

# Phase-resolved ocean wave forecast with ensemble-based data assimilation

Guangyao Wang<sup>1,2</sup> and Yulin Pan<sup>2,†</sup>

<sup>1</sup>State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China

<sup>2</sup>Department of Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA

(Received 22 November 2020; revised 18 February 2021; accepted 9 April 2021)

Through ensemble-based data assimilation, we address one of the most notorious difficulties in phase-resolved ocean wave forecast, regarding the deviation of numerical solution from the true surface elevation due to the chaotic nature of and underrepresented physics in the nonlinear wave models. In particular, we develop a coupled approach of the high-order spectral (HOS) method with the ensemble Kalman filter (EnKF), through which the measurement data can be incorporated into the simulation to improve the forecast performance. A unique feature in this coupling is the mismatch between the predictable zone and measurement region, which is accounted for through a special algorithm to modify the analysis equation in EnKF. We test the performance of the new EnKF–HOS method using both synthetic data and real radar measurements. For both cases (though differing in details), it is shown that the new method achieves much higher accuracy than the HOS-only method, and can retain the phase information of an irregular wave field for an arbitrarily long forecast time with sequentially assimilated data.

**Key words:** surface gravity waves, computational methods

## 1. Introduction

Accurate prediction of ocean waves plays a significant role in the industries of shipping, oil and gas, aquaculture, ocean renewable energy, coastal and offshore construction. In the past few decades, both phase-averaged and phase-resolved wave models have been developed. The phase-averaged wave models, which provide statistical descriptions in terms of the wave spectrum, have been widely used in the operational forecast of global and regional sea states (Booij, Ris & Holthuijsen 1999; Tolman *et al.* 2009). Despite their

† Email address for correspondence: [yulinpan@umich.edu](mailto:yulinpan@umich.edu)

wide applications and success, phase-averaged models have the limitation of providing no information on the individual deterministic waves. For example, rogue waves, which often appear sporadically and potentially cause enormous damage to offshore structures and ships (Broad 2006; Nikolkina *et al.* 2011), cannot be captured. On the other hand, phase-resolved models can predict the evolution of individual waves, but have received much less attention historically, partly due to the difficulty in obtaining the phase-resolved ocean surface as initial conditions. This has now been largely ameliorated with the recent development of sensing technologies and wave field reconstruction algorithms (e.g. Reichert *et al.* 2004; Nwogu & Lyzenga 2010; Gallego *et al.* 2011; Nougier, Grilli & Guérin 2013; Lyzenga *et al.* 2015; Qi, Xiao & Yue 2016; Desmars *et al.* 2018; Qi *et al.* 2018a). For example, the Doppler coherent marine radars have been applied to measure the radial surface velocity field, based on which the field of both velocity potential and surface elevation can be reconstructed in real time (Nwogu & Lyzenga 2010; Lyzenga *et al.* 2015).

Given the reconstructed surface elevation and velocity potential as initial conditions, the evolution of the wave field can be predicted by linear or nonlinear phase-resolved wave models. Although the linear models yield low computational cost, their prediction horizon is severely limited (e.g. Blondel *et al.* 2010; Qi *et al.* 2017). For nonlinear models, the Euler equations governing the free surface need to be numerically integrated. One efficient numerical algorithm to achieve this goal, based on the high-order spectral (HOS) method, is developed by Dommermuth & Yue (1987) and West *et al.* (1987), with later variants such as Craig & Sulem (1993) and Xu & Guyenne (2009). The novelties of these algorithms lie in the development of an efficient spectral solution of a boundary value problem involved in the nonlinear wave equations, which is neglected in the linear wave models with the sacrifice of accuracy. In recent years, HOS has been developed for short-time predictions of large ocean surface, taking radar measurements as initial conditions (Xiao 2013). However, due to the significant uncertainties involved in realistic forecast (e.g. imperfect initial free surface due to measurement and reconstruction errors; the effects of wind, current, etc., that are not accurately accounted for) as well as the chaotic nature of the nonlinear evolution equations, the simulation may deviate quickly from the true wave dynamics (Annenkov & Shrira 2001). Because of this critical difficulty, operational phase-resolved wave forecast has been considered as a ‘hopeless adventure’ to pursue (Janssen 2008).

The purpose of this paper, however, is to show that the dilemma faced by the phase-resolved wave forecast can be largely addressed by data assimilation (DA), i.e. a technique to link the model to reality by updating the model state with measurement data (Evensen 2003, 2009; Bannister 2017). Mathematically, the principle of DA is to minimize the error of analysis (i.e. results after combining model and measurements), or in a Bayesian framework, to minimize the variance of the state posterior given the measurements (Evensen 1994, 2003; Carrassi *et al.* 2018). Depending on formulations and purposes, two categories of DA algorithms exist, namely the variational-based and the Kalman-filter-based approaches. Among the limited studies to couple DA with phase-resolved wave models, most use the variational-based method, where the purpose is to find the optimal initial condition to minimize a cost function measuring the distance between the model prediction and data in future times (Aragh & Nwogu 2008; Qi *et al.* 2018a; Fujimoto & Waseda 2020). These methods, however, are not directly applicable to operational forecast due to their requirement of future data far after the analysis state (in contrast to the realistic situation where data becomes available sequentially in time). On the other hand, the Kalman-filter-based approach allows data

to be sequentially assimilated, by updating the present state as a weighted average of prediction and data according to the error statistics. While Kalman-filter-based approaches have been commonly applied in phase-averaged models to improve the forecast accuracy (Komen *et al.* 1996; Pinto, Bernadino & Pires Silva 2005; Emmanouil, Galanis & Kallos 2012; Almeida, Rusu & Guedes Soares 2016), their development and application for phase-resolved wave models is still at a very early stage. The only attempt (based on the authors' knowledge) to couple such an approach with a phase-resolved wave model is Yoon, Kim & Choi (2015), which, however, assumes linear propagation of the model error covariance matrix, thus limiting its application only to wave fields of small steepness. More robust methods based on the ensemble Kalman filter (EnKF) (i.e. with error statistics estimated by an ensemble of model simulations), which have led to many recent successes in the geosciences (Carrassi *et al.* 2018), have never been applied to phase-resolved wave forecast. Moreover, most existing work, if not all, uses only synthetic data for the validation of their methods, which ignores the realistic complexity that should be incorporated into the forecast framework, such as the mismatch between the predictable zone and measurement region, and the under-represented physics in the model.

In the present work, we develop the sequential DA capability for nonlinear wave models, by coupling EnKF with HOS. The coupling is implemented in a straightforward manner due to the non-intrusive nature of EnKF, i.e. the HOS code can be directly reused without modification (Evensen 2003, 2009). The new EnKF–HOS solver is able to handle long-term forecast of the ocean surface ensuring minimized analysis error by combining model prediction and measurement data. The possible mismatch of the predictable zone (spatial area theoretically predictable given the limited range of initial conditions, which shrinks in time) and measurement region (spatial area covered by the marine radar which moves with, say, ship speed) is accounted for by a new analysis equation in EnKF. To improve the robustness of the algorithm (i.e. address other practical issues such as misrepresentation of the error covariance matrix due to finite ensemble size and underrepresented physics in the model), we apply both adaptive covariance inflation and localization, which are techniques developed elsewhere in the EnKF community (Anderson & Anderson 1999; Hamill & Whitaker 2005; Anderson 2007; Carrassi *et al.* 2018). We test the performance of the EnKF–HOS method against synthetic and realistic radar data, which shows consistent and significant improvement in forecast accuracy over the HOS-only method in both cases. For the former, we further characterize the effect of parameters in EnKF on the performance. For the latter, we show that the EnKF–HOS method can retain the wave phases for arbitrarily long forecast time, in contrast to the HOS-only method which loses all phase information in a short time.

The paper is organized as follows. The problem statement and detailed algorithm of the EnKF–HOS method are introduced in § 2. The validation and benchmark of the method against synthetic and realistic radar data are presented in § 3. We give a conclusion of the work in § 4.

## 2. Mathematical formulation and methodology

### 2.1. Problem statement

We consider a sequence of measurements of the ocean surface in spatial regions  $\mathcal{M}_j$ , with  $j = 0, 1, 2, 3, \dots$  the index of time  $t$ . In general, we allow  $\mathcal{M}_j$  to be different for different  $j$ , reflecting a mobile system of measurement, e.g. a shipborne marine radar or moving probes. We denote the surface elevation and surface potential, reconstructed from the measurements in  $\mathcal{M}_j$ , as  $\eta_{m,j}(\mathbf{x})$  and  $\psi_{m,j}(\mathbf{x})$  with  $\mathbf{x}$  the two-dimensional (2-D) spatial

coordinates, and assume that the error statistics associated with  $\eta_{m,j}(\mathbf{x})$  and  $\psi_{m,j}(\mathbf{x})$  is known *a priori* from the inherent properties of the measurement equipment.

In addition to the measurements, we have available a wave model that is able to simulate the evolution of the ocean surface (in particular  $\eta(\mathbf{x}, t)$  and  $\psi(\mathbf{x}, t)$ ) given initial conditions. Our purpose is to incorporate measurements  $\eta_{m,j}(\mathbf{x})$  and  $\psi_{m,j}(\mathbf{x})$  into the model simulation sequentially (i.e. immediately as data become available in time) in an optimal way such that the analysis of the states  $\eta_{a,j}(\mathbf{x})$  and  $\psi_{a,j}(\mathbf{x})$  (thus the overall forecast) are most accurate.

### 2.2. The general EnKF–HOS coupled framework

In this study, we use HOS as the nonlinear phase-resolved wave model, coupled with the EnKF for DA. Figure 1 shows a schematic illustration of the proposed EnKF–HOS coupled framework. At initial time  $t = t_0$ , measurements  $\eta_{m,0}(\mathbf{x})$  and  $\psi_{m,0}(\mathbf{x})$  are available, according to which we generate ensembles of perturbed measurements,  $\eta_{m,0}^{(n)}(\mathbf{x})$  and  $\psi_{m,0}^{(n)}(\mathbf{x})$ ,  $n = 1, 2, \dots, N$ , with  $N$  the ensemble size, following the known measurement error statistics (see details in § 2.3). A forecast step is then performed, in which an ensemble of  $N$  HOS simulations are conducted, taking  $\eta_{m,0}^{(n)}(\mathbf{x})$  and  $\psi_{m,0}^{(n)}(\mathbf{x})$  as initial conditions for each ensemble member  $n$  (§ 2.4), until  $t = t_1$  when the next measurements become available. At  $t = t_1$ , an analysis step is performed where the model forecasts  $\eta_{f,1}^{(n)}(\mathbf{x})$  and  $\psi_{f,1}^{(n)}(\mathbf{x})$  are combined with new perturbed measurements  $\eta_{m,1}^{(n)}(\mathbf{x})$  and  $\psi_{m,1}^{(n)}(\mathbf{x})$  to generate the analysis results  $\eta_{a,1}^{(n)}(\mathbf{x})$  and  $\psi_{a,1}^{(n)}(\mathbf{x})$  (§ 2.5). The analysis step ensures minimal uncertainty represented by the analysis ensembles (figure 1), which is mathematically accomplished through the EnKF algorithm. A new ensemble of HOS simulations are then performed taking  $\eta_{a,1}^{(n)}(\mathbf{x})$  and  $\psi_{a,1}^{(n)}(\mathbf{x})$  as initial conditions, and the procedures are repeated for  $t = t_2, t_3, \dots$  until the desired forecast time  $t_{max}$  is reached. The details of each step are introduced next in the aforementioned sections, with the addition of an inflation/localization algorithm to improve the robustness of EnKF included in § 2.6, and treatment of the mismatch between predictable zone and measurement region by modifying the EnKF analysis equation in § 2.7. The full process is finally summarized in § 2.8 with algorithm 1.

### 2.3. Generation of the ensemble of perturbed measurements

As described in § 2.2, ensembles of perturbed measurements are needed at  $t = t_j$ , as the initial conditions of  $N$  HOS simulations for  $j = 0$ , and the input of the analysis step for  $j \geq 1$ . As shown in Burgers, Jan van Leeuwen & Evensen (1998), using an ensemble of measurements (instead of a single measurement) in the analysis step is essential to obtain the correct ensemble variance of the analysis result. We collect and denote these ensembles by

$$\mathbf{S}_{m,j} = \left[ \mathbf{s}_{m,j}^{(1)}, \mathbf{s}_{m,j}^{(2)}, \dots, \mathbf{s}_{m,j}^{(n)}, \dots, \mathbf{s}_{m,j}^{(N-1)}, \mathbf{s}_{m,j}^{(N)} \right] \in \mathbb{R}^{d_j \times N}, \quad (2.1)$$

where  $\mathbf{s}$  represents the state variables of surface elevation  $\eta$  or surface potential  $\psi$ , and  $\mathbf{S}$  the corresponding ensemble. This simplified notation will be used hereafter when necessary to avoid writing two separate equations for  $\eta$  and  $\psi$ .  $\mathbf{s}_{m,j}^{(n)}$  with  $n = 1, 2, \dots, N$  is the  $n$ th member of the perturbed measurements. Here  $d_j$  denotes the number of elements in the measurement state vector of either  $\eta$  or  $\psi$  at  $t = t_j$ . Without loss of generality, in this

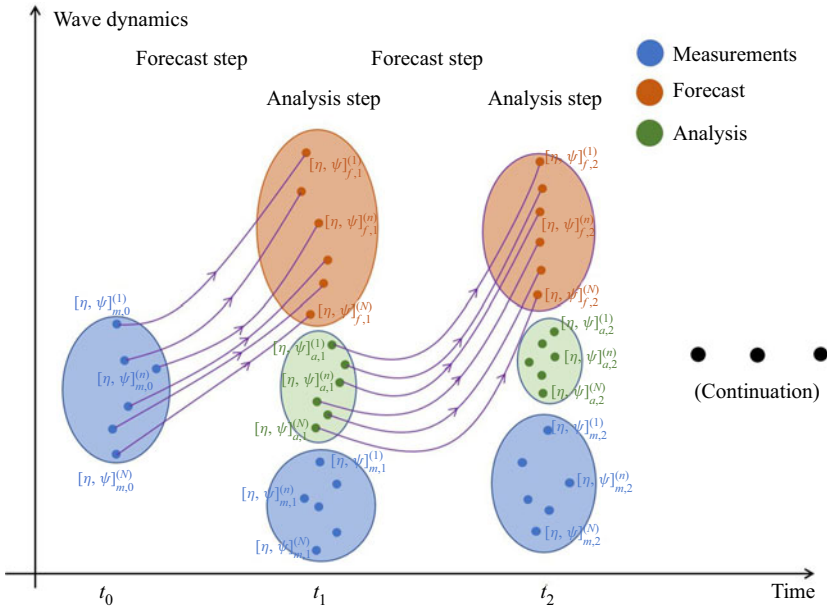


Figure 1. Schematic illustration of the EnKF-HOS coupled framework. The size of ellipse represents the amount of uncertainty. We use short notations  $[\eta, \psi]_{*j}^{(n)}$  to represent  $\eta_{*j}^{(n)}(\mathbf{x}), \psi_{*j}^{(n)}(\mathbf{x})$  with  $* = m, f, a$  for measurement, forecast and analysis, and  $j = 0, 1, 2$ .

work, we use constant  $d_j = d$  for  $j \geq 1$ , and choose  $d_0$  for the convenience of specifying the model initial condition (see details in § 3).

