# A new methodology for using buoy measurements in sea wave data assimilation

George Emmanouil · George Galanis · George Kallos

**Abstract** One of the main drawbacks in modern sea wave data assimilation models is the limited temporal and spatial improvement obtained in the final forecasting products. This is mainly due to deviations coming either from the relevant atmospheric input or from the dynamics of the wave model, resulting to systematic errors of the forecasted fields of numerical wave models, when no observation is available for assimilation. A potential solution is presented in this work, based on a combination of advanced statistical techniques, data assimilation systems, and wave models. More precisely, Kalman filtering algorithms are implemented into the wave model WAM and the results are assimilated by an Optimum Interpolation Scheme, in order to extend the beneficial influence of the latter in time and space. The case studied concerns a 3-month period in an open sea area near the South-West coast of the USA (Pacific Ocean).

**Keywords** Data assimilation · Kalman filters · Wave modeling

G. Emmanouil · G. Galanis · G. Kallos (✉)
Division of Applied Physics, Atmospheric Modeling and Weather
Forecasting Group, School of Physics, University of Athens,
University Campus, Bldg. PHYS-V,
Athens 15784, Greece
e-mail: kallos@mg.uoa.gr

G. Emmanouil
Hellenic National Meteorological Service,
E. Venizelou 14,
Helliniko 16777, Greece

G. Galanis
Section of Mathematics, Hellenic Naval Academy,
Chatzikyriakion,
Piraeus 18539, Greece

## 1 Introduction

The need of high-quality information and forecasting related to sea state conditions is continuously increasing for different applications. Searouting and ship safety, commercial transportation, marine pollution, climate change issues, wave energy production, and other important applications are heavily dependent on an accurate sea state knowledge.

Nowadays, wave analysis and predictions are mainly based on the combined use of Numerical Wave and Atmospheric models, as well as on Data Assimilation (DA) systems (see Abdalla et al. 2005; Siddons 2007; Greenslade and Young 2004; Breivik and Reistad 1994; Greenslade 2001; Voorrips et al. 1997). The latter exploits properly any available sea state information, in order to produce initial conditions of higher quality that results to the improvement of the final wave forecasts. The most common technique in DA systems is the Optimum Interpolation (OI). Other DA methods for ocean waves have also been examined, e.g., an efficient low-rank approximation to the Kalman filter (KF) presented by Voorrips et al. (1999).

However, such systems have usually limited ability of improving the quality of the final predictions, in time and space, especially that of long-period forecasting horizons (Emmanouil et al. 2007). This is mainly due to the fact that the discrepancies in the forecasted atmospheric forcing fields, as well as those coming from the wave model evolution, re-emerge inside the forecasting period when no external information-observation is available to be assimilated. Furthermore, the restricted availability of wave observations contributes to the limited assimilation impact. In this way, the obtained improvement in the forecasting period lasts for only a few hours.

In this work, advanced statistical techniques, based on KFs, are employed in combination with the wave prediction system (WAM) in order to provide additional information that may be used as input ("forecasted") observations for DA inside the forecasting period. These datasets are, in fact, improved model forecasts, properly corrected by the recursive use of recent observations through Kalman filtering techniques. In this way, the temporal and spatial impact of DA systems may be extended. It should be noted that KF are used as an independent statistical methodology that "generates observations" within the forecasting period. This kind of observations is used in a second step, by a classical OI scheme.

Kalman filters (Kalman 1960; Kalman and Bucy 1961; Kalnay 2002) have been employed in combination with observations, at several previous works (see e.g. Evensen 2003, 2004; Galanis and Anadranistakis 2002; Galanis et al. 2006; van der Grijn 2009; Persson 1990) as post-processes for the elimination of the systematic bias from the predicted values of atmospheric and wave models.

The novelty of the proposed methodology is the incorporation of such filters into the wave model and not their use as an external post-procedure. Their results are spatially propagated by the subsequent use of the OI. In this way, the obtained forecasts are not treated just as time series coming from a mathematical method, but they also take into consideration the physical processes simulated by the wave model. This is the main objective of this work: to improve the benefits of DA, in time and space, by using two different methods, Kalman filters and OI-DA.

The paper is organized as follows: the description of the wave model and the data assimilation scheme, as well as the Kalman filter algorithms and the necessary modifications for its introduction into the WAM system are presented in Section 2. The model configuration and the application of different techniques are discussed in Section 3. In Section 4, the results are discussed while the main conclusions are summarized in Section 5.

## 2 Models and post-processes description

### 2.1 Wave model and data assimilation scheme

In this study, the wave model used is WAM Cycle IV (European Centre for Medium-Range Weather Forecasts—ECMWF—version). A detailed description and presentation of the model can be found in WAMDI group (1988), Komen et al. (1994), Jansen (2000) and Bidlot and Janssen (2003), while the configuration used in this study is presented in Section 3.

The analysis fields were corrected by the data assimilation (DA) scheme developed at ECMWF for the WAM model (Lionello et al. 1992), which is based on an OI method, as

outlined in Lorenc (1981). The DA procedure consists of two steps. Firstly, an analyzed field of significant wave heights is created by optimum interpolation. Then, the full two-dimensional wave spectrum is retrieved by this field from a first-guess spectrum, in order to transform the information of a single wave height measurement into different corrections for the wind sea and swell parts of the spectrum. This method identifies the sea state as wind sea or swell without further discretization of the spectral components. Therefore, it corrects the two-dimensional spectrum by introducing appropriate rescaling factors to the energy and frequency scales of the wind sea and swell. Additionally, it updates the local forcing wind speed. The computation of the rescaling factors is performed for two classes of spectra: wind sea spectra, for which the rescaling factors are derived from fetch and duration growth relations, and swell spectra, where it is assumed that the wave steepness is conserved. A more detailed description of the method can be found in Komen et al. (1994).

### 2.2 Kalman filter algorithm and modifications

A brief description of the general form of a Kalman filter is following using the unified notation proposed by Ide et al. (1997). Such filters simulate the evolution in time of an unknown process (state vector), whose observational value at time $t_i$ is denoted by $x^t(t_i)$. The latter is combined with corresponding observations $y_i^O$. The change of $x$ in time and the relation between the observation and the unknown vectors are described by the following (observation and system, respectively) equations:

$$x^t(t_i) = M_{i-1}[x^t(t_{i-1})] + \eta(t_{i-1}), \; y_i^O = H_i[x^t(t_i)] + \varepsilon_i \quad (1)$$

The system operator $M_{i-1}$, the observational one $H_i$ as well as the covariance matrices of the Gaussian (non-systematic) errors $\eta(t_i)$ and $\varepsilon_i$, respectively, have to be determined before the application of the filter. In particular, for the definition of the covariance matrices $Q(t_i)$, of the system equation, and $R(t_i)$, of the observation equation, the following methodology has been adopted: their calculation is based on the sample of the last seven values $\eta(t_i) = x^t(t_{i+1}) - x^t(t_i)$ and $\varepsilon_i = y_i^O - H_i[x^t(t_i)]$, respectively, a choice that allows the fast adjustment to possible change of data and, at the same time, does not increase significantly the computational cost (Galanis et al. 2006):

$$Q(t_i) \equiv \frac{1}{6} \cdot \sum_{i=0}^{6} \left( \left( (x^t(t_{i+1}) - x^t(t_i)) - \left( \frac{\sum_{i=0}^{6}(x^t(t_{i+1}) - x^t(t_i))}{7} \right) \right) \right)^2,$$

$$(2)$$

$$R(t_i) \equiv \frac{1}{6} \cdot \sum_{i=0}^{6} \left( \left( (y_i^O - H_i[x^t(t_i)]) - \left( \frac{\sum_{i=0}^{6} (y_i^O - H_i[x^t(t_i)])}{7} \right) \right) \right)^2. \tag{3}$$

An initial forecast of the state vector x and its error covariance matrix P are given by:

$$x^f(t_i) = M_{i-1}[x^a(t_{i-1})], \tag{4}$$

$$P^f(t_i) = M_{i-1}P^a(t_{i-1})M_{i-1}^T + Q(t_{i-1}),$$

and they are followed up by an update in which the observations available at time $t_i$ are implemented:

$$x^a(t_i) = x^f(t_i) + K_i(y_i^O - H_i[x^f(t_i)]), \tag{5}$$

$$P^a(t_i) = (I - K_iH_i)P^f(t_i)$$

The matrix

$$K_i = P^f(t_i)H_i^T[H_iP^f(t_i)H_i^T + R_i]^{-1} \tag{6}$$

is referred as the *Kalman gain* and describes how easily the filter adjusts to possible new conditions. The superscripts *o*, *t*, *f*, and *a* denote observations, true, forecast, and analysis values correspondingly. Moreover, *T* and −1 are the classical symbols of the transpose and the inverse matrix respectively, while *I* stands for the unitary matrix.

