

# Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of High-Frequency radar surface currents – application to the German Bight

A. Barth<sup>1,2</sup>, A. Alvera-Azcárate<sup>1,2</sup>, K.-W. Gurgel<sup>3</sup>, J. Staneva<sup>4</sup>, A. Port<sup>5</sup>, J.-M. Beckers<sup>1,2</sup>, and E. V. Stanev<sup>4</sup>

<sup>1</sup>GeoHydrodynamics and Environment Research (GHER), MARE, AGO, University of Liège, Liège, Belgium

<sup>2</sup>National Fund for Scientific Research, Belgium

<sup>3</sup>Institute of Oceanography, University of Hamburg, Germany

<sup>4</sup>Institute for Coastal Research, GKSS Research Center, Geesthacht, Germany

<sup>5</sup>Institute for Chemistry and Biology of the Marine Environment (ICBM), University of Oldenburg, Germany

Received: 7 October 2009 – Published in Ocean Sci. Discuss.: 26 October 2009

Revised: 14 January 2010 – Accepted: 18 January 2010 – Published: 4 February 2010

**Abstract.** High-Frequency (HF) radars measure the ocean surface currents at various spatial and temporal scales. These include tidal currents, wind-driven circulation, density-driven circulation and Stokes drift. Sequential assimilation methods updating the model state have been proven successful to correct the density-driven currents by assimilation of observations such as sea surface height, sea surface temperature and in-situ profiles. However, the situation is different for tides in coastal models since these are not generated within the domain, but are rather propagated inside the domain through the boundary conditions. For improving the modeled tidal variability it is therefore not sufficient to update the model state via data assimilation without updating the boundary conditions. The optimization of boundary conditions to match observations inside the domain is traditionally achieved through variational assimilation methods. In this work we present an ensemble smoother to improve the tidal boundary values so that the model represents more closely the observed currents. To create an ensemble of dynamically realistic boundary conditions, a cost function is formulated which is directly related to the probability of each boundary condition perturbation. This cost function ensures that the boundary condition perturbations are spatially smooth and that the structure of the perturbations satisfies approximately the harmonic linearized shallow water equations. Based on those perturbations an ensemble simulation is carried out using the full three-dimensional General Estuarine Ocean Model (GETM). Optimized boundary values are obtained by assimilating all observations using the covariances of the ensemble simulation.

## 1 Introduction

Most ensemble-based assimilation schemes such as the EnKF (Ensemble Kalman Filter, Evensen, 2003, 2004), ESSE (Error Subspace Statistical Estimation, Lermusiaux and Robinson, 1999), SEIK filter (Singular Evolutive Interpolated Kalman filter, Pham, 2001) are sequential: the ensemble is updated using observations at the time they are measured. In this sequential approach however the model undergoes a sometimes vigorous adjustment process when the model is restarted (e.g. Malanotte-Rizzoli et al., 1989) if some assumptions of the underlying assimilation scheme are not verified (e.g. poorly known error covariances, model biases or non-Gaussian pdf). A too frequent assimilation of observations can even lead to the situation where the assimilation degrades the model results due to the High-Frequency motions generated by the assimilation (Talagrand, 1972).

Several approaches have been proposed to reduce this well known problem. Instead of applying the analysis correction at the assimilation time, in the Incremental Analysis Update (Bloom et al., 1996) the correction is added incrementally over several time steps, reducing the generation of spurious gravity waves. This scheme has been used with large-scale ocean models (e.g. Kepenne et al., 2005; Ourmières et al., 2006) but it is questionable if it can also be used in regional models representing fast ocean processes such as tides.

In free-surface ocean models, such transient motions often propagate as fast shallow-water barotropic waves. To prevent the generation of spurious gravity waves by assimilation, Dobricic et al. (2007) proposed an iterative method to damp the divergence in the analysis increment of the flow field. Barotropic gravity waves can also be removed explicitly by



Correspondence to: A. Barth  
(a.barth@ulg.ac.be)

filtering them from the analysis increment (Barth et al., 2007, 2008).

Those approaches are however not appropriate for tidal simulation since the main processes here are barotropic waves. It would be difficult to distinguish between spurious gravity waves and tidal waves which are missing or misrepresented in the model solution.

As a first step for the assimilation of High-Frequency (HF) radar data, we concentrate for simplicity on the M2 tidal signal. Adding other tidal constituents would not add fundamentally any new complexity except that the model integration period would have to be substantially longer to resolve two constituents with a similar frequency as required by the Rayleigh criterion. In order to distinguish two periodic signals of frequency  $\nu_1$  and  $\nu_2$ , the length of the time series has to be at least  $|\nu_1 - \nu_2|^{-1}$  (Emery and Thomson, 1998).

The objective of this paper is twofold: it aims to apply the ensemble generation scheme providing dynamically constrained perturbations proposed in Barth et al. (2009) to a realistic data assimilation case study and to assess the realism of the assimilation results using an ensemble assimilation scheme. The paper aims also to assess the usefulness of the HF radar data and altimetry-based tidal products to estimate tidal boundary conditions using a smoother scheme for data assimilation.

In Sect. 2, the observations used in this assimilation study are presented. A brief description of the model is given in Sect. 3. The assimilation strategy is detailed in Sect. 4. The assimilation results are discussed and validated in Sect. 5. Finally, the conclusions of this study are presented in Sect. 6.

## 2 Observations

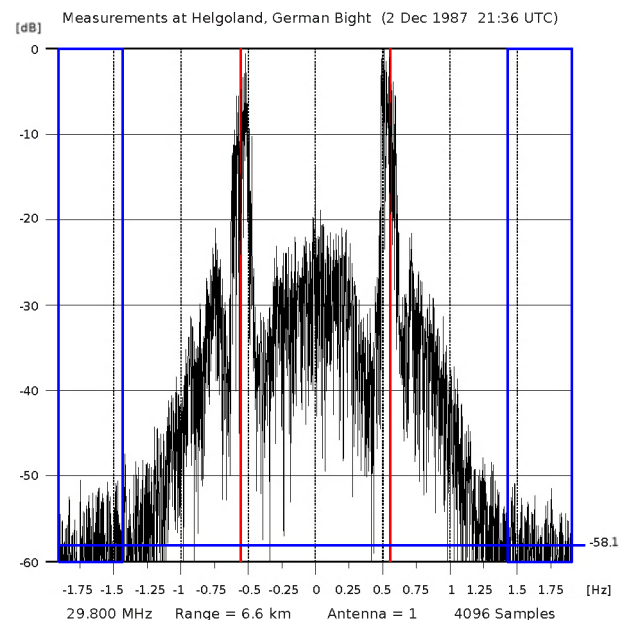
### 2.1 HF radar data

Surface current observation data were provided by the University of Hamburg by means of HF radar measurements carried out in the context of the PRISMA project (PRISMA, 1994). Two systems were installed at a distance of approx. 50 km, the first on the island of Helgoland (54.19° N, 7.88° E) and the second on the mainland coast near the town of St. Peter-Örding (54.34° N, 8.59° E). They were based on an early CODAR design developed at NOAA (Barrick et al., 1977), which had been modified at the University of Hamburg.

