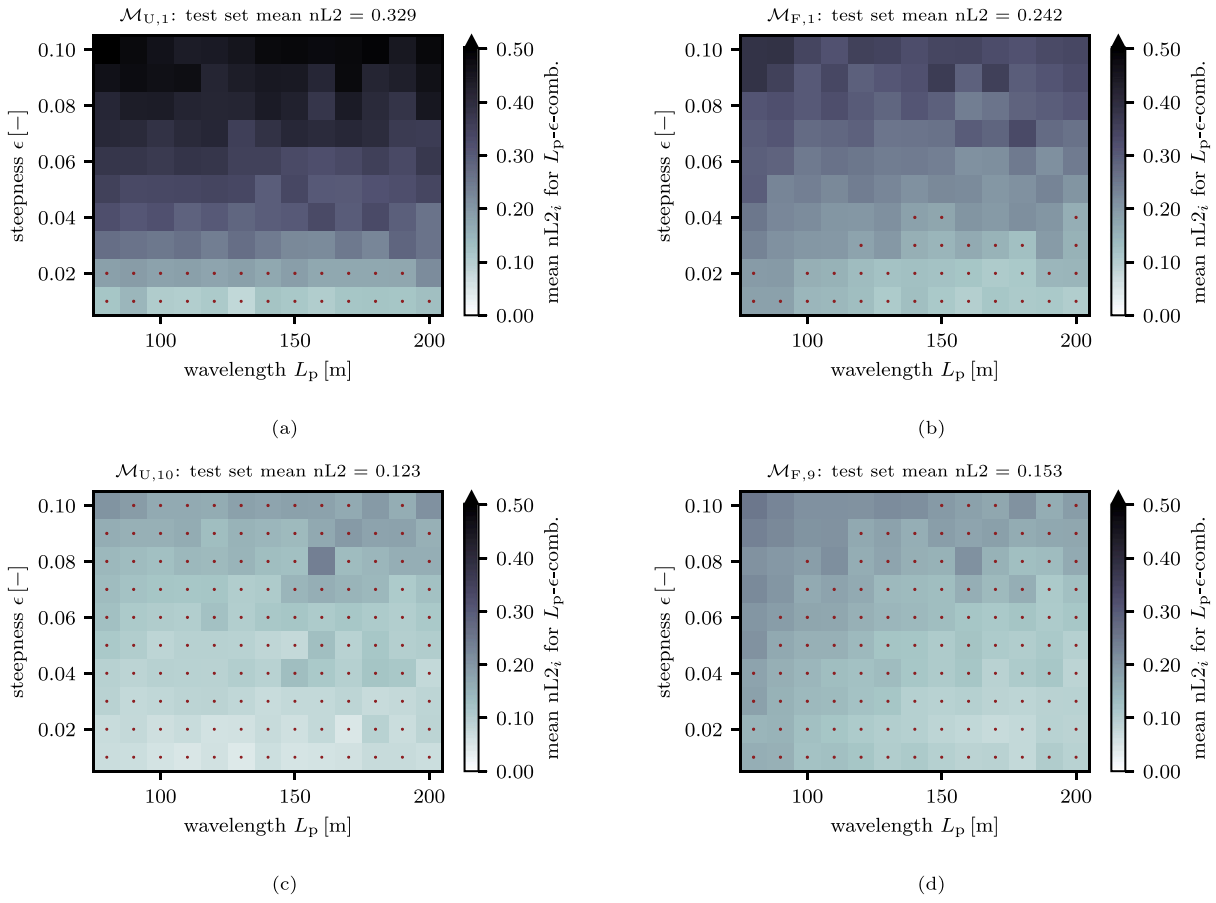

Fig. 12. Error surfaces generalizing the previous observations for sample-specific errors nL2$_i$ of the four investigated models $\mathcal{M}$ depending on the $L_\mathrm{p} - \epsilon$-combination of the samples from the test set. Red dots indicate parameter combinations that meet the Criterion 2 of reconstruction errors SSP$_i \leq 0.10$. The upper subplots illustrate the result of (a) the U-Net-based model and (b) the FNO-based model, both trained with only one radar snapshot ($n_\mathrm{s} = 1$) in each input $\mathbf{x}_i \in \mathbb{R}^{n_\mathrm{r} \times 1}$. The same architectures were trained with multiple historic radar snapshots in each input $\mathbf{x}_i \in \mathbb{R}^{n_\mathrm{r} \times n_\mathrm{s}}$, as demonstrated in the lower subplots, where (c) shows the U-Net-based model trained with $n_\mathrm{s} = 10$ and (d) the FNO-based model trained with $n_\mathrm{s} = 9$.

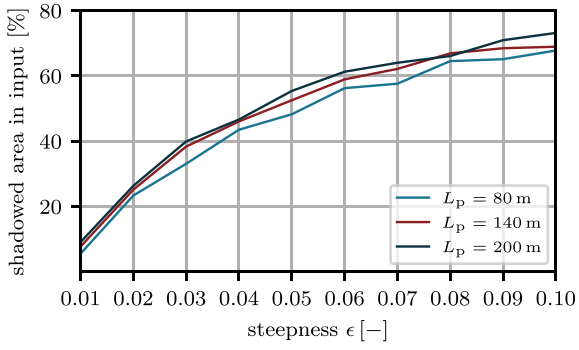

Fig. 13. Graphs visualizing the average proportion of each input $\mathbf{x}_i$ being affected by shadowing modulation in dependency of the samples wave steepness values $\epsilon = 0.01 - 0.10$ for the shortest, one medium and the longest peak wavelength L_p occurring in the test set.

more information about the wave on average, as the reconstruction quality significantly improves compared to $\mathcal{M}_{\mathrm{U},1}$. Therefore we infer that the spatial and temporal shifts of the additional radar intensities acquired at $t_j = \sum_{j=0}^{n_\mathrm{s}-1} t_\mathrm{s} - j\Delta t_\mathrm{r}$ can be compensated successfully. This may be attributed to the fact that each kernel applied to the input has its own channel for each snapshot, allowing for separate processing to counterbalance the shift first, followed by the addition of results to one feature map utilized as part of the input for the next layer. The improved reconstruction observed for $\mathcal{M}_{\mathrm{U},10}$ is further supported by its performance generalization shown in Fig. 12(c). Compared to Fig. 12(a), the mean nL2 error is substantially smaller and sample-specific reconstruction errors nL2$_i$ are more evenly distributed across the $L_\mathrm{p} - \epsilon$-space, resulting in a satisfactory SSP$_i$ value Criterion 2 for almost all samples. Although there is still a slight increase in the error for samples with higher L_p and ϵ, the proposed model $\mathcal{M}_{\mathrm{U},10}$ can accurately reconstruct samples with varying wave characteristics and degrees of shadowing, thus supporting .

Motivated by the inherent patterns in wave data and the successful application of the Fourier neural operator (FNO) to systems exhibiting certain periodic properties, we conducted a comparative analysis of the global inductive bias of this network architecture with the local inductive bias of the CNN-based U-Net. As discussed in Section 4.2.1, our observations indicate that the FNO-based model $\mathcal{M}_{\mathrm{F},1}$ trained with only one snapshot ($n_\mathrm{s} = 1$) outperforms the U-Net-based $\mathcal{M}_{\mathrm{U},1}$ in reconstructing shadowed areas in the input, as evidenced for example by comparing the reconstruction in Figs. 6(b) to 9(b). This observation is generalizable to the entire test data set, as shown in Fig. 12(b). Although errors in the FNO error surface still increase with higher steepness ϵ and consequently with an increase in the percentage of shadowing according to Fig. 13, the increase is much less severe than that obtained by $\mathcal{M}_{\mathrm{U},1}$ shown in Fig. 12(a).