Equations 1–6 update the Kalman algorithm from time $t_{i-1}$ to $t_i$ (see also Kalman 1960; Kalman and Bucy 1961; Persson 1990; Kalnay 2002; Galanis and Anadranistakis 2002; Galanis et al. 2006).

In most of the cases, the Kalman algorithm, described earlier, was utilized as post process and a single filter was

employed for each area of interest. The use of such filters as post-processes for the elimination of systematic errors was studied in Galanis et al. (2009). In the proposed approach, the KF are used within the forecasting period, at the southwest coast of USA (shown in Fig. 1), resulting to improved WAM forecasts, as illustrated in Fig. 2. The issue addressed here is the exploitation of these results as "forecasted" observations, in combination with an OI assimilation scheme so, to improve wave prediction by maximizing the gain of KF and assimilation.

In this study, the Kalman filter and the wave model were modified, in order to adapt the filter into the wave system. These modifications were important so to:

- Extend the data sources that may be used by the DA scheme.
- Make feasible the use of Kalman filters with the available observational timeseries in real time.
- Use discrete Kalman filters in different areas.
- Optimize the calculation of the covariance matrices Q and R in an operational environment.

More precisely, the Kalman filters were implemented in WAM after the time integration of the two-dimensional frequency-direction wave spectra and before the DA, as shown in Fig. 3a. During the assimilation window (between $T_{start}$ and $T_0$ in Fig. 3b), the Kalman filter is running in a training mode (using the same timestep as the DA) and it is calculating the necessary parameters (e.g., correlation matrices) based on the relevant prediction of the model (value 1 in Fig. 3b) and the observations (value 2). The output of the KF in the assimilation window is denoted with number 3. Then, the DA scheme is employing the WAM value and the observation. It corrects the SWH (value 4 in the same figure), in order to produce corrected



Fig. 1 The study area and the buoys used for assimilation (*B*, *C*, *E*, *F*) and for evaluation (*A*, *D*, *G*). The *black rectangle* denotes the area of the comparison with satellite data (map from Google Earth)
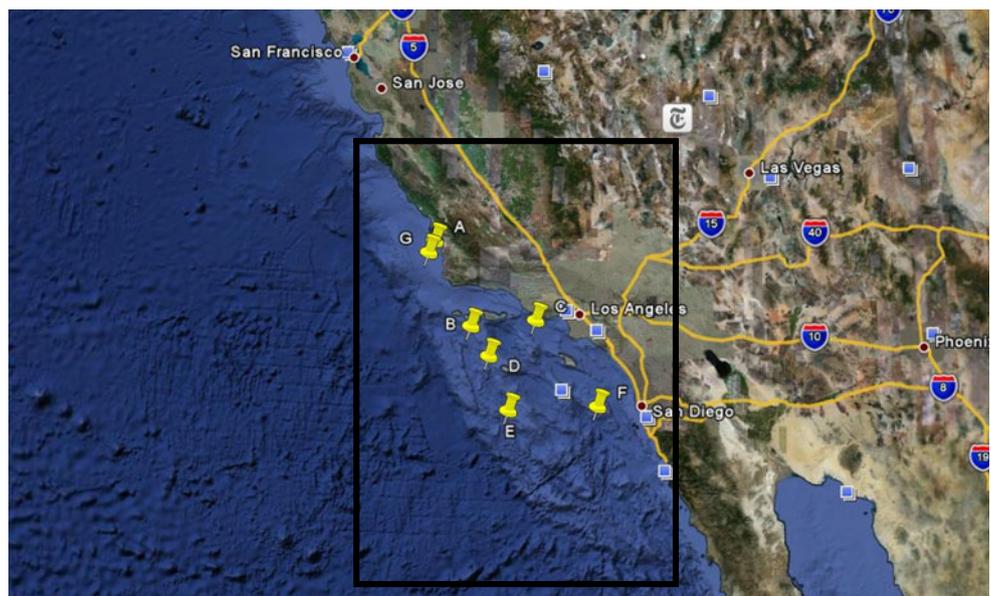
**Fig. 2** Application of KF as
post process for the production
of forecasted observations for
the areas of: **a** buoy C
and **b** buoy F
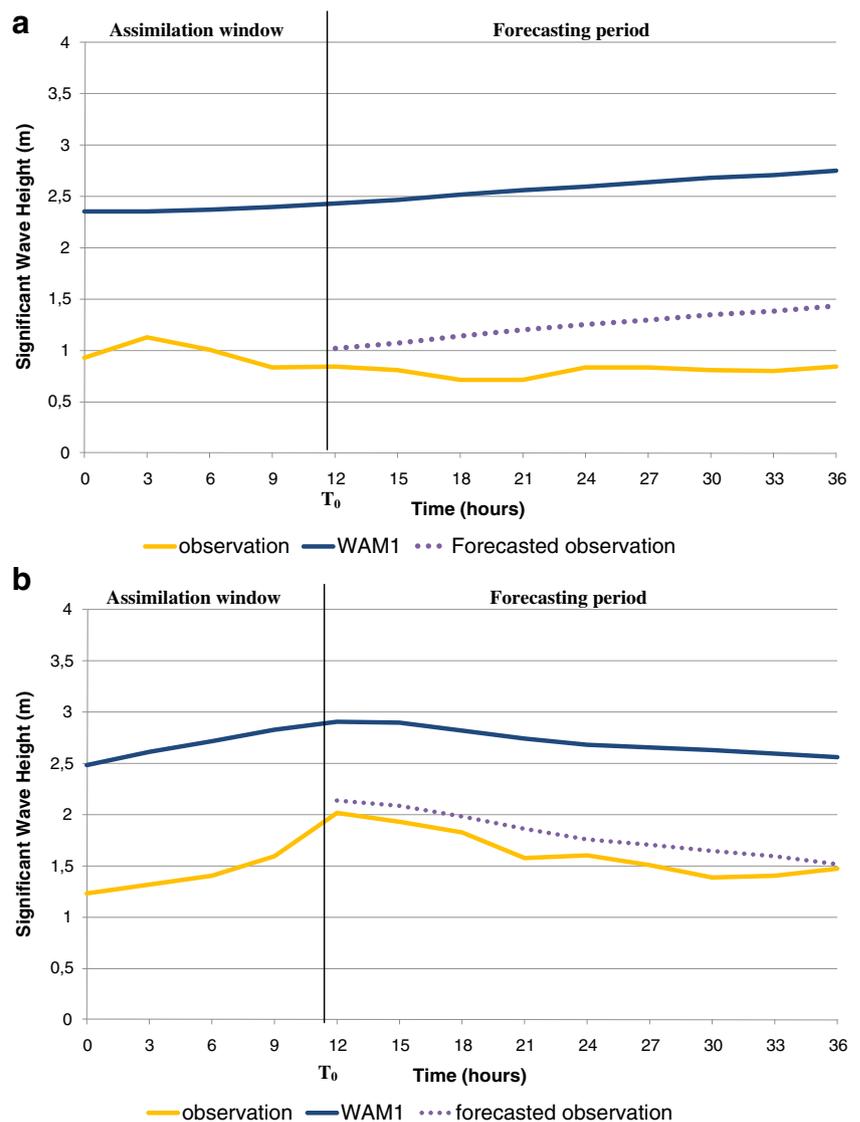


initial fields (analysis) at the analysis time $T_0$ (value 8), by
the combination of the WAM and the observational values
5 and 6 accordingly in Fig. 3b, when Kalman outputs the
value 7. At time $T_0$, the forecasting period begins and there
is no availability of observations. Though, the DA scheme
remains activated and uses as input the "forecasted
observations" computed by KF values 10, produced at
each assimilation time step, and the WAM values 9. The
assimilated value that results from this procedure is 11.
Summarizing the above procedure, the model predictions
are bias corrected by the use of the Kalman filters for areas
where there are observations. Then, the corrected values are
treated as "forecasted observations" and they are used as
input for DA.

