

EXPERIMENTAL EVALUATION OF DATA-DRIVEN DETERMINISTIC WAVE GENERATION

Mathies Wedler^{1,*}, Marc-André Pick², Robert Seifried², Norbert Hoffmann³, Sören Ehlers¹, Marco Klein¹

¹German Aerospace Center (DLR), Institute of Maritime Technologies and Propulsion Systems, Geesthacht, Germany

²Hamburg University of Technology, Institute of Mechanics and Ocean Engineering, Hamburg, Germany

³Hamburg University of Technology, Dynamics Group, Hamburg, Germany

ABSTRACT

This paper explores the applicability of data-driven deterministic wave generation in an experimental environment. Various datasets that comprise different combinations of experimental- and synthetic data are used to train a fully convolutional neural network on the task of predicting the wave-maker control signal associated to a desired target wave sequence at a predefined target location in the wave tank. The trained models are evaluated in an experimental program, where the predicted control signals are executed in a wave tank, the actual wave is measured and compared against the predefined target wave. The standard wave modeling approach for the generation of tailored wave sequences, which is based on linear theory, is conducted for reference. The test dataset contains 950 irregular sea states with a wide range of characteristic sea state parameters, as well as seven tailored extreme waves, including the well-known “Draupner Wave”. It is shown that the data-driven approach consistently outperforms the standard method in accuracy across the studied range of sea states, emphasizing its effectiveness as a flexible and precise alternative for wave basin experiments.

Keywords: machine learning, wave generation, experimental evaluation, tailored wave sequences

1. INTRODUCTION

Model tests on waves in laboratory environments are still essential for assessing the impact of waves on ships and offshore structures, thus advancing research in hydrodynamics, coastal engineering, and renewable energy systems. Stochastic irregular sea states based on design wave spectra allow for realistic representations of the chaotic natural sea and can be tuned in terms of the characteristic sea state parameters to involve wave lengths and -heights that are assumed to be critical. In order to reduce statistical uncertainties in terms of short-term and long-term statistics of the system response, a set of irregular sea states with different initial phase distributions is typically investigated.

A more elaborate approach to model tests is the generation of critical wave groups, i.e., phase-resolved sequences of waves that trigger critical system responses. Approaches to generate tailored wave sequences range from short critical wave groups derived from classical design wave concepts [1–4] or preceding numerical simulations [5–8] to embed critical wave sequences, e.g., single high waves or dangerous steep wave groups, into irregular wave sequences of short duration. The standard ocean wave model [9] is widely used in test facilities due to its simplicity. It represents a sea state as the superposition of independent linear harmonic wave components governed by the linear dispersion relation, allowing the sea state to be described in both space and time. Within this framework, a wave sequence defined at a specific location in the wave tank (the target location) can be transformed using linear wave theory. To derive the wave maker control signal, the desired wave sequence is therefore translated to the wave maker position and scaled by the wave maker response amplitude operator [10–12]. More advanced models are necessary to accurately generate or reproduce predefined wave sequences at specific locations in a wave tank, as steep wave groups violate the assumptions of linear wave theory. Consequently, wave sequences produced using the standard model can exhibit significant deviations at the target location due to nonlinear wave propagation. These advanced approaches aim to compute a wave maker control signal by adjusting wave amplitudes and phases so that the wave train generated at the wave board evolves into the desired sequence at the target position. However, existing methods are largely complex, empirical, or computationally demanding, which limits their applicability as a simple plug-and-play solution for model test facilities (e.g., [13–29]).

Recent studies propose the use of machine learning (ML) to learn the nonlinear relation between the target wave and the wave at the wave-maker [30], or even the wave-maker control signal itself [31], from data. ML methods have been shown to be *universal function approximators* that are able to learn any continuous function to arbitrary accuracy from observations [32, 33].

*Corresponding author: mathies.wedler@dlr.de

TABLE 1: Integration of previous work in this study. The model architecture is adopted from [30, 31] either 1:1 (*full*) or in a *reduced* form, cf. Fig. 3. The composition of the respective training data is broken down into exact number of samples. *The synthetic data from [30] is combined with the RAO of the wave-maker [31, 35].

FCNN model		training data			
name	variant	synthetic* [30]		experimental [31]	
$\mathcal{M}_{100:0}$	full	63 000	(100 %)	0	(0 %)
$\mathcal{M}_{0:100}$	full	0	(0 %)	9 669	(100 %)
$\mathcal{M}_{100:100}$	full	63 000	(100 %)	9 669	(100 %)
$\mathcal{M}_{0:5}$	reduced	0	(0 %)	483	(5 %)
$\mathcal{M}_{0:15}$	reduced	0	(0 %)	1450	(15 %)
$\mathcal{M}_{100:5}$	reduced	63 000	(100 %)	483	(5 %)
$\mathcal{M}_{100:15}$	reduced	63 000	(100 %)	1450	(15 %)

Supervised ML methods require labeled data, i.e., pairs of inputs and outputs that encode the behavior of the underlying system. Klein et al. [31] propose to exploit this general approach to learn the relation between the target wave and the associated wave-maker control signal. In fact, the data-driven identification of an experimental wave tank setup by means of neural networks bypasses the need for explicit hydrodynamic modeling by learning the complex relations between wave-maker dynamics, wave surface dynamics, and wave basin response from measurement data.

In this paper, a fully convolutional neural network (FCNN) [34] is deployed to generate tailored waves in an experimental setup. The FCNN architecture is trained to map the wave sequence at a target location to the associated wave-maker control signal. The same architecture is fit to a variety of training datasets that comprise different combinations of experimental- and synthetic data. The trained ML models follow a naming convention, where $\mathcal{M}_{a:b}$ indicates the FCNN trained with a % of available synthetic data and b % of available experimental data. For instance, $\mathcal{M}_{100:100}$ is trained on the largest dataset comprising all synthetic and all experimental data. Table 1 lists all model variants and the associated training data studied throughout this paper. The baseline FCNN architecture was first introduced in [30] to map the target wave to the wave at the location of the wave-maker, and subsequently re-used in [31] to map the target wave to the control signal of the wave-maker. The synthetic data was generated in [30] and is here combined with the response amplitude operator (RAO) of the wave-maker [31, 35] to establish a direct mapping between the target wave and the wave-maker control signal. All models studied in this paper are trained from scratch using the training data listed in Tab. 1.

The trained models are compared against the standard approach based on analytical wave-maker transfer functions (linear) and linear wave theory (LWT) to propagate the wave from the target location to the wave-maker as starting point of the wave generation process. All methods are validated experimentally by generating short deterministic wave sequences that are compared to the predefined target wave at target location. The experimental evaluation is conducted on 950 irregular waves within

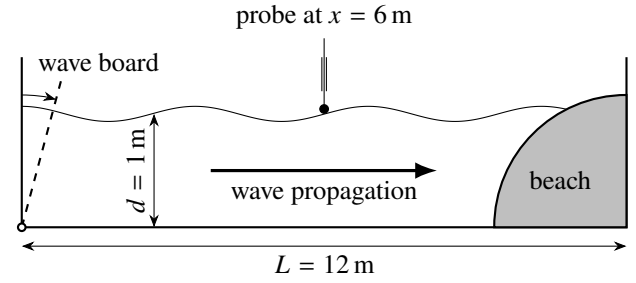


FIGURE 1: Schematic of the wave tank that was used for the experimental data generation [30, 31] and validation. The wave probe measures the time series of the surface elevation at $x = 6$ m.

a wide range of sea state parameters spanned by varying JON-SWAP [36] peak enhancement factors, angular peak frequencies and wave steepness. Furthermore, a set of 7 tailored extreme waves, including the famous Draupner Wave, is used to evaluate the methods on the generation of extreme events.

This paper starts off with Sect. 2 describing the generation of the synthetic- and experimental datasets used for training of the supervised ML approach. The ML framework, including the model architecture and the training setup, is subsequently introduced in Sect. 3, before the results are presented in Sect. 4. Final conclusions as well as an outlook are given in Sect. 5.

2. DATA GENERATION

The field of ML aims at building prediction models by extracting knowledge from data. Classical supervised ML methods have no physical priors but attempt to learn causal relations between inputs and outputs purely from observations. In other words, supervised ML relies on labeled data, i.e., pairs of reference inputs and target outputs that implicitly describe the governing dynamics of an arbitrary system. Within this data-centered perspective, the quality of the data becomes crucial, as it eventually determines what the model is presented during training and thus able to learn. In general, ML models have limited capabilities to extrapolate beyond their training distribution. Consequently, the training data acts as an essential resource but, at the same time, also as a limiting factor.

This paper aims at the data-driven generation of tailored waves in a wave tank. The underlying ML models were trained to predict the wave-maker control signal based on a desired wave sequence at a target location. The data generation was based around the wave tank located at Hamburg University of Technology. For the experimental evaluation, the predicted control signals were executed in the wave tank, where the actual wave was measured and then compared against the target wave. Figure 1 provides a longitudinal schematic of the wave tank, which has a length of $L = 12$ m, a profile of 1.5×1.5 m, and a water depth of $d = 1$ m. It uses a flap-type wave-maker with the pivot point at the sea bed. A beach is installed at the end of the tank facing the wave-maker to attenuate wave reflections.

A *synthetic* and an *experimental* approach to the data generation form the baseline for this paper. Both data types were generated independently in the scope of previous works [30, 31] and are here used either separately or in combination. Figure 2 provides

a schematic overview of the data generation processes, marks the either linear or nonlinear relation between quantities, and indicates which signals are ultimately used for supervised ML. The synthetic data generation process and the resulting trained ML model is marked in light blue, the experimental counterpart is marked in yellow. Trained models based on any combination of both datasets are marked in green.

The synthetic approach employed the high-order spectral (HOS) method [37, 38] as a non-linear numerical wave tank [30], which mimicked the actual wave tank, cf. Figure 1. The spatio-temporal non-linear wave simulation was sampled at (i) the position of the wave-maker $\eta_0(t)$, and (ii) the target location $\eta_t(t)$. In addition, the time series of the wave at the position of the wave-maker was transformed to the control signal $\xi(t)$ via linear RAO of the wave-maker. Thus, the control signal $\xi(t)$ and the wave sequence at target location $\eta_t(t)$ formed the data base for the synthetic approach which included the linear relation between the wave sequence at wave-maker and wave-maker motion as well as non-linear (HOS) wave propagation between wave-maker and target location.

The experimental data generation approach involved an extensive experimental program, where a large number of irregular sea states with varying characteristic sea state parameters were generated in the wave tank itself [31]. The wave generation process followed the standard approach by defining the respective wave sequence at target location, transforming the target wave to the wave-maker linearly and multiplying the wave sequence at the wave-maker with the wave-maker RAO to obtain the control signal. However, the control signal $\xi(t)$ that was used to generate the wave and the actual wave measured at a target location ($\eta_t(t)$) form the basis for supervised ML training.

In the following, the generation of training datasets for ML based on the synthetic and experimental approach are described in detail. For an extensive in-depth description of the data generation procedure, the interested reader is referred to the original publications on the synthetic- [30] and experimental datasets [31]. The derivation of the wave-maker RAO used in this paper can be found here [31, 35].

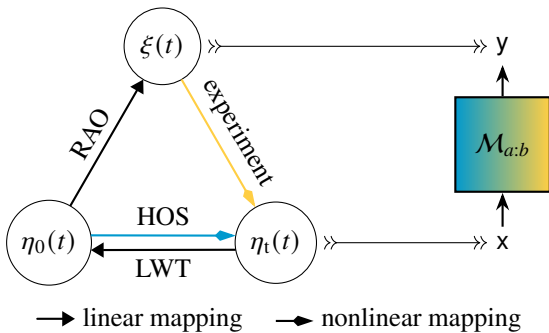


FIGURE 2: Schematic of the synthetic and experimental data generation for supervised ML. Both approaches share the same linear RAO [31, 35] to translate the wave at the wave-maker location $\eta_0(t)$ to the control signal $\xi(t)$. The nonlinear relation between the target wave $\eta_t(t)$ and the control signal $\xi(t)$ originates in the HOS method, or the actual experimental run, respectively.