The improved ability of the FNO-based model $\mathcal{M}_{\mathrm{F},1}$ in reconstructing shadowed areas from a single-snapshot input can be attributed to its mode of operation outlined in Section 3.2. Although the latent representation v_0 in Fig. 5 is usually not explicitly known, we can infer that the layer P with $n_\mathrm{s} = 1$ input nodes and n_w output nodes only performs linear transformations to each radar input $\mathbf{x}_i$. As the radar inputs exhibit kinks at the transitions from visible to shadowed areas, v_0 will have similar characteristics along the range direction. These transitions result in peaks for specific wavenumbers k in the spectrum $F(k)$. However, the desired wave outputs $\mathbf{y}_i$ of the training data samples possess smooth periodic properties, without peaks at the kink-related wavenumbers in $F_\mathbf{y}(k)$. Since the R_j matrices in the Fourier layers scale the radar input spectrum to the wave output spectrum, they learn small coefficients for the corresponding entries to reduce the peaks. Therefore, the FNO's global inductive bias, combined with the data structure of wave surfaces, can efficiently correct sparse, shadowed regions in spectral space, resolving the issue of insufficient local information for reconstruction that arises with the U-Net-based model $\mathcal{M}_{\mathrm{U},1}$ in Euclidean space. Thus it can be also stated that the FNO explicitly hard-encodes prior knowledge about physical wave properties through its network structure and thus can be assumed to be a *physics-guided design of architecture* (cf. Willard et al., 2022; Wang and Yu, 2023) for our problem.

Despite the better performance of the FNO-based model $\mathcal{M}_{\mathrm{F},1}$ compared to $\mathcal{M}_{\mathrm{U},1}$ in reconstructing shadowed radar inputs that already supports Hypothesis 2, the red dots in Fig. 12(b) still reveal that most of the test set samples fail to meet the Criterion 2 of $\mathrm{SSP}_i \leq 0.10$. However, this issue was resolved by training a FNO-based model $\mathcal{M}_{\mathrm{F},9}$ with $n_\mathrm{s} = 9$ historical radar snapshots in each input $\mathbf{x}_i$. This was demonstrated for the two test set examples in Fig. 11 and is generalized in Fig. 12(d). We observe from that Figure, that the slightly higher mean error across the entire test set of $\mathcal{M}_{\mathrm{F},9}$ compared to $\mathcal{M}_{\mathrm{U},10}$, is primarily caused by the individual errors $\mathrm{nL2}_i$ of samples with low steepness ϵ or short wavelengths L_p. It is worth noting, that the observed minimal increase in errors for short wavelengths cannot be attributed to a truncation at an insufficient number of Fourier series modes n_m in the Fourier layers. In this work, n_m is determined as 64 and the spectral representation is discretized by $\Delta k = \frac{2\pi}{n_r \cdot \Delta r} = 0.00351\,\mathrm{m}^{-1}$. The highest peak wavenumber of $k_\mathrm{p} = 0.0785\,\mathrm{m}^{-1}$ in our data set is reached for samples with $L_\mathrm{p} = 80\,\mathrm{m}$. The spectral density around k_p has decayed almost completely at $k_\mathrm{filt} = n_\mathrm{m} \cdot \Delta k = 0.2264\,\mathrm{m}^{-1}$, such that no important wave components are filtered out, as is visualized in Fig. C.17. Therefore, the small unequal tendency in error distribution achieved by $\mathcal{M}_{\mathrm{F},9}$ in Fig. 12(d) for samples described by different $L_\mathrm{p} - \epsilon$, is likely caused by other factors than by an unsuitable network hyperparameter n_m. Moreover, we observed in the loss curve shown in Fig. B.16(b) that further training for more than 800 epochs could potentially improve the model's performance, whereas the best performance on the test set for $\mathcal{M}_{\mathrm{U},10}$ seems to be already reached, as the model begins to overfit the training data, as depicted in Fig. B.15(a).

4.3.2. Discussion of reconstruction uniformity

So far the generalization of the reconstruction quality has been evaluated based on samples-specific $\mathrm{nL2}_i$ or SSP_i values across the entire spatial r-domain. However, Table 1 indicates that the FNO-based model $\mathcal{M}_{\mathrm{F},9}$ achieves a more uniform reconstruction between shadowed and visible areas. This is demonstrated by the mean ratio of $\frac{\mathrm{nL2}_\mathrm{shad}}{\mathrm{nL2}_\mathrm{vis}} = 1.381$ across all samples in the test set, while the U-Net-based model $\mathcal{M}_{\mathrm{U},10}$ still struggles with reconstructing shadowed areas as inferred by its $\frac{\mathrm{nL2}_\mathrm{shad}}{\mathrm{nL2}_\mathrm{vis}} = 1.755$. Therefore, the $\frac{\mathrm{nL2}_{\mathrm{shad}_i}}{\mathrm{nL2}_{\mathrm{vis}_i}}$-ratio error distribution is displayed in Fig. 14 for each test set sample based on their $L_\mathrm{p} - \epsilon$-combination. The model $\mathcal{M}_{\mathrm{U},10}$ generates an error surface shown in Fig. 14(a) that exhibits broadly varying levels of uniformity in the reconstruction even for samples with neighbouring $L_\mathrm{p}-\epsilon$-combinations. In some cases, the reconstruction errors in shadowed areas exceed those in visible areas by more than 2.5 times. This undesired effect is much less pronounced for the FNO-based model $\mathcal{M}_{\mathrm{F},9}$, as a comparison with Fig. 14(b) reveals.

4.3.3. Final comparison

For a final evaluation, either the general reconstruction quality nL2 can be chosen as the main performance criterion, which in our case would argue for the selection of the U-Net-based model $\mathcal{M}_{\mathrm{U},10}$, or instead the uniformity of the reconstruction indicated by $\frac{\mathrm{nL2}_\mathrm{shad}}{\mathrm{nL2}_\mathrm{vis}}$, which would argue for the FNO-based model $\mathcal{M}_{\mathrm{F},9}$. This decision should be made based on the application case. If the ML-reconstructed wave surface is intended to be used as an initial condition for subsequent prediction with the HOS method, we would expect a more uniform reconstruction to represent a more physical result, and consequently, the FNO-based reconstruction to be less likely to affect the subsequent wave prediction in a negative way. Moreover, we observed that the global approach of the FNO-based models would allow for a reasonably more meaningful reconstruction of shadowed areas even with fewer historical radar snapshots n_s contained in each input $\mathbf{x}_i$. This is not necessarily the case using the U-Net-based models.