In this work, it was necessary to have one multiple filter,
containing different filters for different areas where obser-
vations are available. This was essential so to exploit

observations that represent different local characteristics
(such as bathymetry and coastline) and different wave
climate regions (wind sea or swell dominated). In this
way, the output of the Kalman-filtered results will better
represent the special wave behavior of each observa-
tional area. For this purpose, the relevant system and
observation equations of the $\nu$ observation area takes
the form:

$$x^t(t_{i+1}, v) = M_{i-1}(v)x^t(t_i, v) + n(t_i, v), \tag{7}$$

$$y_i^{o,v} = H_i(v)[x^t(t_i, v)] + \varepsilon_i(v),$$

The calculation method of the covariance matrices $Q$
and $R$ was also modified. More precisely, in a classical
Kalman application, the estimation of these matrices is
based on an already-known time series of observations
(Eqs. 2 and 3). This was not possible in this study since

**Fig. 3 a** Flowchart of the new WAM. **b** Time integration in the new WAM with Kalman filters and DA. The steps inside the assimilation window are: *1–4*; at the analysis time: *5–8*, and during the forecasting period: *9–11*
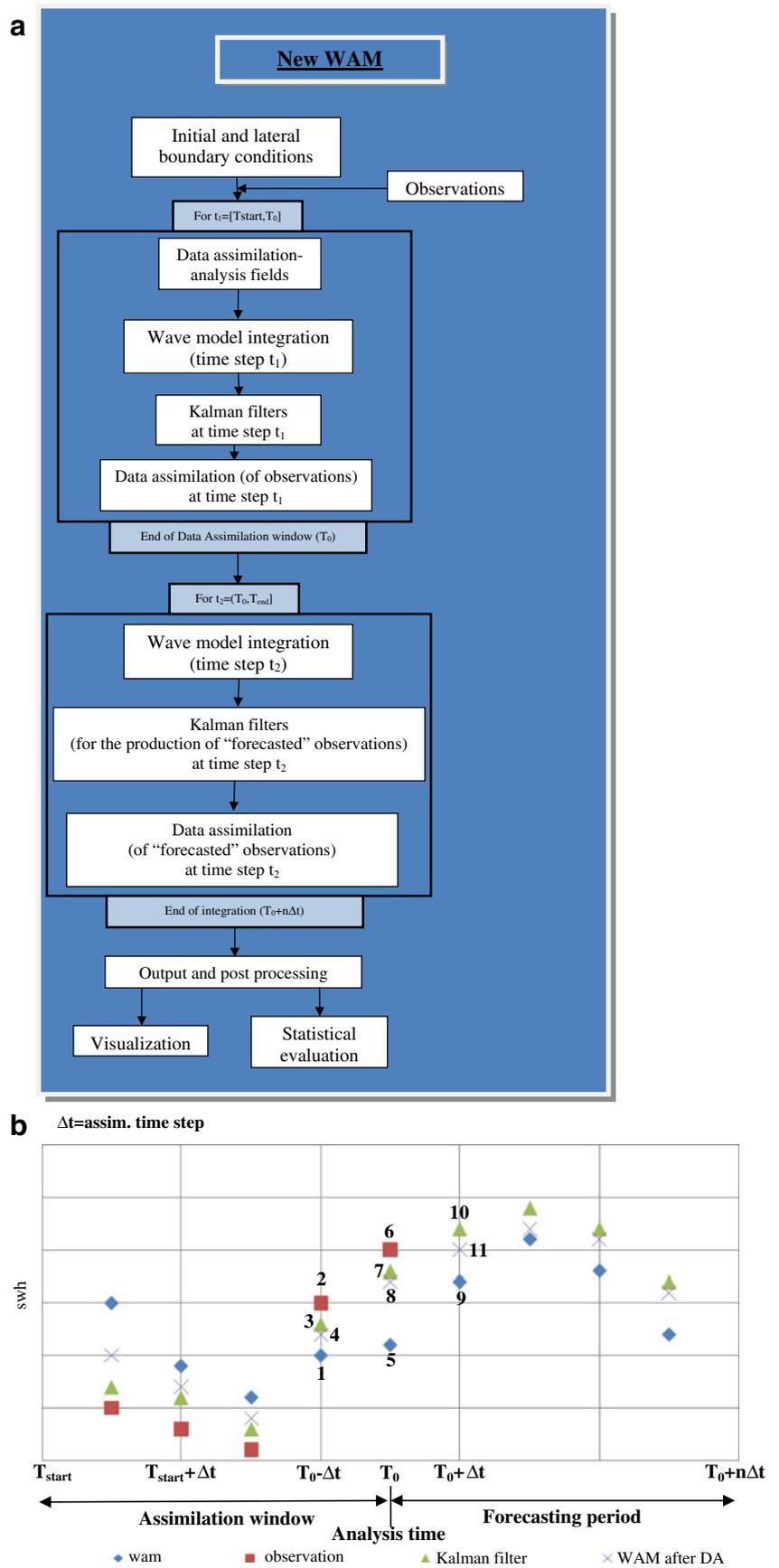
**Table 1** Buoy coordinates

| Buoy Label | Lat | Lon | Water depth (m) |
|---|---|---|---|
| A | N 34.88 | W 120.87 | 204 |
| B | N 33.65 | W 120.20 | 1,004 |
| C | N 33.75 | W 119.08 | 881 |
| D | N 33.22 | W 119.88 | 335 |
| E | N 32.43 | W 119.53 | 1,393 |
| F | N 32.50 | W 118.00 | 1,856 |
| G | N 34.27 | W 120.70 | 384 |

the Kalman filter is activated inside the forecasting period as part of the wave system, and therefore no observations are available. To overcome this problem, alternative ways of calculation were adopted and are described in detail in Section 3. On the other hand, the equations concerning the Kalman gain for the $\nu$ observational location become:

$$P^f(t_i, \nu) = M_{i-1}(\nu) P^a(t_{i-1}, \nu) M_{i-1}^T(\nu) + Q(t_{i-1}, \nu), \quad (8)$$

$$K_i(\nu) = P^f(t_i, \nu) H_i^T(\nu) \left[ H_i P^f(t_i, \nu) H_i^T(\nu) + R_i(\nu) \right]^{-1} \quad (9)$$

The Kalman filter was applied to a single forecasted parameter: the significant wave height (SWH). The corresponding bias $y_i^O$ was estimated as a function of the forecasting model direct output $SWH_i$, as proposed by Galanis et al. (2006):

$$y_i^o(\nu) = a_{0,i}(\nu) + a_{1,i}(\nu) swh_i(\nu) + a_{2,i}(\nu) SWH_1^2(\nu) + \varepsilon_i(\nu), \quad (10)$$

and the coefficients $\{\alpha_{0,i}(\nu) \; \alpha_{1,i}(\nu) \; \alpha_{2,i}(\nu)\}$ have to be estimated by the filter. Parameter $\varepsilon_i$ stands for the Gaussian (non-systematic) error of the previous procedure. In this way the state vector of the filter becomes $x(t_i,\nu)=[\alpha_{0,i}(\nu) \; \alpha_{1,i}(\nu) \; \alpha_{2,i}(\nu)]^T$, the bias $y_i^O(\nu)$ is used as known parameter, the observation matrix takes the form $H_i(\nu)=[1 \; SWH_i(\nu)$

$SWH_i^2(\nu)]$ and the system matrix is the three-dimensional identity. Therefore, the system and observation Eq. 7 take the following initial values:

$$x = 0, y_0^O(\nu) = \varepsilon_0, \; P(t_0, \nu) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \quad (11)$$

$Q(t_0, \nu) = I_3, \; R(t_0, \nu) = 6.$

No correlations between different coordinates of the state vector $x$ are assumed. The high values for $R$ and the diagonal elements of $P$ indicate low credibility of the first guess and ensure fast adjustment to new conditions. The obtained KF-estimated bias $y_i^o$ is then used for the correction of the forecasted significant wave height.