2.1. Synthetic data generation

The HOS wave model [37, 38] is a numerically efficient method to simulate the propagation of an initial wave field including all, resonant and non-resonant, physically existing nonlinear interactions. The HOS method is based on potential flow theory and assumes an irrotational and incompressible Newtonian fluid, such that the velocity potential in the fluid domain ϕ satisfies the Laplace equation, $\nabla^2 \phi = 0$. The HOS method employs the Hamiltonian formalism [39] to express the motion at the free surface as a boundary value problem in the kinematic-

$$\frac{\partial \eta}{\partial t} = -\nabla \phi^S \nabla \eta + W \left(1 + |\nabla \eta|^2 \right), \quad (1)$$

and the dynamic boundary conditions

$$\frac{\partial \phi^S}{\partial t} = -g\eta - \frac{1}{2} |\nabla \phi^S|^2 + \frac{1}{2} W^2 \left(1 + |\nabla \eta|^2 \right), \quad (2)$$

with ϕ^S denoting the velocity potential at the free surface and W the vertical velocity in the fluid domain. Under the assumption that ϕ^S and η are quantities of $\mathcal{O}(\eta^n)$, the solution for W can be expressed as a Taylor series expansion around $z = 0$, yielding

$$W^{(m)} = - \sum_{n=1}^m \frac{\eta^n}{n!} \frac{\partial^{n+1}}{\partial z^{n+1}} \phi^{(m-n)} \quad (3)$$

where

$$\phi^{(m)} = \begin{cases} \phi^S & \text{for } m = 0, \text{ and} \\ - \sum_{n=1}^m \frac{\eta^n}{n!} \frac{\partial^n}{\partial z^n} \phi^{(m-n)} & \text{for } m > 0, \end{cases} \quad (4)$$

with $M = m + 1$ denoting the order of nonlinearity. The HOS method has been validated against wave tank experiments [40, 41], and is widely employed in academia for the physically accurate simulation of two- [30, 42–44] and three-dimensional wave fields [45–47].

To generate the synthetic data, the HOS method with series expansion up to the fourth order ($M = 4$) was initialized with irregular sea states generated based on the TMA form [48] of the JONSWAP spectrum [36], which accounts for finite water depth. The parameter space was spanned by three JONSWAP peak enhancement factors $\gamma \in \{1, 3, 6\}$, seven peak angular frequencies $\omega_p \in \{3, 4, 5, 6, 7, 8, 9\} \text{ rad s}^{-1}$ and six wave steepness $\varepsilon = \frac{\pi H_s}{L_p} \in \{0.10, 0.075, 0.05, 0.0375, 0.025, 0.0125\}$, resulting in a total of 126 unique sea state parameter combinations. For each combination, 500 irregular sea states were initialized with a random phase, such that overall, 63 000 unique irregular sea states were used as initial condition for the HOS simulation. A stratified data split w.r.t. the governing sea state parameters was applied, such that the final training dataset consisted of 44 086 spatio-temporal wave simulations $\eta(x, t)$. Next, $\eta(x, t)$ was evaluated at $x = 0 \text{ m}$ to obtain the time series of the wave at the wave-maker $\eta_0(t)$, and $x = 6 \text{ m}$ to get the time series at the target location $\eta_t(t)$ [30]. Linking back to Fig. 2, $\eta_0(t)$ formed the foundation of the synthetic data generation process, based on which (i) the target wave was derived by nonlinear propagation using the HOS method, and (ii) the control signal $\xi(t)$ was obtained

by applying the linear RAO of the wave-maker to the Fourier transform of $\eta_0(t)$,

$$\eta_t(t) \xleftarrow[\text{nonl.}]{\text{HOS}} \eta_0(t) \xrightarrow[\text{lin.}]{\text{RAO}} \xi(t).$$

Ultimately, the labeled dataset for supervised ML is given by $(x, y) = (\eta_t(t), \xi(t))$.

2.2. Experimental data generation

As aforementioned, the experimental dataset was generated in an extensive experimental campaign. For this purpose, a fully automated wave generation process, covering the determination of a target wave, the calculation of the respective control signal, execution of the experimental run, recording and storage of the measurement data, and a waiting period to allow the water to calm in between experiments, was employed. The target wave $\eta_t(t)$ was generated based on the TMA form [48] of the JONSWAP spectrum [36]. The space for characteristic sea state parameters was spanned by three JONSWAP peak enhancement factors $\gamma \in \{1, 3, 6\}$, seven peak angular frequencies $\omega_p \in \{3, 4, 5, 6, 7, 8, 9\} \text{ rad s}^{-1}$ and ten wave steepness $\varepsilon \in \{0.17, 0.16, 0.15, 0.14, 0.13, 0.12, 0.11, 0.10, 0.075, 0.05\}$. The wave steepness range for the experimental data generation was chosen to complement the range for the synthetic data generation, which is intentionally limited to $\varepsilon \leq 0.1$ to ensure that the steepness of all waves in the randomly initialized sea state is below the wave breaking limit [31]. The time series of the target wave were initialized with random phase. The length of the time series were dynamically adapted to the underlying sea state parameter (with an upper limit of 60 s) to mitigate reflections for smaller ω_p , i.e., higher group velocities. To calculate the associated control signal $\xi(t)$, the target wave $\eta_t(t)$ was transferred to Fourier space using the discrete Fourier transform, where all frequency components were linearly propagated to the position of the wave-maker. Here, the linear RAO of the wave-maker was applied, before the physical control signal $\xi(t)$ was obtained via the inverse discrete Fourier transform. Then, the derived control signal was used to execute an experimental run. A wave probe measured the wave surface elevation time series at the target location, cf. Fig. 1. Finally, the measured wave $\hat{\eta}_t(t)$ was stored alongside the characteristic sea state parameters. A waiting period of 5 minutes, allowing the water to calm down in between experiments, marked the end of each experimental cycle. A stratified data split w.r.t. the governing sea state parameters was applied to the 13 814 experimental runs that were identified as valid. The final training dataset consists of 9 669 measured waves $\hat{\eta}_t(t)$ and the respective control signals $\xi(t)$. To align with Fig. 2, the process includes three sequential steps

$$\eta_t(t) \xrightarrow[\text{lin.}]{\text{LWT}} \eta_0(t) \xrightarrow[\text{lin.}]{\text{RAO}} \xi(t) \xrightarrow[\text{nonl.}]{\text{exp.}} \hat{\eta}_t(t),$$

where $(x, y) = (\hat{\eta}_t(t), \xi(t))$ is the labeled dataset for supervised ML training. Overall, the experimental data generation process was in line with the standard approach for wave generation, with the decisive difference, that the usage of the measured wave $\hat{\eta}_t(t)$ for the training induces a nonlinear relation to the control signal

$\xi(t)$. In the following, the subscript $[\cdot]_t$ is dropped and all instances of the target wave η and the measured wave $\hat{\eta}$ refer to the respective quantity at the target location at $x = 6 \text{ m}$.

3. MACHINE LEARNING FRAMEWORK

In this section, the ML framework that was employed to learn the relation between the target wave $\eta(t)$ and the associated control signal $\xi(t)$ for the wave-maker is introduced. First, the model architecture that was adopted from previous works [30, 31] is described. Then, the training environment is introduced. This section closes with a brief evaluation of the training of the models trained with fully synthetic-, fully experimental-, and a merged dataset including all synthetic and experimental data in terms of the loss curves.

3.1. Model architecture

In this study, a FCNN [34] with an encoder-decoder-structure [49] is used to learn a mapping between the target wave sequence $\eta(t)$ and the control signal $\xi(t)$. Convolutions are designed to efficiently extract local features from coherent data that is structured in arrays. A convolutional layer applies a trainable filter bank, the so-called kernel, to the input. By convolving a small kernel with the input signal, the same set of weights is applied to all positions of the input. This unique way of parameter sharing makes convolutional layers *equivariant to translation*, which refers to the ability to detect recurring pattern at all positions in the input [49, 50]. The output of a convolutional layer can be interpreted as a map that indicates where a particular feature is detected in the input, and is therefore commonly called *feature map*. Fig. 3 shows the architecture of the FCNN, where the intermediate feature maps are schematically depicted by white square. The model follows a common design philosophy, where the number of channels increases with increasing network depth. The encoder path is built from blocks, that extract features on the current scale of the data and perform a down-sampling operation by means of average pooling, thus bisecting the sequence length. In other words, the encoder trades sequence length or *physical resolution* for feature resolution, until the feature density reaches its maximum in the bottleneck. Here, the information in the initial input x is expressed as a set of rich high-level features. Based on this dense information, the decoder has the task to model the desired output y . The decoder path consists of five blocks that perform a convolution and then up-sample the information by a factor of two using a transposed convolution operator (Conv^T) [51]. The actual dimensions of the feature maps in the forward path of the model (sequence length $\times$ channels) depend on the depth of the model and the size of the input, cf. Fig. 3. All convolutional layers whose main purpose is the extraction of feature use exponential linear unit (ELU) activation [52], the transposed convolutions used for up-sampling deploy linear activation. Overall, the model has 110 825 trainable parameter.

3.2. Model training

The training is conducted using the surface similarity parameter loss function [42, 53, 54]

$$\text{SSP} : \mathbb{R}^N \rightarrow \mathbb{R}^+ : (y, \hat{y}) \mapsto \frac{\|y - \hat{y}\|}{\|y\| + \|\hat{y}\|} \in [0, 1], \quad (5)$$

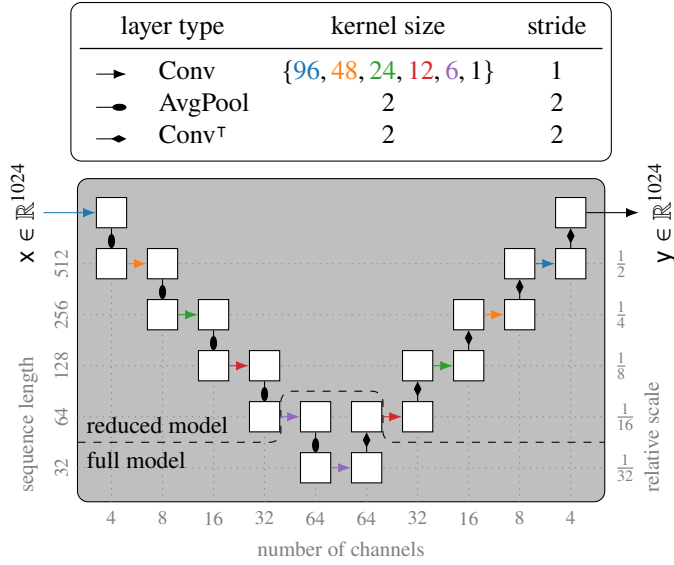


FIGURE 3: Fully convolutional neural network with encoder-decoder structure used for the sequence-to-sequence prediction task. The grid indicates the number of channels and the sequence length (relative scale of the data w.r.t. the input x) in the respective layer of the model. Two variants of the architecture are used, a *full* model (depth of five) and a *reduced* model (depth of four).

where $SSP = 0$ indicates perfect agreement among y and $\hat{y}$, and $SSP = 1$ can only be obtained for phase inverted signals and comparison of a non-zero signal against zero. Being a normalized error metric, the SSP is invariant to the scale or amplitude of the signal. The SSP is based on the Sobolev norm [55]

$$\|y(\cdot)\|^2 = \int_{-\infty}^{\infty} |Y(f)|^2 (1 + |f|^2)^k df, \quad (6)$$

which for $k = 0$ induces an isometry between the physical space and the complex Fourier space L^2 norms. Consequently, despite being formulated in the physical space, the SSP has the unique property of calculating the error between two signals y and $\hat{y}$ in the complex Fourier space, thus inherently penalizing deviations in magnitude and phase. When used as a loss function in supervised ML, a weight update has to improve the prediction in terms of both quantities in order to reduce the loss. In comparison, Euclidean distance-based loss metrics like the mean squared error (MSE) are overly sensitive to amplitude errors and penalize phase errors only implicitly [42]. The rather strict loss formulation of the SSP pairs well with the Adam optimizer [56], which builds up momentum when the gradients are near zero for an extended number of consecutive training steps. Here, the Adam optimizer is employed with learning rate $\alpha = 0.001$ and decay rates $\beta_1 = 0.9$, $\beta_2 = 0.999$. The training is conducted with an early stopping policy, which terminates the training process either once the validation loss has converged (no improvement for 100 consecutive epochs), or after a maximum of 5000 training epochs [31].

