Ocean Dynamics (2009) 59:523–535
DOI 10.1007/s10236-009-0191-8

A new methodology for the extension of the impact of data assimilation on ocean wave prediction

George Galanis · George Emmanouil · Peter C. Chu · George Kallos

Received: 21 July 2008 / Accepted: 6 March 2009 / Published online: 1 April 2009

Abstract It is a common fact that the majority of today's wave assimilation platforms have a limited, in time, ability of affecting the final wave prediction, especially that of long-period forecasting systems. This is mainly due to the fact that after "closing" the assimilation window, i.e., the time that the available observations are assimilated into the wave model, the latter continues to run without any external information. Therefore, if a systematic divergence from the observations occurs, only a limited portion of the forecasting period will be improved. A way of dealing with this drawback is proposed in this study: A combination of two different statistical tools—Kolmogorov–Zurbenko and Kalman filters—is employed so as to eliminate any systematic error of (a first run of) the wave model results. Then, the obtained forecasts are used as artificial observations that can be assimilated to a follow-up model simulation inside the forecasting period. The method was successfully applied to an open sea area (Pacific Ocean) for significant wave height forecasts using the wave model WAM and six different buoys as observational stations. The results were encouraging and led to the extension of the assimilation impact to the entire forecasting period as well as to a significant reduction of the forecast bias.

Keywords Assimilation · Kalman filters · Kolmogorov–Zurbenko filters · Wave modeling

Responsible Editor: Jin-Song von Storch

G. Galanis (✉)
Section of Mathematics, Naval Academy of Greece,
Xatzikyriakion,
Piraeus 18539, Greece
e-mail: ggalanis@mg.uoa.gr

G. Galanis · G. Emmanouil · G. Kallos
School of Physics, Division of Applied Physics,
Atmospheric Modeling and Weather Forecasting Group,
University of Athens,
University Campus, Building PHYS-V,
Athens 15784, Greece

P. C. Chu
Department of Oceanography,
Graduate School of Engineering and Applied Science,
Naval Postgraduate School,
Monterey, CA 93943, USA

1 Introduction

The use of assimilation techniques for the correction of initial wind and wave conditions has significantly contributed to the improvement of the final outputs of numerical wave models over the last number of years. Several studies are devoted to the presentation, clarification, and operational use of such type of modules. We refer the reader to Janssen et al. (1987); Lionello et al. (1992); Lionello et al. (1995); Greenslade and Young (2005); Abdalla et al. (2005); and Skandrani et al. (2004).

Despite the substantial progress achieved in this field, a serious disadvantage of the assimilation systems remains, namely, that their impact is, in most cases, limited in time. This is mainly due to the fact that the wave models used today are heavily dependent on the corresponding predicted wind input. As a result, any biases coming from wind prediction will force the subsequent wave forecasted fields to diverge when no external information is available. Moreover, the limited number of available and quality-controlled wave observations contributes further to the above-mentioned problem. In this way, only a short part of the forecasting results is improved.

A way out of this seemingly unavoidable problem is proposed in this paper using assimilation techniques in

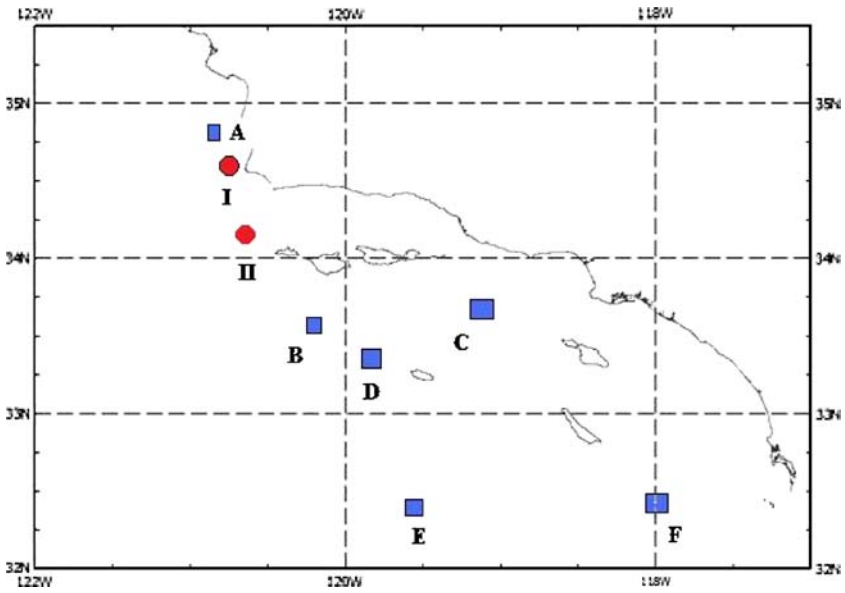

Fig. 1 The study area. The buoys used for data assimilation are denoted by *rectangles* while those with *circles* were used solely for evaluation

numerical wave modeling. Based on a combination of Kolmogorov–Zurbenko (KZ) and Kalman filters, we improve the direct outputs of an initial model run based on available observations. More precisely, KZ filters are employed in order to reduce the different qualitative characteristics between the two time series used, i.e., observations and the model forecast. The first are raw measurements usually highly variable in contrast to model forecasts which, being smoothed by the model itself in space and time, have a milder evolution. The KZ-filtered series can be used as input to Kalman filters in order to eliminate possible systematic errors. In this way, we obtain accurate local predictions without any systematic errors. These improved forecasts are then assimilated in a second model run, having the role of "artificial" observations inside the forecasting period. In this way, the assimilation impact spreads to the whole forecasting data set.

The proposed technique has been tested on the wave model WAM for a 3-month period (October–December 2006) over the Pacific Ocean near the southwest coast of the United States. The necessary observations were obtained from National Oceanic and Atmospheric Administration (NOAA) buoys and the corresponding results were promising.

Table 1 Buoy coordinates

Buoy label	Latitude	Longitude	Depth (m)
A	34.88 N	120.87 W	204
B	33.65 N	120.2 W	1,004.6
C	33.75 N	119.08 W	881.5
D	33.22 N	119.88 W	335
E	32.43 N	119.53 W	1,393.5
F	32.5 N	118 W	1,856.2
Buoy I	34.71 N	120.97 W	384.1
Buoy II	34.27 N	120.70 W	632

2 Models and post processes description

2.1 The wave model

The numerical wave model used in this paper is WAM (cycle 4). More precisely, the version of the European Centre for Medium-Range Forecasts (ECMWF) was employed, subsequently modified by the Atmospheric Modeling and Weather Forecasting Group of the University of Athens to run in parallel on Linux clusters. This is a third-generation wave model which solves the wave transport equation explicitly without any assumptions regarding the shape of the wave spectrum. Detailed descriptions may be found in WAMDIG (1988) and Komen et al. (1994), while an explicit presentation of the WAM parameters used in the present study is presented in Section 3.

