A probabilistic approach for the quantification of prediction error in deterministic phase-resolved wave forecasting

Fabio Fucile[1,2] (fabio.fucile@phd.units.it), Gabriele Bulian[1] (gbulian@units.it), Claudio Lugni[2,3] (claudio.lugni@cnr.it)

[1] Department of Engineering and Architecture, University of Trieste, Trieste, Italy
[2] CNR-INSEAN, Marine Technology Institute, Rome, Italy
[3] AMOS/NTNU, Centre for Autonomous Marine Operations and Systems, Trondheim, Norway

Highlights

- Deterministic sea wave predictions through phase-resolved linear wave models are considered
- A semi-analytical Linear Estimator of Prediction Error (LEPrE) is presented
- The approach allows accounting for different fitting and propagation strategies
- The exact spectral shape for mono/multi-directional waves is taken into account, as well as possible measurement error
- Example applications and comparisons with Monte Carlo simulations are reported

Abstract

This paper presents a semi-analytical methodology for the determination of prediction error statistics in deterministic sea wave predictions (DSWP), based on linear wave models. The underlying wave elevation is modelled as a Gaussian stochastic process and the coefficients of the wave propagation model are assumed to be determined by linear fitting on available measurements in time and/or space. The possible data contamination due to measurement error is also explicitly considered. The resulting approach eventually provides a Linear Estimator of Prediction Error (LEPrE) in time and space, in terms of prediction error standard deviation, given the fitting procedure and the sea spectrum. The presented approach allows supplementing deterministic predictions based on phase-resolved linear wave models with a sound prediction error measure, and allows defining the concept of "Predictability Region" in a consistent probabilistic framework. Example applications are reported, both for long-crested and short-crested waves, with verification through Monte Carlo simulations. Single point wave gauge/wave buoy measurements as well as spatial extent wave radar measurements have been considered as simulated examples. The developed methodology is also compared with existing approaches highlighting and discussing both the differences and the interesting qualitative commonalities.

Keywords: *deterministic sea wave prediction (DSWP); predictability region; phase-resolved wave models; prediction error; wave radar.*

1 Introduction

The nowadays interest about deterministic wave propagation models based on the marine wave radar technology is encouraged by the outlook of possible applications to real time waves and ship motion forecasting. The development of early warnings, guidance and decision-support systems based on deterministic prediction procedures could possibly have a positive impact for the safety and operability at sea. A main asset in this kind of emerging short-term forecasting technology (with temporal horizon of the order of minutes, and spatial horizon of the order of hundreds meters) is the marine wave radar. In fact, the marine wave radar has been shown to be potentially capable of scanning the sea surface and retrieving the instantaneous images of the nearby wave field in a wide spatial range (Dankert and Rosenthal 2004; Nieto Borge et al., 2004; Serafino et al. 2011; Naaijen and Wijaya 2014). It is however to be noted that challenges in modelling of the associated basic electromagnetic backscattering mechanism still require evolutions of this technology to obtain very accurate measurements, at least when used to feed DSWP model. The LIDAR technology has also been explored for the measurement of wave elevation (Belmont et al. 2007; Grilli et al., 2011; Nouguier et al. 2014). In principle LIDAR could be considered as an alternative to wave radar. However, presently available research on corresponding local wave elevation measurements (Belmont et al. 2007; Grilli et al., 2011; Nouguier et al. 2014) indicate a yet too limited spatial extent of the measurement region. As a result, the application of such technology in case of deterministic predictions in realistic sea states characterised by long waves in open sea becomes difficult. Nevertheless, an extension of the LIDAR wave measurement range could allow this technology to become a possible alternative to wave radar.

Once wave elevation data are (assumed to be) available from a suitable wave measurements system, a phase-resolved propagation procedure can then be applied to perform a deterministic forecasting. The procedure is required to be fast if the use is intended for real-time applications. Furthermore it is required to have a prediction time horizon compatible with the operational needs. Particularly due to computational speed requirements, linear deterministic wave propagation models are often preferred (Hilmer and Thornhill, 2015), particularly for intended uses in real-time applications, and different aspects of their implementation have been investigated in the past (Belmont et al., 2006, 2014; Blondel-Couprie and Naaijen, 2012; Connell et al., 2015; Naaijen and Blondel-Couprie, 2012; Naaijen et al., 2014; Naaijen and Huijsmans, 2008). Linear Deterministic Sea Wave Prediction (DSWP) procedures usually consist of two main steps. First, in the *fitting step (FS)*, the wave elevation data are analysed in the measurement domain by means of Fourier decomposition techniques, either based on the DFT (Morris et al. 1998; Naaijen and Blondel-Couprie, 2012; Vettor, 2010) or on a least-squares approach (Connell et al., 2015; Naaijen et al., 2009; Vettor, 2010). Afterwards a linear propagation model is defined in the *propagation step (PS)*. The extensive use of the FFT, both in the FS and in the PS steps, is deeply discussed for short-crested sea applications by Blondel-Couprie and Naaijen (2012) and Naaijen and Blondel-Couprie (2012). Different implementations of linear fitting and propagation procedures, still based on a Fourier analysis, are presented by Abusedra and Belmont (2011) and Belmont et al. (2006).

Nonlinear phase-resolved wave propagation models have been proposed by several authors (e.g. Blondel et al., 2010; Blondel-Couprie et al., 2013; Nouguier et al. 2014; Wu, 2004; Zhang et al., 1999a,b; Yoon et al., 2016). The main issue of these techniques is represented by the costly fitting/initialization step. In fact, in general, the measured wave elevation data has to be pre-processed before being actually available for the propagation model. The pre-processing step, and its consequent complexity, depend mainly on the nature of the available measurement and on the nonlinear model considered. The reconstruction of the initial conditions of the nonlinear wave model can require iterative procedures on the measured data (Zhang et al., 1999a,b) or data assimilation procedures as in Wu (2004), Blondel et al. (2010), Blondel-Couprie et al. (2013) or

Yoon et al. (2016). In particular, the variational data assimilation procedure proposed by Wu (2004) and Blondel et al. (2010) can be considered as an optimization problem for the initial conditions of the model, with cost function defined as a suitable measure of the distance between the wave elevation given by the nonlinear model and the measured wave elevation data. Also Nouguier et al. (2014) used the minimization of a cost function, representing the average squared difference between measured wave elevation and wave model to be propagated, for the identification of the free parameters of the wave propagation model. Actually, such approach, in addition of being used for forecasting purposes, served also the purpose of wave elevation reconstruction procedure for LIDAR measurements (Grilli et al., 2011; Nouguier et al., 2014).

A common issue to all DSWP methods is related to the need of providing an estimation of the region where the deterministic prediction can be considered to be sufficiently reliable for the intended purposes. In fact, any DSWP procedure is inevitably affected by prediction errors with respect to the true wave elevation, which indirectly define the limits of application of this kind of procedures. One source of prediction error is the inherent limitation of the assumed propagation model which does not exactly represent the underlying wave elevation field. As a result, even when a propagation model perfectly fits the true wave elevation at some discrete sampling points in time and/or space, the predicted (or reconstructed) wave elevation at different locations in time and/or space will differ from the true one. In addition, in real applications, the wave measurements themselves are affected by measurement errors, which bring into the problem an additional source of uncertainty, an aspect which is often overlooked. This means that a key aspect of DSWP should be the capability of providing not only an estimation of the predicted wave elevation, but also some information regarding the prediction error. However, although the assessment of the prediction error is crucial for a consistent deterministic wave prediction, the problem is rarely addressed specifically. In this context, the idea of using brute force tools such as massive Monte Carlo simulations to estimate the expected prediction errors statistics is, in general, practically unfeasible due to the time consuming computations that eventually will go to detriment of a direct use in-real time applications. Therefore concepts of faster and more direct application are required.

The most widespread concept related to the prediction performances of DSWP approaches is the so-called “Predictability Region”. The Predictability Region is considered to be the region of space and time where it is considered “possible” to predict the wave elevation, ideally without errors. It is therefore, originally, a binary concept, which split the time/space domain in a region where the prediction “is possible”, and a region where the prediction “is not possible”. In the past, a matter of discussion has been whether to use the group velocity or the phase velocity of the waves for the identification of the Predictability Region (e.g. Abusedra & Belmont 2011; Edgar et al., 2000; Morris et al., 1998; Naaijen et al., 2014; Wu, 2004). For example, based on wave propagation considerations, Morris et al. (1998) selected the wave phase velocity for the determination of the region where the propagation of information, and the corresponding deterministic prediction, can be considered possible (see also Edgar et al., 2000), in case of long-crested seas. Instead, Wu (2004) used the wave group speed for the determination of the Predictability Region and further extended the concept to the case of short-crested seas. According to Wu (2004) (see also Naaijen et al. (2014)), the Predictability Region is defined using the group velocity of the fastest and slowest wave components of the considered sea spectrum. However, Abusedra and Belmont (2011) have shown that the use of wave group velocity cannot be completely justified, and they also challenged those previous justifications for such use which were based on asymptotic stationary phase approximation. The concept of Predictability Region has been further developed by Wu (2004) and Naaijen et al. (2014) with the introduction of the “Predictability Indicator”: a measure of the prediction capability at a generic point in time and space, given the sea spectrum. The Predictability Indicator takes into account the actual shape of the wave spectrum, and this represents an advance with respect to the standard Predictability Region, which, instead, accounts only for the (assumed)

lowest and highest frequency limits of the spectrum. Naaijen et al. (2014) verified the Predictability Indicator method to be qualitatively consistent with Monte Carlo simulations for which every realization of the sea states is ideally measured and then propagated to create a reference statistics for ensemble analysis. The encouraging results showed by Naaijen et al. (2014) and the simple formulation of the method makes the Predictability Indicator an interesting tool for a more advanced, and potentially more precise, definition of Predictability Region compared to the original concept. However, the Predictability Indicator still lacks a consistent statistical background theory able to provide a clear probabilistic interpretation of the obtained quantitative values.

It is then useful to make a step forward in the definition of the concept of Predictability Region, with a view to more soundly account for the prediction error from a probabilistic perspective. To this end, a theoretical approach for providing a consistent probabilistic measure of prediction error for deterministic phase-resolved linear wave prediction models, is herein presented. The approach is based on the description of the sea as a Gaussian stationary stochastic process. The features of the fitting procedure and of the prediction model are naturally embedded in the formulation. Furthermore, the formulation takes into account, in an analytic way, the actual shape of the spectrum for long-crested and short-crested waves. On top of this, the proposed framework also allows taking consistently into account the possible presence of additional measurement noise.

The paper is structured as follows. First, the theoretical background is presented starting from the definition of the fitting model and then providing the definition of the prediction error as a stochastic process. The assessment of the ensemble variance of the error process leads to the natural definition of a Linear Estimator of Prediction Error (LEPrE), which accounts also for the possible contribution of measurement noise. A section then follows, containing three different simulated verification test cases, considering long-crested and short-crested sea states, to show how the LEPrE can be used in identifying the level of prediction error. Reported results from the application of LEPrE are verified along with corresponding sets of Monte Carlo simulations. Eventually, some comparison with the classical (binary) Predictability Region and the Predictability Indicator by Wu (2004) and Naaijen et al. (2014) are also provided, in order to show how the presented LEPrE compares with existing approaches. In order to provide consistent comparisons between simulations and LEPrE predictions for verification purposes, simulated wave fields in all synthetic applications are based on the hypothesis of linear waves. The first application is conceived to present an ideal, though realistic, laboratory experiment where the measurement device, used for the propagation step, is affected by measurement noise. The second application shows details of a comparison between LEPrE approach, Monte Carlo simulations, (binary) Predictability Region and Prediction Error Indicator for a simplified bimodal spectrum made of two separated band limited white noise regions. In the third and last application the LEPrE methodology is applied to a more realistic scenario where an offshore structure is considered and the wave elevation, used for the propagation model, is assumed to be measured by means of a wave radar device. Following the verifying applications section, a specific section is dedicated to a discussion regarding nonlinear effects, which are not addressed by the LEPrE approach, also in view of possible future studies on this topic. The paper eventually provides some concluding remarks.

2 Theoretical background

As anticipated, the scope of this work is to provide a methodology for addressing the so-called Predictability Region in phase-resolved wave propagation models from a consistent probabilistic perspective. The aim is to provide an approach representing a step forward with respect to the present common practice of identifying a conventionally defined Predictability Region by means of the loosely defined “slowest” and “fastest” components of the sea spectrum (see, e.g. Blondel-

Couprie and Naaijen, 2012; Blondel et al., 2010; Naaijen and Huijsmans, 2008; Wu, 2004), and with respect to the more advanced, yet partially semi-empirical, Predictability Indicator (Wu, 2004; Naaijen et al., 2014)

To this end, this section presents a methodology for the semi-analytical estimation of prediction error when a model for the propagation of the wave elevation field is fitted on a set of data corresponding to measured wave elevation in space and/or time. The framework presented herein is general and applicable to both long-crested and short-crested waves, with and without presence of measurement noise, and for generic linear fitting procedures. More specific applications are reported in the application section. Preliminary results for the long-crested (1D) case in absence of measurement noise and using a Fourier fitting have been reported by Fucile et al. (2016b).

2.1 Linear estimator of prediction error (LEPrE) and definition of predictability region

It is firstly assumed that the true wave elevation field $\eta(\underline{x},t)$ is a stationary Gaussian process for which a single generic realization can be represented as follows:

$$\begin{cases} \eta(\underline{x},t) = \sum_{i=1}^{N_\eta} a_i \cos(\underline{k}_i \cdot \underline{x} - \omega_i t) + b_i \sin(\underline{k}_i \cdot \underline{x} - \omega_i t) = \underline{p}_\eta(\underline{x},t)\,\underline{\alpha} \\ \underline{p}_\eta(\underline{x},t) = \left(\cos(\underline{k}_1 \cdot \underline{x} - \omega_1 t), \sin(\underline{k}_1 \cdot \underline{x} - \omega_1 t), ..., \cos\left(\underline{k}_{N_\eta} \cdot \underline{x} - \omega_{N_\eta} t\right), \sin\left(\underline{k}_{N_\eta} \cdot \underline{x} - \omega_{N_\eta} t\right)\right) \\ \underline{\alpha} = \left(a_1, b_1, ..., a_{N_\eta}, b_{N_\eta}\right)^T \\ \underline{p}_\eta(\underline{x},t) \in \mathbb{R}^{1,2N_\eta}; \underline{\alpha} \in \mathbb{R}^{2N_\eta,1} \end{cases} \quad (1)$$

In the expression for the wave elevation $\eta(\underline{x},t)$, $\underline{k}_i = \left(k_x, k_y\right)^T$ is the wave number vector for the i-th harmonic component, which is linked to the wave frequency ω_i by a suitable dispersion relation. In the case of linear waves and infinite water depth $\omega_i = \sqrt{g\left\|\underline{k}_i\right\|}$. The space and time dependent vector $\underline{p}_\eta(\underline{x},t)$ can be referred as the "propagator vector" for η (see also Connell et al., 2015) since it propagates the wave elevation field in time and space if the coefficients vector $\underline{\alpha}$ is known. In (1) a finite, but sufficiently large, number of harmonic components, N_η, is considered, while the actual stochastic process is in principle recovered in the limit $N_\eta \to \infty$. It is also assumed that $\eta(\underline{x},t)$ can be measured at certain points in space ($\underline{x} \in \mathbb{R}^{2,1}$) and/or time ($t \in \mathbb{R}$), and that the measured wave elevation $\eta_M(\underline{x},t)$ is associated with a certain measurement error δ_η, i.e.:

$$\eta_M(\underline{x},t) = \eta(\underline{x},t) + \delta_\eta(\underline{x},t) \quad (2)$$

The case of measurements without error represents a special case of (2) where $\delta_\eta(\underline{x},t) = 0$ for all positions and time instants. Herein the wording "measured" is used for sake of simplicity, although, in general, the wave elevation can be either measured (e.g. at a wave gauge or wave buoy) or estimated (e.g. from the inversion of wave radar images, Connell et al., 2015; Dankert and Rosenthal, 2004; Fucile et al., 2016a; Nieto Borge et al., 2004; Wijaya et al., 2015). It is now

assumed that the true wave elevation is fitted through a phase-resolved wave model $\zeta(\underline{x},t)$ having the following expression, similar to that of $\eta(\underline{x},t)$:

$$\begin{cases} \zeta(\underline{x},t)=\sum_{h=1}^{N_\zeta}\tilde{a}_h\cos\left(\underline{\tilde{k}}_h\cdot\underline{x}-\tilde{\omega}_h t\right)+\tilde{b}_h\sin\left(\underline{\tilde{k}}_h\cdot\underline{x}-\tilde{\omega}_h t\right)=\underline{p}_\zeta(\underline{x},t)\,\underline{\beta} \\ \underline{p}_\zeta(\underline{x},t)=\left(\cos\left(\underline{\tilde{k}}_1\cdot\underline{x}-\tilde{\omega}_1 t\right),\sin\left(\underline{\tilde{k}}_1\cdot\underline{x}-\tilde{\omega}_1 t\right),...,\cos\left(\underline{\tilde{k}}_{N_\zeta}\cdot\underline{x}-\tilde{\omega}_{N_\zeta} t\right),\sin\left(\underline{k}_{N_\zeta}\cdot\underline{x}-\omega_{N_\zeta} t\right)\right) \\ \underline{\beta}=\left(\tilde{a}_1,\tilde{b}_1,...,\tilde{a}_{N_\zeta},\tilde{b}_{N_\zeta}\right)^T \\ \underline{p}_\zeta(\underline{x},t)\in\mathbb{R}^{1,2N_\zeta};\underline{\beta}\in\mathbb{R}^{2N_\zeta,1} \end{cases} \tag{3}$$

where, in general, the harmonic components of $\zeta(\underline{x},t)$ differ from those of $\eta(\underline{x},t)$ and, in addition, typically $N_\zeta << N_\eta$. The next assumption is that, given a set of N_M measurements $\eta_M\left(\underline{x}_j,t_j\right)\ j=1,...,N_M$ at different points in space and/or time, the coefficients vector $\underline{\beta}$ can be determined by a linear transformation of the available measurements through an appropriate matrix $\underline{\underline{T}}_M$, i.e.:

$$\begin{cases} \underline{\beta}=\underline{\underline{T}}_M\,\underline{\eta}_M \\ \underline{\underline{T}}_M\in\mathbb{R}^{2N_\zeta,N_M} \\ \underline{\eta}_M=\left(\eta_M\left(\underline{x}_1,t_1\right),...,\eta_M\left(\underline{x}_{N_M},t_{N_M}\right)\right)^T\in\mathbb{R}^{N_M,1} \end{cases} \tag{4}$$

Such an assumption covers at least three important cases. The first case is the classical Fourier analysis on rectangular grids, in which case $\underline{\underline{T}}_M$ is the DFT matrix (Golub and Van Loan, 2013). The second one is associated with a direct least-squares fitting process (see Connell et al., 2015; Vettor, 2010). In such case the fitting is firstly setup as follows:

$$\begin{cases} \underline{\underline{P}}_{\zeta,M}\,\underline{\beta}=\underline{\eta}_M \\ \text{with } \underline{\underline{P}}_{\zeta,M}=\begin{bmatrix}\underline{p}_\zeta\left(\underline{x}_1,t_1\right)\\ \vdots \\ \underline{p}_\zeta\left(\underline{x}_{N_M},t_{N_M}\right)\end{bmatrix}\in\mathbb{R}^{N_M,2N_\zeta} \\ N_M\geq 2N_\zeta \end{cases} \tag{5}$$

where $\underline{\underline{P}}_{\zeta,M}$ can be referred to as the propagator matrix (Connell et al., 2015) at the measurement points in time and space, and $\underline{\underline{T}}_M$ is the Moore-Penrose pseudoinverse of $\underline{\underline{P}}_{\zeta,M}$, i.e.

$$\underline{\underline{T}}_M=\underline{\underline{P}}^{+}_{\zeta,M}=\left(\underline{\underline{P}}^{T}_{\zeta,M}\,\underline{\underline{P}}_{\zeta,M}\right)^{-1}\underline{\underline{P}}^{T}_{\zeta,M} \tag{6}$$

The third case covered by the assumption (4) is the regularised least-square fitting using Tikhonov regularisation (Hansen, 1998; Vogel, 2002), where a solution $\underline{\beta}$ is sought such to minimize the following objective function:

$$\left\| \underline{\underline{P}}_{\zeta,M}\,\underline{\beta} - \underline{\eta}_M \right\|^2 + \left\| \underline{\underline{\Lambda}}\,\underline{\beta} \right\|^2 \tag{7}$$

In (7), $\underline{\underline{\Lambda}}$ is an $2N_\zeta \times 2N_\zeta$ regularisation matrix, which is often chosen as $\underline{\underline{\Lambda}} = \lambda \underline{\underline{I}}_{2N_\zeta,2N_\zeta}$, being $\underline{\underline{I}}_{2N_\zeta,2N_\zeta}$ the $2N_\zeta \times 2N_\zeta$ identity matrix and λ the regularisation parameter. In the general case of Tikhonov regularization, the matrix $\underline{\underline{T}}_M$ then becomes:

$$\underline{\underline{T}}_M = \left(\underline{\underline{P}}_{\zeta,M}^T \underline{\underline{P}}_{\zeta,M} + \underline{\underline{\Lambda}}^T \underline{\underline{\Lambda}} \right)^{-1} \underline{\underline{P}}_{\zeta,M}^T \tag{8}$$

It is worth noticing that the classical least-squares case is a special case of the Tikhonov regularisation, while the classical Fourier case can also be seen as a special case of both the other two.