The operating frequency was 29.85 MHz which made the system to couple to 5.02 m long ocean waves. The radar measures the radial component of the surface current by analyzing the additional Doppler shift caused by the Bragg-resonant waves of the underlying current field. Due to the decrease in orbital motion of these waves, they couple to ocean currents down to about 0.5 m below the sea surface. As a result the radar provides surface currents averaged over the top 0.5 m of the water column.

The modified CODAR system used a transmit pulse length of 16  $\mu$ s which results in a range resolution (range cell depth) of 2.4 km. However, the range cells were sampled every 8  $\mu$ s giving interpolated data every 1.2 km. Azimuthal resolution was provided by means of a four-element array with the antennas arranged in a square at  $0.5 \lambda$  diagonal spacing (5.02 m). The angle of arrival of the backscattered signal was determined by direction-finding in the frequency domain based on the phase difference between the antennas. After internal pre-processing the sampling time of the CODAR was 0.262 s. In each data run 4096 samples were acquired, resulting in an 18 min “coherent integration time” (CIT).

Figure 1 shows a typical backscatter spectrum acquired during a previous experiment. The red vertical lines mark the Doppler frequency expected from Bragg-resonant waves without any additional shift due to the current field (Bragg frequency).



**Fig. 1.** An example of a typical backscatter spectrum acquired during a previous experiment (Helgoland, German Bight, 2 December 1987, 21:36 UTC). The vertical red lines mark the frequency shift due to the phase velocity of the Bragg-resonant ocean waves traveling towards and away from the radar. The areas marked in blue indicate the part of the spectrum the noise level is calculated from, the blue horizontal line indicates the noise level found.

#### 2.1.1 Error statistics of the current velocity

After a fast Fourier transform, the backscatter spectra from the four antennas consist of  $N_{\text{FFT}} = 4096$  lines each and the additional Doppler shift caused by the current field is analyzed. As the radial component of the current velocity varies with angle, the radial component of current field is spread

over numerous spectral lines. The direction of arrival is calculated for each spectral line with a signal-to-noise ratio  $S/N \geq 6$  dB and the respective radial current velocity  $u_r(i)$  as well as the associated  $S/N(i)$  is then sorted into “direction-boxes” at  $1^\circ$  increments.

The noise level of the backscatter spectrum is calculated as the average power of the areas marked in blue in Fig. 1, the  $S/N$  value is the ratio of the power of a signal line in the spectrum to the noise level. This calculation is done for all range cells.

In a next step, a  $27 \times 21$  Cartesian grid with 3 km horizontal resolution is defined. To transform the measurements from radar coordinates (range, azimuth) to this Cartesian grid, all data within a circle of 3 km radius around a grid point is copied from the “direction-boxes” into a “grid-box”. After this transformation, there are e.g.  $K$  samples of a radial current velocity  $u_r(i)$  with a signal-to-noise ratio  $S/N(i)$  in a “grid-box”.

From these  $K$  samples, the radial current velocity  $\bar{u}_r$  is calculated as

$$\bar{u}_r = \frac{\sum_{i=1}^K u_r(i) S/N(i)}{\sum_{i=1}^K S/N(i)} \quad (1)$$

This weighting with the  $S/N$  implements a “center of gravity” technique which gives preference to strong echoes and provides a more stable estimate of  $\bar{u}_r$ .

The variance of the radial current velocity  $\sigma_r^2$  is calculated as

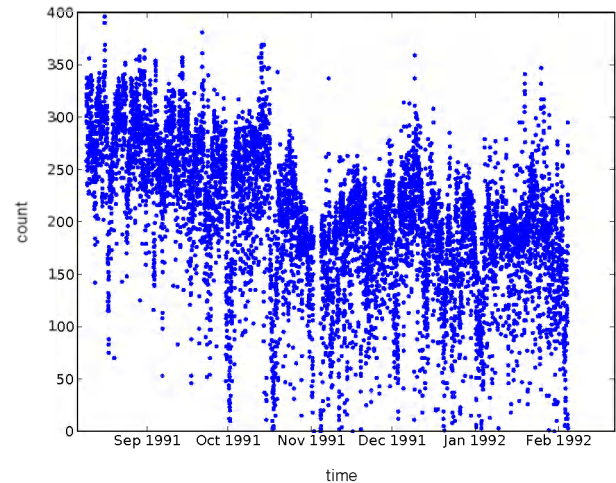
$$\sigma_r^2 = \frac{\sum_{i=1}^K u_r(i)^2 S/N(i)}{\sum_{i=1}^K S/N(i)} - \bar{u}_r^2 \quad (2)$$

Note, that the variance contains measurement errors as well as the variability due to temporal changes in the current field within the CIT caused by e.g. tides. Finally, the accuracy of the radial current velocity  $\text{Acc}_r$  is calculated as

$$\text{Acc}_r = \frac{\sigma_r}{\sqrt{K}} \quad (3)$$

Because the value of  $K$  is not very large, a correction based on the Student's  $t$  distribution is applied. The radial current velocity  $\bar{u}_r$  as well as its variance  $\sigma_r^2$  and accuracy  $\text{Acc}_r$  are calculated for all points of the Cartesian grid, as long as  $K \geq 3$ .

In a final step, the radial current velocities measured by two or more CODARs installed at different locations are combined to form the surface current vector  $\mathbf{u}$ . The algorithm is published by Gurgel (1994) and is based on a least-squares-fit. It makes use of the error statistics of the radial current velocities to be combined and thus also provides the variances  $\sigma^2(u)$  and  $\sigma^2(v)$  of the meridional ( $u$ ) and zonal ( $v$ ) components of the surface current  $\mathbf{u}$ . The algorithm is outlined in appendix B. These variances include the influence of the geometry (the angle between the radial components) similar to the GDOP (Geometrical Dilution Of Precision) known from the GPS (Global Positioning System).



