Highlights

Data-driven, high resolution ocean wave forecasting and extreme wave predictions

Thomas Breunung,Balakumar Balachandran

- Novel data-driven methodology for ocean wave forecasting
- Combination of physical modelling with data-driven techniques
- Extensive validation of the proposed methodology with experimental data and field measurements
- Accurate forecasts in real-time
- Prediction of rogue waves up to one minute in advance

Data-driven, high resolution ocean wave forecasting and extreme wave predictions⋆

Thomas Breunung[a,*], Balakumar Balachandran[b]

[a]Department of Mechanical Engineering, University of Maryland, College Park, Maryland 20742, USA

ARTICLE INFO

Keywords:
ocean waves
extreme waves
data-driven method
machine learning
forecasting
buoy data

ABSTRACT

Reliable ocean wave forecasts are critical for reducing the risk posed by extreme waves for ships and offshore infrastructure. Despite insights gained from theoretical investigations and constantly improving operational wave forecasting systems, the emergence of extreme waves remain unpredictable. In this article, the authors propose a data-driven modeling approach to generate high resolution wave forecasts for capturing individual waves, including rogue waves. To this end, a decomposition of the sea surface into a set of rapid oscillations and slowly varying amplitudes is utilized. The slow amplitude variations are subsequently forecasted by using universal, data-driven methods. In this approach, the extrapolation range of the data-driven techniques is extended by the slowness of the amplitude variations. The method's capabilities are demonstrated by using available measurements from an experimental wave tank and field data from ocean buoys.

1. Introduction

Ocean wave forecasting and rouge wave[1] predictions in particular are of paramount importance for the safety of maritime operations and infrastructure in the ocean. Such extreme waves continue to endanger ships, damage offshore structures, and cause serious injuries to sailors (Didenkulova et al., 2006; Dysthe et al., 2008; Kharif and Pelinovsky, 2003).

Modern, operational ocean wave forecasting systems rely on ocean wave spectra (Komen et al., 1996). Two well-known examples are the WAM model (The Wamdi Group, 1988) and Wavewatch III (Tolman, 1991). These models can yield accurate predictions over multiple days and can provide a globally consistent picture (Janssen et al., 2005). The core of such models is an action balance equation governing the time evolution of the wave spectra. The action balance equation can be parameterized in various forms to include numerous physical effects such as wind input, dissipation, and nonlinear wave interactions (Cavaleri et al., 2007). Especially, the nonlinear wave interactions have been identified as a crucial driving mechanism for wave growth (Hasselmann et al., 1973). A parameterized form of the resonant four-wave interactions (Hasselmann, 1962) is commonly used to model wave interactions in spectral wave models. However, a scaling analysis indicates that these terms do not capture the Benjamin-Feir instability (Janssen, 2003), a prominent candidate for explaining the formation of extreme waves. More importantly, the action balance equation is parameterized to govern waves spectra with a spatial resolution of several kilometers and time steps ranging hours. This coarse model resolution fundamentally limits the use of these spectral models to resolve individual waves such as rogue waves, which are inherently localized in time and space. Finally, with these spectral wave models, one can not resolve the phase of ocean waves, which are necessary for the prediction of individual waves (Alam, 2014).

A more detailed picture on individual waves can be obtained from tools developed for computational fluid dynamics. For example, higher order spectral methods (Toffoli et al., 2010; Bitner-Gregersen et al., 2014), boundary element methods (Fochesato et al., 2007), smoothed particle hydrodynamics (Chakraborty and Balachandran, 2021), or harmonic polynomial cell methods (Zhao et al., 2020) have been used to study and model extreme waves. Therein, researchers consider idealized scenarios, such as modeling an experimental wave tank or a periodic domain. These idealizations impede straightforward extensions of such computational results to the real ocean. Despite the development of fast numerical solvers and efficient utilization of computational resources, resolving a sizable part of an ocean under realistic condition remains currently infeasible with tools from computational fluid dynamics. Furthermore, the computational times even for small, idealized domains remain significant. Hence, real time forecasts cannot be obtained.

Another frequent starting point to study ocean waves is the classical analysis of Stokes (Stokes, 1880). A perturbation expansion of Stokes' solution can lead to the nonlinear Schrödinger equation (Kharif and Pelinovsky, 2003; Farazmand and Sapsis, 2019). For certain boundary conditions (e.g. periodic domain), this equation can be solved in closed form via the inverse scattering transformation (Osborne, 2002). Besides classical soliton solutions (Ablowitz and Clarkson, 1991), many other solutions featuring extreme waves have been found (Chabalko et al., 2014; Moitra

⋆This is an extension of the preliminary conference report Breunung and Balachandran (2022).

*Corresponding author

thomasbr@umd.edu (T. Breunung); balab@umd.edu (B. Balachandran)
ORCID(s):

[1]Rogue or freak waves are individual waves that are significantly higher than the surrounding waves. Mathematically, rogue waves have been defined as waves with a crest height η_c exceeding the significant wave height H_s by a factor of 1.25 ($\eta_c/H_s > 1.25$) (Haver, 2000; Dysthe et al., 2008). Further, H_s is defined as four times the standard deviation of the surface elevation. Alternative criteria relate the wave height H (measured from trough to crest) to the significant wave height (Haver, 2000; Dysthe et al., 2008).

et al., 2016). Moreover, further extensions of the nonlinear Schrödinger to include higher order terms have been proposed (Dysthe, 1979). Recently, a stochastic phase model has been proposed (Birkholz et al., 2016; Wang and Balachandran, 2018), which can accurately reproduce a-priori known extreme waves. These analysis can lead to important insights into rogue wave formations. For example the Benjamin-Feir instability (Benjamin and Feir, 1967; Yuen and Lake, 1980), linear focusing, and wave-current interaction can be used to explain the occurrence of an extreme wave (Kharif and Pelinovsky, 2003). Within these models, however, assumption such as unidirectionality, stationarity, or a narrow band process are made, all of which do not hold in reality. Thus, the applicability of such models to realistic scenarios remains limited. Indeed, two recent extensive data analysis on buoy data (Cattrell et al., 2018; Häfner et al., 2021) reveal that classical rogue wave indicators such as the Benjamin-Feir index (Mori et al., 2011), which has been designed to indicate the aforementioned Benjamin-Feir instability, poorly correlate with rogue wave occurrences.

A plethora of data-driven methods have been proposed and utilized to analyze general time series. Their application to fluid mechanics has been recently summarized (Brunton et al., 2020). Methods such as singular spectrum analysis (SSA) (Vautard and Ghil, 1989) or linear embeddings via a dynamics mode decomposition (DMD) (Schmid, 2010) are based on delay embedding or the trajectory matrix (Broomhead and King, 1986). A theoretical underpinning of such approaches can be the Koopman operator (Rowley et al., 2009). Similarly, recurrent neural networks (Goodfellow et al., 2016), temporal convolutional neural networks (Lea et al., 2016) and, recently, transformers (Vaswani et al., 2017) are appealing black box models for general time series forecasting. In essence the promise of such approaches is to approximate any arbitrary input-output relationship given that enough data for parameter fitting is available. While the aforementioned approaches are fundamentally interpolative, forecasting requires extrapolation. Hence, the applicability of such methods for wave forecasting remains unclear. While a recurrent neural network has recently been utilized to forecast rogue waves (Kagemoto, 2020, 2022), within this approach only the crest heights and trough depth are forecasted but no information when a wave occurs is provided. Moreover, the forecast range is not specified and it unclear whether this approach can yield forecasts in real time. Moreover, due to their black-box character only limited insights on the physical driving mechanisms for extreme wave formation can be obtained.

In summary, available ocean wave models are designed for time and length scales larger than individual extreme waves, whereas tools from computational fluid dynamics remain too expensive to generate meaningful forecasts. Simplified equations on the other hand seem not to capture the physics leading to rogue waves in realistic scenarios. Moreover, black-box data-driven methods suffer from limited interpretabilty and poor extrapolation capabilities.

To overcome some of the aforementioned shortcomings and enable wave forecasting with a high temporal resolution[2] on intermediate time scales, the authors combine the strengths of established ocean waves models with recent data-driven methods. To this end a decomposition of the ocean surface elevation into rapid oscillations with slowly varying amplitudes is utilized. The time variations of the amplitudes are subsequently forecasted by fitting linear auto regressive models via a dynamic mode decomposition (DMD) (Schmid, 2010) and recurrent neural networks with long short-term memory cells (LSTM) (Hochreiter and Schmidhuber, 1997). The decomposition into slowly varying amplitudes, in turn, extends the extrapolation range of the data-driven methods utilized to forecast the time varying amplitudes. The approach's capabilities to accurately forecast ocean waves in realistic scenarios are demonstrated by applying it to measurements from an experimental wave tank (Eeltink, 2022) and available buoy data (CDIP Buoy data, 2022).

2. Methods

Only a few measuring systems can be used to accurately capture individual waves, in particular, rogue waves (Dysthe et al., 2008) and the methods within this article are developed and tested keeping the current limitations and challenges of measuring ocean surface elevations in mind. Arguably the most accessible and best controlled recordings of sea surface elevation are single point observations stemming from surface buoys. Therefore, the following developments focus on this type of measurements. However, the proposed methodology can, at least in principle, be extended to incorporate spatially distributed observations.

The real time ocean wave forecasting methodology proposed consists of three main steps. These steps are sketched in Figure 1. At first, a wave model is fitted to the recorded data. Within this model, the ocean surface elevation is *decomposed* into rapid oscillations with a slowly varying spectrum. The wave model is described and motivated in Section 2.1 and the fitting procedure is detailed in Section 2.2.

The slowly varying amplitudes are then *forecasted* with data-driven methods. Therein, the extrapolation capabilities of the black-box, data driven methods are extended by the gradual nature of the amplitude variation. These data-driven forecasting methods are described in Section 2.3.

Finally, the extrapolated slowly varying amplitudes are *synthesized* to yield real time forecasts of the sea surface elevation. Therein, the computational time spent to generate forecasts diminishes the effective forecasting horizon. To

[2]To resolve extreme waves, the temporal resolution needs to be fine enough to capture individual waves. Thus, for ocean waves, which are primarily observed within the frequency band between 0.02 Hz and 0.6 Hz (Hasselmann et al., 1973), the sampling rate should be higher than 1.2 Hz. Indeed, the buoy data (CDIP Buoy data, 2022) utilized within this manuscript stems from Datawell directional wave riders (Datawell B.V., 2020), which generate recordings of the ocean surface elevation at a sampling rate of 1.28 Hz.

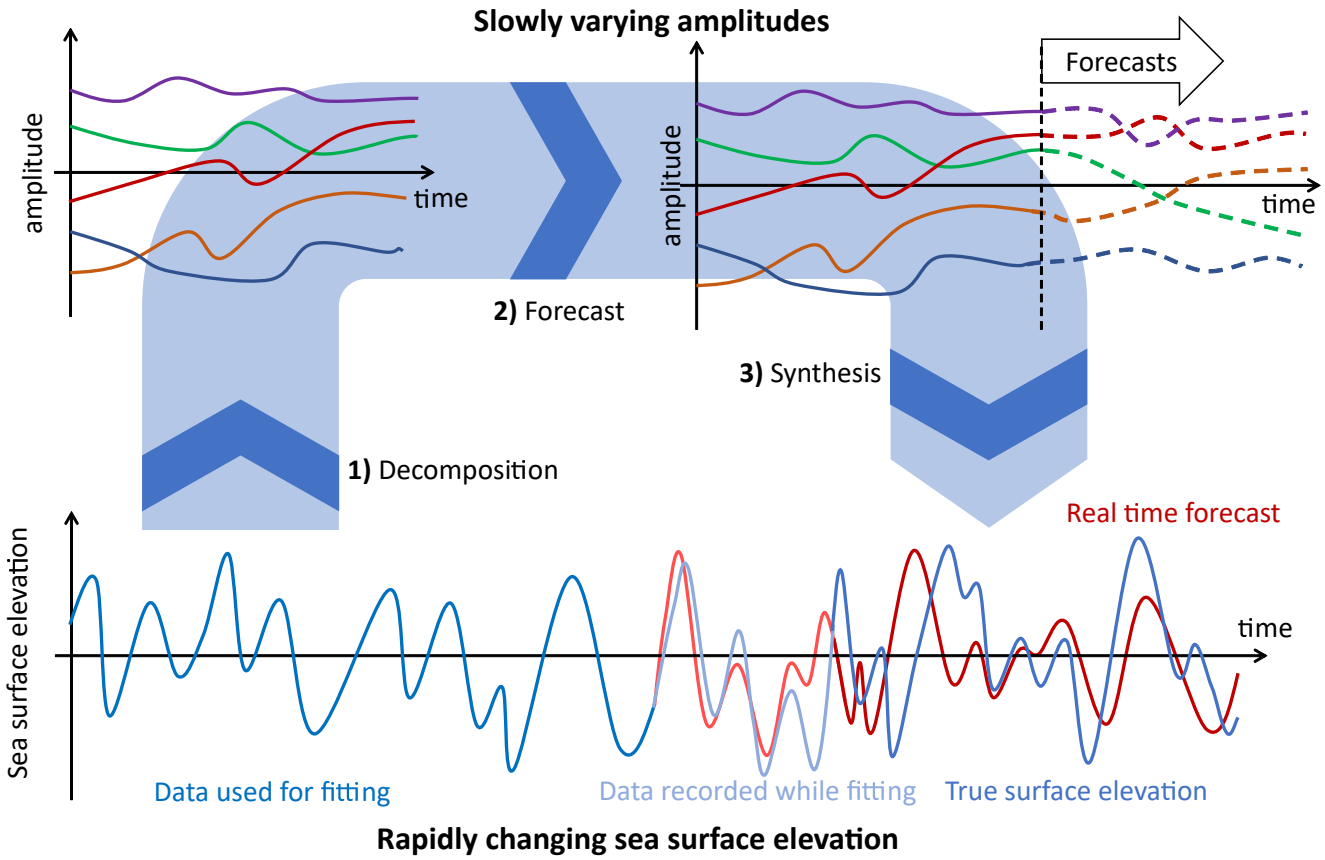

Figure 1: Sketch illustrating the three main steps involved for real time ocean wave forecasting. In the bottom the measured sea surface elevation is shown in blue, while the forecast is shown in red.

minimize this time, a multi-step procedure enabling real time forecasts of ocean waves is described in Section 2.4.