To generate each ensemble member  $s_{m,j}^{(n)}$  from measurements  $s_{m,j}$ , we first produce  $\eta_{m,j}^{(n)}$  from

$$\eta_{m,j}^{(n)}(\mathbf{x}) = \eta_{m,j}(\mathbf{x}) + w^{(n)}(\mathbf{x}), \tag{2.2}$$

where  $w^{(n)}(\mathbf{x})$  is the random noise following a zero-mean Gaussian process with spatial correlation function (Evensen 2003, 2009)

$$C(w^{(n)}(\mathbf{x}_1), w^{(n)}(\mathbf{x}_2)) = \begin{cases} c \exp\left(-\frac{|\mathbf{x}_1 - \mathbf{x}_2|^2}{a^2}\right), & \text{for } |\mathbf{x}_1 - \mathbf{x}_2| \leq \sqrt{3}a, \\ 0, & \text{for } |\mathbf{x}_1 - \mathbf{x}_2| > \sqrt{3}a. \end{cases} \tag{2.3}$$

In (2.3),  $c$  is the variance of  $w^{(n)}(\mathbf{x})$  and  $a$  the decorrelation length scale, both of which practically depend on the characteristics of the measurement devices (and thus assumed known *a priori*). The perturbed measurement of surface potential  $\psi_{m,j}^{(n)}$  is reconstructed from  $\eta_{m,j}^{(n)}$  based on the linear wave theory,

$$\psi_{m,j}^{(n)}(\mathbf{x}) \sim \int \frac{i\omega(\mathbf{k})}{|\mathbf{k}|} \tilde{\eta}_{m,j}^{(n)}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) d\mathbf{k}, \tag{2.4}$$

where  $\tilde{\eta}_{m,j}^{(n)}(\mathbf{k})$  denotes the  $n$ th member of perturbed surface elevation in Fourier space, and  $\omega(\mathbf{k})$  is the angular frequency corresponding to the vector wavenumber  $\mathbf{k}$ . We use ‘ $\sim$ ’ instead of ‘ $=$ ’ in (2.4) as the sign of the integrand relies on the wave travelling direction (and the complex conjugate relation that has to be satisfied for modes  $\mathbf{k}$  and  $-\mathbf{k}$ ).

We remark that this linear construction of  $\psi_{m,j}^{(n)}(\mathbf{x})$  is the best available given the current radar technology, since  $\psi_{m,j}(\mathbf{x})$  is either not directly measured or holds a linear relation with  $\eta_{m,j}(\mathbf{x})$  (Stredulinsky & Thornhill 2011; Hilmer & Thornhill 2015; Lyzenga *et al.* 2015). The additional error introduced by (2.4), however, can be remedied by the EnKF algorithm as will be shown in § 3.

Although the error statistics of the measurements can be fully determined by (2.3) and (2.4), it is a common practice in EnKF to compute the error covariance matrix directly from the ensemble (2.1) (in order to match the same procedure which has to be used for the forecast ensemble). For this purpose, we define an operator  $\mathfrak{C}$  applied on the ensemble  $\mathbf{S}$  (such as  $\mathbf{S}_{m,j}$ ) such that

$$\mathfrak{C}(\mathbf{S}) = \mathbf{S}'(\mathbf{S}')^T, \tag{2.5}$$

where

$$\mathbf{S}' = \frac{1}{\sqrt{N-1}} \left[ s^{(1)} - \bar{s}, s^{(2)} - \bar{s}, \dots, s^{(n)} - \bar{s}, \dots, s^{(N-1)} - \bar{s}, s^{(N)} - \bar{s} \right], \tag{2.6}$$

$$\bar{s} = \frac{1}{N} \sum_{n=1}^N s^{(n)}. \tag{2.7}$$

Therefore, applying  $\mathfrak{C}$  on  $\mathbf{S}_{m,j}$ ,

$$\mathbf{R}_{s,j} = \mathfrak{C}(\mathbf{S}_{m,j}) \tag{2.8}$$

gives the error covariance matrix of the measurements.

### 2.4. Nonlinear wave model by HOS

Given the initial condition  $s_{m,j}^{(n)}$  for each ensemble member  $n$ , the evolution of  $s^{(n)}(\mathbf{x}, t)$  is solved by integrating the surface wave equations in Zakharov form (Zakharov 1968), formulated as

$$\begin{aligned} \frac{\partial \eta(\mathbf{x}, t)}{\partial t} + \frac{\partial \psi(\mathbf{x}, t)}{\partial \mathbf{x}} \cdot \frac{\partial \eta(\mathbf{x}, t)}{\partial \mathbf{x}} - \left[ 1 + \frac{\partial \eta(\mathbf{x}, t)}{\partial \mathbf{x}} \cdot \frac{\partial \eta(\mathbf{x}, t)}{\partial \mathbf{x}} \right] \phi_z(\mathbf{x}, t) &= 0, \tag{2.9} \\ \frac{\partial \psi(\mathbf{x}, t)}{\partial t} + \frac{1}{2} \frac{\partial \psi(\mathbf{x}, t)}{\partial \mathbf{x}} \cdot \frac{\partial \psi(\mathbf{x}, t)}{\partial \mathbf{x}} + \eta(\mathbf{x}, t) - \frac{1}{2} \left[ 1 + \frac{\partial \eta(\mathbf{x}, t)}{\partial \mathbf{x}} \cdot \frac{\partial \eta(\mathbf{x}, t)}{\partial \mathbf{x}} \right] \phi_z(\mathbf{x}, t)^2 &= 0, \tag{2.10} \end{aligned}$$

where  $\phi_z(\mathbf{x}, t) \equiv \partial \phi / \partial z|_{z=\eta}(\mathbf{x}, t)$  is the surface vertical velocity with  $\phi(\mathbf{x}, z, t)$  being the velocity potential of the flow field, and  $\psi(\mathbf{x}, t) = \phi(\mathbf{x}, \eta, t)$ . In (2.9) and (2.10), we have assumed, for simplicity, that the time and mass units are chosen so that the gravitational acceleration and fluid density are unity (e.g. Dommermuth & Yue 1987).

The key procedure in HOS is to solve for  $\phi_z(\mathbf{x}, t)$  given  $\psi(\mathbf{x}, t)$  and  $\eta(\mathbf{x}, t)$ , formulated as a boundary value problem for  $\phi(\mathbf{x}, z, t)$ . This is achieved through a pseudo-spectral method in combination with a mode-coupling approach, with details included in multiple papers such as Dommermuth & Yue (1987) and Pan, Liu & Yue (2018).

2.5. DA scheme by EnKF

Equations (2.9) and (2.10) are integrated in time for each ensemble member to provide the ensemble of forecasts at  $t = t_j$  (for  $j \geq 1$ ):

$$\mathbf{S}_{f,j} = \left[ \mathbf{s}_{f,j}^{(1)}, \mathbf{s}_{f,j}^{(2)}, \dots, \mathbf{s}_{f,j}^{(n)}, \dots, \mathbf{s}_{f,j}^{(N-1)}, \mathbf{s}_{f,j}^{(N)} \right] \in \mathbb{R}^{L \times N}, \tag{2.11}$$

where  $L$  is the number of elements in the forecast state vector and  $\mathbf{s}_{f,j}^{(n)}(\mathbf{x}) \equiv \mathbf{s}_f^{(n)}(\mathbf{x}, t_j)$  is the  $n$ th member of the ensembles of model forecast results. The error covariance matrix of the model forecast can be computed by applying the operator  $\mathfrak{C}$  on  $\mathbf{S}_{f,j}$ :

$$\mathbf{Q}_{s,j} = \mathfrak{C}(\mathbf{S}_{f,j}). \tag{2.12}$$

An analysis step is then performed, which combines the ensembles of model forecasts and perturbed measurements to produce the optimal analysis results (Carrassi *et al.* 2018):

$$\mathbf{S}_{a,j} = \mathbf{S}_{f,j} + \mathbf{K}_{s,j} [\mathbf{S}_{m,j} - \mathbf{G}_j \mathbf{S}_{f,j}], \tag{2.13}$$

$\begin{matrix} L \times N & L \times N & L \times d & d \times N & d \times L \times L \times N \end{matrix}$

where

$$\mathbf{K}_{s,j} = \mathbf{Q}_{s,j} \mathbf{G}_j^T \left[ \mathbf{G}_j \mathbf{Q}_{s,j} \mathbf{G}_j^T + \mathbf{R}_{s,j} \right]^{-1} \tag{2.14}$$

is the optimal Kalman gain matrix of the state (for  $s = \eta$  or  $s = \psi$ ). Here  $\mathbf{G}_j$  is a linear operator, which maps a state vector from the model space to the measurement space:  $\mathbb{R}^L \rightarrow \mathbb{R}^d$ . In the present study,  $\mathbf{G}_j$  is constructed by considering a linear interpolation (or Fourier interpolation (Grafakos 2008)) from the space of model forecast, i.e.  $\mathbf{s}_{f,j}^{(n)}$ , to the space of measurements, i.e.  $\mathbf{s}_{m,j}^{(n)}$ .

While we have now completed the formal introduction of the EnKF–HOS algorithm (and all steps associated with figure 1), additional procedures are needed to improve the robustness of EnKF and address the possible mismatch between the predictable zone and measurement region. These will be discussed, respectively, in §§ 2.6 and 2.7, with the former leading to a (heuristic but effective) correction of  $\mathbf{S}_{f,j}$  and  $\mathbf{Q}_{s,j}$  before (2.13) and (2.14) are applied, and the latter a modification of (2.13) when the mismatch occurs.

2.6. Adaptive inflation and localization

With  $N \rightarrow \infty$  and exact representation of physics by (2.9) and (2.10), it is expected that (2.11) and (2.12) capture the accurate statistics of the model states and (2.13) provides the true optimal analysis. However, due to the finite ensemble size and the underrepresented physics in (2.9) and (2.10), errors associated with statistics computed by (2.12) may lead to suboptimal analysis and even (classical) filter divergence (Evensen 2003, 2009; Carrassi *et al.* 2018). These errors have been investigated by numerous previous studies (Hansen 2002; Evensen 2003; Lorenc 2003; Hamill & Whitaker 2005; Houtekamer *et al.* 2005; Evensen 2009; Carrassi *et al.* 2018), which are characterized by (i) underestimates of the ensemble variance in  $\mathbf{Q}_{s,j}$  and (ii) spurious correlations in  $\mathbf{Q}_{s,j}$  over long spatial distances. To remedy this situation, adaptive inflation and localization (respectively, for error (i) and (ii)) are usually applied as common practices in EnKF to correct  $\mathbf{S}_{f,j}$  and  $\mathbf{Q}_{s,j}$  before they are used in (2.13) and (2.14).

In this work, we apply the adaptive inflation algorithm (Anderson & Anderson 1999; Anderson 2007) in our EnKF–HOS framework. Specifically, each ensemble member in

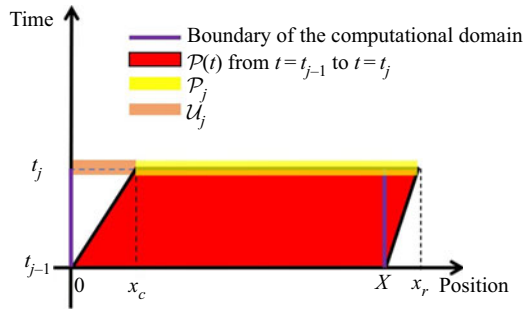


Figure 2. The spatial–temporal predictable zone (red) for a case of unidirectional waves which are assumed to travel from left to right, with initial data in  $[0, X]$  at  $t = t_{j-1}$ . The boundary of the computational domain is indicated by the purple vertical lines. The left and right boundary moves with speed  $c_g^{max}$  and  $c_g^{min}$ , respectively. At  $t = t_j$ , the spatial predictable zone  $\mathcal{P}_j$  (yellow) is located in  $[x_c, x_r] = [c_g^{max}(t_j - t_{j-1}), X + c_g^{min}(t_j - t_{j-1})]$  (or in  $[x_c, X]$  within the computational domain), and the unpredictable zone  $\mathcal{U}_j$  (orange) is located in  $[0, x_c]$ .