It is worth noticing that the selection of a non-linear function in KF (Eq. 10) in the current study compensates, at least partially, the disadvantage of the application of such linear filters in wave models.

## 3 Model configuration and applications

In this study, the wave model WAM (cycle IV) was running globally, with horizontal resolution of $1.0\times1.0°$. The main target was to test the proposed methodology in a simple configuration and not to build an operational system. It is worth noticing that analogous studies performed for atmospheric models, showed that the use of successful KF do not really depend on the resolution used (Galanis et al. 2006). The wave spectrum was discriminated on 30 frequencies and 24 directions. The lowest frequency was defined to 0.0417 Hz, while the propagation time step was set to 300 s. The model ran on a deep water mode with no refraction. The necessary atmospheric input (10-m wind speed and direction) was obtained by NCEP/GFS global model (horizontal grid resolution $1.0\times1.0°$) by a time step of 6 h.

The area of interest was the southwest coast of USA (Fig. 1). This choice was made because this is an open sea area, where one may study in detail the response of the proposed system with low dependency on local

**Table 2** Characteristics of the WAM versions used

| | Data assimilation | Kalman filters | Calculation of matrices Q and R of the Kalman filters | | |
|---|---|---|---|---|---|
| | | | Continuous-dynamic | Semi continuous | Climatological |
| WAM1 | x | x | x | x | x |
| WAM2 | √ | x | x | x | x |
| WAM3 | √ | √ | √ | x | x |
| WAM4 | √ | √ | x | √ | x |
| WAM5 | √ | √ | x | x | √ |

**Fig. 4** The results of one cycle of WAM2 at the areas of buoys B (**a**) and C (**b**)
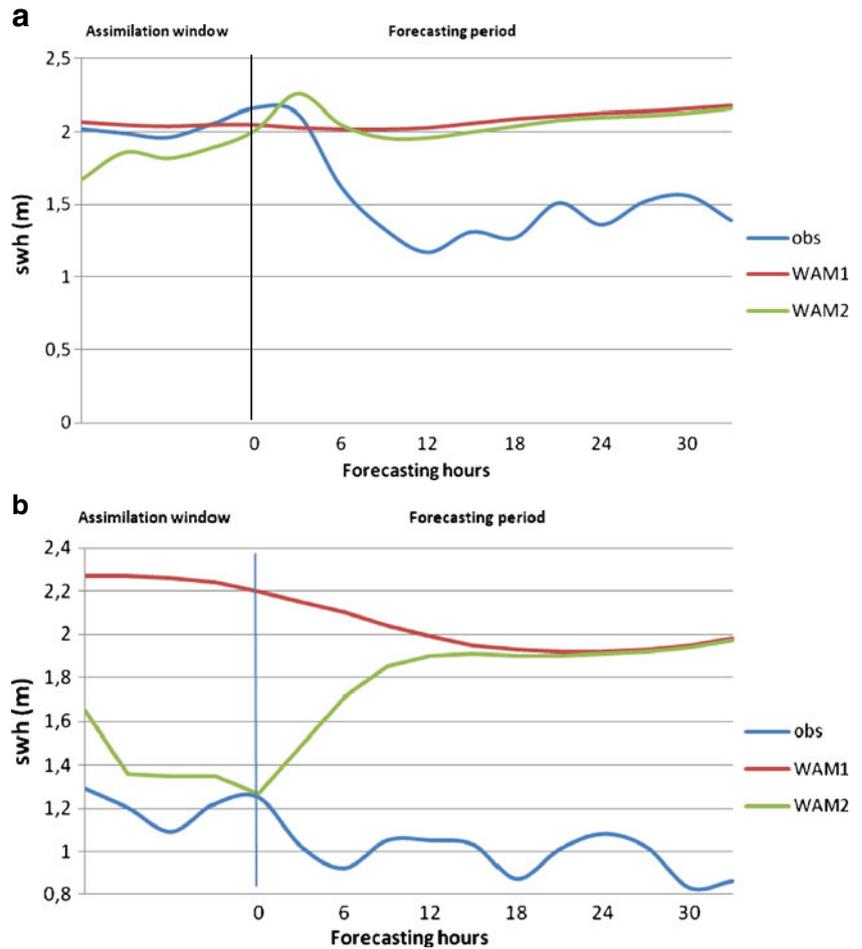


characteristics. The position of the buoys (NOAA/ National Data Buoy Center network) employed as observational sources and for evaluation purposes are indicated also in Fig. 1. The buoys selected for assimilation were chosen so to cover the major part of the study area. On the other hand, the independent buoys target to the evaluation of the area around the "assimilated" ones (buoy D), as well as to explore the spatial impact in the neighborhood (buoys A and G). Their exact coordinates are listed in Table 1.

**Fig. 5** Results of WAM experiments for buoy B. In WAM3-5 experiments, apart from the observations employed within the assimilation window, Kalman filter projections in time are assimilated inside the forecasting period
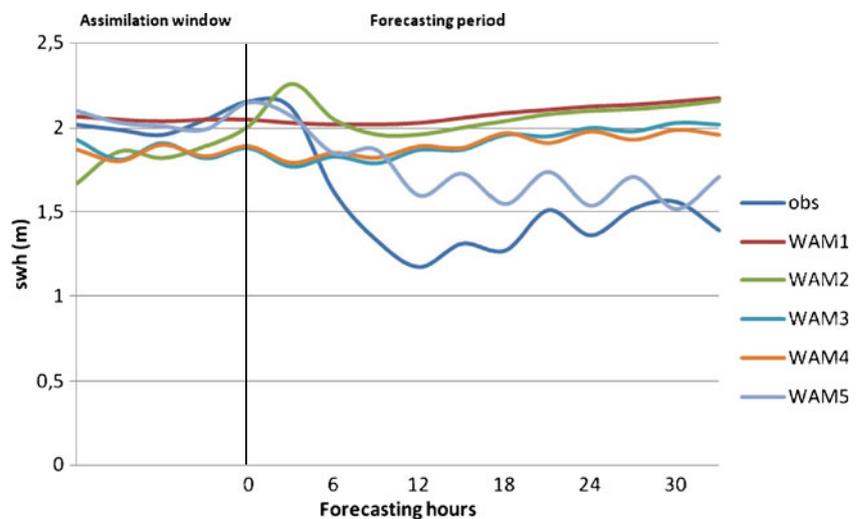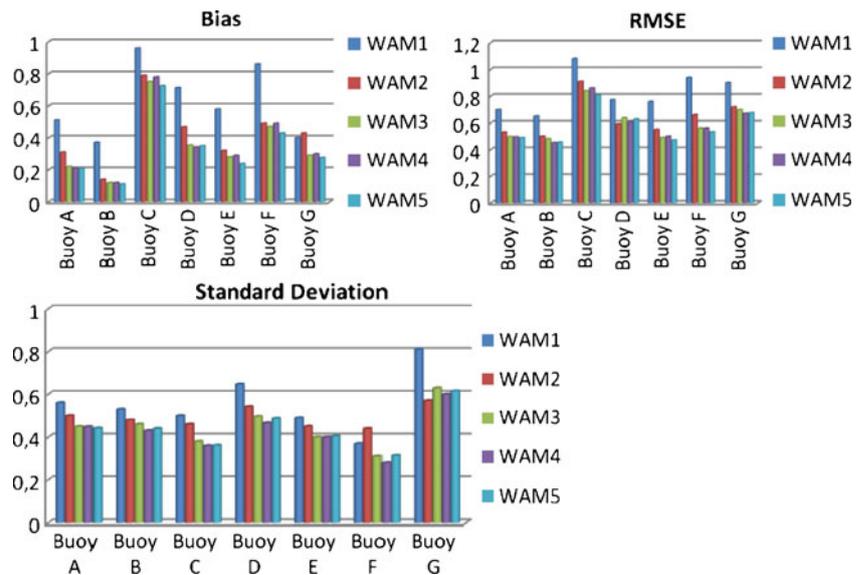
Five different experimental versions of WAM were employed for a 3-month period (October–December 2006), as shown in Table 2:

