Applied Mathematical Modelling 34 (2010) 1984–1999

Contents lists available at ScienceDirect

Applied Mathematical Modelling

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

Non-linear wave data assimilation with an ANN-type wind-wave model and Ensemble Kalman Filter (EnKF)

Ahmadreza Zamani [a,*], Ahmadreza Azimian [a], Arnold Heemink [b,1], Dimitri Solomatine [c]

[a] *Mechanical Engineering Department, Isfahan University of Technology, Isfahan, Iran*
[b] *Delft Institute of Applied Mathematics, Delft University of Technology, The Netherlands*
[c] *Core of Hydroinformatics, UNESCO-IHE Institute for Water Education, and Section of Water Management, Delft University of Technology, The Netherlands*

ARTICLE INFO

Article history:
Received 28 January 2009
Received in revised form 5 October 2009
Accepted 13 October 2009
Available online 20 October 2009

Keywords:
Wave forecasting
Data assimilation
Surrogate model
Ensemble Kalman Filter
Wind-waves

ABSTRACT

Non-linear data assimilation for a wind-wave dynamical surrogate model in a reduced space is presented. A dynamic artificial neural network is used for surrogate modeling. It provides a fast emulation of a wind-wave model which is used for the evaluation of the system states during a small period of time. The system state consists of wave height and wave direction in the reduced space which is affected by the reduced space wind field. The projection from the full space to the reduced one is performed by a principal component analysis. Ensemble methods require the evaluation of dynamics for a large number of statistical ensembles, so coupling this surrogate (instead of a full model) with an Ensemble Kalman Filter (EnKF) leads to computational efficiency. Application of the procedure is demonstrated through 6 month hindcast study of wind waves over the Caspian Sea using the third-generation wave model and the analysis of the ECMWF wind field. The trained network is embedded into the stochastic environment. Then, the EnKF is used to find estimate of the system states. Experiments show that the proposed data assimilation technique can correct the prediction of the wind-waves requiring just a modest execution time.

1. Introduction

Wave forecasting is important for a wide range of marine activities in oceanic and coastal seas. The dynamics of ocean waves have been studied for a long time. Basics of the wave theory were developed in the 19th century, but it was not until the 20th century that the first practical laws were found and the first wave models were built. Since then the knowledge of the ocean waves has increased significantly and the models have become more and more advanced. Nowadays state-of-the art third-generation wave models [1–3] provide wave information and wave forecasts at regional or global scales.

In spite of great progress in the wind-wave modeling, these models are not perfect and it is not possible to represent all the physical processes in the numerical wave model. In addition, the numerical wave model is dependent on the initial conditions, boundary conditions and driving forces which all of them contain errors. The accuracy of wave forecast decreases with time due to the aforementioned source of errors. Errors will accumulate during the execution of the wave model and make the forecasts useless. Therefore the improvement of wave models output takes a prominent place in the wind-wave forecasting. Whenever the observations are available, an improvement in the wave predictions can be obtained.

* Corresponding author. Tel.: +98 311 3912539; fax: +98 311 3912518.
E-mail addresses: arzamani@cc.iut.ac.ir (A. Zamani), azimian@cc.iut.ac.ir (A. Azimian), A.W.Heemink@ewi.tudelft.nl (A. Heemink), d.solomatine@unesco-ihe.org (D. Solomatine).
[1] Tel.: +31 15 2785813.

0307-904X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.apm.2009.10.013

Application of the observations in order to improve the performance of a model is referred to as data assimilation (DA). In DA, model results and observations are combined to calibrate the model parameters, to improve initial conditions for a new model run or to correct the outputs of model. Two families of DA techniques are commonly used for wave models. The first family includes the variational method. This method is based on the minimization of the cost function which measures the difference between the model results and measurements. The minimization problem is implemented by solving the adjoint problem. This method adjusts the model solution to all available observations in a certain period. The adjustment should be applied simultaneously to all time steps and all time steps are influenced by all observations. Derivation of adjoint is a tedious job. Especially the derivation of the non-linear interactions is very difficult. Various authors [4,5] have developed adjoint model via an analytical adjoint with some simplifications. This has been done for WAM [1] model by De las Haras [4] where the advection in the wave model was not taken into account. Also De Valk [5] left the non-linear wave interaction out of the adjoint model. Giering and Kaminski [6] have developed a compiler (TAMC) which can generate the adjoint code automatically. The TAMC compiler has been used by Hersbach [7] to obtain an adjoint version of wave model WAM. In another research Vos [8] investigated the possibilities of operational wave forecasting with variational DA using a limited number of control variables to avoid adjoint model. The wind field was favored over other parameters such as model constants as it has a significant influence on the accuracy of forecasted wave field. In total, five different wind correction methods were tried.

Sequential methods such as optimal interpolation and Kalman filter belong to the second family of methods. In these methods, observations are used as soon as they are available to correct the present state of the model. At each time step the model results are corrected in such a way that the model results agree with the latest observations as much as possible. So they are suitable to improve the model's skill for operational or real time forecasting.

Optimal interpolation [9–14] is a statistical interpolation technique which does not include any constraint posed by numerical wave model. The error covariances are prescribed without using the model. The best or analyzed model results at each assimilation step are obtained by a linear superposition of the first guess vector (produced by model) and weighted errors between the observed data and corresponding model data. The advantage of this method is that it is fast and only one model run is needed. However a serious drawback is that the observations only influence the model results at later times and not at pervious times. Another disadvantage is that the sequence of corrections is not consistent with the model dynamic.

The KF is similar to optimal interpolation, except that the error covariances are not a priori specified but are computed using the numerical model. The technique is formulated by Kalman [15] for linear systems. Although the KF theory is designed for linear systems but there are many variants which can cope with non-linear models. Some of these variants include the Extended Kalman Filter [16], Reduced Rank Square Root Filter [17], Ensemble Kalman Filter [18,19] and Sigma-point Kalman Filter [20].

The KF is an excellent technique for data assimilation as it takes maximum advantage of observations. On the other hand this method has a few practical problems when it is used in operational applications for large non-linear models. In these situations the computation of the forecast error covariance matrix and Kalman Gain can become enormous job both in computer power and computer storage. For example for the non-linear extension of KF that use ensemble methods [31] it is required to evaluate the dynamics over statistical ensembles of at least $2n+1$ members with n as the dimension of state-space. As an example, evaluating of $2n+1$ members with a 10^3 dimensional state and 20 min forward run time of model will approximately take 2.4×10^6 s which is equivalent to 667 h.

The Kalman Filter has not been used extensively in wave data assimilation. Voorrips et al. [21] used an Extend Kalman Filter with a truncated second order filter. Also in order to estimate wind corrections based on wave observations, the filter has been extended with a fixed lag smoother. They used limited number of grid points due to huge memory demands of KF. Pinto et al. [22] applied the KF method in a two dimensional domain with deep water condition and relatively small area without wind forcing or dissipation. The assimilation procedure is tested for swell propagation where the KF is used to assimilate significant wave height.

It has been demonstrated in literature how the non-linear regression models or artificial neural networks (ANN) can help in building the DA models. Incorporation of artificial neural networks and data assimilation technique into a third generation wind-wave model for wave forecasting is studied by [23]. In above reference, there are four ANN for four different location which they mimic the relation between the model output and corrected output. The corrected output is calculated by optimal interpolation.

Babovic et al. [24] provide a scheme for error correction of a predictive ocean wave model using local model approximation. The approach is based on application of a method inspired by chaos theory for building of local model. Local model acts as an error correction tool (by combining the model predictions with the observations). This scheme is only used to correct the model output for observation stations.

There are the following restrictions for direct implementation of EnKF or other ensemble based data assimilation methods for large models:

- execution time of numerical model;
- calculation of Kalman gain;
- the wave model source code has to be changed for this purpose.

If we ignore the role of dynamic equation, then it is possible to use simple output correction methods such as optimal interpolation, Ensemble optimal interpolation and optimal interpolation of partitions which are examined by various authors. These methods need one model run for each updating of the model.

