




## A Simplified Assimilation Scheme for a Coastal Wave Model Using Concepts of Particle Filter

SUCHANDRA AICH BHOWMICK,<sup>1</sup>  SMITHA RATHEESH,<sup>1</sup> RASHMI SHARMA,<sup>1</sup> SUJIT BASU,<sup>1</sup> and RAJ KUMAR<sup>1</sup>

**Abstract**—In this work, observations of ocean wave height from satellite altimeters are assimilated into the coastal wave model simulating waves nearshore (SWAN) operating in the Indian coastal waters. The study has two distinctive features. The most important is the use of certain concepts of the modern particle filter technique, which does not represent the model probability density function (PDF) by a Gaussian. The other feature is the joint assimilation of data from three altimeters. The method starts by generating an initial ensemble by the bootstrap technique in which the significant wave height (SWH) field of a control run is perturbed by randomly adding bias to produce a member of the ensemble. At the first assimilation time, a weight-based resampling of the individual members, known as particles, is performed. Stronger particles are retained, while the weaker ones are discarded. In order to keep the ensemble size constant, the algorithm replicates a few strong members. After this resampling, a single model run is employed in the forecast step using a kind of averaging. At the next assimilation time, a synthetic ensemble is again formed by the same bootstrapping, and observations are assimilated. The forecast–assimilation cycle is repeated until the observations are exhausted. Assimilation experiments were conducted for 6 months from February through July in 2016. The power of the technique is evaluated by validating results with altimeter data as well as independent data sets from moored buoys. The results are found to be extremely encouraging for the use of this method in carrying out coastal wave forecasting.

**Key words:** Altimeter data assimilation, particle filter, coastal wave model.

### 1. Introduction

Coastal zones are important from the point of view of fisheries, ship routing, tourism and even naval operations. All these activities are largely dependent on accurate prediction of waves in the coastal oceans. Various coastal wave models can be

used to perform this prediction, a very powerful one being simulating waves nearshore, or SWAN (Booij et al. 1999). The model has been utilized extensively for a number of global coasts for wave prediction and harnessing of renewable wave energy (Dykes et al. 2002; Silva et al. 2016). However, the coastal wave models in general, and SWAN in particular, suffer from incomplete knowledge of initial and boundary conditions and imperfect parameterization of non-linear processes. A partial solution is provided by the assimilation of high-quality observations from satellite altimeters and buoys. Following the pioneering study by Esteva (1988), many researchers have carried out data assimilation in numerical models (Janssen et al. 1989; Lionello et al. 1992; Lionello and Hasselmann 1997; Greenslade 2001; Bhatt et al. 2005; Bhowmick et al. 2016; Rusu and Soares 2015). While all these studies used the method of optimal interpolation or variants thereof, the study by Rusu and Soares (2014) employed a local data assimilation scheme based on recursive successive correction for assimilating buoy data in the vicinity of Portuguese ports. There are also more elegant ensemble-based methods in which the model probability density function (PDF) is allowed to evolve in time. There are two such types of methods, namely the ensemble Kalman filter (EnKF) method of assimilation (Evensen 2009) and the particle filter method (van Leeuwen 2009). Almeida et al. (2016) used EnKF to assimilate data in the SWAN model. Unlike the EnKF, the a priori assumption of the Gaussianity of the model PDF is not made with the particle filter technique. Rather, the model PDF is represented here by several randomly chosen particles shot through the state space of the model, and this freedom makes the method an attractive and more efficient one compared

<sup>1</sup> Space Applications Centre, Indian Space Research Organization, Ahmedabad, India. E-mail: suchandra81@yahoo.com

with the suboptimal EnKF (van Leeuwen 2010). Because of the use of ensembles, both these methods are, however, extremely expensive from a computational point of view. In the case of EnKF, there is a cheaper alternative known as ensemble optimal interpolation (EnOI), wherein the forecast covariance matrix is time-independent as in optimal interpolation, but is formed using an ensemble of model runs. In the case of ocean circulation models, it is found to work reasonably well (e.g., Ratheesh et al. 2014). The search for an analogous simplification of the particle filter for assimilating wave height data led to the present work.

In this study, we have assimilated observations from three different altimeters in SWAN in the coastal waters of India to improve its predictive capability. As mentioned earlier, the method is a simplification of the particle filter and obviously borrows certain concepts from it. The months of February through July (inclusive of both months) of 2016 are chosen for assimilation. The choice is not arbitrary, however. In the month of February, the wave condition in the coastal waters of India generally remains very calm, while in July it is extremely rough. The entire period from February through July shows how Indian Ocean waves evolve from a calm to a rough sea state, with an intermediate moderate sea state. Thus the method can be checked for its efficiency in all possible wave conditions occurring in the Indian Ocean by confining to these 6 months.

## 2. Model and Data

The SWAN model is based on the wave action balance equation with source and sink terms (Booij et al. 1999). Mathematically it is expressed as:

$$\frac{\partial N}{\partial t} + \frac{\partial(C_x N)}{\partial x} + \frac{\partial(C_y N)}{\partial y} + \frac{\partial(C_\sigma N)}{\partial \sigma} + \frac{\partial(C_\theta N)}{\partial \theta} = \frac{S}{\sigma} \quad (1)$$

$$S(\sigma, \theta) = S_{in}(\sigma, \theta) + S_{nl}(\sigma, \theta) + S_{wc}(\sigma, \theta) + S_{bf}(\sigma, \theta) + S_{dib}(\sigma, \theta) \quad (2)$$

where  $\sigma$  and  $\theta$  are the relative frequency and direction of the propagating wave, respectively.  $N(\sigma, \theta)$  is

the two-dimensional wave action density spectrum. The terms  $C_x, C_y$  are the propagation velocities in the geographic space, while  $C_\theta, C_\sigma$  are the propagation velocities in  $\theta$  and  $\sigma$  space, respectively.  $S(\sigma, \theta)$  contains the source and sink terms.  $S_{in}(\sigma, \theta)$  represents the wind input,  $S_{wc}(\sigma, \theta)$  deals with the energy dissipation due to white-capping processes,  $S_{nl}(\sigma, \theta)$  is a nonlinear interaction term that is responsible for the energy redistribution to higher- and lower-frequency waves, and  $S_{bf}(\sigma, \theta)$  and  $S_{dib}(\sigma, \theta)$  represent the energy dissipation due to bottom friction and the depth-induced breaking, respectively. The model is nested in the global wave model (WAM) which is run between 70°S–30°N and 0°–160°E, covering the entire Southern Ocean, so that the impact of incoming swell into the study area (0–25°N and 75–90°E) could be taken into account. The WAM model has a nominal resolution of 25 km, whereas SWAN operates in the inner domain and has a nominal resolution of 5 km. The model nested domain is shown in Fig. 1.

With regard to the data used, it was already mentioned that data from three altimeters have been assimilated. The altimeters are Jason-2, Jason-3, and SARAL/AltiKa. Apart from altimeters, buoy data have been also used for the purpose of validation. We now briefly describe these data sets.