2.1. Wave Model

Harmonic functions are natural candidates to model the everlasting recurrence and shapes of ocean waves. Indeed, Stokes' classic perturbation analysis (Stokes, 1880) of the Bernoulli equation for irrotational flows with a free surface leads to waves expressed in terms of harmonic functions. In the same vein, wave spectra (The Wamdi Group, 1988; Tolman, 1991) prescribing the amplitudes and frequencies of ocean waves are central to the state-of-the-art operational wave forecasting systems. In such models, ocean waves are separated into rapid oscillations (frequencies) and slowly varying spectrum (amplitudes) (Komen et al., 1996). This separation of time scales is motivated by theoretical considerations (Hasselmann, 1962; Komen et al., 1996) and can be confirmed by field observations (Hasselmann et al., 1973) indicating that the spectrum varies slowly in time and space. Similarly, analyses leading to the nonlinear Schrödinger equation and associated extensions (Kharif and Pelinovsky, 2003; Farazmand and Sapsis, 2019; Dysthe, 1979) are based on slowly perturbed harmonic waves. For a single point, fixed observer all of the aforementioned modeling approaches and field observations can be summarized in the form of the following wave model

$$\eta_f(t) = \sum_{n=1}^{N} \left(u_n(\varepsilon t)\cos(\omega_n t) + v_n(\varepsilon t)\sin(\omega_n t) \right) + u_0(\varepsilon t), \qquad 0 < \varepsilon \ll 1, \tag{1}$$

where η_f denotes the sea surface elevation, ω_n are frequencies and the integer N is the number of frequencies. The amplitudes u_n and v_n vary in time and u_0 is used to model a slowly varying drift. The small parameter ε is used to describe the slow time variations of the amplitudes. A detailed comparison of common wave models with the wave model (1) is included in Appendix A.

It is noted, that with the model (1), there is a nonlinear dependence on the frequencies ω_n and a linear dependence on the amplitudes. Instead of two amplitudes for each frequency model (1) can be rewritten to have one amplitude and a phase ϕ_n for each frequency. However, such a reformulated model would be nonlinear in the phase variables ϕ_n.

To employ the model (1) for forecasting, its parameters, more specifically the ampliltudes $u_n(\varepsilon t)$ and $v_n(\varepsilon t)$ as well as the frequencies ω_n need to obtained. Within this article, these parameters are obtained by fitting the surface elevation $\eta_f(t)$ from the model (1) to measured surface elevations $\eta(t)$. The fitting procedure employed will be detailed in the next section.

2.2. Fitting the wave model (1)

To generate forecasts, the modelled surface elevation η_f of the wave model (1) is needs to be fitted to measured surface elevations $\eta(t)$. Technically, the function (1) can be used to fit any arbitrary continuous function over a finite time interval with arbitrary precision. Since the aim of this work is to forecast ocean waves, more specifically their crest heights and trough depths, the intermediate values between the troughs and crests are of secondary importance. Thus, the intermediate values are discarded and only the trough depths and crest heights are kept for fitting in the following. This procedure reduces the number of data points within a given time interval and thus a more accurate fit can be achieved with less parameters. The obtained time series is non-uniform in time and will be denoted by $\eta(t_j)$, where t_j denotes the j-th time instance.

The model (1) is fitted by minimizing the following cost function[3]

$$\begin{aligned}\Xi(\omega_n, u_n(t), v_n(t)) := & \underbrace{\|\eta(t) - \eta_f(t)\|^2}_{\text{accurate fit}} \\ & + \underbrace{\frac{1}{\varepsilon}\left[\|\dot{u}_0(t)\|^2 + \left[\sum_{n=1}^{N}\left(\|\dot{u}_n(t)\|^2 + \|\dot{v}_n(t)\|^2\|\right)\right]\right]}_{\text{slowly varying amplitudes}} \\ & + \underbrace{\alpha\left\|\sum_{n=1}^{N} -u_n(t)\omega_n \sin(\omega_n t) + v_n(t)\omega_n \cos(\omega_n t)\right\|^2}_{\text{extreme value}},\end{aligned} \tag{2}$$

where the norm $\|\ \|^2$ denotes the sum of squares over all time instances t_j, $\|\eta(t)\|^2 := \sum \eta(t_j)^2$.

Each summand in equation (2) is used to ensure that the model (1) accurately represents the ocean surface as indicated. The first term ensures that the fit η_f is close to the measured surface elevation η. The second term penalizes time variations in the amplitudes, and enforces a slowly varying spectrum. It is scaled by the inverse of the parameter ε which, in turn, controls the slowness of the amplitude variations. As such ε has to be set to a small value to be consistent with the theory (Hasselmann, 1962; Komen et al., 1996) and observations (Hasselmann et al., 1973). The last term is the time derivative of the fitting function η_f, in which the small time variations of the amplitudes have been neglected. Enforcing a small time derivative of η_f at the time instances t_j ensures that the values $\eta_f(t_j)$ are indeed extreme values, namely, crests and troughs. The scaling parameters α and ε are treated as user defined input parameters.

[3] In the cost function (2), the time dependence of the amplitudes on the slow time scale εt it not explicitly included. The notation in equation (1) follows the mathematical convention to ensure a small time derivative of the amplitudes as follows $\mathrm{d}/\mathrm{d}t\ u_n(\varepsilon t) = \varepsilon \partial_t u_n(\varepsilon t)$. In the constructed cost function (2), this slow time variation is ensured by scaling the second summand by $1/\varepsilon$ without requiring the amplitudes to depend on a artificially introduced slow time scale. Hence, the dependence on the amplitudes on the slow time scale is not explicitly shown in the following.

Often the amplitudes of ocean wave models are obtained by the means of Fourier transformation, in particular, the algorithmic implementation of it, namely, the fast Fourier transformation (FFT). While the Fourier transformation is defined for signals over infinite long time intervals, measurements on the other hand are only obtainable for finite times. Thus, transforming a measurement into the Fourier domain always requires a continuation of the finite time measurements outside of the measurement window. Commonly, a periodic continuation is assumed. Such an assumption is avoided by obtaining the frequencies via the minimization

$$\arg \min_{\omega_n, u_n(t), v_n(t)} \Xi(\omega_n, u_n(t), v_n(t)), \tag{3}$$

which is nonlinear in the frequencies ω_n, since $\eta_f(t)$ depends nonlinearly on the frequencies ω_n (cf. equation (1)). To reduce computational burden and avoid spurious solutions, the minimization (3) is solved in two steps.

In the first step, the amplitudes are kept constant ($u_n(t) = u_n$ and $v_n(t) = v_n$) which yields

$$\arg \min_{\omega_n, u_n, v_n} \Xi(\omega_n, u_n, v_n). \tag{4}$$

The minimization (4) can be numerically solved to obtain an approximate, local minimum. To this end, MATLAB's nonlinear least squares solver `lsqnonlin` is used. Therein, it is observed that the obtained minimum strongly depends on the initial values for frequencies ω_n provided to the nonlinear optimization routine. To avoid non-optimal solutions with large residuals, a Monte-Carlo strategy is employed. The optimization (4) is repeated with multiple initial guesses for the frequencies. The initial values are selected by sampling the uniform distibution within the interval between 0.02 Hz and 0.6 Hz, which is the typical frequency band for ocean waves (Hasselmann et al., 1973). Subsequently, the parameters yielding the lowest residual are selected.

In the second step, the frequencies are fixed to the values obtained in the first step and the amplitudes are allowed to vary in time, yielding

$$\arg \min_{u_n(t), v_n(t)} \Xi(\omega_n, u_n(t), v_n(t)). \tag{5}$$

Since the wave model (1) is linear in the amplitudes, the first and third summand in the cost function (2) are quadratic. Approximating the time derivatives of the amplitudes in the cost function (2) with finite differences yields a quadratic cost function. Thus, the minimization (5) can be solved in closed form. Obtaining this solution is computational significantly cheaper than the solution strategy employed for the minimization (4).

The outcomes of the described fitting procedure are N frequencies ω_n and $2N + 1$ time series of slowly varying amplitudes u_n and v_n. To obtain forecasts, the model (1) needs to be evaluated outside the fitting interval. The frequencies ω_n are constant by assumption, but the slowly varying amplitudes $u_n(t)$ and $v_n(t)$ need to be extrapolated. The methods employed for this forecasting task are detailed in the next section.

2.3. Forecasting

Forecasting requires to evaluate the wave model (1) outside the fitting interval. To this end, the slowly varying amplitudes $u_n(t)$ and $v_n(t)$ need to be extrapolated. Due to the lack of established models prescribing the time variations of the slowly varying amplitudes, data-driven, black-box methods are utilized. These methods can be most concisely summarized as the mapping

$$\mathbf{z}(t_{j+1}) = \mathbf{f}(\mathbf{z}(t_j), \mathbf{p}), \tag{6}$$

which maps the current, abstract state $\mathbf{z}(t_j)$ forward in time. The vector $\mathbf{p}$ is a fixed parameter vector. The state vector $\mathbf{z}(t_j)$ will generally consist of the following:

i. the amplitudes u_n and v_n at the current time instance t_j ($u_n(t_j)$ and $v_n(t_j)$)
ii. amplitudes at previous time instances $u_n(t_{j-1})$, $u_n(t_{j-2})$, ..., $u_n(t_{j-M})$ and $v_n(t_{j-1})$, $v_n(t_{j-2})$, ..., $v_n(t_{j-M})$
iii. a set of hidden states.

Equation (6) is a dynamical system in the discrete time domain. If a function $\mathbf{f}$ and its parameters $\mathbf{p}$ are selected, then the forecasts are generated by applying the dynamical systems recursively to data. More specifically, given an initial state $\mathbf{z}(t_j)$ consisting of slowly varying amplitudes u_n and v_n and, potentially, hidden states, the dynamical system (6) is used to generate the forecast $\mathbf{z}(t_{j+1})$. From this state the new amplitudes $u_n(t_{j+1})$ and $v_n(t_{j+1})$ can be deduced. Applying $\mathbf{f}$ to the new state $\mathbf{z}(t_{j+1})$ yields the amplitudes $u_n(t_{j+2})$ and $v_n(t_{j+2})$. This procedure is repeated until a specified forecasting horizon is reached.

To use function (6) for forecasting the functional relationship $\mathbf{f}$ and its associated parameters $\mathbf{p}$ need to be determined. For the functional form $\mathbf{f}$ (*model selection*) linear auto regressive models and recurrent neural networks with LSTM cells (Hochreiter and Schmidhuber, 1997) are used. It is anticipated that the slow variations of the amplitudes extend the extrapolation range of these fundamentally interpolative approaches. Once a model is selected, the parameters $\mathbf{p}$ need to be obtained (*parameter fitting/tuning*). The two distinct model choices and the therein employed tuning methods are detailed in the next two sections.

2.3.1. Linear models

Within linear autoregressive models of the form

$$u_n(t_l) = \sum_{m=1}^{M} \beta_m u_n(t_{l-m}), \tag{7}$$

it is assumed that the current value $u_n(t_l)$ is given by a linear combination of the previous M values. The integer M denotes the order of the autoregressive model. The coefficients β_m are scalars and these coefficients are usually obtained by fitting the model (7) to data. For this purpose, the non-uniform time series of the amplitudes (cf. Section 2.1) are smoothly interpolated (spline interpolation) to yield uniform time series $u_n(t)$ and $v_n(t)$.

The coefficients β_m are obtained with two different methods. First, they are obtained by constructing the linear system

$$\underbrace{\begin{bmatrix} u_n(t_M) \\ u_n(t_{M+1}) \\ u_n(t_{M+2}) \\ \vdots \end{bmatrix}}_{\mathbf{u}_{u_n}} = \underbrace{\begin{bmatrix} u_n(t_{M-1}) & u_n(t_{M-2}) & \dots & u_n(t_0) \\ u_n(t_M) & u_n(t_{M-1}) & \dots & u_n(t_1) \\ u_n(t_{M+1}) & u_n(t_M) & \dots & u_n(t_2) \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}}_{\mathbf{A}_{u_n}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}}_{\boldsymbol{\beta}_{u_n}}, \tag{8}$$

where the amplitudes $u_n(t_j)$ are known and the coefficients β_m are unknown. The coefficients β_m are then obtained by minimizing the error $\mathbf{u}_{u_n} - \mathbf{A}_{u_n}\boldsymbol{\beta}_{u_n}$ in the least squares sense. The solution of this minimization is available in closed form and can be computed with ease. Owing to the least squares minimization, the obtained model is referred to as *LSQ* model.