The present work is another approach to wave data assimilation using EnKF. It is based on the idea of a reduced dimension Kalman filter which recently was applied to non-linear dynamics encountered in benchmark estuarine flows [25]. To overcome the problems with the existing DA algorithms, an artificial neural network (ANN) surrogate is trained to approximate the forward wave model. Also in order to be computationally feasible, the entire system operates on a low dimensional subspace obtained by principal component projection of wave and wind fields. In this case the ANN emulates the physical relationship between the wind and wave by recognition of the patterns in the data presented to the network. This ANN is fast but its accuracy may deteriorate through time. Especially, in the test period, the ANN wave model should simulate the wind-wave conditions which they have not seen beforehand. The quality of the ANN then is determined by the quality of the wind-wave realizations in the training data sets. The use of fast ANN in combination with EnKF makes it possible to perform many runs of the dynamic model during each time step of the assimilation.

Combination of ANN and EnKF has a number of additional attractive features important for practice. They allow for:

- Optimization of the location of existing measurement networks.
- Demonstration of the effect of number of measurement stations before deployment of wave measurement systems.
- Comparison of performance of Ensemble based methods with the same explicit dynamic equation. (ANN can provide explicit dynamic equation for wind-waves and various data assimilations can be combined with it).

The data availability for this research was not optimal, so it should be seen as a study into feasibility of the approach. It can be further extended for other research concerning the combination machine learning methods, wave models and data assimilation methods.

The structure of the present paper is as follow. First, a brief introduction to spectral wave model and foundation of model reduction are given. Then the dynamic ANN surrogate and EnKF algorithm are explained. Some technical issues of the problem are presented as well. In the following section, the results of wave model and its surrogate, type of observations and their locations and effect of DA algorithm are specified. Finally, a conclusions are drawn and the further directions for future research are outlined.

2. Wave modeling

The spectral wave model calculates the evolution of the energy spectrum for ocean waves by solving action balance or energy balance equation explicitly without any presumption on the shape of the wave spectrum. The wave spectrum describes the state of sea surface at certain time and location. The wind driving forces are not constant and the wave spectra will therefore change according to changing forces. The energy balance equation becomes

$$\frac{\partial}{\partial t}F(f,\theta,\vec{r},t)+\nabla\cdot(\vec{C}_g F(f,\theta,\vec{r},t))+\frac{\partial}{\partial\theta}(\dot{\theta}F(f,\theta,\vec{r},t))=S_{tot} \tag{1}$$

where F is the wave energy spectrum in terms of frequency f and propagation direction θ at the position vector $\vec{r}$ and at time t. Also $\vec{C}_g$ is the group velocity. The first term on the left hand side of Eq. (1) represents the local rate of change of energy density in time. The second term represents propagation of energy in geographic space with propagation velocity of density $\vec{C}_g$. The third term represents depth induced and current induced refraction with propagation velocity of θ° .The right hand side source term S_{tot}, represents all physical processes which affect the wave spectrum at position $\vec{r}$. This term includes wind input, non-linear wave-wave interaction, dissipation due to wave breaking and bottom friction.

The energy balance Eq. (1) is solved in discretized form using a semi-implicit time integration scheme and a first order upwind advection scheme. Time is discretized with a constant time step for simultaneous integration of propagation term and source terms. For small scale computations Cartesian coordinates can be used and geometric space is discretized with rectangular grids with constant resolution.On solving Eq. (1) the direction wave spectrum in each spatial grid location at every time step has been estimated. The wave parameters such as significant wave height (H_s), the mean wave period (T_m) and mean wave direction (θ_m) are obtained from spectrum analysis at each grid point. These parameters are first guess model values in full space for given wind field. A detailed presentation of the theory can be found in Komen et al. [26] and Holthuijsen [27].

In the present study, the hindcast wave data at the Caspian Sea [28] is used. The data set is generated by execution of MIKE 21-SW wave model. MIKE 21-SW is a 3rd generation spectral wind-wave model that simulates the growth, decay and transformation of wind generated waves and swell in offshore and coastal areas.

3. Model in reduced space and EOF analysis

It is not possible to train the model surrogate in the full space. There are a number of methods to extract patterns with a specific spatial scale out of the full space. Empirical Orthogonal functions (EOF) [29] are widely used for data analysis in meteorology and oceanography. This technique can analyze the variability in the data set and decompose this information into a set of orthogonal spatial patterns with time dependent amplitude. Most of the variance of the patterns is captured in the first few patterns. This is a very useful tool for data reduction because the original data can be reasonably well represented with only the first few patterns. Use $X(k) \in \Re^n$ and $X_s(k) \in \Re^r$ to denote the state space of the full numerical model and states in subspace, respectively. Also k shows index of each time step. The EOF subspace achieved by the following mapping

$$X_s(k) = \Pi_r \times (X(k) - \mu) \quad X \in \Re^n, \quad X_s \in \Re^r \tag{2}$$

$$\Pi_r = [\varphi_1, \varphi_2, \ldots, \varphi_r] \tag{3}$$

where Π_r is $r \times n$ EOF projection matrix. This matrix is characterized by the r leading eigen vectors φ_i of the covariance matrix for a data set of state vectors $X(k)$. Also μ is the mean of data set. It is possible to reconstruct the reduced space into the full space as

$$X(k) = [\Pi^T \Pi \times (X(k) - \mu) + \mu] + \varepsilon \tag{4}$$

with ε is the reconstruction error.

The above mentioned procedure can be applied to both wind $(U(k))$ and wave fields $(X^H(k))$ as below:

$$U_s(k) = \Pi_{r1} \times (U(k) - \mu^U) \tag{5}$$

$$X_s^H(k) = \Pi_{r2} \times (X^H(k) - \mu^H) \tag{6}$$

$$U(k) = [u_1(k), v_1(k), u_2(k), v_2(k), \ldots, u_n(k), v_{n1}(k)] \tag{7}$$

$$X^H(k) = [H_1(k), T_1(k), \Theta_1(k), H_2(k), T_2(k), \Theta_2(k), \ldots, H_{n2}(k), T_{n2}(k), \Theta_{n2}(k)] \tag{8}$$

where $U_s(k)$ and $X_s^H(k)$ are reduced space wind and wave fields with dimensions r_1 and r_2. Also μ^U and μ^H are the mean of wind and wave field data sets. In Eq. (7) the two components of wind velocity at each grid point denoted by u_i and v_i which total number of grids are n_1.

In a similar manner, in Eq. (8) the integral parameters of wave field at each grid point are denoted by H_i, T_i and θ_i which total number of grids is n_2.

It is necessary to treat wind or wave direction in the EOF analysis. There is a discontinuity in the direction around the north direction. For both of the wind and wave field, we have used the components of wind velocity and wave height in the east and north direction. On the other hand, a simple change of variables is done to avoid the problems encountered in direct usage of direction.

Also, we have used integral wave parameters instead of spectrum. The reduction of the wave model space is to go from integral parameters of 388 grid points to their EOFs. Also the reduction of wind 388 of u, v components of wind velocity to their EOFs is done separately.

The reduction of space from full spectrum to reduced space spectrum can be used as another alternative which have not considered here. In our hindcast data base there were no spectrum data and we have access to integral wave parameters.

4. Surrogate model in subspace

Surrogate models (or emulators, or meta-models) are approximate models that mimic the behavior of large scale model as closely as possible while being computationally cheap. Surrogate models are constructed using data driven methods. It is assumed that only the input–output behavior is important. A surrogate is constructed based on modeling the response of large scale model to a limited number of intelligently chosen data points. An artificial neural network (ANN) is normally used to map random input vector into the corresponding output vector. The physics of underlying system need not to be known beforehand. This feature makes ANN a suitable tool to approximate non-linear function relationships without a pre-existing model and with only diminutive knowledge about the physics of the system. ANN assumes complicated function involving multiple sums over some specified functions such as sigmoid function etc. See Haykin [30] for comprehensive foundation of ANN.

In the context of ocean modeling, one interesting application of the ANN surrogate is described in [31,32]. They used ANN surrogate that provides extremely fast and accurate emulation of dynamics of the large scale circulation model. Their circulation code has large degree of freedom with high non-linearity. It was driven by ocean and atmospheric and river influences in boundaries.