2.2 Assimilation method

The analysis fields used were corrected by an assimilation method developed at the Norwegian Meteorological Institute (Breivik and Reistad 1994). The analysis scheme is based on Bratseth (1986) that converges towards the results of the classical statistical interpolation method with a proper choice of parameters for the analysis weights.

More precisely, the analysis begins with the direct numerical model output for significant wave height (SWH^P) as a first guess field. This is corrected by the use

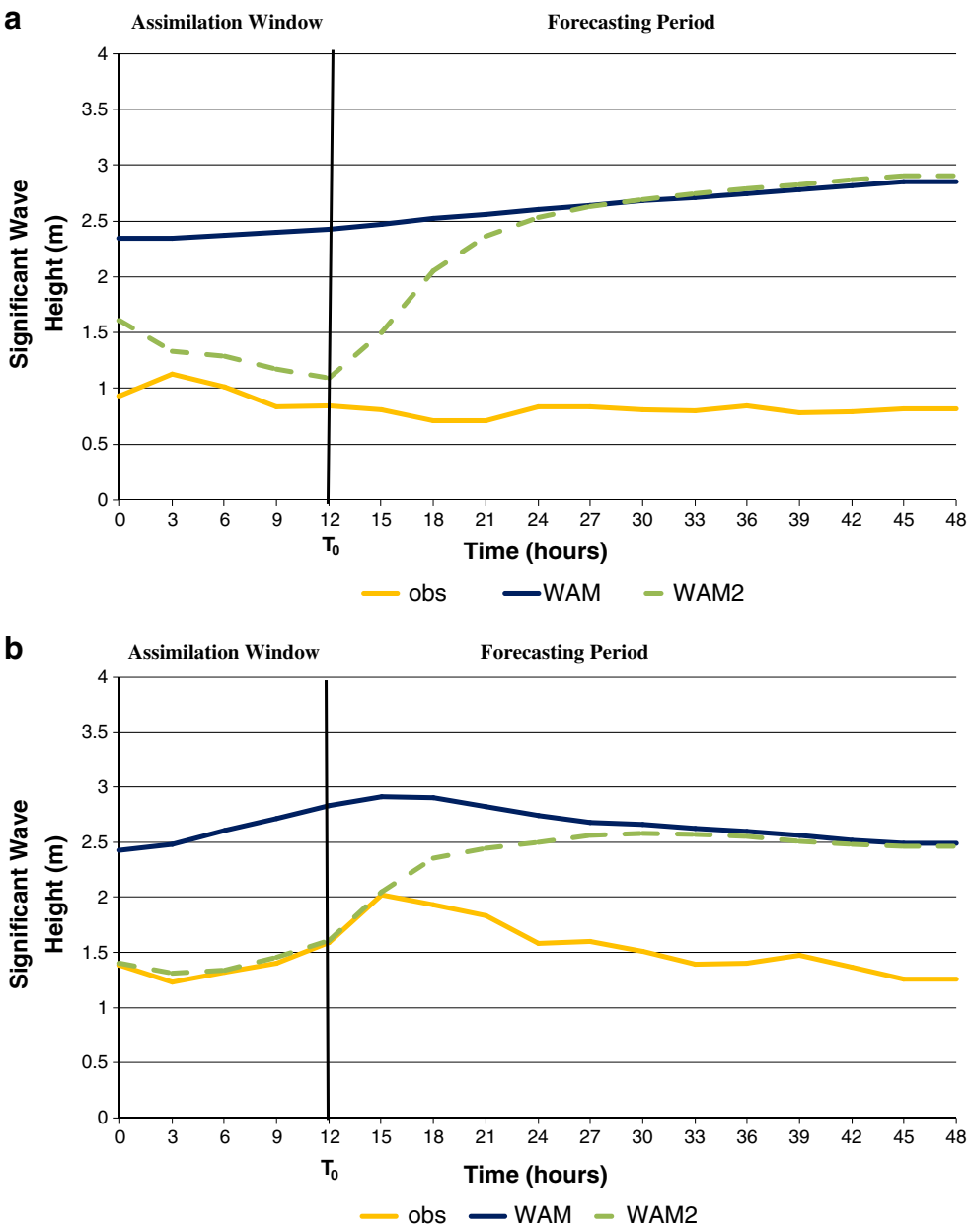

Fig. 2 **a** The results of one cycle of WAM2 in the area of buoy C. **b** The results of one cycle of WAM2 in the area of buoy F

of available corresponding observations (SWH^O) based on the following iterative equations:

$$SWH_i^A(k+1) = SWH_i^A(k) + \sum_{j=1}^{N} a_{ij}\left(SWH_j^O - SWH_j^A(k)\right), \quad (1)$$

$$SWH_x^A(k+1) = SWH_x^A(k) + \sum_{j=1}^{N} a_{xj}\left(SWH_j^O - SWH_j^A(k)\right). \quad (2)$$

Here, the analysis weights are given by:

$$a_{ij} = \left(m_{ij} + d_{ij}\right)/M_j,\ a_{xj} = m_{xj}/M_j \quad (3)$$

where the subscripts i and j refer to the observation points, x to the grid points, and the superscripts O, P, and A to the observed, first guess, and analyzed values. Moreover, N is the number of observations, k is an iteration counter, m_{ij} denotes the model's error covariances, and d_{ij} denotes the observation's error covariances, which are assumed to be uncorrelated to each other. Finally, M_j is a function of m_{ij} and d_{ij} calculated for each observation separately in a way that ensures the convergence of the above system of equations. More precisely, in this way, the iteration limit