The matrix $\mathbf{A}_{u_n}$ is also known as trajectory matrix (Broomhead and King, 1986) and plays an essential part in the singular spectrum analysis (Vautard and Ghil, 1989) as well as in the dynamic mode decomposition (DMD) (Schmid, 2010). This decomposition is based on the matrix equation

$$\underbrace{\begin{bmatrix} u_n(t_M) & u_n(t_{M-1}) & \dots & u_n(t_1) \\ u_n(t_{M+1}) & u_n(t_M) & \dots & u_n(t_2) \\ u_n(t_{M+2}) & u_n(t_{M+1}) & \dots & u_n(t_3) \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}}_{\mathbf{A}^2_{u_n}} = \underbrace{\begin{bmatrix} u_n(t_{M-1}) & u_n(t_{M-2}) & \dots & u_n(t_0) \\ u_n(t_M) & u_n(t_{M-1}) & \dots & u_n(t_1) \\ u_n(t_{M+1}) & u_n(t_M) & \dots & u_n(t_2) \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}}_{\mathbf{A}^1_{u_n}} \underbrace{\begin{bmatrix} \beta_1 & 1 & 0 & \dots \\ \beta_2 & 0 & 1 & \dots \\ \beta_3 & 0 & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}}_{\boldsymbol{\beta}^*_{u_n}}, \tag{9}$$

wherein the matrix $\boldsymbol{\beta}^*_{u_n}$ advances the trajectory matrix $\mathbf{A}^1_{u_n}$ by one time step. Introducing the singular value decomposition of the generally non rectangular matrix $\mathbf{A}^1_{u_n} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T$, where $\boldsymbol{\Lambda}$ is a diagonal, rectangular matrix and $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and right multiplying equation (9) with $\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T$ yields

$$\boldsymbol{\beta}^*_{u_n} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{A}^2_{u_n}, \tag{10}$$

whereby the coefficients β_m can be obtained. The resulting model will be denoted as *DMD* model. DMD and its variants have been successfully applied to numerous data sets in fluid dynamics (Schmid, 2010, 2022).

2.3.2. Recurrent Neural Networks

In contrast to linear autoregressive models (cf. Section 2.3.1), the relationship between the current and previous values can be nonlinear for recurrent neural networks. The strength of neural networks is the capability to approximate

any functional relationship between input and output. Thus, at least in principle, a neural network can approximate the current values from previous values, if such a functional relationship exits. Machine learning approaches have been successfully applied in numerous areas, including fluid dynamics (Brunton et al., 2020) and ocean waves (Agrawal and Deo, 2002). To overcome the vanishing gradient problem while training recurrent neural networks for long time series, LSTM cells have been proposed (Hochreiter and Schmidhuber, 1997). Like any other recurrent neural network, LSTM cells feature internal states. As a distinction, LSTM cells are designed to flexibly retain or erase the internal states. Thereby, dependencies between values distant (in time) can be detected. The input-output relationship for LSTM cells is nonlinear and can be found, for example, in the references (Hochreiter and Schmidhuber, 1997; Goodfellow et al., 2016).

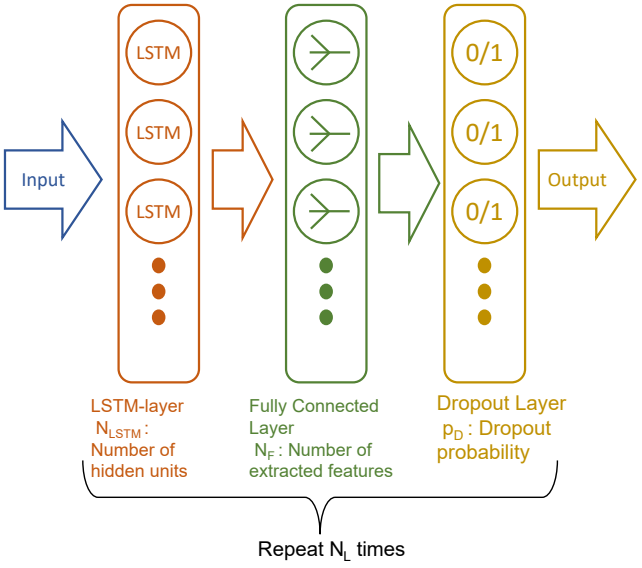

Figure 2: Architecture of the neural network utilized for ocean wave forecasting.

Within this work, the neural network architecture sketched out in Figure 2 is utilized. The first layer consists of N_{LSTM} LSTM cells in parallel, commonly, denoted as hidden units. Subsequently, a fully connected layer follows, through which, one selects N_f features from the output of the LSTM cells. This layer is connected to a dropout layer, which sets each feature to zero with a probability of p_D. The dropout layer is designed to prevent overfitting, a common issue with neural networks (Shalev-Shwartz, 2014). These three layers (LSTM, fully connected and dropout) are then repeated for N_L times. The weights of the involved functions are tuned with the stochastic optimizer Adam (Kingma and Ba, 2015)[4].

[4]The various hyper parameters have been selected based on the best performance. More specifically, the number of layers, number of hidden LSTM units, the dropout rate, and the feature size of the fully connected layer have been varied. The performance of the selected combination was found to be optimal for the trials conducted within this manuscript.

2.4. Real time forecasting strategy for buoy data[5]

As indicated in Figure 1, for each supplied measurement, an individual wave model (1) and forecasting method (linear methods or LSTM network, cf. Section 2.3) are parameterized. Thus, the time required to fit the wave model (1), reduces the useful forecasting horizon and hence it should be kept minimal (cf. also Figure 1). To this end, a multi-step forecasting strategy that is illustrated in Figure 3 is employed. All durations given in Figure 3 are approximate and the actual durations will vary depending on the selected parameters, such as the number of harmonics N or ϵ as well as the underlying data employed for the fitting[6].

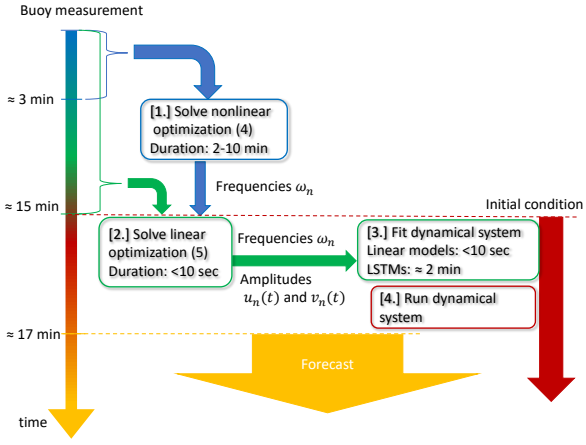

Figure 3: Forecasting strategy to enable real time forecasts.

The proposed methodology is constructed such that it can run in parallel while measurements are being made. The following time line lists the subsequent steps yielding ocean wave forecasts in real time:

Step 1 (0-3 minutes): For the first three minutes, the surface elevation is measured and no computations are carried out, since no data are available.

Step 2 (3-15 minutes): After three minutes of ocean surface elevations are recorded, the nonlinear optimization (4) is solved with this data. This step takes about two to twelve minutes, due to the employed Monte-Carlo strategy. During this time, data recording is of course continued; however, the recorded data is not fed into the minimization (4).

Step 3 (15-17 minutes): The frequencies ω_n obtained in the first optimization (4) are then utilized within the second subsequent optimization (5). Since the minimum is available in closed form, this step only takes a few seconds. The

[5]The employed strategy is only used for the field measurements from ocean buoys (cf. Section 3.2). Real time forecasts for wave tank data (cf. Section 3.1) are not obtained, since the available measurements are rather short (about 40 seconds).

[6]All computations have been performed by using MATLAB R2021a on Windows PC with Intel Core i7-8700 @ 3.2 GHz and 8GB RAM. Of course, the hard- and software of a real forecasting system will most likely differ. The observations and timings in this sections merely indicate that the computational burden is manageable and forecasts in real time are within reach. Indeed, the forecasts for the field measurements are obtained in real time with the computational hardware used.

data recorded while solving equation (4) are used within the quadratic minimization (5).

Once the time varying amplitudes $u_n(t)$ and $v_n(t)$ have been obtained, the parameters for the forecasting methods (cf. Section 2.3) are tuned. The computational burden to fit the linear systems is minimal. Since the LSTM network requires a nonlinear minimization, it takes significantly more time to obtain the parameters. After parameterizing the dynamical systems, time series can be generated almost instantly. The dynamical systems are initialized at the last known sample point (before fitting them). Thus, about the first two minutes of the generated time series are not forecasts as this time is already in the past, in practice.

Step 4 (after 17 minutes): All time series produced after about 17 minutes are then meaningful forecasts as indicated in Figure 3

It is noted, that the minimization (4) is solved only for the first three minutes of measurement while the optimization (5) is carried out for the first fifteen minutes. Of course, the obtained frequencies optimized for the first three minutes will be generally different than the frequencies optimized over the whole fifteen minutes. However, as previously mentioned, observations and most wave models indicate that the dominant frequencies in the ocean change only slowly. Thus, it is reasonable to expect that the frequencies obtained by optimizing with three minutes of measurements will also be a good choice to decompose the whole fifteen minute measurement window[7].

3. Results

To demonstrate the applicability and capability of the proposed methodology in Section 2, it is first applied to recordings from an experimental wave tank (Eeltink, 2022). The data from this well controlled experiment serves as a first viability test. Second, this methodology is applied to field measurements from ocean buoys (CDIP Buoy data, 2022) to demonstrate the applicability of the developed approach in realistic conditions. This demonstration exemplifies that the proposed methodology is robust to noise and other disturbances present in field measurements.

3.1. Experimental wave tank

The proposed approach (cf. Section 2) is applied to recordings from an experimental wave tank (Eeltink, 2022)[8]. This data stems from an investigation on wave breaking (Eeltink et al., 2022). The wave tank is 40 m long, 2.7 m wide, and 0.8 m deep, and the waves are generated by a piston wave maker positioned at one end of the tank (cf. Figure 4). In the experiments, a perturbed plane wave is generated to trigger the modulation instability and the surface elevation has been measured by twelve waves gauges. The sampling rate for these measurements is 400 Hz. More details on the experimental setting and generated waves can be found in (Eeltink et al., 2022).

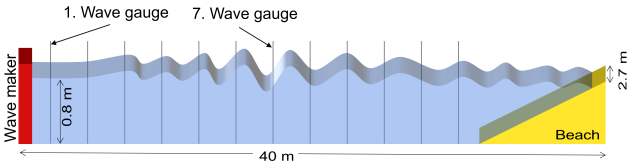

Figure 4: Sketch of the experimental wave tank. The first wave gauge is closest to the wave maker, whereas the seventh wave gauge is located towards the middle of the wave tank.

The proposed methodology is first applied to individual recordings from the first and seventh wave gauges (cf. Figure 4). After that, all 258 available measurements are considered and average statistics are presented.

3.1.1. *Individual time series*

In the following, two individual time series from one experiment are considered. In the experiment, the wave maker is used to generate a wave with a carrier band of 1.3 Hz and two equal side bands with a modulation frequency of 0.13 Hz and normalized amplitudes of 0.025. First, the measurements from wave gauge closest to the wave maker are considered (cf. Figure 4). For this wave gauge, the recorded waves are similar to the generated waves by the wave maker. Second, the measured surface elevation from the seventh wave gauge is forecasted with the proposed methodology. Since complex processes such as modulation instability and wave breaking alter the generated wave, the wave measurements from the seventh wave gauge differ significantly from the initiated part of a wave.

First wave gauge: Considering an individual recording from the wave gauge closest to the wave maker, the obtained spectrum is shown in Figure 5a. The depicted spectrum of the surface elevation is estimated by Welch's method (Welch, 1967). The carrier frequency at 1.3 Hz and the side bands are clearly discernible in Figure 5a.

As a first step, the wave model (1) is fitted to the data (Eeltink, 2022) as described in Section 2.1. For this purpose, one Fourier mode ($N = 1$), the parameter controlling the slowness $\varepsilon = 10^{-3}$ and $\alpha = 10$ are selected. Moreover, the nonlinear minimization (4) is started from one thousand choices for the initial frequency[9]. As shown Figure 5b, the fitted surface elevation closely resembles those of the actual measurements. The relative mean square error is about 0.1 percent.

[7] Indeed, within this work, it is observed that solving the minimizations (4) and (5) for fifteen minutes leads only to minor improvements compared to the employed strategy. For example, decomposing ten 15-minute windows of buoy measurements with the employed strategy leads to an average mean square error of 3.8%, whereas performing both minimizations (4) and (5) for fifteen minutes yields an average error of 3.5%.

[8] As noted in Section 2.4, the forecasting strategy shown in Figure 3 is not employed, since the available measurements are relatively short (about 40 seconds). Rather, the minimizations (4) and (5) are solved over the same time span.

[9] Since the carrier frequency is known in this example (1.3 Hz) this steps could be skipped. In reality, the carrier frequency is unknown, and hence it is treated as an unknown in the following. The minimization (4) indeed yields a dominant frequency of 1.305, which differs by less than one percent from the actual value.