Some other applications of ANN to improve computational efficiency of the numerical oceanic models are studied by Krasnopolsky et al. [33] and Tolman et al. [34]. Although their networks were not used to emulate the dynamics of full system state directly, but they, however, replaced some computational blocks of a high dimensional environmental models with an ANN model which learn complex parameterization of several types of physical processes. Krasnopolsky [35], Krasnopolsky et al. [36] developed ANN emulator of major components of climate models as well. The application of radial bias function neural network is applied to emulate an Extened Kalman Filter (EKF) in a data assimilation scenario by Harter and Velho [37]. They used one dimensional shallow water equation in their study.

Looking at the wind-wave process as a dynamic system, the state of the system (wave field) at time k depends on the system's state of one or more preceding time steps as well as present or preceding forcing of system (wind speed). To account for temporal evolution, the dynamic model must be used in 'state-space' form. Such a state space form can be achieved by designing a recurrent network. At the present work, the following time lagged ANN surrogate is used to build the relationship between the wind and waves in reduced space

$$\hat{X}_s^H(k+1) = F_{NN}[\hat{X}_s^H(k), \hat{X}_s^H(k-1), \ldots, \hat{X}_s^H(k-\tau_H), U_s(k), U_s(k-1), \cdots, U_s(k-\tau_U)] \tag{9}$$

where F_{NN} denotes the surrogate model. $\hat{X}_s^H(k+1)$ is a predicted output of network, So the surrogate will lead to the one step forecast. τ_U and τ_H are time embedding indexes. The other parameters are as defined before.

The schematic representation of the network is shown in Fig. 1. According to this figure and Eq. (9), this network is a dynamic recurrent network with feedback from output to input. Input vectors consisting of current and past system states in reduced dimension wind field. The output of network $\hat{X}_s^H(k+1)$ is fed back through a unit delay operator Z^{-1} to the input of the state delay line. The problem associated here is to forecast wave height with respect to the previous information about the system. The available information can be wind speed vector and a priori knowledge of the wave height. This structure makes it possible to include time-delay behavior into the data driven ANN model.

Furthermore, the proposed ANN is a standard multi layer percepteron with a single hidden layer with hyperbolic tangent activation function and a linear output layer. This combination of activation functions is recommended for function approximation problems [38].The output of network can re-embed into the full space in order to use for any application.

The third-generation wave models will produce full spectral data with much higher quality than the ANN. As far as quality is concerned, these models perform best in most situations. ANN is a model approximator (i.e. it is a model of a model). Obviously its accuracy is lower than the main model but it can increase by using more data, clustering of data and using modular learning methods. We have not used the spectral components, but integral wave parameters. The role of physics is preserved in the wave field which is computed directly from spectral components. Also this scheme is based on a data driven model (ANN) and it is not sensitive to type of wave model. It means that, the scheme can be set up by using the data of the second generation wave models.

5. Ensemble Kalman Filter for non-linear system

This section will briefly review the mathematical formulation of EnKF for non-linear systems. Consider the system of a general, non-linear stochastic model M and the observations [39]:

$$x^f(k) = M[x^f(k-1), U(k-1)] + w(k-1) \tag{10}$$

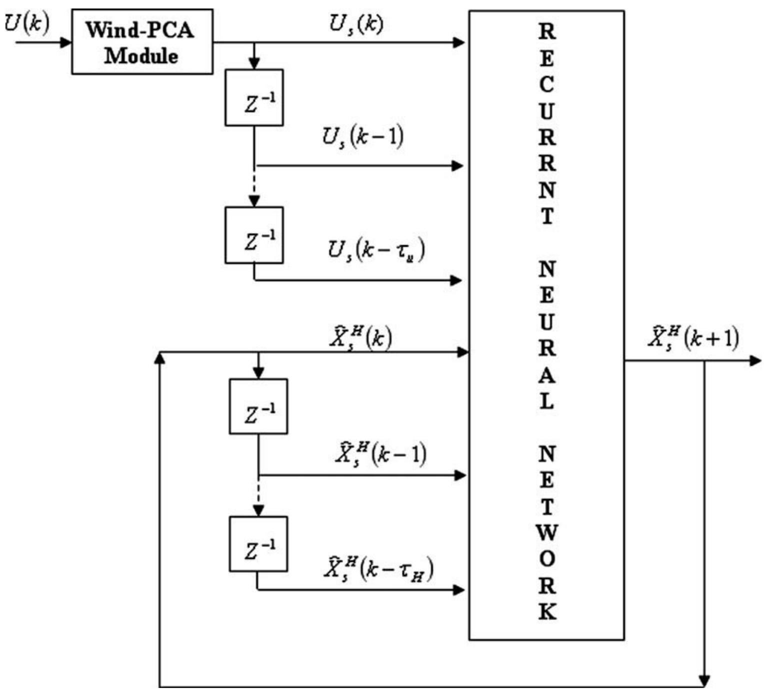

Fig. 1. Schematic presentation of ANN surrogate.

$$y^o(k) = H(k)x^f(k) + v(k) \tag{11}$$

where $x^f(k) \in \Re^n$ denotes the forecast of the system state at time k, $U(k)$ is forcing of the system and M represents one time step of the model. A normal distributed system noise $w(k) \in \Re^n, w(k) \sim N(0, Q)$ is introduced to take the uncertainties of the model into account. The vector $y^o \in \Re^{ro}$ represents the measurements which are supposed to be a linear combination of the states represented by the operator H. The observation noise is represented by $v(k) \sim N(0, R)$. In order to obtain an optimal estimate, it is needed to combine the measurement taken from the actual system modeled by Eq. (11) with the information given by the system model (Eq. (10)). The forecast state at time k, denoted by x^f, is the forecast from observation time $k-1$ to observation time k by the following equation:

$$x^f(k) = M[x^a(k-1), U(k-1)] \tag{12}$$

where $x^a(k-1)$ is analyzed system state. At time k, observation $y^o(k)$ is available and the estimate is updated by the analysis step:

$$x^a(k) = x^f(k) + K[y^o(k) - H(k)x^f(k)] \tag{13}$$

where

$$K(k) = P^f(k)H(k)^T[H(k)P^f(k)H(k)^T + R]^{-1} \tag{14}$$

is the minimum variance gain and $P^f(k)$ is the forecast error covariance matrix. In the EnKF method, the covariance can be calculated by a finite number of randomly generated system states. This scheme introduced by Evensen [18] and later clarified by Burgers and Evensen [40]. It has been successfully implemented and used in different types of applications [19]. The EnKF is a Monte Carlo approach based on the representation of the probability density of the state which is estimated by a finite number of randomly generated system states.

For the initial state estimate x_0, the uncertainty is expressed by an ensemble $\xi_i^a,\ i = 1, \ldots, N$ of randomly generated states. The ensemble members are propagated from one time step to another using the original model operator:

$$\xi_i^f(k) = M[\xi_i^a(k-1), U(k-1)] + w_i(k-1) \tag{15}$$

with $w_i(k)$ realizations of the noise process $w(k)$, i.e. noise is added to the most uncertain parts of the model to estimate the covariance between observations and the model variables. The ensemble mean

$$\bar{x}^f(k) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^f(k) \tag{16}$$

represents the state estimate at time k. Using this estimate the error covariance can be computed as:

$$E^{f,EN}(k) = [\xi_1^f(k) - \bar{x}^f(k), \ldots, \xi_N^f(k) - \bar{x}^f(k)] \tag{17}$$

$$P^f(k) = \frac{1}{N-1}E^{f,EN}(k)[E^{f,EN}(k)]^T \tag{18}$$

With the error covariance P^f calculated, the Kalman gain $K(k)$ is obtained from Eq. (14) and the update equations for the analyzed ensemble are:

$$\xi_i^a(k) = \xi_i^f(k) + K(k)[y^o(k) - H(k)\xi_i^f(k) + v_i(k)] \tag{19}$$

where $v_i(k)$ represents realizations of the measurement noise process $v(k)$.

In the standard form of EnKF as described in Eq. (10)–(19), we are dealing with time step k and $k+1$. The state vectors in these steps are related by dynamic Eq. (10). On the other hand the ANN surrogate as described by Eq. (9) consists current and time delayed system states. In the next section the new definition of system state will present.