Using the model (3) and the assumption (4), the fitted model can be evaluated at any point in space and time as follows:

$$\zeta_{F,\delta}\left(\underline{x},t\right) = \underline{p}_\zeta\left(\underline{x},t\right)\underline{\beta} = \underline{p}_\zeta\left(\underline{x},t\right)\underline{\underline{T}}_M\,\underline{\eta}_M \tag{9}$$

The subscript "F,δ" indicates that the model is using coefficients which have been determined through fitting taking into account the presence of measurement error δ_η. It is now possible to determine the error between the fitted model and the true wave elevation $\eta\left(\underline{x},t\right)$ as follows:

$$\varepsilon_\delta\left(\underline{x},t\right) = \zeta_{F,\delta}\left(\underline{x},t\right) - \eta\left(\underline{x},t\right) = \underline{p}_\zeta\left(\underline{x},t\right)\underline{\underline{T}}_M\,\underline{\eta}_M - \underline{p}_\eta\left(\underline{x},t\right)\underline{\alpha} \tag{10}$$

Furthermore, using (1) and (2), the vector of measured wave elevation can be written as:

$$\begin{cases} \underline{\eta}_M = \underline{\eta} + \underline{\delta}_{\eta,M} = \underline{\underline{P}}_{\eta,M}\,\underline{\alpha} + \underline{\delta}_{\eta,M} \\ \underline{\underline{P}}_{\eta,M} = \begin{bmatrix} \underline{p}_\eta\left(\underline{x}_1,t_1\right) \\ \vdots \\ \underline{p}_\eta\left(\underline{x}_{N_M},t_{N_M}\right) \end{bmatrix} \in \mathbb{R}^{N_M,2N_\eta} \\ \underline{\delta}_{\eta,M} = \left(\delta_\eta\left(\underline{x}_1,t_1\right),\ldots,\delta_\eta\left(\underline{x}_{N_M},t_{N_M}\right)\right)^T \in \mathbb{R}^{N_M,1} \end{cases} \tag{11}$$

Combining (10) and (11), the error $\varepsilon_\delta\left(\underline{x},t\right)$ can therefore be written as

$$\begin{cases} \varepsilon_\delta\left(\underline{x},t\right)=\underline{q}\left(\underline{x},t\right)\underline{\alpha}+\underline{n}\left(\underline{x},t\right)\underline{\delta}_{\eta,M} \\ \text{with} \\ \underline{q}\left(\underline{x},t\right)=\underline{p}_\zeta\left(\underline{x},t\right)\underline{\underline{T}}_M\underline{\underline{P}}_{\eta,M}-\underline{p}_\eta\left(\underline{x},t\right) \\ \underline{n}\left(\underline{x},t\right)=\underline{p}_\zeta\left(\underline{x},t\right)\underline{\underline{T}}_M \\ \underline{q}\left(\underline{x},t\right)\in\mathbb{R}^{1,2N_\eta}\ ;\ \underline{n}\left(\underline{x},t\right)\in\mathbb{R}^{1,N_M} \end{cases} \tag{12}$$

From the first equation in (12), it can be seen that the error $\varepsilon_\delta\left(\underline{x},t\right)$ is due to two sources. The first term, $\underline{q}\left(\underline{x},t\right)\underline{\alpha}$, represents the error due to the fact that the fitted model, in general, differs from the actual process. The second term, $\underline{n}\left(\underline{x},t\right)\underline{\delta}_{\eta,M}$, represents the propagation of the measurement error through the fitted model. Both error sources also embed the effect of the fitting procedure.

It is now worth noticing that the error $\varepsilon_\delta\left(\underline{x},t\right)$ as obtained in (12) is a linear function of the amplitudes of the harmonic components of the true wave elevation $\eta\left(\underline{x},t\right)$. Furthermore, taking into account the fact that $\eta\left(\underline{x},t\right)$ is, actually, a stochastic process and the fact that $\underline{\delta}_{\eta,M}$ is a random vector, the error $\varepsilon_\delta\left(\underline{x},t\right)$ can be interpreted as a stochastic process. In all the following considerations, the set of N_M measurement points is assumed to be deterministically fixed in the ensemble domain, i.e., across multiple realisations. Furthermore, it is assumed that other possible fitting parameters (e.g. the regularization matrix $\underline{\underline{\Lambda}}$ in (8)) are also deterministically fixed. These two latter assumptions mean that the matrices $\underline{\underline{T}}_M$ and $\underline{\underline{P}}_{\eta,M}$ are deterministic matrices. As a result, such assumptions allow to consider $q\left(\underline{x},t\right)$ and $\underline{n}\left(\underline{x},t\right)$ in (12) as deterministic functions of space and time.

By using the assumption that $\eta\left(\underline{x},t\right)$ is a (discretised) linear Gaussian process, similarly to the one dimensional case (Tucker et al., 1984), the amplitudes a_i and b_i of the harmonic components are considered as zero mean independent Gaussian variables. They are linked to the (single side) directional wave energy spectrum $S_\eta\left(\underline{k}_i\right)$ as follows:

$$\begin{cases} a_i\sim N\left(0,\sigma_i^2\right)\ ;\ b_i\sim N\left(0,\sigma_i^2\right)\ ;\ \sigma_i^2=S_\eta\left(\underline{k}_i\right)\Delta k_{x,i}\Delta k_{y,i} \\ COV\left\{a_i,a_j\right\}=\begin{cases}0 & i\neq j\\ \sigma_i^2 & i=j\end{cases}\ ;\ COV\left\{b_i,b_j\right\}=\begin{cases}0 & i\neq j\\ \sigma_i^2 & i=j\end{cases} \\ COV\left\{a_i,b_j\right\}=0\ \ \forall i,j \end{cases} \tag{13}$$

where $N\left(0,\sigma_i^2\right)$ indicates a normal distribution with zero mean and variance σ_i^2, $\Delta k_{x,i}$ and $\Delta k_{y,i}$ are wave number intervals associated with the assumed discretization of the spectrum $S_\eta\left(\underline{k}_i\right)$ in N_η wave components, and $COV\left\{.,.\right\}$ indicates the covariance operator. According to (13) the wave amplitudes vector $\underline{\alpha}$ is a zero mean Gaussian random vector with diagonal covariance matrix, that is:

$$\begin{cases} \underline{\mu}_{\underline{\alpha}} = E\left\{\underline{\alpha}\right\} = \underline{0} \\ \underline{\underline{C}}_{\underline{\alpha},\underline{\alpha}} = E\left\{\left(\underline{\alpha} - \underline{\mu}_{\underline{\alpha}}\right)\left(\underline{\alpha} - \underline{\mu}_{\underline{\alpha}}\right)^T\right\} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma_1^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{N_\eta}^2 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_{N_\eta}^2 \end{bmatrix} = \\ = diag\left(\sigma_1^2, \sigma_1^2, ..., \sigma_{N_\eta}^2, \sigma_{N_\eta}^2\right) \end{cases} \tag{14}$$

where $E\{.\}$ is the expected value operator. In addition, it is assumed that the measurement error $\delta_\eta\left(\underline{x},t\right)$ is a zero mean Gaussian process and that it is independent of η. As a consequence, given the specific set of N_M measurement points in space and time, the vector $\underline{\delta}_{\eta,M}$ is a zero mean Gaussian vector with a given covariance matrix depending on the assumed measurement error characteristics, i.e.:

$$\begin{cases} \underline{\mu}_{\underline{\delta}_{\eta,M}} = E\left\{\underline{\delta}_{\eta,M}\right\} = \underline{0} \\ \underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}} = E\left\{\left(\underline{\delta}_{\eta,M} - \underline{\mu}_{\underline{\delta}_{\eta,M}}\right)\left(\underline{\delta}_{\eta,M} - \underline{\mu}_{\underline{\delta}_{\eta,M}}\right)^T\right\} \end{cases} \tag{15}$$

By using (12), (14) and (15), the assumption that $\delta_\eta\left(\underline{x},t\right)$ is independent of η, and recalling that, from the considered assumptions, $\underline{q}\left(\underline{x},t\right)$ and $\underline{n}\left(\underline{x},t\right)$ are deterministic vector functions, it is now possible to provide a full probabilistic characterisation of the prediction error $\varepsilon_\delta\left(\underline{x},t\right)$. In fact, from (12), it follows that $\varepsilon_\delta\left(\underline{x},t\right)$ is a Gaussian process, since it is a linear combination of random Gaussian vectors. Furthermore, the mean and the variance of $\varepsilon_\delta\left(\underline{x},t\right)$ can be determined as follows:

$$\begin{cases} E\left\{\varepsilon_\delta\left(\underline{x},t\right)\right\} = 0 \\ Var\left\{\varepsilon_\delta\left(\underline{x},t\right)\right\} = \sigma_{\varepsilon_\delta}^2\left(\underline{x},t\right) = \underline{q}\left(\underline{x},t\right)\underline{\underline{C}}_{\underline{\alpha},\underline{\alpha}}\underline{q}^T\left(\underline{x},t\right) + \underline{n}\left(\underline{x},t\right)\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}\underline{n}^T\left(\underline{x},t\right) \end{cases} \tag{16}$$

Where $E\{.\}$ is the expected value operator, $Var\{.\}$ is the variance operator, and $\sigma_{\varepsilon_\delta}^2\left(\underline{x},t\right)$ is the variance of the prediction error at $\left(\underline{x},t\right)$. From the considered assumptions, it follows that the prediction error has zero mean. In addition, similarly to what was noticed when commenting (12), it can be noticed that also in (16) the variance of the prediction error is composed of two terms: the first term is associated with the inherent difference between the fitted model and the true process, while the second term represents the effect of the measurement error.

If it is assumed that the measurement errors δ_η at two different points in time and space are independent (actually it is sufficient they are uncorrelated), then the covariance matrix $\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}$ becomes diagonal:

$$\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}} = \begin{bmatrix} \sigma^2_{\delta_\eta}\left(\underline{x}_1,t_1\right) & 0 & \cdots & 0 \\ 0 & \sigma^2_{\delta_\eta}\left(\underline{x}_2,t_2\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{\delta_\eta}\left(\underline{x}_{N_M},t_{N_M}\right) \end{bmatrix} \tag{17}$$

The generic term $\sigma^2_{\delta_\eta}\left(\underline{x}_j,t_j\right)$, with $j=1,...,N_M$, represents the measurement error variance at point $\underline{x}_j$ in space and at time t_j. If, in addition, it is assumed that the measurement error is a uniform noise with variance $\sigma^2_{\delta_\eta}$, then the covariance matrix $\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}$ simplifies to:

$$\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}} = \sigma^2_{\delta_\eta}\,\underline{\underline{I}}_{N_M,N_M} \tag{18}$$

Making use of the fact that the covariance matrix $\underline{\underline{C}}_{\underline{\alpha},\underline{\alpha}}$ is diagonal, and if it is assumed that $\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}$ is also diagonal, and by using (13), the expression for the variance of the prediction error can be simplified to:

$$\begin{aligned} &\sigma^2_{\varepsilon_\delta}\left(\underline{x},t\right) = \sum_{i=1}^{N_\eta}\left(q^2_{2i-1}\left(\underline{x},t\right)+q^2_{2i}\left(\underline{x},t\right)\right)\sigma^2_i + \sum_{j=1}^{N_M} n^2_j\left(\underline{x},t\right)\sigma^2_{\delta_\eta}\left(\underline{x}_j,t_j\right) = \\ &= \sum_{i=1}^{N_\eta}\left(q^2_{2i-1}\left(\underline{x},t\right)+q^2_{2i}\left(\underline{x},t\right)\right)S_\eta\left(\underline{k}_i\right)\Delta k_{x,i}\Delta k_{y,i} + \sum_{j=1}^{N_M} n^2_j\left(\underline{x},t\right)\sigma^2_{\delta_\eta}\left(\underline{x}_j,t_j\right) \end{aligned} \tag{19}$$

Each term in the first summation in (19) represents the contribution to the prediction error variance at $\left(\underline{x},t\right)$ coming from the spectral energy around the wave number vector $\underline{k}_i$. Each term in the second summation in (19) represents, instead, the contribution to the prediction error variance at $\left(\underline{x},t\right)$ coming from the measurement error at the measurement point $\left(\underline{x}_j,t_j\right)$. In the limit the number of harmonics assumed in the model of $\eta\left(\underline{x},t\right)$ increases towards infinity ($N_\eta \to \infty$), the first term in (19) tends to an integral involving the spectrum of η, and the variance of the prediction error can be rewritten as:

$$\sigma^2_{\varepsilon_\delta}\left(\underline{x},t\right) = \iint\left(f^2_1\left(\underline{k},\underline{x},t\right)+f^2_2\left(\underline{k},\underline{x},t\right)\right)S_\eta\left(\underline{k}\right)dk_x dk_y + \sum_{j=1}^{N_M} n^2_j\left(\underline{x},t\right)\sigma^2_{\delta_\eta}\left(\underline{x}_j,t_j\right) \tag{20}$$

where $f^2_1\left(\underline{k},\underline{x},t\right)$ and $f^2_2\left(\underline{k},\underline{x},t\right)$ are functions depending on the fitting procedure. It is important, at this point, to underline that the single side energy spectrum $S_\eta\left(\underline{k}\right)$ appearing in (19) and (20) is the spectrum of the true wave elevation η and it must not be confused with the spectrum, assuming it could be consistently defined, of the measured wave elevation. In fact, the measured wave elevation is contaminated, in general, by the effect of the measurement error process.

For a realistic measurement, the assumption of diagonal covariance matrix $\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}$ may be considered as a reasonable (or at least convenient) approximation when measurement points are

sufficiently far each other in space and time. However, in the general case, the matrix $\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}$ cannot be assumed to be diagonal. In fact, it is expectable that there may be a correlation in time and space between measurement error at different points, particularly when such measurement points are close in space or time. In essence, the covariance matrix $\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}$ provides a probabilistic characterization of the measurement error for the assumed measurement system. Although the specification of such a matrix is expected to be a complex problem, in the general case when $\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}$ is not diagonal equations (19) and (20) keep their more general form, which is based only on the fact that $\underline{\underline{C}}_{\underline{\alpha},\underline{\alpha}}$ is always diagonal, i.e.

$$\begin{cases} \sigma_{\varepsilon_\delta}^2\left(\underline{x},t\right) = \sum_{i=1}^{N_\eta}\left(q_{2i-1}^2\left(\underline{x},t\right)+q_{2i}^2\left(\underline{x},t\right)\right)S_\eta\left(\underline{k}_i\right)\Delta k_{x,i}\Delta k_{y,i} + \underline{n}\left(\underline{x},t\right)\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}\underline{n}^T\left(\underline{x},t\right) \\ \sigma_{\varepsilon_\delta}^2\left(\underline{x},t\right) = \iint\left(f_1^2\left(\underline{k},\underline{x},t\right)+f_2^2\left(\underline{k},\underline{x},t\right)\right)S_\eta\left(\underline{k}\right)dk_x dk_y + \underline{n}\left(\underline{x},t\right)\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}\underline{n}^T\left(\underline{x},t\right) \end{cases} \tag{21}$$

A closed analytical formulation for the functions $f_1^2\left(\underline{k},\underline{x},t\right)$ and $f_2^2\left(\underline{k},\underline{x},t\right)$ is difficult to derive in the general case, for it depends on the fitting model as it has to embed the solution step in (4). However, some specific consideration can be given to the case where the fitting model is derived from a Discrete Fourier Transform (DFT) approach. In fact, choosing a suitable orthogonal basis in the sense of DFT, possibly accounting only for a subset of allowed Fourier frequencies, the calculation of the $\underline{\underline{T}}_M$ matrix and of the product $\underline{\underline{T}}_M\underline{\underline{P}}_{\eta,M}$ can be carried out analytically, leading, eventually, to analytical expressions for the resulting functions $f_1^2\left(\underline{k},\underline{x},t\right)$ and $f_2^2\left(\underline{k},\underline{x},t\right)$. Further details about the calculation of the functions $f_1^2\left(\underline{k},\underline{x},t\right)$ and $f_2^2\left(\underline{k},\underline{x},t\right)$ in such special case are provided in the Appendix.

Moreover, it is worth to mention, that, once specified, the fitting model completely defines the matrix $\underline{\underline{T}}_M$ that, consequently, will remain a constant quantity of the problem. The majority of the computational burden is then limited to the calculation of the matrix $\underline{\underline{T}}_M$ at the beginning of the procedure, while other quantities can then be easily calculated afterwards.