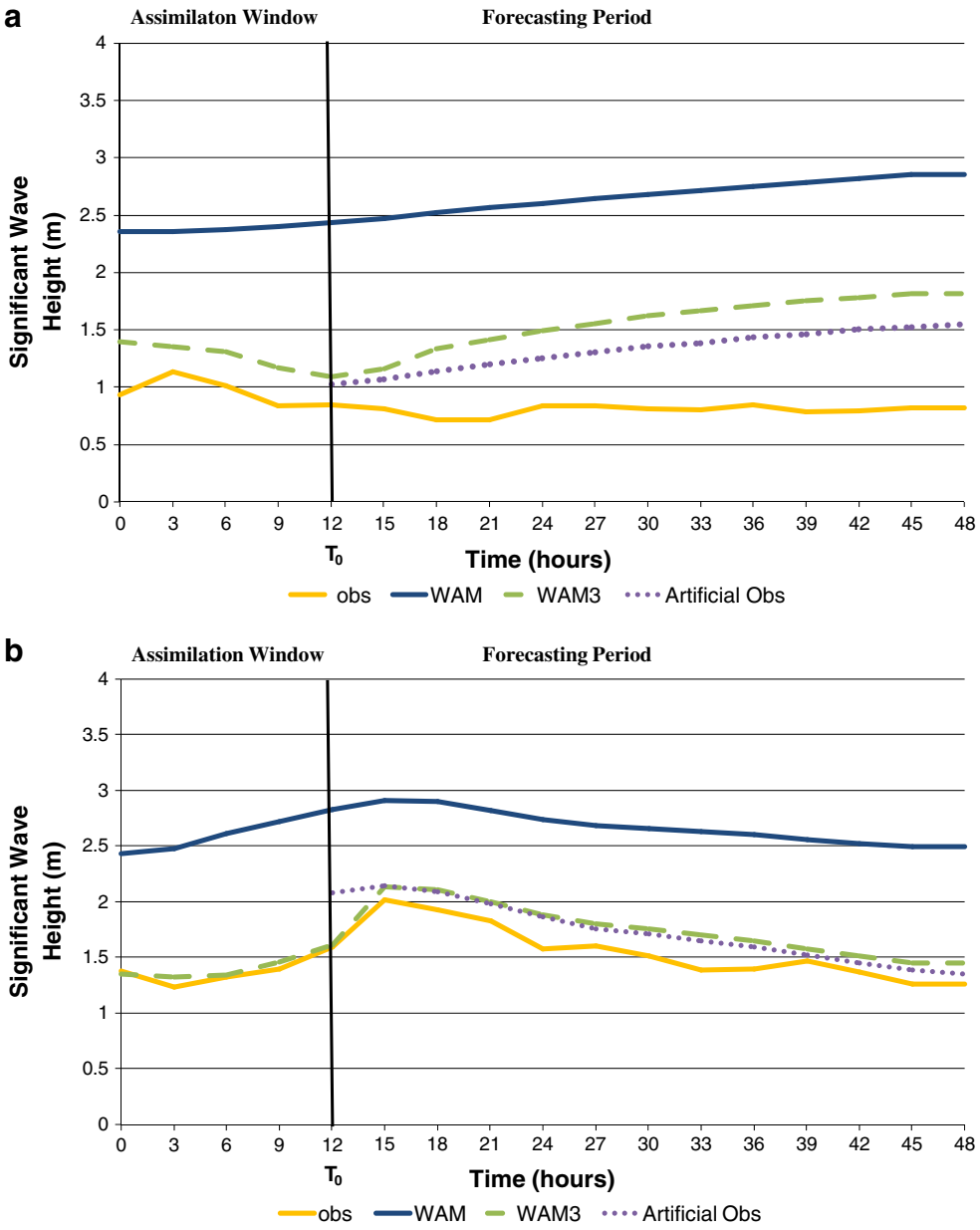

Fig. 3 **a** Results of WAM3 for the same period as in Fig. 2a (buoy C). Apart from the real observations assimilated within the assimilation window, the improved, via KZ and Kalman filters, forecasts of WAM are also assimilated inside the forecasting period as artificial observations. **b** The same results for the area of buoy F

becomes equal to the solution obtained by optimal interpolation (see Bratseth 1986 and Breivik and Reistad 1994).

The iterations begin with the initial values $\mathrm{SWH}_x^{\mathrm{A}}(1) = \mathrm{SWH}_x^{\mathrm{P}}$ and $\mathrm{SWH}_i^{\mathrm{A}}(1) = \mathrm{SWH}_x^{\mathrm{P}}$ and the coefficients m_{ij} are calculated by:

$$m_{ij} = E_i^{\mathrm{P}} a^{\mathrm{P}}(r_{ij}) E_j^{\mathrm{P}} \tag{4}$$

where E_i^{P} is the standard deviation of the model first guess. The parameter a^{P} is given by $a^{\mathrm{P}}(r_{ij}) = \exp\left(-0.5 \times r_{ij}^2 \Big/ b^2\right)$ where r_{ij} is the distance between the observation points and b is the horizontal resolution of the first guess field error correlation determining the scale of the observation's influence. The observational error covariances have the form:

$$d_{ij} = E_i^{\mathrm{O}} a^{\mathrm{O}}(r_{ij}) E_j^{\mathrm{O}} \tag{5}$$

where E_i^{O} is the standard deviation of the observations. In most of the cases, the matrix $a^{\mathrm{O}}(r_{ij})$ is assumed to be unitary, based on the assumption that the observations are unbiased and uncorrelated.

The above-described method leads to a corrected SWH field. Moreover, the total wave model spectrum is updated through a division into a swell and a wind sea part, which are scaled according to the analyzed SWH. Additionally, in

