Ensemble Kalman Filter Based Data Assimilation in Wave Models

Paula Etala
Naval Hydrographic Service *

Pablo Echevarría
National Weather Service, Buenos Aires City, Argentina

1. Introduction

The continuous increase in the accuracy of operational wave models is widely recognized. Still, there is room for improvement in the definition of wave conditions by making use of the available information globally provided by remote sensing. Lefèvre (2012) showed the benefit of significant wave height, wind speed and spectral observations in a last-generation wave prediction system.

Since Evensen (1994) first proposed the Ensemble Kalman Filter (EnKF) as an alternative to the Extended Kalman Filter, where flow-dependent background errors were based on an ensemble, several approaches to the model update or analysis step have been introduced. More recently, Ott et al. (2004) proved the feasibility of an entirely local scheme in the so-called Local Ensemble Kalman Filter (LEKF) and Hunt et al. (2007) improved the efficiency of calculations in the Local Ensemble Transform Kalman Filter (LETKF). The 4D-LETKF code shared by T. Miyoshi (Miyoshi and Yamane 2007) and updates has been experimentally implemented in a storm surge ensemble prediction system with very encouraging results (Etala et al. 2012) merging altimeter and tidal gauge storm surge observations. In this article, we will describe an exploratory application of the same scheme to the multivariate surface vector wind and significant wave height data assimilation in a coarse NOAA/NCEP WAVEWATCH III®(Tolman 2009) global implementation.

It is widely known that in its forcing problem, wind sea quickly looses memory of its initial state. A rather short-living impact along the forecast range is then achieved by the assimilation. On the contrary, the improvement of wind and wave mean conditions during development can provide more long-lasting effects in the resulting swell. Improvement in the 6-hour forecast is demonstrated, although such matter is not especifically addressed for longer ranges in this article.

The data assimilation algorithm is described in section 2, where we will briefly explain the basics of the advanced method used in the mean parameters analysis, as well as our approach to the update of wave spectra. The assimilation performed on two storm events and assesed by independent buoy data contributes to illustrate some features in section 3. We will discuss the results in section 4 and, finally, we will summarize conclusions and our thoughts on future work in section 5.

2. The Data Assimilation Method

The general approach of the EnKF combines the flow-dependent background errors provided by an ensemble prediction and the observations to build the analysis ensemble, including the analysis uncertainty. The analysis ensemble so obtained provides the initial state to a new ensemble forecast cycle. Given an n-dimensional model state x and an m-member model ensemble, δX^f is the $m \times n$ matrix containing the m perturbations of the ensemble. The $i_s t$ member perturbation is defined as its difference to the ensemble mean $\delta x_i = x_i - \bar{x}$. We will denote vectors in lowercase and matrices in uppercase.

The forecast step is performed globally in the model space and it is common to any EnKF:

$$X_i^f = M(X_{i-1}^a) \tag{1}$$

Then, the forecast error covariance matrix provided by the ensemble is

$$P^f \approx (m-1)^{-1} \delta X^f (\delta X^f)^T \tag{2}$$

The observation operator H applied to the model variable provides the "model observation", i.e., the model in the observation space $y^f = H(x^f)$. In the ensemble, $\delta Y = H(\delta X^f)$ is the perturbation of the model observation to the ensemble mean.

The p available observations y^o introduce the new information in the observational increment or innovation

* *Corresponding author address:* Paula Etala, Naval Hydrographic Service, Av. Montes de Oca 2124, Buenos Aires City, C1270ABV, Argentina. Email: paulaetala@hidro.gov.ar

in $y^o - y^f$. The way these observations are considered in the analysis step to update the background (section a) is what distinguishes LETKF from other EnKF methods.

Some localization in space and time is required to override spurious background perturbations covariance, not related to the local background uncertainty. An arbitrary local patch is defined at every analysis point where the "smoothed localization" approach increases the observational error with distance to the analysis point, so that the influence of an observation is half its value at a distance σ while decreasing exponentially. σ is called the "localization scale". The weight function 3 is applied to the inverse R matrix in the calculations.

$$W_{(dist)} = e^{-dist^2/2\sigma^2} \tag{3}$$

An analogous localization in time limits the observation influence within the assimilation window to a certain time localization scale. Furthermore, any corrections introduced by the observations only take place in the scale of the significant wave height (h_s), zonal wind component (u_w) and meridional wind component (v_w) ensemble perturbations covariance and cross-covariance, as will be shown in section 3.