1. A first one (referred from now on as WAM1) did not use any assimilation system.
2. The second version (WAM2) used the DA system described in Section 2 (Lionello et al. 1992) that is widely used from operational centers and meteorological services worldwide and is based on an OI technique. This DA scheme assimilates buoy observations, available until the start of the 36-h forecasting period (analysis time) as illustrated in Fig. 4.

   It becomes obvious from the previous figure that during the assimilation window (until the analysis time), the model results are significantly improved, approaching the corresponding observation values. However, this positive impact decreases by time and almost disappears after a period of 12-h. This is due to the fact that there are no available observations to be assimilated during the forecasting period and the initially emerged discrepancies appear again. It should be noted that the buoy measurements have fluctuations in time. The latter are not simulated by the wave model, since it is known that the model results are smoothed in time and space.

3. A third version of WAM (WAM3) assimilated two different observation types:

   a. The buoy observations (also used by WAM2) in the assimilation window,
   b. Improved-filtered forecasts of WAM, obtained by Kalman filters, which are used as "generated observations" inside the 36-h forecasting period (Fig. 5), and then employed by the independent OI assimilation system (as shown in Fig. 3). In these values, an important part of the systematic error has been removed. By this way the assimilation impact was extended to the entire forecasting period. Moreover, one may notice that the time fluctuations of the buoy measurements mentioned earlier (WAM2) are now better simulated by the wave model.

   The Kalman filter covariance matrices $Q(t_i)$, $R(t_i)$, of the system and observation equation respectively (Eqs. 2 and 3), are calculated by the last seven values of $n(t_i)$ and $\varepsilon_i$ (which are available at 3-h intervals at the present study), a period that resulted as the optimum one after a series of relevant tests (Galanis et al. 2006). On the other hand, this choice allows fast adjustment to possible

**Table 3** Mean values at the areas of the buoys used by the DA (buoys B, C, E, and F)

| Buoy SWH average=2.9m | WAM1 | WAM2 | WAM3 | WAM4 | WAM5 |
|---|---|---|---|---|---|
| Bias | 0.69 | 0.44 | 0.41 | 0.42 | 0.38 |
| RMSE | 0.86 | 0.66 | 0.59 | 0.59 | 0.57 |
| Nbias | 0.62 | 0.41 | 0.40 | 0.41 | 0.38 |
| SD | 0.47 | 0.46 | 0.39 | 0.37 | 0.38 |

**Table 4** Improvement percentage at the areas of the buoys used by the DA (buoys B, C, E, and F)

|  | WAM3 vs. WAM1 | WAM3 vs. WAM2 | WAM4 vs. WAM1 | WAM4 vs. WAM2 | WAM5 vs. WAM1 | WAM5 vs. WAM2 |
|---|---|---|---|---|---|---|
| Bias | 40.5 | 6.8 | 39.1 | 4.5 | 44.9 | 13.6 |
| RMSE | 31.4 | 10.6 | 31.4 | 10.6 | 33.7 | 13.6 |
| Nbias | 35.5 | 2.4 | 33.9 | 0 | 38.7 | 7.3 |
| SD | 17.0 | 15.2 | 21.3 | 19.6 | 19.1 | 17.4 |

changes of the time series in study. The values used for this calculation are either observations (into the assimilation window), or "forecasted observations" from the Kalman filter (into the forecasting period). The aim of this methodology is to explore the advantages obtained by the continuous-dynamical calculation of the previous matrices and, therefore, to study the adjustment of the filter to the new conditions appearing in the forecasting period. Finally, it should be mentioned that each observational area employs its own characteristic values for the Kalman filter parameters, since a different filter is employed at each observational area, ensuring the best interpretation of the local environment.

4. The fourth test (WAM4) is the same with WAM3, using again KF—"forecasted observations" as input to the OI—assimilation scheme, except from the fact that the Kalman covariance matrices ($Q$ and $R$) are updated only when observations are available (inside the assimilation window). During the forecasting period, a mean value is used which is calculated from the last seven observations (Fig. 5). By this way, the use of the most recent observational behaviur compared with the one of the model is maximizing the possibility for better results in the next hours.

5. Finally, WAM5 uses again KF-forecasted observations as referred in WAM3 and 4, while for the calculation of KF covariance matrices $Q$ and $R$, it was used a climatological calculation based on the observations and the modeled values over a period of more than 2 years (Aug 2004–Sept 2006).

The results obtained from the above simulations (WAM1-5) were evaluated both for their forecast accuracy and their assimilation impact in time and space. The statistical analysis was based on the following parameters:

1. Bias of forecasted values:

$$\text{Bias} = \frac{1}{k} \cdot \sum_{i=1}^{k} \left( \text{for}(i) - \text{obs}(i) \right) \quad (12)$$

Here obs*(i)* denotes the recorded (observed) value at time $i$, for*(i)* the respected forecast and $k$ the size of the sample.

2. Normalized Bias (N.Bias):

$$\text{N.Bias} = \frac{1}{k} \sum_{i=1}^{k} \left| \frac{\text{for}(i) - \text{obs}(i)}{\text{obs}(i)} \right| \quad (13)$$

where | | stands for the absolute value, revealing the normalized divergence of the forecasts as a proportion of the observations.

3. Root mean square error (RMSE) and standard deviation (SD) of the error (two classical variation and divergence measures respectively):

$$\text{RMSE} = \sqrt{\frac{1}{k} \cdot \sum_{i=1}^{k} \left( \text{for}(i) - \text{obs}(i) \right)^2},$$
$$\text{SD} = \sqrt{\frac{1}{k} \cdot \sum_{i=1}^{k} \left( \left( \text{for}(i) - \text{obs}(i) \right) - \text{Bias} \right)^2} \quad (14)$$

## 4 Results and discussion

As it has been mentioned in previous sections, the main purpose of this study is to propose a new technique for the extension in time and space of the positive impact of the

**Table 5** Mean values at the areas of the buoys used by the DA (buoys B, C, E, and F)

| Buoy SWH average=3.3m | WAM1 | WAM2 | WAM3 | WAM4 | WAM5 |
|---|---|---|---|---|---|
| Bias | 0.54 | 0.40 | 0.29 | 0.28 | 0.28 |
| RMSE | 0.79 | 0.61 | 0.61 | 0.59 | 0.60 |
| Nbias | 0.42 | 0.29 | 0.29 | 0.28 | 0.28 |
| SD | 0.67 | 0.54 | 0.53 | 0.50 | 0.52 |