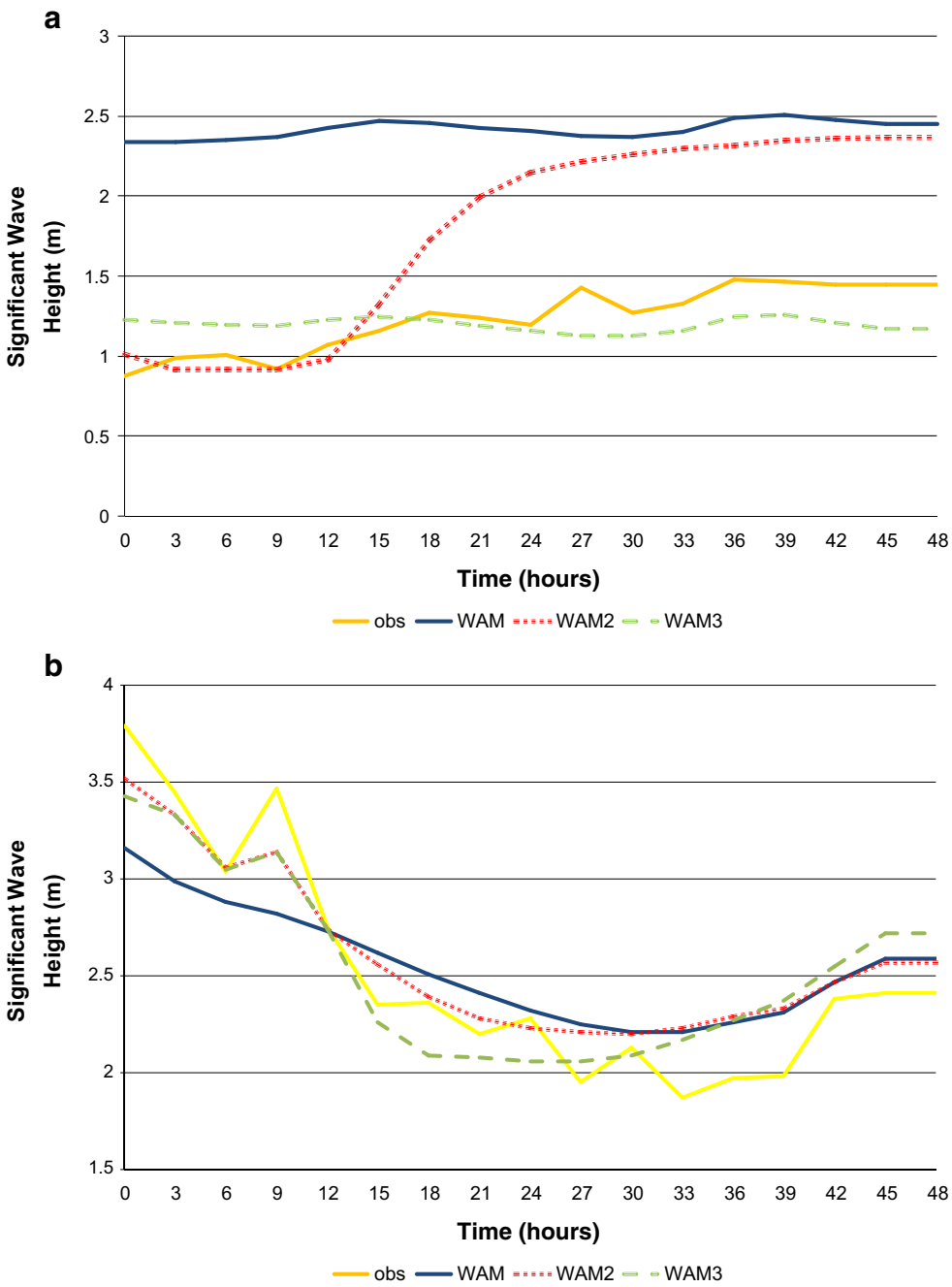

Fig. 4 **a** Adjustment of the Kalman-filtered forecasts (WAM3) to the observations in a 48-h wave model cycle (2–3 November 2006) with constant systematic model error (a selected period in the area of buoy F). **b** Dynamical adjustment of the Kalman-filtered forecasts (WAM3) to the observations in a 48-h wave model cycle (10–11 November 2006) with variable model error (a selected period in the area of buoy E)

order to avoid possible imbalances, the wind at the analysis time is updated accordingly. A more detailed description of the assimilation method is given in Breivik and Reistad (1994).

2.3 Kolmogorov–Zurbenko filters

Focusing now on the statistical methodologies employed, a short description of KZ filters is given (for details, see Eskridge et al. 1997; Rao et al. 1997). They consist of a series of iterative moving averages aiming at the removal of high-frequency variations from the initial data and, therefore, modifying the observations and the corresponding model forecasts into a compatible form.

More precisely, if $\left(x_i^0\right)_i$ are the initial values of a given series, then the first iteration of the filter results to:

$$x_i^1 = \frac{1}{2q+1}\sum_{j=-q}^{q} x_{i+j}^0 \tag{6}$$

where MA=$2q+1$ is the filter (moving average) window. In the second step of the method, these new values

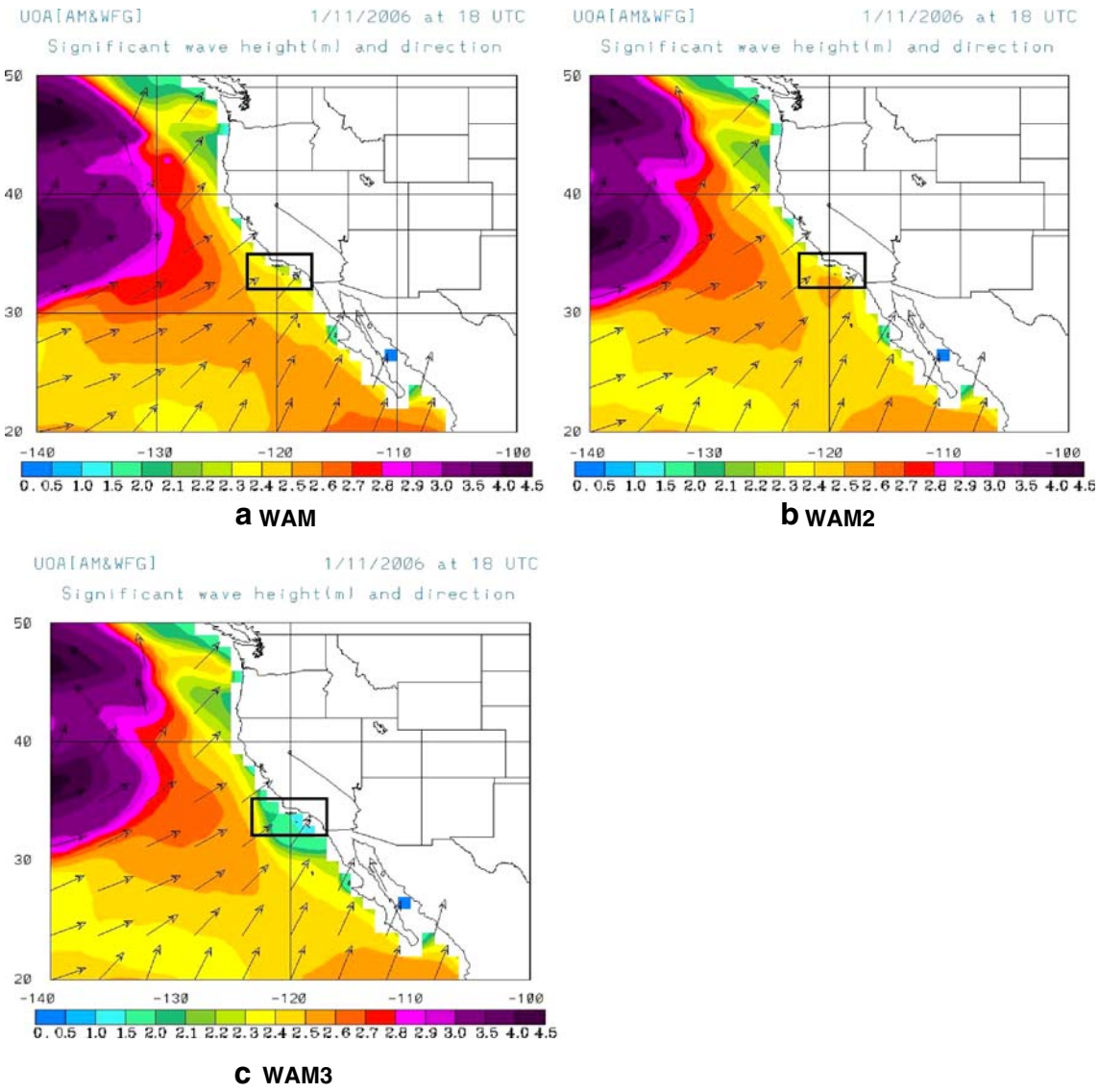

Fig. 5 Spread of the artificial information. The *internal bold rectangle* denotes the buoy area as shown in Fig. 1