Each of the datasets introduced in Sect. 2 is used to train the FCNN introduced in Sect. 3.1. Figure 4 visualizes the training distribution for all models in terms of characteristic sea state parameters by means of a 2D histogram over the peak angular frequency ω_p and the steepness ε . The color encodes the sam-

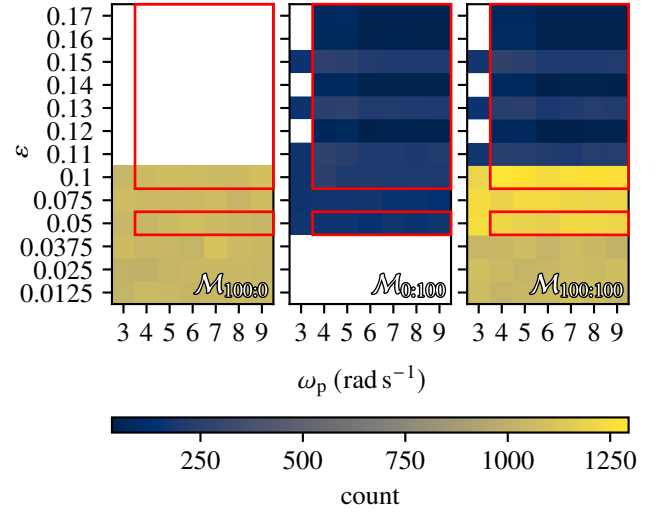


FIGURE 4: Training distribution for ML models (valid for $\gamma \in \{1, 3, 6\}$). The color indicates the number of samples for a specific combination of ω_p and ε . Areas outlined by red indicate the distribution for experimental evaluation.

ple count (blue: few samples, yellow: many samples). Areas outlined in red indicate the distribution of characteristic sea state parameters in the test dataset that is later used for the experimental evaluation. In particular, although the test dataset has the same characteristic sea state parameters as the experimental training dataset, it contains irregular sea states with random phases that are not present in any of the training datasets. The depicted distributions are valid for all JONSWAP peak enhancement factors $\gamma \in \{1, 3, 6\}$ ¹. Other than the training data, all models share the same hyperparameters. Figure 5 shows the loss curves for $M_{100:0}$, $M_{0:100}$, and $M_{100:100}$. The loss curve visualizes the evolution of the training- and validation loss over the training epochs, and is a basic but important tool to assess the training progress as well as the data fit. A potential problem with large ML models is their massive capacity, which allows them to memorize the training data. If this is the case, the model will perform poorly on unseen data, a behavior that will ultimately manifest in diverging training- and validation loss curves. Here, the validation loss (dashed lines) follows the training loss (solid lines) for all models, indicating no overfitting but satisfactory data fits. All loss curves exhibit a similar shape, where the loss decreases rapidly within approximately the first 10 % of training epochs before it slowly converges towards a constant SSP limit. The different number of training epochs until convergence aligns with the relative size of the underlying training datasets: $M_{100:0}$ arrives at a minimal validation loss of $SSP = 0.0202$ after 1104 training epochs. In contrast, $M_{0:100}$ takes 791 training epochs to arrive at a minimal validation loss of $SSP = 0.0877$. $M_{100:100}$ is trained for the largest number of epochs and arrives at a minimal validation loss in between $M_{100:0}$ and $M_{0:100}$ ($SSP = 0.0352$ after 1539 training epochs). It is important to revisit the fact that all models, in particular $M_{100:0}$ and $M_{0:100}$, were trained on a different data

¹The data for the experimental evaluation is limited to $\gamma \in \{1, 6\}$.

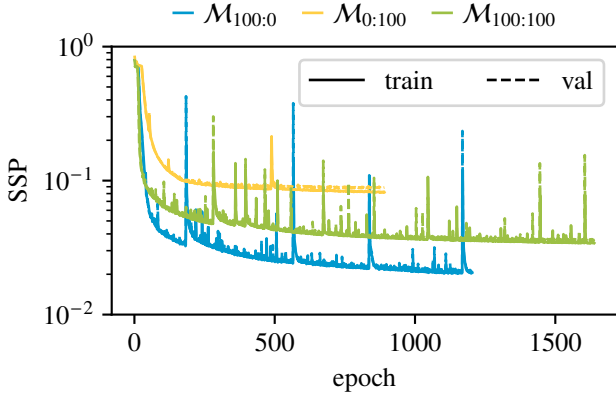


FIGURE 5: Loss curves for ML models. Solid lines indicate the training loss, dashed lines the validation loss.

basis, including a different distribution of characteristic sea state parameters, cf. Fig. 4. To recall, the synthetic data covers wave steepness up to $\varepsilon = 0.10$. In contrast, the wave steepness in the experimental data ranges within $\varepsilon \in [0.05, 0.17]$, with finer resolution for $\varepsilon \geq 0.10$. Since the impact of nonlinear effects correlates with the severity of the sea state, the experimental data overall contains way more nonlinear interactions, making it generally more complex. Consequently, a comparison between the loss curves of the different models has only little significance.

In the end, all models were trained to predict the control signal necessary to generate a certain target wave. The HOS method used to generate the synthetic wave data, albeit being validated against wave tank experiments [40, 41], remains a numeric method that relies on model assumptions. Thus, it is valid to assume that the synthetic data does not represent the real-world water wave dynamics to the same extent as the experimental data does. Moreover, the training distribution of the synthetic data overlaps only partly with the distribution for experimental evaluation, cf. red areas in Fig. 4. Generally, the performance of an ML model correlates with the amount of training data. However, the fact that the synthetic training data covers a different parameter space than the test data ultimately means that the models trained on synthetic data have to generalize way beyond their “comfort zone”. If the synthetic data would cover the same parameter space used for the experimental evaluation, the models trained with synthetic data would be assumed to perform best, simply because they saw about four times the amount of unique samples during training, cf. Sect. 2. Nonetheless, as the parameter space for the experimental evaluation is fully covered by the experimental training data, cf. Fig. 4, it is expected that the model trained on the experimental data will perform best. While the accuracy of the predicted control signal is generally a good indicator for the accuracy of the target wave, cf. Fig. 5, the real-world impact must be assessed in the physical world, i.e., the wave tank. For this, the predicted control signals are used in the experimental setup to generate a wave, measure it at the target location, and compare it to the predefined target wave (used as input for the different ML models).

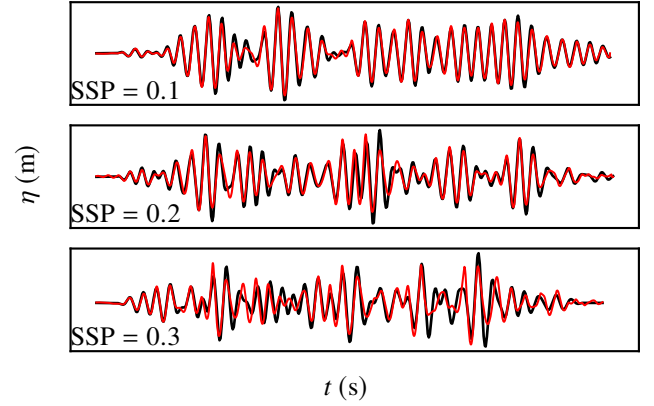


FIGURE 6: Exemplary target- (black) and measured waves (red) visualize deviations quantified by $\text{SSP} \in \{0.1, 0.2, 0.3\}$.

4. EXPERIMENTAL EVALUATION

For the evaluation, all trained ML models were inferred on the same test dataset to predict the control signal $\hat{\xi}(t)$ to a predefined target wave $\eta(t)$,

$$\eta(t) \xrightarrow{\mathcal{M}_{a:b}} \hat{\xi}(t).$$

The predicted control signals $\hat{\xi}(t)$ are then used to generate a wave in the experimental wave tank setup, cf. Fig. 1. The generated wave $\hat{\eta}(x, t)$ is measured at the target location $x = 6$ m, and compared against the target wave $\eta(t)$ using the SSP metric, cf. Eq. (5). For reference, the traditional workflow for tailored wave generation is conducted, where the target wave sequence is linearly propagated from the target location to the location of the wave-maker, before it is translated into the control signal $\hat{\xi}(t)$ using the linear RAO of the wave-maker [9]. To provide some intuition on the SSP metric, Fig. 6 shows three exemplary pairs of target waves (black) and measured waves (red). From an empirical standpoint, the depicted deviations mark the boundaries for satisfactory ($\text{SSP} \leq 0.1$), acceptable ($0.1 < \text{SSP} < 0.2$), and unacceptable ($\text{SSP} \geq 0.2$) accuracy.

The distribution of characteristic sea state parameters in the test dataset are outlined in red in Fig. 4. In contrast to the training datasets, the test dataset is limited to JONSWAP peak enhancement factors $\gamma \in \{1, 6\}$. It is important to note that the parameter space of the experimental training dataset covers the whole test dataset, while the distribution of the synthetic training dataset and the test dataset overlap for $\varepsilon \in \{0.05, 0.10\}$ only. The different distribution of sea state parameters in the training datasets was initially motivated by the assumption that the HOS provides sufficient fidelity in the low steepness range, thus effectively rendering the time-consuming and thus costly generation of an experimental dataset dispensable [31]. In a real-world scenario, the comparatively cheap synthetic dataset would be used complementary to the costly experimental dataset. However, as briefly discussed at the end of Sect. 3, the vastly different coverage of the training datasets w.r.t. characteristic sea state parameter in the test dataset create an advantage for the model trained with experimental data.

This section first addresses postprocessing routines applied to the raw measurement data. Then, the evaluation first focuses

on the phase-resolved generation of irregular sea states, before the tailored generation of extreme waves, including the Draupner Wave, is investigated.

4.1. Postprocessing of raw measurement data

The raw measurement data exhibit a non-zero z -offset, indicating that the wave gauge sensor had some constant offset to the mean water level in the wave tank. The mean difference over the first 1000 time steps (where the target wave sequence is zero and the measured wave exhibits negligible zero-mean fluctuations) is used to estimate and compensate the z -offset in each measured wave. Moreover, a time delay between the measured and the target wave was observed. From an application-driven perspective, if the actual wave form matches the target wave, a small time delay could be considered irrelevant. The SSP metric, on the other hand, is very sensitive to phase-errors and would consequently yield misleading results for time-delayed signals. To account for this, all measured signals $\hat{\eta}$ are shifted relative to the target wave η . The optimal time-delay compensation within ± 0.2 s is obtained by minimizing

$$\min_{\kappa} \text{SSP}(\eta(t), \hat{\eta}(t + \kappa \delta t)), \quad \kappa \in \mathbb{Z}, \quad -20 \leq \kappa \leq 20, \quad (7)$$

with $\delta t = 0.01$ s. Fig. 7 shows the histogram of the optimal κ for all waves. Both Sturges' rule [57] and the Freedman-Diaconis rule [58] are used to calculate the optimal number of histogram bins based on statistics of the underlying distribution. Under the assumption of a normally distributed bin count, a probability density function (PDF) is fitted to each data. The histogram of the discrete time delay steps κ varies strongly between models, cf. Fig. 7. While the distribution for $\mathcal{M}_{100:0}$ is approximately zero-mean, the distributions of the remaining ML models exhibit a mean of $\kappa \approx -6$ ($\kappa \approx 5$ for LWT). In other words, the waves generated using $\mathcal{M}_{0:100}$ and $\mathcal{M}_{100:100}$ arrived on average prior to the actual wave at the target location, while the waves generated using the linear approach arrived with a slight delay. The positive time delay for the linear approach can be linked to the fact that linear theory underestimates the wave celerity due to the neglect of non-linear, steepness-dependent corrections. However, all data-driven models were trained on data that involves a nonlinear relation between the target wave and the associated control signal. Still, the time delay does exist for models that involve experimental data only. In fact, a similar behavior has been observed in preliminary work, where 36 % of the the control signals predicted by the model trained with experimental data were contaminated by an almost constant time delay of $|dt| = 0.05$ s [31]. This leads to the assumption, that the negative time delay observed for models $\mathcal{M}_{0:100}$ and $\mathcal{M}_{100:100}$ is rooted in a not fully synchronized measurement setup during the experimental data generation. The fact that the observed time delay of $-6\delta t = -0.06$ s aligns well with the time delay observed in [31] suggests that the models inherently learned a temporal offset that is encoded in the experimental training data. This effect is less pronounced for the model trained on the merged dataset due to the impact of both experimental- (temporal offset) and synthetic (no temporal offset) data. A last postprocessing step involves the isolation of the actual wave, i.e., the removal of calm sea before and after the wave sequence. For

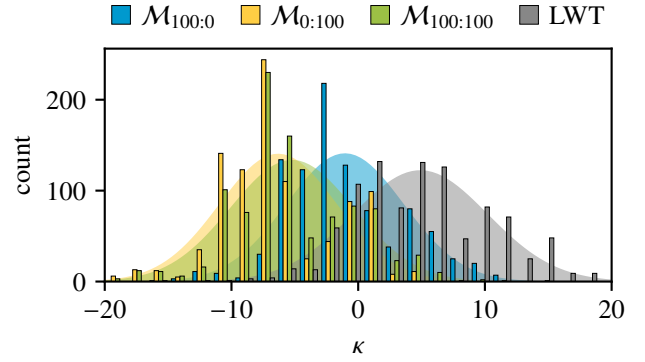


FIGURE 7: Histogram of shifts applied to the predictions in order to minimize the SSP.

this, a binary mask is derived from the time derivative of the target wave $\eta(t)$, indicating either calm sea (0) or wave (1). The valid area of the binary mask (1) is extended by ± 1.5 s in each direction. Then, all sampling points of the target- and measured waves where the mask is (0) are removed, leaving only the isolated wave for evaluation (see, e.g., time series depicted in Fig. 6). Hereafter, the offset- and time-delay compensated measured wave sequences are used exclusively for evaluation.

4.2. Evaluation on irregular sea states

The evaluation of the proposed data-driven methods to phase-resolved wave generation in a wave tank was conducted purely on SSP, cf. Eq. (5). Based on the differences in training distributions and particularly the aforementioned coverage of the test distribution, models trained with experimental data are expected to perform better than the model trained exclusively on synthetic data. More specifically, $\mathcal{M}_{0:100}$ and $\mathcal{M}_{100:100}$ are expected provide more accurate results than $\mathcal{M}_{100:0}$. The merged dataset enriches the experimental dataset in the low steepness range, cf. Fig. 4. At the same time, it adds a huge number of samples that lie outside the test distribution. The following evaluation will show, whether the synthetic samples outweigh the relatively small number of experimental samples in the merged dataset, or if the datasets do indeed complement each other.