Besides, the FNO-based model $\mathcal{M}_{\mathrm{F},9}$ in this work allows for much faster inference speed than the U-Net-based $\mathcal{M}_{\mathrm{U},10}$, even though $\mathcal{M}_{\mathrm{F},9}$ is constructed as a custom implementation and contains more weights compared to $\mathcal{M}_{\mathrm{U},10}$, which uses standard layers from the PyTorch library that are in addition probably optimized. More specifically, using the hardware specifications outlined in Section 3.3, our $\mathcal{M}_{\mathrm{F},9}$ is able to generate reconstructions $\hat{\mathbf{y}}_i$ for an input sample $\mathbf{x}_i$ in an average time of $1.9 \cdot 10^{-3}$ s which is approximately 20 times faster than the average time of $3.7 \cdot 10^{-2}$ s required by $\mathcal{M}_{\mathrm{U},10}$ for the same task.

5. Conclusion

This work introduces a novel machine learning-based approach for the phase-resolved reconstruction of ocean wave surface elevations from sparse radar measurements. To evaluate the performance of our approach, we generate synthetic nonlinear wave surface data for a wide range of sea states and corresponding radar surface data by incorporating both tilt- and shadowing modulation mechanisms. Two neural network architectures based on the U-Net or the Fourier neural operator are trained, both provided with varying amounts of spatio-temporal radar surface measurement input.

Our results and discussion indicate that both models are capable of producing high-quality wave surface reconstructions with average errors below $\mathrm{SSP} \leq 0.10$ when trained with a sufficient amount of $n_\mathrm{s} = 10$ or 9 consecutive radar snapshots. Furthermore, both models generalize well across different sea states. On average, the U-Net-based model achieves slightly smaller errors across the entire spatial domain of each reconstructed wave sample, while the FNO-based model produces a more uniform wave reconstruction between areas that were shadowed and visible in the corresponding radar input. This observation is further confirmed by the edge case of instantaneous inversion, i.e. if the networks are trained with only a single radar snapshot in each input. The weakness in the reconstruction of shadowing-affected areas of the U-Net-based model can be attributed to the local operation of the network architecture, where its small convolutional kernels do not receive processable information when shifted across shadowed input areas with zero intensities only. The problem can be circumvented using the FNO-based network that learns a global mapping between radar input and wave output in the Fourier space. This network structure already encodes prior physical knowledge about the periodic data structure apparent in ocean waves and is therefore possibly better suited for our use case.

Our findings suggest that the FNO-based network may offer additional advantages, especially concerning smaller training datasets and noisy input radar data. Furthermore, future research could delve into the reconstruction of two-dimensional ocean wave surfaces, as the FNO network can also be implemented using 2D-FFTs. However, due to the different propagation directions of the component waves in short-crested, two-dimensional sea states, we anticipate a potential degradation in the reconstruction performance compared to the

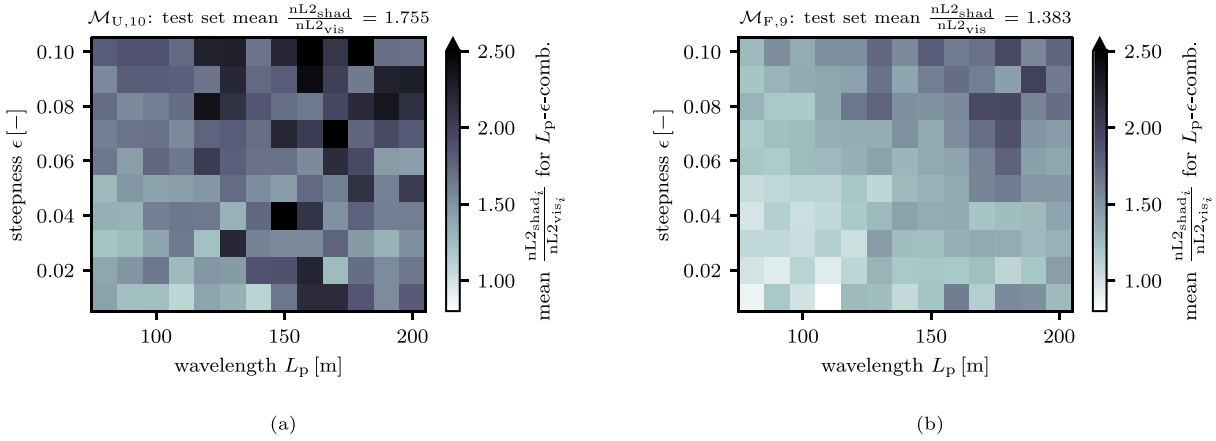

Fig. 14. Error surfaces depicting the ratio $\frac{\mathrm{nL2_{shad}}_i}{\mathrm{nL2_{vis}}_i}$ between the reconstruction quality achieved on shadowed and visible areas depending on the specific $L_\mathrm{p} - \epsilon$-combination of the samples from the test set. The individual cell entries display the mean ratio across the 4–5 samples available for each specific parameter combination. The uniformity of the reconstructions achieved by the U-Net-based model $\mathcal{M}_{\mathrm{U},10}$ in (a) thus is compared to the one achieved by the FNO-based model $\mathcal{M}_{\mathrm{F},9}$ in (b).

one-dimensional scenario explored in this study. This performance degradation could be mitigated through appropriate countermeasures, such as employing FNOs with increased capacity, conducting longer training runs, or applying suitable regularization techniques.

Moreover, the current methodology solely relies on synthetic radar input and the corresponding wave output data. Although it can be presumed that the HOS method generates wave surfaces that exhibit a reasonable degree of realism, radar imaging mechanisms for marine X-Band radar are not yet fully understood, such that state-of-the-art radar models are associated with higher uncertainties. Consequently, on the one hand, a model trained solely on synthetic radar-wave data pairs cannot be applied for inference using real-world radar data. On the other hand, the acquisition of real-world radar-wave pair samples to train the neural networks is associated with high operational costs due to the necessity of deploying a dense grid of buoys for capturing wave snapshots. These data issues currently limit the application of the developed machine-learning-based reconstruction approach for real-world applications. In future research, we endeavour to tackle this issue through two opportunities: Firstly, we aim to enhance the realism of synthetic radar data models to improve their accuracy. Alternatively, we intend to investigate the feasibility of physics-informed learning approaches as a tool to overcome the challenges associated with measuring real-world high-resolution wave output data.

CRediT authorship contribution statement

Svenja Ehlers: Conceptualization, Methodology, Software (ML), Investigation, Writing – original draft, Visualization. **Marco Klein:** Software (HOSM), Project administration, Writing (Review Introduction and Wave Data). **Alexander Heinlein:** Conceptualization, Formal analysis, Validation, Writing – review & editing. **Mathies Wedler:** Data Curation, Writing (Review ML Methods and Results). **Nicolas Desmars:** Software (Radar), Writing (Review Radar Data). **Norbert Hoffmann:** Supervision, Project administration, Resources, Funding acquisition, Writing (Review). **Merten Stender:** Conceptualization, Supervision, Project administration, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Declaration of Generative AI and AI-assisted technologies in the writing process

The manuscript was completely written by the authors. Once the manuscript was completed, the authors used ChatGPT in order to improve its grammar and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgements

This work was supported by the Deutsche Forschungsgesellschaft (DFG - German Research Foundation) [project number 277972093: Excitability of Ocean Rogue Waves].