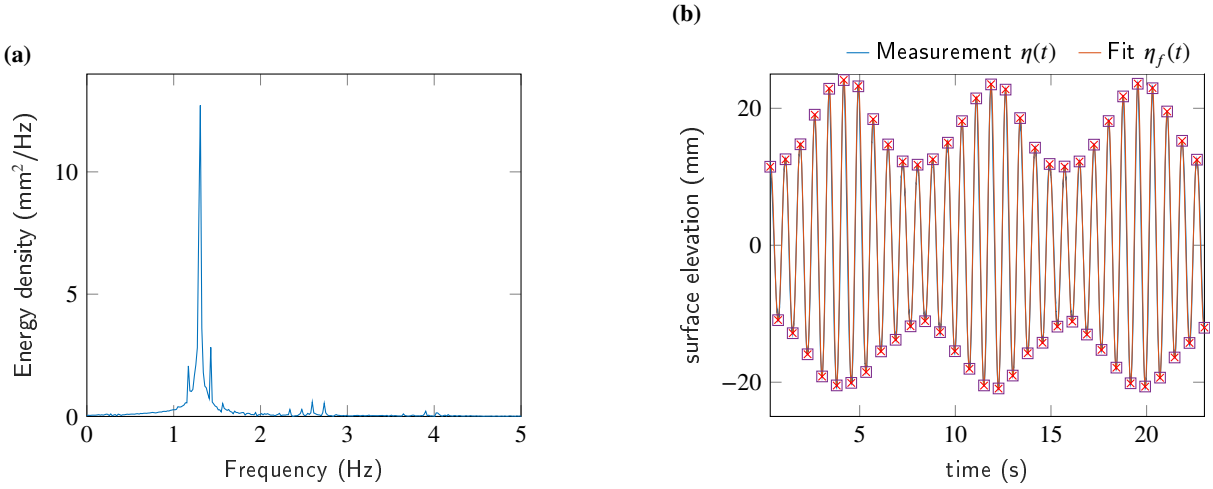

Figure 5: Wave measurements at gauge closest to the wave maker. The wave maker is used to initiate a wave of the form $\eta_{wm}(t) = 0.95\sin(1.3 \cdot 2\pi t) + 0.025\sin((1.3+0.13)2\pi t) + 0.025\sin((1.3-0.13)2\pi t)$. 5a: Spectrum of surface elevation. 5b: Measured and fitted surface elevation for the first 22 seconds.

The slowly varying amplitudes are shown in Figure 6a. As enforced by the cost function (2), the amplitudes vary slowly compared to the fast variation of the surface elevation (cf. Figure 5b). After fitting the model (1) for the first 22 seconds, the three different dynamical models presented in Section 2.3 are utilized to forecast the time evolutions of the amplitudes. The order for the linear autoregressive models was set to $M = 100$ for both, the LSQ model and the DMD model. The LSTM architecture consist of three repeated layers. The number of LSTM cells N_{LSTM} is set to 32 for the first, 16 for the second, and 8 for the last LSTM layer. The number of features for the three fully connected layers N_f was 100 for the first, and 50 for the second and third layer. The dropout portability p_D was kept constant at 5%.

The forecasted amplitude variations are included in Figure 6a. All three models pick up the general variations of the amplitudes. Their performance is discernible from the mean square error of three forecasts shown in Figure (6b). Therein, a fit of model (1) over all 44 seconds is considered as the ground truth. Additionally, the errors of two benchmarks are included. Within these two benchmarks, the variability of the amplitudes is ignored and only a single constant value for each amplitude is forecasted. For the first benchmark, the predicted value is the last known value after about 22 seconds, and hence, the name *last sample*. Furthermore, the forecast *training mean* is for the mean value over the first 22 seconds for the upcoming amplitudes. As indicated in Figure 6b, all three dynamical models can be used to improve the predictions from two simple benchmarks. Hence, the dynamical models can be used to successfully forecast the slowly evolving envelope. Especially, in the first ten seconds, the DMD model and LSTM model perform significantly better than the benchmarks and the LSQ model.

As the final outcome of the forecasting methodology, time series of surface elevations as the final outcome are shown in Figures 6c and 6d. In both forecasts, the rising and falling wave envelope is accurately captured, although some deviations are visible. The average error in the predicted wave crests and trough depths (mean difference between the yellow diamonds and blue crosses in Figures 6c and 6d) is about three millimeters.

Seventh wave gauge: While the wave dynamics close to the wave maker is clearly dominated by single harmonic (cf. Figure 5a), this spectrum evolves and gets distorted further away from the wave maker. Besides the triggered modulation instability, more complex processes such as wave breaking are expected to alter the spectrum. Indeed, the wave spectrum recorded at the seventh wave gauge has more dominant side bands as well as a broader frequency spectrum (cf. Figure 7a). These features could challenge the simplistic wave model (1). However, as shown in Figure 7b, the surface elevation at the seventh wave gauge can be accurately fitted by using the same parameters as for the first wave gauge; that is, $\alpha = 10$, $\varepsilon = 10^{-3}$ and, most importantly, a single frequency $N = 1$. It is remarkable that the wave model (1) can resolve the complex spectrum shown in Figure 7a with a single frequency.

The fitted time evolutions of the amplitudes for the first 22 seconds and the subsequent forecasts from the three dynamical models are shown in Figure 8a. Similarly as for the first wave gauge, the dynamical models capture the variations from the fitted data. The mean square error shown in Figure 8b reveals that LSTM model performs the best, while the LSQ model improves the simple benchmark from the training mean only marginally.

Additionally, the predicted surface elevations from the DMD model respectively LSTM model are depicted in Figures (8c) and (8d). Both predictions are accurate for the next twenty seconds, although some deviations in the LSTM model are visible. The average error of the forecasted wave crests and trough depths is 3 mm for the DMD model and

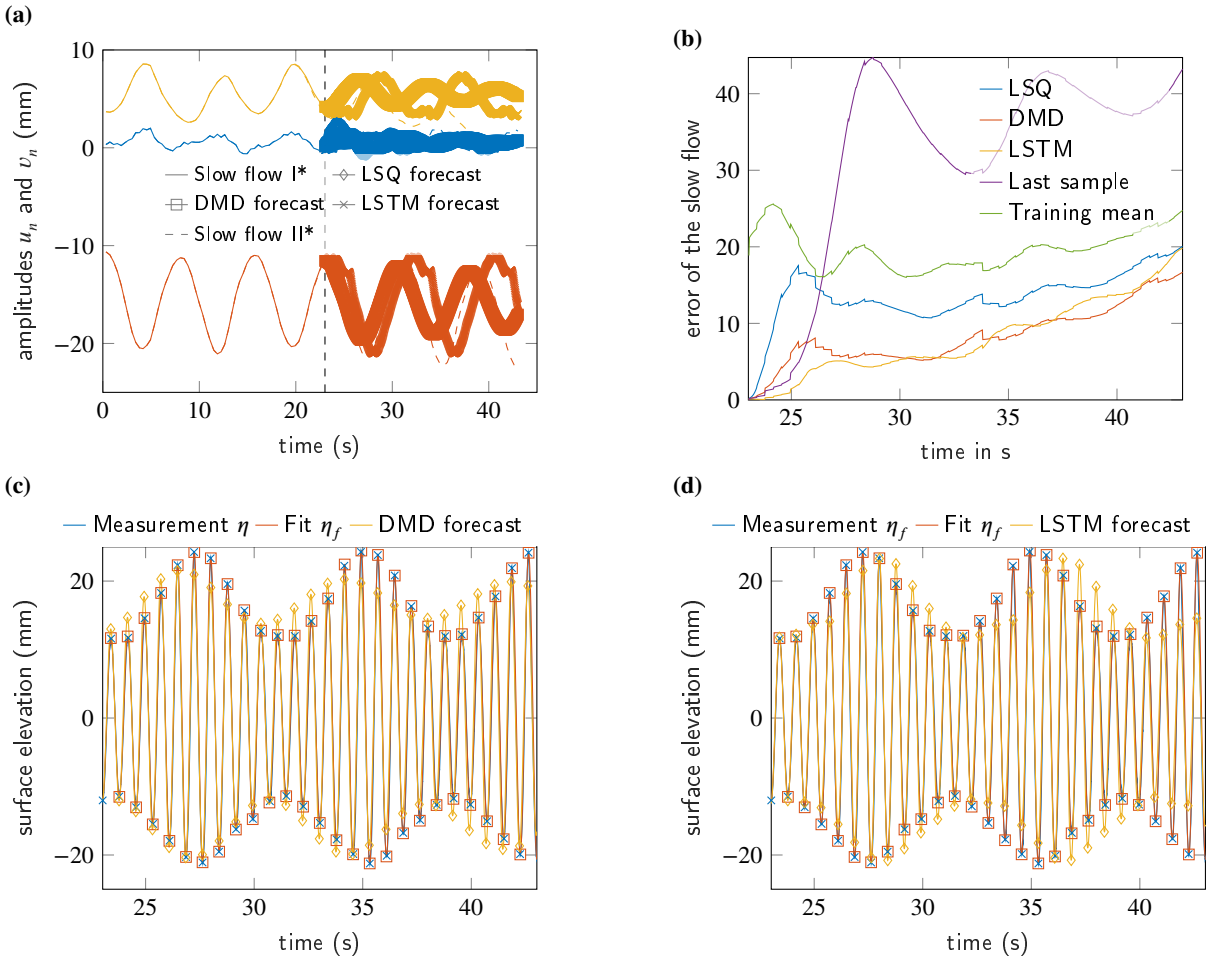

Figure 6: Forecasts for the first wave gauge. 6a: Fitted and forecasted time varying amplitudes. *The slow flow I is obtained for the first 22 seconds only and used for fitting the dynamical models. Slow flow II is fitted over the whole 44 second interval and used to compute the error shownin Figure 6b. 6b: Mean square errors in the time varying amplitudes for the various dynamical models (cf. Section 2.3) and two benchmarks. 6c: Forecasted (DMD model), fitted, and measured surface elevations. 6d: Forecasted (LSTM model), fitted and measured surface elevation.

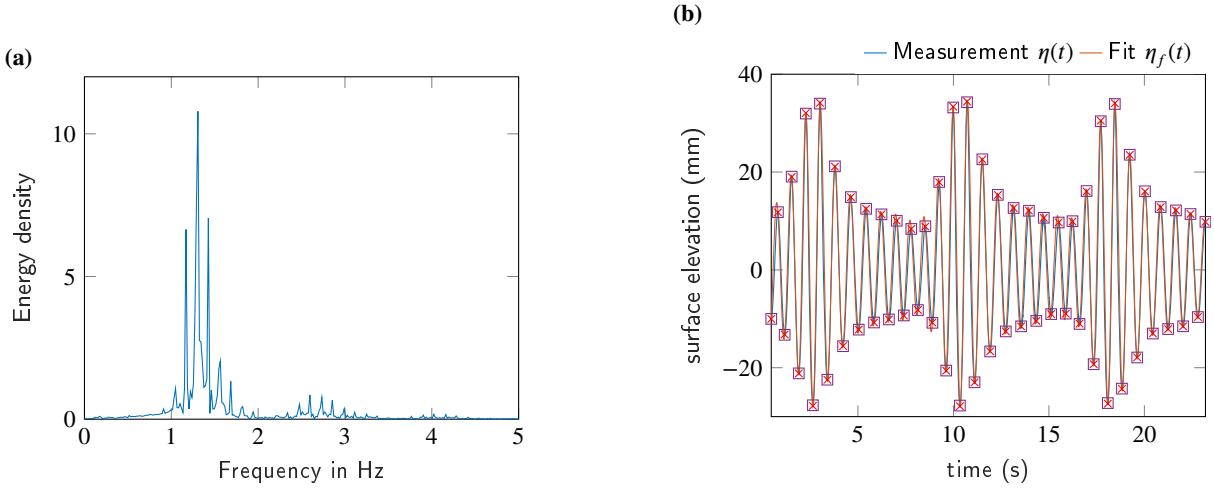

Figure 7: Wave measurements at the seventh wave gauge. The wave maker is used to initiate a wave of the form $\eta_{wm}(t) = 0.95\sin(1.3\cdot 2\pi t) + 0.025\sin((1.3+0.13)2\pi t) + 0.025\sin((1.3-0.13)2\pi t)$. 7a: Spectrum of surface elevation. 7b: Measured and fitted surface elevation for the first 22 seconds.