a. *The analysis step: LETKF*

The LETKF determines the analysis ensemble locally in the space spanned by the ensemble, as a linear combination of the background perturbations. Therefore, the updated model state becomes:

$$x^a = \bar{x}^f + (\delta x)^f w^a \tag{4}$$

where w is a gaussian random vector, defined so that its ensemble mean $\bar{w}^f = 0$ and covariance $\tilde{P}^f = (k-1)^{-1}I$. The authors demonstrate that the solution for w also minimizes the original analysis cost function and analysis equations analogous to EnKF are solved for the w ensemble in the local $m \times m$ ensemble space, substantially simplified by the variable transformation.

$$\tilde{P}^a = [(m-1)I + (\delta Y)^{fT} R^{-1} (\delta Y)^f]^{-1} \tag{5}$$

where R is the observational error covariance matrix of the locally used observations, modified by the localization 3.

The LETKF belongs to the class of the so-called "deterministic" or square-root EnKF. It updates in a single step the ensemble mean (equation 6) and retrieves the analysis ensemble perturbations from the covariance matrix in equation 7.

$$\bar{w}^a = \tilde{P}^a (\delta Y)^{fT} R^{-1} (y^o - \bar{y}^f) \tag{6}$$

$$\delta w^a = [(m-1)\tilde{P}^a]^{1/2} \tag{7}$$

Equations 5 to 7 for the transformation weights W are those actually solved by the algorithm in the local space of the ensemble. The full analysis ensemble is then built through 4 back into the global model space. Alternatively, we may choose to update the analysis mean or the deterministic model in a hybrid-type approach by

$$\bar{x}^a = \bar{x}^f + (\delta x)^f \bar{w}^a \tag{8}$$

The analysis ensemble perturbations in this scheme are close to the original background ensemble perturbations, as

$$(\delta x)^a = (\delta x)^f (\delta w)^a \tag{9}$$

from 4 and 8.

b. *Updating the Wave Model Spectra*

Several techniques have been intended to determine the local wave spectra from mean parameters and also to extend the impact of the assimilation along the forecast cycle. For instance, Voorrips et al. (1999) determined the wind from the wave height to use it as a smoother to the significant wave field. Here, we adapted the approach by Lionello et al. (1996) and Breivik and Reistad (1994), updated as in ECMWF (2013), except that we do not make any assumptions on wind duration. Instead,we introduce the analysed and background surface wind field into the calculations, assuming no drag change in the analysis. We consider the total wind sea fraction as provided by the model when splitting the analysed h_s into windsea and swell.

The updated spectrum, where f is frequency, while the f superscript denotes forecast or background, is

$$E^a = AE^f(Bf, \theta) \tag{10}$$

From the dimensionless wave energy $\varepsilon_* = Eg^2/u_*^4$ and $\varepsilon_* = 5.054 * 10^{-4} \bar{f}_*^{-2.959}$ with the dimensionless mean frequency $\bar{f}_* = u_* \bar{f}/g$, where u_* is the friction velocity and $E = (H/4)^2$, being H the partitioned significant wave height and U the wind speed, then

$$B = \frac{\bar{f}^f}{\bar{f}^a} = \left(\frac{U^f}{U^a}\right)^{0.3518} \left(\frac{H_{ws}^a}{H_{ws}^f}\right)^{0.6759} \tag{11}$$

and

$$A = B\left(\frac{H_{ws}^a}{H_{ws}^f}\right)^2 \tag{12}$$

for the windsea part of the spectrum, and

$$B = \left(\frac{H_{sw}^a}{H_{sw}^f}\right)^{\frac{1}{2}} \tag{13}$$

and

$$A = B\left(\frac{H^a_{sw}}{H^f_{sw}}\right)^2 \tag{14}$$

for the swell part of the spectrum.

The procedure described in this section completes the wind and wave data assimilation process into the wave model. Analysed vector wind and significant wave height fields are the input to the assimilation routine where local 2-D spectra are inferred for every grid point using 10. The integral parameter fields and spectra after the wave spectra retrieval herein described are the final result of the assimilation and provide the ensemble restart to the new forecast cycle.