Fig. 8 shows the deviation between the target wave $\eta(t)$ and the measured wave $\hat{\eta}(t)$ quantified in terms of SSP for $\gamma = 1$, resolved over the characteristic sea state parameters angular peak frequency ω_p and wave steepness ε . Fig. 8 separated into four panels, each dedicated to either a trained ML model, or the linear reference approach (cf. label in lower right corner of the panels). The color of the tiles indicates the mean deviation in terms of SSP over all sea states characterized by the respective sea state parameter combination (ω_p, ε) , with dark colors indicating low SSP and bright colors indicating higher SSP. Moreover, each panel states the mean SSP and standard deviation over all tiles, i.e., the studied parameter space. With a mean error of $\text{SSP} = 0.11$, $\mathcal{M}_{0:100}$ exhibits the overall best results, followed by $\mathcal{M}_{100:100}$ ($\text{SSP} = 0.14$), $\mathcal{M}_{100:0}$ ($\text{SSP} = 0.19$), and finally the linear reference approach with $\text{SSP} = 0.24$. A shared trend is visible among the panels, where the SSP-quantified error between the target and the measured wave is generally lowest for high ω_p and low steepness ε ,

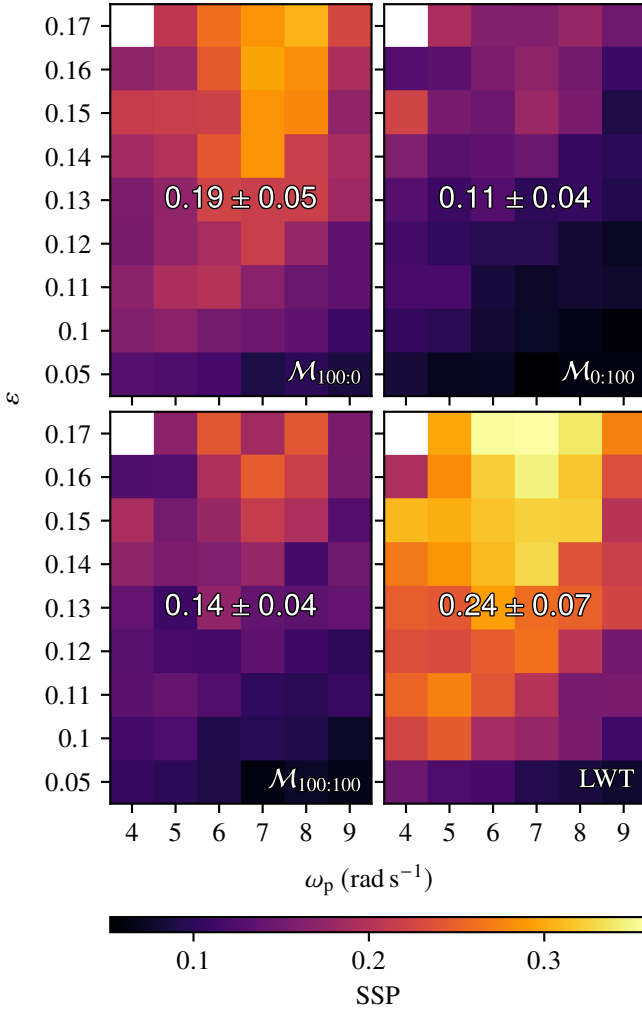


FIGURE 8: Deviation between target wave and measured wave at target location in terms of SSP resolved over the characteristic sea state parameters ω_p and ε for $\gamma = 1$. Each panel states the respective model as well as the mean and standard deviation SSP.

and largest for within the area of the parameter space spanned by $\omega_p \in [6, 8]$ and $\varepsilon \geq 0.14$. Likewise, Fig. 9 shows the results for $\gamma = 6$. While the results qualitatively resemble the same characteristics described for $\gamma = 1$, the SSP-quantified errors are overall approximately 21 % (27 % for $M_{0:100}$) lower. As a result, the only instance where the mean error undercuts $\text{SSP} = 0.1$ is observed for model $M_{0:100}$, indicating overall very high agreement among the target wave and the measured wave. The correlation between the JONSWAP peak enhancement factor and the deviation between the target and actual wave is most certainly rooted in the more focused spectrum for higher γ [36], and has been observed in prior works [43, 45, 47]. A more focused spectrum is equivalent to a more focused energy distribution among frequency modes, which ultimately leads to less wave components that contribute to a sea state, such that individual uncertainties become less relevant.

Fig. 10 provides a statistical perspective on the results, where the SSP-quantified error for all 950 sea states in the test dataset

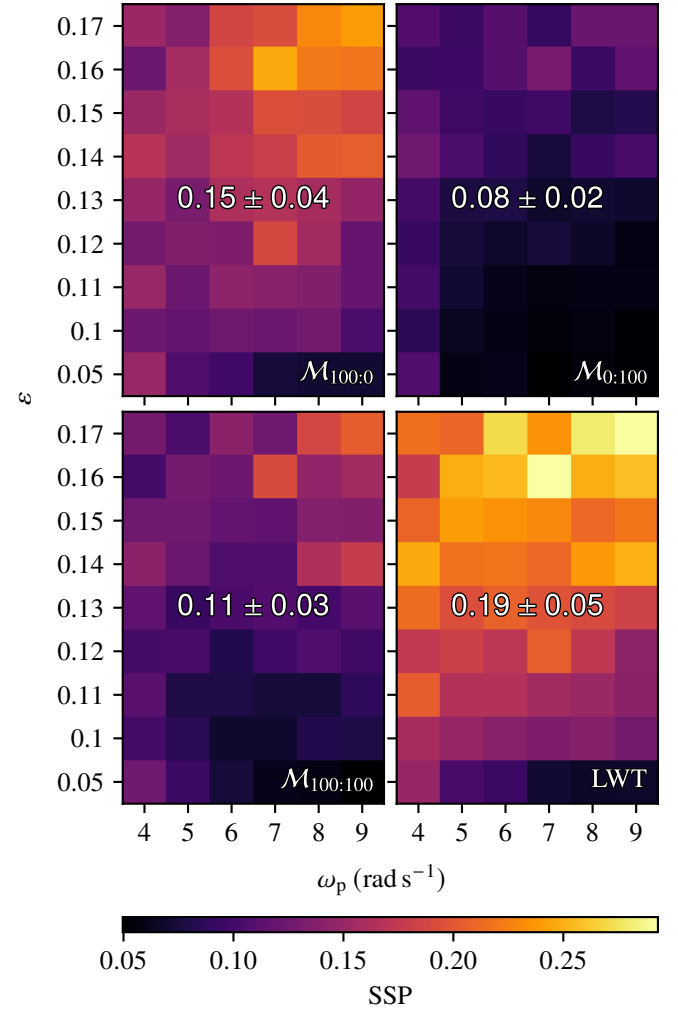


FIGURE 9: Deviation between target wave and measured wave at target location in terms of SSP resolved over the characteristic sea state parameters ω_p and ε for $\gamma = 6$. Each panel states the respective model as well as the mean and standard deviation SSP.

are visualized in a box-plot, grouped by the model that was used to generate the control signal. The statistics for each model are reported twofold, where the left-hand side half of each box (darker hue) shows the results for $\gamma = 1$, and the right-hand side half (lighter hue) for $\gamma = 6$. Statistical outliers, i.e., all values $\{x \in X \mid x \notin [Q_1(X) - 1.5 \text{ IQR}(X), Q_3(X) + 1.5 \text{ IQR}(X)]\}$ with Q_i the i^{th} quartile and IQR the interquartile range of the distribution X , are marked with triangles. Additionally, a violin-plot outlines the kernel density estimation of the underlying distribution, thus ensuring unimodality, hence confirming that all distributions have a dominant central cluster.

While the kernel density estimate of SSP-quantified errors approaches a normal distribution for LWT, the distributions of the data-driven methods are skewed towards lower SSP values. As a result, the data-driven methods exhibit far more statistical outlier towards the high SSP range compared to the linear approach. Overall, the box-plots confirm that $M_{0:100}$ provides the best results among the investigated methods. For $\gamma = 6$, Q_3 is

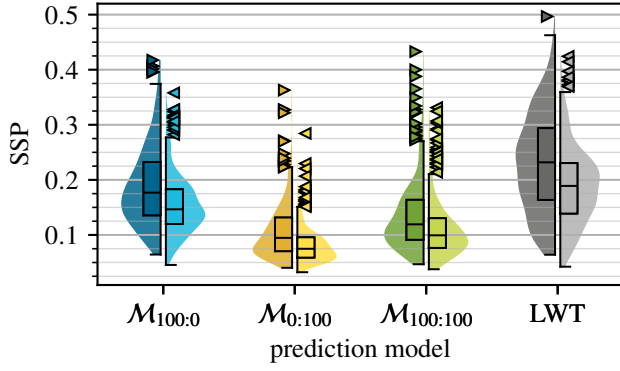


FIGURE 10: Distribution of deviation in terms of SSP between target wave and measured wave at target location for irregular waves. Each box is divided into the statistics for $\gamma = 1$ (left, darker hue) and $\gamma = 6$ (right, lighter hue).

at $SSP < 0.10$, indicating that 75 % of all measured waves are attributed an error $SSP < 0.10$. For $\gamma = 1$, Q_3 is at $SSP \approx 0.125$, which is slightly worse but still indicates high agreement among the target wave and the measured wave. Furthermore, the interquartile ranges do not overlap among $M_{100:0}$ and $M_{0:100}$, indicating a pronounced separation in the central tendency between the results of the respective models. Within the models trained on experimental data, the distribution for $\gamma = 1$ of $M_{0:100}$ closely resembles the distribution for $\gamma = 6$ of $M_{100:100}$, indicating the same level performance on the arguably more complex case.

4.3. Evaluation on extreme waves

In this section, all models are evaluated on a selection of extreme events. Extreme events evolve when the energy of multiple wave components is focused in a single point in space and time. The resulting single large wave usually imposes the highest structural loads, and is therefore of particular interest for model tests, while the surrounding sea state is less important. To evaluate the fidelity of the measured wave locally, a sliding-window measure based on the SSP is defined as

$$SSP(\eta, \hat{\eta})(\tau) = SSP(\eta_{\tau:\tau+n-1}, \hat{\eta}_{\tau:\tau+n-1}) \quad (8)$$

with window length n . A cosine-tapered window [59] is applied to the windowed signals to impose periodic boundary conditions and thus preventing the excitation of high modes before the SSP is calculated. The resulting sequence hold $N - n + 1$ SSP values which are left aligned with the original time series of length N . Padding the time-distributed error SSP to the left- and right-hand side ensures matching length and correct alignment relative to the time series $\eta(t)$ and $\hat{\eta}(t)$. The window size in Eq. (8) is defined twofold: A narrow window of size T_p is designed to approximately match the temporal extent of the extreme event itself, while a wider window of size $2.5T_p$ covers the target wave, as well as the crest and trough of the leading and tailing wave.