**Table 6** Improvement percentage at the areas of the buoys used by the DA (buoys B, C, E, and F)

|  | WAM3 vs. WAM1 | WAM3 vs. WAM2 | WAM4 vs. WAM1 | WAM4 vs. WAM2 | WAM5 vs. WAM1 | WAM5 vs. WAM2 |
|---|---|---|---|---|---|---|
| Bias | 46.3 | 28.0 | 48.1 | 30.0 | 48.1 | 30.0 |
| RMSE | 22.3 | 0 | 25.3 | 3 | 24 | 1.5 |
| Nbias | 31 | 0 | 33.3 | 3 | 33.3 | 3 |
| SD | 20.9 | 2 | 25.4 | 7 | 22.4 | 4 |

**Fig. 7** Scatter diagrams for buoy location E
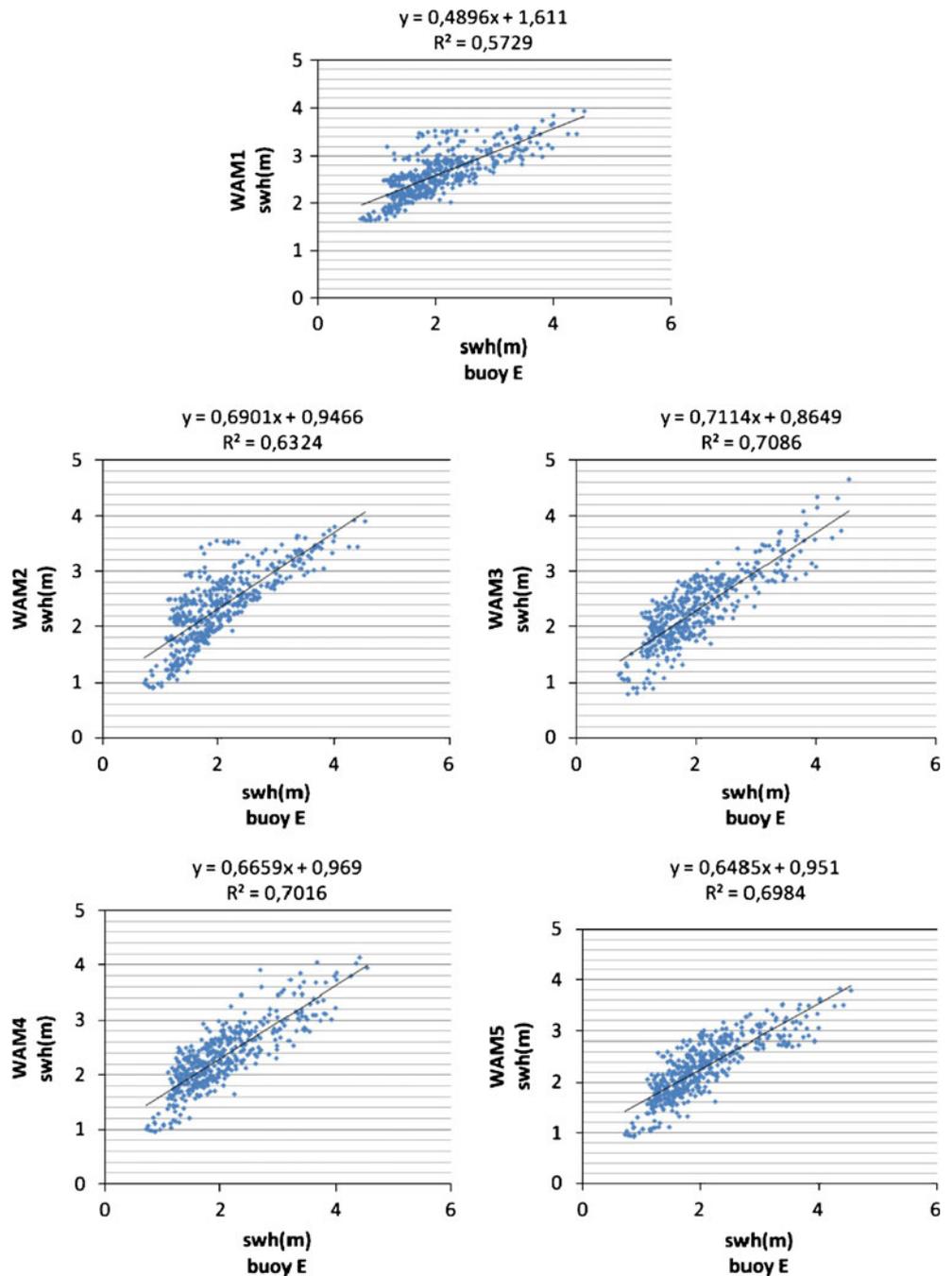
**Table 7** Evaluation against ENVISAT RA-2 data

| RA2 SWH average=2.5m | WAM1 | WAM2 | WAM3 | WAM4 | WAM5 |
|---|---|---|---|---|---|
| Bias | 0.50 | 0.37 | 0.06 | 0.08 | 0.05 |
| RMSE | 0.74 | 0.70 | 0.53 | 0.53 | 0.51 |
| Nbias | 0.39 | 0.34 | 0.26 | 0.26 | 0.25 |
| SD | 0.55 | 0.59 | 0.54 | 0.54 | 0.50 |

assimilated observations, since, in wave modeling, classical DA schemes influence the quality of wave forecasting for only few hours (as already illustrated in Fig. 4).

A step forward to the elimination of this drawback was presented in Galanis et al. (2009). In that study, the way that Kalman filters can be used to improve the direct outputs of an initial wave model run WAM2, where DA is used in the assimilation window employing the available observations, as shown in Fig. 3 was described. The most important advantage gained by Kalman filters was the significant reduction of possible systematic errors in WAM forecasts. These filtered forecasts were then assimilated (by OI) in a second model run inside the forecasting period.

In the present work, Kalman filters are incorporated into the WAM model and, by this way, it is introduced a new integrated wave prediction system. In this system, KF "forecasted observations" are assimilated by OI inside the forecasting period. Of course, these forecasts are not as accurate as the buoy observations. However, the proper use of the Kalman filters, as described in Section 2, leads to the reduction of the systematic error. Therefore, the assimilation of these values inside the forecasting period—where no other information exists—leads to a considerable extension of the assimilation impact, as well as to the improvement of the final forecast.

A main advantage of this methodology is the correction of the bias between model direct forecasts and buoy measurements, even in cases where such discrepancies do not have a constant behavior. As a matter of fact, such deviations may change in time due to the change of the weather and wave patterns during the forecasting period. Kalman filters can dynamically correct the model forecasts, as shown in the characteristic example of Fig. 5. Moreover, the subsequent use of these improved forecasts as "observations" (produced by the Kalman filter projection in time) by the DA scheme, ensures that the relevant correction will spread over a larger area than the point location of buoys.

Concerning the time extension of the assimilation impact, the results illustrated in Fig. 5, for buoy B are characteristic. WAM3-5 forecasts are more accurate than WAM1-2 and due to the new methodology applied. WAM5—with climatic calculation of the covariance matrices—seems to "follow" better the observed sea state conditions. This leads to the conclusion that the best interpretation of the wave conditions is succeeded when the Kalman filter covariance matrices (Q and R) are calculated from a long time series. However, the dynamical way of calculation (WAM3-4) also improves the forecasts. Similar performance was obtained for the rest of the buoys.

A direct result of this extended effect of the DA was the reduction of the bias in the final forecasted values for the major part of the area of interest. In Fig. 6, statistical results are presented comparing the model output and the corresponding observed values from all the buoys for the entire 3-month study period. It should be noted that these comparisons concerned only the forecasting period. In general, for the present application the wave model over-estimated the SWH in comparison with the relevant measurements. This may be attributed to the dynamics of the wave model (Emmanouil et al. 2007) and the atmospheric input. Moreover, the results seem to be better for deep water areas, since the configuration used fits well to such cases (global grid).