3. Experiments and results

The data assimilation experiments presented in this section correspond to the period of 17 to 29 December 2012. They were set on a 6-hour ensemble forecast cycle for a global $1° \times 1°$ WAVEWATCH III®model ensemble. 4 daily global ensemble wind fields where obtained from the NCEP Global Ensemble Forecast System (GEFS) in the THORPEX Interactive Grand Global Ensemble (TIGGE). Buoys available from the JCOMM Wave Forecast Verification Project (WFVP) are used here as independent data for assessment. Vector wind observations from ASCAT scatterometer and significant wave heights from altimeters on Jason 1 and Jason 2 and the predicted model outputs were distributed in hourly slots within every 9-hour assimilation window ($T - 6hrs$, $T + 3hrs$). Thus, the observational departure from the background ensemble mean ($y^o - \bar{y}^f$) in equation 6, as well as the "model observation" perturbation $(\delta Y)^f$ reflect an evolving background field in this 4-D approach. The time localization scale (equation 3) used in the experiments for every observation is 1 time slot, forward and backward. In agreement with results elsewhere, the data assimilation presents a spin-up period of the order of 10 - 12 cycles (2.5 - 3 days). In figure 1 we present the average observational departure from the background ensemble mean vs. the assimilation cycle.

The localization parameter σ in equation 3 is 7° lat/lon in these runs, while the local patch to select the observations used at each grid point is aproximately 25° lat/lon. The column of the error covariance matrix P^f (equation 2) corresponding to the point $50°N$, $140°W$ is partially shown in figure 2 to illustrate the scale of the perturbations in the NE Pacific storm event on 19 December 18:00. All times referred are GMT. The normalized $\overline{\delta h_s \delta h_s}$, $\overline{\delta u_w \delta h_s}$ and $\overline{\delta v_w \delta h_s}$ are presented in the upper, central and lower panel, respectively.

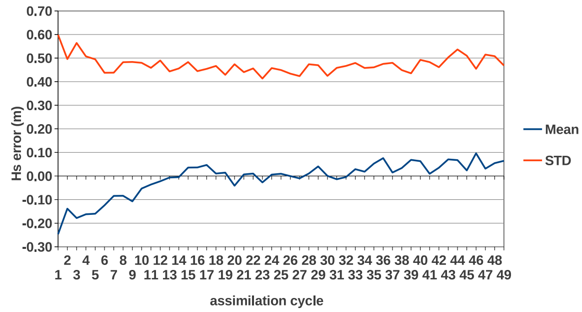

FIG. 1. Evolution of the background observational departure mean and standard deviation with the assimilation cycle. The observations considered are those used by the analysis. The spin-up period was about 2-3 days (8-12 cycles).

The observations used in three of the assimilation cycles, together with the h_s analysis increments produced by the 4D-LETKF are plotted in 3. The analysis starts correcting the NE Pacific storm noticeably, mainly position and size, on 19 December 06:00 (upper panel). It is worth highlighting that only wind information was available on site at 18:00 (center panel). Figure 4 shows the 6-hour forecast $h_s$ensemble mean produced by this full-assimilation run and that from the control no-assimilation cycle at buoys 46184 and 46208 offshore the Canadian West coast (buoys locations are plotted in figure 5).

Supporting the initial correction to wave generation, the waves generated were further updated in succeeding cycles by the scatterometer and altimeters at day 20 00:00 (figure 3lower panel) and the southward swell is still corrected by altimeters one day later (not shown). On the other hand, the h_s contrast between the full and control runs two days later is shown in 5. The swell delay without the cycle-by-cycle analysis update is more remarkable. Buoy 51101 North of Hawaii confirms these results after day 21 in figure 6 (buoy location in figure 5).

Similarly, the assimilation is assessed and validated at buoy 56006 in figure 7, Australia SW coast (buoy location in figure 8 upper right panel). On 19 December, the 6-hour h_s prediction is impacted by the overall assimilation process. Waves produced by a storm moving northeastward along a track located south of the continent are detected at the western coast around 06:00. The normalized 2-D wave energy spectra (not shown) provide evidence of the presence of waves due to the storm passing south. Even though