Figure 11 shows the target time series and the respective power spectral densities (PSD) of all tailored extreme waves studied in this paper, namely the Draupner Wave, Alwyn North, and two envelope solitons. The famous Draupner Wave has been

registered on January 1st, 1995 at the Draupner platform in the North Sea [60]. It is characterized by an extraordinarily high single wave with a maximum crest height of $H_{\max} = 25.63$ m in a surrounding sea state of $H_s = 11.92$ m. The Alwyn North wave was recorded in 1997 at the Alwyn North oil and gas field in the northern North Sea, about 100 miles east of the Shetland islands [61]. Similar to the Draupner Wave, it is characterized by an isolated high single wave. Both the Draupner and the Alwyn North target waves are designed in model scales 1:140 and 1:280. For the smaller scale versions, the peak angular frequency is shifted towards the higher frequency range, such that a noteworthy share of the spectral energy lies beyond the upper limit of the training distribution in ω , cf. red areas in Fig. 11. Moreover, the Draupner Wave in scale 1:140 is investigated with two different time series length. The shorter time series, besides introducing less reflections in the experimental evaluation, matches the length of the time series in the training data better, and generally contains less information. It is therefore expected that the data-driven models perform better on the short time series. In addition, all models are evaluated on a parametrized envelope soliton, which can be considered a prototype for critical wave groups at offshore structures [62]. Here, the soliton solution is investigated with $\omega_p = 6 \text{ rad s}^{-1}$ and steepness $\varepsilon \in \{0.25, 0.30\}$. The red vertical lines in the PSD plots in Fig. 11 indicate the range of ω_p covered by the training distributions, cf. Fig. 4. It is important to note that extreme waves were not explicitly portrayed in any training dataset. Consequently, even though the peak angular frequencies are covered by the training distributions, the models have to perform beyond the cases they saw during training, i.e., irregular waves. Nevertheless, statistically, the training datasets may inadvertently contain extreme waves due to the initialization of the sea state with a random phase. The occurrence of extreme waves in the training data and their impact on the trained models is not part of this paper and will be investigated in future studies. The left-hand side of Tab. 2 summarizes the wave parameters, including the model scale of the wave and the length of the time series (if applicable). The right-hand side of Tab. 2 provides the deviation between target wave and measured wave within the two studied intervals around the peak of the extreme event itself. The best result per row and interval is marked in bold font. Overall, the results indicate a strong correlation between the SSP in the narrow and wider interval around the peak, indicating that the extreme event and the surrounding sea is replicated to approximately the same fidelity. At this, $M_{0:100}$ exhibits equal or slightly better performance on the narrow interval in six out of seven cases, indicating that the model has a tendency to replicate the extreme event itself ever so slightly better than the surrounding sea state. In contrast, $M_{100:0}$ performs equal or slightly better on the wider window in six out of seven cases. The model trained on the merged dataset, $M_{100:100}$, is overall well balanced with a slight tendency towards better results on the wider interval spanning $2.5T_p$. The assessment of the overall accuracy aligns well with the results in irregular sea states presented in Sect. 4.2, where the models utilizing the experimental data for training perform best. In the case of extreme events, $M_{0:100}$ achieves the best performance in five out of seven waves. For the remaining two waves, the Draupner Wave (scale 1:280)

and the Alwyn North wave (scale 1:140), $\mathcal{M}_{100:100}$ performs best and $\mathcal{M}_{0:100}$ 2nd best. $\mathcal{M}_{100:0}$ performs overall well, but the SSP values spike up for the Draupner Wave and Alwyn North wave in scale 1:280. A similar behavior can be observed for $\mathcal{M}_{0:100}$, where in particular the performance on the Draupner Wave with scale 1:280 is way below its average SSP score. This observation aligns well with the PSD plots depicted in Fig. 11, which reveal that 26 % of the total energy of the Draupner Wave (1:280) (39 % for Alwyn North, 1:280) is distributed over angular frequencies beyond the range of ω_p covered by the training distribution. Nevertheless, $\mathcal{M}_{100:100}$ performs surprisingly well in those cases with $\text{SSP} = 0.08$ for both investigated intervals. The linear approach yields the worst results across all studied waves except for Alwyn North (1:280), where the performance of $\mathcal{M}_{100:0}$ spikes to $\text{SSP} = 0.43$ (0.33). This is expected, as the linear approach does not model nonlinear terms, which become more impactful with increasing wave steepness. Overall, the results emphasize the benefit of the proposed data-driven approach over standard approach for the accurate generation of extreme events. Owing to the fact that no training distribution explicitly contained extreme events, all data-driven models performed remarkably well. In fact, the results strongly suggest that all models learned the nonlinear dynamics of water waves that is deeply encoded in the training data, as extreme waves, despite being a rare event, do obey the laws of physics.

In the following, the performance on the famous Draupner Wave (1:140, short time series length) is investigated in greater depth. The upper panel in Fig. 12 shows the target wave (black) alongside the measured waves generated using the control signal predicted by $\mathcal{M}_{100:0}$, $\mathcal{M}_{100:100}$, and $\mathcal{M}_{0:100}$, as well as the linear reference approach (LWT). The temporal extent of the windows spanning T_p (solid line) and $2.5T_p$ (dashed line) are marked in red, the associated SSP sequences obtained from Eq. (8) are shown in the lower panel. The larger window of $2.5T_p$ has a smoothing effect on SSP, which manifests in minimal fluctuations in the dashed lines. To recall, the accuracy within the T_p interval around the peak obtained by the studied methods quantify to

$$\text{SSP} = \begin{cases} 0.10 & \text{for } \mathcal{M}_{100:0}, \\ 0.08 & \text{for } \mathcal{M}_{0:100}, \\ 0.14 & \text{for } \mathcal{M}_{100:100}, \text{ and} \\ 0.22 & \text{for LWT.} \end{cases} \quad (9)$$

The accuracies evaluated in the $2.5T_p$ interval around the peak exhibit negligible deviations within $\Delta\text{SSP} = \pm 0.01$ to the results stated in Eq. (9), and are thus not discussed in detail here.

Within the limited scope framing the extreme event, all measured waves generally align well in phase with the target wave, cf. Fig. 12. However, the magnitude of the measured waves deviate strongly among the methods. The measured wave associated to $\mathcal{M}_{0:100}$ exhibits high correlation with the peak of the Draupner Wave and the preceding trough, i.e., the trough to the right-hand side of the peak in Fig. 12. This is further emphasized by the SSP sequences shown in the lower panel, where the solid line yellow associated with $\mathcal{M}_{0:100}$ undercuts the remaining lines for $t \in [31.5, 33.2]$ s. $\mathcal{M}_{100:100}$ manages to recreate the trough and wavefront of the Draupner Wave to high accuracy, but fails at

the actual peak. Likewise, the wave associated to $\mathcal{M}_{100:0}$ shows high correlation for the trough and the wavefront, but underestimates the peak of the Draupner Wave even more. From visual inspection, $\mathcal{M}_{100:100}$ recreates the actual extreme event better than $\mathcal{M}_{100:0}$. Remarkably, the SSP-quantified errors in Eq. (9) indicate a superior performance for $\mathcal{M}_{100:0}$. This is attributed to the tailing slope of the wave (i.e., left to the peak in Fig. 12), which is far better captured by the measured wave associated to $\mathcal{M}_{100:0}$. Consequently, comparing the SSP sequences for later time steps, i.e., once the tailing slope is not covered by the window anymore, the results for $\mathcal{M}_{100:100}$ yield lower SSP values than the results for $\mathcal{M}_{100:0}$. The respective cross-over in the SSP sequences happens at $t \approx 33.0$ s, cf. lower panel in Fig. 12. The linear approach produces an “extreme event” that does not stand out from the surrounding waves, neither in terms of steepness, nor crest height. Moreover, the measured wave associated to LWT exhibits a sudden peak around $t = 31.5$ s, while all data-driven methods correctly provide a smooth slope. Overall, the results presented in Fig. 12 emphasize the superiority of nonlinear data-driven approaches in comparison to the classical linear approach. For a global perspective, Fig. 14 in Appx. A shows the full time series of the measured and target waves studied here.

4.4. Investigation of dataset composition

The results presented in Sect. 4.2 and 4.3 show that the control signal predicted by the ML model trained exclusively on experimental data yields the most accurate tailored wave at the target location. However, the experimental generation of a dataset that meets the quantitative and qualitative requirements to train an ML model is time-consuming, therefore costly, and thus generally not feasible for real-world applications. This section investigates the effect of the size and composition of the training dataset on the accuracy of the generated wave in an experimental environment. For this, two subsets incorporating 5 % (483 training samples) and 15 % (1450 training samples) are randomly sampled from the available experimental data, both maintaining full coverage of the parameter space in γ , ω_p and ε . There is an overlap of 337 samples that exist in both subsets, i.e., there are 1113 samples that are unique to the larger dataset and 146 samples that are unique to the smaller dataset. Both subsets are merged with all available synthetic data and an FCNN prediction model is trained on the four new datasets. The severe difference in size among the datasets motivated the use of a simplified model architecture for the 5 % and 15 % experimental datasets. This simplified model qualitatively resembles the FCNN introduced in Sect. 3.1, but uses a network depth of four levels, while the models trained on the larger merged datasets employ the original architecture with network depth five levels. All hyperparameter remain consistent with the initial model architecture. Following the naming convention introduced in Sect. 3, the trained models are referred to as $\mathcal{M}_{0:5}$, $\mathcal{M}_{0:15}$, $\mathcal{M}_{100:5}$, and $\mathcal{M}_{100:15}$.

Fig. 13 shows the deviations in terms of SSP between the target wave and the measured waves for irregular sea states with JONSWAP peak enhancement factor $\gamma = 6$. Remarkably, the models trained on the now very limited experimental datasets provide the best accuracy w.r.t. the measured waves, with $\mathcal{M}_{0:15}$ performing slightly better than $\mathcal{M}_{0:5}$. This being said, all models

TABLE 2: Wave parameters and results (deviation between target wave and measured wave in terms of SSP) for extreme waves. The SSP is evaluated in a window spanning T_p ($2.5T_p$) s around the peak of the wave.

Name	Wave parameters				SSP within T_p ($2.5T_p$) around extreme event							
	ω_p	ε	scale	length	$\mathcal{M}_{100:0}$		$\mathcal{M}_{0:100}$		$\mathcal{M}_{100:100}$		LWT	
Alwyn North	6 rad s ⁻¹	0.14	1:140	baseline	0.12	(0.11)	0.10	(0.10)	0.08	(0.09)	0.36	(0.29)
Alwyn North	8 rad s ⁻¹	0.15	1:280	baseline	0.43	(0.33)	0.12	(0.14)	0.14	(0.14)	0.35	(0.30)
Draupner Wave	6 rad s ⁻¹	0.08	1:280	baseline	0.24	(0.20)	0.22	(0.19)	0.08	(0.08)	0.28	(0.27)
Draupner Wave	4 rad s ⁻¹	0.08	1:140	baseline	0.10	(0.09)	0.06	(0.07)	0.18	(0.18)	0.23	(0.25)
Draupner Wave	4 rad s ⁻¹	0.08	1:140	short	0.10	(0.10)	0.08	(0.08)	0.14	(0.13)	0.22	(0.23)
Envelope Soliton	6 rad s ⁻¹	0.30	—	—	0.13	(0.13)	0.07	(0.07)	0.12	(0.10)	0.16	(0.21)
Envelope Soliton	6 rad s ⁻¹	0.25	—	—	0.12	(0.13)	0.04	(0.05)	0.10	(0.07)	0.12	(0.15)

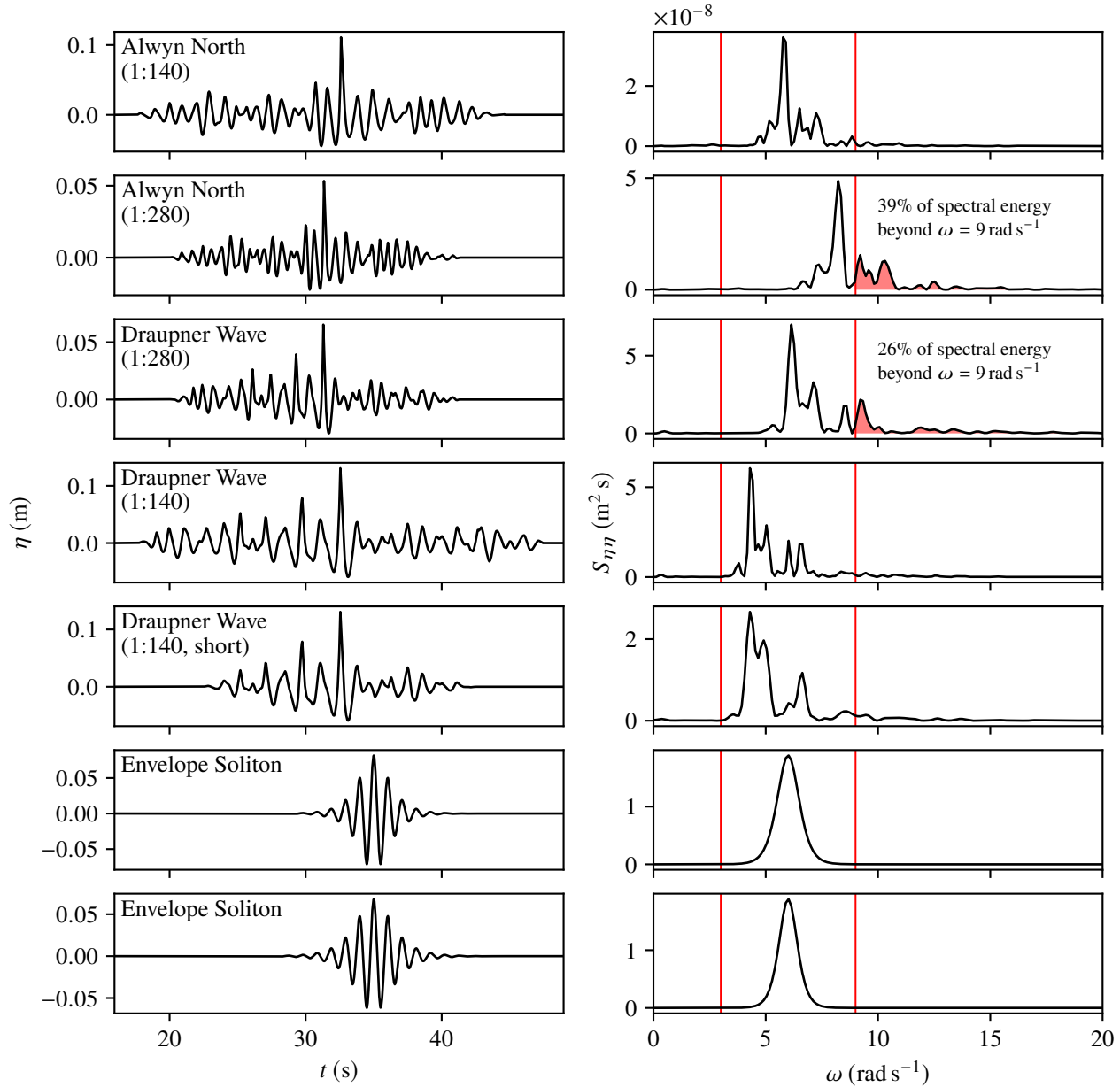


FIGURE 11: Tailored extreme waves studied in this paper (left) alongside the corresponding power spectral densities (PSD, right). The red vertical lines in the PSD plots indicate the range of ω_p in the training distribution. Instances where more than 25% of the total energy is beyond the training distribution in terms of ω_p are marked in red. Additional parameter for the extreme waves are summarized in Tab. 2.