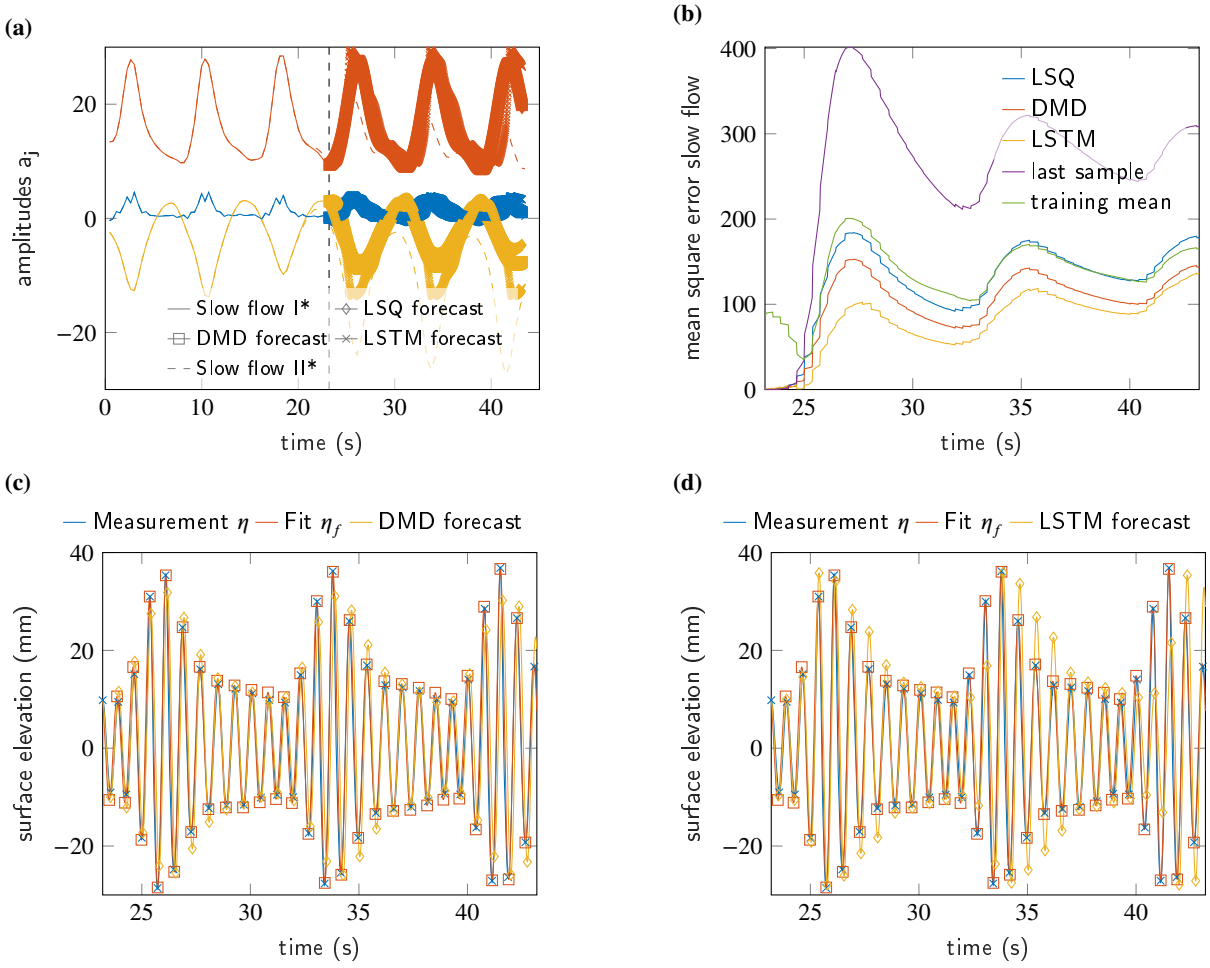

Figure 8: Forecasts for the seventh wave gauge. 8a: Fitted and forecasted time varying amplitudes. *The slow flow I is obtained for the first 22 seconds only and used for fitting the dynamical models. Slow flow II is fitted over the whole 44 second interval and used to compute the error shown in Figure 8b. 8b: Mean square error of the time varying amplitudes for the various dynamical models (cf. Section 2.3) and two benchmarks. 8c: Forecasted (DMD model), fitted, and measured surface elevations. 8d: Forecasted (LSTM model), fitted and measured surface elevation.

4 mm for the LSTM model. It is interesting to note that the LSTM model performs better than the DMD model when considering the mean square error of the varying amplitudes. However, the LSTM model performs worse when considering the actual wave crests and troughs.

After considering two individual recordings of a single experiment, in the next section, the proposed methodology is applied to all available recordings of regular waves from (Eeltink, 2022).

3.1.2. Comprehensive comparison including all available measurements

As a final demonstration, all 258 measurements of regular waves from (Eeltink, 2022) are used to test the proposed forecasting strategy. Each measurement is divided into two sequences of equal length. The first half is used to fit the model (1) and train the dynamical models (cf. Section 2.3). Subsequently, these models are employed to forecast a surface elevation for the remaining half. The crest heights and trough depths are extracted from these time series. These values are then compared to the crests and troughs from the measurements and an average error is computed. In addition to the LSQ, DMD, and LSTM model, a benchmark is included. In this benchmark, a single wave train is forecasted. Its amplitude corresponds to the mean of the absolute value of the crest heights and trough depths and the period is twice the mean duration between consecutive crests and troughs. Moreover, the wave model (1) is fitted for the whole measurement. The error from this fitting indicates how well the wave model (1) can be used to capture the wave dynamics.

Averaging the mean crest height and trough depth error over all 258 measurements yield the average errors shown in Figure 9. The average crest heights and trough depths errors for the wave gauge closest to the wave maker are depicted in Figure 9a. In this case, the average error from the DMD model is almost 50% less than the error from the simple benchmark. For the seventh wave gauge shown in

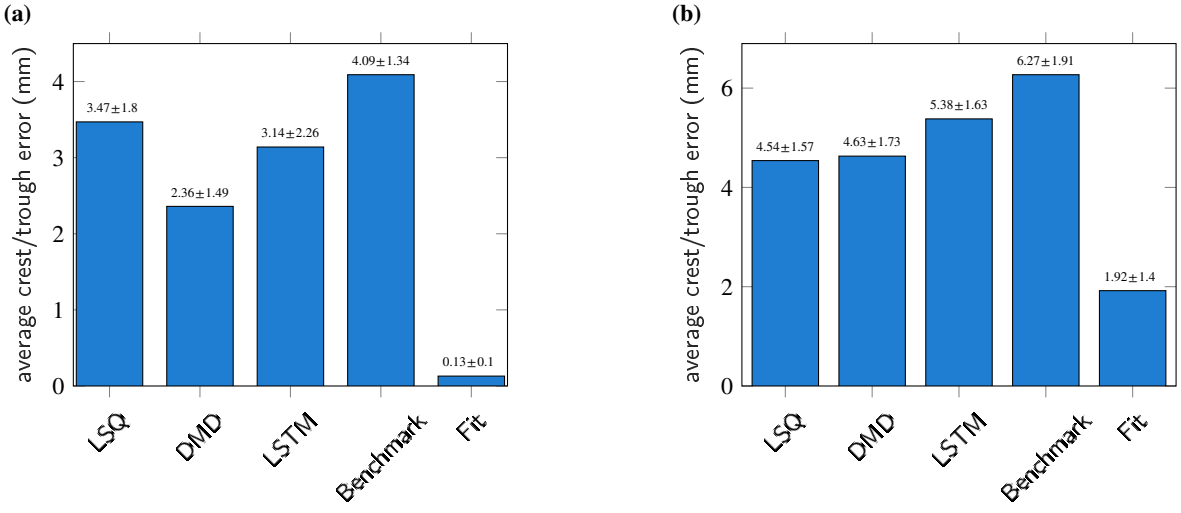

Figure 9: Average crest heights and trough depths error for all 258 regular wave measurements from (Eeltink, 2022). The numbers denote mean and standard deviation for the various models. 9a: Wave gauge closest to the wave maker. 9b: Seventh wave gauge.

Figure 9b, an improvement in the linear models of about 30% is obtained compared to the benchmark. Moreover, it is clearly discernible that the obtained fit for the seventh wave gauge is less accurate than for the first wave gauge. A reason for this discrepancy is the more complex wave profiles measured at the seventh wave gauge (compare also Figures 5a and 7a). Allowing for more than one frequency in the wave model (1) ($N > 1$) would allow to fit more complex wave envelopes and therefore could reduce the fitting error significantly. This approach will be examined further in the next section featuring buoy measurements (CDIP Buoy data, 2022).

In conclusion, the proposed forecasting strategy yields accurate forecasts for the experimental wave tank data (Eeltink, 2022). The slowly varying amplitudes can be extrapolated to yield forecasts for a duration as long as the training data. An improvement in terms of wave height forecasting of almost 50% is achieved. This application demonstrates that the envisioned strategy can work with wave measurements.

The authors emphasize that the performance demonstrated should not be taken as an absolute measure. Indeed, the performance of the proposed models can be, most certainly, significantly improved by tuning the various involved hyperparameters. Especially, given the very regular and recurrent wave patterns a better performance of the LSTM networks seems achievable. In reality, however, the ocean will drastically differ from the unidirectional and regular patterns created in the wave tank. Thus, such an envisioned numerical exercise would have only limited implications for practical ocean wave forecasting. To this end, an application to field measurements is inevitable. This is considered in the following section.

3.2. Field measurements from ocean buoys

After validating the proposed wave forecasting methodology on measurements obtained in an experimental wave tank, field measurements from ocean buoys (CDIP Buoy data, 2022) are analyzed. These measurements are recorded by Datawell directional waverider MkIII (Datawell B.V., 2020). Among other quantities, MkIII buoys measure vertical accelerations, which are subsequently processed, filtered and integrated to yield sea surface elevations. The sampling rate for the buoy's vertical displacement measurements is 1.28 Hz. These sea surface elevation measurements are used for fitting and forecasting in the following[10]. This data serves as a realistic benchmark to also demonstrate that the proposed methodology is suitable for the noise levels and disturbances present in ocean surface measurements.

In the following, the buoy data (CDIP Buoy data, 2022) is decomposed by using $N = 16$ Fourier modes in the model (1). Moreover, the parameters $\epsilon = 10^{-3}$ and $\alpha = 0.5$ for the cost function (2) are selected. The order for the linear dynamical systems was set to $M = 50$ for the DMD model and $M = 100$ for the LSQ model. The LSTM network consists of one stacked layer with $N_{LSTM} = 200$ LSTM units, a fully connected layer extracting $N_f = 100$ features, and a dropout layer with probability $p_D = 0.1$.

At first, measurements from a buoy near San Nicholas Island are considered. Subsequently, the capabilities of the proposed methodology are demonstrated for alternative buoy locations, including open, deep water and coastal, shallow water. The discussion is concluded by forecasting two rogue waves measured by the buoy near San Nicholas Island.

[10]It is noted, that surface buoys move laterally and tend to avoid large wave crests (Dysthe et al., 2008). Since, MkIII-buoys also pick up lateral displacements (Datawell B.V., 2020), the model (1) could, at least in principle, be extended to incorporate these measurements. However, to keep the wave model (1) simple and allow for fast computations, the lateral buoy movement is ignored in this study. It is implicitly assumed that the measurements are from a laterally fixed observer such as the wave gauges in the experimental wave tank in Section 3.1.

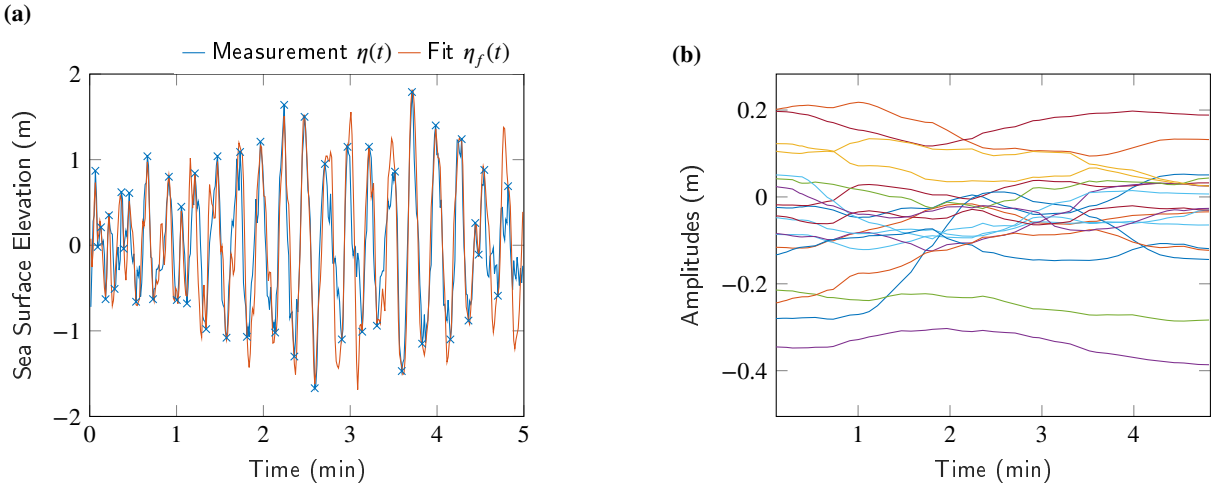

Figure 10: Obtained fit of measured ocean surface elevation (CDIP Buoy data, 2022) for five minutes. The parameters for the cost function (2) are set to $N = 16$, $\varepsilon = 10^{-3}$, and $\alpha = 0.5$. 10a: Measured and fitted sea surface elevations. 10b: Slowly varying amplitudes.

3.2.1. San Nicholas Island (CDIP Buoy 067)

At first, measurements from a Datawell directional waverider MkIII (Datawell B.V., 2020) near San Nicholas Island off the coast from Los Angeles (CDIP Buoy 067) between Feb., 9th 2018 and Mar., 5th 2019 are considered. A representative outcome of the employed fitting strategy is shown in Figure 10a. Therein, the ocean surface measurements (CDIP Buoy data, 2022) are decomposed into ocean waves with slowly varying spectrum, whereby only the wave crests and troughs are kept for fitting (blue crosses in Figure 10a). As enforced by the cost function (2), the fitted model (1) resembles closely the measured sea surface elevation. Moreover, the amplitudes indeed vary slowly as illustrated in Figure 10b. The relative error for the time series shown in Figure 10a is less than one percent. Repeating the fitting for one hundred randomly selected measurements with the aforementioned parameters yields an average relative error of 3.7 percent with a standard deviation of 2.9 percent.

A forecast for the amplitude u_7 is displayed in Figure 11. Therein, the wave model (1) and dynamical systems are fitted for the first thirteen minutes and, subsequently, three forecasts are obtained. Whereas the time series generated by the linear models fluctuate less, the more complex LSTM model is found to forecast some of the variations in true time series (Slow flow II). Overall, the correct trend is predicted by all three models.