Appendix A. Influence of neural network hyperparameters

To mitigate the high cost of obtaining a larger data set, a four-fold cross-validation approach with an independent test set was utilized for finding the network hyperparameters, as recommended for example by Raschka (2018). The data set of $N = 3120$ samples was divided into a fixed and independent test set comprising 20% or $N_\mathrm{test} = 624$ samples, with the remaining 2496 samples partitioned into four equal-sized parts based on the governing sea state parameters (L_p, ϵ) using a stratified data split technique to ensure equal representation of each wave characteristic in the resulting subsets. During each cross-validation step, one part with $N_\mathrm{val} = 624$ samples was used as the

Table A.2
Results of the hyperparameter study for the U-Net-based architecture. Each investigated architecture is characterized by a depth n_d, a kernel size s_k and an approach for the number of convolutional kernels n_k in each layer. The approaches for n_k explored doubling the number of kernels per layer with increasing encoder depth (and in reverse halving in the decoder) and another keeping n_k the same in all convolutional layers. The performance, measured by the mean nL2 value on validation set displays only slight variations among architectures, indicating that the U-Net is not highly sensitive to these changes. Nevertheless, we select the best model according to the validation nL2, that highlighted in blue, as our $\mathcal{M}_{U,1}$.

U-Net hyperparameters			#weights	Epochs	nL2		$\frac{nL2_{shad}}{nL2_{vis}}$		SSP	
Depth n_d	Kernels each layer n_k	k. size s_k			Train	Val	Train	Val	Train	Val
3	[8, 16, 32]	3	12,401	785	0.392	0.395	2.124	2.119	0.208	0.210
		5	18,305	791	0.360	0.367	2.512	2.526	0.189	0.193
4	[8, 16, 32, 64]	3	48,433	645	0.352	0.371	2.475	2.527	0.185	0.195
		5	72,769	189	0.347	0.361	2.536	2.575	0.182	0.189
5	[8, 16, 32, 64, 128]	3	192,177	278	0.342	0.360	2.601	2.644	0.179	0.189
		5	290,241	154	0.303	0.344	2.276	2.438	0.158	0.179
6	[8, 16, 32, 64, 128, 256]	3	766,385	223	0.336	0.356	2.557	2.601	0.176	0.186
		5	1,159,361	78	0.326	0.356	2.365	2.452	0.170	0.186
3	[16, 32, 64]	3	49,121	779	0.388	0.392	2.149	2.144	0.206	0.208
		5	72,705	675	0.353	0.366	2.543	2.582	0.185	0.193
4	[16, 32, 64, 128]	3	192,865	310	0.356	0.371	2.439	2.479	0.187	0.195
		5	290,177	99	0.349	0.364	2.520	2.551	0.183	0.191
5	[16, 32, 64, 128, 256]	3	767,073	140	0.343	0.365	2.490	2.532	0.181	0.192
		5	1,159,297	76	0.298	0.342	2.364	2.476	0.155	0.178
6	[16, 32, 64, 128, 256, 512]	3	3,062,369	225	0.338	0.358	2.568	2.611	0.177	0.188
		5	4,634,241	124	0.346	0.366	2.419	2.487	0.183	0.194
3	[32, 32, 32]	3	38465	778	0.388	0.392	2.135	2.137	0.206	0.208
		5	52865	724	0.351	0.363	2.581	2.611	0.185	0.191
4	[32, 32, 32, 32]	3	49,825	569	0.350	0.364	2.555	2.583	0.184	0.192
		5	70,369	199	0.345	0.358	2.631	2.663	0.181	0.188
5	[32, 32, 32, 32, 32]	3	61,185	224	0.344	0.357	2.654	2.663	0.180	0.187
		5	87,873	183	0.310	0.341	2.536	2.659	0.161	0.177
6	[32, 32, 32, 32, 32, 32]	3	72,545	395	0.350	0.362	2.548	2.562	0.184	0.191
		5	105,377	217	0.333	0.350	2.693	2.741	0.174	0.183
3	[64, 64, 64]	3	152,705	739	0.383	0.389	2.186	2.189	0.203	0.206
		5	210,177	394	0.351	0.365	2.539	2.568	0.185	0.192
4	[64, 64, 64, 64]	3	197,953	333	0.351	0.367	2.482	2.524	0.184	0.193
		5	280,001	120	0.343	0.359	2.611	2.646	0.179	0.187
5	[64, 64, 64, 64, 64]	3	243,201	191	0.342	0.356	2.687	2.694	0.179	0.186
		5	349,825	182	0.316	0.348	2.482	2.603	0.165	0.182
6	[64, 64, 64, 64, 64, 64]	3	288,449	255	0.345	0.362	2.532	2.568	0.181	0.190
		5	419,649	138	0.323	0.349	2.598	2.684	0.168	0.182

validation set, and the remaining three parts with $N_{train} = 1872$ samples constituted the training set.

Tables A.2 and A.3 present the results of the four-fold cross-validation hyperparameter studies for the U-Net- and FNO-based architectures. For both network types, the same fixed test set was excluded from this investigation. The metrics (nL2, $\frac{nL2_{shad}}{nL2_{vis}}$, SSP) and the number of epochs necessary to attain the best performance represent average values across all four folds.

Appendix B. Loss curves

After determining appropriate hyperparameters for the models in Appendix A, the train and validation data from the four-fold cross-validation were merged. This combined data set was then used to train the models $\mathcal{M}_{U,n_s}$ and $\mathcal{M}_{F,n_s}$, with one radar snapshot in each samples input ($n_s = 1$) or either $n_s = 9$ or $n_s = 10$ historical radar snapshots. The performance evaluation of these models was conducted on the previously excluded test set of $N_{test} = 624$ samples. The loss curves depicted in Figs. B.15(a)–B.16(b) illustrate the model performance and the impact of different values for n_s throughout the training epochs. Deviation between the train and test loss curves indicates overfitting, caused by excessive adaptation to the training data, resulting in poor generalization to new samples. Consequently, the best models were selected based on the lowest test loss within the 800 training epochs.

Appendix C. Visualization of spectral representation

During the investigations on the FNO models (see Fig. 5), a concern arose regarding the chosen number of Fourier series modes $n_m = 64$ in the R_i-matrices, which might lead to the omission of significant frequency components in the wave data. To address this concern, we visualized the JONSWAP spectra employed to initialize the HOS wave simulation for a specific steepness value ϵ (since different $\epsilon = 0.08$ just scale the amplitude of spectral density) and all peak wavelengths $L_p \in \{80, 90, \ldots, 190, 200\}$ m, each corresponding to a specific ω_p and k_p. Based on the findings depicted and explained in Fig. C.17, we conclude that this assumption is invalid.