Jason-2 is at an altitude of 1336 km. The orbit is circular at a 66° inclination angle covering 95% of ice-free ocean. Its repeat cycle is 10 days. The altimeter is a dual-frequency one operating at a nominal frequency of 13.6 GHz in Ku-band and 5.3 Hz in C-band. Jason-3 is placed at an elliptical orbit with an inclination of 66.05°, with a perigee of 1331.7 km and apogee of 1343.7 km. SARAL, a joint ISRO-CNES mission, consists of payloads that include a Ka-band high-frequency altimeter (AltiKa). The satellite orbit is sun-synchronous, with an inclination of 98.55° at a height of 814 km. It has a repeat cycle of 35 days. Because of the high frequency, it has a very small footprint and hence is more suitable for coastal studies including assimilation in coastal models. In fact, an attempt was already made to assimilate AltiKa data in SWAN during the tropical cyclone Phailin, with encouraging results (Bhowmick et al. 2016). However, as already stated, in the current study the focus is on a more improved assimilation technique.

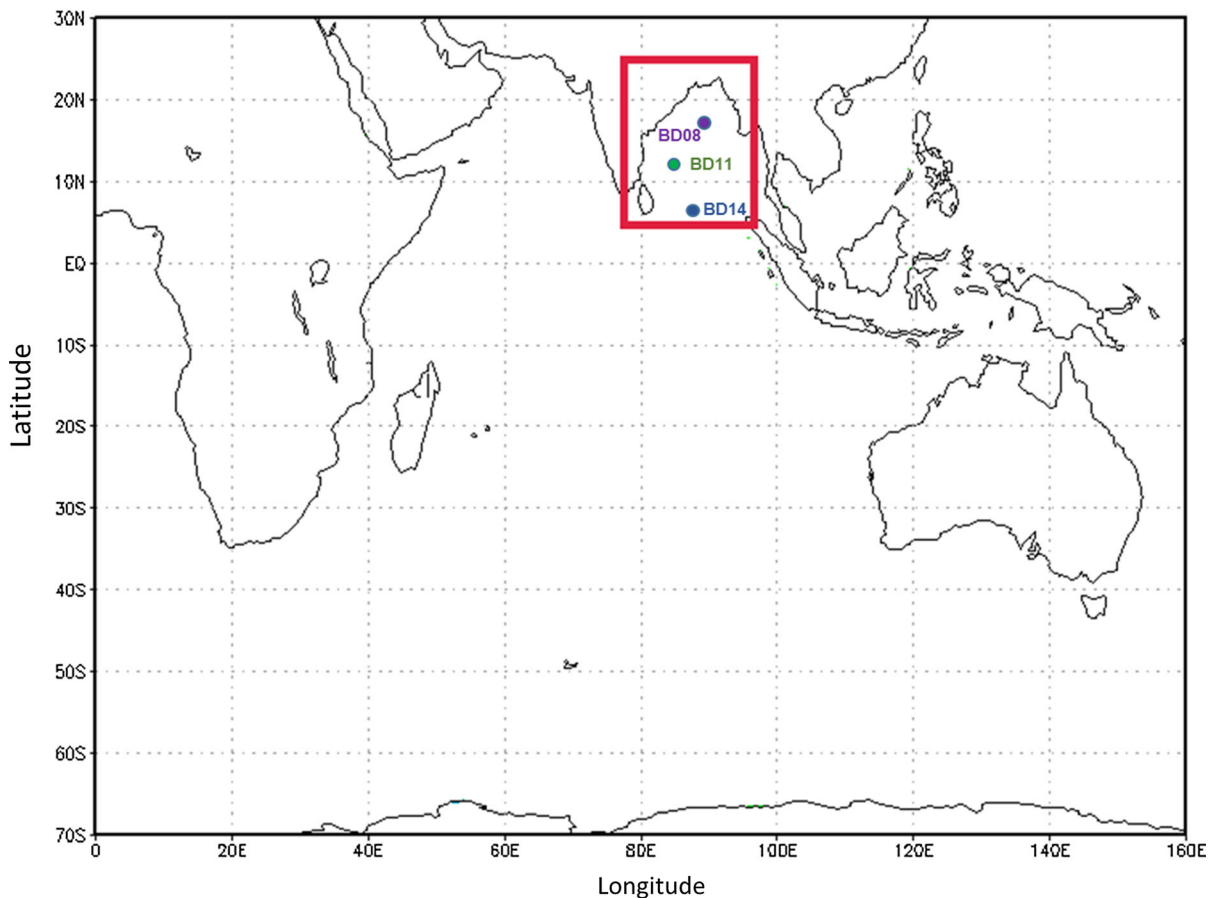


Figure 1

The domain of the SWAN model nested in WAM. Also shown are the locations of the buoys used for validation

Apart from altimeter data, buoy data have also been used for validating the results. In the Indian Ocean, these buoys are deployed by the Indian National Centre for Ocean Information Services (INCOIS). These are moored buoys, and the data consist of meteorological data along with wave data. The wave data from three moored buoys have been used in the study. The buoy identification numbers with their locations are presented in Table 1.

Table 1  
List of buoys used for validation

Buoy ID	Latitude (°)	Longitude (°)
BD08	18.14	89.66
BD11	13.48	84.01
BD14	6.99	88.03

### 3. Assimilation Scheme and Its Implementation

As mentioned earlier, some concepts have been borrowed from the particle filter, which is an ensemble-based technique. The novelty of particle filter stems from the fact that it does not impose any restriction on the form of the model PDF. In particular, it does not assume the PDF to be Gaussian, which is the standard assumption of EnKF, and which is often violated in practice (van Leeuwen 2010). The particle filter is described in many references (Ristic et al. 2004; Dowd 2006; van Leeuwen 2009). The technique has been applied in recent times for assimilating chlorophyll data in biophysical ocean models, with remarkable success (Mattern et al. 2013; Ratheesh et al. 2016). Referring the reader to these references for a thorough understanding of the

method, we confine ourselves to the description of its basic concepts and how they have been implemented in this study.

Particle filter, like any other ensemble-based data assimilation scheme, makes the assumptions that the model state  $\psi$  is described by a multivariate PDF  $p_m(\psi)$ . As in van Leeuwen (2009), we further denote the PDF of observations by  $p_d(d)$ .

The cornerstone of the particle filter is Bayes' theorem, which, for our case, reads:

$$p_m(\psi|d) = \frac{p_d(d|\psi)p_m(\psi)}{p_d(d)} \quad (3)$$

where standard conditional probability notations have been used for the PDF of the model given the observations and vice versa. Following van Leeuwen (2009), hereafter we drop the subscripts in the PDFs, and it will be assumed that the arguments will clarify which particular PDF is being talked about. The PDF in the denominator can be easily calculated from the numerator by integration:

$$p(d) = \int p(d, \psi) d\psi = \int p(d|\psi)p(\psi) d(\psi) \quad (4)$$

and is simply the normalization of the posterior PDF taking into account the observations. Thus, in principle, the calculation of the posterior PDF needs only the knowledge of the PDF of observations given the model and the a priori model PDF. The first can be calculated by a plausible assumption, more often than not a Gaussian one. However, other plausible assumptions are possible, as will be shown presently. The difficulty lies in the calculation of the a priori PDF, and particularly its storage, since the dimension of the state space is prohibitively large. Ensemble-based techniques address this problem in a variety of ways. In particle filtering, the model PDF is represented by a number of random draws from the model state space, called ensemble members, or particles. If there are  $N$  such particles, namely  $\psi_i$  with the index  $i$  spanning the range 1 to  $N$ , the model PDF becomes:

$$p(\psi) = \frac{1}{N} \sum_{i=1}^N \delta(\psi - \psi_i) \quad (5)$$

Substituting for the model PDF from Eq. (5) into the basic Eq. (3), we obtain:

$$p(\psi|d) = \sum_{i=1}^N w_i \delta(\psi - \psi_i) \quad (6)$$

with the weights being given by:

$$w_i = \frac{p(d|\psi_i)}{\sum_{i=1}^N p(d|\psi_i)} \quad (7)$$

The numerator is the probability density of the observations given the model state  $\psi_i$  and is known as likelihood. Weight computation thus requires computation of the likelihood which lies at the core of particle filtering.