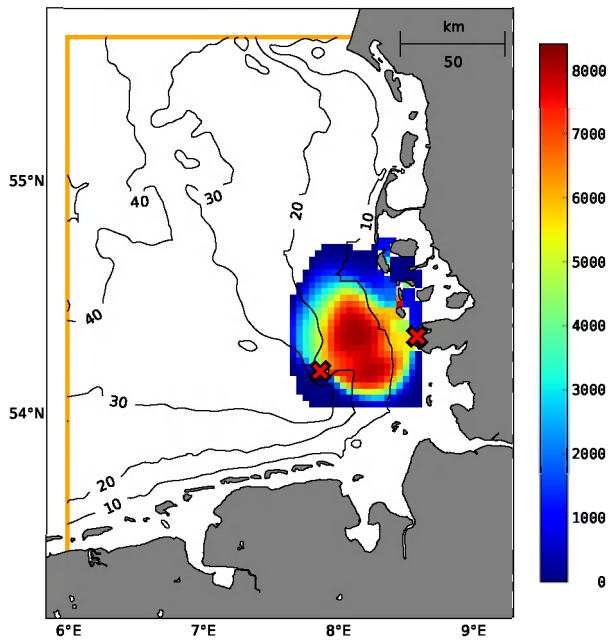
**Fig. 2.** Temporal coverage of the HF radar zonal and meridional surface velocity observations: the number of observation grid points with valid data available is shown for each sample time. The Cartesian observation grid has 567 points, but due to land coverage and distance from the antennas, the maximum spatial coverage is 396 grid cells.

Close to the line connecting the two radar sites, the GDOP becomes very large because both radial measurements basically reflect the same radial component of the current field. If the GDOP is larger than 5, the measurement with the higher  $S/N$  is used and the missing information perpendicular to the connecting line is interpolated from surrounding grid points.

### 2.1.2 The data set used for assimilation

The final data set comprises 8414 samples in time from 9 August 1991 until 4 February 1992 and has a time resolution of 30 min. Due to land coverage and distance from the antennas, the maximum spatial coverage is 396 of 567 available grid cells. The coverage shows high variation over time (see Fig. 2) with a mean spatial coverage of 202 grid cells. The mean coverage over time for each individual point on the observation grid is shown in Fig. 3. This variation is caused by a changing sea state and also due to Radio Frequency Interference (RFI) at the radar frequency. Medium wave height (1–2 m) results in the largest range of the radar measurements, while extremely calm or rough sea reduces the range. This observation is discussed in more detail in Gurgel et al. (1999).

Only observations from 1 September 1991 to 10 October 1991 are used in this work for data assimilation. Observations from 10 October 1991 to 30 October 1991 will be used for validation. The expected error variance obtained from the processing of the radar data is used in the assimilation. The error correlation between different velocity components is ignored here for simplicity. The corresponding error covariance matrix  $\mathbf{R}_{\text{HF}}''$  is thus diagonal (but its diagonal elements depend on space).



**Fig. 3.** Spatial coverage of the HF radar zonal and meridional surface velocity observations: the number of samples available at each observation grid point is color-coded according to the color-bar. The entire data set comprises 8414 samples in time. Contour lines show depth in meters as used in the hydrodynamic model's bathymetry. The crosses show the location of HF radar antennas.

The representation error accounts for missing processes in the model (but present in the observations), and errors that cannot be corrected modifying only the M2 tidal boundary conditions. It also includes the fact that the model and observation error covariances are approximations. The representation error  $S_{\text{HF}}^2$  is assumed proportional to the identity matrix and must be added to this error covariance:

$$\mathbf{R}_{\text{HF}} = \mathbf{R}'_{\text{HF}} + S_{\text{HF}}^2 \mathbf{I} \quad (4)$$

Figure 4 shows the tidal analysis of the zonal and meridional component of the velocity observations. Since close to the line joining the HF radar sites data have partly been interpolated, the tidal parameters are less accurate near this line (Fig. 5).

## 2.2 Empirical Ocean Tides (EOT08a)

The EOT08a (Savcenko and Bosch, 2008) is a global data set for amplitude and tides of the major ocean tidal constituents, four diurnal tides (K1, O1, P1, and Q1), five semi-diurnal tides (M2, S2, N2, K2, and 2N2), and one non-linear tidal constituent (M4). It is based on empirical analysis of altimeter data of multiple satellite missions and is obtained by a harmonic analysis of the residual of the altimetry data relative to the ocean tidal model FES2004 (Lettellier et al., 2004; Lyard et al., 2006). It has a spatial resolution of  $7.5' \times 7.5'$  and it is available at <ftp://ftp.dgfi.badw.de/pub/EOT08a/>

In this study only the M2 component is used (Fig. 6). The real and imaginary parts of the complex tidal parameters are represented as separate elements in the observation vector. For simplicity, the observational error covariance  $\mathbf{R}_{\text{EOT}}$  is assumed to be diagonal with a constant value  $S_{\text{EOT}}^2$  on its diagonal:

$$\mathbf{R}_{\text{EOT}} = S_{\text{EOT}}^2 \mathbf{I} \quad (5)$$

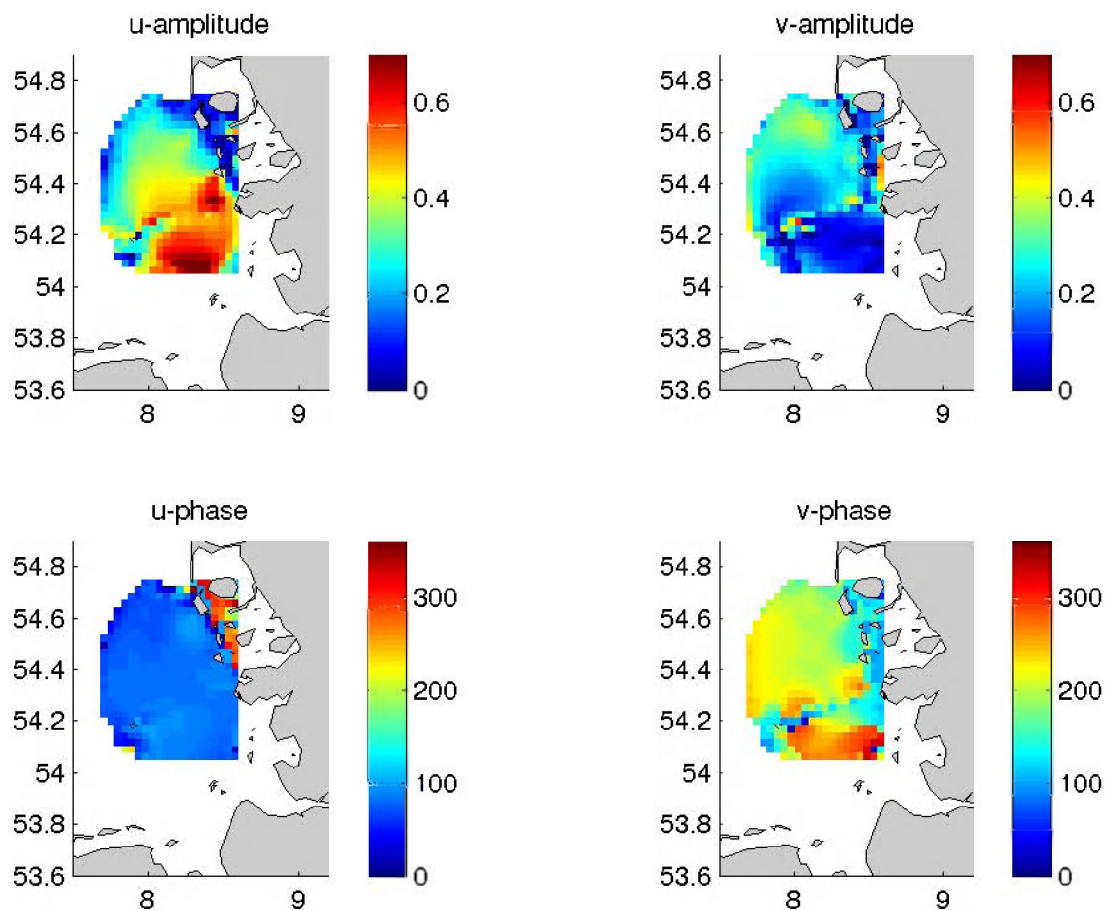
The amplitude and phase of the M2 tidal signal is a time invariant field.