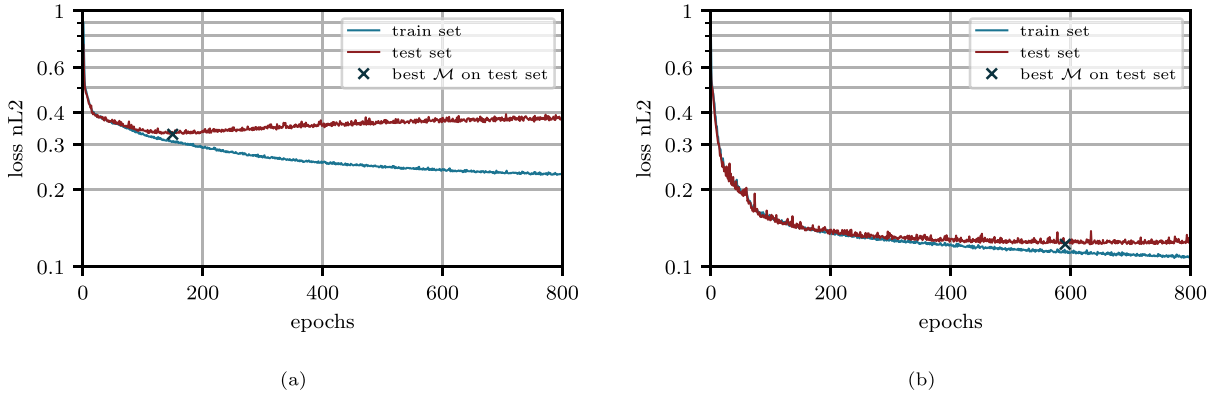

Fig. B.15. Loss curves for training of the U-Net-based model. Subfigure (a) depicts the loss of model $\mathcal{M}_{\mathrm{U},1}$ trained with one one snapshot $n_{\mathrm{s}} = 1$ in the radar input, where the best performance nL2 = 0.329 on test set for model evaluation is reached after 150 epochs. Afterwards the model would tend to overfit the training data. Subfigure (b) depicts model $\mathcal{M}_{\mathrm{U},10}$ trained with $n_{\mathrm{s}} = 10$ instead, which strongly increases performance, resulting in nL2 = 0.123 after 592 epochs of training.

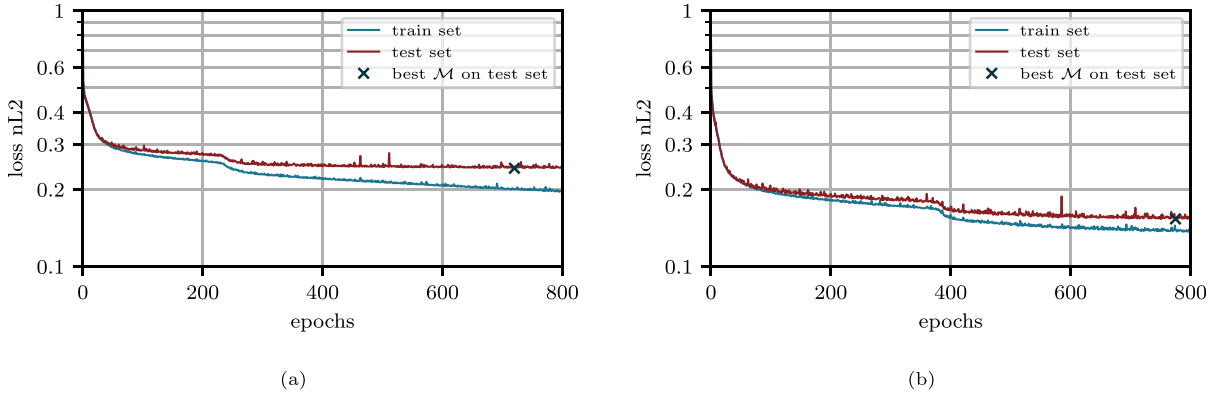

Fig. B.16. Loss curves for training of the FNO-based model. Subfigure (a) depicts the loss of model $\mathcal{M}_{\mathrm{F},1}$ trained with one one snapshot $n_{\mathrm{s}} = 1$ in the radar input, where the best performance nL2 = 0.242 on test set for model evaluation is reached after 721 epochs. Compared to the U-Net based model $\mathcal{M}_{\mathrm{U},1}$, $\mathcal{M}_{\mathrm{F},1}$ does not seem to be susceptible to overfitting . Subfigure (b) depicts model $\mathcal{M}_{\mathrm{F},9}$ trained with $n_{\mathrm{s}} = 9$ instead, which increases performance, resulting in nL2 = 0.153 after 776 epochs of training. It can be expected that training beyond 800 epochs would further slightly increase the best performance on test set.

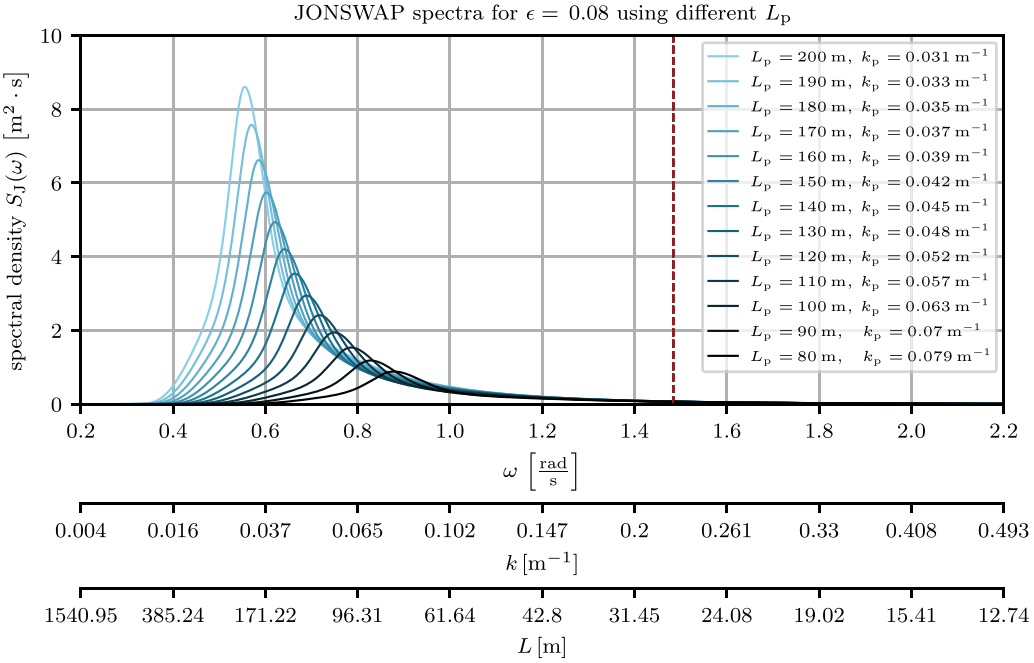

Fig. C.17. JONSWAP spectra used in the data generation for one exemplary steepness value ϵ, but varying peak wavelengths $L_{\mathrm{p}} = 80 - 200\,\mathrm{m}$. The shortest peak wavelength of $L_{\mathrm{p}} = 80\,\mathrm{m}$ corresponds to the highest peak wavenumber of $k_{\mathrm{p}} = 0.079\,\mathrm{m}^{-1}$. The filtering wavenumber of $k_{\mathrm{filt}} = n_{\mathrm{m}} \cdot \Delta k = 64 \cdot 0.00351\,\mathrm{m}^{-1} = 0.2246\,\mathrm{m}^{-1}$, which is indicated by the dotted red line and defined by the Fourier layers in this work, consequently does not truncate important wave components in our data set-up.

Table A.3
Results of the hyperparameter study for the FNO-based architecture. Each investigated architecture is characterized by a number of Fourier layers n_f, a width of the latent representation n_w and a number of modes n_m for truncation of the layers' Fourier series. Our observations reveal that the performance on the validation set does not show relevant improvement for $n_w > 32$, and that $n_f = 3$ lead to slightly better performance than the larger number of layers. The performance, measured by the mean nL2 value on validation set, displays a little more distinct variation among architectures than the U-Net before. Nevertheless, neighbouring parameter combination of the FNO result in almost the same performance, suggesting that the FNO, is also not overly sensitive to hyperparameter changes. We select the best model according to the smallest mean validation sets nL2 highlighted in blue as our $\mathcal{M}_{F,1}$.