6. Coupling of surrogate model and EnKF

In the combined method, the output of the ANN will be considered as state vectors. The state vectors are provided to the data assimilation module. This module works based on EnKF. By help of the observations, the EnKF will correct the output of the ANN to find the best estimate of the system (or analyzed state). The analyzed state will return to the inputs of the ANN for the next time step.

There is a different between inputs of network which come from forcing or from feedback loop. One should carefully note that the inputs of the EnKF are different from inputs of the ANN. To deal with the time delayed states in Eq. (9), The Kalman filter formulation proposed by Gibson et al. [41] is employed and an extend vector is taken as follows:

$$\chi^k = [X_s^{H,k}, X_s^{H,k-1}, \ldots, X_s^{H,k-\tau_H}]^T \tag{20}$$

This new state vector expands the dynamic equation to the following equation:

$$\begin{bmatrix} X_s^{H,k+1} \\ X_s^{H,k} \\ X_s^{H,k-1} \\ \vdots \\ X_s^{H,k-\tau_H+1} \end{bmatrix} = \begin{bmatrix} & & F_{NN}(\cdots) & & & \\ I & 0 & 0 & \cdots & 0 & 0 \\ 0 & I & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} X_s^{H,k} \\ X_s^{H,k-1} \\ \\ \vdots \\ X_s^{H,k-\tau_H} \end{bmatrix} + \begin{bmatrix} w_s^k \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{21}$$

In Eq. (21), I is identity matrix and $F_{NN}(\cdots)$ is described by Eq. (9). It can easily show that the compact form of the Eq. (21) is in accordance with the general formulation of the Kalman Filter equations. Eq. (21) can be rewritten as:

$$\chi^{k+1} = M_s[\chi^k] + W^k \tag{22}$$

$$W^k = [w_s^k, 0, 0, \ldots, 0]^T \tag{23}$$

where W^k is the extend noise vector and M_s is a new dynamic operator.

In the standard KF, the observation is formulated by Eq. (11). On the other hand, the states are projected to the subspace. Therefore the observation equation should be replaced with following equation:

$$y^o(k) = H(k)\Pi_{r2}^T X_s^{H,k} + \upsilon_s \tag{24}$$

where $\upsilon_s(k)$ is measurement noise in the reduced space. Correspondingly

$$\upsilon_s(t_k) \approx H(t_k)(X^{H,k} - \Pi_{r2}^H(\Pi_{r2}^H)^T X^{H,k} + \mu^H) + \upsilon(t_k) \tag{25}$$

See Lu et al. [25] for derivation of the measurement noise in the reduced space. The data assimilation corrects the underlying model and makes the quality of the model less important. One extension of this work could be the design of scheme for building a relationship between wave model and its corrected output from Coupling of ANN model and EnKF.

7. Results

In the present study, input–output patterns for training of a network are from a 6 month hindcast study at the Caspian Sea [28]. A third generation MIKE 21-SW wave model is set up in the domain spanning longitude of 46.5–55° and latitude of 36.5–47.5° with the resolution of 0.125° for southern part of Caspian Sea and 0.5° for reminder region. The 6 hourly wind fields are extracted from ECMWF with resolution of 0.5°.

The period of available data is from January 7, 2000 to July 7, 2000. During the wave model execution the wave spectrum is discretized in 16 equidistance directions and 25 frequency ranging from 0.052 to 0.60 Hz, on a logarithmically equidistance.

The 3 hourly intervals were selected for the execution of wave model. So the total number of data sets is 1500. Also the data of 388 grid points are selected for study. The wave model results are a proxy for true wave states. The forward run time of the wave model takes 184 s. Fig. 2a shows the study area and location of selected grid points and observation points. *Due* to limited number of available observation points in southern part of Caspian, we simulate the situation with approximate location of buoys. Also the buoy data of other neighboring countries are not available and there is no information about their networks. Also Fig. 2b shows the time series plot of wave height of two arbitrary grid points for 1500 time index. For other grid points it is possible to produce similar figures, so one may say that indeed there are extreme events in the time series.

To illustrate the EOF method, an analysis is applied to both wind and wave fields on 1400 number of data set. Fig. 3 shows the variance distribution of the EOF's. According to this figure, the first eigenvector contains the highest percentage of the total variance followed by the second mode, and so on. The larger variances represent the large scale patterns with slowly varying amplitudes, representing more stable signal in data set. Smaller variances are often connected to the smaller scale patterns with highly varying amplitudes, representing more unstable signal in data set. To prepare the surrogate model, the total data set is divided in two groups. The first group is used for training which contains 1400 number of data sets. The second group is used for test of network which contains 100 numbers of data sets.

In this study, 15 EOF's for representing the reduced space of wind wave fields have been chosen. Also 3 time step lag has been considered for input and outputs of network. The resulting network topology is 90 inputs and 15 outputs. The supervised algorithm of resilient back propagation is used for training of the network. The weights of ANN network are initialized from zero. Due to limited number of data set and to prevent overfitting in training of network, the n-fold cross validation method is used [42]. In this method n different sets of training-validation data should be prepared based on the original training data with the length of N_t. In each set (N_t/n) number of data vectors is used for validation and the $(1-\frac{1}{n})N_t$ of them for training. n models are built, and for each model training and validation sets are different. For model i the validation set comprise (N_t/n) vectors starting from vector $\frac{N_t^*(i-1)}{n}+1$, and the rest of the data set is used for training. The stopping criteria were one of the following: minimum error in validation set, mean square error in training reaching threshold of 0.0001 or the number of epochs reaching 25,000.

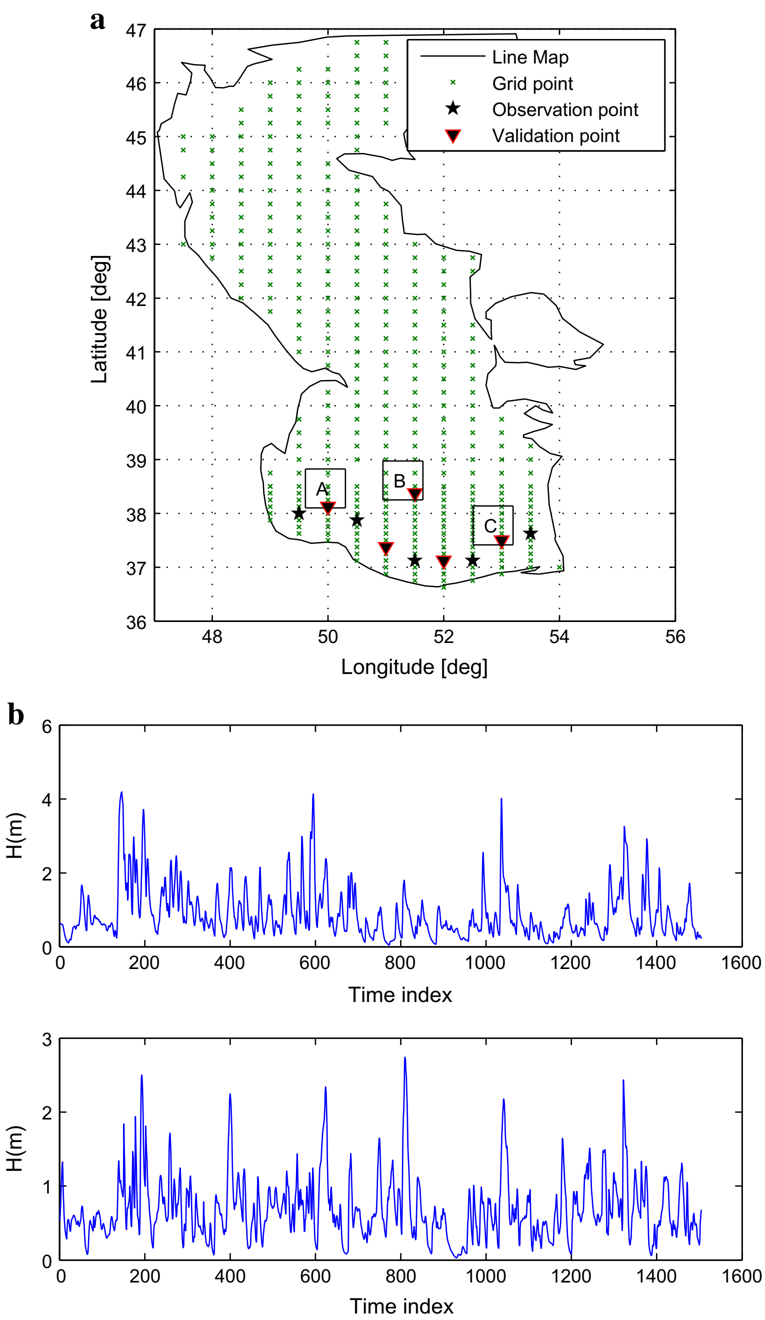

Fig. 2. (a) Location of study area and grid points and validation points. (b) Time series plot of wave height for two arbitrary grid points.