Note that the weights are already normalized (so that their sum is unity). This fact is guaranteed by Eq. (7). Often the likelihood is taken to be a Gaussian one, but we want to stress that there is no compelling reason to do so. Of course, if the analytical form of the altimeter error distribution was explicitly known, we could have tried to calculate the likelihood without making any ad hoc assumptions. But we generally know only a typical altimeter data error (typically about 0.5 m), and in studies related to data assimilation (at least in the most prevalent OI ones), the observation error covariance matrix is more often than not taken to be a diagonal one (actually a scalar one with all the diagonal values, i.e. the variances, being identical, typically taken to be about 0.25 m<sup>2</sup>). We feel that this is quite heuristic. We avoid this complexity here by looking upon the weights in Eq. (7) as some “probabilities”, normalized automatically to unit sum. We realize, however, that they cannot be arbitrary, and somehow have to take into account the effect of observations. It is reasonable to suppose that the “probabilities” would be greater if a particle were closer (in some chosen metric) to the observation set (at a given time), i.e. a strong particle. In the case of weaker particles, i.e. those relatively far from the same observation (in the same metric), these “probabilities” (weights) should be less. We make the reasonable assumption (Mattern et al. 2013; Ratheesh et al. 2016) that these weights are inversely proportional to the “distances” (to be defined later) between a specific particle and the observation. Our implementation follows the steps mentioned below (van Leeuwen 2009).

1. Randomly draw  $N$  initial particles.

2. Integrate forward these particles using model equations.
3. Calculate the weights from Eq. (7) using some recipe for likelihood and attach the weights to their corresponding particles.
4. Repeat steps 2 and 3 until all observations are exhausted.

The scheme described so far is called importance sampling and was followed in the earlier days of particle filtering. However, there is a price to pay. A straightforward application results in filter degeneracy (van Leeuwen 2009). This means that after a few assimilation steps, one particle has practically all the weight, while the others have very low weight, and the statistical information content is not a meaningful one. To remedy this situation, the idea of resampling emerged. The basic idea is to discard particles with low weight and to retain multiple copies of particles with higher weight so that the total number of particles remains the same. Thus step (3) of the previous scheme is replaced by (3)<sup>new</sup>.

(3)<sup>new</sup> Calculate the weights and resample the particles and again assign them equal weights ( $1/N$ ). This weight calculation and resampling is the process by which observations get assimilated in the model. Step (4) remains the same as before.

Resampling can be done in many ways. A decision has to be taken about exactly how many copies of a particle with a relatively high weight have to be retained. We have used the concept of stochastic universal sampling (SUS), as outlined by van Leeuwen (2009). Here, all the  $N$  weights are put after each other on a line  $[0, 1]$ , after which a random number is drawn from a uniform density on  $[0, 1/N]$ . Then  $N - 1$  line pieces starting from this number with interval length  $1/N$  are laid on the line  $[0, 1]$ . A particle is chosen when one of the end points of these line pieces falls in the weight bin of that particle.

Now we come to the specific method of calculating the weights as employed in the present study. The method is the same as in Ratheesh et al. (2016), where the authors took the root-mean-square difference (RMSD) between the colocated model and observation pairs as a measure of the distance. In our case, of course, the observations are those of significant wave height (SWH) by three different

altimeters. The following simple collocation procedure has been adopted for collocating altimeter observation and model prediction. Satellite data within 3 h and within 5 km of a model grid point are assumed to be colocated with the model predictions. Assimilation is done every 6 h. Thus, after every 6 h interval, we compute the distances  $d_i$  between observation and  $i$ th particle. The raw weights  $w_i^{\text{raw}}$  are just inverses of these distances ( $1/d_i$ ). In other words, we have taken the likelihood to be proportional to the inverse of distance. The constant of proportionality does not concern us at all, since the weights will be normalized following Eq. (7), and the constant (being present in both numerator and denominator) will be cancelled out. This normalization also ensures that the final weights are dimensionless positive numbers less than unity. But before directly applying Eq. (7), we do further manipulation (Mattern et al. 2013; Ratheesh et al. 2016). We calculate intermediate weights  $W_i = (w_i^{\text{raw}}/\max w_i^{\text{raw}})^a$  where  $a$  is a tuning parameter which spreads or contracts the distribution of the weights before they enter into the resampling process. The final weights are:

$$w_i = \frac{W_i}{\sum_{i=1}^N W_i} \quad (8)$$

This, somewhat heuristic, introduction of a tuning parameter and an intermediate weight has a specific purpose (Mattern et al. 2013). The parameter  $a$  brings the weights closer together (if  $< 1$ ), or spreads them apart (if  $> 1$ ). Note that it does not change the relative ranking of the weights. The parameter  $a$  is not a theoretical requirement, rather a practical necessity. If the raw weights are tightly clustered (as is the case generally for weights coming from satellite observations), then without this tuning, the particles are resampled at almost equal probabilities, surely an undesirable feature in particle filtering. With tightly packed weights,  $a < 1$  brings them progressively closer, which again is undesirable. If  $a > 1$ , the weights are spread apart and the resampling is statistically meaningful. A word of caution is required, however. A high parameter value will cause only the very few particles with very high weight to be more likely to be picked and replicated in resampling, leading eventually to ensemble collapse. Thus a careful analysis is required to assign a value to this

tuning parameter such that observations are fully exploited without causing ensemble collapse. Although in our simplification, in the forecast step between two successive assimilations, we will use only one “average” ensemble member (thus without any risk of collapse), we have retained this feature of tuning parameters for an improved performance of the scheme. This simplification, i.e. the use of a single model run in the forecast step, is the simplification of the conventional particle filter that we made, and this is how we make the technique computationally an inexpensive one. An analogous role was played by EnOI while simplifying the traditional EnKF.

Here we must mention an important point of caution to future researchers regarding the distance  $d_i$ . As mentioned, in our case, the likelihood is proportional to the inverse of distance. Thus it is possible that the distance of observation and model ensemble in some hypothetical scenario may be nil, causing the likelihood to be infinity. Therefore, to deal with such hypothetical situation, one may typically normalize the distance with some reference distance to make it dimensionless and then add this dimensionless quantity to unity before considering the reciprocal of it as the weight.

#### 4. Assimilation Experiments and Results

It was already mentioned that the entire assimilation exercise involves four steps, with step 3 now consisting of calculation of weights with subsequent SUS. The first step means generation of initial particles randomly from the state space of the model. Conventionally, the initial particles are generated by running the model several times with different values of forcing parameter (for example wind for the current model), or a few adjustable internal parameters. However, in the present case, the model resolution is quite high (5 km) and the domain is large. This makes conventional ensemble generation a herculean task. Hence we adopt an easier alternative. The initial model state is generated by running the model from some well-defined pre-forecast time. Next, the model SWH is perturbed using additive bias, which is not allowed to be more than the standard deviation of

model SWH known from past experiments. From past studies with the SWAN model in Indian coastal waters, it is known that this standard deviation can be as high as 1 m in the month of July and for the cyclone-induced waves. Thus the bias level is kept between 0 and 1 m for the 128 members constituting the model ensemble. The choice of the bias for an individual particle is completely random. The bias for an individual particle is a pseudorandom number between 0 and 1 following uniform distribution. For an individual particle, the bias is the same at all grid points. This extreme simplification has been made bearing in mind the computational burden, and can be possibly relaxed in the future by making the bias grid-dependent. The impact of the introduced bias on the root-mean-square error (RMSE) of each particle is shown in Fig. 2 for some arbitrarily chosen time. The plot clearly shows that for this particular instant of time, the RMSE is lowest when the bias is between 0.2 and 0.4 m.