FNO hyperparameters			#weights	Epochs	nL2		$\frac{nL2_{shad}}{nL2_{vis}}$		SSP	
Layers n_f	Modes n_m	Width n_w			Train	Val	Train	Val	Train	Val
3	32	16	52,321	793	0.316	0.334	1.380	1.437	0.163	0.173
		32	204,225	790	0.262	0.313	1.338	1.520	0.134	0.160
		48	455,969	540	0.216	0.296	1.371	1.685	0.110	0.150
		64	807,553	500	0.214	0.296	1.362	1.681	0.109	0.151
3	40	16	64,609	791	0.284	0.306	1.469	1.547	0.145	0.157
		32	253,377	703	0.229	0.290	1.402	1.654	0.116	0.148
		48	566,561	480	0.212	0.291	1.380	1.701	0.108	0.148
		64	1,004,161	395	0.207	0.288	1.390	1.719	0.105	0.146
3	48	16	76,897	788	0.272	0.296	1.481	1.573	0.139	0.152
		32	302,529	605	0.219	0.279	1.419	1.675	0.111	0.142
		48	677,153	342	0.216	0.285	1.373	1.661	0.110	0.145
		64	1,200,769	258	0.214	0.286	1.360	1.642	0.109	0.146
3	56	16	89,185	787	0.243	0.270	1.602	1.704	0.124	0.137
		32	351,681	714	0.202	0.265	1.499	1.817	0.102	0.135
		48	787,745	397	0.208	0.265	1.511	1.767	0.105	0.134
		64	1,397,377	307	0.204	0.267	1.475	1.753	0.104	0.136
3	64	16	101,473	798	0.237	0.265	1.644	1.767	0.102	0.135
		32	400,833	534	0.199	0.256	1.561	1.837	0.101	0.130
		48	898,337	349	0.199	0.257	1.560	1.844	0.101	0.130
		64	1,593,985	276	0.190	0.261	1.501	1.822	0.096	0.133
3	72	16	113,761	742	0.234	0.267	1.588	1.750	0.119	0.136
		32	449,985	600	0.197	0.257	1.575	1.887	0.100	0.131
		48	1,008,929	367	0.189	0.258	1.568	1.889	0.096	0.132
		64	1,790,593	244	0.187	0.258	1.523	1.851	0.095	0.131
4	32	16	68,977	789	0.281	0.316	1.386	1.520	0.144	0.162
		32	270,817	520	0.220	0.296	1.355	1.668	0.112	0.151
		48	605,777	394	0.190	0.289	1.358	1.760	0.097	0.147
		64	1,073,857	233	0.185	0.292	1.303	1.711	0.094	0.148
4	40	16	85,361	787	0.250	0.291	1.469	1.658	0.128	0.148
		32	336,353	608	0.193	0.281	1.397	1.817	0.098	0.143
		48	753,233	331	0.188	0.283	1.360	1.771	0.095	0.144
		64	1,336,001	167	0.206	0.293	1.314	1.638	0.105	0.149
4	48	16	101,745	720	0.245	0.289	1.408	1.612	0.125	0.148
		32	401,889	327	0.212	0.282	1.365	1.662	0.103	0.143
		48	900,689	185	0.205	0.282	1.355	1.676	0.104	0.144
		64	1,598,145	431	0.144	0.282	1.324	1.785	0.073	0.142
4	56	16	118,129	703	0.216	0.265	1.533	1.767	0.110	0.135
		32	467,425	518	0.172	0.269	1.420	1.873	0.087	0.136
		48	1,048,145	285	0.171	0.269	1.406	1.834	0.087	0.137
		64	1,860,289	156	0.177	0.269	1.385	1.778	0.089	0.136
4	64	16	134,513	784	0.204	0.258	1.555	1.843	0.103	0.131
		32	532,961	270	0.186	0.258	1.497	1.846	0.095	0.131
		48	1,195,601	149	0.187	0.260	1.476	1.831	0.094	0.132
		64	2,122,433	114	0.176	0.259	1.450	1.826	0.089	0.131
4	72	16	150,897	740	0.189	0.256	1.528	1.903	0.095	0.130
		32	598,497	262	0.180	0.257	1.496	1.903	0.091	0.130
		48	1,343,057	181	0.168	0.257	1.454	1.879	0.085	0.130
		64	2,384,577	121	0.169	0.259	1.406	1.817	0.086	0.132

References

Airy, G., 1849. Tides and Waves. In: Encyclopædia metropolitana, John Joseph Griffin and Company.

Asma, S., Sezer, A., Ozdemir, O., 2012. MLR and ANN models of significant wave height on the west coast of India. Comput. Geosci. 49, 231–237. http://dx.doi.org/10.1016/j.cageo.2012.05.032.

Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V.F., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., 2018. Relational inductive biases, deep learning, and graph networks. arXiv:1806.01261.

Bertero, M., Boccacci, P., de Mol, C., 2022. Introduction to Inverse Problems in Imaging. CRC Press.

Blondel-Couprie, E., 2009. Reconstruction et prévision déterministe de houle à partir de données mesurées (Ph.D. thesis). Ecole Centrale de Nantes (ECN).

Borge, J.C.N., Reichert, K., Dittmer, J., 1999. Use of nautical radar as a wave monitoring instrument. Coast. Eng. 37 (3), 331–342. http://dx.doi.org/10.1016/S0378-3839(99)00032-0.

Borge, J.N., Rodrıguez, G.R., Hessner, K., González, P.I., 2004. Inversion of marine radar images for surface wave analysis. J. Atmos. Ocean. Technol. 21 (8), 1291–1300. http://dx.doi.org/10.1175/1520-0426(2004)021<1291:IOMRIF>2.0.CO;2.

Bouws, E., Günther, H., Rosenthal, W., Vincent, C.L., 1985. Similarity of the wind wave spectrum in finite depth water: 1. Spectral form. J. Geophys. Res. 90, 975–986.

Chen, X., Huang, W., 2022. Spatial–temporal convolutional gated recurrent unit network for significant wave height estimation from shipborne marine radar data. IEEE Trans. Geosci. Remote Sens. 60, 1–11. http://dx.doi.org/10.1109/TGRS.2021.3074075.

Dankert, H., Rosenthal, W., 2004. Ocean surface determination from X-band radar-image sequences. J. Geophys. Res.: Oceans 109 (C4), http://dx.doi.org/10.1029/2003JC002130.

Deo, M., Jha, A., Chaphekar, A., Ravikant, K., 2001. Neural networks for wave forecasting. Ocean Eng. 28 (7), 889–898. http://dx.doi.org/10.1016/S0029-8018(00)00027-5.

Desmars, N., 2020. Real-time reconstruction and prediction of ocean wave fields from remote optical measurements. no. 2020ECDN0036 (Ph.D. thesis). École centrale de Nantes.

Desmars, N., Hartmann, M., Behrendt, J., Klein, M., Hoffmann, N., 2021. Reconstruction of Ocean Surfaces From Randomly Distributed Measurements Using a Grid-Based Method. In: International Conference on Offshore Mechanics and Arctic Engineering. Vol. Volume 6: Ocean Engineering. http://dx.doi.org/10.1115/OMAE2021-62409.