The linear estimator of prediction error (LEPrE) in (19)/(21) (or its continuous version (20)/(21)) can now be used for providing a sound definition of Predictability Region. In fact, being $\varepsilon_\delta\left(\underline{x},t\right)$ a zero mean Gaussian process (in space and time), it is completely characterised, for each single point $\left(\underline{x},t\right)$, by its variance. The Predictability Region can therefore be defined as that region Π, in time and space, where the variance of the prediction error is sufficiently small. Given a threshold τ_ε for the standard deviation of the prediction error, the predictability region can then be defined as:

$$\Pi\left(\tau_\varepsilon\right) = \left\{\left(\underline{x},t\right) : \sigma_{\varepsilon_\delta}^2\left(\underline{x},t\right) \le \tau_\varepsilon^2\right\} \tag{22}$$

It can be noticed that, in the probabilistic framework developed herein, there is no single Predictability Region, and the Predictability Region depends on which level of prediction error is considered to be acceptable through the specification of the threshold τ_ε. A similar definition of Predictability Region was also considered by Naaijen et al. (2014), although they based their definition on a threshold level for the Predictability Indicator.

2.2 Implementation and usage

Expression (19)/(20) (or the more general forms (21)) can be used in deterministic phase-resolved wave predictions to supplement the prediction of wave elevation along with an estimation of the prediction error associated with the specific fitting and prediction procedure.

To this end, in general, a fitting model (3) is firstly chosen, and this corresponds to the definition of a certain set of wave number vectors $\tilde{\underline{k}}_h \;\; h=1,...,N_\zeta$. Then, it is necessary to define the set of N_M points, in time and/or space, which are used for the fitting of the propagation model according to the decided fitting procedure. To further proceed it is necessary to know, or at least to have an as good as possible estimation, of the sea elevation spectrum $S_\eta\left(\underline{k}\right)$. In some applications, such as laboratory experiments or numerical simulations, the spectrum $S_\eta\left(\underline{k}\right)$ is known with sufficient accuracy. However, in applications at sea the estimation of $S_\eta\left(\underline{k}\right)$ can be a critical issue; depending on the specific application, methods based on wave radar processing (Nieto Borge et al., 2004), wave buoys, or ship-as-a-wave-buoys (e.g. Tannuri et al., 2012; Mas-Soler et al., 2018), can be exploited to estimate $S_\eta\left(\underline{k}\right)$. It is however crucial that the estimator of the true wave elevation spectrum $S_\eta\left(\underline{k}\right)$ is depurated, as much as possible, from the effects of measurement noise. In this respect, if the measurement noise is assumed to have a flat spectrum S_{δ_η} additive to $S_\eta\left(\underline{k}\right)$, it follows that the spectrum of the measured wave elevation becomes $S_{\delta_\eta}+S_\eta\left(\underline{k}\right)$. In such case, the analysis of very low and very high wave number regions could be used to estimate the contamination S_{δ_η} which should then be removed from the spectrum of the measured wave elevation, to get, eventually, the (depurated) estimation of $S_\eta\left(\underline{k}\right)$. Another step for the application of the proposed methodology is the definition of a modelling for the covariance matrix of the error in the measurement of the wave elevation (see (15)). In case a flat spectral noise level S_{δ_η} is determined, the variance of the measurement error can be obtained from the integration of S_{δ_η} over the region of $\underline{k}$ up to Nyquist limits from the Fourier analysis. Alternatively, the measurement error can be defined, for instance, from the knowledge of the measurement system, being it, e.g. a wave gauge, a wave buoy, or a wave radar, as appropriate. Other application-specific methods can, of course, also be used.

From the description above, almost all the major information would be available at this stage. However, in order to determine (an estimator of) $\sigma^2_{\varepsilon_\delta}\left(\underline{x},t\right)$ it is necessary to make some additional assumption regarding the process $\eta\left(\underline{x},t\right)$. If the fitting procedure is such that analytical expressions for the functions $f_1^2\left(\underline{k},\underline{x},t\right)$ and $f_2^2\left(\underline{k},\underline{x},t\right)$ can be determined, then expression (20) can be directly used, combined with a suitable numerical integration procedure. Alternatively the discrete formulation (19) can be directly used in the general case. However, when using (19), it is necessary to specify a certain number of harmonic components N_η which are assumed to represent the underlying (unknown) stochastic process $\eta\left(\underline{x},t\right)$. In order for $\eta\left(\underline{x},t\right)$ to be a sufficiently good approximation of the true underlying sea elevation Gaussian process, N_η should be sufficiently large. From application it has been noticed that, when the number of components N_η is sufficiently

large, and in particular sufficiently larger than the number of components N_ζ used for the fitted model, the behaviour of $\sigma^2_{\varepsilon_\delta}(\underline{x},t)$ converges, sufficiently for practical purposes, to the value which would be obtained with $N_\eta \to \infty$. This is expectable, since (part of) $\sigma^2_{\varepsilon_\delta}(\underline{x},t)$ represents the prediction error for $\zeta(\underline{x},t)$ which, loosely speaking, is associated with an insufficient frequency resolution in $\zeta(\underline{x},t)$ to resolve "all" the components of $\eta(\underline{x},t)$. The wording "all" is in quotation marks since, in reality, $\eta(\underline{x},t)$, as a stochastic process, has infinite harmonic components of infinitesimal amplitude. This convergence can also be understood from a numerical perspective when (19) is considered as a discretised version of the integral in (20). The drawback of using the discrete formulation (19) is that an increase of N_η tends to increase the dimensions of some of the matrices involved in the calculations, with a consequent increase of memory consumption and computational time. In order for the procedure to be practically viable with present resources, it is necessary to find a trade-off for N_η such that it has a sufficiently large value to have a sufficiently good/converged estimation of $\sigma^2_{\varepsilon_\delta}(\underline{x},t)$, but without exceeding the available computational time and resources. It is nevertheless to be noted that the problem of memory consumption can be practically solved by calculating the functions $\left(q^2_{2i-1}(\underline{x},t), q^2_{2i}(\underline{x},t)\right)$ in (19)/(21), or $\left(f_1^2(\underline{k},\underline{x},t), f_2^2(\underline{k},\underline{x},t)\right)$ in (20)/(21), separately, for each generic wave number $\underline{k}$, and then, respectively, performing the summation in (19)/(21), or applying usual numerical integration routines for calculating the integral in (20)/(21).

With all data available, $\sigma^2_{\varepsilon_\delta}(\underline{x},t)$ can therefore be estimated in a suitable range of space and/or time. This allows, first, to supplement the prediction $\zeta(\underline{x},t)$ with an associated Gaussian confidence interval based on $\sigma_{\varepsilon_\delta}(\underline{x},t)$. Furthermore, given a limit threshold level τ_ε, the predictability region $\Pi(\tau_\varepsilon)$ can eventually be determined in accordance with (22). It is also to be noted that the possibility of estimating $\sigma^2_{\varepsilon_\delta}(\underline{x},t)$ allows, in addition, to devise "optimum" prediction strategies and/or fitting and prediction setups in order to provide wave elevation predictions with minimum/smaller confidence bounds.

3 Verification through simulated applications

Three simulated applications exploiting the described LEPrE methodology are herein reported as examples. The first two applications are simplified long-crested sea cases, and they are used to show and discuss the main characteristics and capabilities of the LEPrE methodology. In the third application, a more realistic scenario, dealing with the prediction of water wave elevation in short-crested sea, is investigated.

The considered applications are meant to verify the developed LEPrE approach and, as a result, they are based on the same main assumptions on which the LEPrE approach is based. Such applications are, therefore, not to be considered as full validations of the method. In fact, a full validation exercise would require comparisons with experimental data or, as an alternative, a comparison with data based on nonlinear wave field simulations (see, e.g., Blondel-Couprie et al., 2013). Such a validation process is considered to be part of future development activities.

The first application presents a simulated laboratory experiment where the wave elevation is assumed to be produced by a wave maker. The purpose of this application is to set up a controlled and yet realistic environment where to test the LEPrE methodology dealing also with the effects of measurement error (noise). The main results are then verified with respect to a set of Monte Carlo simulations and the effect of the measurement noise is discussed.

The second application has been conceived to show how the LEPrE methodology compares with respect to the Predictability Region (Wu, 2004) and Predictability Indicator (Wu, 2004; Naaijen et al., 2014) techniques in the case of a simplified archetypal strongly bimodal spectrum. The considered simplified spectrum is used for comparison and description purposes, in order to emphasize the effects which might appear, less evidently, in a smoother realistic wind waves plus swell bimodal case.

In the last simulated application the LEPrE methodology is applied to a more realistic scenario where an offshore structure is considered and the wave elevation, used for the propagation model, is assumed to be measured by means of a wave radar device. Although the wave radar inversion problem and related errors are out of the scope of this work (on this topic see, e.g. Fucile et al. (2016a) who compared the performance of radar inversion based on the Modulation Transfer Function (MTF) technique, with those of a least-squares technique using Tikhonov regularisation), the main quantities involved in the fitting and propagation process are chosen to be representative of a realistic scenario involving wave radar measurement. The LEPrE methodology is used to provide the standard deviation of prediction error for the propagated wave elevation process. Different spreading angles are considered, and the corresponding effect on the wave elevation prediction error is discussed. Results are then verified against a corresponding set of Monte Carlo simulations.

3.1 Simulated laboratory experiments with wave gauge measurement

Seakeeping experiments may require measuring the wave elevation at particular locations along the wave tank or at the tested model surroundings. In some cases, a direct instrumental measurement may be prevented by the presence of obstacles, like the tested model itself, or the setup of the wave probes can turn out to be particularly inconvenient. A possible solution is to measure the wave elevation at suitable distance upstream from the points of interest and then derive an indirect measure of the (undisturbed) wave elevation where this information is needed. This kind of reconstruction process will eventually be affected by errors, and therefore it must be supplemented with a sound error analysis. If the measurement error is assumed to be uncorrelated, which is the case assumed in this example, then the LEPrE methodology can be applied to this scenario in its definition (19) or (20), which allows dealing with deterministic predictions in presence of measurement noise. The complete forms (21) could in principle be applied if a general, possibly non-diagonal, covariance matrix $\underline{\underline{C}}_{\underline{\delta}_{\eta,M},\underline{\delta}_{\eta,M}}$ for the measurement error is available.

For this example application a wave probe is assumed to be placed in the middle of a wave tank with depth $D_T = 3.8m$, see Figure 1. The indirectly reconstructed wave elevation is required at two downstream probing points at distance $d_1 = 3.0m$ and $d_2 = 6.0m$ from the measuring wave probe. According to (19)/(20), the LEPrE methodology requires the knowledge of the wave spectrum S_η. In general, the "true" wave spectrum is unknown and it has to be assessed from the wave measurement. However, for sake of simplicity, in this case the nominal spectrum is assumed to be known, since it is a typical input for the experimental wave generation. In particular, the spectrum

considered in this example is a JONSWAP spectrum with peak period $T_p = 9.30s$ (corresponding to peak spectral frequency $\omega_p = 0.676 rad/s$) and peak enhancement factor $\gamma = 3.3$. It is assumed that this spectrum is realized in the experimental laboratory at a model scale $\lambda = 50$. At this scale, which is the one at which results presented in this section are reported, the peak period of the spectrum is 1.32s. The wave probe measurements are considered to be affected by Gaussian random noise with a standard deviation proportional to the significant wave height H_s. Two different levels of noise are considered, namely $\sigma_n = 0.02H_s$ and $\sigma_n = 0.10H_s$. Accounting for the linearity of the approach, all results reported hereinafter for this example are normalized by the significant wave height. They can therefore be considered to represent outcomes per unitary H_s, within the limits of the linearity assumption.

The characteristics of the fitting model are defined by the recording time window, assumed to have a length $T_M = 10 \cdot T_p$, and the wave probe sampling frequency, assumed to be 50Hz, corresponding to $\Delta t = 0.02s$. With the considered time step, the measurement is then defined to start at $t = 0s$ and to finish at $t = T_M = 13.14s \approx 10 \cdot T_p$. The fitting/propagation model is based on DFT analysis. The selected Fourier frequencies used for fitting model are $\tilde{\omega}_j = j2\pi/(N_M \Delta t)$ with $j = 1,...,N_M/2-1$ where $N_M = 658$, thus, zero and Nyquist wave frequencies are not taken into account in the calculations. For the propagation model the finite depth dispersion relation $\tilde{\omega}_j = \sqrt{g\tilde{k}_j \tanh(\tilde{k}_j D_T)}$ has been used, with D_T being the wave tank depth.

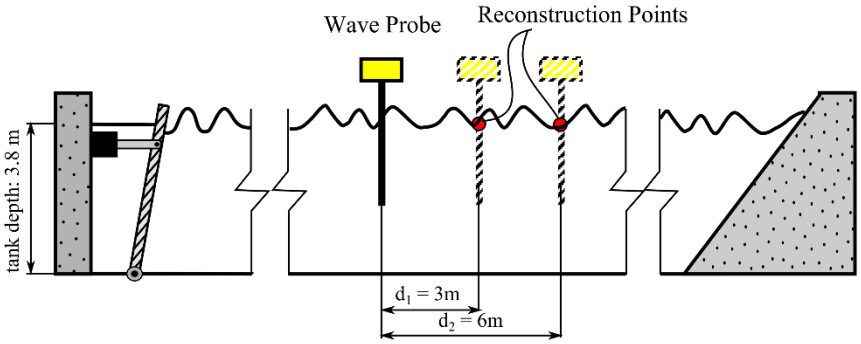

Figure 1: Representative scheme of the wave tank and of the considered idealised laboratory experimental setup.

Results from the LEPrE methodology are verified against a set of 500 Monte Carlo realizations and a 95% confidence interval is provided to account for sampling uncertainty of Monte Carlo simulations. Realizations of wave elevation have been generated following Tucker et al. (1984), using a total of 1026 harmonic components in the range of wave frequencies $0.3 \cdot \omega_p \le \omega \le 3 \cdot \omega_p$. The error standard deviation $\sigma_{\varepsilon_\delta}$ is presented in non-dimensional form, i.e. by normalization with respect to H_s.

First, Figure 2 reports the reconstruction of the wave elevation, normalised by H_s, for four different random realizations of the considered sea state at the probe location $x = d_1 = 3m$ (with $x = 0$ at the measuring wave probe) and with the measurement noise level of $\sigma_n = 0.02H_s$. The true wave elevation is also reported. The LEPrE methodology is used to provide the reconstructed wave elevation with an error bound $\pm 2\sigma_{\varepsilon_\delta}$ corresponding, approximately, to the 95% confidence interval

(the exact 95% confidence interval would correspond to $\pm 1.96\sigma_{\varepsilon_\delta}$). It is possible to note that there is a time interval where the reconstructed wave elevation matches particularly well the true wave elevation and this time window is independent from the realization. Assuming a conventional threshold level $\sigma_{\varepsilon_\delta} < 0.05 H_s$, a predictability time window can therefore be defined on the basis of the LEPrE prediction error standard deviation. In the context of this specific example, the prediction error is to be intended as a "reconstruction error", and the "predictability time window" is to be intended as the region of time where the reconstruction error is sufficiently small at a particular location. Such time window is indicated in Figure 2. It can be noticed that, consistently for all the reported realizations, within the defined predictability time window, the reconstructed wave elevation well reproduces the true wave elevation, whereas, outside, it tends to deviate. It is important to underline that, as described before, it is herein assumed that measurement are taken at the wave probe for a time interval $t_{meas} = [0s, T_M = 13.14s]$ at model scale. It therefore means that for the probe location at $x = 3m$ in Figure 2, the best reconstruction is obtained for time instants before the end of the measurement instant. While this is not an acceptable situation in case of a forecasting application, in the case of the present example this is not posing any problem, because the scope, herein, is to provide an a-posteriori, not real-time, reconstruction of the wave elevation field.

Figure 3 and Figure 4 show the time history of the prediction error standard deviation as calculated at the two probe locations for the two considered levels of noise. Furthermore, the "optimal time delay" δt_{opt} is also identified and highlighted. This is the time delay from the beginning of the measurement window at which the standard deviation of prediction error, $\sigma_{\varepsilon_\delta}$, shows its minimum for the considered probe location. This information could be useful in devising optimum reconstruction strategies, based on the minimization of the prediction error. LEPrE results in Figure 3 and Figure 4 are almost indistinguishable from Monte Carlo simulations and clearly within the confidence interval, indicating that the LEPrE methodology has been properly implemented. Looking at Figure 3 and Figure 4 it can be seen how the distance from the measurement wave probe changes the behaviour of $\sigma_{\varepsilon_\delta}$, with the region of minimum error shifting forward in time and higher in value as the reconstruction point is farther from the measurement probe in the direction of wave propagation. A comparison between results of Figure 3 and Figure 4 indicates that, as the noise level increases, the prediction error standard deviation tends to show a flattening of the region of minimum error. However, at least for the two tested cases characterised by small and moderate noise levels, the optimal time δt_{opt} is independent of the noise level and it only depends on the probe location. It is worth underlining that, in case of the location at $x = 3m$ the optimum reconstruction time occurs before the end of the measurement time window, whereas the opposite occurs for the farther probe at $x = d_2 = 6m$.

Figure 5 and Figure 6 provide a wider picture, by showing the space/time error maps for both the noise levels using the LEPrE methodology. In the figures, the locations of the probe are highlighted with vertical red lines and the points of the minimum prediction error standard deviation are also indicated by a circular yellow symbol. The comparison of the two maps shows that the LEPrE methodology consistently accounts for the presence of higher measurement noise: the increase of the noise level causes an increase of the values of $\sigma_{\varepsilon_\delta}$ in the whole space/time domain.