Also the optimal size of the hidden layer was found by systematically increasing the number of hidden neurons until the network performance on the validation sets did not improve significantly. Results showed that there is no longer any improvement in the performance of the model for more than 53 hidden neurons.

The test data set includes 100 data sets which have not been used during the training. In spite of the well trained network, the surrogate could not predict the exact wave field in the test period. To illustrate this, we investigate the real wave field, reconstructed wave filed and simulated wave field. Fig. 4 presents true wave field at an arbitrary time index ($k = 50$) in the test interval. It is assumed that we have an exact wind forecast during the test period.

This field is projected into the reduced space and it can be re-embed to full space to reconstruct the original wave field. The reconstructed wave field and its difference with the true wave field are shown in Figs. 5 and 6, respectively. Fig. 7 also shows the surrogate (simulated) wave field which is not similar to true wave field due to both reconstruction and simulation

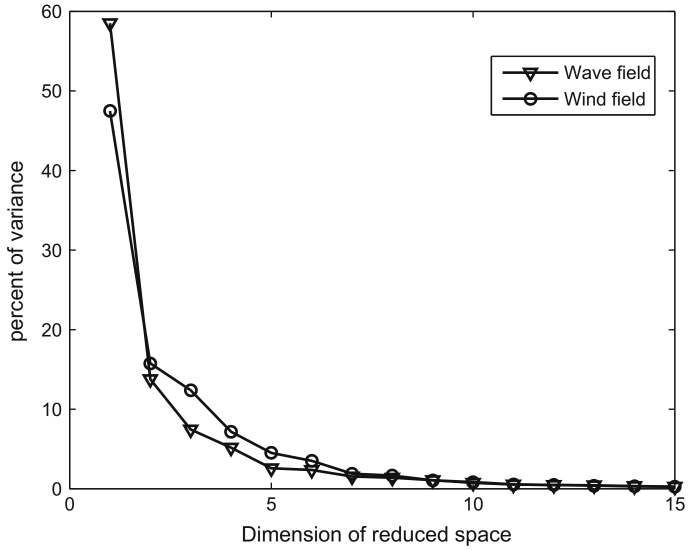

Fig. 3. Variance of individual EOF's.

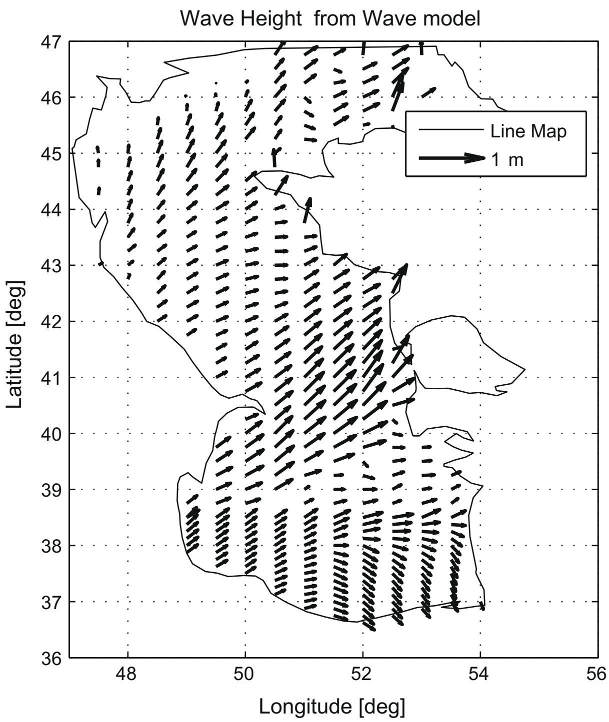

Fig. 4. Wave height from wave model.

errors. To reduce the reconstruction and simulation errors, observations of wave height and wave direction at some selected points are assimilated into the surrogate wave model. The locations of observation locations are indicated in Fig 2a. Time series of the observed wave height and the observed wave direction are extracted from the true wave field. The analysis wave field is presented in Fig 8. This figure is based on 5000 ensemble members and an assimilation-forecast cycle for 3 h. Difference between true wave field and corrected wave field can be seen in Fig 9. As expected, it can be seen that the most of corrections occurred at the high resolution part of model and near the observation points.

Unfortunately the available hindcast data has different resolution. Also our study is limited to measurement points in the southern part of Caspian Sea. In case of using higher resolution grid points and using more measurement stations through

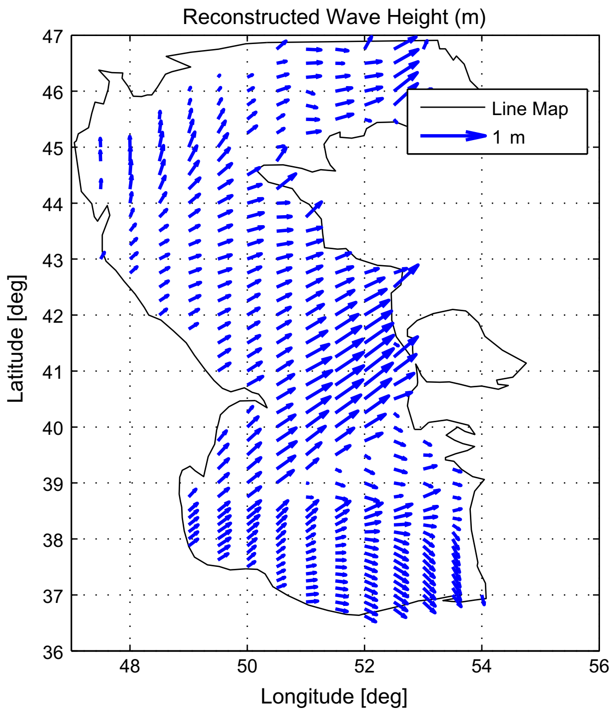

Fig. 5. Reconstructed wave height.

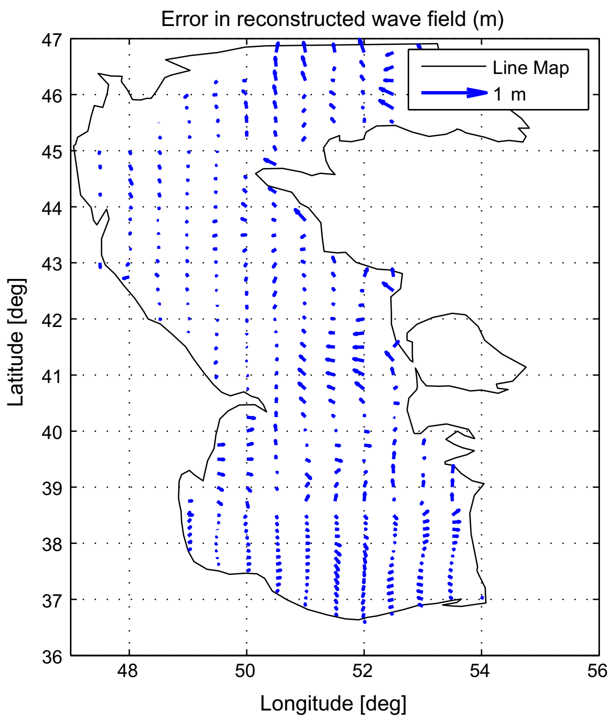

Fig. 6. Errors in reconstructed wave field.

the entire domain, the results can considerably improved. Having the true wave field, it is possible to calculate the root mean square error of the reconstructed wave field, simulated wave field and analyzed wave field. Fig. 10 illustrates the root mean square error for aforementioned fields for all time steps in the test interval. The horizontal axis is time and the vertical axis belongs to average of error in significant wave height in all grid points (the hole space).