## 3 Model

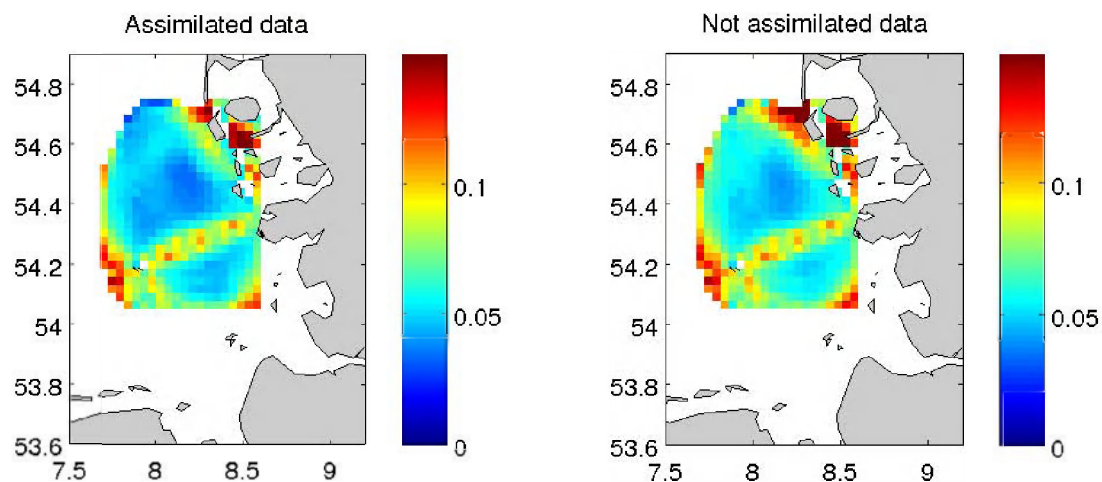
The model used is the General Estuarine Ocean Model (GETM Burchard and Bolding, 2002). It solves the 3-D primitive equations with a free-surface on an Arakawa C-grid (Arakawa and Lamb, 1977). In the vertical, the present configuration uses 21  $\sigma$  levels. It covers the German Bight with a resolution of about 0.9 km. Its boundary conditions are extracted from a 5-km resolution North Sea-Baltic Sea model (Staneva et al., 2009). The bathymetric data for the different model configurations are prepared using the ETOPO-1 topography (Amante and Eakins, 2009), together with observations made available from the German Hydrographic Service (Bundesamt für Seeschifffahrt und Hydrographie, BSH, Germany, Dick et al., 2001). The larger scale model and the nested model include tides. The sea surface elevations of the open boundary of the North Sea-Baltic Sea model are generated using tidal constituents obtained from the TOPEX/Poseidon data via the OSU Tidal Inversion Software (Egbert and Erofeeva, 2002). Atmospheric fluxes are estimated by the bulk formulation using 6-hourly ECMWF re-analysis data. The model is also forced by hourly river run-off data provided by the BSH. More details about the large-scale model and the model nesting can be found in Staneva et al. (2009).

The amplitude and phase of the M2 tidal velocity (Fig. 7) are computed for a 60-day model run starting the 1 September 1991 (without data assimilation). The tidal parameters are shown for the same region as for Fig. 4. The overall structure of the  $u$ -component amplitude (amplitude of the meridional component of the currents) is in reasonable agreement with the observations. However, a significant phase difference is observed in the  $u$ -components. For the  $v$ -component (zonal component of the currents), the velocity amplitude is similar to the observations except in the northern part where it is underestimated in the model. The phase differences are however the largest in the southeastern part of the region covered by the HF radar observations.





**Fig. 4.** M2 amplitude (in m/s) and phase (in degrees) of zonal ( $u$ ) and meridional ( $v$ ) surface currents of the observations.



**Fig. 5.** The square root of the averaged expected observational error variance (without the representation error) for the observations from 1 September 1991 to 10 October 1991 (assimilated; left panel) and from 10 October 1991 to 30 October 1991 (not assimilated; right panel). Units are m/s.

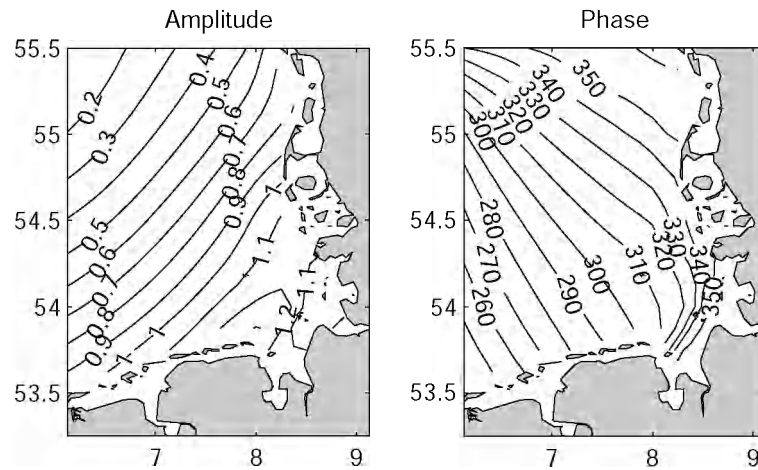


Fig. 6. M2 amplitude (in m) and phase (in degrees) of EOT08a for the German Bight