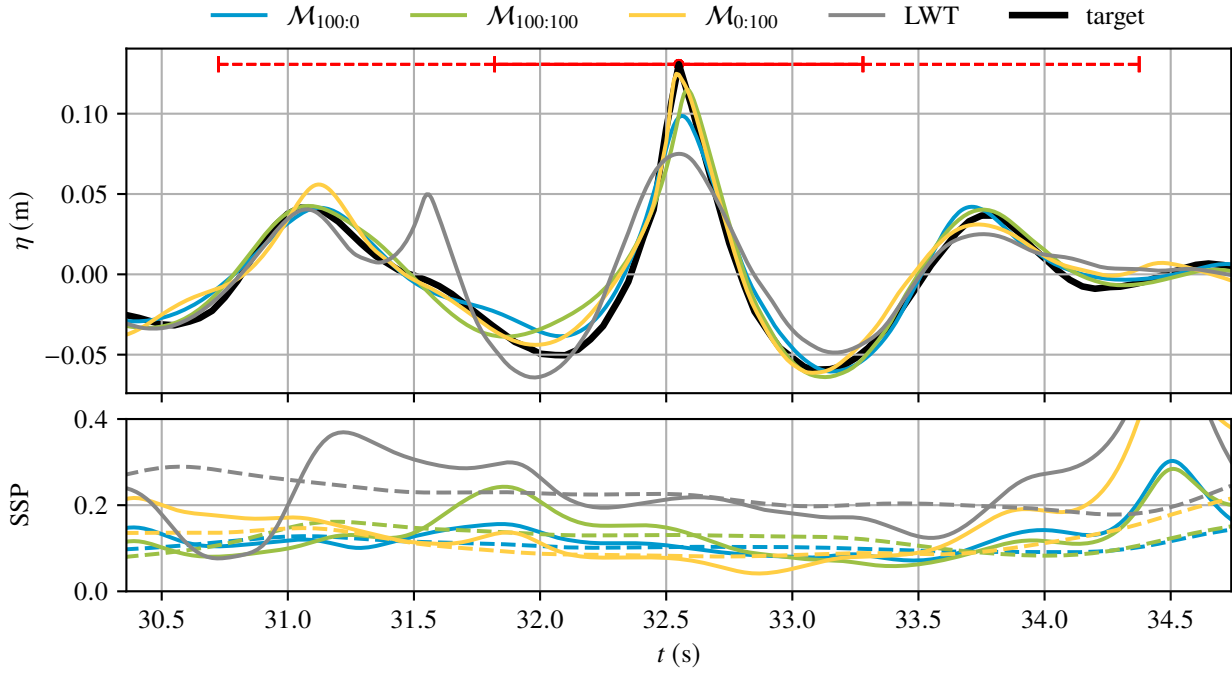


FIGURE 12: Draupner Wave (1:140), short time series. The upper panel shows the wave in time domain, the lower panel shows the SSP evaluated over a sliding window spanning $1T_p$ (solid lines) and $2.5T_p$ (dashed lines).

perform within a narrow band $SSP \in [0.10, 0.14]$, which suggests no significant differences in accuracy among the models. The qualitative trends priorly observed in Sect. 4.2, namely best performance for high ω_p and ε and worst performance for high ε , remain.

From an application-driven perspective, the additional cost of generating a larger experimental dataset does not pay off, as the difference in performance among $M_{0:15}$ and $M_{0:5}$ is negligible. Moreover, the incorporation of synthetic data results in worse performance. One reason (besides the different model architecture) may be the parameter distribution of the synthetic data, which, as already mentioned before, does not cover the test distribution, cf. Fig. 4. The comparatively small number of waves in the reduced experimental datasets, which do share the same parameter distribution as the test dataset, may be outweighed by the immense number of numerically generated waves, effectively shifting the focus of the model out of the test distribution and thus ultimately leading to worse performance despite way more training data.

The results for extreme waves are summarized in Tab. 3. The fidelity of the measured wave is again quantified within a window of T_p (and $2.5T_p$) around the peak of the extreme event. Overall, $M_{0:5}$ achieves the highest accuracy in four out of seven of the studied waves. $M_{100:5}$ achieves the best results on the Solitons, $M_{0:15}$ yields the most accurate wave for the New Year Wave with scale 1:280. Despite being trained with the largest amount of data, $M_{100:15}$ performs worst in five out of seven cases. Figure 15 in Appx. A shows the full time series of the measured and target waves studied here.

Overall, the results are quite surprising, in particular with respect to $M_{0:5}$. A rule of thumb in ML states that the per-

formance of the model generally correlates with the amount of training data. Since all experimental data originates from the same source, homogeneous quality can be assumed. Moreover, both fully experimental datasets share the same probability density w.r.t. the characteristic sea state parameters. Hence, the only valid explanation for the observed behavior lies in a *lucky sampling*, i.e., the 5 % sampled from the experimental data, or even the 114 waves that are unique to the smallest dataset, are presumably characterized by high variance in terms of the contained deterministic wave sequences. In the end, the results suggest that not only the size of the dataset, but also its composition are a key factor for success.

5. CONCLUSIONS AND OUTLOOK

This paper presented the experimental evaluation of a data-driven approach to deterministic wave generation. A shared fully convolutional neural network (FCNN) model architecture was trained on various datasets, consisting in synthetic data that was generated using the high-order spectral (HOS) method, experimental dataset that was measured in extensive wave tank experiments, and merged datasets that combine the two aforementioned sources of data. An FCNN was trained on each dataset to predict the wave-maker control signal based on a desired target wave. Moreover, a classical approach that relies on linear wave theory to map the target wave to the control signal of the wave-maker was employed for reference. The predicted control signals obtained from all approaches were executed in a wave-tank and the actual wave at the target location was measured and compared with the desired target wave. The test data involved 950 irregular sea states over a wide range of characteristic sea state parameters, as well as 7 tailored extreme waves, including the famous Draup-

TABLE 3: Wave parameters and results (deviation between target wave and measured wave in terms of SSP) for extreme waves. The SSP is evaluated in a window spanning T_p ($2.5T_p$) s around the peak of the wave.

Name	Wave parameters				SSP within T_p ($2.5T_p$) around extreme event							
	ω_p	ε	scale	length	$\mathcal{M}_{0:15}$		$\mathcal{M}_{0:5}$		$\mathcal{M}_{100:15}$		$\mathcal{M}_{100:5}$	
Alwyn North	6 rad s ⁻¹	0.14	1:140	baseline	0.11	(0.15)	0.08	(0.10)	0.17	(0.19)	0.14	(0.13)
Alwyn North	8 rad s ⁻¹	0.15	1:280	baseline	0.22	(0.19)	0.21	(0.19)	0.27	(0.24)	0.22	(0.20)
Draupner Wave	6 rad s ⁻¹	0.08	1:280	baseline	0.14	(0.12)	0.18	(0.17)	0.23	(0.20)	0.20	(0.17)
Draupner Wave	4 rad s ⁻¹	0.08	1:140	baseline	0.12	(0.11)	0.09	(0.08)	0.17	(0.19)	0.13	(0.12)
Draupner Wave	4 rad s ⁻¹	0.08	1:140	short	0.14	(0.12)	0.08	(0.09)	0.21	(0.21)	0.09	(0.09)
Envelope Soliton	6 rad s ⁻¹	0.30	—	—	0.15	(0.13)	0.12	(0.10)	0.13	(0.14)	0.10	(0.09)
Envelope Soliton	6 rad s ⁻¹	0.25	—	—	0.12	(0.09)	0.07	(0.07)	0.09	(0.10)	0.05	(0.08)

ner Wave (single wave with crest height of $H_{\max} = 25.63$ m) in model scale.

The experimental evaluation showed that the machine learning (ML) models trained on the experimental data provided the most accurate results for both test cases, followed by the models trained on the merged dataset, then came the models trained on the synthetic dataset, and finally the linear approach. The results emphasize that a nonlinear mapping between the target wave and the control signal is essential, in particular for the generation of tailored extreme waves. The satisfactory performance of all data-driven models coupled with the fact that extreme events were *not explicitly portrayed* in the training data strongly suggests that the models were able to learn the nonlinear dynamics of water wave propagation. Afterwards, a brief study on the composition of the datasets was presented which showed that the models trained on greatly reduced versions of the experimental datasets still performed better than models trained on merged datasets, which combined the reduced experimental datasets with all available synthetic data. One outcome of this study suggests that the variance among the training samples outweighs the sheer number of training samples.

The HOS method employed for the generation of the synthetic data is only valid until the wave breaking limit and thus imposed an upper limit on the wave steepness in the synthetic dataset. The choice of characteristic sea state parameter for the experimental evaluation had only minimal overlap with the distribution of characteristic sea state parameters in the synthetic training data, in particular the wave steepness. As a consequence, the models trained on synthetic data had to generalize way beyond their training distribution, leading to sub-optimal performance. In theory, synthetic data does hold immense potential to augment a small experimental dataset at very low cost. However, the results indicated the model trained on the merged dataset did not benefit from the additional data. This may be due to the large amount of synthetic data beyond the parameter space relevant for this study outweighing the comparatively small amount of data in the relevant parameter space. Future studies may therefore involve a more balanced dataset to explore the potential of merged datasets for real-world experiments.

REFERENCES

[1] Torhaug, Rune, Winterstein, Steven R. and Braathen, Arne.

- “Nonlinear Ship Loads: Stochastic Models for Extreme Response.” *Journal of Ship Research* Vol. 42 No. 01 (1998): pp. 46–55. DOI [10.5957/jsr.1998.42.1.46](https://doi.org/10.5957/jsr.1998.42.1.46).
- [2] Tromans, Peter S., Anaturk, Ali R. and Hagemeyer, Paul. “A New Model For The Kinematics Of Large Ocean Waves-Application As a Design Wave.” Vol. The First International Offshore and Polar Engineering Conference: pp. ISOPE–I–91–154. 1991.
- [3] Hansen, Peter Friis and Nielsen, Lars Peter. “On the New Wave Model for Kinematics of Large Ocean Waves.” *Proceedings of the ASME 1995 14th International Conference on Ocean, Offshore and Arctic Engineering*. 1995. American Society of Mechanical Engineers Digital Collection.
- [4] Clauss, Günther F and Kühnlein, Walter L. “A new tool for seakeeping tests - nonlinear transient wave packets.” (1997).
- [5] Oberhagemann, Jan, Ley, Jens, Shigunov, Vladimir and el Moctar, Ould. “Efficient Approaches for Ship Response Statistics using RANS.” 2012. DOI [10.13140/2.1.4689.0884](https://doi.org/10.13140/2.1.4689.0884).
- [6] Ley, Jens, Oberhagemann, Jan, Amian, Christoph, Langer, Markus, Shigunov, Vladimir, Rathje, Helge and Schellin, Thomas. “Green Water Loads on a Cruise Ship.” *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE*, Vol. 7. 2013. DOI [10.1115/OMAE2013-10132](https://doi.org/10.1115/OMAE2013-10132).
- [7] Klein, Marco. “Tailoring critical wave sequences for response based design.” Ph.D. Thesis, Technische Universität Berlin. 2015. DOI [10.14279/depositonce-4504](https://doi.org/10.14279/depositonce-4504). Accepted: 2015-11-21T00:41:10Z.
- [8] Kim, Dae-Hyun and Troesch, Armin W. “Statistical Estimation of Extreme Roll Responses in Short Crested Irregular Head Seas.” Vol. SNAME Maritime Convention: p. D011S005R006. 2013. DOI [10.5957/SMC-2013-T32](https://doi.org/10.5957/SMC-2013-T32).
- [9] Dinis, Manley St. and Pierson, Willard J. “ON THE MOTIONS OF SHIPS IN CONFUSED SEAS.” 1953.
- [10] Biesel, Francis and Suquet, F. “LES APPAREILS GÉNÉRATEURS DE HOULE EN LABORATOIRE.” *Houille Blanche-revue Internationale De L Eau* (1951): pp. 147–165.
- [11] Svendsen, I. A. *Physical modeling of in coastal engineering*. Balkema (1985): Chap. Physical modeling of water waves, pp. 13–47.