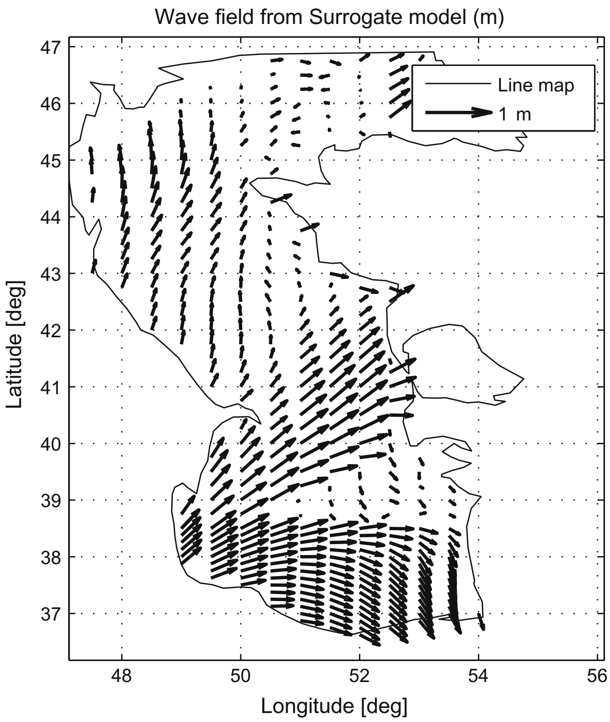

Fig. 7. Simulated wave field.

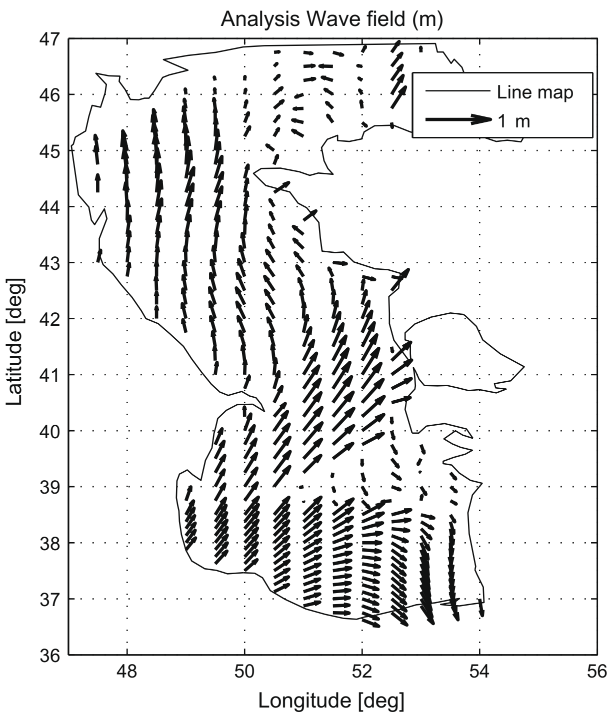

Fig. 8. Analysis wave field.

Although the temporal behavior of the errors shows that the surrogate performing better in some time indexes, but there are some regions with high errors. This problem is most likely caused by under regularization of the network due to the high level of non-stationarities of the forcing in the regions with higher errors.

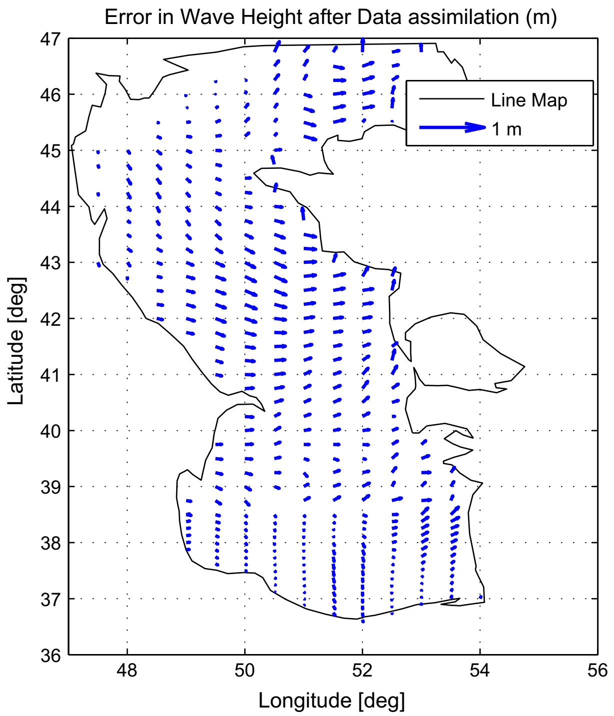

Fig. 9. Errors after data assimilation.

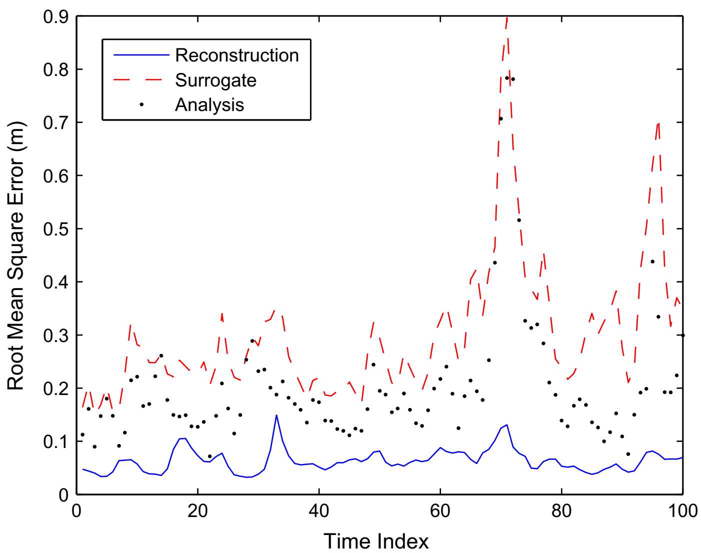

Fig. 10. Root mean square errors for various wave fields.

To see the effect of the data assimilation, all fields are shown for the small region in the Fig 11.

This region is selected close to the observation points and in the high resolution part of the model. It is clear that after assimilation of the data, the magnitude of the wave height and wave direction goes towards the true waves.

The effect of the data assimilation can also be seen from the time series plot of wave height at each point. The time series plot of the wave height for one of the observation points is shown in Fig 12. This plot shows the time variation of the true wave, the reconstructed wave, the simulated wave and analyzed wave at point (A) of Fig 2a. Again in this figure the analyzed wave obtained by EnKF algorithm with 5000 ensembles. It is obvious that the observation at this point has a large effect on the predictions of the surrogate model. The effect of the assimilation can be studied also in other points than observation

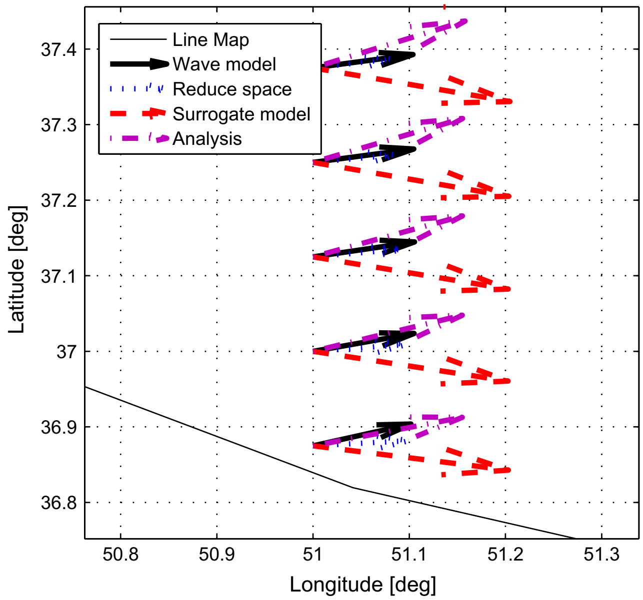

Fig. 11. Small region of study area for all wave fields.