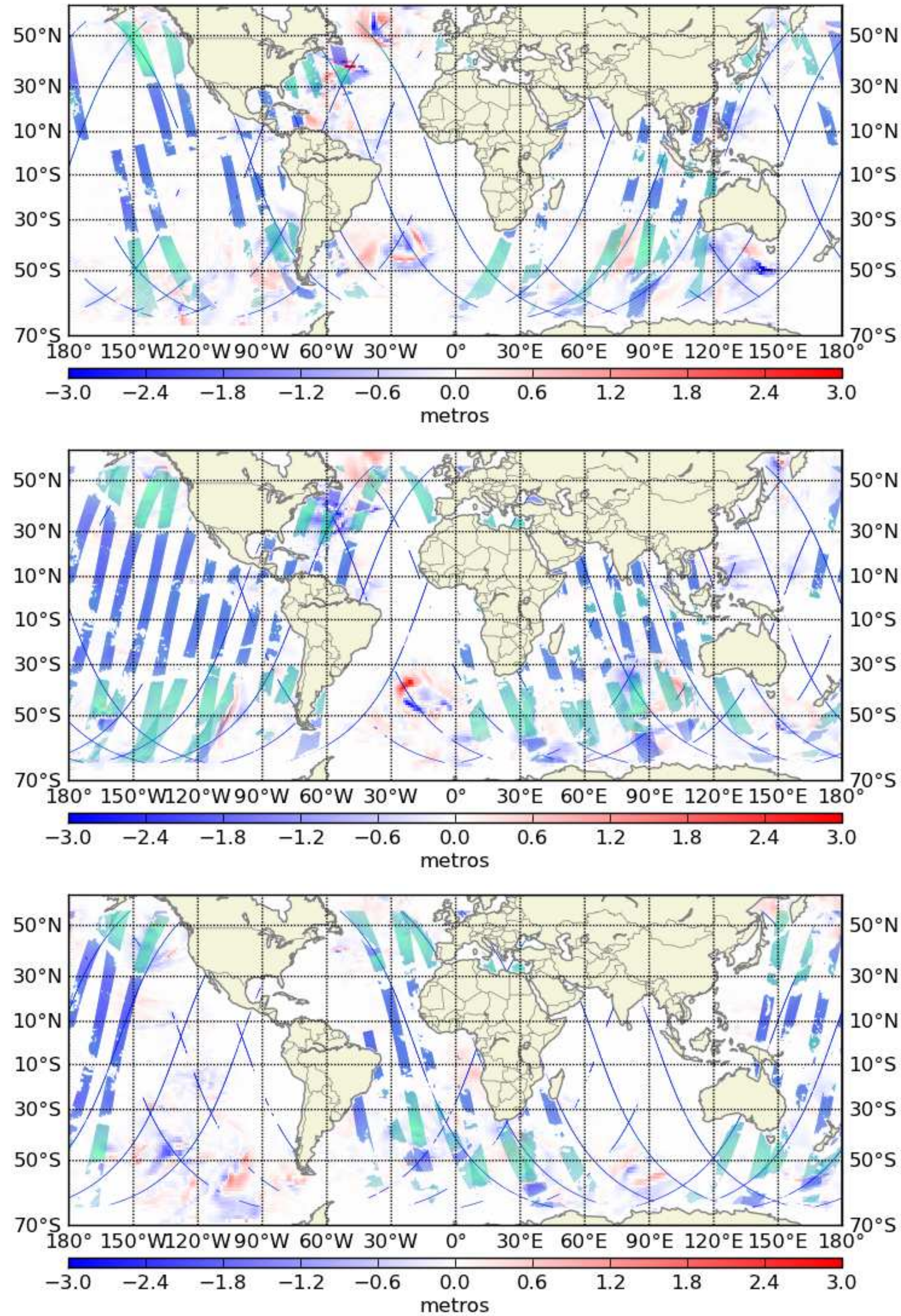

FIG. 3. h_s analysis increments (analysis - background) together with the altimeters and scatterometer tracks used by the analyses in the assimilation windows for 19 December 2012 06:00, 18:00 and 20 December 2012 00:00 (upper, center and lower panel, respectively).