In step 2, the particles are to be marched forward in time using model equations. However, SWH is not a prognostic variable of the model. Thus one must first find a way to calculate the wave spectrum (the prognostic variable) from the SWH in order to carry forward the particles in time. Unfortunately, again, there is no unique way to obtain a spectrum field consistent with a given wave height field. The approach used in this study follows Bhowmick et al. (2016), which in turn was inspired by an earlier work (Greenslade 2001). Ideal wave spectra contain energy due to wind sea or locally generated wind, whose energy is much stronger and represents the higher-frequency part of the spectra. The low-frequency part of the spectra is due to swells. This scheme assumes that the model predicts the distribution of energy in wave and swell accurately. Thus the model-generated wave spectrum is scaled using the ratio of the square of the perturbed wave height to that of the original (unperturbed) wave height. This eventually changes the slope of the spectra without altering the shape of the spectra.

As outlined earlier, in the next step, first the raw weights are calculated from the distances between observation and each particle (actually the model counterpart of observation calculated from each individual particle, namely the SWH). Next, the

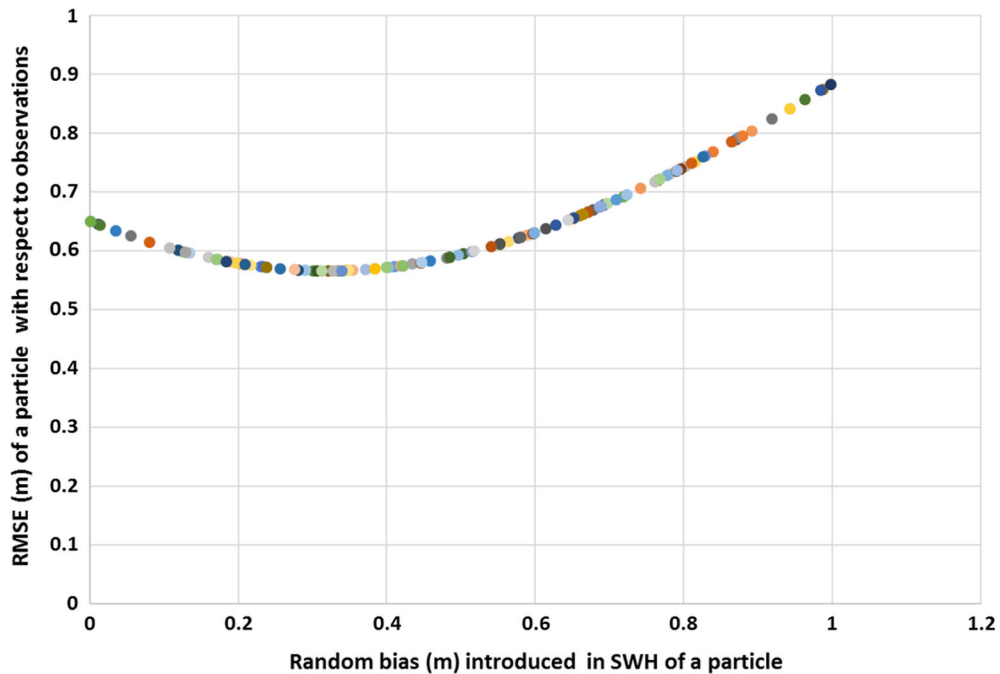


Figure 2  
Variation in RMSE of each particle vs. random bias introduced

intermediate weights are calculated using the raw weights and the adjustable parameter  $a$ . The variation in the intermediate weights with respect to random bias introduced in SWH for several combinations of  $a$  is shown in Fig. 3. Clearly, for  $a = 2, 4, 8$ , the weights are still tightly clustered. As already explained, this is an unacceptable feature of assimilation. However, for  $a = 16$  and  $a = 32$ , the weights for 128 members are reasonably spread apart. The reason for choosing the value 16 is that for  $a = 32$ , the fall in weight is sudden, making the weights extremely sensitive to the bias introduced in SWH. On the other hand, for  $a = 16$ , the fall is gradual and the sensitivity is much less pronounced.

The final weights are simply the normalized intermediate weights from Eq. (8). A resampling is next performed using SUS as explained earlier. Strong ensemble members (with relatively large weights) are kept, while weaker ones are discarded. In order to keep the number of members constant, some strong particles are drawn more than once during resampling. After the resampling is completed, we form a mean wave height field by simple

averaging. As at the start of the procedure, we form a modified wave spectrum field from this mean wave height field, hereafter called the analysed field. We now apply time marching to this modified “average” spectrum. In this way, we have used a single model run instead of running the entire ensemble, thus saving precious computer time. The entire process of distance-based weight computation and subsequent resampling constitutes the process by which the observations are assimilated. This forecast–assimilation cycle continues for the entire duration of the experiment. We would again like to point out how our method differs from a traditional particle filter. After generating the initial ensemble, the particles are marched forward in time until the first observation time. Then the entire observation set is assimilated using the procedure outline above. In the traditional particle filter, time marching is applied to all the particles constituting a new ensemble after assimilation, whereas in our method only a single “average” ensemble is marched forward after assimilation, thus greatly reducing the computational burden. This is the all-important simplification of the particle filter

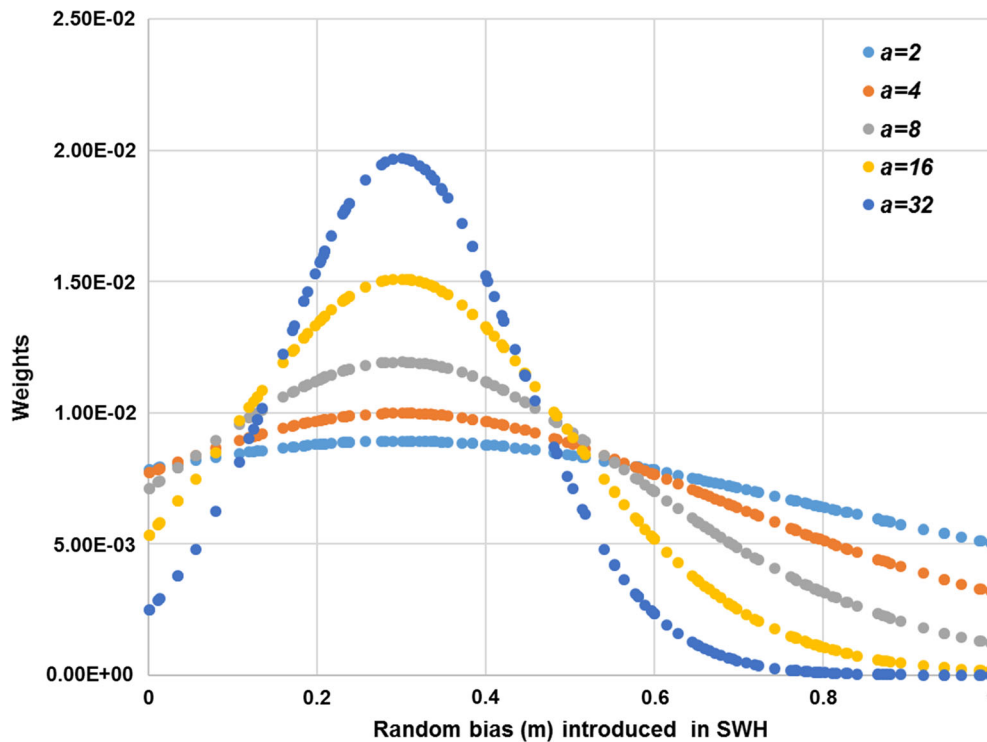


Figure 3

Schematics showing the variation in weight with random bias introduced for various values of the parameter ‘a’. It brings the weights closer in the case that they are widely separated (with  $a < 1$ ), or it spreads them apart (with  $a > 1$ ) if they are tightly packed, similar to this case