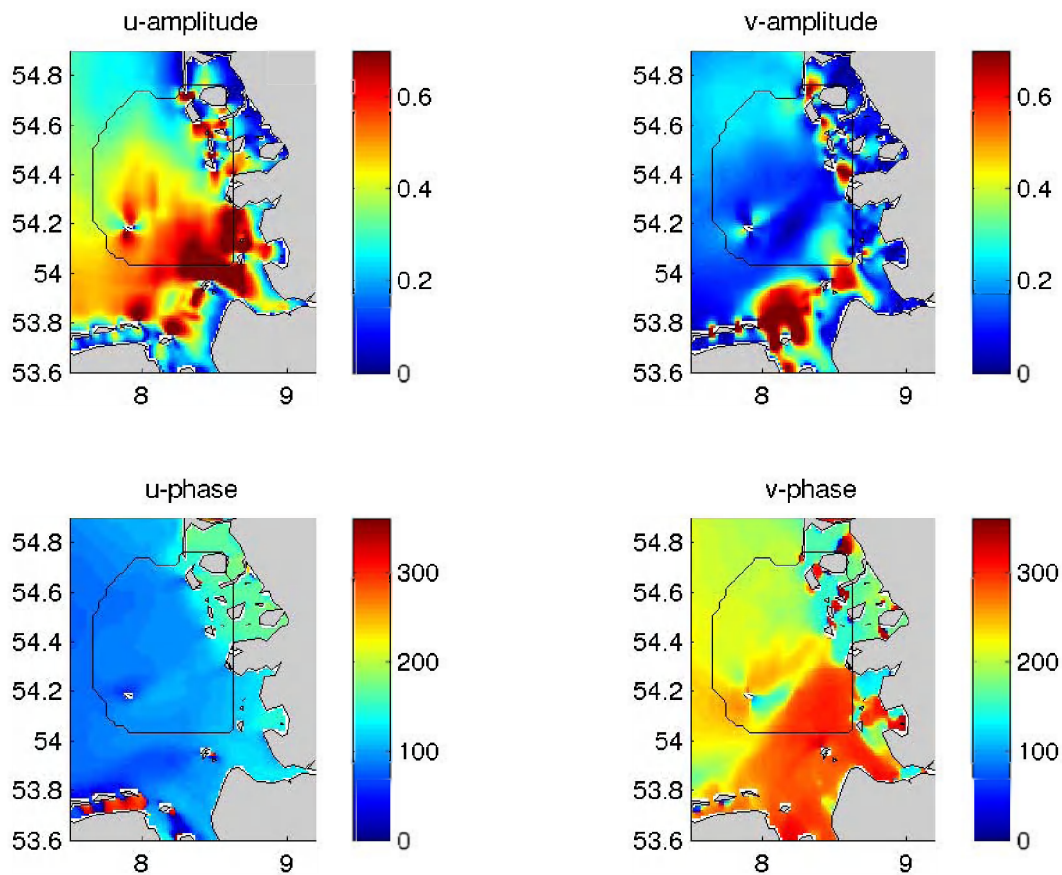


Fig. 7. M2 amplitude (in m/s) and phase (in degrees) of zonal ( $u$ ) and meridional ( $v$ ) surface currents of the free model run.

## 4 Data assimilation

### 4.1 Ensemble perturbations

The tides are governed by the shallow water equations which provide a strong dynamical link between elevation and depth-averaged currents. An ensemble of tidal boundary conditions

is created as in Barth et al. (2009). In the following we briefly present this approach. The probability of a perturbation  $\mathbf{x}$  is assumed to be Gaussian distributed:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \|\mathbf{B}\|^{1/2}} \exp(-J(\mathbf{x})) \quad (6)$$

where  $n$  is the dimension of  $\mathbf{x}$  and the function  $J$ , called a cost function in variational analysis, is a quadratic function of  $\mathbf{x}$  given by:

$$2J(\mathbf{x}) = \mathbf{x}^T \mathbf{W}_E \mathbf{x} + (\mathbf{D}\mathbf{x})^T \mathbf{W}_D (\mathbf{D}\mathbf{x}) + (\mathbf{M}\mathbf{x})^T \mathbf{W}_M (\mathbf{M}\mathbf{x}) \quad (7)$$

$$= \mathbf{x}^T \mathbf{B}^{-1} \mathbf{x} \quad (8)$$

where  $\mathbf{W}_M$ ,  $\mathbf{W}_D$  and  $\mathbf{W}_E$  are, for simplicity, diagonal weighting matrices,  $\mathbf{D}$  is a diffusion operator and  $\mathbf{M}$  is an a priori linear constraint. The subscript of  $\mathbf{W}_E$  refers to energy. The three terms in this cost function ensure respectively that the perturbations have a finite energy, are smooth and satisfy a linear constraint. The matrix  $\mathbf{B}$  is the underlying covariance matrix of the ensemble perturbations. Its inverse is also called the Hessian matrix of the cost function  $J$ .

The precise definition of the matrices  $\mathbf{W}_M$ ,  $\mathbf{W}_D$ ,  $\mathbf{W}_E$  and  $\mathbf{M}$  and of the vector  $\mathbf{x}$  is specific to a particular application. In the present setup, only the M2 tidal constituent is perturbed since it is by far the largest in the German Bight (Schirmer et al., 1994). The perturbation vector  $\mathbf{x}$  in this case is composed by the complex M2 tidal parameters of elevation  $\zeta'$ , and the depth-averaged currents  $u'$  and  $v'$  of the entire model domain:

$$\zeta(x, y, t) = e^{i\omega t} \zeta'(x, y) \quad (9)$$

$$u(x, y, t) = e^{i\omega t} u'(x, y) \quad (10)$$

$$v(x, y, t) = e^{i\omega t} v'(x, y) \quad (11)$$

where  $\omega$  is the M2 tidal frequency.

The energy constraint is based on the barotropic wave energy governed by the shallow water equations integrated over the open boundary of the model domain:

$$\mathbf{x}^T \mathbf{W}_E \mathbf{x} = \frac{\alpha}{2} \int_{\partial S} g ||\zeta'||^2 + h ||u'||^2 + h ||v'||^2 dx \quad (12)$$

where  $g$  is the acceleration due to gravity and  $h$  is the water depth. The parameter  $\alpha$  (a dimensional) determines how strong this constraint has to be enforced. In continuous form, the smoothness constraint is written as:

$$(\mathbf{D}\mathbf{x})^T \mathbf{W}_D (\mathbf{D}\mathbf{x}) = \frac{L^4}{2} \int_{\partial S} g \left\| \frac{\partial^2 \zeta'}{\partial l^2} \right\|^2 + h \left\| \frac{\partial^2 u'}{\partial l^2} \right\|^2 + h \left\| \frac{\partial^2 v'}{\partial l^2} \right\|^2 dx \quad (13)$$