(2.11) is linearly inflated before the subsequent computation, i.e.

$$s_{f,j}^{(n),inf} = \sqrt{\lambda_j}(s_{f,j}^{(n)} - \bar{s}_{f,j}) + \bar{s}_{f,j}, \quad n = 1, 2, \dots, N, \tag{2.15}$$

where  $\lambda_j \geq 1$  is referred to as the covariance inflation factor. The purpose of (2.15) is to amplify the underestimated ensemble variance in  $\mathbf{Q}_{s,j}$ , especially when  $\bar{s}_{f,j}$  is far from  $s_{m,j}$ , therefore, to avoid ignorance of  $\mathbf{S}_{m,j}$  in (2.13) (i.e. filter divergence) due to the overconfidence in the forecast. The appropriate value of  $\lambda_j$  can be determined at each  $t = t_j$  through the adaptive inflation algorithm (Anderson 2007), which considers  $\lambda_j$  as an additional state variable maximizing a posterior distribution  $p(\lambda_j|\eta_{m,j})$ . The detailed algorithm is presented in Appendix A.

After obtaining the inflated  $\mathbf{Q}_{s,j}$ , a localization scheme is applied, which removes the spurious correlation by performing the Schur product (i.e. element-wise matrix product) between  $\mathbf{Q}_{s,j}$  and a local-correlation function  $\boldsymbol{\mu}$  (Carrassi *et al.* 2018),

$$\mathbf{Q}_{s,j}^{loc} = \boldsymbol{\mu} \circ \mathbf{Q}_{s,j}, \tag{2.16}$$

with  $\boldsymbol{\mu}$  defined as the Gaspari–Cohn function (see Appendix B for details).

### 2.7. Interplay between predictable zone and measurement region

The predictable zone is a spatiotemporal zone where the wave field is computationally tractable given an observation of the field in a limited space at a specific time instant (Naaijen, Trulsen & Blondel-Couprie 2014; Köllisch *et al.* 2018; Qi *et al.* 2018b). Depending on the wave travelling direction, the boundary of the spatial predictable zone moves in time with speed  $c_g^{max}$  or  $c_g^{min}$ , which are the maximum and minimum group speeds within all wave modes of interest (see figure 2 for an example of predictable zone  $\mathcal{P}(t)$  and unpredictable zone  $\mathcal{U}(t)$  for a unidirectional wave field starting with initial data in  $[0, X]$ ).

In practice, for a forecast from  $t_{j-1}$  to  $t_j$ , the predictable zone  $\mathcal{P}(t)$  at  $t = t_j$  only constitutes a subregion of the computational domain (see caption of figure 2), and there is no guarantee that the measurement region  $\mathcal{M}_j$  overlaps with the predictable zone  $\mathcal{P}_j = \mathcal{P}(t_j)$ . This requires a special treatment of the region where the measurements are available but the forecast is untrustworthy, i.e.  $x \in (\mathcal{U}_j \cap \mathcal{M}_j)$ , where  $\mathcal{U}_j = \mathcal{U}(t_j)$ .



To address this issue, we develop a modified analysis equation which replaces (2.13) when considering the interplay between  $\mathcal{P}_j, \mathcal{U}_j$  and  $\mathcal{M}_j$ .

We consider our computational region as a subset of  $\mathcal{P}_j \cup \mathcal{M}_j$ , so that the analysis results at all  $\mathbf{x}$  can be determined from the prediction and/or measurements. We further partition the forecast and analysis state vectors  $\mathbf{s}_{f,j}^{(n)}$  and  $\mathbf{s}_{a,j}^{(n)}$  (in the computational domain), as well as the measurement state vector  $\mathbf{s}_{m,j}^{(n)}$  (in  $\mathcal{M}_j$ ), according to

$$\mathbf{s}_{*,j} \in \mathbb{R}^{L \times N} = \begin{bmatrix} \mathbf{S}_{*,j}^{\mathcal{P}} \in \mathbb{R}^{L^{\mathcal{P}} \times N} \\ \mathbf{S}_{*,j}^{\mathcal{U}} \in \mathbb{R}^{L^{\mathcal{U}} \times N} \end{bmatrix}, \quad \mathbf{s}_{m,j} \in \mathbb{R}^{d \times N} = \begin{bmatrix} \mathbf{S}_{m,j}^{\mathcal{P}} \in \mathbb{R}^{d^{\mathcal{P}} \times N} \\ \mathbf{S}_{m,j}^{\mathcal{U}} \in \mathbb{R}^{d^{\mathcal{U}} \times N} \end{bmatrix}, \quad (2.17a,b)$$

where  $*$  =  $f, a$ . The variables with superscript  $\mathcal{U}$  ( $\mathcal{P}$ ) represent the part of state vectors for which  $\mathbf{x} \in \mathcal{U}_j$  ( $\mathbf{x} \in \mathcal{P}_j$ ), with associated number of elements  $L^{\mathcal{U}}$  and  $d^{\mathcal{U}}$  ( $L^{\mathcal{P}}$  and  $d^{\mathcal{P}}$ ) for  $\mathbf{s}_{*,j}^{(n)}$  and  $\mathbf{s}_{m,j}^{(n)}$ . The modified analysis equation is formulated as

$$\mathbf{S}_{a,j}^{\mathcal{P}} = \mathbf{S}_{f,j}^{\mathcal{P}} + \mathbf{K}_{s,j}^{\mathcal{P}} [ \mathbf{S}_{m,j}^{\mathcal{P}} - \mathbf{G}_j^{\mathcal{P}} \mathbf{S}_{f,j}^{\mathcal{P}} ], \quad (2.18a)$$

$L^{\mathcal{P}} \times N \quad L^{\mathcal{P}} \times N \quad L^{\mathcal{P}} \times d^{\mathcal{P}} \quad d^{\mathcal{P}} \times N \quad d^{\mathcal{P}} \times L^{\mathcal{P}} \quad L^{\mathcal{P}} \times N$

$$\mathbf{S}_{a,j}^{\mathcal{U}} = \mathbf{H}_j \mathbf{S}_{m,j}, \quad (2.18b)$$

$L^{\mathcal{U}} \times N \quad L^{\mathcal{U}} \times d \quad d \times N$

where  $\mathbf{K}_{s,j}^{\mathcal{P}} = \mathbf{K}_{s,j}(1 : L^{\mathcal{P}}, 1 : d^{\mathcal{P}})$  and  $\mathbf{G}_{s,j}^{\mathcal{P}} = \mathbf{G}_{s,j}(1 : d^{\mathcal{P}}, 1 : L^{\mathcal{P}})$  are submatrices of  $\mathbf{K}_{s,j}$  and  $\mathbf{G}_{s,j}$  associated with  $\mathbf{x} \in \mathcal{P}_j$  in both measurement and forecast (analysis) spaces,  $\mathbf{H}_j$  is a linear operator which maps a state vector from measurement space to unpredictable zone in the analysis space:  $\mathbb{R}^d \rightarrow \mathbb{R}^{L^{\mathcal{U}}}$  (based on linear/Fourier interpolation).

By implementing (2.18a),  $\mathbf{S}_{f,j}^{\mathcal{U}}$  is discarded in the analysis to compute  $\mathbf{S}_{a,j}^{\mathcal{P}}$ ; and by (2.18b),  $\mathbf{S}_{a,j}^{\mathcal{U}}$  is determined only from the measurements  $\mathbf{S}_{m,j}$  without involving  $\mathbf{S}_{f,j}^{\mathcal{U}}$ . Therefore, the modified EnKF analysis equation provides the minimum analysis error when considering the interplay among  $\mathcal{P}_j, \mathcal{U}_j$  and  $\mathcal{M}_j$ .

### 2.8. Pseudocode and computational cost

Finally, we provide a pseudocode for the complete EnKF–HOS coupled algorithm in [algorithm 1](#). For an ensemble forecast by HOS, the algorithm takes  $O(NL \log L)$  operations for each time step  $\Delta t$ . For the analysis step at  $t = t_j$ , the algorithm has a computational complexity of  $O(dLN)$  for (2.13) and  $O(dL^2) + O(d^3) + O(d^2L)$  for (2.14) (if Gaussian elimination is used for the inverse). Therefore, the average computational complexity for one time step is  $O(NL \log L) + O(dL^2)/(\tau/\Delta t)$  (for  $L \gg N$  and  $L \gg d$ ), with  $\tau$  the DA interval.

### 3. Numerical results

To test the performance of the EnKF–HOS algorithm, we apply it on a series of cases with both synthetic and real ocean wave fields. For the former, we use a reference HOS simulation to generate the true wave field, on top of which we superpose random errors to generate the synthetic noisy measurements. For the latter, we use real data collected by a shipborne Doppler coherent marine radar (Nwogu & Lyzenga 2010; Lyzenga *et al.* 2015). The adaptive inflation and localization algorithms are only applied for the latter case, where the under-represented physics in (2.9) and (2.10) significantly influences the model statistics. The perturbed measurement ensemble (2.1) are generated with parameters

---

**Algorithm 1** Algorithm for the EnKF–HOS method.

---

```

1: Input:  $\eta_{m,0}$ ,  $\psi_{m,0}$  (initial measurements),  $t_{max}$  (final computational time),  $N$ 
   (ensemble size),  $\mathcal{T} = \{t_1, t_2, t_3 \dots\}$  (time instants of DA),  $\bar{\lambda}_0$  and  $\sigma_0^2$  (initial guess
   of the inflation factor, see Appendix A)
2: Begin
3: initialize:
4:      $t = t_0, j = 0$ 
5:     Generate  $\mathbf{S}_{m,0}$  with (2.1) ~ (2.4).
6: time loop:
7:     while  $t \leq t_{max}$  do
8:          $j = j + 1$ .
9:         Solve (2.9) and (2.10) until  $t = t_j$  to obtain  $\mathbf{S}_{f,j}$ .
10:        read  $\eta_{m,j}$  (measurements)
11:        Generate  $\mathbf{S}_{m,j}$  with (2.1) ~ (2.4).
12:        Calculate  $\mathbf{R}_{s,j}$  with (2.8).
13:        Perform adaptive inflation with (2.15) and calculate  $\mathbf{Q}_{s,j}$  with (2.12).
14:        Perform covariance localization on  $\mathbf{Q}_{s,j}$  with (2.16).
15:        Calculate  $\mathbf{S}_{a,j}$  with (2.13) (or (2.18a) when considering  $\mathcal{P}_j, \mathcal{U}_j$  and  $\mathcal{M}_j$ ).
16:        Output  $\bar{s}_{a,j}$  (ensemble mean of  $\mathbf{S}_{a,j}$ ).
17:         $\mathbf{S}_{f,j} = \mathbf{S}_{a,j}$ .
18:    end
19: end

```

---

$c = 0.0025\sigma_\eta^2$  (where  $\sigma_\eta$  is the standard deviation of the surface elevation field) and  $a = \lambda_0/8$  (where  $\lambda_0$  is the fundamental wavelength in the computational domain) in (2.3) in all cases unless otherwise specified. We remark that these choices may not reflect the true error statistics of the radar measurement, which unfortunately has not been characterized yet. In all HOS computations, we use a nonlinearity order  $M = 4$  to solve (2.9) and (2.10) considering sufficiently deep water. The results from EnKF–HOS simulations are compared with HOS-only simulations (both taking noisy measurements as initial conditions) to demonstrate the advantage of the new EnKF–HOS method.

### 3.1. Synthetic cases

We consider the synthetic cases where the true solution of a wave field ( $\eta^{true}(\mathbf{x}, t)$  and  $\psi^{true}(\mathbf{x}, t)$ ) is generated by a single reference HOS simulation starting from the (exact) initial condition. The (noisy) measurements of surface elevation are generated by superposing random error on the true solution,

$$\eta_{m,j}(\mathbf{x}) = \eta^{true}(\mathbf{x}, t_j) + v(\mathbf{x}), \quad j = 0, 1, 2, \dots, \tag{3.1}$$

where  $v(\mathbf{x})$  is a random field, which represents the measurement error and shares the same distribution as  $w^{(n)}$  (see (2.2) and (2.3)). For simplicity in generating the initial model ensemble, we use  $\eta_{m,0} \in \mathbb{R}^L$  in (3.1) (i.e.  $d_0 = L$  in (2.1)), and  $\eta_{m,j} \in \mathbb{R}^d$  for  $j \geq 1$  with  $d$  specified in each case below. Similar to  $\psi_{m,j}^{(n)}(\mathbf{x})$ ,  $\psi_{m,0}(\mathbf{x})$  is generated based on the linear wave theory,

$$\psi_{m,0}(\mathbf{x}) \sim \int \frac{i\omega(\mathbf{k})}{|\mathbf{k}|} \tilde{\eta}_{m,0}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) \, d\mathbf{k}, \tag{3.2}$$

where  $\tilde{\eta}_{m,0}(\mathbf{k})$  denotes the measurement of surface elevation in Fourier space at  $t = 0$ . Although the linear relation (3.2) may introduce initial error in  $\psi_{m,0}(\mathbf{x})$ , the EnKF algorithm ensures the sequential reduction of the error with the increase of time.

Depending on how the true solution is generated, we further classify the synthetic cases into idealistic and realistic cases. In the idealistic case, the true solution is taken from an HOS simulation with periodic boundary conditions, so that the entire computational domain is predictable. In the realistic case, we consider the true solution as a patch in the boundless ocean (practically taken from a patch in a much larger domain where the HOS simulation is conducted), and the interplay between  $\mathcal{M}_j$  and  $\mathcal{P}_j$  discussed in § 2.7 is critical. Correspondingly, we apply the modified EnKF analysis equation (2.18) only in the realistic cases.