attempted in this study. It is somewhat analogous to the EnOI simplification of the EnKF method. This fact is reflected in the title of this paper. It is worth noting that, instead of averaging after resampling (which introduces some noise), we could have taken a weighted “average” of the ensemble before resampling, thereby saving some computer time. This might have been possible. But the fact that resampling replicates some strong members while discarding some weaker ones was quite appealing to us, and we retained it. Nevertheless, since we are running just one average member and not the entire ensemble, the necessary computations are already reduced considerably. Thus we could afford this small luxury.

The experiment begins with short spin-up run of the model, for January 2016. This is followed by two parallel model runs in two different modes, one in control mode (without any assimilation) and another in assimilation mode from the months of February through July 2016. The National Centre for Medium Range Weather Forecasting (NCMRWF) wind at

25 km horizontal resolution is used for the purpose of forcing the model. The assimilation is carried out using the technique described earlier. AltiKa data along with Jason-2 and Jason-3 data are used for assimilation. For each day, the assimilation is carried out four times at 6 h intervals. This is done for the entire 6 months from February through July of 2016. For illustrative purposes, we show the result for 31 July in Fig. 4.

The altimeter tracks on 31 July 2016 are overlaid on an analysed field (obtained after assimilation) and control field separately in Fig. 4. The two black circles in the two panels encircle the superimposed altimeter tracks. In the lower panel, the altimeter tracks are clearly demarcated, indicating the failure of the control run to correctly predict the wave height at the track location. On the other hand, the altimeter tracks are almost indistinguishable from the analysed field in the upper panel, illustrating the power of the particle filter technique.

We have assimilated altimeter data, and the results have been compared with data from the same



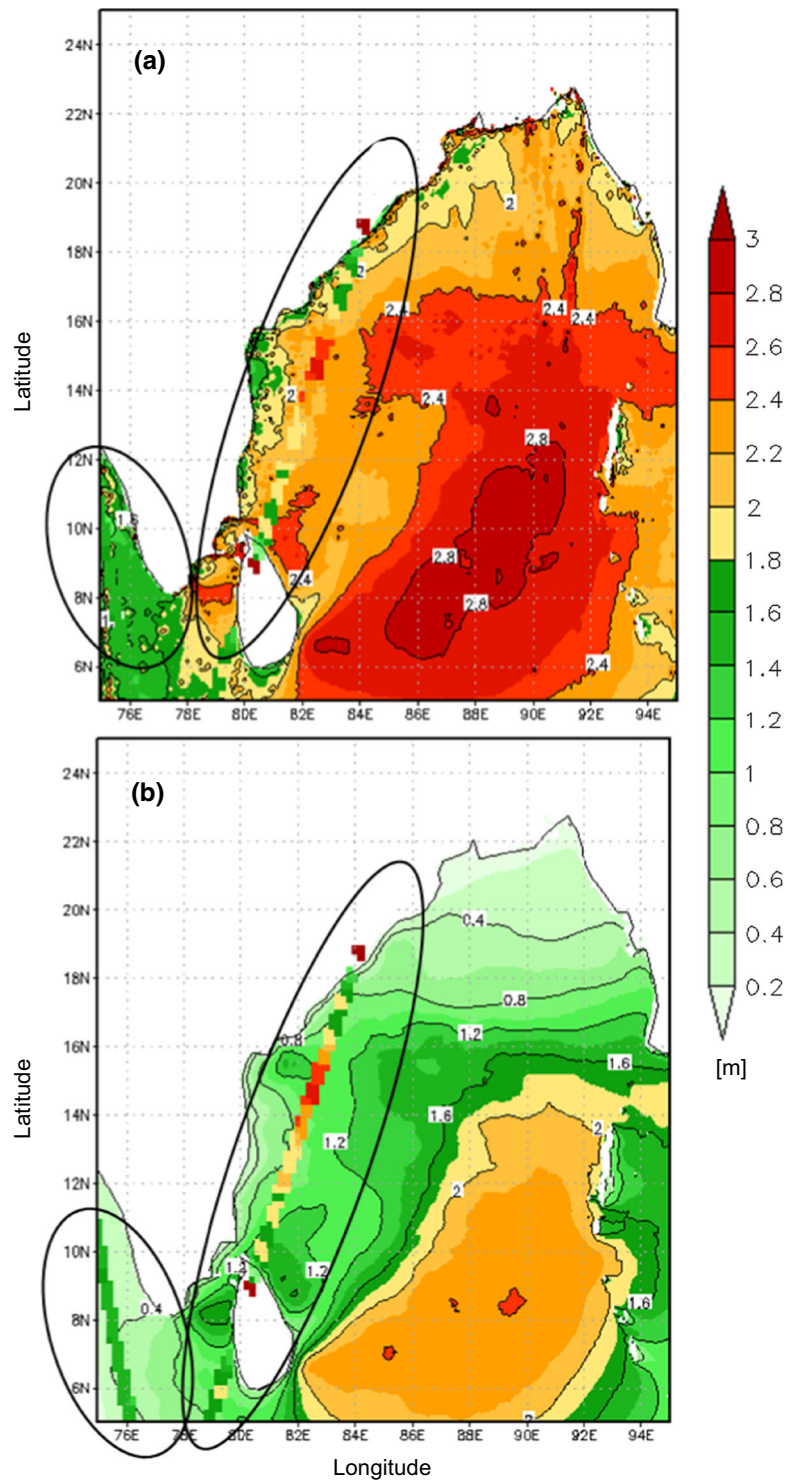


Figure 4

The SWH ( $m$ ) over the Bay of Bengal on 31 July 2016 at 00:00 GMT from **a** analysed field and **b** control field with overlaid SWH from altimeter tracks

source. As a confidence-building measure, we now interpret our results by comparing them with an independent data set, coming from moored buoys, the locations of which are mentioned in Table 1. The model-simulated SWH from the control run and the assimilation run are extracted at the buoy locations and are compared with the buoy-measured SWH. The time series of this comparison from the months of February through July are shown in Fig. 5. Clearly, at all the buoy locations, the SWH from the assimilation run outperforms that from the control run. The impact appears more prominent for the month of July, when

the wave heights are relatively large. In the month of February, however, in the initial part of the time series, the assimilation run mildly overestimates the wave heights. The scenario, however, soon changes in the later part of February, when the control run underestimates the observations to which now the assimilation is quite close. To better quantify the improvement, statistical parameters including RMSE, mean absolute error (MAE), bias with respect to observation and scatter index (SI) are calculated for SWH at each buoy location, and the results are shown in Table 2. The scatter of observed and simulated