$\left(x_i^1\right)_i$ become the input for the second iteration $\left(x_i^2 = \frac{1}{2q+1}\sum_{j=-q}^{q} x_{i+j}^1\right)$ and so on. The total number N_1 of iterations depends on the portion of the variability that one wants to exclude. To be more precise, if we want to subtract any variability emerging in P_1 time steps, then the following condition should be satisfied:

$$N_1 \leq (P_1/(2q+1))^2. \tag{7}$$

2.4 Kalman filters

The time series obtained after the use of the KZ filter, which disposes of high frequencies, has the same qualitative characteristics and can be used as input for a Kalman filter procedure. A detailed description of the general form of a Kalman filter using the unified notation proposed by Ide et al. (1997) is presented here. These filters simulate the evolution in time of an unknown process (state vector) whose "true" value at time t_i is denoted here by $x^t(t_i)$. The latter is combined with corresponding observations y_i^O. The change of x in time is given by the (system) equation:

$$x^t(t_i) = L_{i-1}[x^t(t_{i-1})] + \eta(t_{i-1}), \tag{8}$$

while the relation between the observation and the unknown vectors are described by the (observation) equation:

$$y_i^O = H_i[x^t(t_i)] + \varepsilon_i. \tag{9}$$

The system operator L_i, the observational one H_i as well as the covariance matrices $\mathbf{Q}(t_i)$ and $\mathbf{R}(t_i)$ of $\boldsymbol{\eta}(t_i)$ and $\boldsymbol{\varepsilon}_i$,

Table 2 Bias, RMSE, standard deviation, normalized bias, and scatter index of the three different runs for all buoy locations

	Buoy A (*N*=487)			Buoy B (*N*=531)			Buoy C (*N*=709)			Buoy D (*N*=511)			Buoy E (*N*=531)			Buoy F (*N*=531)		
	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3
Bias	0.51	0.21	0.02	0.39	0.08	−0.19	0.99	0.72	0.33	0.68	0.39	0.10	0.62	0.27	−0.02	0.88	0.44	0.02
RMSE	0.73	0.57	0.48	0.64	0.49	0.47	1.10	0.89	0.55	0.80	0.60	0.38	0.79	0.56	0.42	0.96	0.65	0.27
Normalized bias	0.42	0.28	0.21	0.33	0.22	0.17	1.05	0.73	0.40	0.44	0.27	0.16	0.43	0.24	0.15	0.74	0.38	0.15
St_Dev	0.53	0.53	0.47	0.51	0.48	0.43	0.49	0.53	0.44	0.41	0.45	0.37	0.48	0.49	0.42	0.37	0.48	0.27
Scatter index	0.29	0.29	0.26	0.26	0.25	0.22	0.41	0.44	0.37	0.21	0.23	0.19	0.25	0.25	0.22	0.27	0.36	0.20

The number (*N*) of observations used is declared in parentheses

respectively, have to be determined before the application of the filter.

An initial forecast (first guess) of the state vector $\boldsymbol{x}$ and its error covariance matrix **P** are given by:

$$x^{\mathrm{f}}(t_i) = L_{i-1}[x^{\mathrm{a}}(t_{i-1})], \tag{10}$$

$$P^{\mathrm{f}}(t_i) = L_{i-1}P^{\mathrm{a}}(t_{i-1})L_{i-1}^{T} + Q(t_{i-1}). \tag{11}$$

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

$$x^{\mathrm{a}}(t_i) = x^{\mathrm{f}}(t_i) + K_i\left(y_i^{\mathrm{O}} - H_i\left[x^{\mathrm{f}}(t_i)\right]\right), \tag{12}$$

$$P^{\mathrm{a}}(t_i) = (I - K_iH_i)P^{\mathrm{f}}(t_i). \tag{13}$$

The matrix:

$$K_i = P^{\mathrm{f}}(t_i)H_i^{T}\left[H_iP^{\mathrm{f}}(t_i)H_i^{T} + R_i\right]^{-1} \tag{14}$$

is referred to as the Kalman gain and describes how easily the filter adjusts to possible new conditions. The superscripts o, t, f, and a correspondingly denote observations, true, forecast, and analysis values. Moreover, T and -1 are the classical symbols of the transpose and the inverse matrix, respectively, while I stands for the unitary matrix. Equations 8, 9, 10, 11, 12, 13, and 14 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).

3 The case studied

The present study was based on a global version of the WAM model integrated on 30 frequencies, 24 directions and horizontal resolution of 1.0×1.0°. The first integration frequency was set to 0.0417 Hz, while the propagation time step was defined to 300 s. WAM, running on a deep water mode with no refraction, was driven by six-hourly wind input (10 m winds speed and direction) obtained from the NCEP/GFS global model with a horizontal grid resolution 1.0×1.0°.

Concerning the parameters used in the data assimilation scheme, based on the experience gained by previous applications of the assimilation system (Breivik and Reistad 1994), the number of iterations was set to k=5 and the decorrelation radius b (grid distances) to 4. The maximum influence radius of assimilation was eight grid points (this is only a practical limit set when the correlation following this expression is close to zero). The standard deviation of the model was E^{P}=1.0 m and the observation error was E^{O}=0.5 m.

Table 3 Average statistics and percentage of improvement versus direct model outputs (WAM) and the classical assimilation scheme (WAM2)

	Average values over buoys A–F			Improvement (%) of WAM3 against	
	WAM	WAM2	WAM3	WAM	WAM2
Bias	0.68	0.35	0.04	94	88
RMSE	0.83	0.62	0.43	49	31
Normalized bias	0.57	0.35	0.21	64	41
St_Dev	0.47	0.49	0.40	14	19
Scatter index	0.28	0.30	0.24	14	20

The KZ filter used was a (5,5)-one, i.e., parameters MA, P_1 (see Section 2.3) were both defined equal to 5. This filter was applied to both observations and wave forecasts aiming at the homogenization of their basic characteristics.