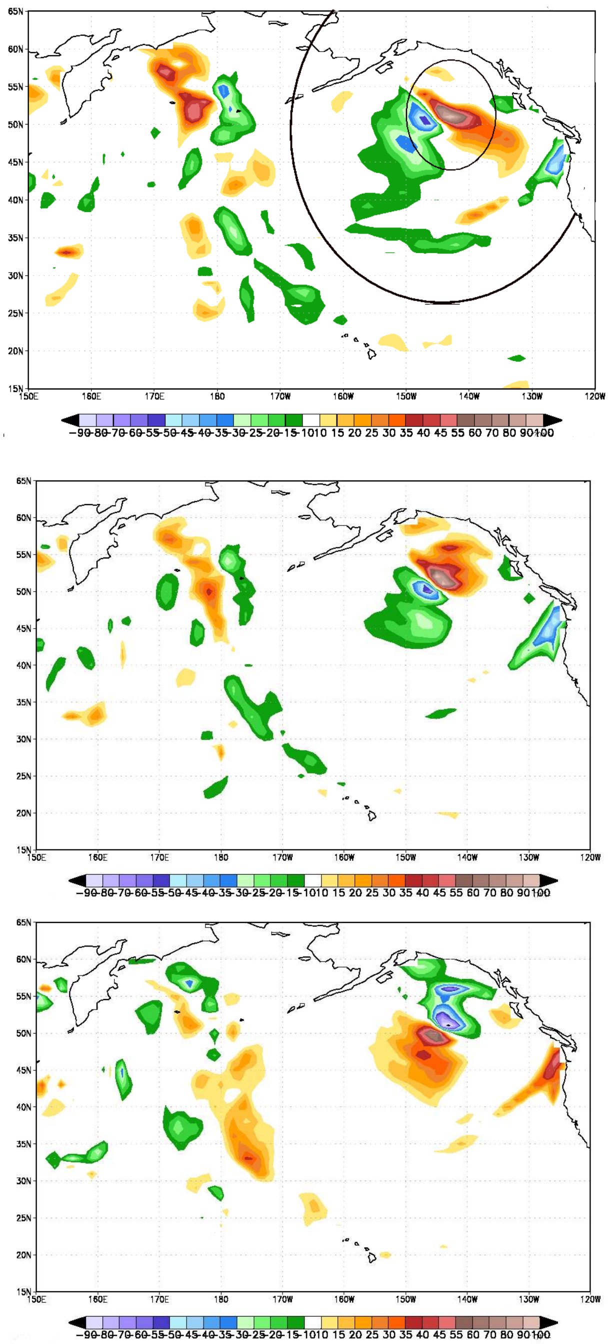

FIG. 2. $\overline{\delta h_s \delta h_s}$ (upper), $\overline{\delta u_w \delta h_s}$ (center) and $\overline{\delta v_w \delta h_s}$ (lower) ensemble perturbations covariances corresponding to point $52°N$, $142°W$ on 19 December 2012 18:00 GMT, normalized with their maximum values. Circles in the upper panel refer to localization scale (inner) and local patch (outer).

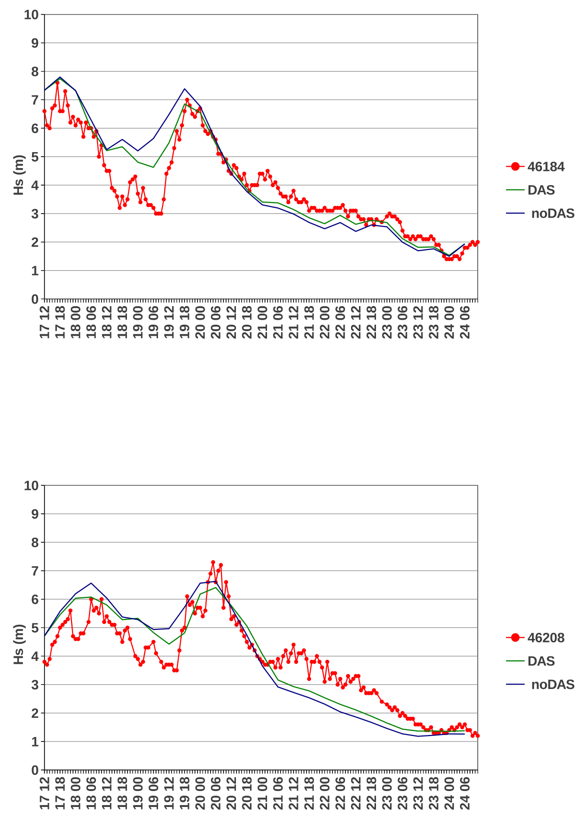

FIG. 4. 17 to 24 December 2012 h_s(m) from buoys 46184 (upper) and 46208 (lower), off-shore North America West coast: red line and dots; green: 6-hour forecast in the data assimilation runs; blue: id., no data assimilation.