Desmars, N., Hartmann, M., Behrendt, J., Klein, M., Hoffmann, N., 2022. Nonlinear Reconstruction and Prediction of Regular Waves. In: International Conference on Offshore Mechanics and Arctic Engineering. Vol. Volume 5B: Ocean Engineering; Honoring Symposium for Professor Günther F. Clauss on Hydrodynamics and Ocean Engineering. http://dx.doi.org/10.1115/OMAE2022-78988.

Desouky, M.A., Abdelkhalik, O., 2019. Wave prediction using wave rider position measurements and NARX network in wave energy conversion. Appl. Ocean Res. 82, 10–21. http://dx.doi.org/10.1016/j.apor.2018.10.016.

Dommermuth, D.G., Yue, D.K.P., 1987. A high-order spectral method for the study of nonlinear gravity waves. J. Fluid Mech. 184, 267–288. http://dx.doi.org/10.1017/S002211208700288X.

Duan, W., Ma, X., Huang, L., Liu, Y., Duan, S., 2020a. Phase-resolved wave prediction model for long-crest waves based on machine learning. Comput. Methods Appl. Mech. Engrg. 372, 113350. http://dx.doi.org/10.1016/j.cma.2020.113350.

Duan, W., Yang, K., Huang, L., Ma, X., 2020b. Numerical investigations on wave remote sensing from synthetic X-Band radar sea clutter images by using deep convolutional neural networks. Remote Sens. 12 (7), http://dx.doi.org/10.3390/rs12071117.

Ducrozet, G., Bonnefoy, F., Le Touzé, D., Ferrant, P., 2007. 3-D HOS simulations of extreme waves in open seas. Nat. Hazards Earth Syst. Sci. 7 (1), 109–122. http://dx.doi.org/10.5194/nhess-7-109-2007.

Eichinger, M., Heinlein, A., Klawonn, A., 2022. Surrogate convolutional neural network models for steady computational fluid dynamics simulations. Electron. Trans. Numer. Anal. 56, 235–255.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, http://www.deeplearningbook.org.

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., Meerburg, A., Mller, P., Olbers, D., Richter, K., Sell, W., Walden., H., 1973. Measurements of wind-wave growth and swell decay during the joint north sea wave project (JONSWAP). Ergänzung zur Deutschen Hydrographischen Zeitschrift, Reihe A (8) 12, 1–95.

Hendrycks, D., Gimpel, K., 2016. Bridging nonlinearities and stochastic regularizers with Gaussian error linear units. CoRR abs/1606.08415, arXiv:1606.08415.

Hilmer, T., Thornhill, E., 2015. Observations of predictive skill for real-time deterministic sea waves from the WaMoS II. In: OCEANS 2015 - MTS/IEEE Washington. pp. 1–7.

Isola, P., Zhu, J., Zhou, T., Efros, A.A., 2016. Image-to-image translation with conditional adversarial networks. arXiv:1611.07004.

James, S.C., Zhang, Y., O'Donncha, F., 2018. A machine learning framework to forecast wave conditions. Coast. Eng. 137, 1–10. http://dx.doi.org/10.1016/j.coastaleng.2018.03.004.

Kagemoto, H., 2020. Forecasting a water-surface wave train with artificial intelligence-A case study. Ocean Eng. 207, 107380. http://dx.doi.org/10.1016/j.oceaneng.2020.107380.

Kharif, C., Slunyaev, A., Pelinovsky, E., 2009. Rogue Waves in the Ocean, first ed. Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-540-88419-4_5.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. http://dx.doi.org/10.48550/ARXIV.1412.6980, arXiv.

Klein, M., Dudek, M., Clauss, G.F., Ehlers, S., Behrendt, J., Hoffmann, N., Onorato, M., 2020. On the deterministic prediction of water waves. Fluids 5 (1), http://dx.doi.org/10.3390/fluids5010009.

Klein, M., Stender, M., Wedler, M., Ehlers, S., Hartmann, M., Desmars, N., Pick, M.-A., Seifried, R., Hoffmann, N., 2022. Application of Machine Learning for the Generation of Tailored Wave Sequences. In: International Conference on Offshore Mechanics and Arctic Engineering. Vol. Volume 5B: Ocean Engineering; Honoring Symposium for Professor Günther F. Clauss on Hydrodynamics and Ocean Engineering. http://dx.doi.org/10.1115/OMAE2022-78601.

Köllisch, N., Behrendt, J., Klein, M., Hoffmann, N., 2018. Nonlinear real time prediction of ocean surface waves. Ocean Eng. 157, 387–400. http://dx.doi.org/10.1016/j.oceaneng.2018.03.048.

Law, Y., Santo, H., Lim, K., Chan, E., 2020. Deterministic wave prediction for unidirectional sea-states in real-time using artificial neural network. Ocean Eng. 195, 106722. http://dx.doi.org/10.1016/j.oceaneng.2019.106722.

LeCun, Y., et al., 1989. Generalization and network design strategies. In: CoNnectionism in Perspective. Vol. 19, no. 143–155. North Holland, p. 18.

Li, Z., Kovachki, N.B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A.M., Anandkumar, A., 2020. Fourier neural operator for parametric partial differential equations. arXiv:2010.08895.

Li, Z., Peng, W., Yuan, Z., Wang, J., 2022. Fourier neural operator approach to large eddy simulation of three-dimensional turbulence. Theor. Appl. Mech. Lett. 12 (6).

Li, H., Xu, Z., Taylor, G., Studer, C., Goldstein, T., 2018. Visualizing the loss landscape of neural nets. In: Advances in Neural Information Processing Systems. Vol. 31.

Liu, D., Wen, B., Liu, X., Wang, Z., Huang, T.S., 2018. When image denoising meets high-level vision tasks: A deep learning approach. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. IJCAI '18, AAAI Press, pp. 842–848.

Liu, Y., Zhang, X., Chen, G., Dong, Q., Guo, X., Tian, X., Lu, W., Peng, T., 2022. Deterministic wave prediction model for irregular long-crested waves with recurrent neural network. J. Ocean Eng. Sci. http://dx.doi.org/10.1016/j.joes.2022.08.002.

Lünser, H., Hartmann, M., Desmars, N., Behrendt, J., Hoffmann, N., Klein, M., 2022. The influence of characteristic sea state parameters on the accuracy of irregular wave field simulations of different complexity. Fluids 7 (7), http://dx.doi.org/10.3390/fluids7070243.

Mitchell, T.M., 1980. The Need for Biases in Learning Generalizations. Citeseer.

Mohaghegh, F., Murthy, J., Alam, M.-R., 2021. Rapid phase-resolved prediction of nonlinear dispersive waves using machine learning. Appl. Ocean Res. 117, 102920. http://dx.doi.org/10.1016/j.apor.2021.102920.

Morris, E., Zienkiewicz, H., Belmont, M., 1998. Short term forecasting of the sea surface shape. Int. Shipbuil. Prog. 45 (444), 383–400.

Naaijen, P., Wijaya, A., 2014. Phase resolved wave prediction from synthetic radar images. In: International Conference on Offshore Mechanics and Arctic Engineering. Vol. 45509. American Society of Mechanical Engineers, V08AT06A045.