$l$  is the spatial dimension tangent to the open boundary. The weighting between variables is thus the same as for the energy constraint. As before, the parameter  $L$  determines the importance of the smoothness constraint relative to other constraints. The exponent is chosen such that  $L$  represents a length-scale. The smoothness constraint is also only enforced at the open sea boundaries. The dynamical constraint is active over the whole domain and can be written as:

$$(\mathbf{M}\mathbf{x})^T \mathbf{W}_M (\mathbf{M}\mathbf{x}) = \frac{1}{2} \int_S g ||\varepsilon_\zeta||^2 + h ||\varepsilon_u||^2 + h ||\varepsilon_v||^2 dx \quad (14)$$

where  $\varepsilon_\zeta$  is the misfit of the shallow-water continuity equation and  $\varepsilon_u$  and  $\varepsilon_v$  are the misfits of the shallow-water momentum equations (with a linear bottom drag):

$$i\omega \zeta' + \frac{\partial(hu')}{\partial x} + \frac{\partial(hv')}{\partial y} = \omega \varepsilon_\zeta \quad (15)$$

$$i\omega u' - f v' + g \frac{\partial \zeta'}{\partial x} + c_u u' = \omega \varepsilon_u \quad (16)$$

$$i\omega v' + f u' + g \frac{\partial \zeta'}{\partial y} + c_v v' = \omega \varepsilon_v \quad (17)$$

The covariance matrix  $\mathbf{B}$  does not need to be formed explicitly to create ensemble perturbations, only its inverse. Since the constraints use derivatives that are approximated as finite differences, the matrix  $\mathbf{B}^{-1}$  is a sparse matrix. Only non-zero elements of this matrix are stored, which allows for an efficient implementation (Barth et al., 2009).

Unlike the approach presented in Barth et al. (2009), the energy constraints and the smoothness constraints are applied only at the boundaries. The spatial structure of the perturbations within the domain is thus given only by the dynamical constraint which is sufficient to ensure smooth perturbations.

## 4.2 Non-sequential assimilation scheme

In this sub-section, the general assimilation approach is detailed while in the next sub-section it is shown how this assimilation scheme is applied to the German Bight setup to estimate tidal boundary conditions.

A non-sequential assimilation scheme can be derived from the classical analysis scheme by embedding the time dimension in the observation vector. All observations within the model integration period are thus grouped into a single observation vector ( $\mathbf{y}^o$ ) with their corresponding error covariance (matrix  $\mathbf{R}$ ).

The vector  $\mathbf{x}^{(k)}$ , where  $k$  is the ensemble member index, includes all unknown forcings and parameters of the model. An ensemble of forcing fields are created by perturbing them within the range of their uncertainty. If the perturbations depend on time, then its time evolution is also grouped together into a single vector in a similar way as the observations.

For every perturbation, the model is integrated forward in time. But unlike classical Kalman Filters, the ensemble members are not influenced at this state by the observations. For each ensemble member, the observed part of the model state is extracted. Formally, this extraction can be expressed as a non-linear “observation operator”  $h(\cdot)$  applied to the ensemble perturbation  $\mathbf{x}^{(k)}$  performing the following operations: adding the perturbation to the background estimate, integrate the model and then interpolate the model to the location of the observations. The model dynamics are thus part of the observations operator. Every element in  $h(\mathbf{x}^{(k)})$  can thus be directly compared to its corresponding element in the observation vector  $\mathbf{y}^o$ .

In sequential data assimilation, the variables estimated by the assimilation are often the entire model state. However, different definitions can be more appropriate depending on the assimilation problem. Here we choose to include the forcing fields, which is a common strategy in variational data assimilation (e.g. Hoteit et al., 2009). Assimilation can also be used to estimate model parameters for e.g. biological models (Spitz et al., 1998) or weights given to individual models in super-ensemble techniques (Rixen et al., 2009).

The definition of the observation operator depends equally on how the assimilation strategy is implemented. It can be a simple operator interpolating the model results at the location of the observations or a more complex transformation. For example, the observation operator can be an operator extracting the spectral information from the model to assimilate tomographic data (e.g. Rémy et al., 2002), an EOF projection to assimilate EOF amplitudes (e.g. Barth et al., 2006), or a complete radiance sub-model to assimilate radiance observations in a numerical weather model (e.g. Greenwald et al., 2002).

From the ensemble simulation, we derive the following matrices whose columns represent the deviation of the ensemble members around the ensemble mean (**S**) and the observations (**E**):

$$(\mathbf{S})_k = (N-1)^{-\frac{1}{2}} (\mathbf{x}^{(k)} - \langle \mathbf{x} \rangle) \quad (18)$$

$$(\mathbf{E})_k = (N-1)^{-\frac{1}{2}} (h(\mathbf{x}^{(k)}) - \langle h(\mathbf{x}) \rangle) \quad (19)$$

where the index  $k$  refers to the ensemble member and  $\langle \cdot \rangle$  is the ensemble average. These matrices are scaled such that the following products represent covariance:

$$\mathbf{SE}^T = \text{cov}(\mathbf{x}^b, h(\mathbf{x}^b)) \quad (20)$$

$$\mathbf{EE}^T = \text{cov}(h(\mathbf{x}^b), h(\mathbf{x}^b)) \quad (21)$$

The matrices  $\mathbf{SE}^T$  and  $\mathbf{EE}^T$  would be  $\mathbf{P}^b \mathbf{H}^T$  and  $\mathbf{H} \mathbf{P}^b \mathbf{H}^T$  if the observation operator could be represented by the linear operator **H**. Those covariance matrices do not need to be formed explicitly as the analysis is performed in the subspace defined by the ensemble members (e.g. Nerger et al., 2005).

The Kalman filter analysis with non-linear observation operators as in Chen and Snyder (2007) provides the optimal perturbation minimizing its expected uncertainty:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{SE}^T (\mathbf{EE}^T + \mathbf{R})^{-1} (\mathbf{y}^o - h(\mathbf{x}^b)) \quad (22)$$

$$= \mathbf{x}^b + \mathbf{S} (\mathbf{E}^T \mathbf{R}^{-1} \mathbf{E} + \mathbf{1})^{-1} \mathbf{E}^T \mathbf{R}^{-1} (\mathbf{y}^o - h(\mathbf{x}^b)) \quad (23)$$

for a given background estimate  $\mathbf{x}^b$ . The analysis  $\mathbf{x}^a$  provides the “optimal” correction to the unknown forcings based on the observations. The model is then rerun with this perturbation applied. The observations influence thus the model solution only by choosing the optimal combination of the perturbations.