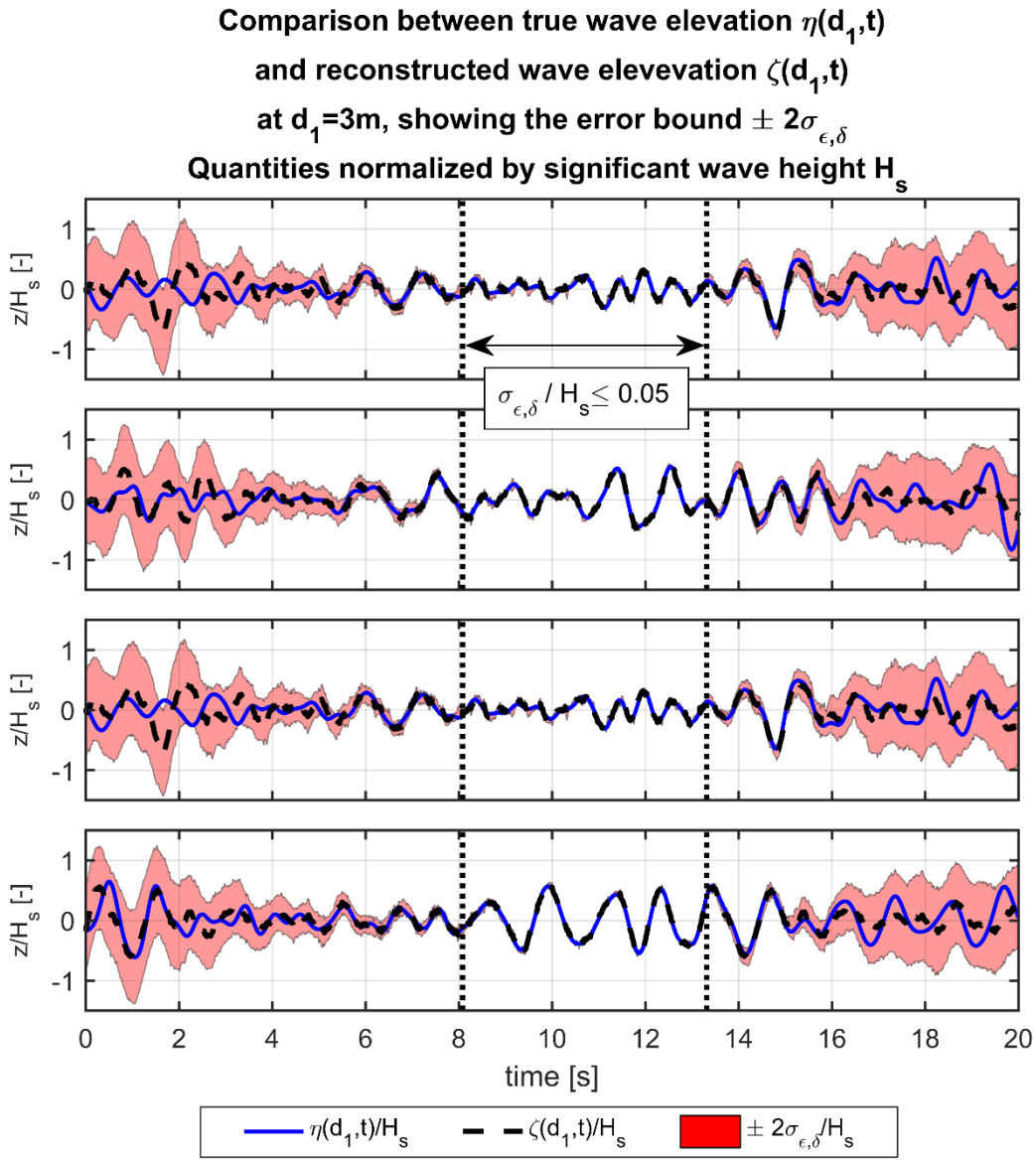

Figure 2: Comparison between the normalized true wave elevation, $\eta(d_1,t)/H_s$, and the normalized reconstructed wave elevation, $\zeta(d_1,t)/H_s$, at the first probing point $d_1 = 3m$. Time histories for four different realizations of the same sea state. Measurement noise level: $\sigma_n = 0.02H_s$. The error bound $\pm 2\sigma_{\varepsilon_\delta}$ is reported as a red band. The two vertical dotted black lines indicate the limit of the time window where the prediction error standard deviation from LEPrE is smaller than the considered threshold value, i.e. $\sigma_{\varepsilon_\delta} \le \tau_\varepsilon = 0.05H_s$.

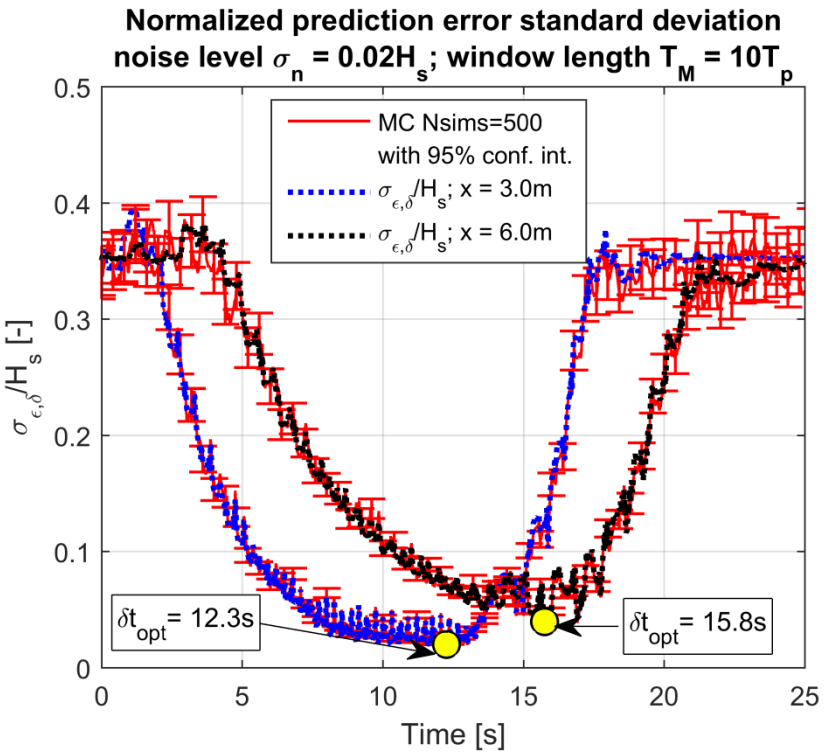

Figure 3: Normalized prediction error standard deviation as a function of time for two different probe locations. Measurement noise level: $\sigma_n = 0.02H_s$.

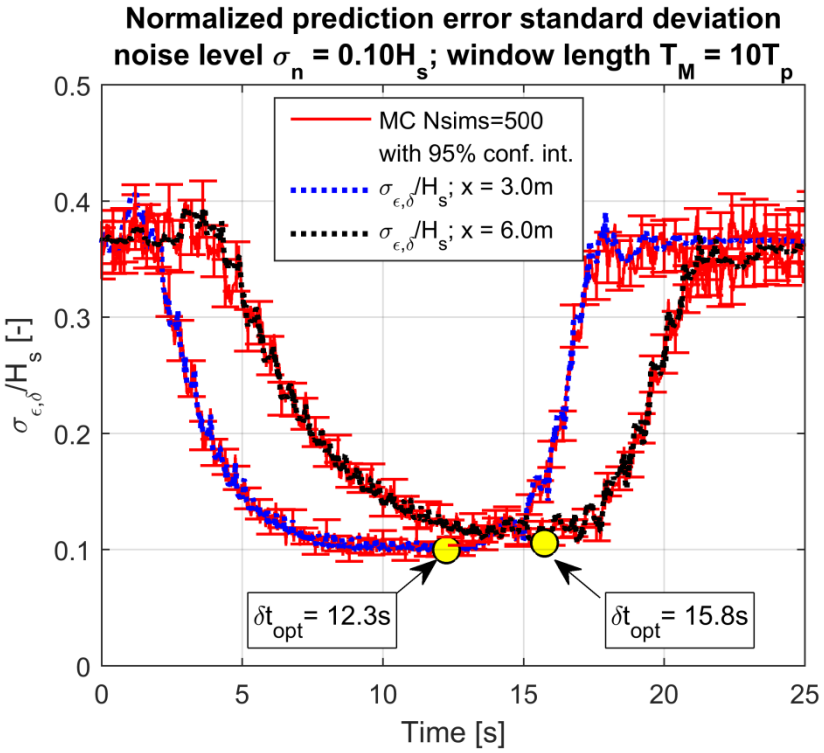

Figure 4: Normalized prediction error standard deviation as a function of time for two different probe locations. Measurement noise level: $\sigma_n = 0.10H_s$.

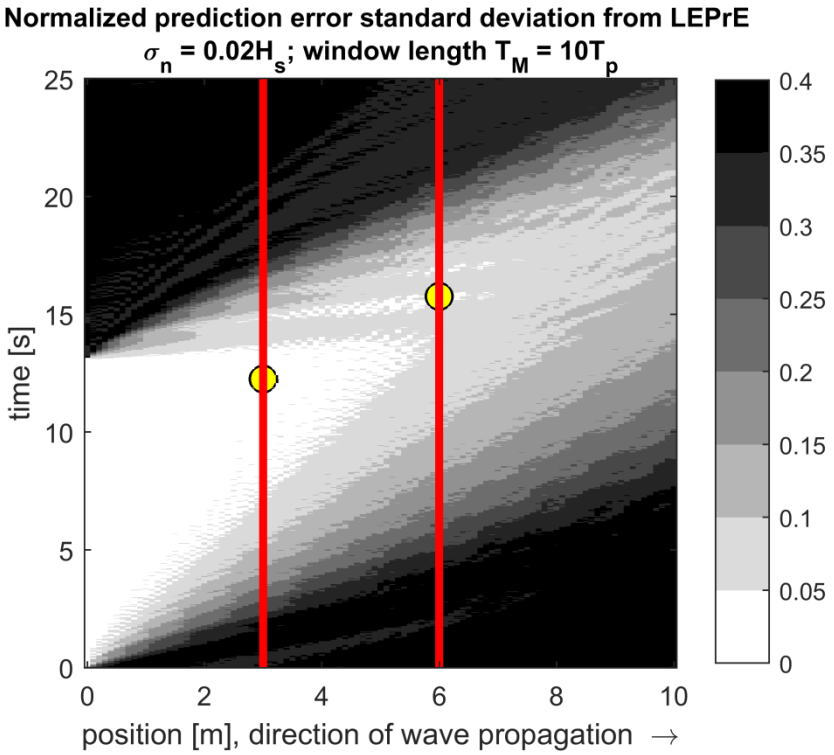

Figure 5: Map of normalized prediction error standard deviation $\sigma_{\varepsilon_\delta} / H_s$. Measurement noise level: $\sigma_n = 0.02 H_s$.

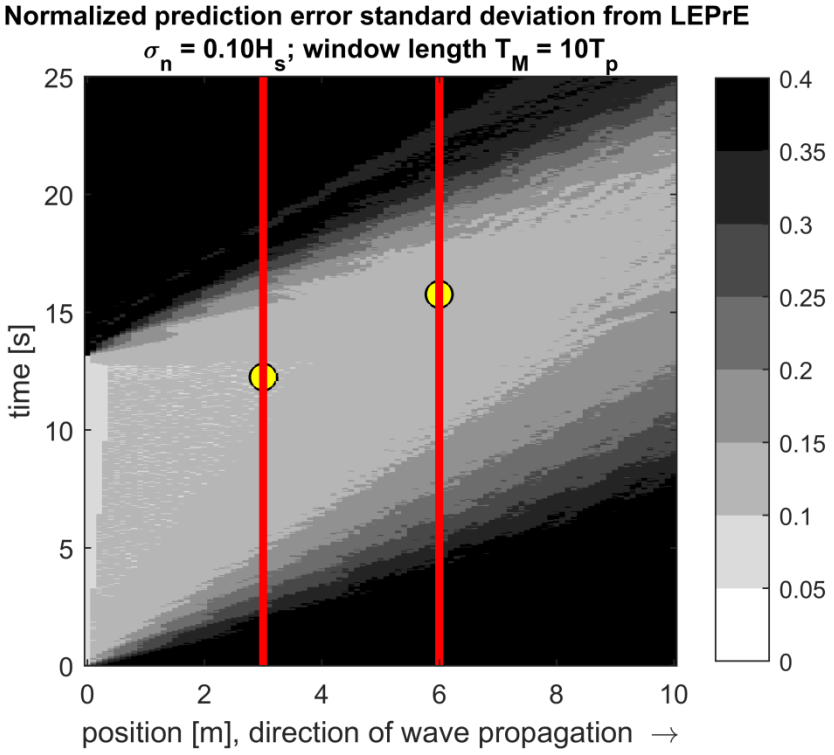

Figure 6: Map of normalized prediction error standard deviation $\sigma_{\varepsilon_\delta} / H_s$. Measurement noise level: $\sigma_n = 0.10 H_s$.

In addition to the analysis of the wave reconstruction at locations different from the wave measurement probe, it is also interesting to provide some more information regarding the

measurement position itself. Data regarding the prediction error at the measurement location can be obtained directly from Figure 5 and Figure 6. However, it is worth providing some more specific detail showing the peculiar behaviour of the quantities of interest at the measurement point. To this end, Figure 7 shows the reconstruction of the wave elevation, normalised by H_s, for four different random realizations of the considered sea state at the measuring wave probe for the case with the measurement noise level of $\sigma_n = 0.02H_s$. This figure is to be compared with Figure 2.

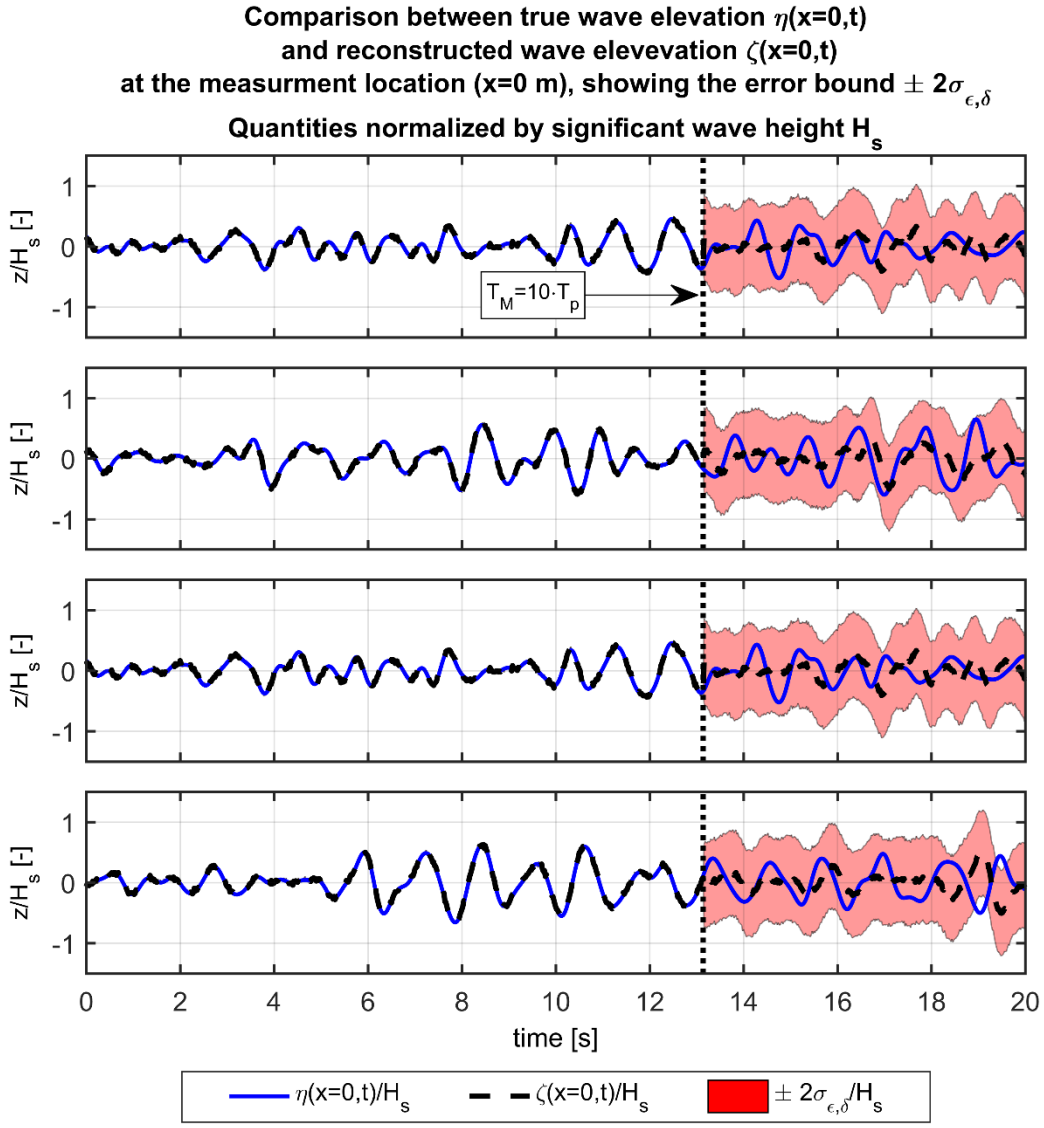

Figure 7: Comparison between the normalized true wave elevation, $\eta(x=0,t)/H_s$, and the normalized reconstructed wave elevation, $\zeta(x=0,t)/H_s$, at the wave measurement probe. Time histories for four different realizations of the same sea state. Measurement noise level: $\sigma_n = 0.02H_s$. The error bound $\pm 2\sigma_{\varepsilon_\delta}$ is reported as a red band.

From Figure 7 it can be seen that, at the measurement location, the model $\zeta(x=0,t)$ fits the wave elevation within the measurement time window $[0,T_M]$, as expected by the described fitting and propagation procedure. As a result, in the measurement time window, the difference between the true wave elevation $\eta(x=0,t)$ and the model $\zeta(x=0,t)$ is basically due to the considered measurement noise. However, for time instants after the end of the measurement time window, the

model $\zeta(x=0,t)$ loses its prediction capabilities, as it is expectable. This is clearly represented by the time dependence of the normalized prediction error standard deviation at the wave measurement location, as reported in Figure 8 (to be compared with Figure 3), which shows a characteristic jump at $t = T_M$. This sudden increase of the prediction error standard deviation corresponds to the sharp widening of confidence intervals in Figure 7. It can be noted that, also for this location, results from the application of LEPrE are verified by the Monte Carlo simulations.

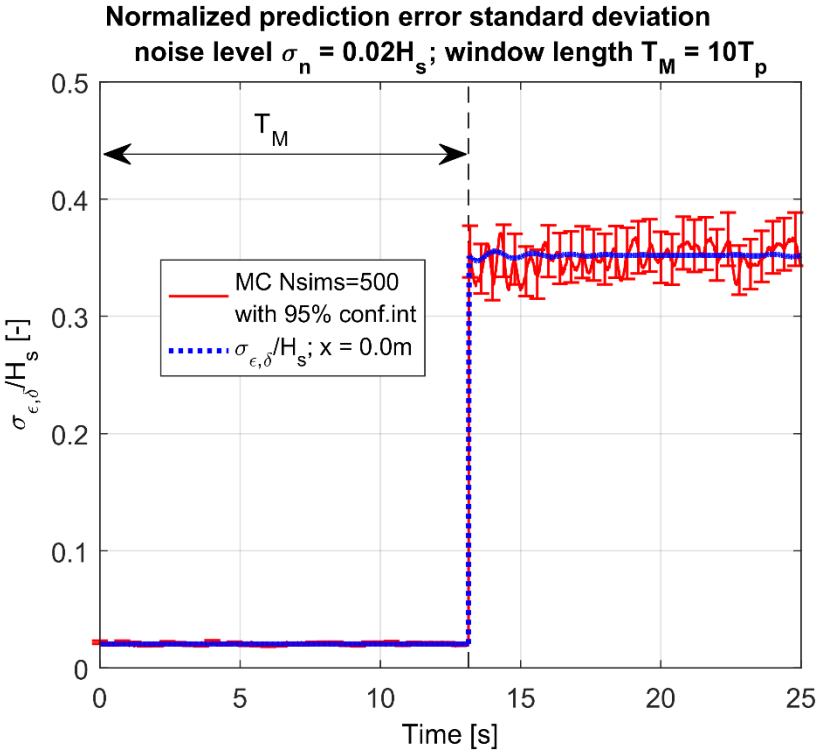

Figure 8: Normalized prediction error standard deviation as a function of time for the wave measurement location x=0m. Measurement noise level: $\sigma_n = 0.02H_s$.

It is finally worth underlining that results obtained in this section are in line with those obtained by Naaijen et al. (2014), who carried out deterministic sea wave predictions, and corresponding prediction error analyses, through simulations and experiments in conditions that are qualitatively similar to those used in the simulated example reported herein.