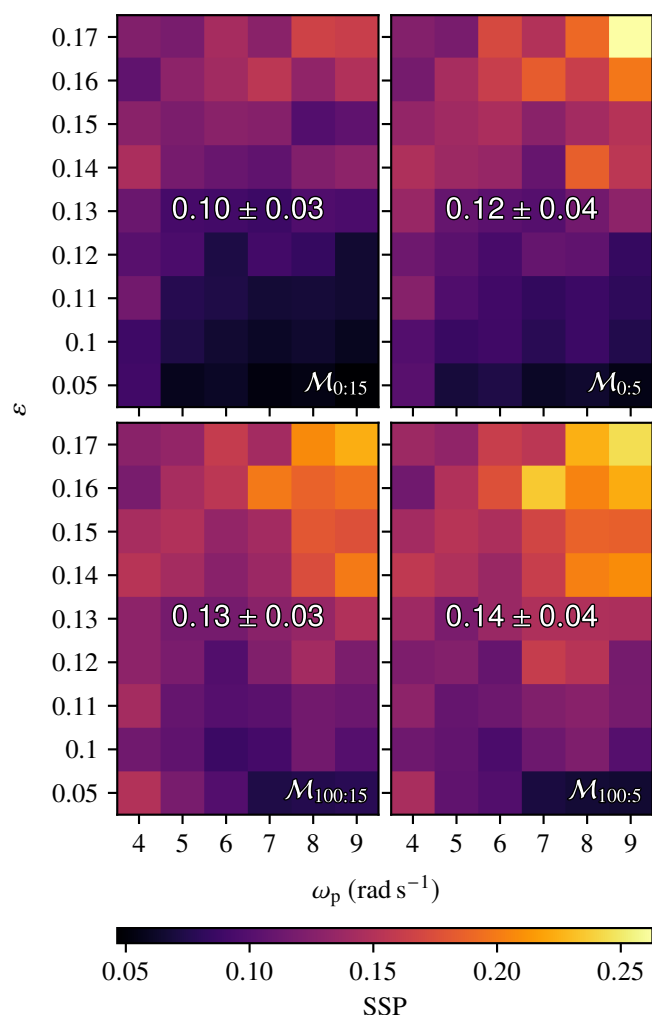


FIGURE 13: Deviation between target wave and measured wave at target location in terms of SSP resolved over the characteristic sea state parameters ω_p and ϵ for $\gamma = 6$. Each panel states the respective model as well as the mean and standard deviation SSP.

[12] Schäffer, H. A. "Second-Order Wavemaker Theory for Irregular Waves." *Ocean Engineering* Vol. 23 No. 1 (1996): pp. 47–88.

[13] Zakharov, V. E. "Stability of periodic waves of finite amplitude on the surface of a deep fluid." *translated in: Journal of Applied Mechanics and Technical Physics*, Vol. 9 (1968): pp. 190–194.

[14] Hasimoto, H. and Ono, H. "Nonlinear modulation of gravity waves." *Journal of Physical Society of Japan*, Vol. 33 (1972): pp. 805–811.

[15] Dysthe, K. B. "Note on a modification to the nonlinear Schrödinger equation for application to deep water waves." *Proceedings of the Royal Society of London, Ser. A* 369 (1979): pp. 105–114.

[16] Chaplin, J. R. "On Frequency-Focusing Unidirectional Waves." *International Journal of Offshore and Polar Engineering*, 6(2) (1996): pp. 131–137.

[17] Shemer, L., Kit, E., Jiao, Haiying and Eitan, O. "Exper-

iments on Nonlinear Wave Groups in Intermediate Water Depth." *Journal of Waterway, Port, Coastal, and Ocean Engineering* Vol. 124 No. 6 (1998): pp. 320–327. DOI [10.1061/\(ASCE\)0733-950X\(1998\)124:6\(320\)](https://doi.org/10.1061/(ASCE)0733-950X(1998)124:6(320)).

[18] Trulsen, K., Kliakhadler, I., Dysthe, K. B. and Verlade, M. G. "On weakly nonlinear modulation of waves in deep water." *Phys. Fluids* 12 (10) (2000): pp. 2432–2437.

[19] Trulsen, K. and Stansberg, C. T. "Spatial Evolution of Water Surface Waves: Numerical Simulation And Experiment of Bichromatic Waves." *The Eleventh International Offshore and Polar Engineering Conference*. 2001. International Society of Offshore and Polar Engineers, Stavanger, Norway.

[20] Shemer, Lev, Kit, Eliezer and Jiao, Haiying. "An experimental and numerical study of the spatial evolution of unidirectional nonlinear water-wave groups." *Physics of Fluids* Vol. 14 No. 10 (2002): pp. 3380–3390. DOI [10.1063/1.1501908](https://doi.org/10.1063/1.1501908). URL <https://doi.org/10.1063/1.1501908>.

[21] Dysthe, K. B., Trulsen, K., Krogstad, H. E. and Socquet-Juglard, H. "Evolution of a narrow-band spectrum of random surface gravity waves." *Journal of Fluid Mechanics* Vol. 478 (2003): pp. 1–10. DOI [10.1017/S0022112002002616](https://doi.org/10.1017/S0022112002002616).

[22] Schmittner, C. E. "Rogue Wave Impact on Marine Structures." *Dissertation*, Technische Universität Berlin (D 83) (2005).

[23] Clauss, G. F. Schmittner, C. E. and Klein, M. "Generation of Rogue Waves with Predefined Steepness." *25th International Conference on Offshore Mechanics and Arctic Engineering*. 2006. Hamburg, Germany. OMAE2006-92272.

[24] Schmittner, C. Kosleck, S. and Hennig, J. "A Phase-Amplitude Iteration Scheme for the Optimization of Deterministic Wave Sequences." *28th International Conference on Ocean, Offshore and Arctic Engineering*. 2009. Honolulu, Hawaii. OMAE2009-80131.

[25] Clauss, G.F. and Klein, M. "The New Year Wave: Spatial Evolution of an Extreme Sea State." *Journal of Offshore Mechanics and Arctic Engineering* Vol. 131 No. 4 (2009).

[26] Clauss, G. Klein, M. Sprenger, F. and Testa, D. "Evaluation of Critical Conditions in Offshore Vessel Operation by Response Based Optimization Procedures." *29th International Conference on Offshore Mechanics and Arctic Engineering*. 2010. ASME, Shanghai, China. OMAE2010-21071.

[27] Schmittner, C. and Hennig, J. "Optimization of short-crested deterministic wave sequences via a phase-amplitude iteration scheme." *31st International Conference on Ocean, Offshore and Arctic Engineering*. 2012. Rio de Janeiro, Brazil. OMAE2012-83150.

[28] Clauss, G.F. Klein, M. Guedes Soares, C. and Fonseca, N. "Response based Identification of Critical Wave Scenarios." *Journal of Offshore Mechanics and Arctic Engineering* Vol. 135 (2013). DOI [10.1115/1.4024269](https://doi.org/10.1115/1.4024269).

[29] Clauss, G.F. and Klein, M. "How safe is 'safe'? Ship dynamics in critical sea states." *Ocean Engineering* Vol. 72 (2013): pp. 87–97. DOI [10.1016/j.oceaneng.2013.06.021](https://doi.org/10.1016/j.oceaneng.2013.06.021).

[30] Klein, Marco, Stender, Merten, Wedler, Mathies, Ehlers, Svenja, Hartmann, Moritz, Desmars, Nicolas, Pick, Marc-

- André, Seifried, Robert and Hoffmann, Norbert. "Application of Machine Learning for the Generation of Tailored Wave Sequences." *Proceedings of the ASME 2022 41st International Conference on Ocean, Offshore and Arctic Engineering*: p. 12. 2022. American Society of Mechanical Engineers Digital Collection. DOI [10.1115/OMAE2022-78601](https://doi.org/10.1115/OMAE2022-78601).
- [31] Klein, Marco, Wedler, Mathies, Pick, Marc-André, Seifried, Robert, Ehlers, Svenja, Stender, Merten and Hoffmann, Norbert. "Data-Driven Generation of Tailored Wave Sequences." *Proceedings of the ASME 2024 43rd International Conference on Ocean, Offshore and Arctic Engineering*: p. 12. 2024. American Society of Mechanical Engineers Digital Collection. DOI [10.1115/OMAE2024-129690](https://doi.org/10.1115/OMAE2024-129690).
- [32] Cybenko, G. "Approximation by superpositions of a sigmoidal function." *Mathematics of Control, Signals and Systems* Vol. 2 No. 4 (1989): pp. 303–314. DOI [10.1007/BF02551274](https://doi.org/10.1007/BF02551274).
- [33] Hornik, Kurt, Stinchcombe, Maxwell and White, Halbert. "Multilayer feedforward networks are universal approximators." *Neural Networks* Vol. 2 No. 5 (1989): pp. 359–366. DOI [10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- [34] Long, Jonathan, Shelhamer, Evan and Darrell, Trevor. "Fully convolutional networks for semantic segmentation." *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015. IEEE.
- [35] Muharini, Wuryansari, Karjanto, N. and Klopman, Gert. "Linear theory for single and double flap wavemakers." *Journal of the Indonesian Mathematical Society (JIMS)* Vol. 12 (2006).
- [36] Hasselmann, Klaus, Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., Meerburg, A., Muller, Peter, Olbers, Dirk, Richter, K., Sell, W. and Walden, H. "Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP)." *Deut. Hydrogr. Z.* Vol. 8 (1973): pp. 1–95.
- [37] Dommermuth, Douglas G. and Yue, Dick K. P. "A high-order spectral method for the study of nonlinear gravity waves." *Journal of Fluid Mechanics* Vol. 184 (1987): pp. 267–288. DOI [10.1017/S002211208700288X](https://doi.org/10.1017/S002211208700288X).
- [38] West, Bruce J., Brueckner, Keith A., Janda, Ralph S., Milder, D. Michael and Milton, Robert L. "A new numerical method for surface hydrodynamics." *Journal of Geophysical Research* Vol. 92 No. C11 (1987): p. 11803. DOI [10.1029/JC092iC11p11803](https://doi.org/10.1029/JC092iC11p11803).
- [39] Zakharov, V. E. "Stability of periodic waves of finite amplitude on the surface of a deep fluid." *Journal of Applied Mechanics and Technical Physics* Vol. 9 No. 2 (1968): pp. 190–194. DOI [10.1007/BF00913182](https://doi.org/10.1007/BF00913182).
- [40] Bonnefoy, Félicien, Ducrozet, Guillaume, Le Touzé, David and Ferrant, Pierre. *TIME DOMAIN SIMULATION OF NONLINEAR WATER WAVES USING SPECTRAL METHODS*. WORLD SCIENTIFIC (2010): pp. 129–164. DOI [10.1142/9789812836502_0004](https://doi.org/10.1142/9789812836502_0004).
- [41] Ducrozet, Guillaume, Bonnefoy, Félicien, Le Touzé, David and Ferrant, Pierre. "A modified High-Order Spectral method for wavemaker modeling in a numerical wave tank." *European Journal of Mechanics - B/Fluids* Vol. 34 (2012): pp. 19–34. DOI [10.1016/j.euromechflu.2012.01.017](https://doi.org/10.1016/j.euromechflu.2012.01.017).
- [42] Wedler, Mathies, Stender, Merten, Klein, Marco, Ehlers, Svenja and Hoffmann, Norbert. "Surface similarity parameter: A new machine learning loss metric for oscillatory spatio-temporal data." *Neural Networks* Vol. 156 (2022): pp. 123–134. DOI [10.1016/j.neunet.2022.09.023](https://doi.org/10.1016/j.neunet.2022.09.023).
- [43] Wedler, Mathies, Stender, Merten, Klein, Marco and Hoffmann, Norbert. "Machine learning simulation of one-dimensional deterministic water wave propagation." *Ocean Engineering* Vol. 284 (2023): pp. 115–222. DOI [10.1016/j.oceaneng.2023.115222](https://doi.org/10.1016/j.oceaneng.2023.115222).
- [44] Ehlers, Svenja, Klein, Marco, Heinlein, Alexander, Wedler, Mathies, Desmars, Nicolas, Hoffmann, Norbert and Stender, Merten. "Machine learning for phase-resolved reconstruction of nonlinear ocean wave surface elevations from sparse remote sensing data." *Ocean Engineering* Vol. 288 (2023): pp. 116–059. DOI [10.1016/j.oceaneng.2023.116059](https://doi.org/10.1016/j.oceaneng.2023.116059).
- [45] Klein, M., Dudek, M., Clauss, G.F., Ehlers, S., Behrendt, J., Hoffmann, N. and Onorato, M. "On the deterministic prediction of water waves." *Fluids* Vol. 5 No. 1 (2020): p. 19. DOI [10.3390/fluids5010009](https://doi.org/10.3390/fluids5010009).
- [46] Desmars, Nicolas, Hartmann, Moritz, Behrendt, Jasper, Klein, Marco and Hoffmann, Norbert. "Reconstruction of Ocean Surfaces From Randomly Distributed Measurements Using a Grid-Based Method." *Volume 6: Ocean Engineering*: p. V006T06A059. 2021. American Society of Mechanical Engineers, Virtual, Online. DOI [10.1115/OMAE2021-62409](https://doi.org/10.1115/OMAE2021-62409).
- [47] Lünser, Helene, Hartmann, Moritz, Desmars, Nicolas, Behrendt, Jasper, Hoffmann, Norbert and Klein, Marco. "The Influence of Characteristic Sea State Parameters on the Accuracy of Irregular Wave Field Simulations of Different Complexity." *Fluids* Vol. 7 No. 7 (2022): p. 243. DOI [10.3390/fluids7070243](https://doi.org/10.3390/fluids7070243). Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
- [48] Bouws, E., Günther, H., Rosenthal, W. and Vincent, C. L. "Similarity of the wind wave spectrum in finite depth water: 1. Spectral form." *Journal of Geophysical Research: Oceans* Vol. 90 No. C1 (1985): pp. 975–986. DOI [10.1029/JC090iC01p00975](https://doi.org/10.1029/JC090iC01p00975).
- [49] Goodfellow, Ian, Bengio, Yoshua and Courville, Aaron. *Deep Learning*. MIT Press (2016). Accessed 2023-11-22, URL <http://www.deeplearningbook.org>.
- [50] LeCun, Yann, Bengio, Yoshua and Hinton, Geoffrey. "Deep learning." *Nature* Vol. 521 No. 7553 (2015): pp. 436–444. DOI [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [51] Dumoulin, Vincent and Visin, Francesco. "A guide to convolution arithmetic for deep learning." (2016). DOI [10.48550/ARXIV.1603.07285](https://doi.org/10.48550/ARXIV.1603.07285).
- [52] Clevert, Djork-Arné, Unterthiner, Thomas and Hochreiter, Sepp. "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)." (2015). DOI [10.48550/ARXIV.1511.07289](https://doi.org/10.48550/ARXIV.1511.07289).