For all the test cases, the use of the classical DA scheme (WAM2) leads to an improvement of the final forecasts, but not to the considerable elimination of the existing bias. The additional use of the proposed methodology (WAM3-5)

**Table 8** Average statistics of all buoys of the five different runs for: the first forecasting period (FP1: 0–12 h of forecasts), the second forecasting period (FP2: 12–24 h of forecasts), and the third forecasting period (FP3: 24–36 h of forecasts)

| | WAM1 | WAM2 | WAM3 | WAM4 | WAM5 |
|---|---|---|---|---|---|
| Total statistics (FP1) | | | | | |
| Bias | 0.63 | 0.42 | 0.40 | 0.40 | 0.36 |
| RMSE | 0.86 | 0.66 | 0.61 | 0.61 | 0.59 |
| Nbias | 0.54 | 0.40 | 0.38 | 0.37 | 0.35 |
| SD | 0.54 | 0.48 | 0.45 | 0.43 | 0.44 |
| Total statistics (FP2) | | | | | |
| Bias | 0.62 | 0.42 | 0.35 | 0.36 | 0.34 |
| RMSE | 0.87 | 0.68 | 0.60 | 0.60 | 0.58 |
| Nbias | 0.53 | 0.24 | 0.36 | 0.37 | 0.36 |
| SD | 0.57 | 0.50 | 0.46 | 0.44 | 0.45 |
| Total statistics (FP3) | | | | | |
| Bias | 0.63 | 0.43 | 0.30 | 0.32 | 0.30 |
| RMSE | 0.86 | 0.66 | 0.56 | 0.55 | 0.55 |
| Nbias | 0.52 | 0.37 | 0.31 | 0.32 | 0.31 |
| SD | 0.56 | 0.49 | 0.42 | 0.40 | 0.42 |
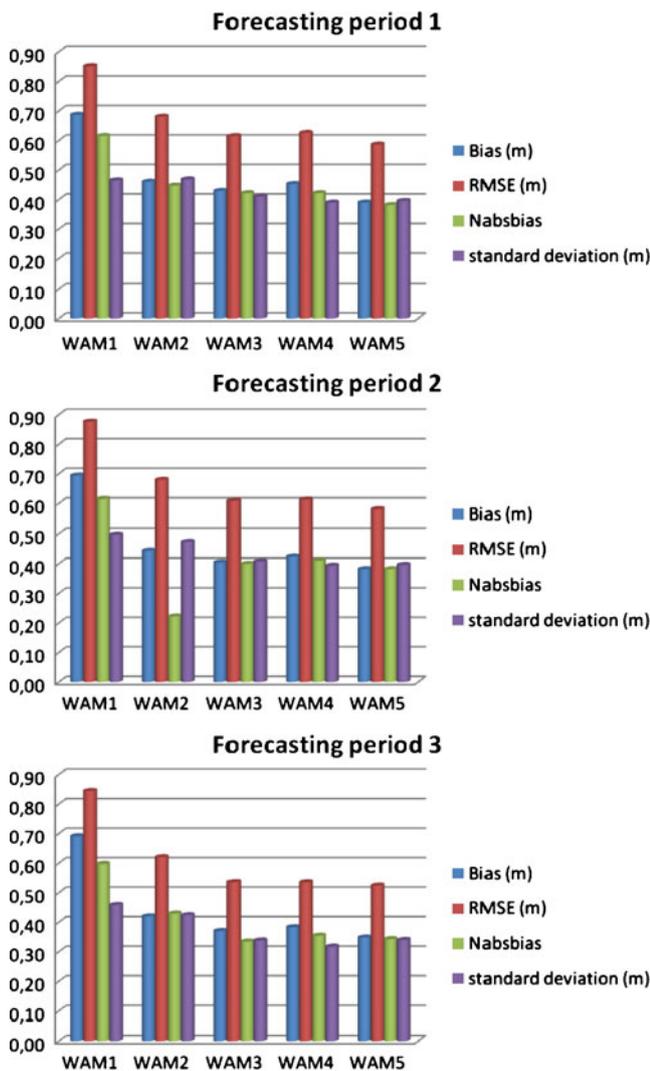
Fig. 8 Histograms for the buoy locations used by the DA

reduction of 15–20% at the standard deviation. The significantly improved results of WAM3-5 against the plain wave model WAM1 should be also underlined. The minimization of the divergence may be attributed to the use of the Kalman filters, when the reduction of the scatter is a result of the subsequent use of the DA into the forecasting period.

The above comparisons reveal the improvement of the forecasts at the location of the buoys. However, it is worth noticing that this gain is spread to a wider region by the use of the DA system that follows the Kalman filters within the forecasting period. This is illustrated in Tables 5 and 6, where the results of the different model simulations are evaluated against the independent buoys in the experimental area. The coordinates of each buoy are listed in Table 1. More precisely, all the statistical measures employed, including those referring to variability, have been improved (bias 46–48% against WAM1 and 28–30% against WAM2, RMSE 22–25% and up to 3% respectively, standard
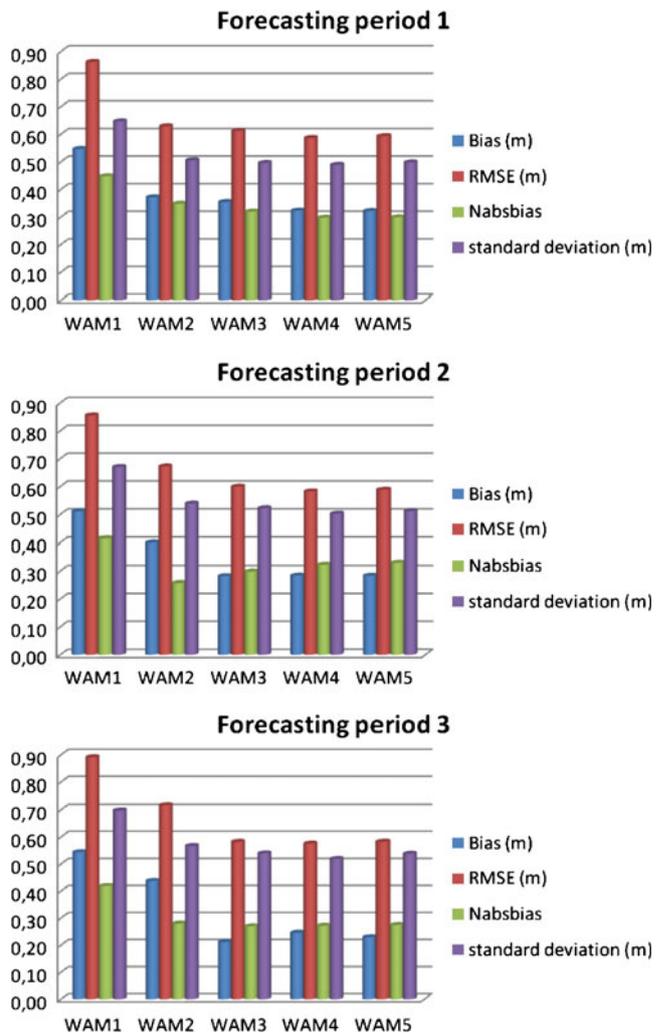


Fig. 9 Histograms for the location of the independent buoys

reduces further the bias, decreases the RMSE and the standard deviation and improves the Normalized Bias values. It is worth noticing that this reduction is most obvious in the WAM5 experiment, while WAM4 follows. These relatively better results of WAM5 can be attributed to the prevailing weather conditions in the experimental area during the testing period that were usual to the climatological pattern of this part of the Pacific Ocean (westerly–southwesterly winds and swell waves). All the above results are summarized in Tables 3 and 4, where the mean statistics from the comparison of the proposed methodology to the assimilated buoys are presented.