Neill, S.P., Hashemi, M.R., 2018. Chapter 7 - in situ and remote methods for resource characterization. In: Neill, S.P., Hashemi, M.R. (Eds.), Fundamentals of Ocean Renewable Energy. In: E-Business Solutions, Academic Press, pp. 157–191. http://dx.doi.org/10.1016/B978-0-12-810448-4.00007-0.

Niekamp, R., Niemann, J., Schröder, J., 2023. A surrogate model for the prediction of permeabilities and flow through porous media: A machine learning approach based on stochastic Brownian motion. Comput. Mech. 71 (3), 563–581. http://dx.doi.org/10.1007/s00466-022-02250-2.

Ongie, G., Jalal, A., Metzler, C.A., Baraniuk, R.G., Dimakis, A.G., Willett, R., 2020. Deep learning techniques for inverse problems in imaging. IEEE J. Sel. Areas Inf. Theory 1 (1), 39–56.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E.Z., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library. arXiv:1912.01703.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A., 2016. Context encoders: Feature learning by inpainting. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, IEEE Computer Society, Los Alamitos, CA, USA, pp. 2536–2544. http://dx.doi.org/10.1109/CVPR.2016.278.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., et al., 2022. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. arXiv:2202.11214.

Peng, W., Yuan, Z., Wang, J., 2022. Attention-enhanced neural network models for turbulence simulation. Phys. Fluids 34 (2), 025111. http://dx.doi.org/10.1063/5.0079302.

Perlin, M., Bustamante, M., 2014. A robust quantitative comparison criterion of two signals based on the Sobolev norm of their difference. J. Eng. Math. 101, 115–124.

Raschka, S., 2018. Model evaluation, model selection, and algorithm selection in machine learning. arXiv:1811.12808.

Rashid, M.M., Pittie, T., Chakraborty, S., Krishnan, N.A., 2022. Learning the stress-strain fields in digital composites using Fourier neural operator. iScience 25 (11), 105452. http://dx.doi.org/10.1016/j.isci.2022.105452.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Computer and information sciences, FOS: Computer and information sciences, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

Salcedo-Sanz, S., Nieto Borge, J., Carro-Calvo, L., Cuadra, L., Hessner, K., Alexandre, E., 2015. Significant wave height estimation using SVR algorithms and shadowing information from simulated and real measured X-band radar images of the sea surface. Ocean Eng. 101, 244–253. http://dx.doi.org/10.1016/j.oceaneng.2015.04.041.

Stender, M., Ohlsen, J., Geisler, H., Chabchoub, A., Hoffmann, N., Schlaefer, A., 2023. U^p-Net: A generic deep learning-based time stepper for parameterized spatio-temporal dynamics. Comput. Mech. http://dx.doi.org/10.1007/s00466-023-02295-x.

Stoian, A., Poulain, V., Inglada, J., Poughon, V., Derksen, D., 2019. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. Remote Sens. 11 (17).

Valenzuela, G.R., 1978. Theories for the interaction of electromagnetic and oceanic waves — A review. Bound.-Lay. Meteorol. 13, 61–85. http://dx.doi.org/10.1007/BF00913863.

Vicen-Bueno, R., Lido-Muela, C., Nieto-Borge, J.C., 2012. Estimate of significant wave height from non-coherent marine radar images by multilayer perceptrons. EURASIP J. Appl. Signal Process. 2012 (1), http://dx.doi.org/10.1186/1687-6180-2012-84.

Wang, F., Eljarrat, A., Müller, J., Henninen, T.R., Erni, R., Koch, C.T., 2020. Multi-resolution convolutional neural networks for inverse problems. Sci. Rep. 10 (1), 1–11.

Wang, R., Yu, R., 2023. Physics-guided deep learning for dynamical systems: A survey. arXiv:2107.01272.

Wedler, M., Stender, M., Klein, M., Ehlers, S., Hoffmann, N., 2022. Surface similarity parameter: A new machine learning loss metric for oscillatory spatio-temporal data. Neural Netw. 156, 123–134. http://dx.doi.org/10.1016/j.neunet.2022.09.023.

Wedler, M., Stender, M., Klein, M., Hoffmann, N., 2023. Machine learning simulation of one-dimensional deterministic water wave propagation. Ocean Eng. 284, 115222. http://dx.doi.org/10.1016/j.oceaneng.2023.115222.

Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A., Benson, S.M., 2022. U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow. Adv. Water Resour. 163, 104180. http://dx.doi.org/10.1016/j.advwatres.2022.104180.

West, B.J., Brueckner, K.A., Janda, R.S., Milder, D.M., Milton, R.L., 1987. A new numerical method for surface hydrodynamics. J. Geophys. Res.: Oceans 92 (C11), 11803–11824. http://dx.doi.org/10.1029/JC092iC11p11803.

Wijaya, A., Naaijen, P., Andonowati, van Groesen, E., 2015. Reconstruction and future prediction of the sea surface from radar observations. Ocean Eng. 106, 261–270. http://dx.doi.org/10.1016/j.oceaneng.2015.07.009.

Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V., 2022. Integrating scientific knowledge with machine learning for engineering and environmental systems. arXiv:2003.04919.

Williamson, J., Brigido, H., Ladeira, M., Souza, J.C.F., 2022. Fourier neural operator for image classification. In: 2022 17th Iberian Conference on Information Systems and Technologies. CISTI, pp. 1–6. http://dx.doi.org/10.23919/CISTI54924.2022.9820128.

Wu, G., 2004. Direct Simulation and Deterministic Prediction of Large-Scale Nonlinear Ocean Wave-Field (Ph.D. thesis). Massachusetts Institute of Technology. Department of Ocean Engineering.

Wu, M., Stefanakos, C., Gao, Z., 2020. Multi-step-ahead forecasting of wave conditions based on a physics-based machine learning (PBML) model for marine operations. J. Mar. Sci. Eng. 8 (12), http://dx.doi.org/10.3390/jmse8120992.

Yan, B., Chen, B., Robert Harp, D., Jia, W., Pawar, R.J., 2022. A robust deep learning workflow to predict multiphase flow behavior during geological CO2 sequestration injection and post-injection periods. J. Hydrol. 607, 127542. http://dx.doi.org/10.1016/j.jhydrol.2022.127542.

Yang, Y., Gao, A.F., Castellanos, J.C., Ross, Z.E., Azizzadenesheli, K., Clayton, R.W., 2021. Seismic wave propagation and inversion with neural operators. Seism. Rec. 1 (3), 126–134.

Yevnin, Y., Toledo, Y., 2022. A deep learning model for improved wind and consequent wave forecasts. J. Phys. Oceanogr. 52 (10), 2531–2537. http://dx.doi.org/10.1175/JPO-D-21-0280.1.

You, H., Zhang, Q., Ross, C.J., Lee, C.-H., Yu, Y., 2022. Learning deep implicit Fourier neural operators (IFNOs) with applications to heterogeneous material modeling. Comput. Methods Appl. Mech. Engrg. 398, 115296. http://dx.doi.org/10.1016/j.cma.2022.115296.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 818–883.

Zhang, J., Zhao, X., Jin, S., Greaves, D., 2022. Phase-resolved real-time ocean wave prediction with quantified uncertainty based on variational Bayesian machine learning. Appl. Energy 324, 119711. http://dx.doi.org/10.1016/j.apenergy.2022.119711.

Zhang, K., Zuo, W., Gu, S., Zhang, L., 2017. Learning deep CNN denoiser prior for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3929–3938.