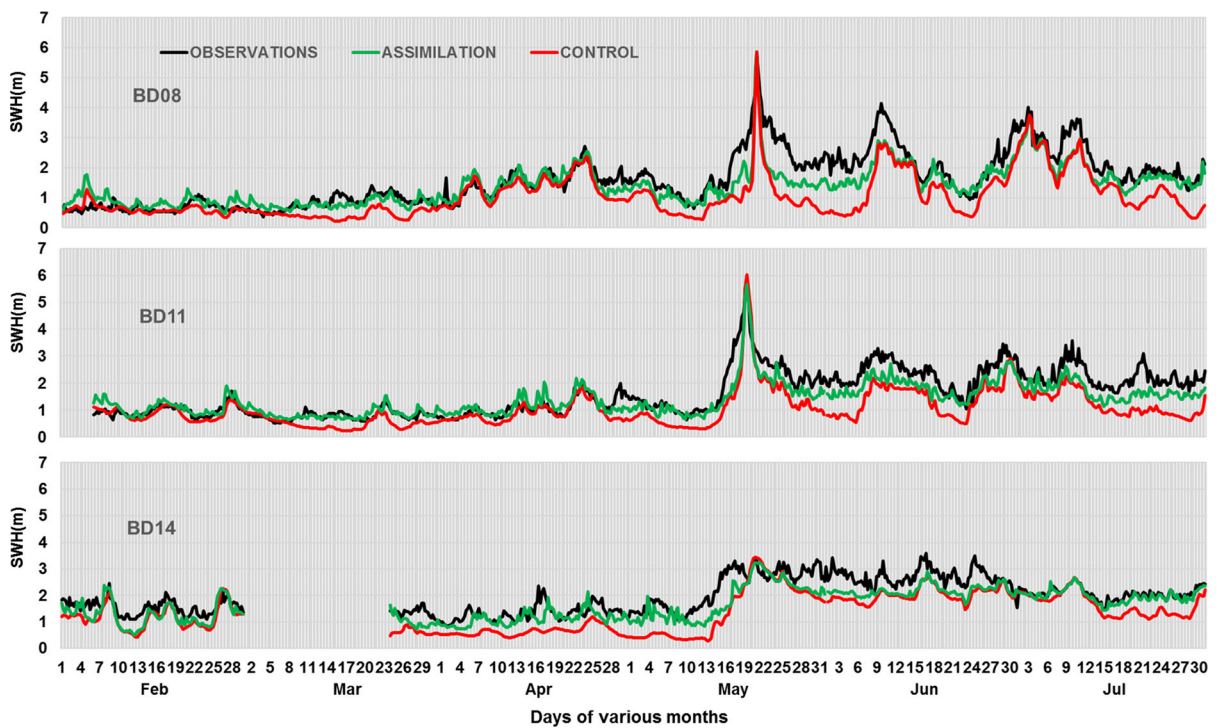


Figure 5

Time variation in SWH from observations, assimilation and control runs at three buoy locations in the Bay of Bengal from February through July 2016

Table 2

Statistical parameters related to SWH at three buoy locations in the control and assimilation runs from February through July of 2016

Buoy ID/statistics	BD08 ( $n = 726$ )		BD11 ( $n = 705$ )		BD14 ( $n = 634$ )	
	Control	Assimilation	Control	Assimilation	Control	Assimilation
MAE (m)	0.488	0.290	0.490	0.291	0.556	0.325
Bias (m)	0.461	0.136	0.477	0.139	0.533	0.272
RMSE (m)	0.610	0.370	0.622	0.365	0.623	0.406
SI (%)	38.14	23.13	38.80	22.81	31.85	20.77

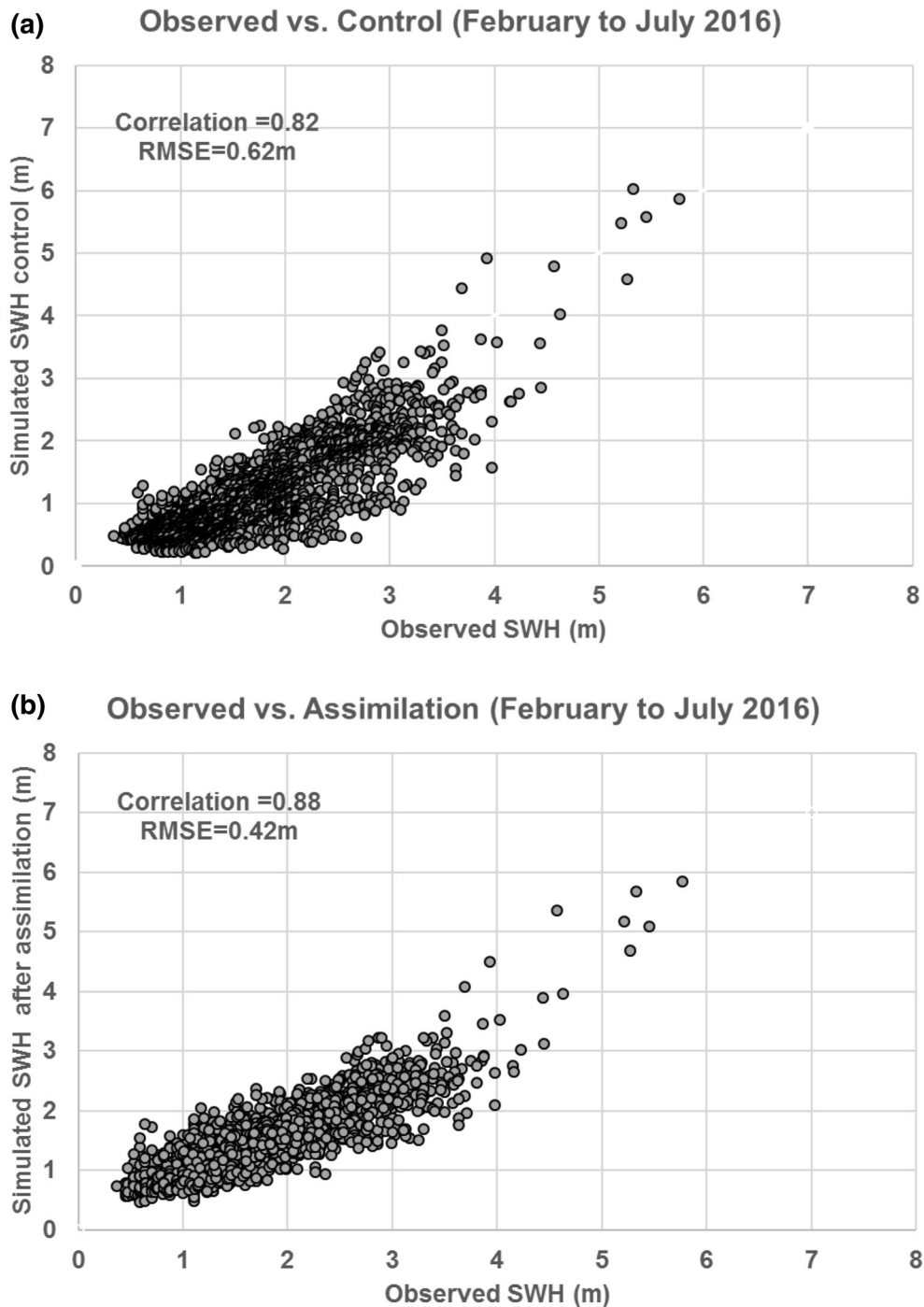


Figure 6

The scatter of SWH from observations vs. **a** control run and **b** assimilation run for all three Bay of Bengal buoys from February through July 2016 (total number of points = 2061)