It is important to underline the reduction of the bias values from 5% to 14% against the classical DA scheme and around 40% against the WAM1 experiment. The same conclusion holds for the other variability measures: more than 10% improvement for RMSE against WAM2, slight reduction of the normalized bias and, finally, an important

deviation 20–25% and up to 7%, respectively). The statistics from the comparison against independent buoys are similar for the three proposed methodologies (WAM3-5) with the WAM4 experiment giving slightly better results.

The proposed methodology improved the correlation and the corresponding statistical figures of the linear relationship between WAM results and the relevant observations as illustrated in Fig. 7. As it is seen in this figure, WAM overestimates the SWH at lower-height waves and underestimates the higher ones. By the classical DA scheme, the results are improved at low SWH. On the contrary, the new methodology improves the forecasted SWH values in all cases. This is due to the Kalman filters that detect the bias and improve (in connection with the DA) the divergence.

In order to secure the validity of these results, more tests and comparison with data from satellite platforms were performed. More specifically, the Envisat-RA2 altimeter measurements were used to compare with model results over the greater area of the test case (as shown in Fig. 1). A considerable reduction of bias, RMSE and normalized bias and a slight improvement of the standard deviation is evident. These results are summarized in Table 7. This is expected since the KF reduce systematic errors and especially bias. The fact that the satellite records used for evaluation cover a wide area, exceeding the scale of the observational area employed by the DA scheme, reconfirms the extended impact of the proposed methodology. It is worth noticing that the above statistics are obtained by the comparison of the mean value of available satellite records at each grid box area against the corresponding WAM grid value.

It is also important to underline that the positive impact remains through the whole forecasting horizon and it is not limited in the first forecasting hours, as in the classical DA (Table 8). The statistics are becoming better with time for areas with observations used by Kalman filters through the new scheme as illustrated in Fig. 8. The experiment with the best results is WAM5. Similar results were succeeded in surrounding areas, where the three sensitivity tests (WAM3-5) gave comparable improvement (Fig. 9).

## 5 Conclusions

The results of wave forecasting systems are improved by Data Assimilation for only a limited time period. This is mainly due to the fact that biases from the atmospheric input or from the dynamics of the wave models lead to the reappearance of the initially emerged discrepancies. In this study, a new technique is proposed in order to reduce the consequences of this drawback. The new approach is the production of "forecasted observations" that is improved

model forecasts, obtained by the use of Kalman filters as part of the wave system. Afterwards, these "observations" are utilized by the DA scheme, inside the forecasting period. This technique leads to the reduction of the systematic error, spreading at the same time this positive impact over a greater area compared with the observations. On the other hand, the use of Kalman filters as a part of the wave model guarantees the compatibility of the relevant outputs with the physics of the simulated wave system.

The proposed methodology was applied to an open ocean area (southwest US coast) for a 3-month period. The Kalman filter algorithms were applied in three different ways: in the first scenario, a dynamical calculation of the covariance matrices was used in a continuous way, covering the whole simulation period. In the second scenario, the previous calculation was performed only during the DA window (where observations were available) and mean values were applied during the forecasting period. Finally, a climatological calculation, at each buoy location, was performed and the obtained values were applied in the whole testing period.

The results are promising, leading to the extension of the assimilation impact to the whole forecasting period. The best statistics were obtained by the climatological calculation of the covariance matrices. In general, an important reduction of the magnitude and variability of the discrepancies between final forecasts and observations was achieved.

However, it should be noted that there are some limitations in the application of this methodology, since it can be applied only in the presence of a continuous flow of observational data (e.g., buoys). At the same time, the buoy network is not very dense in the open ocean and it is mainly located near coastlines. Despite this, such type of data is available and important for areas of increased interest, like big harbors, touristic coasts, commercial areas, etc. In these cases, the proposed technique can contribute in the improvement of the final forecasts in a considerable way.

## References

Abdalla S, Bidlot J, Janssen P (2005) Assimilation of ERS and ENVISAT wave data at ECMWF. In: Proceedings of the 2004 ENVISAT & ERS Symposium, Salzburg, Austria, 6–10 Sept 2004 (ESA SP-572, Apr. 2005)

Bidlot J, Janssen P (2003) Unresolved bathymetry, neutral winds and new stress tables in WAM. ECMWF Research Department Memo R60.9/JB/0400

Breivik LA, Reistad M (1994) Assimilation of ERS-1 altimeter wave heights in an operational numerical wave model, weather and forecasting. 9(3)

Emmanouil G, Galanis G, Kallos G (2007) Assimilation of radar altimeter data in numerical wave models: an impact study in two different wave climate regions. Ann Geophys 25:1–15

Evensen G (2003) The Ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 53:343–367

Evensen G (2004) Sampling strategies and square root analysis schemes for the EnKF. Ocean Dyn 54:539–560. doi:10.1007/s10236-004-0099-2

Galanis G, Anadranistakis M (2002) A one dimensional Kalman filter for the correction of near surface temperature forecasts. Meteor Appl 9:437–441

Galanis G, Louka P, Katsafados P, Kallos G, Pytharoulis I (2006) Applications of Kalman filters based on non-linear functions to numerical weather predictions. Ann Geophys 24:2451–2460

Galanis G, Emmanouil G, Chu PC, Kallos G (2009) A new methodology for the extension of the impact in sea wave assimilation systems. Ocean Dyn. doi:10.1007/s10236-009-0191-8

Greenslade D (2001) The assimilation of ERS-2 significant wave height data in the Australian region. J Mar Sys 28:141–160

Greenslade DJM, Young IR (2004) Background errors in a global wave model determined from altimeter data. J Geophys Res 109 (C09007):1–24

Ide K, Courtier P, Ghil M, Lorenc AC (1997) Unified notation for data assimilation: operational, sequential and variational. J Meteor Soc 75:181–189

Jansen PAEM (2000) ECMWF wave modeling and satellite altimeter wave data. In: Halpern D (ed) Satellites, oceanography and society. Elsevier, New York, pp 35–36

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME, Ser D 82:35–45

Kalman RE, Bucy RS (1961) New results in linear filtering and prediction problems. Trans ASME, Ser D 83:95–108

Kalnay E (2002) Atmospheric modeling, data assimilation and predictability. Cambridge University Press, Cambridge, p 341

Komen GJ, Cavaleri L, Donelan M, Hasselmann K, Hasselmann S, Janssen PAEM (1994) Dynamics and modelling of ocean waves. Cambridge University Press, Cambridge

Lionello P, Günther H, Janssen PAEM (1992) Assimilation of altimeter data in a global third generation wave model. J Geophys Res 97 (C9):14453–14474

Lorenc AC (1981) A global three-dimensional multivariate statistical interpolation scheme. MonWeather Rev 109:701–721

Persson A (1990) Kalman filtering a new approach to adaptive statistical interpretation of numerical meteorological forecasts. ECMWF Newsletter

Siddons L (2007) Data assimilation of HF radar data into coastal wave models. Ph.D. Thesis of the Department of Applied Mathematics, School of Mathematics and Statistics, The University of Sheffield (UK). pp. 260

van der Grijn G (2009) Use of ECMWF products at MeteoGroup, ECMWF Forecast Products Users Meeting, 10–12 June 2009

Voorrips AC, Makin VK, Hasselmann S (1997) Assimilation of wave spectra from pitch-and-roll buoys in a North Sea wave model. J Geo Res 102:5829–5849

Voorrips AC, Heemink AW, Komen GJ (1999) Wave data assimilation with the Kalman filter. J Mar Syst 19(4):267–291

WAMDIG (The WAM-Development and Implementation Group), Group I, Hasselmann S, Hasselmann K, Bauer E, Bertotti L, Cardone CV, Ewing JA, Greenwood JA, Guillaume A, Janssen PAEM, Komen GJ, Lionello P, Reistad M, Zambresky L (1988) The WAM Model—a third generation ocean wave prediction model. J Phys Oceanogr 18(12):1775–1810