After obtaining forecasts for the slowly varying amplitudes, these time series are inserted into the wave model (1) to yield forecasts for the future sea surface elevation (cf. synthesis in Figure 1). From these predictions, the extreme values are extracted, which yields forecasts for the upcoming wave heights and trough depths.

The performance of the proposed forecasting strategy is evaluated by randomly selecting one hundred windows from the buoy close to San Nicholas Island. Seven minutes

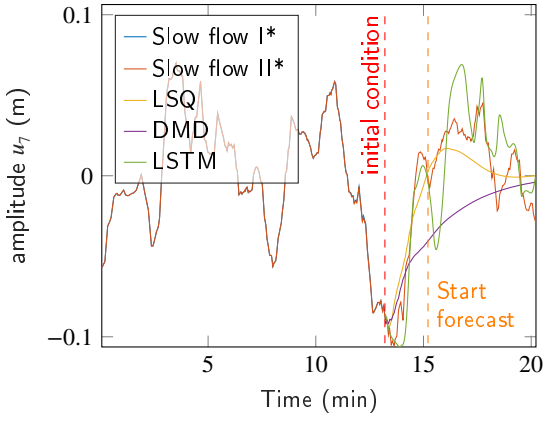

Figure 11: Fitted and forecasted amplitude u_7. *The slow flow I is obtained for the first 13 minutes only and used for fitting the dynamical models. Slow flow II is fitted over the whole 20 minutes.

of sea surface are forecasted and the first two minutes are discarded, which is the duration needed to generate forecasts following the strategy illustrated in Figure 3. Then, the crest heights and trough depths are extracted from the five minute time series and compared to the truth.

The authors have plotted the observed versus the predicted crest heights and trough depths in Figure 12. The line $x = y$ indicating a perfect prediction is included for comparison. The forecasts from the DMD model, shown in Figure 12a, depart farther from the $x = y$-line than the prediction from the LSTM model depicted in Figure 12b. The higher correlation coefficient of the LSTM forecast also indicates a higher accuracy of the predictions from the LSTM model compared to the linear DMD model. Overall,

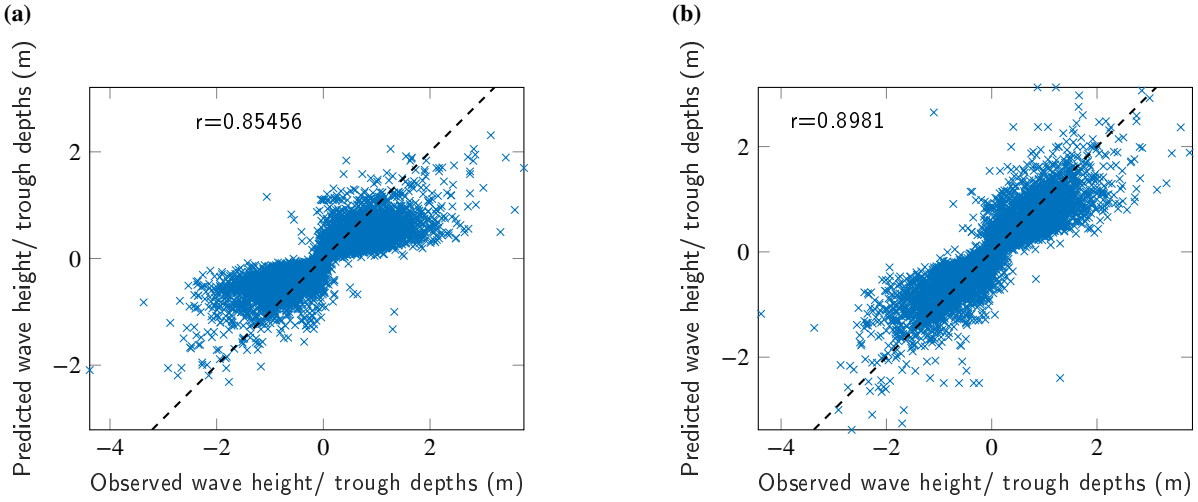

Figure 12: Observed crest heights and trough depths versus the predictions for the measurements from a buoy located near San Nicholas island (water depth: 262m) from (CDIP Buoy data, 2022). The correlation coefficient (r) is also included. Both figures show the forecasts of about 9000 crest heights and trough depths. 12a: Predictions from the DMD model. 12b: Predictions from the LSTM network.

Table 1
San Nicholas Island: Correlation coefficient, root mean square error (RMSE), and mean absolute error (MAE) for the various models for three different forecasting horizons. The statistics are based on up to 9000 crest heights and trough depths.

Forecast	Correlation					RMSE					MAE				
horizon	LSQ	DMD	LSTM	Ben. [1]	Fit	LSQ	DMD	LSTM	Ben.[1]	Fit	LSQ	DMD	LSTM	Ben.[1]	Fit
1 min.	0.88	0.88	0.91	0.85	0.97	0.4	0.4	0.33	0.41	0.18	0.27	0.27	0.21	0.31	0.12
3 min.	0.88	0.87	0.9	0.85	0.97	0.41	0.42	0.34	0.41	0.18	0.27	0.29	0.22	0.31	0.12
5 min.	0.86	0.85	0.9	0.85	0.97	0.42	0.44	0.34	0.41	0.19	0.28	0.29	0.22	0.31	0.12

[1] The benchmark consists of a single wave train. The amplitude and frequency of this wave train are extracted from the training data.

about 9000 predictions of wave heights and trough depths are obtained for each model.

The accuracy of the three models is further summarized in Table 1. As in the earlier analysis of the data from the experimental wave tanks (cf. Section 3.1), a benchmark forecast consisting of a single wave train with the mean crest heights and trough depths from the training data is included. Moreover, the fitting error is also incorporated in Table 1, which allows to assess the accuracy of the wave model (1).

As Table 1 indicates, the linear models (DMD and LSQ) can only slightly outperform the benchmark. Their performance clearly deteriorates with time. The LSTM network on the other hand clearly outperforms the benchmark in both error norms as well as in terms of a higher correlation coefficient. This performance is observed for all time instances. The relatively small fitting error indicate that the wave model (1) can accurately and reliably capture the complex sea surface elevations arising in field measurements.

The mean absolute error corresponds to the mean average crest height/trough depths error which is also shown in Figure 13a. In this error norm, the LSTM model is found to perform the best, improving the benchmark by about 30%.

To further indicate that the proposed strategy can also accurately predict large waves, the highest crest respectively deepest troughs from the forecasts are compared to the measurements in Figure 13b. Also, in this comparison, the LSTM model is found to perform superior, improving the two linear models by about 30%.

3.2.2. *Alternative buoy locations - Deep and shallow water*

The wave model (1) and fitting strategies are not location specific, and hence, this methodology can be applied to measurements from any other buoy location. To demonstrate this universality, two additional buoys are selected. The first buoy is located near Maui, Hawaii (CDIP Buoy 187) and the other one in the Onslow Bay, North Carolina (CDIP Buoy 217). While the buoy in Hawaii is employed in a water depths of 200 m, which is comparable to the buoy close to San Nicholas Island, the buoy in the Onslow bay on the other hand is deployed in significantly more shallow water with a depths of 30 m. In shallow, coastal water, effects such as refraction from the shoreline and shoaling impact the observed waves (Reeve et al., 2018). Considering data from a shallow water buoy helps to verify the capabilities

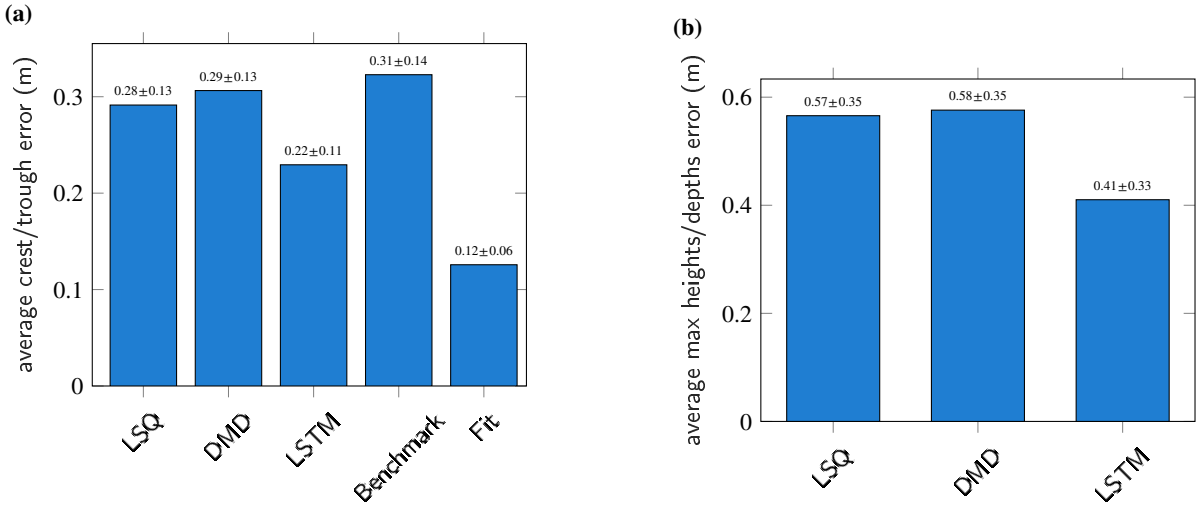

Figure 13: Crest heights and trough depths obtained for one hundred randomly selected windows from a buoy located near San Nicholas island (water depth: 262m) from (CDIP Buoy data, 2022). The numbers denote mean and standard deviation for the various models. 13a: Average error considering all crests and troughs. 13b: Average error of the highest crest/deepest trough.

Table 2
Maui: Correlation coefficient, root mean square error (RMSE), and mean absolute error (MAE) for the various models for three different forecasting horizons. The statistics are based on up to 10000 crest heights and trough depths.

Forecast	Correlation					RMSE					MAE				
horizon	LSQ	DMD	LSTM	Ben.[1]	Fit	LSQ	DMD	LSTM	Ben.[1]	Fit	LSQ	DMD	LSTM	Ben.[1]	Fit
1 min.	0.85	0.84	0.89	0.84	0.98	0.41	0.42	0.34	0.4	0.17	0.25	0.27	0.21	0.3	0.12
3 min.	0.82	0.84	0.91	0.85	0.98	0.42	0.43	0.31	0.39	0.17	0.26	0.27	0.20	0.29	0.12
5 min.	0.75	0.81	0.91	0.85	0.97	0.49	0.44	0.31	0.38	0.18	0.29	0.29	0.20	0.29	0.12

[1] The benchmark consists of a single wave train. The amplitude and frequency of this wave train are extracted from the training data.

of the proposed forecasting approach to adjust to different, realistic environments.

The performance demonstration from the previous section is repeated for the two aforementioned buoy locations. One hundred windows are randomly selected to generate five minute real-time forecast each. Then, the forecasted crest heights and trough depths are compared with the corresponding true values. The arising errors and correlations are listed in Table 2 for the buoy near Maui and in Table 3 for the buoy in the Onslow Bay.

For the data from Maui, the models show a similar performance as in the previous forecasting experiment based on the field data recorded close to San Nicholas Island. The wave model (1) can be used to accurately capture the measured sea surface elevations. The linear models, deteriorate even faster, and after five minutes, the RMSE is lager than the benchmark and the correlation coefficient is less than the benchmark. The LSTM model on the other hand outperforms the benchmark model.

The trends from the previous two experiments is confirmed for the shallow water buoy in the Onslow Bay. It is notable, that the simple benchmark performs very well for this example (cf. Table 3). This indicates a more regular wave pattern for this shallow water example. The mean absolute error corresponding to the mean crest height/trough depths error is also shown in Figure 14 along with the average error for the highest crest/deepest trough for the two locations. In this error norm, the LSTM model is found to perform the best. The prediction from this model are, on the average, about 30% more accurate than the benchmark. Furthermore, the LSTM model is also found to outperform the linear models in predicting large ocean waves.

3.2.3. Rogue wave forecasting

The section is concluded by focusing on the two largest rogue waves in the recordings of the buoy near San Nicholas island. The first rogue wave occurred on Dec. 1st at about 3:00 UTC. A crest height of 6.9 m was measured (wave height: 11.7 m) while the significant wave height H_s was 5.13 m. The second extreme wave occurred on January, 19th at about 1:00 UTC. The measured crest height was 5.5 m (wave height 9 m) and the significant wave height H_s was 3.9 m. The forecast from the LSTM model for the first rogue wave is shown in Figure 15a, while the prediction for the second rogue wave is depicted in Figure 15b. Although the full extent of both rogue waves is clearly underestimated in Figure 15, the rise and fall of the wave envelope are correctly

Table 3
Onslow Bay: Correlation coefficient, root mean square error (RMSE) and mean absolute error (MAE) for the various models for three different forecasting horizons. The statistics are based on up to 13000 crest heights and trough depths.

Forecast	Correlation					RMSE					MAE				
horizon	LSQ	DMD	LSTM	Ben.[1]	Fit	LSQ	DMD	LSTM	Ben.[1]	Fit	LSQ	DMD	LSTM	Ben.[1]	Fit
1 min.	0.83	0.87	0.93	0.87	0.97	0.27	0.25	0.19	0.23	0.12	0.15	0.16	0.12	0.18	0.08
3 min.	0.81	0.78	0.89	0.87	0.97	0.27	0.29	0.21	0.23	0.12	0.16	0.17	0.12	0.18	0.08
5 min.	0.71	0.74	0.86	0.87	0.97	0.34	0.31	0.24	0.23	0.12	0.18	0.18	0.12	0.18	0.08

[1] The benchmark consists of a single wave train. The amplitude and frequency of this wave train are extracted from the training data.