3.2 Simulated long-crested bimodal sea

This example considers the application of LEPrE methodology in case of a simplified, idealised, bimodal spectrum. The main aim is to establish a simple archetypal scenario able to highlight the common behaviours and the main differences between the LEPrE methodology, the (binary) Predictability Region (Wu, 2004) and the Predictability Indicator (Wu, 2004; Naaijen et al., 2014) techniques. The details of these two techniques are not reported herein and major details can be found in Wu (2004) and Naaijen et al. (2014). More precisely, in this example case, results from the LEPrE methodology are compared with the quantity $\sqrt{1-P}$, where P is the Predictability Indicator from Wu (2004) and Naaijen et al. (2014). By borrowing the nomenclature from Naaijen et al. (2014), $\sqrt{1-P}$ is referred to as "Prediction Error Indicator", although, strictly speaking, this name was used by Naaijen et al. (2014) for the quantity $1-P$. For verification purposes, all results

are also compared against a set of 500 Monte Carlo realizations. Realizations of wave elevation have been generated following Tucker et al. (1984), using a total of 2000 harmonic components in the range of frequencies identified by the nonzero values of the bi-modal spectrum, i.e. 1000 harmonic components within each of the two spectral bands.

The two bimodal spectra used for this application are reported in Figure 9. Each spectrum is designed as the sum of two simple band limited flat spectra with a narrow bandwidth $b_w = 0.1k_0$ and with symmetric peaks separated by $\Delta k = [0.25, 0.75] \cdot k_0$ from the central reference wave number $k_0 = 2\pi / \lambda_0 [1/m]$. The corresponding reference period is $T_0 = 2\pi / \omega_0$ and the reference wave frequency in deep water is $\omega_0 = \sqrt{k_0 \cdot g}$ with g the gravity acceleration. These spectra represent an interesting benchmark for a comparison with the Predictability Region and Predictability Indicator. In fact, the Predictability Region approach takes only into account the fastest and lowest component of the spectrum, without accounting for the actual spectral shape, basing the identification of the predictability region on the assumption of using the group velocity as reference speed. On the other hand, the Predictability Indicator approach, although still using the group velocity as reference speed for the computations, is able to take into account the actual spectral shape. Finally, the LEPrE approach accounts, in an analytic way, for all the features of the fitting procedure and for the exact spectral shape of the underlying wave elevation process, without any addition semi-empirical assumption.

The fitting is carried out in the space domain on a measurement region of length $L_M = 13\lambda_0$ with sampling interval $\Delta x = \lambda_0 / 15$. The measurement interval is $0 \le x / \lambda_0 \le 13$. The fitting/propagation model is defined on the basis of DFT analysis. The Fourier wave numbers selected for the fitting model are $\tilde{k}_j = j2\pi / (N_M \Delta x)$ with $j = 1, ..., N_M / 2 - 1$, with $N_M = 196$, thus, zero and Nyquist wave frequencies are not taken into account in the calculations. For the propagation model the infinite depth dispersion relation $\tilde{\omega}_j = \sqrt{g\tilde{k}_j}$ has been used.

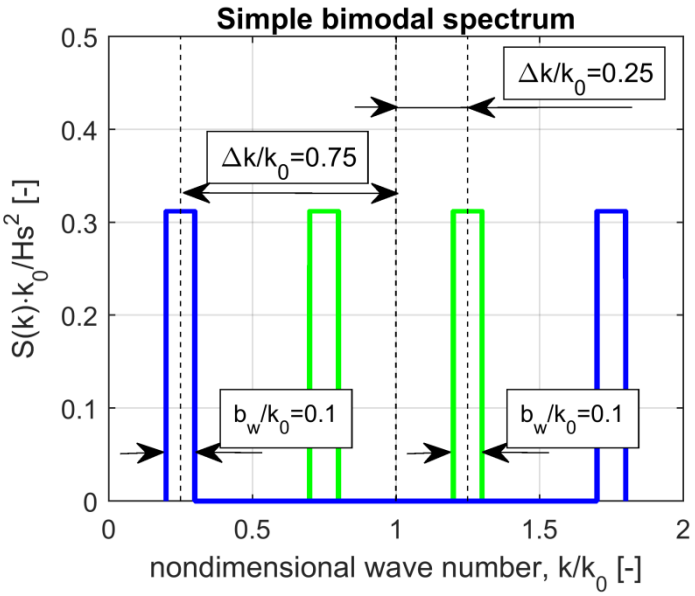

Figure 9: Simple bimodal spectra. Band shift parameter: $\Delta k = 0.25 \cdot k_0$ (green line) and $\Delta k = 0.75 \cdot k_0$ (blue line).

Results from the application of the different methodologies are reported in Figure 10, where surfaces of $\sigma_\varepsilon(x,t)\Big/\left(H_s/2\sqrt{2}\right)$ are shown. Since in this case no measurement error is considered, the standard deviation of prediction error is indicated as σ_ε instead of $\sigma_{\varepsilon_\delta}$. It can be noticed that there is an overall qualitative consistency among the different methodologies. The region where the LEPrE approach shows small values of prediction error standard deviation are in line with the identified Predictability Region according to Wu (2004), and are also in line with the regions where the Prediction Error Indicator, according to Naaijen et al. (2014), is small. Furthermore, the outcomes from the LEPrE methodology are fully verified by the comparison with Monte Carlo simulations, as the two methods show almost undistinguishable prediction error maps. It can be noticed that all methods agree in identifying a reduction of the minimum error region when the bimodal spectrum is characterised by a wider separation between the two bands, which corresponds to the presence of wave components with very different group speeds. The presence of traces in the time/space map associated with the different speeds of the components belonging to the two bands is quite evident, particularly in case of $\Delta k = 0.75 \cdot k_0$.

However, a closer look to results shows how the LEPrE method provides much more information on the prediction error in terms of error pattern, thanks to its capability of implicitly embedding the exact characteristics of the fitting model as well as the exact spectral characteristics of the underlying wave elevation process. It is also worth recalling that, in the limit of the underlying assumptions, the LEPrE formulation stems directly from exact statistical considerations, and it therefore provides clear and unambiguous information about the expected prediction error. On the other hand, the Prediction Error Indicator and the Predictability Region, while qualitatively enabling the identification of regions with “small prediction errors”, they lack a direct relation with a statistical description of the error. However, from the reported comparison, the high level of qualitative agreement between the Prediction Error Indicator and the LEPrE methodology is evident. This suggests the possibility of identifying some underlying semi-empirical relation between the two approaches, and provides a sort of justification to the assumptions used in the determination of the Prediction Error Indicator, at least for the considered fitting procedure setup and for the selected spectral shape.

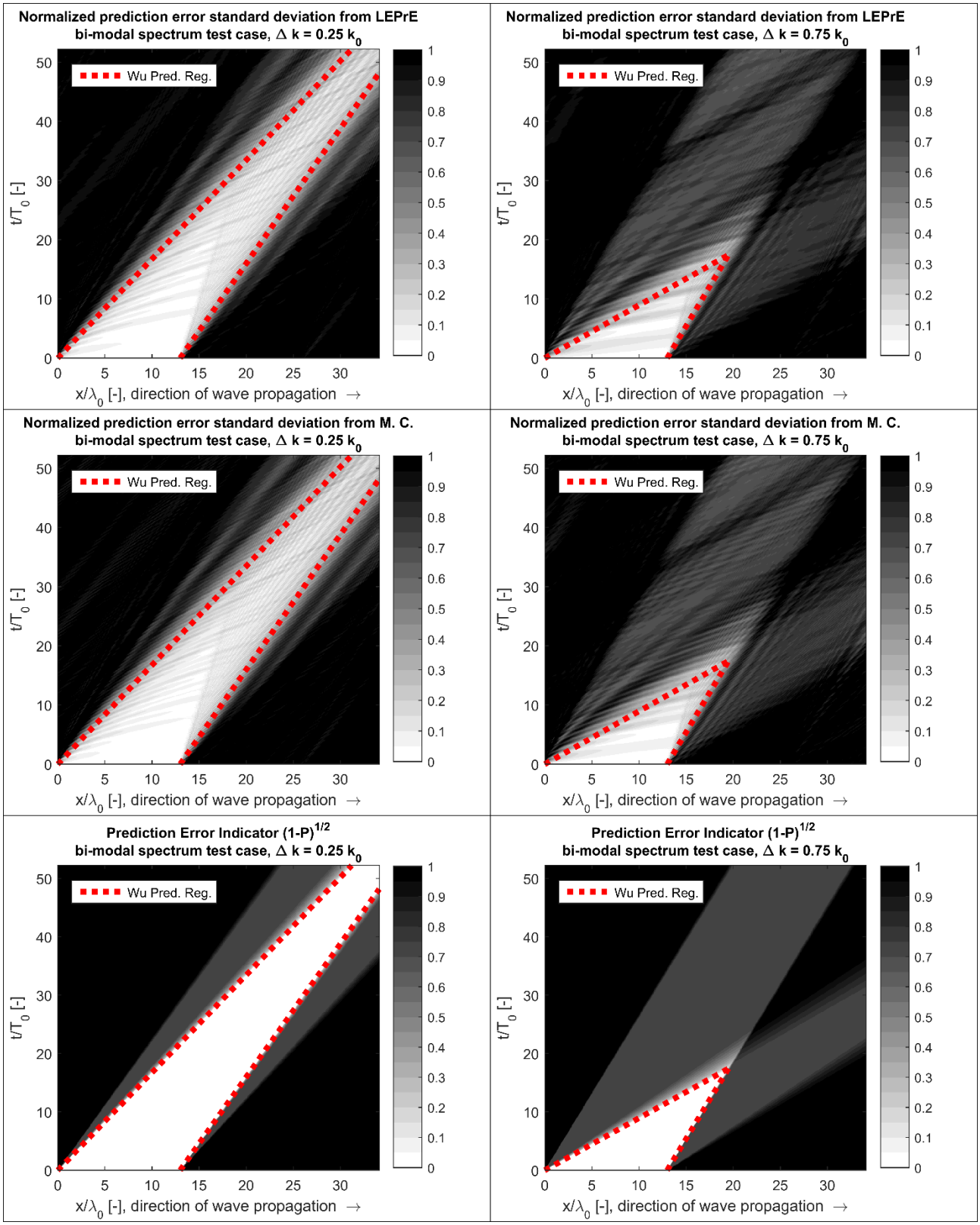

Figure 10: Prediction performance map according to different approaches for two bimodal spectra with band shift parameter $\Delta k = 0.25 \cdot k_0$ (left) and $\Delta k = 0.75 \cdot k_0$ (right). Predictability regions according to Wu (2004) are indicated by red dashed lines in all plots. Normalized prediction error standard deviation, with normalization coefficient $H_s / 2\sqrt{2}$, according to LEPrE (top) and as estimated by means of 500 Monte Carlo realizations (centre). Bottom: Prediction Error Indicator $\sqrt{1-P}$ according to Wu (2004) (see also Naaijen et al. (2014)).

3.3 Simulated short-crested sea with wave radar sensing

In this case, the LEPrE methodology is applied to a realistic scenario where the wave elevation is predicted on the basis of the assumed availability of wave radar measurements. Figure 11 presents the overall scheme of the problem where a wave radar on an offshore structure scans the nearby sea surface. The wave elevation signal is considered to be measured in a annular domain defined by a maximum radar range $R = 3000m$ and limited by a proximity (blind) range $r = 250m$. The notional wave radar imaging bounds are considered to be representative of a usual wave radar measurement. For sake of simplicity, the wave radar is assumed to provide a reconstruction of the instantaneous wave field on the whole domain without errors. Accordingly, and similarly to the previous example application, the prediction error standard deviation is indicated as σ_{ε} and not as $\sigma_{\varepsilon_\delta}$. The fitting procedure is carried on a squared sub-domain of the sensing region with dimensions $L_x = L_y = 1800m$, as reported in Figure 11, and the wave elevation signal is sampled on an evenly spaced grid with spacing $\Delta x = \Delta y = 10m$, which is a typical wave radar resolution. As observed by Blondel-Couprie and Naaijen (2012), the propagation model based on the DFT fitting of an instantaneous wave radar image is particularly sensitive to misinterpretation of the correct propagation direction of the wave components, because of the leakage effects. This leads to an incorrect propagation of waves and is, in general, detrimental to a correct prediction of the wave field. In this work the problem has been dealt with by defining the wave numbers used in the fitting and propagation model to belong to the same quadrant of the main propagation direction of the waves so to have $\tilde{k}_x, \tilde{k}_y > 0$ (Blondel-Couprie and Naaijen, 2012). This assumption basically corresponds to the implementation, in the fitting model, of specific information regarding the directional wave spectrum, and the LEPrE methodology can directly handle this sort of situation.

A grid of $N_x N_y$ points, with $N_x = N_y = 181$, is used for the chosen domain. The time instant of the measurement is conventionally set to $t = 0s$ .The fitting/propagation model is defined on the basis of DFT analysis. The selected Fourier wave numbers for the fitting model are defined in vector form as $\underline{\tilde{k}}_{i,j} = \left(\tilde{k}_{x,i}, \tilde{k}_{y,i}\right)^T$ with $\tilde{k}_{x,i} = i2\pi / (N_x \Delta x)$ and $\tilde{k}_{y,j} = j2\pi / (N_y \Delta y)$ with $i = 0,...,\left(N_x - 1\right)/2$ and $j = 0,...,(N_y - 1)/2$. The zero wave number, $\underline{\tilde{k}} = (\tilde{k}_x = 0, \tilde{k}_y = 0)$ is excluded from the calculations. For the propagation model the infinite depth dispersion relation has been used, so that the wave frequency is calculated as $\tilde{\omega}_{i,j} = \sqrt{g \left\| \underline{\tilde{k}}_{i,j} \right\|}$ and $\left\| \underline{\tilde{k}}_{i,j} \right\| = \sqrt{\tilde{k}_{x,i}^2 + \tilde{k}_{y,j}^2}$.

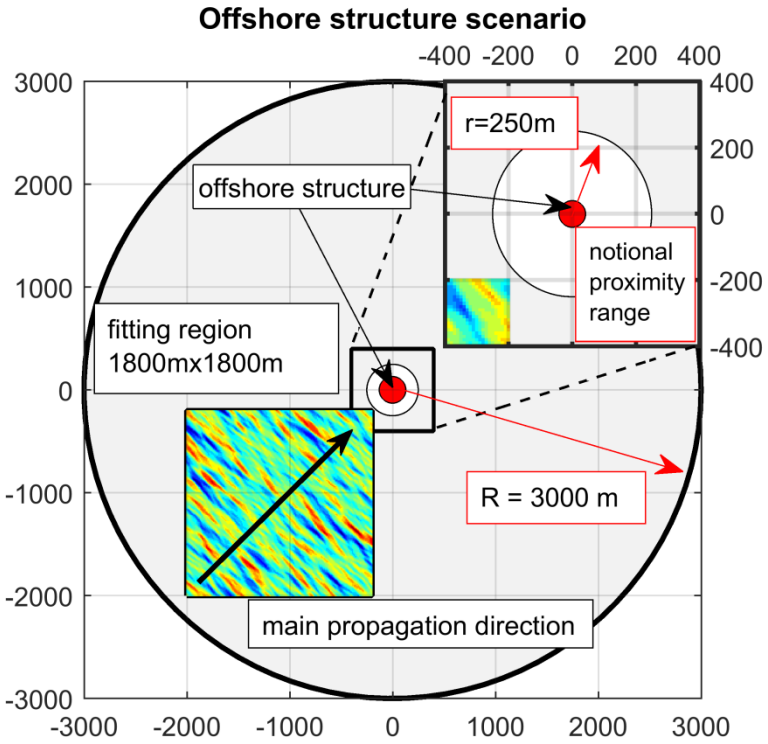

Figure 11: Schematic view of the offshore structure scenario. Radii R (maximum sensing range) and r (proximity range) correspond to limits of the annular sensing region.

A JONSWAP wave elevation spectrum with significant wave height $H_s = 3m$, peak period $T_p = 12s$ (corresponding to peak spectral frequency $\omega_p = 0.524 rad/s$) and peak enhancement factor $\gamma = 3.3$, is considered in this case. The directional wave spectrum $S(k,\theta)$ is defined by using a multiplicative cosine squared spreading function $D(\theta)$ as follows:

$$\begin{cases} S(k,\theta) = S(k)D(\theta) \\ \mathrm{D}(\theta) = \dfrac{2}{\theta_{spread}} \cos\left(\pi \dfrac{\theta - \chi}{\theta_{spread}} \right)^2 \\ \theta \in \left[\chi - \dfrac{\theta_{spread}}{2}, \chi + \dfrac{\theta_{spread}}{2} \right] \end{cases} \tag{23}$$

where θ_{spread} is the spreading angle and χ is the main propagation direction. Different spreading angles are considered, namely $\theta_{spread} = [30, 60, 90]\,\mathrm{deg}$, while the main propagation direction is kept constant to $\chi = 45\,\mathrm{deg}$. An example of directional spectrum is presented in Figure 12.

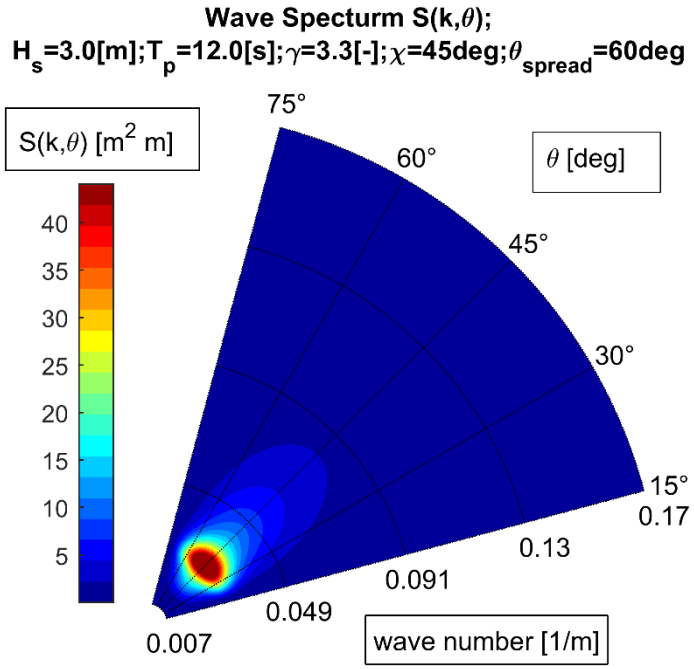

Figure 12: Example of directional wave spectrum used in the calculation. JONSWAP spectrum with cosine squared spreading function: main direction $\chi = 45\,\mathrm{deg}$ and spreading angle $\theta_{spread} = 60\,\mathrm{deg}$, with corresponding range of directions $[15, 75]\,\mathrm{deg}$.