SWH is shown in Fig. 6 for all months of the experiment for all the buoy locations. The figure speaks for itself, vividly demonstrating

superiority of the assimilation run. A comparison of the swell at these buoy locations for the month of July is shown in Fig. 7. The swell has been improved

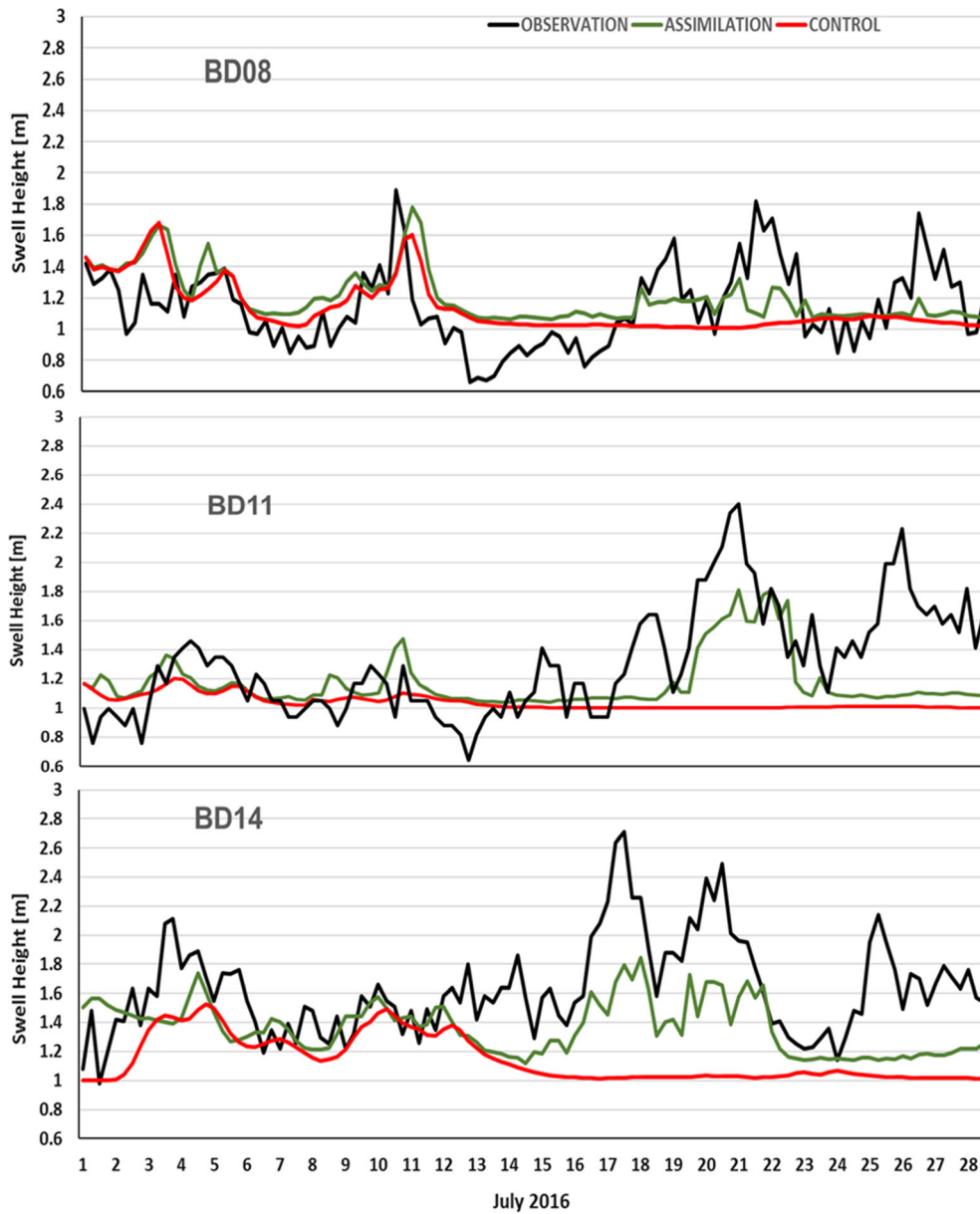


Figure 7

Time variation in swell height from observations, assimilation and control runs at three buoy locations in the Bay of Bengal for the high-swell phase of July 2016

substantially by the assimilation of SWH. However, a better match between observation and simulation is noted in the case of wind waves as compared with that of swells. The SWH measurements from altimeters that are assimilated in the model are influenced by both local and remote wind forcing. Hence one may expect an equal improvement in swell and wind waves after assimilation. Here, however,

one must bear in mind that during this entire assimilation process, a time window of 6 h is considered, i.e. altimeter measurements within an interval of  $\pm 3$  h around a particular time step (after every 6 h) are assimilated in the model. Thus there is a difference between the actual measurement times and the assimilation time. In July, the Indian Ocean has consistent swell inflow from the Southern Ocean due

Table 3

Statistical parameters related to swell heights at three buoy locations in the control and assimilation runs from February through July of 2016

Buoy ID/statistics	BD08 ( $n = 726$ )		BD11 ( $n = 705$ )		BD14 ( $n = 634$ )	
	Control	Assimilation	Control	Assimilation	Control	Assimilation
MAE (m)	0.182	0.193	0.305	0.239	0.492	0.318
Bias (m)	0.06	0.089	0.263	0.130	0.268	0.486
RMSE (m)	0.258	0.269	0.466	0.3423	0.620	0.396
SI (%)	22.67	23.81	35.744	26.24	37.1	24.27

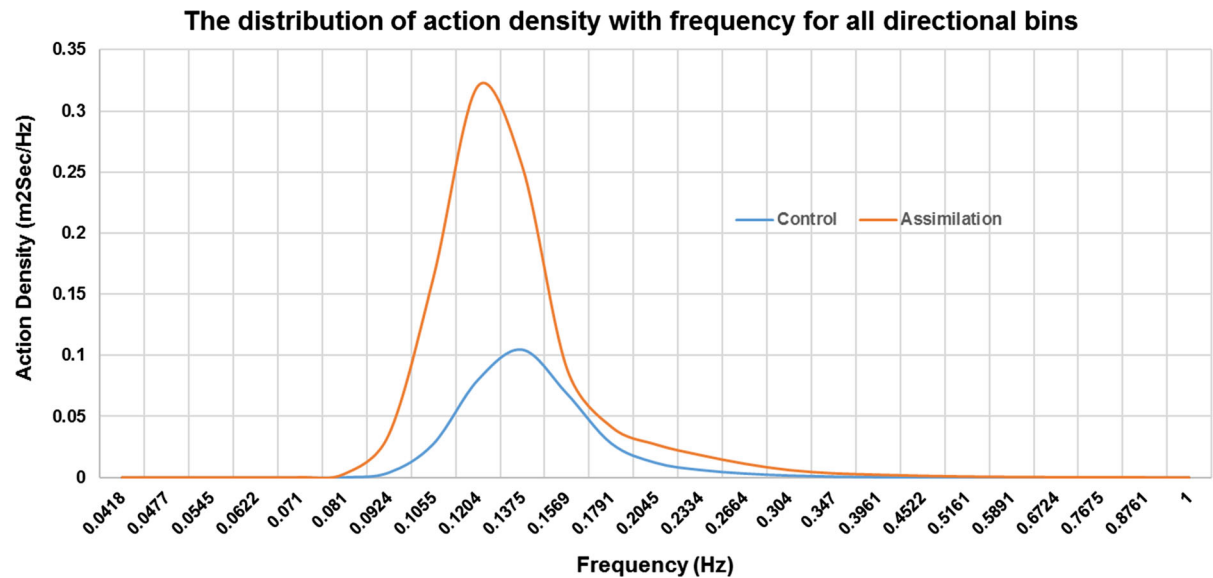


Figure 8

The frequency-wise distribution of action density spectra at some arbitrary location in the Bay of Bengal for 31 July 2016 showing the way assimilation modified the model spectra