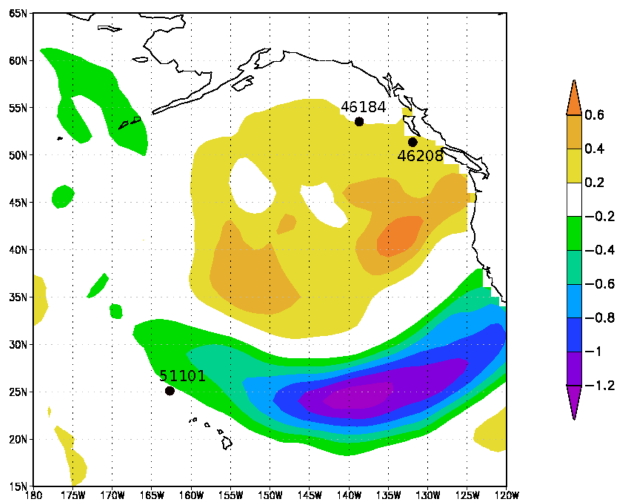

FIG. 5. Difference in the 6-hr predicted h_s(m), data assimilation vs. no data assimilation cycles, 21 December 18:00, together with buoys used in the validation.

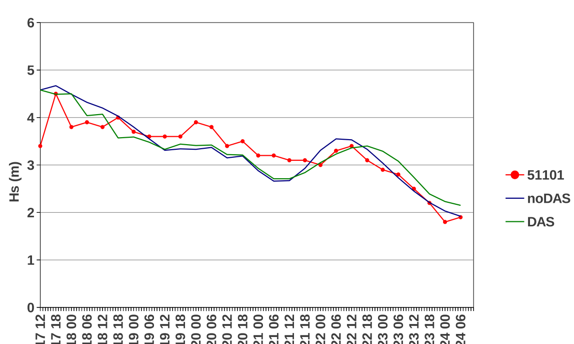

FIG. 6. 17 to 24 December 2012 h_s(m) north of Hawaii. Buoy 51101: red line and dots; green: 6-hour forecast in the data assimilation runs; blue: id., no data assimilation.

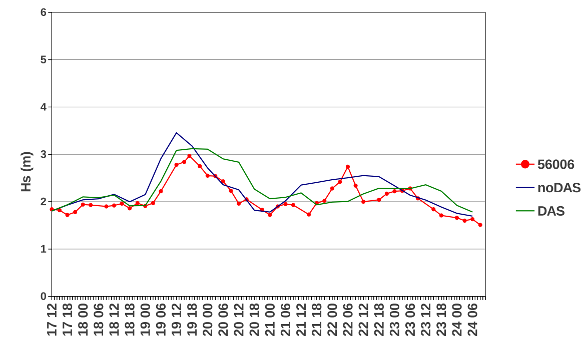

FIG. 7. 17 to 24 December 2012 h_s(m), SW Australia. Buoy 56006: red line and dots; green: 6-hour forecast in the data assimilation runs; blue: id., no data assimilation.

small h_s analysis increments are observed at the buoy location at 06:00 (figure 3 upper panel) in correspondence with a scatterometer swath and altimeter tracks in the assimilation window, the storm has been corrected along its track. The wave train has been sistematically delayed by the assimilation, matching better the buoy observations than the no-assimilation case. Such as in the North Pacific case, we observe a swell that has been corrected at its remote generation area in the past and such correction has been reinforced by sucessive updates.

4. Discusion

The mean statistics of the innovation to the background produced by the altimeter observations, shown in figure 1, measures the discrepancy of the h_s background field with the observations. Evaluated along the assimilation cycles, it may be considered that it evaluates the benefit introduced by the assimilation. We can see in the figure that the mean and standard deviation of the background observational departure is sistematically reduced during the spin-up and it remains quasi-steady thereafter.

The wind observations impact the results in two ways. The cross-covariance of vector wind - wave uncertainties from the ensemble would produce increments to wind wave fields in the analysis step through equation 8, in conjunction with altimeters and even while there were no h_s observations. The update to the windsea part of the model wave spectra (equations 10-12) explicitly introduces the new wind intensity. From figure 3 upper panel we can see that altimeter and scatterometer observations are available in SW Australia for the 19 December 2012 06:00 cycle. We illustrate for that case in 8 how the final h_s after the full assimilation process differs form the intermediate result from the 4D-LETKF. In particular, for the windsea part of the spectra, the analysed wind field and the wave field partitioning further impact the assimilation. We should note that the wind waves field shown in the upper right panel is only illustrative, as it corresponds to the deterministic operational run and not to these experiments. Potential sources of error at this step may be the extra weight given to the background through the conservation of the total windsea fraction and the actual wind to be considered in the partitioning. Some notable changes in storm positioning and/or wind intensity found due to the scatterometer, would justify a revision thereof. Also, in some cases, the change in wind speed would support the need to consider friction velocity in equation 11.