The LEPrE is used here to determine the prediction error standard deviation associated with the assumed fitting and propagation scheme. The determination of the prediction error standard deviation allows also to define, on the one hand, an optimum time delay for the prediction (as discussed before) and, on the other hand, a predictability region given a specified threshold value τ_ε (see (22)). In the following examples a conventional value $\tau_\varepsilon = 0.05 H_s$ is used. Reported results are also verified against a set of 100 Monte Carlo realizations for the case with $\theta_{spread} = 30\,\mathrm{deg}$, which have been analysed in ensemble domain providing estimated expected values and 95% confidence intervals for the prediction error standard deviation. Realizations of wave elevation have been generated by discretizing the spectrum and using random amplitude and phase of each component. The random amplitudes were generated using the local energy of the spectrum, i.e. the local integral of the spectrum associated to each wave component, in accordance with what Tucker et al. (1984) detailed for the long-crested sea case. A total of $3.4 \cdot 10^5$ harmonic components were used in the range of wave frequencies $0.25 \cdot \omega_p \le \omega \le 2.5 \cdot \omega_p$ and in the range of directions $30\,\mathrm{deg} \le \theta \le 60\,\mathrm{deg}$ (see the spreading function $D(\theta)$ in (23)).

Figure 13 shows the comparison of normalized prediction error standard deviation between Monte Carlo simulation and LEPrE results at the offshore structure location. It is noted that the coordinates reported in Figure 13, and in the following ones, are referenced to the lower left corner of the sensing sub-domain shown in Figure 11, and therefore the structure is placed at coordinates $x = y = 2000m$. The very good matching between LEPrE results and Monte Carlo simulations, taking into account the sampling uncertainty reflected by confidence intervals, verifies the implementation of LEPrE also for directional sea state application. Similarly to the previously reported examples, the time history of the normalized prediction error standard deviation shows a minimum, which allows to identify the optimal time delay for a deterministic prediction to be carried out in the considered situation (fitting region, fitting model and sea spectrum). Furthermore, given the considered situation, the behaviour of the prediction error allows also to define the

predictability region, here in the form of a predictability time window. This time window corresponds to the time interval during which the prediction error standard deviation is below the defined threshold $\tau_\varepsilon = 0.05H_s$, i.e. the time window in the future within which the prediction error is considered to be acceptable, according to the threshold limit, from a probabilistic perspective.

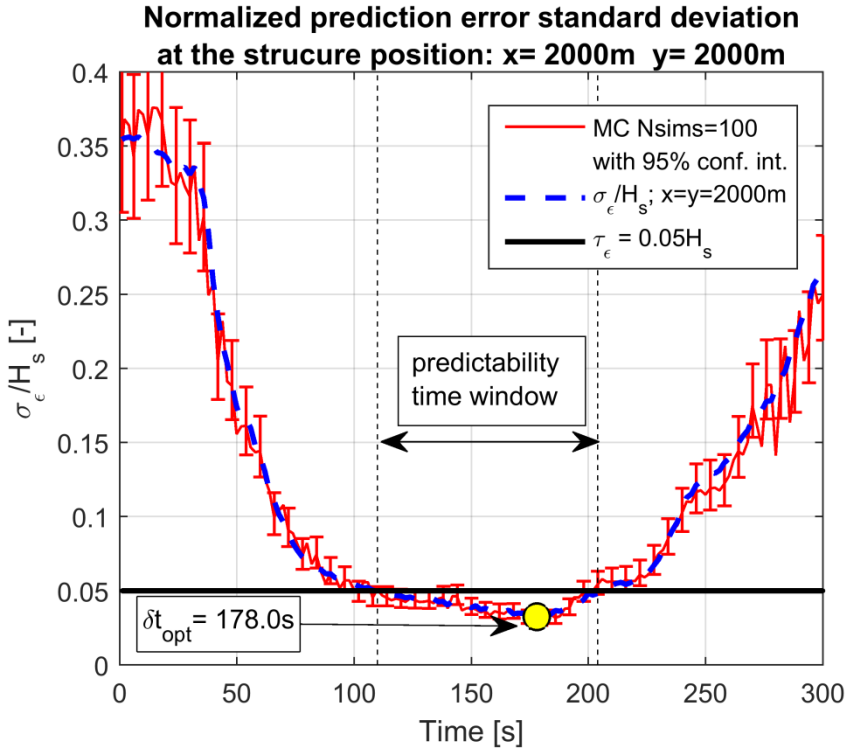

Figure 13: Normalized prediction error standard deviation at the structure location for $\theta_{spread} = 30\,\mathrm{deg}$: comparison between LEPrE and results from 100 Monte Carlo simulations.

Figure 14 presents the LEPrE results, again at the offshore structure location, as obtained for different spreading angles $\theta_{spread} = [30, 60, 90] deg$. Results indicate that an increase of prediction error standard deviation is expected as the spreading angle increases. In the specific case of $\theta_{spread} = 90 deg$ the prediction error standard deviation of the propagation model fails to drop below the specified (although notional) acceptable threshold limit requirements. This indicates that in some cases a predictability region could not exist and this should be taken as an indication about the need of improving the fitting model and/or changing the measurement region.

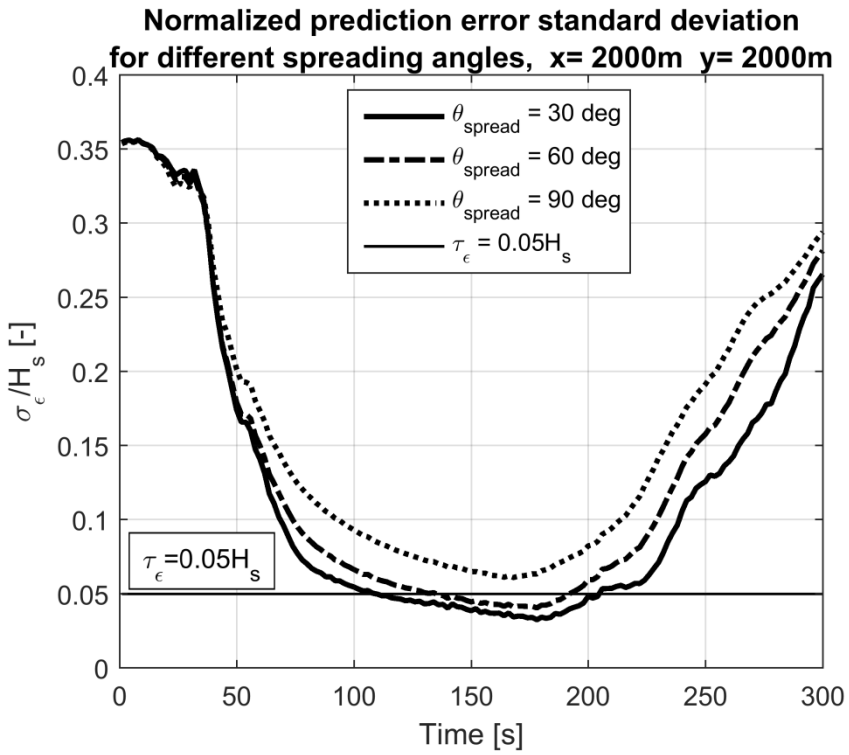

Figure 14: Normalized prediction error standard deviation at the structure location for $\theta_{spread} = [30, 60, 90]\,\mathrm{deg}$: LEPrE results.

Finally, Figure 15 to Figure 19 provide maps of the normalized prediction error standard deviation in the space region surrounding the offshore structure, for different time instants, being $t = 0s$ the instant when the fitting of the propagation model takes place. Normalization in this case is carried out using the normalization factor $H_s / 2\sqrt{2}$. Results from Monte Carlo simulations are also reported as verification of the LEPrE predictions, showing that, considering the natural sampling uncertainty from Monte Carlo simulations, the LEPrE implementation can be considered to be verified. Looking at Figure 15 to Figure 19 as time evolution of the surface of normalized $\sigma_\varepsilon(\underline{x}, t)$, it can be noticed that the region of minimum of the surface, i.e. the region of better prediction from a probabilistic perspective, while modifying its shape in time, propagates along the main wave propagation direction. A smoothing of the "edges" (regions of large gradients) of the surface can also be noted as the time increase, particularly when comparing Figure 15 and Figure 16.

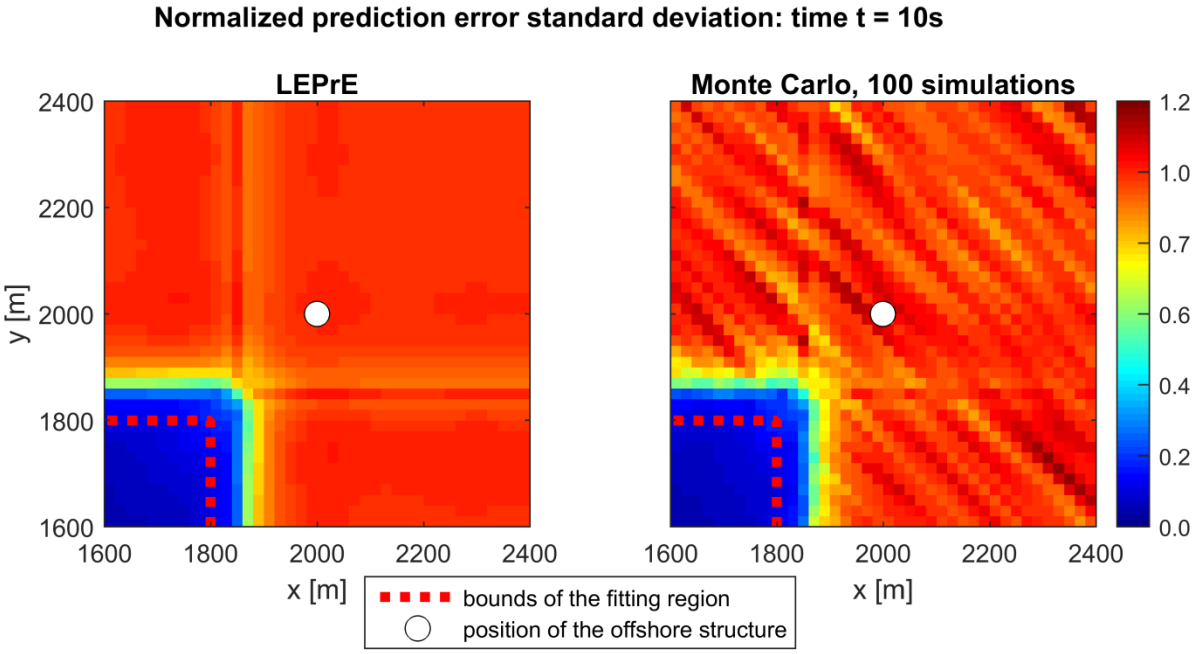

Figure 15: Map of normalized prediction error standard deviation at time $t = 10s$ for a spreading angle of $\theta_{spread} = 30\,\mathrm{deg}$. Normalization coefficient: $H_s / 2\sqrt{2}$. Left: LEPrE results. Right: Monte Carlo simulation results.

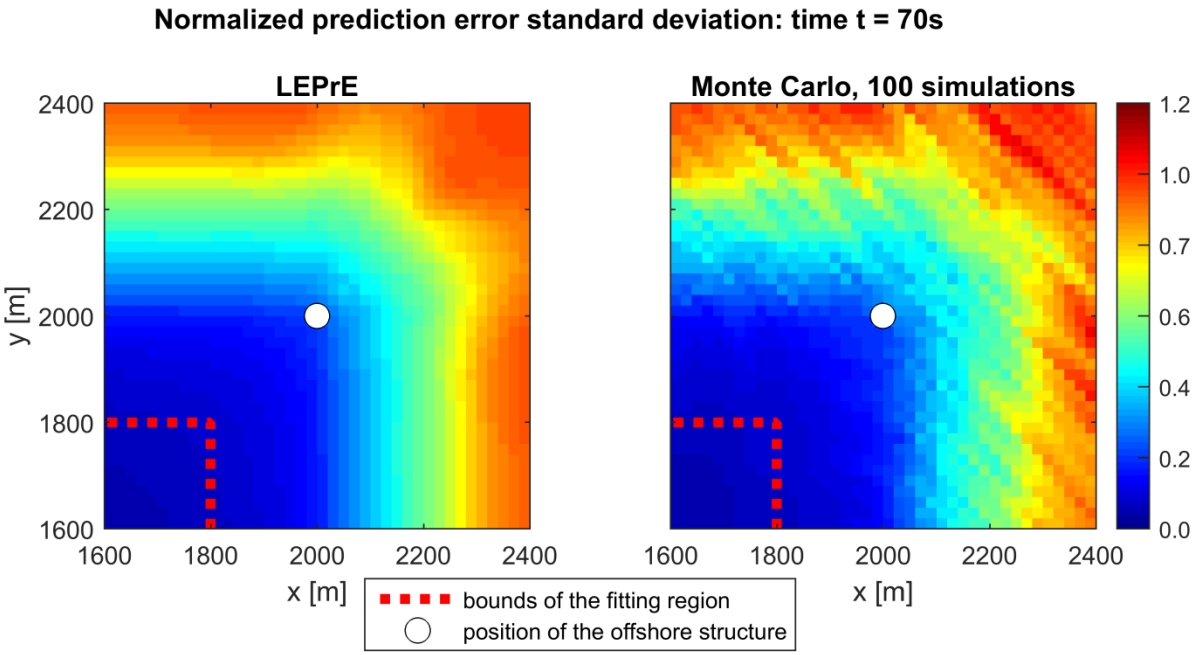

Figure 16: Map of normalized prediction error standard deviation at time $t = 70s$ for a spreading angle of $\theta_{spread} = 30\,\mathrm{deg}$. Normalization coefficient: $H_s / 2\sqrt{2}$. Left: LEPrE results. Right: Monte Carlo simulation results.

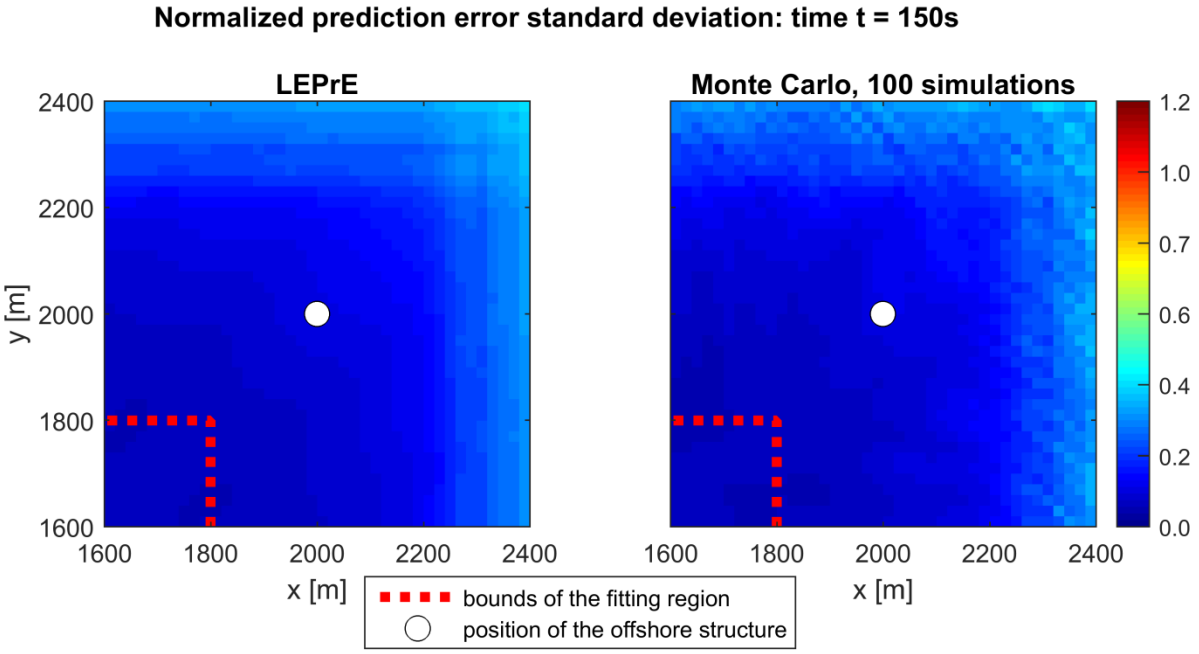

Figure 17: Map of normalized prediction error standard deviation at time $t = 150s$ for a spreading angle of $\theta_{spread} = 30\,\mathrm{deg}$. Normalization coefficient: $H_s / 2\sqrt{2}$. Left: LEPrE results. Right: Monte Carlo simulation results.

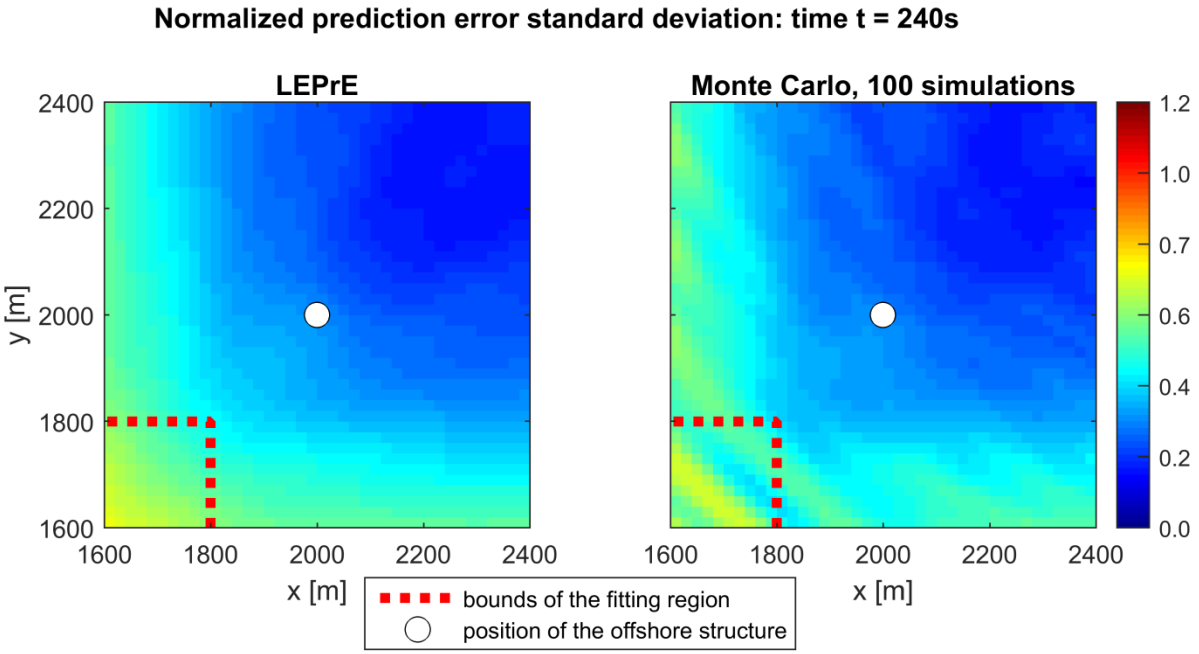

Figure 18: Map of normalized prediction error standard deviation at time $t = 240s$ for a spreading angle of $\theta_{spread} = 30\,\mathrm{deg}$. Normalization coefficient: $H_s / 2\sqrt{2}$. Left: LEPrE results. Right: Monte Carlo simulation results.