To quantify the performance of EnKF–HOS and HOS-only methods, we define an error metric

$$\epsilon(t; \mathcal{A}) = \frac{\int_{\mathcal{A}} |\eta^{true}(\mathbf{x}, t) - \eta^{sim}(\mathbf{x}, t)|^2 d\mathcal{A}}{2\sigma_{\eta}^2 \mathcal{A}}, \tag{3.3}$$

where  $\mathcal{A}$  is a region of interest based on which the spatial average is performed (here we use  $\mathcal{A}$  to represent both the region and its area),  $\eta^{sim}(\mathbf{x}, t)$  represents the simulation results obtained from EnKF–HOS (the ensemble average in this case) or the HOS-only method, and  $\sigma_{\eta}$  is the standard deviation of, say,  $\eta^{true}$  in  $\mathcal{A}$ . It can be shown that the definition (3.3) yields  $\epsilon(t; \mathcal{A}) = 1 - \rho_{\mathcal{A}}(\eta^{true}, \eta^{sim})$ , with

$$\rho_{\mathcal{A}}(\eta^a, \eta^b) = \frac{\int_{\mathcal{A}} \eta^a(\mathbf{x}, t)\eta^b(\mathbf{x}, t)d\mathcal{A}}{\sigma_{\eta}^2 \mathcal{A}}, \tag{3.4}$$

being the correlation coefficient between  $\eta^a$  and  $\eta^b$  (in this case  $\eta^{true}$  and  $\eta^{sim}$ ). Therefore,  $\epsilon(t; \mathcal{A}) = 1$  corresponds to the case that all phase information is lost in the simulation.

In the following, we show results for idealistic and realistic cases of synthetic irregular wave fields. Preliminary results validating the EnKF–HOS method for Stokes waves in the idealistic setting can be found in Wang & Pan (2020).

### 3.1.1. Results for idealistic cases

We consider idealistic cases with both 2-D (with one horizontal direction  $x$ ) and three-dimensional (3-D) (with two horizontal directions  $\mathbf{x} = (x, y)$ ) wave fields. The true solutions for both situations are obtained from reference simulations starting from initial conditions prescribed by a realization of the JONSWAP spectrum  $S(\omega)$ , with a spreading function  $D(\theta)$  for the 3-D case (where  $\omega$  and  $\theta$  are the angular frequency and angle with respect to the positive  $x$  direction).

For the 2-D wave field, we use an initial spectrum  $S(\omega)$  with global steepness  $k_p H_s / 2 = 0.11$ , peak wavenumber  $k_p = 16k_0$  with  $k_0$  the fundamental wavenumber in the computational domain and enhancement factor  $\gamma = 3.3$ . In the (reference, EnKF–HOS and HOS-only) simulations,  $L = 256$  grid points are used in spatial domain  $[0, 2\pi)$ . The noisy measurements  $s_{m,j}$  are generated through (3.1) and (3.2), with a comparison between  $s_{m,0}(x)$  and  $s^{true}(x, t_0)$  shown in figure 3.

Both EnKF–HOS and HOS-only simulations start from initial measurements  $s_{m,0}(x)$ . In the EnKF–HOS method, the ensemble size is set to be  $N = 100$ , and measurements at  $d = 2$  locations of  $x/(2\pi) = 100/256$  and  $170/256$  are assimilated into the model

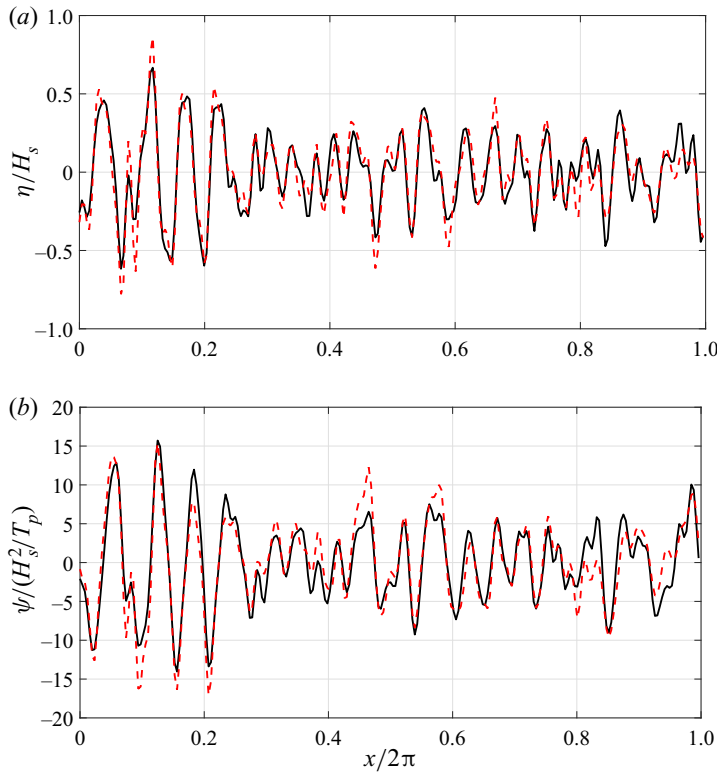


Figure 3. Plots of (a)  $\eta^{true}(x, t_0)$  (—) and  $\eta_{m,0}(x)$  (---, red); (b)  $\psi^{true}(x, t_0)$  (—) and  $\psi_{m,0}(x)$  (---, red).

(assuming from measurements of two buoys) with a constant DA interval  $\tau = t_j - t_{j-1} = T_p/16$ , where  $T_p = 2\pi/\sqrt{k_p}$  from the dispersion relation.

The error  $\epsilon(t; \mathcal{A})$  with  $\mathcal{A} = [0, 2\pi)$  obtained from EnKF–HOS and HOS-only simulations are shown in figure 4. For the HOS-only method, i.e. without DA,  $\epsilon(t; \mathcal{A})$  increases in time from the initial value  $\epsilon(0; \mathcal{A}) \approx 0.05$ , and reaches  $O(1)$  at  $t/T_p \approx 100$ . In contrast,  $\epsilon(t; \mathcal{A})$  from the EnKF–HOS simulation keeps decreasing, and becomes several orders of magnitude smaller than that from the HOS-only method (and two orders of magnitude smaller than the measurement error) at the end of the simulation. For visualization of the wave fields, figure 5 shows snapshots of  $\eta^{true}(x)$  and  $\eta^{sim}(x)$  (with EnKF–HOS and HOS-only methods) at three time instants of  $t/T_p = 5, 45$  and  $95$ , which indicates the much better agreement with  $\eta^{true}(x)$  when DA is applied. Notably, at and after  $t/T_p = 45$ , the EnKF–HOS solution is not visually distinguishable from  $\eta^{true}(x)$ .

The influence of the parameter  $c$  (reflecting the measurement error) on the results from both methods are summarized in table 1. We present the critical time instants  $t^*$  when  $\epsilon(t^*; \mathcal{A})$  reaches  $O(1)$  in the HOS-only method, i.e. when the simulation completely loses the phase information. As expected, all cases lose phase information for sufficiently long time, and the critical time  $t^*$  decreases with the increase of  $c$ . In contrast, for EnKF–HOS method, the error  $\epsilon(t; \mathcal{A})$  decreases with time and reaches  $O(10^{-3})$  at  $t = 100T_p$  in all cases.

We further investigate the effects of EnKF parameters on the performance, including DA interval  $\tau$ , the ensemble size  $N$  and the number of DA locations  $d$ . The errors  $\epsilon(t; \mathcal{A})$  obtained with different parameter values are plotted in figure 6 (for  $\tau$  from  $T_p/16$  to  $T_p/2$ ),

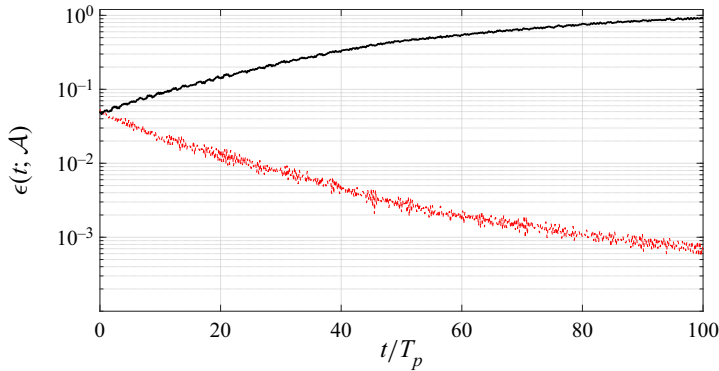


Figure 4. Error  $\epsilon(t; \mathcal{A})$  with  $\mathcal{A} = [0, 2\pi]$  from EnKF-HOS (---, red) and HOS-only (—) methods, for the 2-D idealistic case.

figure 7 (for  $N$  from 40 to 100) and figure 8 (for  $d$  from 1 to 4). In the tested ranges, the performance of EnKF-HOS is generally better (i.e. faster decrease of  $\epsilon(t; \mathcal{A})$  with an increase of  $t$ ) for smaller  $\tau$ , larger  $N$  and larger  $d$ . In addition, for  $\tau = T_p/2$  as shown in figure 6,  $\epsilon(t; \mathcal{A})$  slowly increases with time, indicating a situation that the assimilated data is not sufficient to counteract the deviation of HOS simulation from the true solution (due to the chaotic nature of (2.9) and (2.10)). It is also found that when  $N = 20$ , the error increases with time, mainly due to the filter divergence caused by insufficient ensemble size to capture the error statistics.

Finally, we compare the EnKF-HOS algorithm with the explicit Kalman filter method developed by Yoon *et al.* (2015), where the evolution of the wave field is solved by (2.9) and (2.10) while the propagation of the covariance matrix is linearized. To manifest the contrast between the two approaches, we use as an initial condition a JONSWAP spectrum with a greater global steepness  $k_p H_s/2 = 0.15$ , yet keeping all other parameters the same as the main 2-D idealistic case. Figure 9 compares the errors  $\epsilon(t; \mathcal{A})$  with  $\mathcal{A} = [0, 2\pi]$  from the EnKF-HOS method and the explicit Kalman filter method. Although both errors decrease with time, the EnKF-HOS method shows a faster decreasing rate, with  $\epsilon(100T_p; \mathcal{A})$  approximately one order of magnitude smaller than that from the explicit Kalman filter method.

For the 3-D wave field, we use the same initial spectrum  $S(\omega)$  as in the main 2-D idealistic case, with a direction spreading function

$$D(\theta) = \begin{cases} \frac{2}{\beta} \cos^2\left(\frac{\pi}{\beta}\theta\right), & \text{for } -\frac{\beta}{2} < \theta < \frac{\beta}{2}, \\ 0, & \text{otherwise,} \end{cases} \quad (3.5)$$

where  $\beta = \pi/6$  is the spreading angle. The (reference, EnKF-HOS and HOS-only) simulations are conducted with  $L = 64 \times 64$  grid points. In the EnKF-HOS method,  $N = 100$  ensemble members are used, and data from  $d = 10$  locations (randomly selected with uniform distribution) are assimilated with interval  $\tau = T_p/16$ .

Figure 10 shows the error  $\epsilon(t; \mathcal{A})$  (with  $\mathcal{A} = [0, 2\pi] \times [0, 2\pi]$ ) obtained from the EnKF-HOS and HOS-only methods. Similar to the 2-D case, we see that  $\epsilon(t; \mathcal{A})$  from the EnKF-HOS method decreases with time, and becomes several orders of magnitude smaller than that from the HOS-only method (with the latter increasing with time). A closer scrutiny for error on a snapshot can be obtained by defining a local spatial error

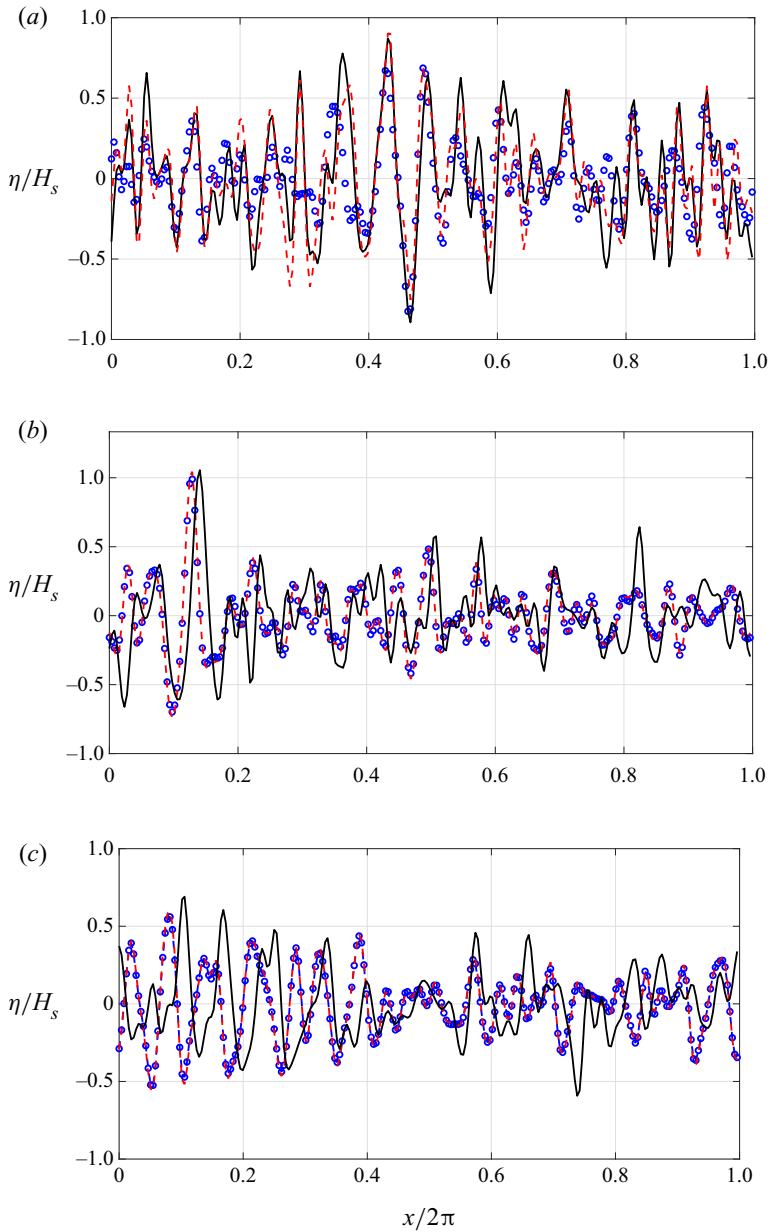


Figure 5. Surface elevations  $\eta^{true}(x)$  ( $\circ$ , blue),  $\eta^{sim}(x)$  with EnKF-HOS (---, red) and HOS-only (—) methods, at (a)  $t/T_p = 5$ , (b)  $t/T_p = 50$  and (c)  $t/T_p = 95$ .

at a time instant  $t$ :

$$e(x; t) = \frac{|\eta^{true}(x, t) - \eta^{sim}(x, t)|}{\sigma_\eta}. \tag{3.6}$$

Three snapshots at  $e(x; t)$  for  $t/T_p = 5, 50$  and  $95$  are shown in [figure 11](#), demonstrating the much smaller error achieved using the EnKF-HOS method especially for large  $t$ , i.e. the superior performance of including DA in the simulation.