Although we have not gone deeply into the update of the wave spectra in this work, we implicitly assigned an extra weight to the wind analysis by not considering any limited wind duration in equation 11. On the other hand, as from the relatively short and disrupted time periods run so far, we could not conclude that the latter would be a particularly reliable parameter to be conserved from the background wind field. This matter should be considered further.

Instead, we did find evidence of the benefit of the use of vector wind observations in the determination of wave generation through the improvement in storm positioning, intensity definition and timing. The wave fields during the storm shown in NE Pacific (figure 4) were repeatedly corrected by both scatterometer and altimeter observations (figure 3) by considering the "errors of the day" in the background fields (figure 2). That led us to a persistent long-term impact on the waves and swells, (figure 5) as detected by the Hawaiian buoy shown in figure 6. Although only 6-hour forecasts have been considered here, it is a reasonable expectation that longer prediction ranges are also impacted in this sense. It should be noted that the results presented here in the shape of a "deterministic" solution are based on the ensemble mean of the coarse wave model ensemble, which has been built as a prototype with the only purpose of assessing the impact of this data assimilation scheme.

Despite our preference so far to emphasize the deterministic approach through the ensemble mean, the way in which this scheme provides ensemble initialization is very important to a persistent impact. The LETKF property in equation 7 to keep closeness of background and analysis perturbations enables us to easily hold separated restarts to the ensemble members throughout the runs. On the other hand, it is well known that the abundance of satellite observations in the swath might act to the detriment of the independence of observation errors, hence the randomness assumption for the diagonal matrix R in equation 5. The analysis ensemble spread lowers in coincidence with scatterometer tracks, even while we used the 25-km resolution product. The spread has been artificially sustained in these experiments by using some additive inflation (20% of the background ensemble spread) while getting results back into the model ensemble space in equation 9.

5. Concluding remarks

We introduced an advanced method for wave data assimilation, which we applied here to the significant wave height. The consideration of flow-dependent errors, including an evolving background error covariance matrix, and multivariate wind-wave analysis provide an up-to-date efficient use of satellite observations, to produce realistic analysis increments even in sparse data cross-track areas. The 4-D algorithm facilitates the consideration of the right time scale for wave evolution while incorporating the new information (innovation) provided by asyncronous observations in the assimilation window. The application of an EnKF approach makes the method and its implementations independent from the prediction model, hence, from model changes. The latter property makes it particularly suitable for operational use.

The joint use of wind and wave observations allowed us to avoid some assumptions in the wave spectra retrieval, even though we have introduced an extra influence of the analysed wind, beyond the multivariate analysis of wind and wind waves. Even while this algorithm can be improved, the need to infer the whole spectra from integral parameters is a very weak point of every data assimilation system, unless more complete sources of spectral information can be made available. In this sense, we suggest that the most promising ways of improvement should be explored by using the spectral partitioning capability in the WAVEWATCH III®model.

Preliminary results are encouraring, but extensive testing and further enhancement are still needed in the scheme. Improved ways to make use of the analysed winds in the wave model shoud be explored. For instance, forcing the model with the new winds at any previous step from the assimilation, could lead to a more realistic wind sea fraction in the partition of the analysed significant wave height for the spectrum update.

The immediately following step would be to also update the deterministic model, as in the usual approach in hybrid