Concerning the Kalman filter, the following details are necessary to be outlined: The filter was applied to a single forecasted parameter: the SWH. The corresponding bias y_i^{O} is estimated as a function of the forecasting model direct output SWH_i:

$$y_i^{\mathrm{O}} = a_{0,i} + a_{1,i} \times \mathrm{SWH}_i + \varepsilon_i \tag{15}$$

where the coefficients $\{a_{0,i},\ a_{1,i}\}$ have to be estimated by the filter. Note that the index i stands for observational time step (3 h in the case studied), i.e., the filter is activated only when a corresponding observation is available. In this way, $a_{0,i} + a_{1,i} \times \mathrm{SWH}_i$ express the systematic part of the bias. Parameter ε_i stands for the Gaussian (nonsystematic) error of the previous procedure. Therefore, $\mathbf{x}(t_i) = \begin{bmatrix} a_{0,i} & a_{1,i} \end{bmatrix}^T$ is the state vector of the filter, the bias y_i^{O} is used as a known parameter, the observation matrix takes the form $H_i = [1 \quad SWH_i]$, and the system matrix is the unitary 1 of dimension 2. As a result, the system and observation equations take the following form:

$$x^{\mathrm{t}}(t_{i+1}) = x^{\mathrm{t}}(t_i) + \eta(t_i), y_i^{\mathrm{O}} = H_i[x^{\mathrm{t}}(t_i)] + \varepsilon_i, \tag{16}$$

while the initial values are:

$$x = 0 \ \ y_0^{\mathrm{O}} = \varepsilon_0, \ \ P(t_0) = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}, \ \ Q(t_0) = I_2, \ \ R(t_0) = 6.$$

In this way, we assume that there are no correlations between different coordinates of the state vector $\boldsymbol{x}$. The relevantly large values for $R(t_0)$ and the diagonal elements of $P(t_0)$ declare low credibility of the first guess and fast adjustment to new conditions.

The variance matrices, $\mathbf{Q}(t_i)$ and $\mathbf{R}(t_i)$, of the system and observation equation, respectively, are calculated from the sample of the last seven values of $\boldsymbol{\eta}(t_i)$ and $\boldsymbol{\varepsilon}_i$, an optimal period resulting from a series of relevant tests (see Galanis et al. 2006).

This model for the description of the systematic error has been also adopted in previous applications of the Kalman filter on different atmospheric parameters with success (see Galanis and Anadranistakis 2002; Galanis et al. 2006). In this work, we extend this approach for sea wave applications.

The study area was located over the Pacific Ocean near the southwest coast of the United States (Fig. 1). The position of buoys (NOAA/National Data Buoy Center network) used as observational sources and for evaluation purposes are indicated on the same map. Their exact coordinates and the water depth at each location are declared in Table 1. It worth noticing that, in all cases, the water depth is in accordance with the deep water mode chosen for the wave model.

For this area, three different forecast experiments were carried out. All experiments performed a forecast of 36 h each second day for a 3-month period (October–December 2006).

The first model run (hereafter referred to as WAM) did not use any assimilation system. The second model simulation (WAM2) used the assimilation system described in Section 2. In this run, the model assimilates buoy observations available in 3-h intervals until the beginning of the 36-h forecasting period (analysis time T_0 in Fig. 2).

Table 4 Evaluation against independent buoys

	Buoy I (N=271)			Buoy II (N=655)		
	WAM	WAM2	WAM3	WAM	WAM2	WAM3
Bias	−0.40	−0.18	0.07	0.57	0.52	0.42
RMSE	1.16	0.72	0.65	0.94	0.74	0.72
Normalized bias	0.34	0.21	0.22	0.38	0.31	0.31
St_Dev	1.08	0.70	0.64	0.75	0.52	0.59
Scatter index	0.38	0.24	0.23	0.34	0.24	0.27

Table 5 Evaluation against ENVISAT RA-2 data

Satellite verification	(*N*=231)		
	WAM	WAM2	WAM3
Bias	0.50	0.37	0.23
RMSE	0.74	0.70	0.63
Normalized bias	0.39	0.34	0.28
St_Dev	0.55	0.59	0.58
Scatter index	0.26	0.28	0.28

The model results (analysis) are significantly improved during the assimilation window, approaching the corresponding observation values (as shown in Fig. 2a, b). However, this positive impact decreases thereafter and almost disappears after a period of 12 h. This is due to the fact that there are no available observations to assimilate during the forecasting period. Note that here, as well as in the subsequent figures, the model output time step is set to 3 h.

Finally, a third version of WAM (WAM3) was run that assimilated two different observation types:

1. The "real" observations used by WAM2 in the assimilation window,
2. Improved-filtered forecasts of WAM, the direct outputs of which were adjusted by a combination of KZ and Kalman filters, resulting to "artificial" observations (every 3 h) inside the 36-h forecasting period (Fig. 3a, b). In these "observations," any systematic error has been removed. This methodology leads to the extension of the assimilation impact to the whole forecasting period.

The examples presented above are representative 2-day simulation results from the study period.

It is important to underline that the whole study was based on a real-time run: The observations of the "current" day (d) are used only as input to the filter and to assimilation modules so as to produce an improved forecast for the next (d+1) day. These new forecasts are evaluated against the observations of the next (d+1) day, ensuring in this way that there is no conflict between assimilation and evaluation data.

The results obtained from the above simulations (WAM, WAM2, and WAM3) were validated for their forecast accuracy as well as for the corresponding assimilation impact. The statistical analysis was based on the following parameters:

Bias of forecasted values:

$$\text{Bias} = \frac{1}{k} \times \sum_{i=1}^{k} (\text{for}(i) - \text{obs}(i)) \tag{17}$$

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

Table 6 Bias, RMSE, standard deviation, normalized bias, and scatter index of the three different runs for all buoy locations concerning the first 12 h of forecast

Forecasting period I (0–12h forecast)																	
	Buoy A			Buoy B			Buoy C			Buoy D			Buoy E			Buoy F		
	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3
Bias	0.54	0.23	0.04	0.37	0.13	−0.20	1.01	0.74	0.37	0.68	0.43	0.09	0.63	0.31	0.00	0.86	0.38	0.01
RMSE	0.75	0.60	0.50	0.59	0.47	0.49	1.11	0.87	0.60	0.79	0.59	0.37	0.79	0.57	0.45	0.94	0.51	0.32
Normalized bias	0.45	0.32	0.27	0.31	0.21	0.17	1.07	0.73	0.44	0.44	0.28	0.17	0.43	0.26	0.18	0.72	0.32	0.16
St_Dev	0.52	0.55	0.50	0.46	0.45	0.44	0.46	0.46	0.47	0.41	0.40	0.36	0.48	0.48	0.45	0.36	0.35	0.32
Scatter index	0.29	0.31	0.28	0.24	0.23	0.23	0.39	0.38	0.40	0.21	0.20	0.18	0.24	0.25	0.23	0.27	0.26	0.24

Table 7 Bias, RMSE, standard deviation, normalized bias, and scatter index of the three different runs for all buoy locations concerning the second 12 h of forecast