	$t^*$	$\epsilon(100T_p; \mathcal{A})$ with EnKF-HOS
$c = 0.0004\sigma_\eta^2$	$150T_p$	$1.65 \times 10^{-3}$
$c = 0.0025\sigma_\eta^2$	$100T_p$	$6.21 \times 10^{-3}$
$c = 0.0100\sigma_\eta^2$	$70T_p$	$7.28 \times 10^{-3}$
$c = 0.0400\sigma_\eta^2$	$40T_p$	$9.02 \times 10^{-3}$

Table 1. Values of  $t^*$  in HOS-only method and  $\epsilon(100T_p; \mathcal{A})$  in EnKF-HOS method for different values of  $c$ .

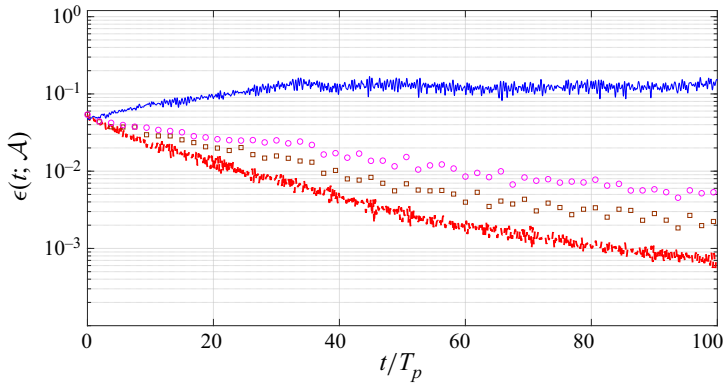


Figure 6. Error  $\epsilon(t; \mathcal{A})$  with  $\mathcal{A} = [0, 2\pi)$  from EnKF-HOS method for  $\tau = T_p/16$  (---, red),  $\tau = T_p/8$  ( $\square$ , brown),  $\tau = T_p/4$  ( $\circ$ , magenta) and  $\tau = T_p/2$  (—, blue). Other parameter values are kept the same as the main 2-D idealistic case.

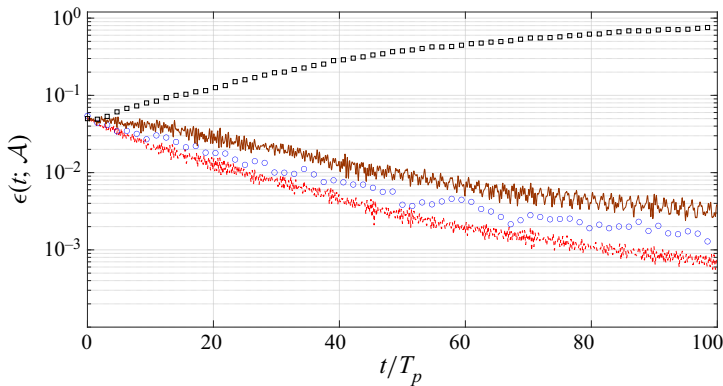


Figure 7. Error  $\epsilon(t; \mathcal{A})$  with  $\mathcal{A} = [0, 2\pi)$  from EnKF-HOS method for  $N = 20$  ( $\square$ ),  $N = 40$  (—, brown),  $N = 70$  ( $\circ$ , blue) and  $N = 100$  (---, red). Other parameter values are kept the same as the main 2-D idealistic case.

### 3.1.2. Results for realistic cases

We consider  $\eta^{true}(\mathbf{x})$  for the realistic case taken from a subregion  $\mathcal{R}$  with quarter edge length of a periodic computational domain  $\mathcal{W}$ , i.e. a patch in the ocean (see figure 12a). The reference simulation in  $\mathcal{W}$  is performed with  $256 \times 256$  grid points, with all other parameters kept the same as the 3-D idealistic simulation. The EnKF-HOS and HOS-only simulations are conducted over  $\mathcal{R} = [0, 2\pi) \times [0, 2\pi)$  with  $L = 64 \times 64$  grid

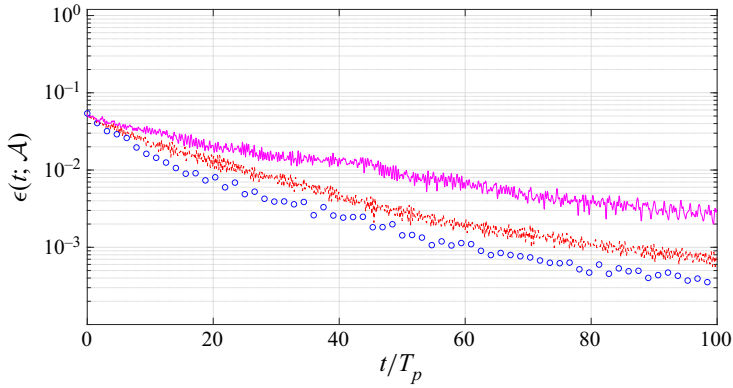


Figure 8. Error  $\epsilon(t; \mathcal{A})$  with  $\mathcal{A} = [0, 2\pi)$  from EnKF-HOS method for  $d = 1$  at  $x/(2\pi) = 100/256$  (—, magenta),  $d = 2$  at  $x/(2\pi) = 100/256$  and  $170/256$  (- - -, red) and  $d = 4$  at  $x/(2\pi) = 100/256, 135/256, 170/256$  and  $205/256$  (o, blue). Other parameter values are kept the same as the main 2-D idealistic case.

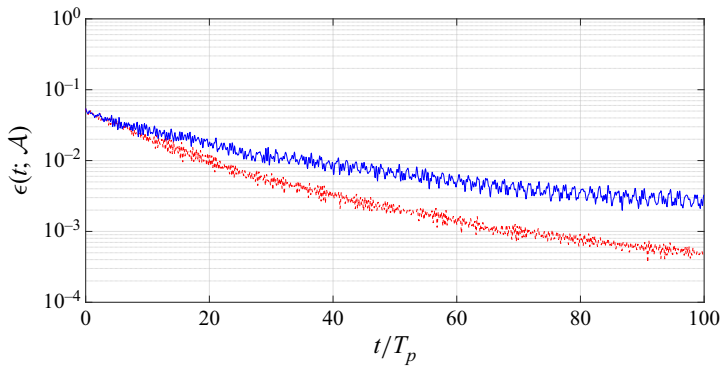


Figure 9. Error  $\epsilon(t; \mathcal{A})$  with  $\mathcal{A} = [0, 2\pi)$  from EnKF-HOS method (- - -, red) and the explicit Kalman filter method by Yoon *et al.* (2015) (—, blue).

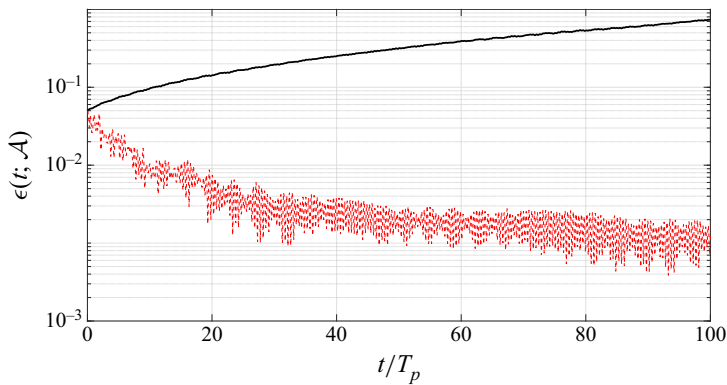


Figure 10. Error  $\epsilon(t; \mathcal{A})$  with  $\mathcal{A} = [0, 2\pi) \times [0, 2\pi)$  from EnKF-HOS (- - -, red) and HOS-only (—) methods, for the 3-D idealistic case.



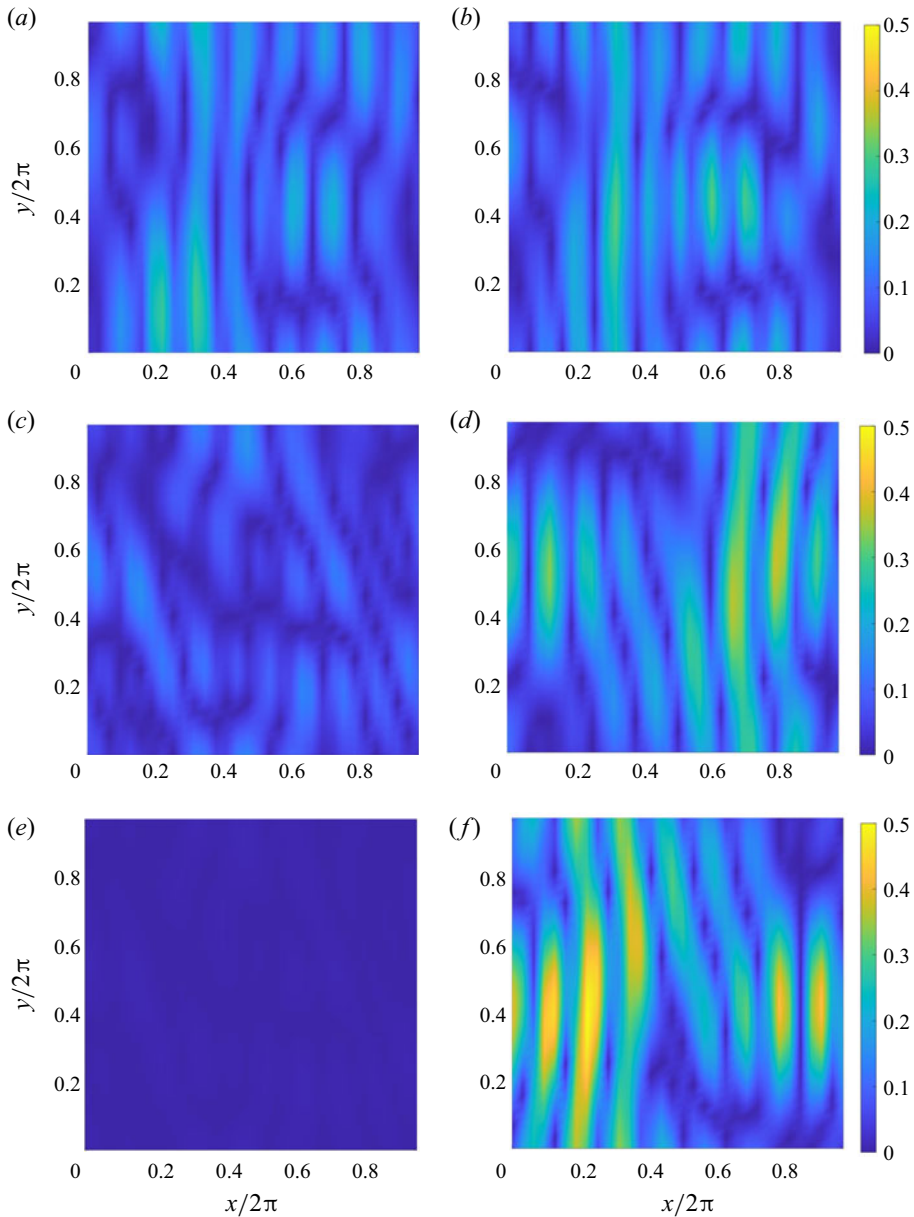


Figure 11. Local spatial error  $e(x; t)$  obtained with EnKF-HOS (a,c,e) and HOS-only (b,d,f) methods at (a,b)  $t/T_p = 5$ , (c,d)  $t/T_p = 50$  and (e,f)  $t/T_p = 95$  for the 3-D idealistic case.

points, starting from initial noisy measurements. For  $j \geq 1$ , We further consider a practical situation where the measurements are obtained from a marine radar and only available in  $\mathcal{M}_j = \mathcal{B}^c \cap \mathcal{R}$ , where  $\mathcal{B} = \{x | x > \pi, \pi/2 < y < 3\pi/2\}$  (say a structure of interest located within  $\mathcal{B}$  preventing the surrounding measurements, see figure 12b). We use  $d = 2176$  (locating on every computational grid point in  $\mathcal{M}_j$ ) and an assimilation interval  $\tau = T_p/4$ .