For a linear model and an infinite large ensemble, Eq. (22) minimizes,

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{b-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^o - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - h(\mathbf{x})) \quad (24)$$

$$= (\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{b-1} (\mathbf{x} - \mathbf{x}^b) + \sum_n (\mathbf{y}_n^o - h_n(\mathbf{x}_n))^T \mathbf{R}_n^{-1} (\mathbf{y}_n^o - h_n(\mathbf{x}_n)) \quad (25)$$

where  $n$  references to the indexed quantities at time  $n$ . The first term of this cost function is the constraint used to defined the dynamically plausible perturbations (Eq. 7.).

The approach used here is closely related to the Ensemble Smoother (van Leeuwen, 2001), 4DEnKF (Hunt et al., 2004, 2007) and the AEnKF (Sakov et al., 2010) where model trajectories (i.e. the model results in space and time), instead of model states, are optimized. In the smoother scheme of van Leeuwen (2001), observations are perturbed as in the standard Ensemble Kalman Filter (Burgers et al., 1998). In 4DEnKF, the observation operator is modified to relate the observations at the time where they are measured to the time where they are assimilated. The AEnKF extends the observations vector and the matrix containing the ensemble members at the location where they are observed by vertically concatenating those vectors and matrices at different time instances. This approach is also used here because it is easier to implement than the 4DEnKF. But both approaches can be seen as equivalent. For an increasing number of ensemble members, the Ensemble Smoother does also converge to the 4DEnKF and AEnKF. In the present study, these approaches are not applied to directly estimate the model trajectory but to the perturbations of the forcings (or the trajectory of the forcing perturbation). Therefore, after the optimal correction of the forcings is computed, the model needs to be rerun to obtain the final model solution. For a linear model, these schemes provide the same results. However, for a non-linear model, the results might be different. The optimal solution from the Ensemble Smoother, 4DEnKF and AEnKF is not guaranteed to satisfy the model equations, while it is per construction the case in the scheme used here. Since the method used here aims to estimate the optimal perturbations, this approach might be called Ensemble Perturbation Smoother.

### 4.3 Application to data assimilation in the German Bight

Errors in the boundary conditions, bathymetry and poorly known bottom friction contribute to errors in the M2 tidal signal inside the model domain. Mourre et al. (2004) showed in a twin experiment how the errors in the bathymetry can be reduced by data assimilation using an Ensemble Kalman Filter. The present study is a first step and focuses on reducing the error generated due to unknown tidal boundary conditions. Also the uncertainties due to errors in the initial conditions, other lateral boundary conditions and atmospheric forcings are not considered here.



The vector  $\mathbf{x}$  is composed in the present implementation by the complex M2 tidal parameter of elevation  $\zeta'$ , zonal  $u'$  and meridional velocity  $v'$  defined over the whole model domain. Those variables vary spatially but are constant in time. This vector has in total 52374 elements. In GETM version 1.6, tidal boundary conditions are implemented such that only elevation at the boundary is used for the ensemble simulation. In future studies, this vector can be augmented by other unknown parameters such as bathymetry or (space dependent) friction coefficients.

The observations vector  $\mathbf{y}^o$  includes the  $u$  and  $v$  components of the HF radar observations  $\mathbf{y}_n^{\text{HF}}$  at all time instances  $t_n$  ( $n = 1, \dots, N_t$  with  $N_t = 1869$ ) within the 40 day model integration at a resolution of 30 min (at 51 time instances no data is available). It also includes the real and imaginary parts of the EOTs elevation M2 tidal parameters  $\mathbf{y}^{\text{EOT}}$  (related to the amplitude and phase). This latter field does not depend on time:

$$\mathbf{y}^o = [\mathbf{y}_1^{\text{HF}T}, \dots, \mathbf{y}_N^{\text{HF}T}, \mathbf{y}^{\text{EOT}T}]^T \quad (26)$$

In total  $\mathbf{y}^o$  has 2 120 396 elements. The observation error covariance matrix  $\mathbf{R}$  is here a diagonal matrix whose diagonal elements are obtained by a concatenation similar to equation (26).

In order to create an ensemble  $\mathbf{x}^{(k)}$  of perturbations following the distribution (6), the Hessian matrix of the cost function ( $\mathbf{B}^{-1}$ ) is decomposed into eigenvectors and eigenvalues:

$$\mathbf{B}^{-1} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (27)$$

Only the 50 eigenvectors with the smallest eigenvalues are retained. The decision how many eigenvectors to retain was based on the distribution of the eigenvalues. Those eigenvectors correspond to the dominant error modes defined by the covariance matrix  $\mathbf{B}$ . From this eigendecomposition, an ensemble of 51 members is created following the 2nd order exact re-sampling strategy of the SEIK Filter (Pham, 2001).

$$\mathbf{x}^{(k)} = \mathbf{U} \mathbf{\Lambda}^{-1/2} \mathbf{A}(\Omega)_k \quad \text{for } k = 1, \dots, N+1 \quad (28)$$

where  $N$  is the number of eigenvectors retained,  $\mathbf{A}$  is a  $(N+1) \times N$  matrix whose column vectors form an orthonormal basis perpendicular to the column vector with all  $N+1$  elements equal to one and  $(\Omega)_k$  is the  $k$ th column of a  $N \times N$  random orthogonal matrix  $\Omega$ . This ensemble will have exactly a zero mean and its ensemble covariance is the covariance matrix  $\mathbf{B}$  reduced to its 50 largest eigenvectors.

The tidal perturbations are added to the GETM boundary conditions and GETM is run for 40 days with each of those perturbed boundary values. The minimum length should be at least one string tide/neap tide cycle. Otherwise the errors in the S2 tidal signal could lead to modification of the M2 boundary conditions. This would result in a high RMS error when the model currents are compared to the validation of surface currents.

From those 51 model runs, the surface currents at the locations measured by the HF radar are extracted and will be compared to the HF radar observations. Also for every surface model grid point, the complex M2 tidal parameters of surface elevation are computed using the T\_TIDE package (Pawlowicz et al., 2002). These complex tidal parameters are interpolated at the grid of the EOT08a data. Once the optimized boundary values following Eq. (22) are obtained, the model is rerun with the corrected boundary values for 60 days.