	Forecasting period II (12–24h forecast)																	
	Buoy A			Buoy B			Buoy C			Buoy D			Buoy E			Buoy F		
	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3
Bias	0.48	0.28	−0.05	0.41	0.27	−0.19	0.96	0.92	0.30	0.74	0.62	0.15	0.62	0.43	−0.03	0.88	0.69	0.01
RMSE	0.72	0.70	0.60	0.66	0.64	0.56	1.06	1.07	0.56	0.81	0.79	0.44	0.78	0.71	0.52	0.94	0.84	0.29
Normalized bias	0.42	0.39	0.28	0.36	0.32	0.20	0.99	0.97	0.42	0.47	0.40	0.21	0.44	0.36	0.21	0.73	0.58	0.18
St_Dev	0.54	0.66	0.60	0.52	0.58	0.53	0.45	0.55	0.47	0.34	0.48	0.42	0.48	0.57	0.52	0.33	0.49	0.29
Scatter index	0.30	0.36	0.33	0.26	0.30	0.27	0.38	0.46	0.39	0.17	0.24	0.21	0.24	0.29	0.27	0.25	0.37	0.22

Table 8 Bias, RMSE, standard deviation, normalized bias, and scatter index of the three different runs for all buoy locations concerning the third 12 h of forecast

	Forecasting period III (24–36h forecast)																	
	Buoy A			Buoy B			Buoy C			Buoy D			Buoy E			Buoy F		
	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3	WAM	WAM2	WAM3
Bias	0.54	0.32	0.02	0.35	0.14	−0.24	0.98	0.93	0.32	0.69	0.47	0.04	0.61	0.43	−0.04	0.89	0.72	0.03
RMSE	0.74	0.71	0.57	0.62	0.60	0.59	1.09	1.08	0.61	0.80	0.70	0.47	0.77	0.72	0.53	0.95	0.87	0.33
Normalized bias	0.44	0.39	0.27	0.33	0.30	0.23	1.05	0.97	0.44	0.44	0.35	0.21	0.42	0.36	0.21	0.75	0.62	0.22
St_Dev	0.51	0.64	0.58	0.51	0.59	0.53	0.47	0.54	0.51	0.41	0.52	0.46	0.46	0.57	0.53	0.33	0.50	0.33
Scatter index	0.28	0.36	0.32	0.26	0.30	0.27	0.40	0.45	0.43	0.21	0.26	0.24	0.24	0.29	0.27	0.24	0.38	0.25

Table 9 Average statistics and percentage of improvement versus direct model outputs (WAM) and the classical assimilation scheme (WAM2) for the first 12 h of forecast

	Forecasting period I (0–12h) average values over buoys A–F			Improvement (%) of WAM3 against	
	WAM	WAM2	WAM3	WAM	WAM2
Bias	0.68	0.37	0.05	92	86
RMSE	0.83	0.60	0.46	45	24
Normalized bias	0.57	0.35	0.23	59	34
St_Dev	0.45	0.45	0.42	6	6
Scatter index	0.28	0.30	0.24	14	20

Normalized bias:

$$\text{Normalized bias} = \frac{1}{k} \times \sum_{i=1}^{k} \left| \frac{\text{for}(i) - \text{obs}(i)}{\text{obs}(i)} \right| \tag{18}$$

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

Root mean square error (RMSE) and standard deviation of the error (two classical divergence measures):

$$\begin{aligned} \text{RMSE} &= \sqrt{\tfrac{1}{k} \times \sum_{i=1}^{k} (\text{for}(i) - \text{obs}(i))^2}, \\ \text{St\_Dev} &= \sqrt{\tfrac{1}{k} \times \sum_{i=1}^{k} ((\text{for}(i) - \text{obs}(i)) - \text{Bias})^2} \end{aligned} \tag{19}$$

Scatter index:

$$\text{Scatter index} = {}^{\text{standard deviation of error}}/_{\text{mean observed wave height}} \tag{20}$$

a variability measure (see, e.g., Janssen et al. 1987).

4 Results

The main issue studied in this work is the time-limited impact of assimilation techniques, which is clearly illustrated in Fig. 2a, b. The assimilation method employed in WAM2 had excellent results which, however, affected only a period of 10–12 h after the analysis time. Throughout the rest of the forecasting period, the assimilation module had no impact and the model forecasts returned to their initial behavior, which, in most of the cases, introduces systematic errors.

The proposed methodology of employing artificial observations inside the forecasting period offers a possible solution to this problem. These observations are obtained from a previous model run followed up by a post process that eliminates any existing systematic error (see Fig. 3). Of course, the artificial assimilation data are not as accurate as the true observations. However, the proper use of the KZ and Kalman filters, as explained in Sections 2 and 3, leads to the elimination of any possible systematic error and increases the accuracy of these forecasts. Therefore, data assimilation of these artificial "observations" 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 natural question that may rise at this point could be if one could just bias-correct the model forecast at buoy locations instead of using such a sophisticated statistical procedure. Could such a simple correction give similar results? The answer is negative, for two main reasons: first of all, the bias between direct model forecasts and buoy measurements is neither known nor constant. It may change in time due to alterations in weather/wave conditions. So, it

Table 10 Average statistics and percentage of improvement versus direct model outputs (WAM) and the classical assimilation scheme (WAM2) for the second 12 h of forecast

	Forecasting period II (12–24h) average values over buoys A–F			Improvement (%) of WAM3 against	
	WAM	WAM2	WAM3	WAM	WAM2
Bias	0.68	0.54	0.03	95	94
RMSE	0.83	0.79	0.50	40	37
Normalized bias	0.57	0.50	0.25	56	50
St_Dev	0.44	0.56	0.47	6	15
Scatter index	0.28	0.30	0.24	14	20

Table 11 Average statistics and percentage of improvement versus direct model outputs (WAM) and the classical assimilation scheme (WAM2) for the third 12 h of forecast

	Forecasting period III (24–36h) average values over buoys A–F			Improvement (%) of WAM3 against	
	WAM	WAM2	WAM3	WAM	WAM2
Bias	0.68	0.50	0.02	97	96
RMSE	0.83	0.78	0.52	38	34
Normalized bias	0.57	0.50	0.26	54	47
St_Dev	0.45	0.56	0.49	9	13
Scatter index	0.28	0.30	0.24	14	20

would be really risky if one tried just to transform the model forecast using a constant. On the contrary, a Kalman filter can dynamically estimate possible discrepancies between the two time series and correct, respectively, the model forecasts even in cases where this deviation changes form. Two characteristic examples are given in Fig. 4a, b where the filter improves the final forecast analogously either in a simple systematic divergence (Fig. 4a) or in a more complex case with variable divergence (Fig. 4b). In such cases, it is obvious that a simple bias subtraction would lead to increased deviation between the forecasts and the corresponding observations. Moreover, a simple data assimilation scheme (WAM2) results to the improvement of the forecasts for only a few hours, as already mentioned earlier.