- [53] Perlin, Marc and Bustamante, Miguel D. “A robust quantitative comparison criterion of two signals based on the Sobolev norm of their difference.” *Journal of Engineering Mathematics* Vol. 101 No. 1 (2016): pp. 115–124. DOI [10.1007/s10665-016-9849-7](https://doi.org/10.1007/s10665-016-9849-7).
- [54] Wedler, Mathies, Stender, Merten, Klein, Marco, Ehlers, Svenja and Hoffmann, Norbert. “Surface Similarity Parameter machine learning loss metric.” (2025). DOI [10.5281/zenodo.16033049](https://doi.org/10.5281/zenodo.16033049).
- [55] Evans, Lawrence. *Partial Differential Equations*. American Mathematical Society (2010). DOI [10.1090/gsm/019](https://doi.org/10.1090/gsm/019).
- [56] Kingma, Diederik P. and Ba, Jimmy. “Adam: A Method for Stochastic Optimization.” (2017). Number: arXiv:1412.6980 arXiv:1412.6980 [cs].
- [57] Sturges, Herbert A. “The Choice of a Class Interval.” *Journal of the American Statistical Association* Vol. 21 No. 153 (1926): pp. 65–66. DOI [10.1080/01621459.1926.10502161](https://doi.org/10.1080/01621459.1926.10502161).
- [58] Freedman, David and Diaconis, Persi. “On the histogram as a density estimator:L2 theory.” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* Vol. 57 No. 4 (1981): pp. 453–476. DOI [10.1007/BF01025868](https://doi.org/10.1007/BF01025868).
- [59] Harris, F.J. “On the use of windows for harmonic analysis with the discrete Fourier transform.” *Proceedings of the IEEE* Vol. 66 No. 1 (1978): pp. 51–83. DOI [10.1109/proc.1978.10837](https://doi.org/10.1109/proc.1978.10837).
- [60] Hayer, Sverre and Andersen, Odd Jan. “Freak Waves: Rare Realizations of a Typical Population Or Typical Realizations of a Rare Population?” *Proceedings of the Tenth International Offshore and Polar Engineering Conference*, Vol. The Tenth International Offshore and Polar Engineering Conference: pp. ISOPE–I–00–233. 2000.
- [61] Wolfram, Julian, Linfoot, B.T. and Stansell, Paul. “Long- and short-term extreme wave statistics in the North Sea: 1994-1998.” .
- [62] Klein, Marco, Clauss, Günther F. and Hoffmann, Norbert. “Introducing envelope soliton solutions for wave-structure investigations.” *Ocean Engineering* Vol. 234 (2021): p. 109271. DOI [10.1016/j.oceaneng.2021.109271](https://doi.org/10.1016/j.oceaneng.2021.109271).

Appendix A. TARGET- VS. MEASURED EXTREME WAVES

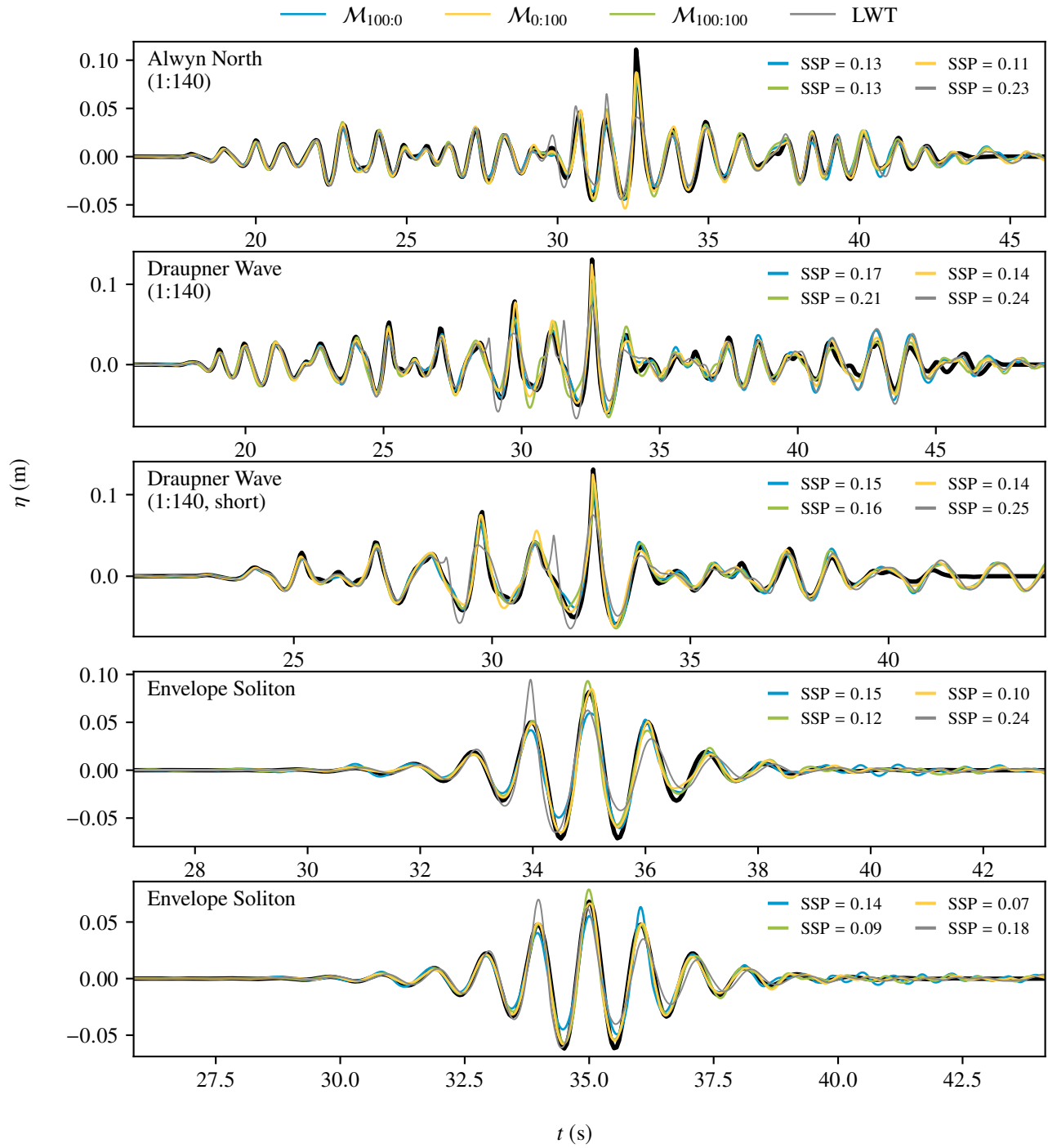


FIGURE 14: Time series of target wave (black) and generated waves for models studied in Sect. 4.3. To provide a global perspective alongside the results in Tab. 2, each panel states the SSP for each model evaluated over the whole time series.

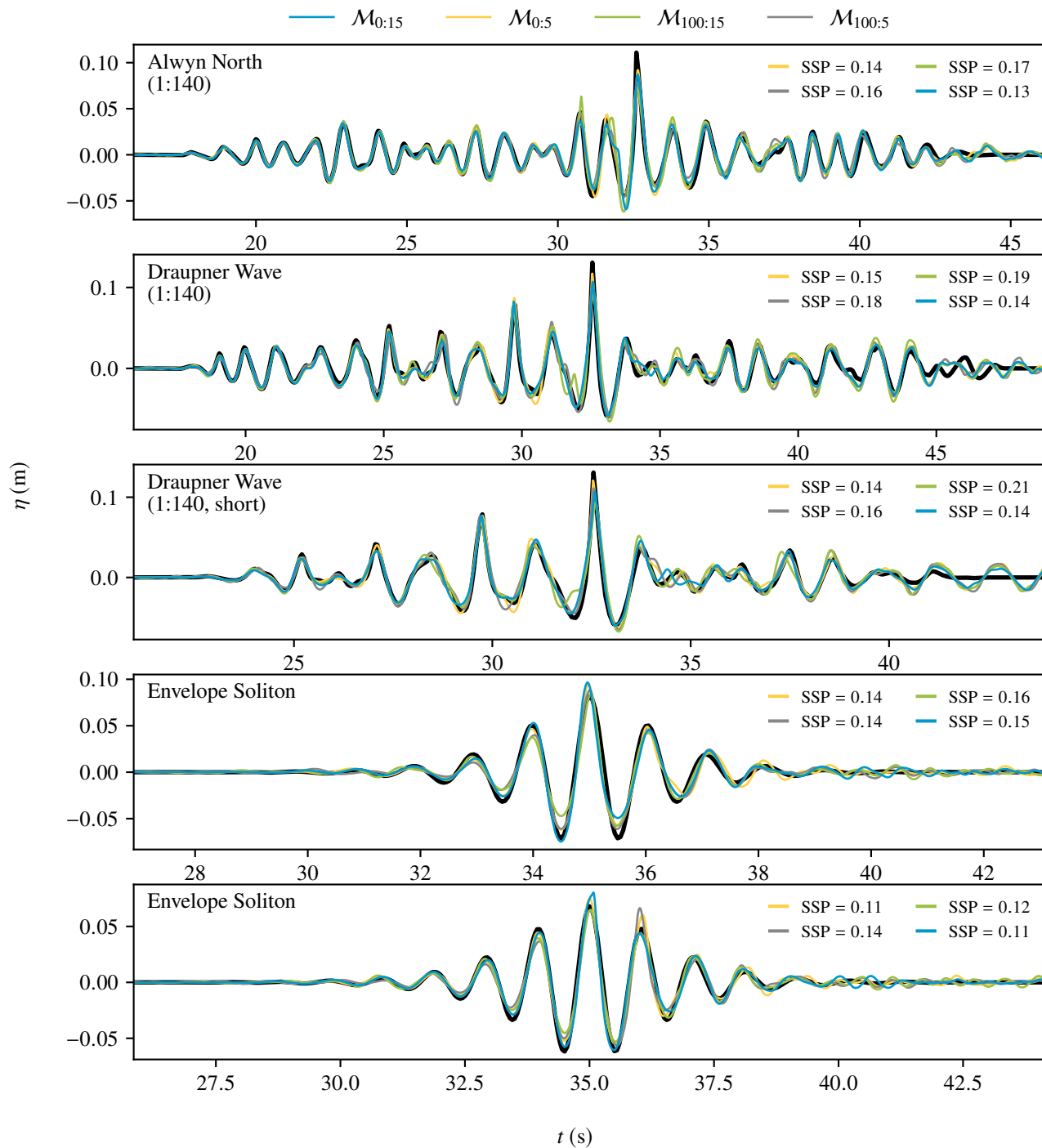


FIGURE 15: Time series of target wave (black) and generated waves for models studied in Sect. 4.4. To provide a global perspective alongside the results in Tab. 3, each panel states the SSP for each model evaluated over the whole time series.