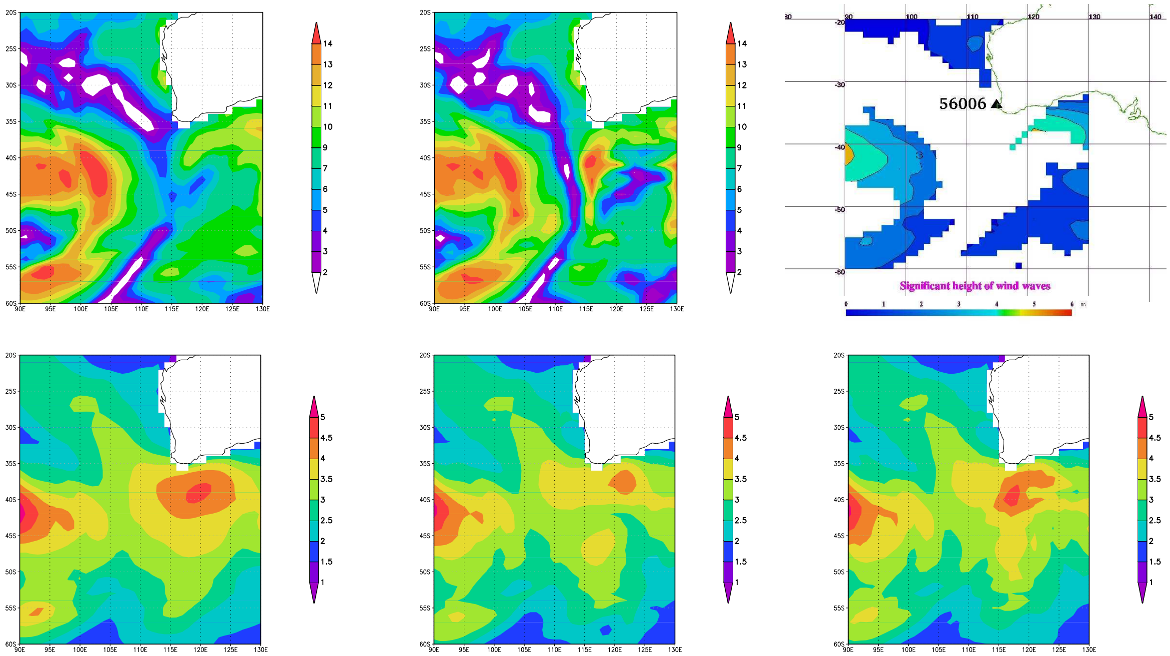

FIG. 8. 19 December 2012 06:00, SW Australia. Upper panel: surface wind intensity (m/s) background (left) and analysis (center), wind waves field as from the deterministic run and buoy location (right). In the lower panel, background h_s (m) (left), analysis after 4D-LETKF (center) and after assimilation into the model (right), respectively. The extra influence of the analysed wind field in the h_s final assimilation is evident.

methods. While in the latter, the ensemble predictions are initialized with any other method, this algorithm would also initialize the ensemble. The proper ensemble initialization to truly represent forecast uncertainties in the short range, so far done with the help of additive inflation, still needs close attention. The matter of how to adequate our analysis methods to high resolution scatterometer data is still to be revised. Moreover, the considerations made by Yang et al. (2009) on the LETKF transformation weights in an atmospheric prediction system suggest that the extension of the analysis to multiscale models could be rather economic and straighforward. Those authors proved that the smoothness of the transformation weights allows to initialize higher resolution models in a single-step approach. Such strategy is still to be tested in our particular problem.

Acknowledgments.

To E. Kalnay and J.H. Alves for fruitful discussions on further improvement. To J. Ruiz for comments and suggestions on this work. The 4D-LETKF/WWIII implementation was supported under PIDDEF grant 046/10, Ministry of Defense. D. Souto and S.M. Alonso provided support to the validation process.

REFERENCES

Breivik, L.-A. and M. Reistad, 1994: Assimilation of ERS-1 Altimeter Wave Heights in an Operational Numerical Wave Model. *Weather and Forecasting*, **9 (3)**, 440 – 451.

ECMWF, 2013: IFS Documentation Cy38r1, Part VII: ECMWF wave model. on-line documentation.

Etala, P., M. Saraceno, and P. Echevarría, 2012: Towards assimilation of satellite altimetry products into a storm-surge model ensemble along the Argentine coast. *6th Coastal Altimetry Workshop*, 112 – 126.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10 14310 162.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, **230 (12)**, 112 – 126.

Lefèvre, J.-M., 2012: Latest developments in wave data assimilation. ECMWF Workshop on Ocean Waves.

Lionello, P., H. Gunter, B. Hansen, P. A. E. M. Janssen, and S. Hasselmann, 1996: *Dynamics and Modelling of Ocean Waves*, chap. VI.5. Cambridge University Press.

Miyoshi, T. and S. Yamane, 2007: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution. *Monthly Weather Review*, **135 (11)**, 3841–3861.

Ott, E., et al., 2004: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. *Tellus*, **56A**, 415–428.

Tolman, H. L., 2009: User manual and system documentation of WAVEWATCH III version 3.14. Technical Note 276, NOAA/NWS/NCEP/MMAB.

Voorrips, A. C., A. W. Heemink, and G. J. Komen, 1999: Wave data assimilation with the Kalman filter. *Journal of Marine Systems*, **19 (4)**, 267 – 291.

Yang, S.-C., E. Kalnay, B. Hunt, and N. E. Bowler, 2009: Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter. *Quarterly Journal of the Royal Meteorological Society*, **135 (638)**, 251–262.