On the other hand, the subsequent use of artificial observations by the data assimilation scheme ensures that the aforementioned correction would spread over a larger area than the point location of buoys. This is clearly demonstrated in Fig. 5 where the impact of the artificial observations, after 30 h of forecast, has spread over a large area significantly exceeding the buoy domain. On the contrary, by the single use of data assimilation scheme (WAM2—Fig. 5b) the forecasting results seem to converge again to those of WAM (Fig. 5a), both as absolute values as well as the spatial distribution of the wave pattern.

A direct result of this extended effect of the assimilation was the radical reduction of the bias in the final forecasted values for the area of interest. In Table 2, statistical results are presented concerning different measures of divergence of the forecasts from the corresponding observed values for the entire 3-month study period.

In all cases, the use of the classical assimilation technique (WAM2) leads to an improvement of the final forecasts but not to the total elimination of the existing bias. The application of the proposed methodology (WAM3) removed the bias almost completely and reduced RMSE, normalized bias, standard deviation, and scatter index. It is worth noticing that this reduction is not as obvious in the case of buoy B due to the fact that at this specific location the major part of the error has been removed by the use of the assimilation (WAM2). However, even in this case, the other four statistical quantities are further reduced in the case of WAM3. Note also that the standard deviation remains almost unchanged if one uses a classical assimilation scheme (WAM2) while the use of artificial observations (WAM3) leads to a 20% reduction on average.

These statistical results are summarized in Table 3 where the percentage of the improvement of the proposed methodology against direct model outputs (WAM) and the use of a classical assimilation method (WAM2) is presented.

It is important to underline the negligible bias values, which represents an 88% improvement on the classical assimilation scheme, as well as the decrease of the variability indices. The 31% improvement of the RMSE in contrast to WAM2 is due entirely to the application of the proposed technique and to the utilization of artificial assimilated data inside the forecasting period. There is no way of achieving this result through the use of an assimilation system only.

The above results demonstrate the improvement obtained by the proposed methodology at the location of the buoys. Furthermore, it is important to emphasize that this positive impact is not restricted only in the area around the buoys, but is spread to a wider region through the use of the assimilation system within the forecasting period. This is illustrated in Tables 4 and 5 where the results of the different model simulations are evaluated against independent buoys (buoys I and II), the location of which is declared in Table 1, as well as against satellite data (Envisat-RA2 altimeter) from the greater area of the test case (longitude 116–122° W, latitude 30–36° N).

As far as the buoys are concerned, all the statistical measures employed, including those referring to variability, have been significantly improved (over 20%). In particular, one may notice the better statistics achieved for buoy I, which may be attributed to the shorter distance to an observational source (buoy A).

On the other hand, comparison with satellite measurements (Table 5) reveals a considerable reduction of bias and RMSE. The fact that the area of satellite data used for evaluation (6×6°) significantly exceeds the observation

domain (2×3°) partly explains the unchanged values of standard deviation. It worth noticing here that these results are based on the comparison of the average value of available satellite records in the area of the closest model grid point against the corresponding WAM output.

Concerning the time extension of the assimilation impact, the results presented in Fig. 3a, b for buoys C and F, respectively, are characteristic. Similar outputs were also obtained for the rest of the buoys. A more general statistical proof of the longer time impact achieved by the use of the proposed method is given in Tables 6, 7, 8, 9, 10, and 11 where a detailed analysis is presented concerning three 12-h forecasting periods starting from the analysis time T_0: $T_0 \rightarrow T_0+12$, $T_0+12 \rightarrow T_0+24$, $T_0+24 \rightarrow T_0+36$. The obtained improvement in the quality of the forecasts does not restrict only in the first period but remains constantly better of WAM and WAM2 relevant results until the end of the forecasting cycle.

5 Conclusions

It is widely known that data assimilation improves the predictions from operational wave forecasting systems only for a limited time period. In order to overcome this drawback, a new combination of techniques is proposed. The approach involves the use of "artificial" observations inside the forecasting period. The latter are obtained from an initial run of the model followed by the use of post processing techniques—KZ and Kalman filters—that leads to the elimination of any possible systematic error.

The method was applied to an open sea area (near the southwestern coast of the US) for a 3-month period and six different buoys from NOAA's network were utilized as sources of observational data. The results were more than satisfactory, leading to the extension of the assimilation impact to the whole forecasting period as well as to the reduction of the magnitude and the variability of the discrepancies between the final forecasts and observations.

Of course it should be noted that there are limitations in the application of the proposed method since it can be only applied when a continuous flow of observational data (e.g., buoys) is available. For instance, the buoy network is not very dense over the global ocean and is mainly located near coastlines. However, in many cases, such type of data could be very important and available for areas of increased interest, like big harbors, touristic coastline, commercial areas, etc. In these cases, the proposed technique can provide an additional tool for improved forecasts.

The authors believe that the same methodology could also be applied to several other atmospheric or wave parameters, such as temperature, wind speed, mean wave period, etc., offering an alternative way of exploiting assimilation systems and contributing to more accurate predictions at the local scale.

References

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

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

Bratseth A (1986) Statistical interpolation by means of successive corrections. Tellus Ser A 38:439–447

Eskridge RE, Ku JY, Rao ST, Porter PS, Zurbenko IG (1997) Separating different scales of motion in time series of meteorological variables. Bull Am Meteorol Soc 78(7):1473–1483

Galanis G, Anadranistakis M (2002) A one dimensional Kalman filter for the correction of near surface temperature forecasts. Meteorol 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

Greenslade D, Young I (2005) The impact of inhomogenous background errors on a global wave data assimilation system. J Atmos Ocean Sci 10(2):61–93

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

Janssen PAEM, Lionello P, Reistad M, Hollingsworth A (1987) A study of the feasibility of using sea and wind information from the ERS-1 satellite, part 2: use of scatterometer and altimeter data in wave modelling and assimilation. ECMWF report to ESA, Reading

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

Lionello P, Günther H, Hansen B (1995) A sequential assimilation scheme applied to global wave analysis and prediction. J Mar Syst 6:87–107

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

Rao ST, Zurbenko IG, Neagu R, Porter PS, Ku JY, Henry RF (1997) Space and time scales in ambient ozone data. Bull Am Meteorol Soc 78(10):2153–2166

Skandrani C, Lefèvre JM, Queffeulou P (2004) Impact of multi-satellite altimeter data assimilation on wave analysis and forecast. Mar Geod 27:1–23

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