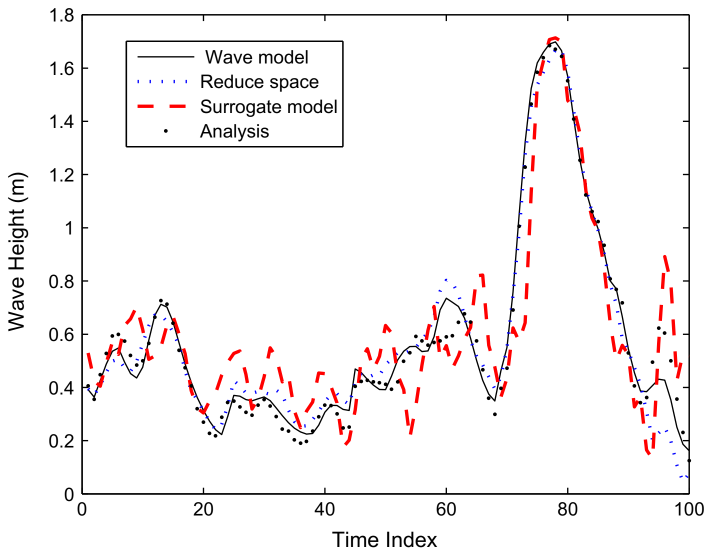

Fig. 12. Wave height time series at observation location (A).

points. This is done for point (B). The position of this point is shown in Fig 1. The time series of the wave height at this point is plotted in Fig 13. This point is far from the observation points and the effect of the corrections is less than at the observation points.

To find the effect of the number of ensemble members the dynamic model is run with different number of ensembles for point (C). Fig. 14 shows the effect of increasing the ensemble members. For this point the results of 100 ensembles and 50 ensembles are close to each other. The results of 100 ensembles are closer to the true wave.

Root mean squares error in the wave height for the model with different ensemble members also is listed in Table 1. Increasing number of ensembles will increase the computational cost but it will create some acceptable results with lower sensitivity even for noisy data.

Depending on the size of domain, the surrogate model is several times faster than full model. The CPU time for execution of surrogate for different number of ensemble members listed in Table 2. Simulations were done at whole Caspian Sea on a single processor Pentum-4 2 GHz PC with 2 Gb of memory.

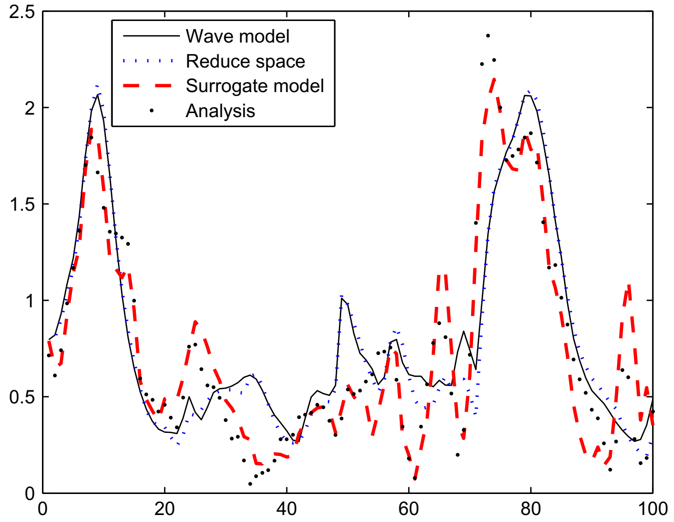

Fig. 13. Wave height time series at validation location (B).

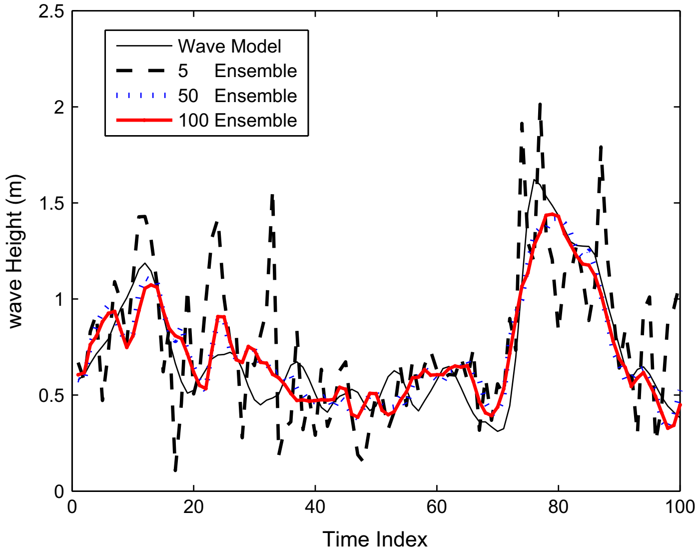

Fig. 14. Effect of ensemble members on wave height time series at validation location (C).

Table 1
Wave height root mean square error for different ensemble members of Fig 14.

Ensemble members	Root mean square error (m)
5	1.387
50	0.126
100	0.110

Table 2
Required CPU time for propagation of surrogate model.

Ensembles members	Wave model CPU time (s)	Surrogate model CPU time (s)
5	925	29.19
50	9250	172.16
500	92,500	1624.40
5000	925,000	14240.45

There is a trade off between the accuracy and speed of the surrogate model. Its accuracy is less than the full numerical model but its computational speed is an important aspect to be taken account for wave data assimilation.

8. Conclusions and future work

An approach to wind-wave data assimilation is presented. It is based on using an artificial neural network as a fast surrogate model combined with an efficient low rank approximation of the Kalman filter algorithm. The dynamic equation for the relation between the state vector and the forcing of system is derived from data at the Caspian Sea in a reduced space. ANN surrogate describes both the spatial and the temporal behavior of the wind waves in this area. This ANN data driven prediction is updated using new observations as soon as they are available. The analyzed or corrected states will then be used for next forecast of the wave field.

In context of data assimilation, accelerating the simulation of the dynamics of the system via surrogate successfully reduce the time needed to propagate the ensemble forecasts. So, the approach based on the non-linear extension of Kalman filter (EnKF) that requires the evolution of the dynamics of the system over a large number of ensembles becomes feasible with widely available computing power. The experiments on the available data set, show that the error of the model forecast can be reduced by incorporating the observed data and due to the predictor corrector behavior of the algorithm, the forecast is more accurate than with the use of the surrogate alone.

In spite of the fact that the data availability for this research was not optimal, it can be concluded that the presented approach has a potential in wind-wave modeling, allowing for efficient corrections of predictions.

9. Future work

The scheme can be further improved. For example one can use the numerical wave model for preparing a lot of realizations of wind wave spectra for the area under study. These spectra can be used as states in the full space and the reduction of space can be implemented directly on the spectral components. Evidently, the required number of modes necessary to approximate the full space will be more. Under this condition, more accurate reduction methods which seek a good trade off between the reconstruction error of the states and the preserved dynamics are required. In the model reduction phase, it may be possible to use non-linear dimension reduction methods [43,44].

Another improvement can be done in the simulation phase. The development of ANN emulator for fast calculations of physical processes depends significantly on the ability to generate a representative training set. Owing to the high dimensionality of the ANN input vector, it is rather difficult to cover the entire domain, and especially its “far corners” associated with rare events, even when we use model-simulated data for ANN training. Also the domain may change in time due to climate variability. In this situation the emulating ANN may be forced to extrapolate beyond its generalization ability and may lead to large errors in the outputs. Krasnopolsky et al. [36] developed a compound parameterization technique to address this problem and make the emulation approach more suitable for numerical modeling applications. Adding this technique to the presented approach may increase it accuracy.

Another possibility is to generate wave data by numerical model according to predefined strategies in the study area. This task is not included in this research.

In the present work, the true wind field was considered to be free of errors. However it is possible to study a more realistic setting assuming the errors in the forcing. These errors can be corrected by using a Kalman-smoother.

Acknowledgment

The authors thank the Port and Shipping Organization of IRAN for providing hindcast wave data.