In this case, the use of modified EnKF analysis (2.18) is critical due to the interplay between  $\mathcal{M}_j$ ,  $\mathcal{P}_j$  and  $\mathcal{U}_j$ . In particular, the left, upper/lower bounds of  $\mathcal{P}(t)$  moves (towards

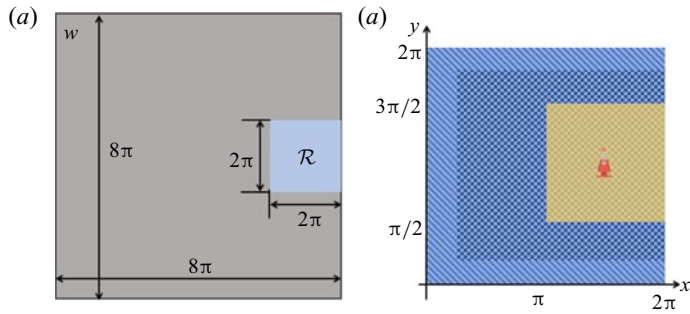


Figure 12. Schematic illustration of spatial domains: (a) periodic computational domain  $\mathcal{W}$  for the reference simulation and subregion  $\mathcal{R}$  for EnKF–HOS and HOS-only simulations; (b) computational region  $\mathcal{R}$  with measurement zone  $\mathcal{M}_j$  marked by blue, region  $\mathcal{B} = \{x|x > \pi, \pi/2 < y < 3\pi/2\}$  with no data available by yellow, predictable zone  $\mathcal{P}_j$  by checker board, and unpredictable zone  $\mathcal{U}_j$  by downward diagonal stripes.

right, down/up), respectively, with speeds  $c_{g,x}^{max}$  and  $c_{g,y}^{max}$ , i.e. the maximum group speeds up in the  $x$  and  $y$  directions (corresponding to the group speed of mode  $\mathbf{k} = (1, 1)$ ). After applying the modified EnKF analysis equation (2.18), which takes into consideration of  $\mathcal{M}_j$ ,  $\mathcal{P}_j$  and  $\mathcal{U}_j$  (see a sketch in figure 12b),  $\mathcal{P}_j$  is recovered to fill in  $\mathcal{R}$  due to the DA. In contrast, in the HOS-only method,  $\mathcal{P}(t) \equiv \mathcal{P}^*(t)$  keeps shrinking and vanishes for sufficient time. We consider two error metrics  $\epsilon(t; \mathcal{R})$  and  $\epsilon(t; \mathcal{P}^*(t))$ , which are plotted in figure 13 for both EnKF–HOS and HOS-only methods. For HOS-only simulation,  $\epsilon(t; \mathcal{R})$  increases rapidly in time and reaches  $O(1)$  at  $t/T_p \sim O(3)$ . This is resulted from the chaotic nature of (2.9) and (2.10), as well as the larger error in  $\mathcal{U}(t)$ . For EnKF–HOS simulation,  $\epsilon(t; \mathcal{R})$  decreases with time and reaches a constant level of  $O(0.002)$  after  $t/T_p \sim O(3)$ . The further reduction of the error is prohibited due to the region  $\mathcal{U}(t)$ , because  $\epsilon(t; \mathcal{U}(t))$  has a lower bound from the measurement error. The general trend of  $\epsilon(t; \mathcal{P}^*(t))$  is similar to  $\epsilon(t; \mathcal{R})$ , but the magnitude of  $\epsilon(t; \mathcal{P}^*(t))$  is smaller than that of  $\epsilon(t; \mathcal{R})$  for both methods. In particular, the growth rate of  $\epsilon(t; \mathcal{P}^*(t))$  for the HOS-only method is comparable to those in 2-D/3-D idealistic cases. This is due to the removal of  $\mathcal{U}(t)$  from  $\mathcal{A}$  in the computation of the error.

To further understand the error characteristics, we plot the local spatial error  $e(\mathbf{x}; t)$  at  $t/T_p = 2.875, 4.875$  and  $7.25$  for both methods in figure 14. While the spatial error generally increases in time for the HOS-only method,  $e(\mathbf{x}; t)$  from the EnKF–HOS method is significantly smaller and exhibits a heterogeneous spatial distribution. Within  $\mathcal{U}_j$ ,  $e(\mathbf{x}; t)$  is relatively high with the same order of the measurement error. In  $\mathcal{P}_j$ ,  $e(\mathbf{x}; t)$  decreases with time and becomes significantly lower than that in  $\mathcal{U}_j$ . Remarkably, this also applies to the region where measurements are not available (i.e.  $\mathcal{M}^c \cap \mathcal{R}$ ) as the waves in this region travel from upstream locations where DA is performed. This result is of practical importance as it shows that the wave forecast at a location of interest in the ocean (say the location of an offshore structure) can be made accurate through DA in the upstream region.

### 3.2. Results with real radar measurements

In this section, we test the performance of the EnKF–HOS method with real radar measurements of the ocean wave field. The measurements are obtained from a shipborne X-band (9.4 GHz) Doppler coherent marine radar off the coast of Southern California.

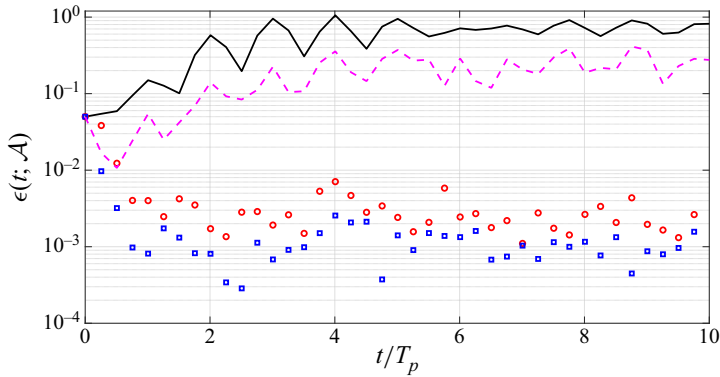


Figure 13. The error metric  $\epsilon(t; \mathcal{R})$  obtained from EnKF-HOS ( $\circ$ , red) and HOS-only (—) methods; and  $\epsilon(t; \mathcal{P}^*(t))$  obtained from EnKF-HOS ( $\square$ , blue) and HOS-only (---, magenta) methods, for the 3-D realistic case.

We consider a patch from the radar-scanned area as our initial domain of interest  $\mathcal{R}_0$ , which covers a  $480 \text{ m} \times 480 \text{ m}$  area resolved on a  $64 \times 64$  grid (figure 15). The starting time of computation is 2013-09-13T01 : 00 : 13Z, with a global wave steepness  $k_p H_s / 2 = 0.027$  from the initial radar data. The direction of waves is determined for each mode (travelling at its phase speed) by finding the direction which correlates better with two sequential snapshots of surface elevation (which gives consistent direction for almost all the modes). In both EnKF-HOS and HOS-only simulations, we rescale the computational domain  $\mathcal{R}_j$  to  $[0, 2\pi) \times [0, 2\pi)$  and use  $L = 64 \times 64$  grid points. For EnKF, we use  $d = 64 \times 64$ , which covers the whole patch, and set the DA interval the same as radar data collection interval, which fluctuates in time around  $T_p/4 = 2.82 \text{ s}$ . The ensemble size is set to be  $N = 100$ . For adaptive inflation, we use  $\lambda_0 = 1$  in the prior distribution of  $\lambda_0$  (see Appendix A) to sequentially determine  $\lambda_j$  in (2.15), which is applied at each  $t = t_j$  together with the localization (2.16).

A critical issue in this case is the movement of  $\mathcal{M}_j$  in time due to the ship speed (around  $0.2 \text{ m s}^{-1}$ ), which results in a mismatch between  $\mathcal{M}_j$  and the computational region  $\mathcal{R}_{j-1}$  at each  $t = t_j$ . To address this issue, we shift the computational region from  $\mathcal{R}_{j-1}$  to  $\mathcal{R}_j$  which matches  $\mathcal{M}_j$ . In the EnKF-HOS method, we further partition  $\mathcal{R}_j$  into  $\mathcal{P}_j$  and  $\mathcal{U}_j$  (using the predictable zone calculated from  $\mathcal{R}_{j-1}$ ) and apply the modified analysis equation (2.18) accordingly. In the HOS-only method, we use Fourier periodic extension (Grafakos 2008) to obtain the wave field covering  $\mathcal{R}_j$ .

Since the true solution is not available in this case, we directly use  $\rho_{\mathcal{A}}(\eta_{m,j}, \eta^{sim})$  (see (3.4)) as the metric to evaluate the performance. Figure 16 plots  $\rho_{\mathcal{A}}(\eta_{m,j}, \eta^{sim})$  with  $\mathcal{A} = \mathcal{R}_j$  and  $\mathcal{A} = \mathcal{P}^*(t)$  (for which the shrinking speed is still taken as the group speed of mode  $\mathbf{k} = (1, 1)$ ) obtained from EnKF-HOS and HOS-only methods. While all time series start from 1 at  $t = t_0$ ,  $\rho_{\mathcal{R}_j}(\eta_{m,j}, \eta^{sim})$  and  $\rho_{\mathcal{P}^*(t)}(\eta_{m,j}, \eta^{sim})$  from the HOS-only simulation quickly approach  $O(0.25)$  and  $O(0.40)$  within one peak period  $T_p$ , indicating the (almost) complete loss of the phase information. This is much faster than any synthetic case, mainly due to the under-resolved physics in (2.9) and (2.10) with respect to the real ocean (which includes extra physical effects of current, wind, etc.). Furthermore,  $\rho_{\mathcal{R}_j}(\eta_{m,j}, \eta^{sim})$  from the HOS-only method is lower in comparison with that reported in Lyzenga *et al.* (2015) (approximately  $0.5 \sim 0.75$  after  $10T_p$ , which we have confirmed using our code). This is mainly due to the different region used in Lyzenga *et al.* (2015), which may feature less

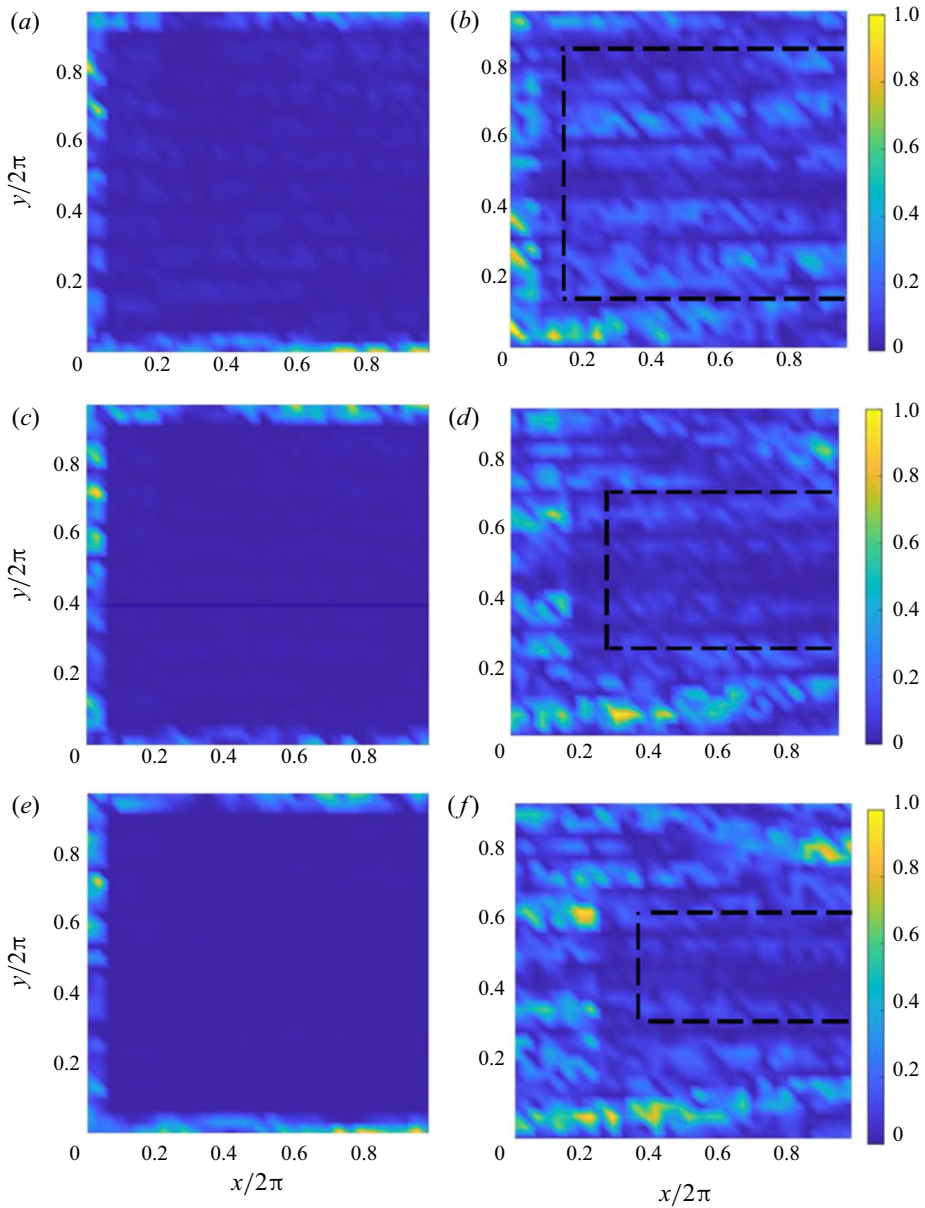


Figure 14. Local spatial error  $e(x; t)$  obtained from (a,c,e) EnKF-HOS and (b,d,f) HOS-only methods for the 3-D realistic case, at (a,b)  $t/T_p = 2.875$ , (c,d)  $t/T_p = 4.875$  and (e,f)  $t/T_p = 7.25$ . The regions bounded by the black dash lines represent  $\mathcal{P}^*(t)$  with the HOS-only method.

significant effects of current and wind, as well as smaller radar error due to its smaller distance from the ship. In contrast to results from the HOS-only method, the correlation  $\rho_{\mathcal{R}_j}(\eta_{m,j}, \eta^{sim})$  and  $\rho_{\mathcal{P}^*(t)}(\eta_{m,j}, \eta^{sim})$  from EnKF-HOS remain at  $O(0.75)$  (with the latter slightly greater than the former), retaining the phase information for arbitrarily long time. Figure 17 further plots the snapshots of  $\eta_{m,j}(\mathbf{x})$  and  $\eta^{sim}(\mathbf{x}, t)$  at two cross-sections of  $y/(2\pi) = 1/3$  and  $2/3$  for both methods at three time instants of  $t/T_p = 1, 5$  and  $9$ . It can

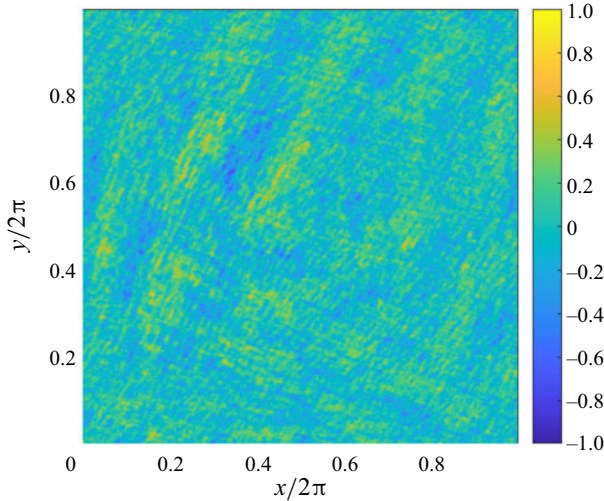


Figure 15. Initial surface elevation  $\eta_{m,0}(\mathbf{x})/H_s$  (with  $H_s = 1.70$  m) measured by radar at  $t = t_0$ , i.e. 2013-09-13T01 : 00 : 13Z.