to monsoon winds. Hence the time lag could be significant enough to cause the discrepancy in the case of swells. Table 3 shows the RMSE of the swell from the control run and assimilation run at the buoy locations during all the months of the experiment. The directional spectra for all frequencies are shown in Fig. 8. This spectrum is at a location within the SWAN domain far from altimeter and buoy observations. At this location, therefore, measurements cannot directly influence the spectrum and hence modify it, and so whatever change is seen in the spectrum after assimilation is a genuine effect of assimilation. This spectrum illustrates the way this technique modifies the energy distribution over various frequencies. Very clearly, the SWH assimilation

has shifted the peak frequency. In the lower-frequency part, the energies are enhanced, which may have contributed to the improvement in the swell waves.

### 5. Conclusions

Space-based satellite observations of waves are crucial for wave forecasting, since a wide network of in situ observations is simply non-existent in the oceans. Assimilation of these observations into wave models is vital for an accurate medium-range forecast of the wave state. Until now, satellite data were assimilated into wave models using traditional

techniques, in which either the model PDF was assumed to be stationary or, even if it was allowed to propagate in time, it was arbitrarily assumed to be Gaussian, thus severely limiting their applicability for assimilating data into high-resolution nonlinear coastal models with PDFs, not necessarily Gaussian ones. In recent years, an appealing alternative, not bound by these restrictions, has emerged. It goes by the name of particle filter, which although ensemble-based does not assume any preassigned form of the model PDF. In this paper we have used concepts from this technique for assimilating satellite altimeter-derived SWH data into a state-of-the-art coastal wave model operating in the Indian coastal waters. Another novel feature of this study is the use of data from three altimeters. Data have been assimilated sequentially at 6 h intervals. Our method is of course simplified because the full power of the particle filter technique has not been exploited, the reason being purely computational. In effect, our technique is a counterpart of EnOI in the particle filter domain. The assimilation run spans the period from February through July (inclusive of both months) of 2016. A parallel control run is also carried out for the same period. The results are first interpreted in terms of comparison with altimeter data, and then, more importantly, with independent buoy observations. The results unequivocally demonstrate the efficiency and power of this simplified technique based on particle filter concepts for assimilating data into nonlinear high-resolution models for forecasting waves in the Indian coastal waters.

#### Acknowledgements

The authors would like to express their sincere gratitude to the Director, Space Applications Centre, for encouragement and to the Group Director, Atmospheric and Oceanic Sciences Group, for motivation. The authors are indebted to the Archiving Validation and Interpretation of Satellite Oceanography (AVISO) data and Meteorology Oceanography Satellite Data Archival Centre (MOSDAC) database systems for providing altimeter data sets. They are also thankful to INCOIS for the buoy data sets.

Finally, they are immensely thankful to the two esteemed reviewers for their helpful comments and suggestions, which helped to greatly improve the quality of the manuscript.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### REFERENCES

- Almeida, S., Rusu, L., & Soares, C. G. (2016). Data assimilation with the ensemble Kalman filter in a high-resolution wave forecasting model for coastal areas. *Journal of Operational Oceanography*, 9(2), 103–114.
- Bhatt, V., Kumar, R., Basu, S., & Agarwal, V. K. (2005). Assimilation of altimeter significant wave height into a third generation global spectral wave model. *IEEE Transactions on Geoscience and Remote Sensing*, 43(1), 110–117.
- Bhowmick, S. A., Basu, S., Sharma, R., & Kumar, R. (2016). Impact of assimilating SARAL/AltiKa SWH in SWAN model during Indian Ocean tropical cyclone Phailin. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3), 1812–1817.
- Booij, N., Ris, R. C., & Holthuisen, L. H. (1999). A third-generation wave model for coastal regions I. Model description and validation. *Journal of Geophysical Research*, 104(4), 7649–7666.
- Dowd, M. (2006). A sequential Monte Carlo approach for marine ecological prediction. *Environmetrics*, 17(5), 435–455.
- Dykes, J.D., Hsu, Y.L., & Rogers, W.E. (2002). The development of an operational SWAN model for NGLI. Conference proceedings, Oceans 02 MTS/IEEE 29–31 Oct 2002 Biloxi, MI, USA.
- Esteve, D. C. (1988). Evaluation of preliminary experiments assimilating Seasat significant wave heights into a spectral wave model. *Journal of Physical Oceanography*, 93, 14099–14105.
- Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Berlin: Springer.
- Greenslade, D. J. M. (2001). The assimilation of ERS-2 significant wave height data in Australian region. *Journal of Marine Systems*, 28, 141–160.
- Janssen, P. A. M., Lionello, P., Reistd, M., & Hollingworth, A. (1989). Hindcasts and data assimilation studies with the WAM model during the Seasat period. *Journal of Geophysical Research*, 94, 973–993.
- Lionello, P., Gunther, H., & Janssen, P. A. M. (1992). Assimilation of altimeter data in a global third generation wave model. *Journal of Geophysical Research*, 97, 14453–14474.
- Lionello, P., & Hasselmann, K. (1997). An optimal interpolation scheme for the assimilation of spectral wave data. *Journal of Geophysical Research*, 102(C7), 15823–15836.
- Mattern, J. P., Dowd, M., & Fennel, K. (2013). Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations. *Journal of Geophysical Research*, 118, 2746–2760.
- Ratheesh, S., Chakraborty, A., Sharma, R., & Basu, S. (2016). Assimilation of satellite chlorophyll measurements into a

- coupled biophysical model of the Indian Ocean with a guided particle filter. *Remote Sensing Letters*, 7(5), 446–455.
- Ratheesh, S., Sharma, R., & Basu, S. (2014). An EnOI assimilation of satellite data in an Indian Ocean circulation model. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7), 4106–4111.
- Ristic, B., Arulampalam, S., & Gordon, N. (2004). *Beyond the Kalman filter: particle filters for tracking applications*. London: Artech House.
- Rusu, L., & Soares, C. G. (2014). Local data assimilation scheme for wave predictions close to the Portuguese ports. *Journal of Operational Oceanography*, 7(2), 45–57.
- Rusu, L., & Soares, C. G. (2015). Impact of assimilating altimeter data on wave predictions in the western Iberian coast. *Ocean Modelling*, 96, 126–135.
- Silva, D., Rusu, E., & Soares, C. G. (2016). High resolution wave energy assessment in shallow water accounting for tides. *Energies*, 9(9), 761.
- van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review*, 137, 4089–4114.
- van Leeuwen, P. J. (2010). Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136, 1991–1999.

(Received March 15, 2019, revised September 30, 2019, accepted October 3, 2019, Published online October 23, 2019)

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.