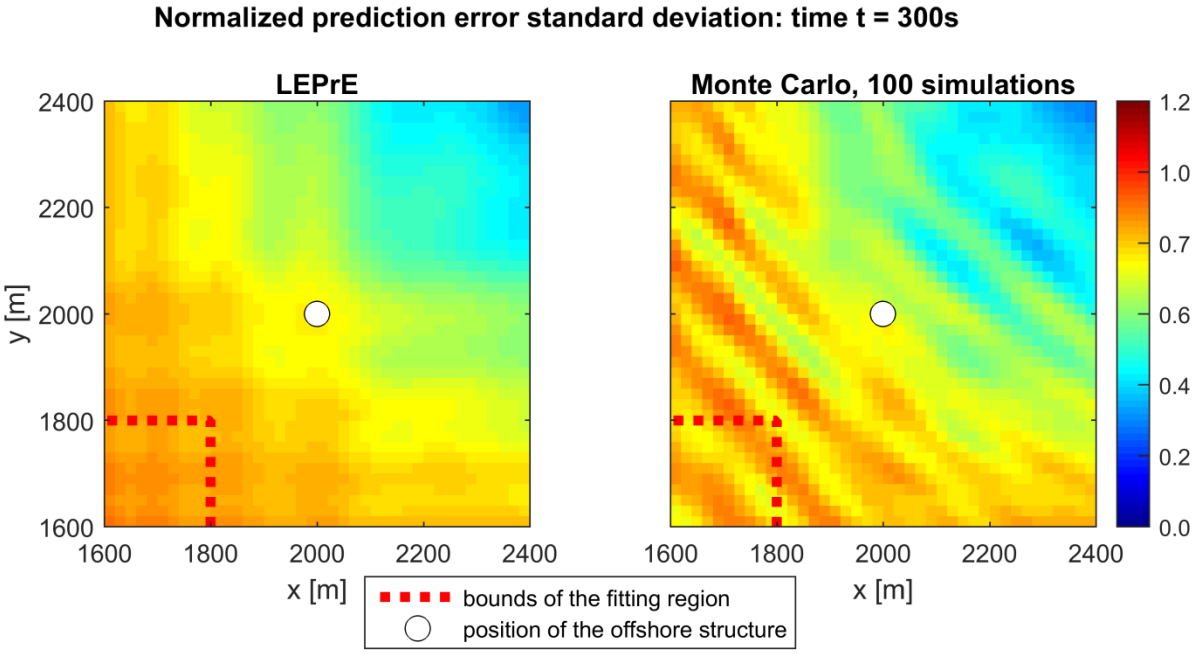

Figure 19: Map of normalized prediction error standard deviation at time $t = 300s$ for a spreading angle of $\theta_{spread} = 30\,\mathrm{deg}$. Normalization coefficient: $H_s / 2\sqrt{2}$. Left: LEPrE results. Right: Monte Carlo simulation results.

4 Some considerations regarding nonlinear effects

The scope of the LEPrE methodology presented herein is to provide a fast tool for supplementing deterministic sea wave predictions (DSWP) with a related estimation of the prediction error characteristics. In order to derive a formulation for prediction error statistics, the LEPrE methodology uses assumptions that are consistent with a linear DSWP framework. As reported in the introduction, the choice of linear DSWP models is usually justified in light of computational efficiency and robustness, although, in principle, it is pertinent only to those situations where wave field nonlinearities are negligible. Nonetheless, even when nonlinear DSWP models are considered, and nonlinear wave contributions are explicitly accounted for, linear tools are still used to provide relevant, though approximate, indications for the higher order models (e.g. Blondel et al., 2010; Blondel-Couprie et al., 2013; Grilli et al., 2011; Nouguier et al., 2014; Wu, 2004). In this respect, the LEPrE methodology, by redefining some key aspect of the linear DSWP framework, represents a step forward compared to existing linear approaches.

Modelling the evolution of a water wave field, simulating its interaction with the environment and catching transient and highly nonlinear water wave phenomena is generally acknowledged as a difficult task. It is, therefore, not surprising that embedding such complexity within fast and efficient DSWP models still represents a prohibitive challenge. Hence, in the DSWP framework some of the most complex nonlinear aspects of the wave dynamics, such as wave breaking or interaction with arbitrary bottom topography, are seldom considered or only partially addressed (Wu, 2004). However, major efforts are done in modelling some of the peculiar nonlinear features of a propagating wave train, such as bound waves, nonlinear dispersion relation or even multiple resonant waves interaction, by using more or less refined nonlinear or weakly nonlinear models. Among them, the most relevant to DSWP are related to the role of the bound waves, for a correct modelling of the wave profile, and to the nonlinear effects on the dispersion relation, for the correct

wave propagation (Blondel et al., 2010; Blondel-Couprie et al., 2013). The former aspect is associated with second- (and higher-) order nonlinear effects, i.e. bound waves, leading, de facto, to a departure from the Gaussianity of the wave elevation field (e.g. Tayfun and Fedele, 2007; Nouguier et al., 2009; Fedele et al. 2017). The modification of the dispersion relation is due to effects arising at least at the third-order.

The number of studies regarding deterministic sea wave predictions based on nonlinear wave models are, however, not numerous. Zhang et al. (1999a,b) developed and successfully applied a directional hybrid wave model based on second-order approach for deterministic sea wave predictions, but they did not compare the outcomes against linear ones. Wu (2004) studied deterministic sea wave forecasting considering a multi-level approach, based on a first-order (linear) model, a second-order model with possibility of accounting for nonlinear dispersion relation and a nonlinear HOS model (with possible arbitrary order of nonlinearly). The last one was also supplemented with a filtering technique aimed to overcome the problems related to the inception of wave breaking and spilling breakers. The analysis by Wu (2004) highlighted the importance of considering nonlinear effects in case of large significant wave steepnesses, whereas the linear model was shown to provide good results in case of milder sea states. In general, for severe sea states, the study by Wu (2004) highlighted the importance of considering the nonlinear wave dispersion relation, and showed that the use of nonlinear prediction models tend to improve the predictions at the cost, however, of an increased computational time. Wu (2004) also highlighted that wave-wave interaction effects become more relevant for wave evolutions on long time horizons. Referring to Wu (2004), Blondel et al. (2010) carried out a numerical study associated with deterministic sea wave predictions. For unidirectional irregular waves they compared the prediction capabilities of: first-order wave model, pure second-order wave model, second-order wave model enhanced with a third-order nonlinear dispersion relation, and a third-order wave model based on HOS. For small significant wave steepenesses, the first-order (linear) model provided results comparable with nonlinear models. At larger significant wave steepness, the relative prediction error by all models increased as well, indicating the effect of nonlinearities (in a fully linear framework the magnitude of the relative prediction error, for a given spectral shape, is independent of the significant wave height). The comparison between prediction capabilities from different models showed that nonlinear models performed better as the significant wave steepness increased. Interestingly, the first-order and the second-order model performed quite similarly, whereas an improvement of the prediction performances was much more noticeable when using the enhanced second-order model and, even further, when using the HOS third-order model. This indicates the importance of a modified, nonlinear, dispersion relation. In particular, the comparison between the predicted and the target wave elevation showed a reasonably good agreement of the wave profile for all models, but the first-order and the second-order model showed a systematic phase shift, indicating the effect of the nonlinear dispersion relation terms. Generally, wave peaks and troughs were better reproduced by the second-order model, this showing a beneficial effect of considering bound waves and wave-wave interactions terms. However, first-order and second order model performed very similarly. A subsequent study by Blondel-Couprie et al. (2013) provided an experimental confirmation of the trends observed by Blondel et al. (2010), on the basis of long-crested irregular waves experiments. Nouguier et al. (2014), referring to synthetic LIDAR data, studied a reconstruction and forecasting approach (see also Grilli et al. (2011)) based on a linear wave model and on a nonlinear Lagrangian "choppy wave model" (Nouguier et al., 2009). Both linear and nonlinear model provided similar reconstruction performance, but the nonlinear model showed better forecasting capabilities.

The highlighted nonlinear effects cannot be directly accounted for by the proposed LEPrE methodology. Indeed, the LEPrE approach is essentially linear and Gaussian; it is therefore representative for describing the free waves contribution. An important characteristic of the LEPrE approach is that, according to the linearity assumption, the harmonics of the single wave

components of the underlying wave field are assumed to be statistically independent random variables in ensemble domain. The introduction of bound waves and wave-wave interactions would break this assumption, because the amplitudes of nonlinearly generated wave components would deterministically depend on those of the free waves. To account for nonlinear effects in the wave elevation, it could be useful to revisit the proposed approach by separating the free waves from the bound waves contributions. At the same time, it is to be noted that the LEPrE approach is not strictly tied to the use of a basic linear dispersion relation. In fact, a general relation between wave number and wave frequency can be used without any restriction. This means that effects such as current could be directly accounted for, but it is also possible to consider the possibility of introducing a nonlinear/corrected dispersion relation. Although not being yet a fully consistent nonlinear extension of the LEPrE approach, such a modification could possibly provide some improvements with respect to the use of a linear approach.

It is clear that a linear model, as the one underlying the proposed LEPrE methodology, has a series of limitations. However, it may be useful, at least in an approximate framework, to try keeping separate the concept of providing an accurate deterministic forecasting of wave elevation, from the concept of providing a corresponding measure of the associated uncertainty (which is the target herein). One of the aims of the LEPrE methodology is to provide a linearly-approximate, but fast and a-priori, estimation of the confidence interval for the forecasting. Notwithstanding the inherent limits, in the interpretation of the prediction during an actual application, the LEPrE estimation could be used either quantitatively (for mild/linear sea states) or, alternatively, as an approximate guide (for more severe, nonlinear, but not strongly nonlinear, sea states). In this latter case it is not unreasonable to imagine that a LEPrE estimation of the uncertainty level could be combined with a deterministic prediction based on nonlinear approaches, at least for not too severe sea states. In such a situation, the high computational effort due to the use of nonlinear wave models would be limited to the single deterministic prediction, while the fast semi-analytical approach from LEPrE would be used for the approximate estimation of uncertainty bounds. If a fully nonlinear approach would be used also for the determination of the confidence interval for the prediction, this would likely require the use of Monte Carlo simulations, which may be impractical for real-time approaches. Beside the scenario of a real time application, it shall also be underlined that an estimation of the prediction error statistics could also be useful as a guide for the preliminary setting up of the parameters of a prediction system. In fact, when a large space of design alternatives is to be explored, the use of nonlinear Monte Carlo approaches could become impractical or basically impossible (depending on the computational effort and the number of parameters to be tested). In such a situation, the use of LEPrE could be of help, at least for a preliminary screening. In all cases, as usual, the gain in computational time, compared to the use of Monte Carlo simulations based on nonlinear models, is paid in the LEPrE methodology by the introduction of a series of simplifying assumptions. On the basis of these considerations, it would therefore be useful to carry out future investigations regarding the qualitative and quantitative comparison between prediction error statistics as obtained from direct nonlinear Monte Carlo simulations and from the proposed LEPrE methodology.

5 Conclusions

Deterministic predictions of wave elevation could contribute to the safety and operability of vessels and other types of units operating at sea, through implementation in early-warning, guidance and decision support systems. While one of the issues in this respect is on how to generate the deterministic predictions, another parallel issue is how to supplement deterministic predictions with a consistent prediction error measure / confidence band.

This paper has specifically addressed the latter problem by presenting a semi-analytical methodology for the determination of prediction error statistics in deterministic sea wave predictions, focusing on linear wave models. The approach is based on the assumption that the underlying wave elevation stochastic process is Gaussian, and that the coefficients of the fitting and propagation model can be obtained from a linear fit of the measured wave elevation in space and time. Furthermore, the presented methodology allows to take into account also the presence of measurement error. Eventually, the presented approach allows to obtain a Linear Estimator of Prediction Error (LEPrE) in space and time, where the prediction error is measured through its ensemble variance. This estimator takes into account the characteristics of the fitting model and of the fitting points in time and/or space, as well as the exact characteristics of the sea spectrum. The approach is applicable in both long-crested and short-crested sea states and allows to directly account for different types of fitting procedures for the coefficients of the propagation model.

The presented LEPrE approach allows to provide a sound and statistically consistent measure of prediction error. As a result, it represents a conceptual step forward with respect to existing approaches (binary Predictability Region, and Prediction Error Indicator), which, while being easier to apply in some situations, are unfortunately characterised by a level of semi-empiricism which could lead to a difficult quantitative interpretation of their outcomes. Overall, the application of LEPrE methodology provides a full picture of the evolution of the standard deviation of prediction error in space and time. This, in turns, allows to properly determine a predictability region given a specified acceptable threshold for the prediction error standard deviation. The capability of determining the prediction error standard deviation also allows the definition of optimum prediction delays in space and/or time, given the fitting model and the underlying sea spectrum, as points of minimum of prediction error standard deviation.

The presented approach has been applied to a series of representative simulated example cases, covering long-crested and short-crested seas, as well as unimodal and (archetypal) bimodal spectra. Corresponding Monte Carlo simulations have also been reported for verification purposes, showing the expected agreement with the LEPrE semi-analytical calculations. The considered example cases are relevant to laboratory applications, as well as applications where the wave elevation could be considered to be measured from a wave elevation sensing device such as a wave radar. The presented approach has also been compared with the existing binary Predictability Region approach and the Prediction Error Indicator, showing differences and highlighting interesting qualitative commonalities.

It is envisioned that the LEPrE approach could be useful in the future for a deeper and more clear understanding of the characteristics of error in deterministic predictions and for the definition of optimal deterministic prediction strategies. Furthermore, the scheme of the presented approach has also been extended by the authors for application to deterministic ship motions predictions, and it is now being tested, with results expected to be reported in the near future.

Acknowledgments

The contribution of Fabio Fucile to this work has been carried out in the framework of his Ph.D. programme at University of Trieste (Ph.D. scholarship D/4-Cycle XXIX funded by CNR-INSEAN).

References

Abusedra, L., Belmont, M.R., 2011. Prediction diagrams for deterministic sea wave prediction and the introduction of the data extension prediction method. International Shipbuilding Progress 58, 59–81.

Belmont, M.R., Horwood, J.M.K., Thurley, R.W.F., Baker, J., 2006. Filters for linear sea-wave prediction. Ocean Engineering 33, pp. 2332–2351.

Belmont, M.R., Horwood, J.M.K., Thurley, R.W.F., Baker, J., 2007. Shallow angle wave profiling Lidar. Journal of Atmospheric and Oceanic Technology 24, pp. 1150–1156.

Belmont, M.R., Christmas, J., Dannenberg, J., Hilmer, T., Duncan, J., Duncan, J.M., Ferrier, B., 2014. An examination of the feasibility of linear deterministic sea wave prediction in multidirectional seas using wave profiling radar: Theory, simulation, and sea trials. Journal of Atmospheric and Oceanic Technology 31, pp. 1601–1614.

Blondel, E., Bonnefoy, F., Ferrant, P., 2010. Deterministic non-linear wave prediction using probe data. Ocean Engineering 37, pp. 913–926.

Blondel-Couprie, E., Bonnefoy, F., Ferrant, P., 2013. Experimental validation of non-linear deterministic prediction schemes for long-crested waves. Ocean Engineering 58, pp. 284–292.

Blondel-Couprie, E., Naaijen, P., 2012. Reconstruction and prediction of short-crested seas based on the application of a 3D-FFT on synthetic waves: Part 2 - Prediction. Proceedings of ASME 31st International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2012), 1-6 July , Rio de Janeiro, Brazil, pp. 55-70.

Connell, B.S.H., Rudzinsky, J.P., Brundick, C.S., Milewski, W.M., Kusters, J.G., Farquharson, G., 2015. Development of an environmental and ship motion forecasting system. Proceedings of ASME 34th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2015), May 31-June 5, St. John's, Newfoundland, Canada, 11p.

Dankert, H., Rosenthal, W., 2004. Ocean surface determination from X-band radar-image sequences. Journal of Geophysical Research 109, C04016.

Edgar, D.R., Horwood, J.M.K., Thurley, R., Belmont, M.R., 2000. The effects of parameters on the maximum prediction time possible in short term forecasting of the sea surface shape. International Shipbuilding Progress 47, pp. 287–301.

Fucile, F., Ludeno, Serafino, F., Bulian, G., Soldovieri, F., Lugni, C., 2016a. Some challenges in recovering wave features from a wave radar system. Proceedings of the 26th International Ocean and Polar Engineering Conference (ISOPE 2016), June 26-July 1, Rhodes, Greece, Vol.3, pp. 319-326.

Fedele, F., Lugni, C, Chawla, A., 2017. The sinking of the El Faro: predicting real world rogue waves during Hurricane Joaquin. Scientific Reports 7, 1188

Fucile, F., Bulian, G., Lugni, C., 2016b. Quantifying error in deterministic predictions based on phase-resolved linear wave models. Proceedings of 26th European Safety and Reliability Conference (ESREL2016), 25-29 September, Glasgow, Scotland, UK, pp. 355-361

Golub, G.H., Van Loan, C.F., 2013. Matrix computations, Fourth edition. Johns Hopkins studies in the mathematical sciences. The Johns Hopkins University Press, Baltimore.

Grilli, S.T., Guérin, C.-H., Goldstein, B., 2011. Ocean Wave Reconstruction Algorithms Based on Spatio-temporal Data Acquired by a Flash LIDAR Camera. Proceedings of the Twenty-first

International Ocean and Polar Engineering Conference (ISOPE 2011), 19-24 June, Maui, Hawaii, USA, Vol. 3, pp. 275-282

Hansen, P.C., 1998. Rank-deficient and discrete ill-posed problems, Mathematical modeling and computation. Society for Industrial and Applied Mathematics.

Hilmer, T., Thornhill, E., 2015. Observations of predictive skill for real-time Deterministic Sea Waves from the WaMoS II. OCEANS 2015 - MTS/IEEE Washington, Washington DC, 19-22 October 2015, 7p.

Mas-Soler, J., Simos, A.N., Tannuri, E., 2018. Estimating on-site wave spectra from the motions of a semi-submersible platform: An assessment based on model scale results. Ocean Engineering 153, pp. 154-172

Morris, E. L., Zienkiewicz, H. K., Belmont, M. R., 1998. Short term forecasting of the sea surface shape. International Shipbuilding Progress 45 (444), pp. 383-400.

Naaijen, P., Blondel-Couprie, E., 2012. Reconstruction and prediction of short-crested seas based on the application of a 3D-FFT on synthetic waves: Part 1 - Reconstruction. Proceedings of ASME 31st International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2012), 1-6 July, Rio de Janeiro, Brazil, pp. 43-55.

Naaijen, P., Huijsmans, R., 2008. Real time wave forecasting for real time ship motion predictions. Proceeding of ASME 27th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2008), 15-20 June, Estoril, Portugal, pp. 607-614.

Naaijen, P., Wijaya, A.P. 2014. Phase resolved wave prediction from synthetic radar images. Proceeding of ASME 33rd International Conference on Ocean, Offshore and Arctic Engineering, (OMAE 2014), 8-13 June, San Francisco, California.

Naaijen, P., Trulsen, K., Blondel-Couprie, E., 2014. Limits to the extent of the spatio-temporal domain for deterministic wave prediction. International Shipbuilding Progress 61, pp. 203–223.

Naaijen, P., van Dijk, R.R.T., Huijsmans, R.H.M., El-Mouhandiz, A.A., 2009. Real time estimation of ship motions in short crested seas. Proceedings of ASME 28th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2009), May 31 - June 5, Honolulu, Hawaii, pp. 243-255.

Nieto Borge, J., Rodriguez, G.R., Hessner, K., González, P.I., 2004. Inversion of marine radar images for surface wave analysis. Journal of Atmospheric and Oceanic Technology 21, pp. 1291–1300.