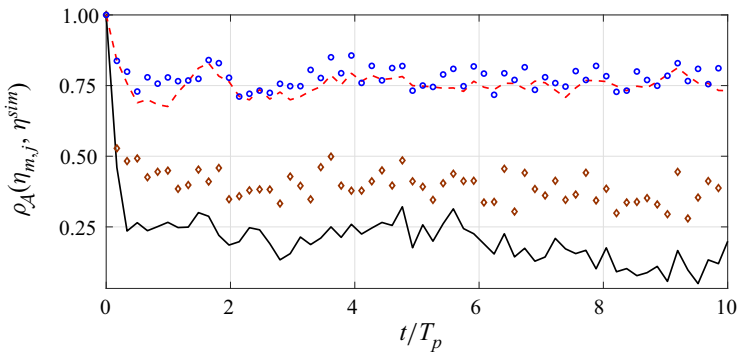


Figure 16. Time series of  $\rho_{R_j}(\eta_{m,j}, \eta^{sim})$  from EnKF-HOS (---, red) and HOS-only (—) methods; and  $\rho_{P^*(t)}(\eta_{m,j}, \eta^{sim})$  from EnKF-HOS ( $\circ$ , blue) and HOS-only ( $\diamond$ , brown) methods.

be visually observed that the EnKF-HOS results are (on average) much closer to  $\eta_{m,j}(\mathbf{x})$  for all cases.

We finally test the effect of parameter  $\bar{\lambda}_0$  on the performance of EnKF-HOS. In general, the value of  $\bar{\lambda}_0$  can be considered as a control of the extent to which the inflation is applied. For larger values of  $\bar{\lambda}_0$ , it is expected that the ensemble variance of (2.11) is amplified to a greater extent, and more weights are assigned to measurements when the analysis (2.18) is applied. Physically, larger values of  $\bar{\lambda}_0$  may be chosen if the model is associated with significant under-represented physics (thus severely underestimates the variance in (2.11)). To elucidate this effect of  $\bar{\lambda}_0$ , we test another value of  $\bar{\lambda}_0 = 2$ , and compare the resulting  $\rho_{R_j}(\eta_{m,j}, \eta^{sim})$  with that from  $\bar{\lambda}_0 = 1$  in figure 18. Indeed, the result for  $\bar{\lambda}_0 = 2$  shows a

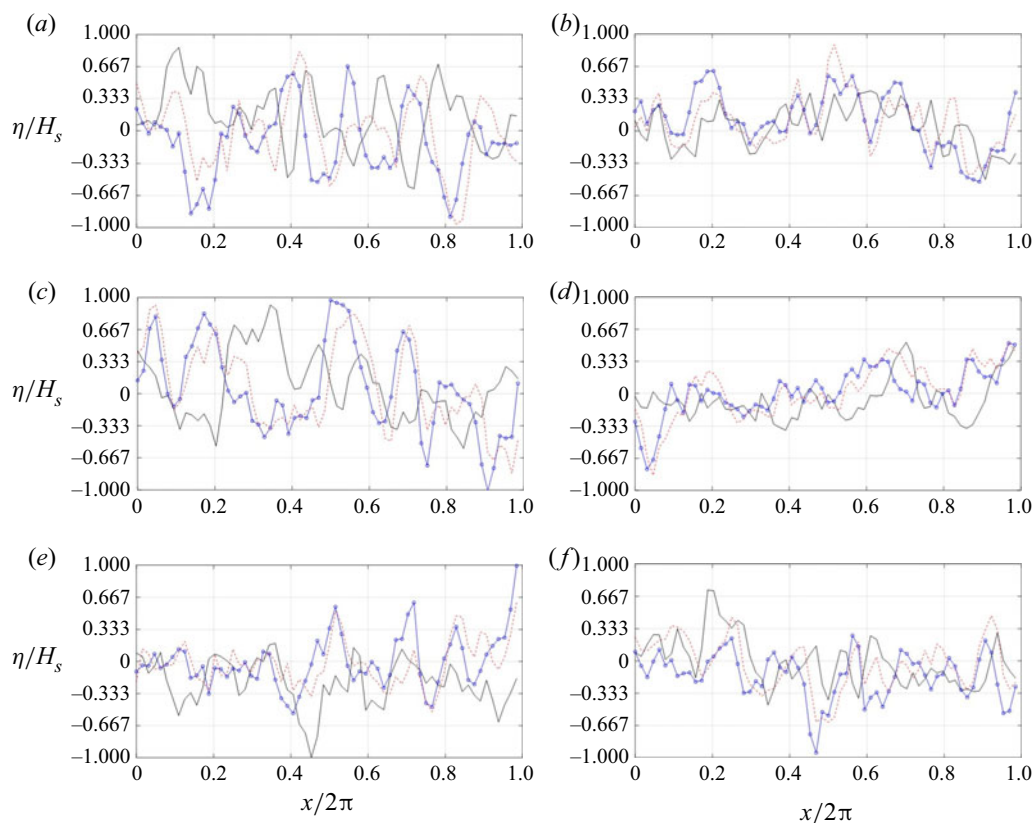


Figure 17. Snapshots of  $\eta_{m,j}(x)$  ( $\ominus$ , blue) and  $\eta^{sim}(x, t)$  from EnKF-HOS (---, red) and HOS-only (—) methods, at two cross-sections (a,c,e)  $y/(2\pi) = 1/3$  and (b,d,f)  $y/(2\pi) = 2/3$ , and time instants (a,b)  $t/T_p = 1$ , (c,d)  $t/T_p = 5$  and (e,f)  $t/T_p = 9$ .

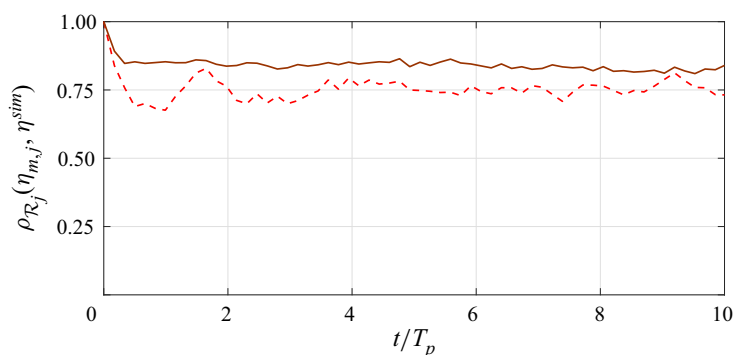


Figure 18. Time series of  $\rho_{R_j}(\eta_{m,j}, \eta^{sim})$  with  $\bar{\lambda}_0 = 1.0$  (---, red) and  $\bar{\lambda}_0 = 2.0$  (—, brown).

higher correlation with the measurement, with  $\rho_{R_j}$  above  $O(0.8)$  for all time. We note that the higher value of  $\rho_{R_j}$  does not imply the closeness of  $\eta^{sim}$  to  $\eta^{true}$ , since the relation between  $\eta_{m,j}$  and  $\eta^{true}$  is not known in this case.

#### 4. Conclusions

In this paper, we develop the ensemble-based DA capability for phase-resolved wave forecast, resulting in a new EnKF–HOS method. A unique consideration in EnKF–HOS is the treatment of the interplay between predictable and measurement zones, which is successfully accounted for through a modified analysis equation. The performance of the EnKF–HOS method is extensively tested and compared with the (traditional) HOS-only method using both synthetic wave fields and real radar data. In all cases, significant advantages are demonstrated by using the EnKF–HOS method, namely the dramatic reduction of forecast error and retaining the phase information for an arbitrarily long time with DA of radar data. In contrast, the phase information is lost within one peak period in the HOS-only method when considering the real ocean waves. The parameters involved in the EnKF–HOS method are carefully benchmarked, including the ensemble size, DA interval, number of DA locations and the inflation factor. The developed EnKF–HOS algorithm is intrinsically parallel and very suitable for implementation on a graphics processing unit (GPU), a compact device that can be conveniently installed in the offshore environment (our ongoing work). Finally, the accuracy of the EnKF–HOS method can be further improved given deeper knowledge about the radar measurement error, which we recommend as one of the next essential focuses for the remote sensing community.

**Acknowledgements.** We would like to thank Dr D. Lyzenga, Dr O. Nwogu and Dr R. Beck in the Department of Naval Architecture and Marine Engineering at the University of Michigan, for providing and interpreting the radar data. We also thank Dr P. Hess from the Office of Naval Research to grant us permission to use this data set for research.

**Funding.** The research is funded by the Michigan Institute for Computational Discovery and Engineering (MICDE) through a 2019 Catalyst Grant. The computational resources are provided by NSF-XSEDE grant TG-CTS200016.

**Declaration of interests.** The authors report no conflict of interest.

**Author ORCIDs.**

 Yulin Pan <https://orcid.org/0000-0002-7504-8645>.

#### Appendix A. Adaptive inflation

We apply the adaptive inflation algorithm developed by Anderson (2007) to determine the values of  $\lambda_j$  in (2.15). The algorithm is applied at each  $t = t_j$  (for  $j \geq 1$ ), and we shall drop the subscript  $j$  on other variables in the following description for simplicity.

The key idea in adaptive inflation is to consider  $\lambda$  as an additional state variable and update its value through Bayes' theorem given the measurements

$$p(\lambda|\eta_m(\mathbf{x})) \sim p(\lambda)p(\eta_m(\mathbf{x})|\lambda), \quad (\text{A1})$$

where  $p(\lambda)$  is the (given) prior distribution and  $p(\eta_m(\mathbf{x})|\lambda)$  is the likelihood function. In principle, our purpose is to find the optimal  $\lambda$  which maximizes the posterior  $p(\lambda|\eta_m(\mathbf{x})) = p(\lambda|\eta_m(1), \dots, \eta_m(d))$ . Intuitively, the inflation (2.15) using such an optimal  $\lambda$  provides sufficient variance of the forecast ensemble to cover the measurements, and thus avoid the overconfidence in the forecast when analysis is performed.

By assuming independent measurement errors, it can be shown that the full Bayes' problem (A1) is equivalent to the sequential problem (which saves significant

computational cost, see Anderson (2007) for details)

$$p(\lambda|\eta_m(1)) = p(\eta_m(1)|\lambda)p(\lambda)/\mathcal{Z}_1, \tag{A2}$$

$$p(\lambda|\eta_m(1), \eta_m(2), \dots, \eta_m(i)) = p(\eta_m(i)|\lambda)p(\lambda|\eta_m(1), \eta_m(2), \dots, \eta_m(i-1))/\mathcal{Z}_i, \tag{A3}$$

where (A3) is applied sequentially for  $i = 2, \dots, d$ , and  $\mathcal{Z}_i$  are the normalization factors (which do not play a role in the computation).

In computing (A2) and (A3), we use a Gaussian prior distribution,

$$p(\lambda) = \mathcal{N}(\bar{\lambda}_0, \sigma_0^2), \tag{A4}$$

with mean  $\bar{\lambda}_0$  and variance  $\sigma_0^2$  (say  $\bar{\lambda}_0 = 1$  and  $\sigma_0^2 = c\bar{\lambda}_0^2/H_s^2$  at  $j = 1$ , with the situation of  $j > 1$  discussed at the end of Appendix A). The key computations in (A2) and (A3) are the likelihood functions  $p(\eta_m(i)|\lambda)$  for  $i = 1, 2, \dots, d$ , which will be discussed below.