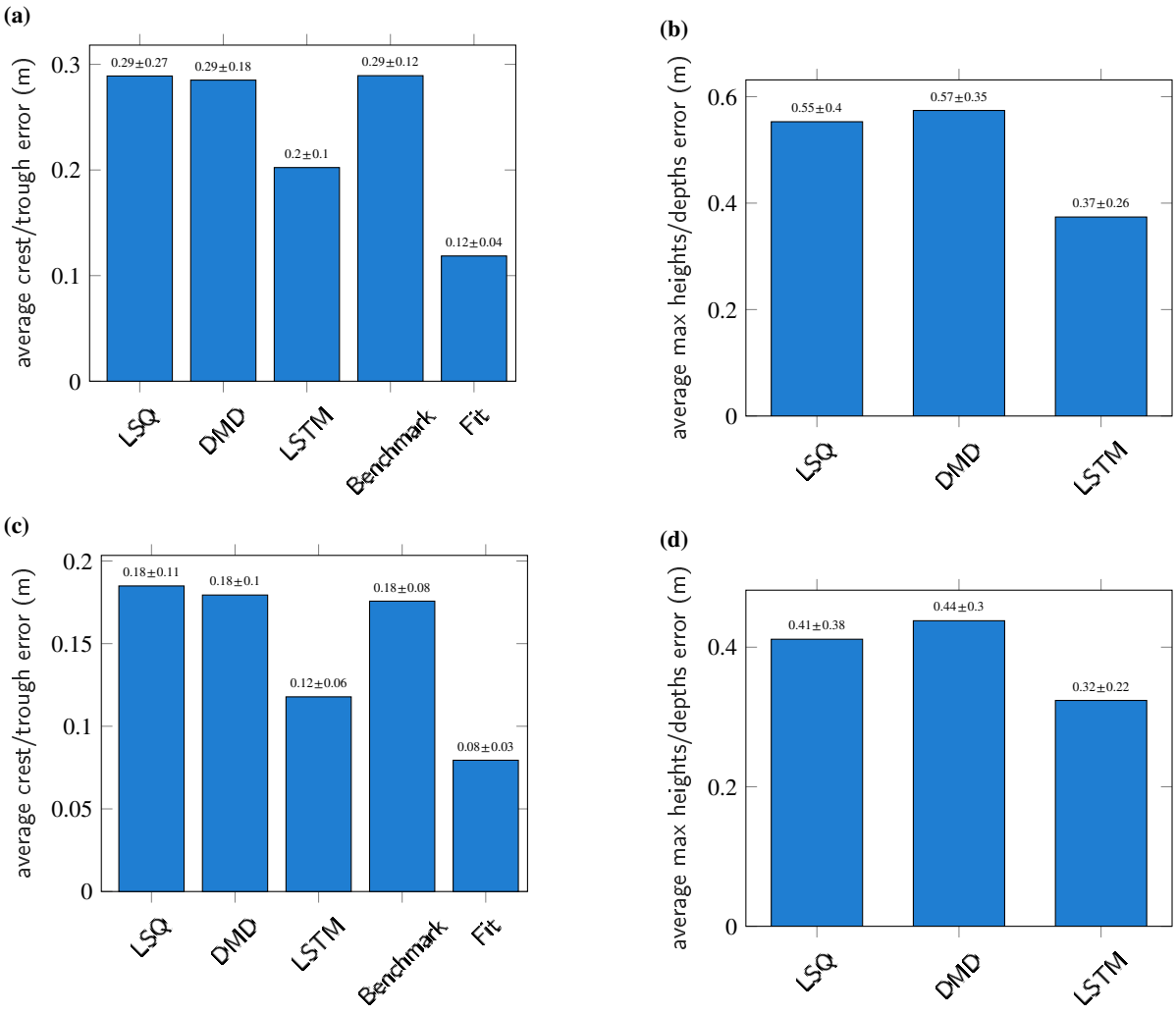

Figure 14: Crest heights and trough depths obtained for one hundred randomly selected windows. The bouys are located near Maui (water depth: 200 m) for the results shown in 14a and 14b and in the Onslow Bay (water depth: 20 m) for the results shown in 14c and 14d. 14a and 14c: Average error considering all crests and troughs. 14b and 14d: Average error of the highest crest/deepest trough.

predicted. In both examples of Figure 15, the rogue wave is predicted about one minute in advance.

4. Discussion

As demonstrated in Section 3, the proposed methodology can be used to accurately predict upcoming ocean waves with a high temporal resolution. This capability ultimately enables forecasts of rogue waves (cf. Figure 15). In the following, the assumptions and limitations of the proposed methodology are summarized and possible extensions are discussed.

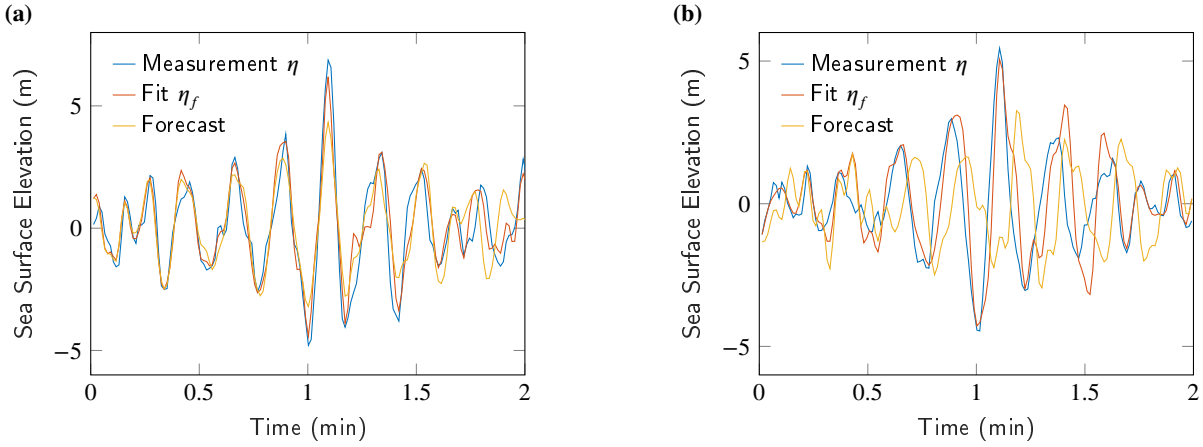

Figure 15: Forecasted sea surface elevation for the two rogue wave events close to San Nicholas Island. The measurements and fit are included for comparison. 15a: Rogue wave occurred on Dec. 1st. 15b: Rogue wave occurred on Jan. 19th.

The developed methodology is proposed for single point observations, in accordance with the current limited, high-resolution (in time) field measurements of sea surface elevation stemming from surface buoys (CDIP Buoy data, 2022). However, the wave model (1) and the utilized fitting strategies can in principle be extended to incorporate spatial dependencies. For this purpose, validated high-resolution (in time and space) sea surface measurements need to be available. In this setting, a predicted rouge wave at a single location could also serve as an extreme wave indicator for locations in the travel direction of the extreme wave. Hence, the time that a rogue wave can be predicted in advance could be increased by spatial observations.

Motivated by the necessity of resolving individual waves for rogue waves predictions, the developed forecasting strategy is used to generate forecasts of the raw sea surface with a sampling rate of about 1 Hz. This sampling interval is considerable shorter than the sampling interval of current operational wave forecasting models based on ocean spectra (about 30 minutes). This capability comes with the price that the forecasts are only meaningful for intermediate time spans of several minutes (cf. Section 3), compared to the multi-day ahead predictions of spectral wave models. Of course, it will be of immediate interest to increase the forecasting horizon as well as accuracy. However, the proposed deterministic approach will be fundamentally limited to intermediate time scales (in the range of minutes), since long-range forecasts with spectral wave models are based on stochastic models. The occurrence of individual extreme wave, however, is not discernible from such models.

Moreover, in the proposed methodology the authors treat the individual supplied measurements completely independent. For each time series, new parameterizations of the wave model (1) and the forecasting methods (linear or LSTM network) are obtained. This procedure is motivated by the fact that individual recordings can be separated by month or even years in time and stem from completely different buoy locations (cf. Section 3.2.2). In this setting, it is unclear whether these measurements are correlated. Within the proposed method, no relationship is assumed, which is clearly a conservative assumption. A more extensive analysis on the buoy data could be used to extract common patterns from distant (in time and location) buoy measurements and utilize those in forecasting tasks.

Finally, the wave model (1) as well as the forecasting methods (cf. Section 2.3) do not require location specific tuning. For the shallow water example in Section 3.2.2 the performance of the linear models deteriorates significantly. On the other hand the performance of benchmark improved, which indicates a more regular behaviour of the measured ocean waves. This regularity suggest that a location specific tuning of the hyperparameter for the linear models could help improving their performance.

5. Conclusions

The wave model (1) is utilized to generate high resolution forecasts of the sea surface elevation in real time. The model (1) is fitted to available measurements by using the cost function (2) and decomposing a ocean wave into fast oscillations and slowly varying amplitudes. This decomposition is inspired from common wave models and available field observations. Moreover, this separation of time scales is beneficially utilized to extend the extrapolation range of the universal, black-box models such as linear autoregressive models and a neural network with LSTM cells utilized to forecast the amplitude variations.

The proposed strategy is first tested on measurements from an experimental wave tank (Eeltink, 2022). The wave tank measurements can be efficiently decomposed and the upcoming wave can be accurately forecasted. Herein, the more complex wave patterns at the seventh wave gauge pose a greater challenge than more regular waves observed at the first wave gauge closest to the wave maker. The linear autoregressive model performance is found to be comparable to the neural network approach for this data set.

The field measurements from ocean buoys (CDIP Buoy data, 2022) can also be efficiently decomposed into fast oscillations and slowly varying amplitudes. Employing the strategy illustrated in Figure 3, real time forecasts were obtained. Therein, the crest heights and trough depths predicted with the LSTM model were found to be more accurate than the linear models and relevant benchmark. Moreover, the proposed strategy is found to perform equally well for measurements from buoys employed in deep, open water and coastal, shallow water. Finally, the work is concluded by obtaining real time forecasts of two rogue waves. These predictions are obtained one minute in advance.

It will be the authors' constant effort to increase the forecasting accuracy as well as the horizon. To this end, the harmonic functions in the wave model (1) can be replaced with more complex functions. These could be either physically motivated such as Stokes waves or solitons (Ablowitz and Clarkson, 1991) or learned from data. Another approach could include alternative neural network architectures, for example, encoder-decoder networks.

Within this work, two rogue waves have been predicted about one minute in advance. To confidently use the proposed approach for rogue wave forecasting, a broader validation on measurements of extreme waves is desirable. Furthermore, the systematic underestimation of the highest crests could be further analyzed and corrective measures could be proposed and tested.

Moreover, the utilized decomposition of data into rapid oscillations and slowly varying spectrum can also be applied in contexts beyond ocean wave forecasting. Indeed, a similar separation of time scales have been observed in other branches of environmental science as well as in various areas such as mechanical vibrations or laser physics. It would be of interest to explore the capabilities of the model (1) along with the cost function (2) for these applications.

Acknowledgements. The authors gratefully acknowledge the support of the National Science Foundation through Grant No. CMMI1854532. Additionally, they are thankful to Samarpan Chakraborty, Samuel Dipasqua, Soheil Feizi, Kayo Ide, and Aya Abdelsalam Ismail of the University of Maryland, College Park, for ongoing discussions on this work. The data (CDIP Buoy data, 2022) were available from the Coastal Data Information Program (CDIP), Integrative Oceanography Division, operated by the Scripps Institution of Oceanography, under the sponsorship of the U.S. Army Corps of Engineers and the California Department of Parks and Recreation. The authors are also thankful to the feedback received from the reviewers, which has helped to improve the quality of the discussion.

Data Availability Statement

The measurements from the experimental wave tank are freely accessible (Eeltink, 2022). The ocean buoy data is publicly available (CDIP Buoy data, 2022). The scripts utilized within this study to obtain the results presented in Section 3 will be made openly available at an appropriate repository.

Conflict of Interest

The authors have no conflicts to disclose.

References

Ablowitz, M.A., Clarkson, P.A., 1991. Solitons, Nonlinear Evolution Equations and Inverse Scattering. London Mathematical Society Lecture Note Series, Cambridge University Press. doi:10.1017/CBO9780511623998.

Agrawal, J., Deo, M., 2002. On-line wave prediction. Marine Structures 15, 57–74. doi:https://doi.org/10.1016/S0951-8339(01)00014-4.

Alam, M.R., 2014. Predictability horizon of oceanic rogue waves. Geophysical Research Letters 41, 8477–8485.

Benjamin, T.B., Feir, J.E., 1967. The disintegration of wave trains on deep water part 1. theory. Journal of Fluid Mechanics 27, 417–430. doi:10.1017/S002211206700045X.

Birkholz, S., Brée, C., Veselić, I., Demircan, A., Steinmeyer, G., 2016. Ocean rogue waves and their phase space dynamics in the limit of a linear interference model. Scientific reports 6, 1–8.

Bitner-Gregersen, E.M., Fernandez, L., Lefèvre, J.M., Monbaliu, J., Toffoli, A., 2014. The north sea andrea storm and numerical simulations. Natural Hazards and Earth System Sciences 14, 1407–1415. doi:10.5194/nhess-14-1407-2014.