Nouguier, F., Guérin, C.A., Chapron, B., 2009. "Choppy wave" model for nonlinear gravity waves. Journal of Geophysical Research 114, C09012

Nouguier, F., Grilli, S. T., Guerin, C. A. 2014. Nonlinear ocean wave reconstruction algorithms based on simulated spatio-temporal data acquired by a flash LIDAR camera. Geoscience and Remote Sensing, IEEE Transactions on 52(3), pp. 1761–1771.

Serafino, F., Lugni, C., Nieto Borge, J. C., Soldovieri, F. 2011. A simple strategy to mitigate the aliasing effect in X-band marine radar data: numerical results for a 2D case. Sensors 11(12), pp. 1009–1027.

Tannuri, E.A., Simos, A.N., Sparano, J.V., Matos, V.L.F., 2012. Motion-based wave estimation: small-scale tests with a crane-barge model. Marine Structures 28, pp. 67–85.

Tayfun, M.A., Fedele, F., 2007. Wave-height distributions and nonlinear effects. Ocean Engineering 34, pp. 1631-1649

Tucker, M. J., Challenor, P. I., Carter, D. J. T, 1984. Numerical simulation of a random sea: a common error and its effect upon wave group statistics. Applied Ocean Research 6(2), pp. 118-122

Vettor, R., 2010. Sviluppo di una metodologia di nowcasting del moto ondoso finalizzata all'operatività di mezzi offshore (Development of wave elevation nowcasting methodology for the operability of offshore units). MSc Thesis (In Italian). University of Trieste.

Vogel, C.R., 2002. Computational Methods for Inverse Problems. Society for Industrial and Applied Mathematics.

Wijaya, A.P., Naaijen, P., Andonowati, van Groesen, E., 2015. Reconstruction and future prediction of the sea surface from radar observations. Ocean Engineering 106, pp. 261–270.
Wu, G., 2004. Direct simulation and deterministic prediction of large-scale nonlinear ocean wave-field. PhD Thesis. Massachusetts Institute of Technology.
Yoon, S., Kim, J., Choi, W., 2016. An explicit data assimilation scheme for a nonlinear wave prediction model based on a pseudo-spectral method. IEEE Journal of Oceanic Engineering 41, pp. 112–122.
Zhang, J., Yang, J., Wen, J., Prislin, I., Hong, K., 1999a. Deterministic wave model for short-crested ocean waves: Part I. Theory and numerical scheme. Applied Ocean Research 21, pp. 167–188.
Zhang, J., Prislin, I., Yang, J., Wen, J., 1999b. Deterministic wave model for short-crested ocean waves: Part II. Comparison with laboratory and field measurements. Applied Ocean Research 21, pp. 189–206.

Appendix: analytical formulation of LEPrE in case of DFT fitting applied to long-crested (1D) and short-crested (2D) waves

This section presents the derivation of functions $f_1^2(k,x,t)$ and $f_2^2(k,x,t)$, in equation (20)/(21), in the special case when a DFT fitting approach is applied. Both long-crested (1D) and short-crested (2D) waves are considered.

The derivation starts from the case of long-crested (1D) sea states, where waves propagate in the same direction. In this case the generic wave number vector $\underline{k}$ can be replaced by its corresponding scalar magnitude k and the generic position vector $\underline{x}$ can be replaced by a scalar coordinate x.

The possible presence of measurement error does not affect the definition of $f_1^2(k,x,t)$ and $f_2^2(k,x,t)$ in (20), since the contribution to the prediction error standard deviation due to the presence of measurement error represents an additive term. Therefore, for the purpose of this appendix, the prediction error can be defined by combining equations (10) and (12) as follows, without considering measurement error:

$$\varepsilon(x,t) = \zeta_F(x,t) - \eta(x,t) = \left[\underline{p}_{\zeta}(x,t) \underline{\underline{T}}_M \underline{\underline{P}}_{\eta,M} - \underline{p}_{\eta}(x,t) \right] \underline{\alpha} = \underline{q}(x,t) \underline{\alpha} \tag{A.1}$$

The measurement of the wave elevation is assumed to be carried out on an evenly spaced 1D grid of M points, with sampling resolution Δx so that the sampled position is $x_m = x_0 + m\Delta x$ with $m = \{0,...,M-1\}$ and x_0 is a given bias position. The corresponding Fourier wave numbers are defined as $\tilde{k}_s = 2\pi s / M\Delta x$, with the index s taking the values on a convenient subset of the set $\{0,1,...,M/2\}$. The fitting model is then assumed to be based on a set of N_ζ wave numbers, $\tilde{k}_i \in \{\tilde{k}_1,...,\tilde{k}_{N_\zeta}\}$, chosen among the previously defined Fourier ones. For sake of simplicity, the analytical solution will not account for the zero wave number, $\tilde{k} = 0$, and for the possible Nyquist wave number $\tilde{k} = \pi / \Delta x$. Exploiting the properties of the DFT matrix, it is straightforward to prove that the matrix $\underline{\underline{T}}_M$ appearing in (4) can be actually written as:

$$\underline{\underline{T}}_M = \frac{2}{M} P_{\zeta,M}^T \tag{A.2}$$

The matrix product $\underline{\underline{T}}_M \underline{\underline{P}}_{\eta,M}$ can now be rewritten in the following block matrix form:

$$\underline{\underline{F}}_M = \begin{bmatrix} \underline{\underline{f}}_{1,1} & \cdots & \underline{\underline{f}}_{1,N_\eta} \\ \cdots & \underline{\underline{f}}_{i,j} & \cdots \\ \underline{\underline{f}}_{N_\zeta,1} & \cdots & \underline{\underline{f}}_{N_\zeta,N_\eta} \end{bmatrix} \quad \text{where} \quad \underline{\underline{f}}_{i,j} = \begin{bmatrix} f_{cc}(\tilde{k}_i, k_j) & f_{cs}(\tilde{k}_i, k_i) \\ f_{sc}(\tilde{k}_i, k_j) & f_{ss}(\tilde{k}_i, k_j) \end{bmatrix} \tag{A.3}$$

The determination of the elements of $\underline{\underline{f}}_{i,j}$ can be carried out analytically, leading to the following results:

$$\begin{aligned} &f_{cc}(\tilde{k}_i, k_j) = \frac{2}{M} \sum_{m=0}^{M-1} \cos(\tilde{k}_i x_m) \cos(k_j x_m) = \\ &= \cos\left(\pi(\mu_i + \nu_j) \cdot \Theta\right) D(\mu_i + \nu_j, M) + \cos\left(\pi(\mu_i - \nu_j) \cdot \Theta\right) D(\mu_i - \nu_j, M) \end{aligned} \tag{A.4}$$

$$\begin{aligned} &f_{ss}(\tilde{k}_i, k_j) = \frac{2}{M} \sum_{m=0}^{M-1} \sin(\tilde{k}_i x_m) \sin(k_j x_m) = \\ &= -\cos\left(\pi(\mu_i + \nu_j)\Theta\right) D\left(\mu_i + \nu_j, M\right) + \cos\left((\pi(\mu_i - \nu_j)\Theta\right) D\left(\mu_i - \nu_j, M\right) \end{aligned} \tag{A.5}$$

$$\begin{aligned} &f_{cs}(\tilde{k}_i, k_j) = \frac{2}{M} \sum_{m=0}^{M-1} \cos(\tilde{k}_i x_m) \sin(k_j x_m) = \\ &= \sin\left(\pi(\mu_i + \nu_j)\Theta\right) D(\mu_i + \nu_j, M) - \sin\left(\pi(\mu_i - \nu_j)\Theta\right) D(\mu_i - \nu_j, M) \end{aligned} \tag{A.6}$$

$$\begin{aligned} &f_{sc}(\tilde{k}_i, k_j) = \frac{2}{M} \sum_{m=0}^{M-1} \sin(\tilde{k}_i x_m) \cos(k_j x_m) = \\ &= \sin\left(\pi(\mu_i + \nu_j)\Theta\right) D(\mu_i + \nu_j, M) + \sin\left(\pi(\mu_i - \nu_j)\Theta\right) D(\mu_i - \nu_j, M) \end{aligned} \tag{A.7}$$

where $\Theta = \left(\frac{(M-1)}{M} + \alpha\right)$ with $\alpha = 2x_0 / (M\Delta x)$. μ_i is a natural number defined as $\mu_i = \frac{\tilde{k}_i}{2\pi} M\Delta x \in \mathbb{N}$ and ν_j is in general a real number defined as $\nu_j = \frac{k_j}{2\pi} M\Delta x \in \mathbb{R}$. The function $D(p, M)$ appearing in the previous expressions is defined as:

$$D(p, M) = \frac{\sin(\pi p)}{M \sin(\pi p / M)} \quad \text{with p} \in \mathbb{R} \tag{A.8}$$

The wave numbers $\tilde{k}_i$ and k_j are, respectively, the Fourier wave number of the fitting model and the assumed wave number of the underlying model. This latter, in general, can be arbitrary, and not necessarily linked to the sampling grid.

The results reported so far are provided assuming that the measurement and fitting is carried out in the space domain. However, the obtained results can be recast in a form suitable for the dual situation where the measurement and fitting procedure is carried out in time domain. In such case, it is necessary to interchange space and time, as well as wave numbers and wave frequencies. In a time domain fitting it shall also be assumed that the fitting frequencies correspond to an orthogonal basis with respect to the measurement domain, and that the wave numbers corresponding to the fitting frequencies are determined from the dispersion relation.

It is worth noticing at this point that, in phase-resolved propagation methods, it is good practice to define the sampling grid in such a way to avoid, as much as possible, aliasing problems. However, in general, the sampling grid is defined by the sampling resolution of the measuring device and some aliasing is unavoidable. In this case the whole fitting and propagation procedure will be affected by aliasing. The LEPrE methodology naturally embeds the fitting model characteristics in its definition and consequently it will provide a prediction error that consistently takes into account the presence of aliasing.

The elements of the vector $\underline{q}(x,t)$ in (A.1) can now be rewritten as:

$$
\begin{aligned}
&q_{2j-1}(x,t) = \\
&= \sum_{i=1}^{N_\zeta}\left[f_{cc}(\tilde{k}_i,k_j)\cos\left(\tilde{k}_i x-\omega(\tilde{k}_i)t\right)+f_{sc}(\tilde{k}_i,k_j)\sin\left(\tilde{k}_i x-\omega(\tilde{k}_i)t\right)\right]-\cos\left(k_j x-\omega(k_j)t\right) \\
&q_{2j}(x,t) = \\
&= \sum_{i=1}^{N_\zeta}\left[f_{sc}(\tilde{k}_i,k_j)\cos\left(\tilde{k}_i x-\omega(\tilde{k}_i)t\right)+f_{ss}(\tilde{k}_i,,k_j)\sin\left(\tilde{k}_i x-\omega(\tilde{k}_i)t\right)\right]-\sin\left(k_j x-\omega(k_j)t\right)
\end{aligned}
\tag{A.9}
$$

However, the terms $q_{2j-1}(x,t)$ and $q_{2j}(x,t)$ are now in a closed form with respect to the generic (arbitrary) wave number k_j. As a result, the subscript j can be dropped and the final formulation of the functions $f_1^2(k,x,t)$ and $f_2^2(k,x,t)$ can be written as:

$$
\begin{aligned}
&f_1^2(k,x,t) = \\
&= \left[\sum_{i=1}^{N_\zeta}\left[f_{cc}(\tilde{k}_i,k)\cos\left(\tilde{k}_i x-\omega(\tilde{k}_i)t\right)+f_{cs}(\tilde{k}_i,k)\sin\left(\tilde{k}_i x-\omega(\tilde{k}_i)t\right)\right]-\cos\left(kx-\omega(k)t\right)\right]^2 \\
&f_2^2(k,x,t) = \\
&= \left[\sum_{i=1}^{N_\zeta}\left[f_{sc}(\tilde{k}_i,k)\cos\left(\tilde{k}_i x-\omega(\tilde{k}_i)t\right)+f_{ss}(\tilde{k}_i,k)\sin\left(\tilde{k}_i x-\omega(\tilde{k}_i)t\right)\right]-\sin\left(kx-\omega(k)t\right)\right]^2
\end{aligned}
\tag{A.10}
$$

The formulation developed so far can be easily extended to the short-crested (2D) case within the hypothesis that the fitting procedure is carried out using an orthogonal Fourier basis. The

measurement of the wave elevation is assumed to be carried out on an evenly spaced 2D grid of MN points, with sampling resolution Δx along the x direction and Δy along the y direction. The grid points $\underline{x}_{m,n} = \left(x_m, y_n\right)^T$ are defined as $x_m = x_0 + m\Delta x$ with $m = \{0,...,M-1\}$ and $y_n = y_0 + n\Delta y$ with $n = \{0,...,N-1\}$, with $\left(x_0, y_0\right)$ being the given bias position. The corresponding Fourier wave numbers, considered in their vector form $\underline{\tilde{k}} = \left(\tilde{k}_x, \tilde{k}_y\right)^T$, are defined as $\tilde{k}_{x,r} = 2\pi r / \left(M\Delta x\right)$ with the index r taking the values on a convenient subset of the set $\{0,1,...,M/2\}$ and $\tilde{k}_{y,s} = 2\pi s / \left(N\Delta y\right)$ with the index s taking the values on $\{0,1,...,N/2\}$. The fitting model is then assumed to be based on a set of N_ζ wave numbers, $\underline{\tilde{k}}_i = \left(\tilde{k}_{x,i}, \tilde{k}_{y,i}\right)^T \in \{\underline{\tilde{k}}_1,...,\underline{\tilde{k}}_{N_\zeta}\}$, chosen among the previously defined Fourier ones, excluding the zero and the Nyquist wave numbers.

In this case the functions $f_{cc}, f_{cs}, f_{sc}, f_{ss}$ presented in equations A.4 to A.7, can be exploited to calculate a set of corresponding functions associated with the 2D case as:

$$g_{cc}(\underline{\tilde{k}}_i, \underline{k}_j) = \frac{2}{MN}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\cos(\tilde{k}_{x,i}x_m + \tilde{k}_{y,i}y_n)\cos(k_{x,j}x_m + k_{y,j}y_n) = \\ = \frac{1}{2}\begin{bmatrix} f_{cc}(\tilde{k}_{x,i}, k_{x,j})f_{cc}(\tilde{k}_{y,i}, k_{y,j}) - f_{cs}(\tilde{k}_{x,i}, k_{x,j})f_{cs}(\tilde{k}_{y,i}, k_{y,j})... \\ -f_{sc}(\tilde{k}_{x,j}, k_{x,j})f_{sc}(\tilde{k}_{y,i}, k_{y,j}) + f_{ss}(\tilde{k}_{x,i}, k_{x,j})f_{ss}(\tilde{k}_{y,i}, k_{y,j}) \end{bmatrix} \tag{A.11}$$

$$g_{ss}(\underline{\tilde{k}}_i, \underline{k}_j) = \frac{2}{MN}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\sin(\tilde{k}_{x,i}x_m + \tilde{k}_{y,i}y_n)\sin(k_{x,j}x_m + k_{y,j}y_n) = \\ = \frac{1}{2}\begin{bmatrix} f_{ss}(\tilde{k}_{x,i}, k_{x,j})f_{cc}(\tilde{k}_{y,i}, k_{y,j}) + f_{sc}(\tilde{k}_{x,i}, k_{x,j})f_{cs}(\tilde{k}_{y,i}, k_{y,j})... \\ +f_{cs}(\tilde{k}_{x,i}, k_{x,j})f_{sc}(\tilde{k}_{y,i}, k_{y,j}) + f_{cc}(\tilde{k}_{x,i}, k_{x,j})f_{ss}(\tilde{k}_{y,i}, k_{y,j}) \end{bmatrix} \tag{A.12}$$

$$g_{cs}(\underline{\tilde{k}}_i, \underline{k}_j) = \frac{2}{MN}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\cos(\tilde{k}_{x,i}x_m + \tilde{k}_{y,i}y_n)\sin(k_{x,j}x_m + k_{y,j}y_n) = \\ = \frac{1}{2}\begin{bmatrix} f_{cs}(\tilde{k}_{x,i}, k_{x,j})f_{cc}(\tilde{k}_{y,i}, k_{y,j}) + f_{cc}(\tilde{k}_{x,i}, k_{x,j})f_{cs}(\tilde{k}_{y,i}, k_{y,j})... \\ -f_{ss}(\tilde{k}_{x,i}, k_{x,j})f_{sc}(\tilde{k}_{y,i}, k_{y,j}) - f_{sc}(\tilde{k}_{x,i}, k_{x,j})f_{ss}(\tilde{k}_{y,i}, k_{y,j}) \end{bmatrix} \tag{A.13}$$

$$g_{sc}(\tilde{k}_i, k_j) = \frac{2}{MN}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\sin(\tilde{k}_{x,i}x_m + \tilde{k}_{y,i}y_n)\cos(k_{x,j}x_m + k_{y,j}y_n) = \\ = \frac{1}{2}\begin{bmatrix} f_{sc}(\tilde{k}_{x,i}, k_{x,j})f_{cc}(\tilde{k}_{y,i}, k_{y,j}) - f_{ss}(\tilde{k}_{x,i}, k_{x,j})f_{cs}(\tilde{k}_{y,i}, k_{y,j})... \\ +f_{cc}(\tilde{k}_{x,i}, k_{x,j})f_{sc}(\tilde{k}_{y,j}, k_{y,j}) - f_{cs}(\tilde{k}_{x,i}, k_{x,j})f_{ss}(\tilde{k}_{y,i}, k_{y,j}) \end{bmatrix} \tag{A.14}$$

Similarly to A.10, the index j associated to the frequency in the underlying assumed wave elevation model η can now be dropped, so that the functions $f_1^2\left(\underline{k}, \underline{x}, t\right)$ and $f_2^2\left(\underline{k}, \underline{x}, t\right)$ for the general short-crested (2D) case can finally be written as:

$$
\begin{aligned}
& f_1^2\left(\underline{k},\underline{x},t\right) = \\
& = \left[\sum_{i=1}^{N_\zeta}\left[g_{cc}(\tilde{\underline{k}}_i,\underline{k})\cos\left(\tilde{\underline{k}}_i\underline{x}-\tilde{\omega}_i t\right)+g_{cs}(\tilde{\underline{k}}_i,\underline{k})\sin\left(\tilde{\underline{k}}_i\underline{x}-\tilde{\omega}_i t\right)\right]-\cos\left(\underline{k}\underline{x}-\omega t\right)\right]^2 \\
& f_2^2\left(\underline{k},\underline{x},t\right) = \\
& = \left[\sum_{i=1}^{N_\zeta}\left[g_{sc}(\tilde{\underline{k}}_i,\underline{k})\cos\left(\tilde{\underline{k}}_i\underline{x}-\tilde{\omega}_i t\right)+g_{ss}(\tilde{\underline{k}}_i,\underline{k})\sin\left(\tilde{\underline{k}}_i\underline{x}-\tilde{\omega}_i t\right)\right]-\sin\left(\underline{k}\underline{x}-\omega t\right)\right]^2
\end{aligned}
\qquad \text{(A.15)}
$$