Given a value of  $\lambda$ , a member in the forecast ensemble at measurement location  $i$  is denoted by  $[\mathbf{G}\eta_f^{(n),inf}](i)$  with  $\eta_f^{(n),inf}$  given by (2.15). This ensemble is assumed to have a Gaussian distribution consistent with the EnKF framework, with mean  $[\mathbf{G}\bar{\eta}_f](i)$  and variance  $\sigma_f(i)^2 = \lambda[\mathbf{G}\mathbf{Q}_\eta\mathbf{G}^T](i, i)$  (note that only  $\sigma_f(i)^2$  is affected by the inflation). Let  $D = \eta_m(i) - [\mathbf{G}\bar{\eta}_f](i)$  be the distance between the mean forecast and measurement at location  $i$ , which is drawn from a zero-mean Gaussian random variable  $\mathcal{D}$  with variance  $\theta_i^2 = \sigma_f(i)^2 + \mathbf{R}_\eta(i, i)$  (here we consider the summation of two independent Gaussian random variables and assume that both the measurement and forecast are unbiased). It follows that

$$p(\eta_m(i)|\lambda) = p(\mathcal{D} = D|\lambda) = \frac{1}{\sqrt{2\pi\theta_i}} \exp\left(\frac{-D^2}{2\theta_i^2}\right). \tag{A5}$$

Therefore, the posterior distribution in each equation of (A2) and (A3) can be formulated as

$$p(\lambda|\eta_m(1), \dots, \eta_m(i)) = \frac{1}{2\pi\theta_i\sigma_{i-1}} \exp\left(-\frac{D^2}{2\theta_i^2} - \frac{(\lambda - \bar{\lambda}_{i-1})^2}{2\sigma_{i-1}^2}\right) / \mathcal{Z}_i. \tag{A6}$$

The formulation (and sequential computation) of (A6) for  $i = 1, 2, \dots, d$  require each posterior  $p(\lambda|\eta_m(1), \dots, \eta_m(i))$  in (A2) and (A3) (and thus the prior for the next sequential equation) to be approximated by Gaussian distribution  $\mathcal{G}(\lambda) \sim \mathcal{N}(\bar{\lambda}_i, \sigma_i^2)$ . We follow Anderson (2007) to set the  $\bar{\lambda}_i$  as the mode of  $p(\lambda|\eta_m(1), \dots, \eta_m(i))$ , and compute  $\sigma_i$  by considering  $\Gamma = p(\bar{\lambda}_i|\eta_m(1), \dots, \eta_m(i))/p(\bar{\lambda}_i + \sigma_i|\eta_m(1), \dots, \eta_m(i)) = \mathcal{G}(\bar{\lambda}_i)/\mathcal{G}(\bar{\lambda}_i + \sigma_i)$ , i.e. the same decay rate of the distribution, which gives  $\sigma_i^2 = -(\sigma_{i-1}^2/2) \ln \Gamma$ .

Finally, we use  $\lambda = \bar{\lambda}_d$  in (2.15) for inflation at  $t = t_j$ , and set  $\mathcal{N}(\bar{\lambda}_d, \sigma_d^2)$  computed at time  $t = t_j$  as the prior  $p(\lambda)$  at  $t = t_{j+1}$ . The computations of (A2) and (A3) are repeated, which completes the full algorithm to determine  $\lambda$  by adaptive inflation at each  $t_j$ .



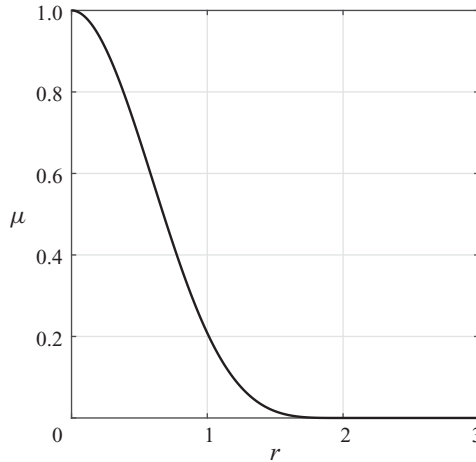


Figure 19. Gaspari–Cohn correlation function  $\mu(r)$  used in the localization.

### Appendix B. Gaspari–Cohn function

The local-correlation function  $\mu$  in the localization equation (2.16) is defined as the Gaspari–Cohn function (Gaspari & Cohn 1999), given by

$$[\mu]_{ij} = \begin{cases} 1 - \frac{5}{3}r^2 + \frac{5}{8}r^3 + \frac{1}{2}r^4 - \frac{1}{4}r^5, & \text{for } 0 \leq r < 1, \\ 4 - 5r + \frac{5}{3}r^2 + \frac{5}{8}r^3 - \frac{1}{2}r^4 + \frac{1}{12}r^5 - \frac{2}{3r}, & \text{for } 1 \leq r < 2, \\ 0, & \text{for } r \geq 2, \end{cases} \quad (\text{B1})$$

where  $r = |\mathbf{x}_i - \mathbf{x}_j|/(\sqrt{3}a/2)$ , with  $a$  taking the same value as in (2.3). A plot of  $\mu(r)$  is provided in figure 19.

### REFERENCES

- ALMEIDA, S., RUSU, L. & GUEDES SOARES, C. 2016 Data assimilation with the ensemble kalman filter in a high-resolution wave forecasting model for coastal areas. *J. Oper. Oceanogr.* **9** (2), 103–114.
- ANDERSON, J.L. 2007 An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A* **59** (2), 210–224.
- ANDERSON, J.L. & ANDERSON, S.L. 1999 A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weath. Rev.* **127** (12), 2741–2758.
- ANNENKOV, S.Y. & SHRIRA, V.I. 2001 On the predictability of evolution of surface gravity and gravity–capillary waves. *Physica D* **152**, 665–675.
- ARAGH, S. & NWOGU, O. 2008 Variation assimilating of synthetic radar data into a pseudo-spectral wave model. *J. Coast. Res.* **10052**, 235–244.
- BANNISTER, R.N. 2017 A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **143** (703), 607–633.
- BLONDEL, E., *et al.* 2010 Experimental validation of deterministic non-linear wave prediction schemes in 2D. In *The Twentieth International Offshore and Polar Engineering Conference*. International Society of Offshore and Polar Engineers.
- BOOIJ, N.R.R.C., RIS, R.C. & HOLTHUIJSEN, L.H. 1999 A third-generation wave model for coastal regions: 1. Model description and validation. *J. Geophys. Res.* **104** (C4), 7649–7666.
- BROAD, W.J. 2006 *Rogue giants at sea*, vol. 11. The New York Times.
- BURGERS, G., JAN VAN LEEUWEN, P. & EVENSEN, G. 1998 Analysis scheme in the ensemble kalman filter. *Mon. Weath. Rev.* **126** (6), 1719–1724.
- CARRASSI, A., BOCQUET, M., BERTINO, L. & EVENSEN, G. 2018 Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *Wiley Interdiscip. Rev.* **9** (5), e535.
- CRAIG, W. & SULEM, C. 1993 Numerical simulation of gravity waves. *J. Comput. Phys.* **108** (1), 73–83.

- DESMARS, N., PÉRIGNON, Y., DUCROZET, G., GUÉRIN, C.-A., GRILLI, S.T. & FERRANT, P. 2018 Phase-resolved reconstruction algorithm and deterministic prediction of nonlinear ocean waves from spatio-temporal optical measurements. In *International Conference on Offshore Mechanics and Arctic Engineering*, vol. 51272, p. V07BT06A054. American Society of Mechanical Engineers.
- DOMERMUTH, D.G. & YUE, D.K.P. 1987 A high-order spectral method for the study of nonlinear gravity waves. *J. Fluid Mech.* **184**, 267–288.
- EMMANOUIL, G., GALANIS, G. & KALLOS, G. 2012 Combination of statistical kalman filters and data assimilation for improving ocean waves analysis and forecasting. *Ocean Model.* **59**, 11–23.
- EVENSEN, G. 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.* **99** (C5), 10143–10162.
- EVENSEN, G. 2003 The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53** (4), 343–367.
- EVENSEN, G. 2009 *Data Assimilation: The Ensemble Kalman Filter*. Springer Science & Business Media.
- FUJIMOTO, W. & WASEDA, T. 2020 Ensemble-based variational method for nonlinear inversion of surface gravity waves. *J. Atmos. Ocean. Technol.* **37** (1), 17–31.
- GALLEGO, G., YEZZI, A., FEDELE, F. & BENETAZZO, A. 2011 A variational stereo method for the three-dimensional reconstruction of ocean waves. *IEEE Trans. Geosci. Remote Sens.* **49** (11), 4445–4457.
- GASPARI, G. & COHN, S.E. 1999 Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125** (554), 723–757.
- GRAFAKOS, L. 2008 *Classical Fourier Analysis*, vol. 2. Springer.
- HAMILL, T.M. & WHITAKER, J.S. 2005 Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Weath. Rev.* **133** (11), 3132–3147.
- HANSEN, J.A. 2002 Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Weath. Rev.* **130** (10), 2373–2391.
- HILMER, T. & THORNHILL, E. 2015 Observations of predictive skill for real-time deterministic sea waves from the wamos II. In *OCEANS 2015-MTS/IEEE Washington*, pp. 1–7. IEEE.
- HOUTEKAMER, P.L., MITCHELL, H.L., PELLERIN, G., BUEHNER, M., CHARRON, M., SPACEK, L. & HANSEN, B. 2005 Atmospheric data assimilation with an ensemble kalman filter: results with real observations. *Mon. Weath. Rev.* **133** (3), 604–620.
- JANSSEN, P.A.E.M. 2008 Progress in ocean wave forecasting. *J. Comput. Phys.* **227** (7), 3572–3594.
- KÖLLISCH, N., BEHRENDT, J., KLEIN, M. & HOFFMANN, N. 2018 Nonlinear real time prediction of ocean surface waves. *Ocean Engng* **157**, 387–400.
- KOMEN, G.J., CAVALERI, L., DONELAN, M., HASSELMANN, K., HASSELMANN, S. & JANSSEN, P.A.E.M. 1996 *Dynamics and modelling of ocean waves*. Cambridge University Press.
- LORENC, A.C. 2003 The potential of the ensemble kalman filter for nwp—a comparison with 4d-var. *Q. J. R. Meteorol. Soc.* **129** (595), 3183–3203.
- LYZENGA, D.R., NWOGU, O.G., BECK, R.F., O'BRIEN, A., JOHNSON, J., DE PAOLO, T. & TERRILL, E. 2015 Real-time estimation of ocean wave fields from marine radar data. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3622–3625. IEEE.
- NAAIJEN, P., TRULSEN, K. & BLONDEL-COUPRIE, E. 2014 Limits to the extent of the spatio-temporal domain for deterministic wave prediction. *Intl Shipbuild. Progr.* **61** (3–4), 203–223.
- NIKOLKINA, I., DIDENKULOVA, I., PELINOVSKY, E. & LIU, P. 2011 Rogue waves in 2006–2010. *Nat. Hazards Earth Syst. Sci.* **11** (11), 2913–2924.
- NOUGUIER, F., GRILLI, S.T. & GUÉRIN, C.-A. 2013 Nonlinear ocean wave reconstruction algorithms based on simulated spatiotemporal data acquired by a flash lidar camera. *IEEE Trans. Geosci. Remote Sens.* **52** (3), 1761–1771.
- NWOGU, O.G. & LYZENGA, D.R. 2010 Surface-wavefield estimation from coherent marine radars. *IEEE Geosci. Remote Sens. Lett.* **7** (4), 631–635.
- PAN, Y., LIU, Y. & YUE, D.K.P. 2018 On high-order perturbation expansion for the study of long–short wave interactions. *J. Fluid Mech.* **846**, 902–915.
- PINTO, J.P., BERNADINO, M.C. & PIRES SILVA, A. 2005 A kalman filter application to a spectral wave model. *Nonlinear Process. Geophys.* **12** (6), 775–782.
- QI, Y., WU, G., LIU, Y., KIM, M.-H. & YUE, D.K.P. 2018a Nonlinear phase-resolved reconstruction of irregular water waves. *J. Fluid Mech.* **838**, 544.
- QI, Y., WU, G., LIU, Y. & YUE, D.K.P. 2018b Predictable zone for phase-resolved reconstruction and forecast of irregular waves. *Wave Motion* **77**, 195–213.
- QI, Y., XIAO, W. & YUE, D.K.P. 2016 Phase-resolved wave field simulation calibration of sea surface reconstruction using noncoherent marine radar. *J. f Atmos. Ocean. Technol.* **33** (6), 1135–1149.

*Phase-resolved ocean wave forecast with data assimilation*

- QI, Y., *et al.* 2017 Phase-resolved reconstruction and forecast of nonlinear irregular wave field based on direct numerical simulations. PhD thesis, Massachusetts Institute of Technology.
- REICHERT, K., HESSNER, K., DANNENBERG, J. & TRÄNKMANN, I. 2004 X-band radar as a tool to determine spectral and single wave properties. In *25th International Conference on Offshore Mechanics and Arctic Engineering*, pp. 683–688. American Society of Mechanical Engineers Digital Collection.
- STREDULINSKY, D.C. & THORNHILL, E.M. 2011 Ship motion and wave radar data fusion for shipboard wave measurement. *J. Ship Res.* **55** (2), 73–85.
- TOLMAN, H.L., *et al.* 2009 User manual and system documentation of wavewatch III TM version 3.14. *Tech. Note, MMAB Contrib.* **276**, 220.
- WANG, G. & PAN, Y. 2020 Data assimilation for phase-resolved ocean wave forecast. In *ASME 2020 39th International Conference on Ocean, Offshore and Arctic Engineering*. American Society of Mechanical Engineers Digital Collection.
- WEST, B.J., BRUECKNER, K.A., JANDA, R.S., MILDER, D.M. & MILTON, R.L. 1987 A new numerical method for surface hydrodynamics. *J. Geophys. Res.* **92** (C11), 11803–11824.
- XIAO, W. 2013 Study of directional ocean wavefield evolution and rogue wave occurrence using large-scale phase-resolved nonlinear simulations. PhD thesis, Massachusetts Institute of Technology.
- XU, L. & GUYENNE, P. 2009 Numerical simulation of three-dimensional nonlinear water waves. *J. Comput. Phys.* **228** (22), 8446–8466.
- YOON, S., KIM, J. & CHOI, W. 2015 An explicit data assimilation scheme for a nonlinear wave prediction model based on a pseudo-spectral method. *IEEE J. ocean. Engng* **41** (1), 112–122.
- ZAKHAROV, V.E. 1968 Stability of periodic waves of finite amplitude on the surface of a deep fluid. *J. Appl. Mech. Tech. Phys.* **9** (2), 190–194.