References

[1] The WAMDI Group, The WAM model-third generation ocean wave prediction model, Journal of Physical Oceanography 18 (12) (1988) 1775–1810.
[2] N. Booij, R.C. Ris, L.H. Holthuijsen, A third generation wave model for coastal regions – part I: model description and validation, Journal of Geophysical Research 104 (1999) 7649–7666.
[3] H.L. Tolman, User Manual and System Documentation of WAVE WATCH-III Version 2-22, NOAA/NCEP/MMAB Technical Note 222, 2002.
[4] M.M. De las Heras, G. Burgers, P.A.E.M. Janssen, Variational wave data assimilation in a third generation wave model, Journal of Atmospheric and Oceanic Technology 11 (1994) 1350–1369.
[5] C.F. De Valk, A wind wave data assimilation scheme based on adjoint technique, in: Komen et al., 1994, pp. 460–468.
[6] R. Giering, T. Kaminski, Recipes for adjoint code construction, ACM Transactions on Mathematical Software 24 (4) (1998) 437–474.
[7] H. Hersbach, Application of the adjoint WAM model to inverse wave modeling, Journal of Geophysical Research 103 (C5) (1998) 10469–10488.
[8] S. Vos, Variational Wave Data Assimilation for Optimal Wave Forecasting, PhD Thesis, Technical University of Delft, Netherlands, 2002.
[9] D.C. Esteva, Evaluation of preliminary experiments assimilation Seasat significant wave heights into a spectral wave model, Journal of Geophysical Research 93 (1988) 14099–14106.
[10] K. Hasselmann, S. Hasselmann, E. Bauer, C. Bruning, S. Lehner, H. Graber, P. Lionello, Development of a Satellite SAR Image Spectra and Altimeter Wave Height Data Assimilation System for ERS-1, ESA Report, Max-Plank Institute fur Meteorologie 19, Hamburg, 1998.
[11] P.A.E.M. Janssen, P. Lionello, M. Reistad, A. Hollingsworth, Hindcasts and data assimilation studies with the WAM model during the Seasat period, Journal of Geophysical Research 94 (1989) 973–993.

[12] P. Lionello, H. Gunther, P.A.E.M. Janssen, Assimilation of altimeter data in a global third generation wave model, Journal of Geophysical Research C97 (1992) 14453–14474.
[13] A.C. Voorrips, C. De Valk, Comparison of two operational wave data assimilation methods, Global Atmosphere and Ocean Systems 7 (1999) 1–46.
[14] D.J.M. Greenslade, I.R. Young, The impact of inhomogeneous background errors on a global wave data assimilation system, Journal of Atmospheric and Ocean Science 10 (2) (2005) 61–93.
[15] R.E. Kalman, A new approach to linear filtering and prediction problems, Transactions on ASME Journal of Basic Engineering 82 (1960) 35–45.
[16] A. Jazwinski, Stochastic process and filtering theory, Mathematics in Science and Engineering, vol. 64, Academic Press, New York, 1970.
[17] M. Verlaan, A.W. Heemink, Reduced rank square root filters for large scale data assimilation problems, in: Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, World Meteorological Organization, 1995.
[18] G. Evensen, Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast model statistics, Journal of Geophysical Research 99 (C5) (1994) 10143–10162.
[19] G. Evensen, The Ensemble Kalman Filter: theoretical formulation and practical implementation, Monthly Weather Review 123 (2003) 1128–1145.
[20] R. Van der Merwe, E. Wan, Sigma-Point Kalman Filers for probabilistic inference in dynamic state-space models, in: Workshop on Advance in Machine Learning, Montreal, Canada, 2003.
[21] A.C. Voorrips, A.W. Heemink, G.J. Komen, Wave data assimilation with Kalman filter, Journal of Marine Systems 19 (1999) 267–291.
[22] J.P. Pinto, M.C. Bernardino, A. Pires Silva, A Kalman filter application to a spectral wave model, Nonlinear Process in Geophysics 12 (2005) 775–782.
[23] Z. Zhang, C. Li, Y. Li, Y. Qi, Incorporation of artificial neural networks and data assimilation techniques into a third-generation wind–wave model for wave forecasting, Journal of Hydroinformatics 7 (2006) 65–76.
[24] V. Babovic, S.A. Samnasiraj, E.S. Chan, Error correction of predictive ocean model using local model approximation, Journal of Marine Systems 1 (2005) 17.
[25] Z. Lu, T.K. Lean, R. VanderMerwe, S. Frolov, A.M. Baptista, Sequential data assimilation with sigma-point Kalman filter on low dimensional Manifold, Technical Report TR-07-001, NSF-STC for Coastal Margin Observation and Prediction, 2007. <http:// www.stccmop.org/files,CMOP-TR-07-001.Pdf>.
[26] G.J. Komen, L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, P.A.E.M. Janssen, Dynamics and Modeling of Ocean Waves, Cambridge University Press, New York, 1994.
[27] L.H. Holthuijsen, Waves in Oceanic and Coastal Water, first ed., Cambridge University Press.
[28] Port and Shipping Organization of Iran (PSO), Modeling of Iranian Waters – Caspian Sea, 2003 (Chapter 4).
[29] H.V. Storch, C. Frankignoul, Empirical modal decomposition in coastal oceanography, in: KH. Brink, A.R. Robinson (Eds.), The Global Coastal Ocean Processes and Methods, The Sea, vol. 10, John Wiley & Sons Inc., New York, 1998, pp. 419–455.
[30] S. Haykin, Neural Networks – A Comprehensive Foundation, second ed., Prentice Hall, New Jersey, 1999.
[31] R. Vander Merwe, T. Leen, Z. Lu, S. Frolov, A. Baptista, Fast Neural Network surrogates for very high-dimensional physics based models in computational oceanography, Neural Networks 20 (4) (2007) 462–478.
[32] S. Frolov, Enabling Technologies for Fast Non-linear Data Assimilation in a Coastal Margin Observatory, PhD Thesis, Oregon Health and Science University, 2007.
[33] V.M. Krasnopolsky, D.V. Chalikov, H.L. Tolman, A neural network technique to improve computational efficiency of numerical oceanic models, Ocean Modeling 4 (2002) 363–383.
[34] H.L. Tolman, V.M. Krasnopolsky, D.V. Chalikov, Neural network approximations for nonlinear interactions in wind wave spectra: direct mapping for wind seas in deep water, Ocean Modeling 8 (2005) 253–278.
[35] V.M. Krasnopolsky, Neural network emulations for complex multi dimensional geophysical mappings: applications of neural network techniques to atmospheric and oceanic satellite retrievals and numerical modeling, Reviews of Geophysics 45 (2007) RG 3009. doi:10.1029/2006RG000200.
[36] V.M. Krasnopolsky, M.S. Fox-Rabinovitz, H.L. Tolman, A.A. Belochitski, Neural network approach for robust and fast calculation of physical processes in numerical environmental models: compound parameterization with quality control of larger errors, Neural Networks 21 (2–3) (2008) 535–549.
[37] F.P. Harter, H.F. Campos Velho, New approach to applying neural network in nonlinear dynamic model, Applied Mathematical Modelling 32 (2008) 2621–2633.
[38] C. Bishop, Neural Network for Pattern Recognition, Oxford University Press, 1995.
[39] R. Hanea, G.J.M. Velders, A.J. Segersc, M. Verlaan, A.W. Heemink, A hybrid Kalman filter algorithm for large scale atmospheric-chemistry data assimilation, Monthly Weather Review 135 (1) (2007) 140–151.
[40] G. Burgers, G. Evensen, On the analysis scheme in the ensemble Kalman filter, Monthly Weather Review (126) (1998) 1719–1724.
[41] J. Gibson, B. Koo, S. Gray, Filtering of colored noise for speech enhancement and coding, IEEE Transactions on Signal Processing 39 (8) (1991).
[42] J. Moody, The effective number of parameters: An analysis of generalization and regularization in non-linear learning systems, in: Advances in Neural Information Processing System, vol. 4, Palo Alto, CA, 1992, 847–854.
[43] W.W. Hsieh, Nonlinear principal component analysis by neural networks, Appeared in Tellus, vol. 53A, 2001, 599–615.
[44] Y. Teh, S. Roweis, Automatic alignment of local representations, Advances in Neural Information Processing Systems 15 (2003).