Breunung, T., Balachandran, B., 2022. Freak wave forecasting: A data-driven approach, in: Proceedings of the ASME 202 2022 41st International Conference on Ocean, Offshore and Arctic Engineering.

Broomhead, D., King, G.P., 1986. Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena 20, 217–236. doi:https://doi.org/10.1016/0167-2789(86)90031-X.

Brunton, S.L., Noack, B.R., Koumoutsakos, P., 2020. Machine learning for fluid mechanics. Annual Review of Fluid Mechanics 52, 477–508. doi:10.1146/annurev-fluid-010719-060214.

Cattrell, A.D., Srokosz, M., Moat, B.I., Marsh, R., 2018. Can rogue waves be predicted using characteristic wave parameters? Journal of geophysical research. Oceans 123, 5624–5636.

Cavaleri, L., Alves, J.H., Ardhuin, F., Babanin, A., Banner, M., Belibassakis, K., Benoit, M., Donelan, M., Groeneweg, J., Herbers, T., et al., 2007. Wave modelling – the state of the art. Progress in oceanography 75, 603–674.

CDIP Buoy data, 2022. Coastal Data Information Program (CDIP), Integrative Oceanography Division, Scripps Institution of Oceanography. https://doi.org/10.18437/C7WC72. Accessed: 2022-22-06.

Chabalko, C., Moitra, A., Balachandran, B., 2014. Rogue waves: new forms enabled by gpu computing. Physics Letters A 378, 2377–2381.

Chakraborty, S., Balachandran, B., 2021. Wave propagation studies in numerical wave tanks with weakly compressible smoothed particle hydrodynamics. Journal of Marine Science and Engineering 9. doi:10.3390/jmse9020233.

Datawell B.V., 2020. Datawell waverider reference manual dwr-mkiii. https://www.datawell.nl/Portals/0/Documents/Manuals/datawell_manual_dwr-mk3_dwr-g_wr-sg_2020-02-01.pdf. Accessed: 2022-05-07.

Didenkulova, I., Slunyaev, A., Pelinovsky, E., Kharif, C., 2006. Freak waves in 2005. Natural Hazards and Earth System Sciences 6, 1007–1015.

Dysthe, K., Krogstad, H.E., Müller, P., 2008. Oceanic rogue waves. Annu. Rev. Fluid Mech. 40, 287–310.

Dysthe, K.B., 1979. Note on a modification to the nonlinear schrodinger equation for application to deep water waves. Proceedings of the Royal Society of London. Series A, Mathematical and physical sciences 369, 105–114.

Eeltink, D., 2022. Data nonlinear wave evolution with data-driven breaking. doi:10.5281/zenodo.6326470. accessed: 2022-22-06.

Eeltink, D., Branger, H., Luneau, C., He, Y., Chabchoub, A., Kasparian, J., van den Bremer, T., Sapsis, T., 2022. Nonlinear wave evolution with data-driven breaking. Nature communications 13, 1–11.

Farazmand, M., Sapsis, T.P., 2019. Extreme Events: Mechanisms and Prediction. Applied Mechanics Reviews 71. doi:10.1115/1.4042065. 050801.

Fochesato, C., Grilli, S., Dias, F., 2007. Numerical modeling of extreme rogue waves generated by directional energy focusing. Wave Motion 44, 395–416. doi:https://doi.org/10.1016/j.wavemoti.2007.01.003.

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

Häfner, D., Gemmrich, J., Jochum, M., 2021. Real-world rogue wave probabilities. Scientific reports 11, 10084–10084.

Hasselmann, K., 1962. On the non-linear energy transfer in a gravity-wave spectrum part 1. general theory. Journal of Fluid Mechanics 12, 481–500. doi:10.1017/S0022112062000373.

Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., Enke, K., Ewing, J., Gienapp, A., Hasselmann, D., Kruseman, P., et al., 1973. Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap). Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A .

Haver, S., 2000. Evidences of the existence of freak waves, in: Olagnon, M., Athanassoulis, G. (Eds.), Rogue Waves 200, Ifremer. pp. 129 – 140.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Janssen, P., Bidlot, J.R., Abdalla, S., Hersbach, H., 2005. Progress in ocean wave forecasting at ECMWF. ECMWF Reading, UK.

Janssen, P.A., 2003. Nonlinear four-wave interactions and freak waves. Journal of Physical Oceanography 33, 863–884.

Kagemoto, H., 2020. Forecasting a water-surface wave train with artificial intelligence-a case study. Ocean Engineering 207, 107380.

Kagemoto, H., 2022. Forecasting a water-surface wave train with artificial intelligence (part 2)–can the occurrence of freak waves be predicted with ai? Ocean Engineering 252, 111205.

Kharif, C., Pelinovsky, E., 2003. Physical mechanisms of the rogue wave phenomenon. European Journal of Mechanics-B/Fluids 22, 603–634.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization, in: ICLR (Poster).

Komen, G.J., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., Janssen, P., 1996. Dynamics and modelling of ocean waves.

Lea, C., Vidal, R., Reiter, A., Hager, G.D., 2016. Temporal convolutional networks: A unified approach to action segmentation, in: Hua, G., Jégou, H. (Eds.), Computer Vision – ECCV 2016 Workshops, Springer International Publishing, Cham. pp. 47–54.

Moitra, A., Chabalko, C., Balachandran, B., 2016. Extreme wave solutions: Parametric studies and wavelet analysis. International Journal of Non-Linear Mechanics 83, 39–47.

Mori, N., Onorato, M., Janssen, P.A.E.M., 2011. On the estimation of the kurtosis in directional sea states for freak wave forecasting. Journal of physical oceanography 41, 1484–1497.

Osborne, A.R., 2002. Nonlinear ocean wave and the inverse scattering transform, in: Scattering. Elsevier, pp. 637–666.

Rahman, M., 1995. Water waves: relating modern theory to advanced engineering applications .

Reeve, D., Chadwick, A., Fleming, C., 2018. Coastal engineering : processes, theory and design practice. Third edition. ed., CRC Press, an imprint of Taylor and Francis, Boca Raton, FL.

Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D.S., 2009. Spectral analysis of nonlinear flows. Journal of Fluid Mechanics 641, 115–127. doi:10.1017/S0022112009992059.

Schmid, P.J., 2010. Dynamic mode decomposition of numerical and experimental data. Journal of Fluid Mechanics 656, 5–28. doi:10.1017/S0022112010001217.

Schmid, P.J., 2022. Dynamic mode decomposition and its variants. Annual Review of Fluid Mechanics 54, 225–254.

Shalev-Shwartz, S., 2014. Understanding machine learning : from theory to algorithms. Cambridge University Press, Cambridge.

Stokes, G.G., 1880. On the theory of oscillatory waves. Mathematical and Physical Papers 1, 197–229. URL: https://archive.org/details/mathphyspapers01stokrich/page/n213/mode/2up.

The Wamdi Group, 1988. The WAM model – a third generation ocean wave prediction model. Journal of Physical Oceanography 18, 1775–1810.

Toffoli, A., Gramstad, O., Trulsen, K., Monbaliu, J., Bitner-Gergersen, E., Onorato, M., 2010. Evolution of weakly nonlinear random directional waves: laboratory experiments and numerical simulations. Journal of Fluid Mechanics 664, 313–336. doi:10.1017/S002211201000385X.

Tolman, H.L., 1991. A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents. Journal of Physical Oceanography 21, 782–797.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I., 2017. Attention is all you need, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Vautard, R., Ghil, M., 1989. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Physica D: Nonlinear Phenomena 35, 395–424. doi:https://doi.org/10.1016/0167-2789(89)90077-8.

Wang, R., Balachandran, B., 2018. Extreme wave formation in unidirectional sea due to stochastic wave phase dynamics. Physics Letters A 382, 1864–1872.

Welch, P., 1967. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on audio and electroacoustics 15, 70–73.

Whitham, G.B., 2011. Linear and nonlinear waves. John Wiley & Sons.

Yuen, H.C., Lake, B.M., 1980. Instabilities of waves on deep water. Annual Review of Fluid Mechanics 12, 303–334. doi:10.1146/annurev.fl.12.010180.001511.

Zakharov, V.E., 1968. Stability of periodic waves of finite amplitude on the surface of a deep fluid. Journal of Applied Mechanics and Technical Physics 9, 190–194.

Zhao, B., Zheng, K., Duan, W., Ertekin, R., Shao, Y., 2020. Time domain simulation of focused waves by high-level irrotational green–naghdi equations and harmonic polynomial cell method. European Journal of Mechanics - B/Fluids 82, 83–92. doi:https://doi.org/10.1016/j.euromechflu.2020.02.006.

A. Comparison of common wave models with the wave model (1)

Stokes analyzed the Bernoulli equation for irrotational flows (Stokes, 1880). Assuming a small wave steepness, he performed a perturbation expansion which, up to the third order, yields the following wave profile

$$\begin{aligned}\eta(x,t) =& a\cos(kx-\omega t)+\frac{a}{2}(ka)\cos(2(kx-\omega t))\\ &+\frac{3a}{8}(ka)^2\cos(3(kx-\omega t)),\end{aligned} \tag{A.1}$$

where a is an amplitude and k is the wave number (cf. also (Whitham, 2011; Rahman, 1995)). The product ka, also denoted as wave steepness is the small parameter in the Stokes' expansion. The wave model (1) can represent a simple superposition of multiple Stokes waves (A.1) for a fixed observer ($x = const.$). In this setting, the amplitudes u_n and v_n are constant and the higher order terms are either negligible (small ka) or can be represented by including additional harmonics in the wave model (1).

In state-of-the-art operational wave forecast systems, one carefully solves parameterized versions of the action-balance equation (Komen et al., 1996). Therein, waves trains

are modelled as

$$\eta(x,t) = \sum_{n=1}^{N} \left(a_n(\varepsilon\mathbf{x},\varepsilon t) e^{i(\mathbf{k}_n(\varepsilon\mathbf{x},\varepsilon t)\mathbf{x} - \omega_n(\varepsilon\mathbf{x},\varepsilon t)t)} + \bar{a}_n(\varepsilon\mathbf{x},\varepsilon t) e^{-i(\mathbf{k}_n(\varepsilon\mathbf{x},\varepsilon t)\mathbf{x} - \omega_n(\varepsilon\mathbf{x},\varepsilon t)t)} \right), \quad 0 < \varepsilon \ll 1, \tag{A.2}$$

where $\bar{a}_n$ is the complex conjugate of the amplitude a_n. The number of Fourier modes N can be finite or infinite depending on the application. In model (A.2), one allows for spatial and temporal variations of the amplitudes $a_n(\varepsilon\mathbf{x},\varepsilon t)$, the frequencies $\omega_n(\varepsilon\mathbf{x},\varepsilon t)$, and the wave numbers $\mathbf{k}_n(\varepsilon\mathbf{x},\varepsilon t)$. These variations are generally slow (Komen et al., 1996) as indicated by the small parameter ε. Assuming a constant observer $\mathbf{x} = \mathbf{x}_c$ and introducing the mean frequencies $\omega_n := 1/T \int_0^T \omega_n(\varepsilon\mathbf{x}_c,\varepsilon t)dt$, equation (A.2) can be rewritten

$$\begin{aligned}\eta(x,t) &= \sum_{n=1}^{N} \Bigg(\underbrace{a_n(\varepsilon\mathbf{x}_c,\varepsilon t) e^{i(\mathbf{k}_n(\varepsilon\mathbf{x}_c,\varepsilon t)\mathbf{x}_c + (\omega_n - \omega_n(\varepsilon\mathbf{x}_c,\varepsilon t))t)}}_{=:\upsilon_n(\varepsilon t)} e^{-i\omega_n t} \\ &\quad + \underbrace{\bar{a}_n(\varepsilon\mathbf{x}_c,\varepsilon t) e^{-i(\mathbf{k}_n(\varepsilon\mathbf{x}_c,\varepsilon t)\mathbf{x}_c + (\omega_n - \omega_n(\varepsilon\mathbf{x},\varepsilon t))t)}}_{\bar{\upsilon}_n(\varepsilon t)} e^{i\omega_n t} \Bigg) \\ &= \sum_{n=1}^{N} \left(\upsilon_n(\varepsilon t) e^{-i\omega_n t} + \bar{\upsilon}_n(\varepsilon t) e^{i\omega_n t} \right),\end{aligned} \tag{A.3}$$

which is the complex equivalent to the wave model (1). While the time variations of the amplitudes are driven by physically motivated source terms, black-box, data-driven methods are utilized in the proposed method.

The nonlinear Schrödinger equation arises in the context of ocean waves as a governing equation of an amplitude of a modulated, single wave train (Kharif and Pelinovsky, 2003). The results are based on a perturbation approach assuming weak nonlinearities and, additionally, a single dominant frequency as well as slowly varying amplitudes (Zakharov, 1968). In essence, the nonlinear Schrödinger equation governs the amplitude $u(\varepsilon x,\varepsilon t)$ of the wave train

$$\eta(x,t) := u(\varepsilon x,\varepsilon t) e^{i(kx-\omega t)} + u(\varepsilon x,\varepsilon t) e^{-i(kx-\omega t)}, \quad 0 < \varepsilon \ll 1. \tag{A.4}$$

Similarly, the higher order extensions of the nonlinear Schrödinger equation such as those discussed in reference (Dysthe, 1979) rely on the same assumption. The ansatz (A.4) is equivalent to the wave model (1) with a single wave train ($N = 1$).