## 5 Results

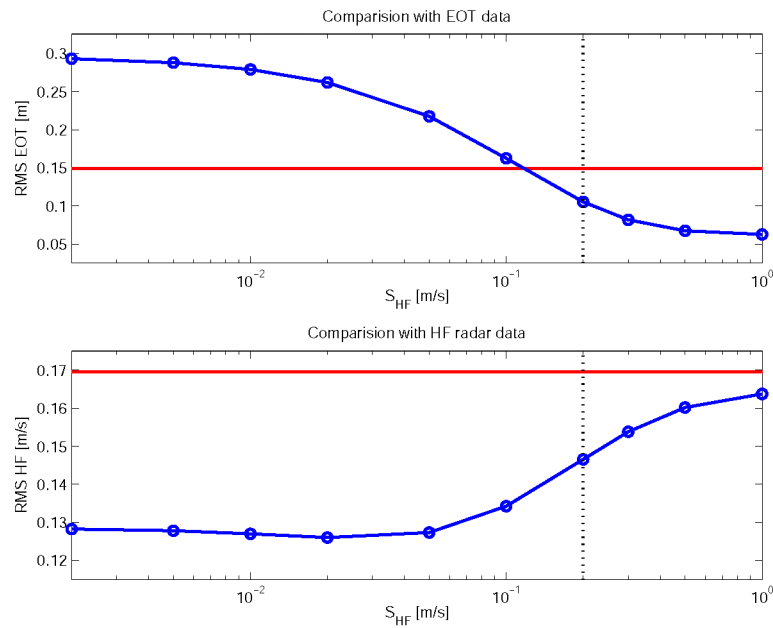
### 5.1 Sensitivity to the observational error covariance

In the present approach, the observational error covariances do not influence the ensemble simulation (unlike Ensemble Kalman Filters). Only at the analysis step, which is carried out only once, the observations are used. Different parameters of the observational error covariances can thus be tested without repeating the ensemble simulation which is by far the most CPU resource-intensive step.

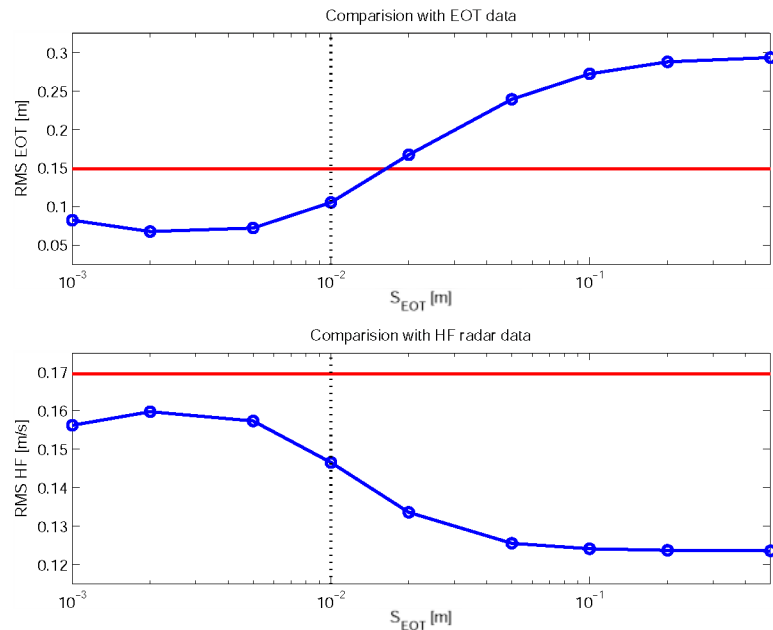
In a first series of experiments, the expected error standard deviation of the EOT data  $S_{\text{EOT}}$  is fixed at 0.01 m (the impact of this value will be discussed later) and 10 different values for  $S_{\text{HF}}$  are used ranging from 0.002 m/s to 1 m/s. Each blue circle in Fig. 8 corresponds thus to a different analysis and a subsequent 60-day run of the model. The RMS error between the EOT M2 tidal parameter for elevation and the model (according to equation in appendix A) and the RMS error between the HF radar and the model surface currents are computed. The corresponding errors of the model run with unperturbed boundary values (hereafter called the free model run) are also included in Fig. 8 for reference (red line).

If the representative error of the HF radar currents is too small, then a degradation of the model results compared to EOT is observed. The HF radar data set covers indeed only a limited portion of the model domain. The correction of the boundary values is extrapolated from the information of the HF radar. For  $S_{\text{HF}} > 0.12$  m/s, an improvement for both data sets is obtained compared to the free model run. As a compromise between improvement relative to HF radar observations and EOT analysis, we choose here a value of  $S_{\text{HF}} = 0.2$  m/s for the following experiment.

In a second series of experiments,  $S_{\text{HF}}$  is thus fixed to 0.2 m/s and  $S_{\text{EOT}}$  is varied between 0.001 m and 0.5 m. As before, the analysis is repeated for those different values and a model run is performed for 60 days starting from 1 September 1991. As long as  $S_{\text{EOT}}$  is smaller than 0.017 m, an improvement is observed compared to both data sets (Fig. 9). This justifies a posteriori the value of 0.01 m chosen previously.



**Fig. 8.** RMS error relative to EOT analysis (in m) and HF radar observations (in m/s) for  $S_{EOT} = 0.01$  m and different values for  $S_{HF}$  (in m/s). Free run is in red and run with data assimilation in blue. The dotted line shows the adopted value for  $S_{HF}$ .

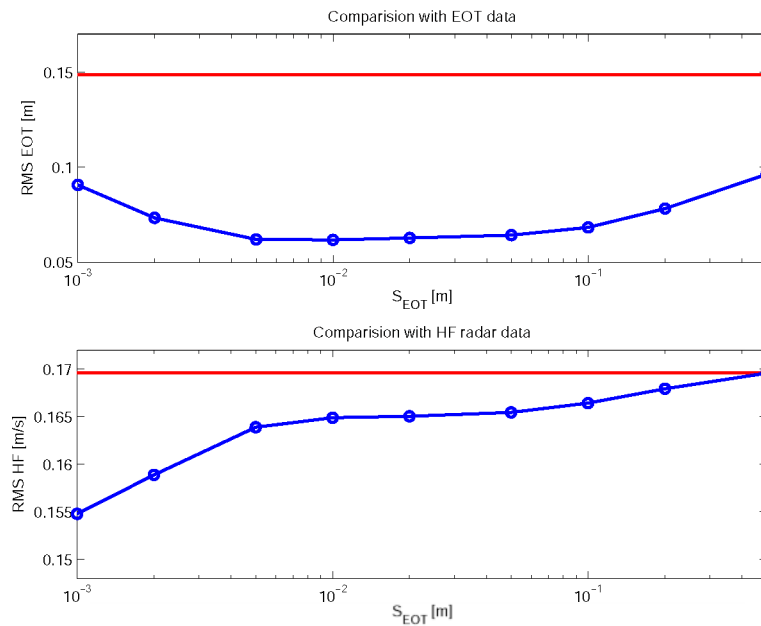


**Fig. 9.** RMS error relative to EOT analysis (in m) and HF radar observations (in m/s) for  $S_{HF} = 0.2$  m/s and different values for  $S_{EOT}$  (in m). Free run is in red and run with data assimilation in blue. The dotted line shows the adopted value for  $S_{EOT}$ .

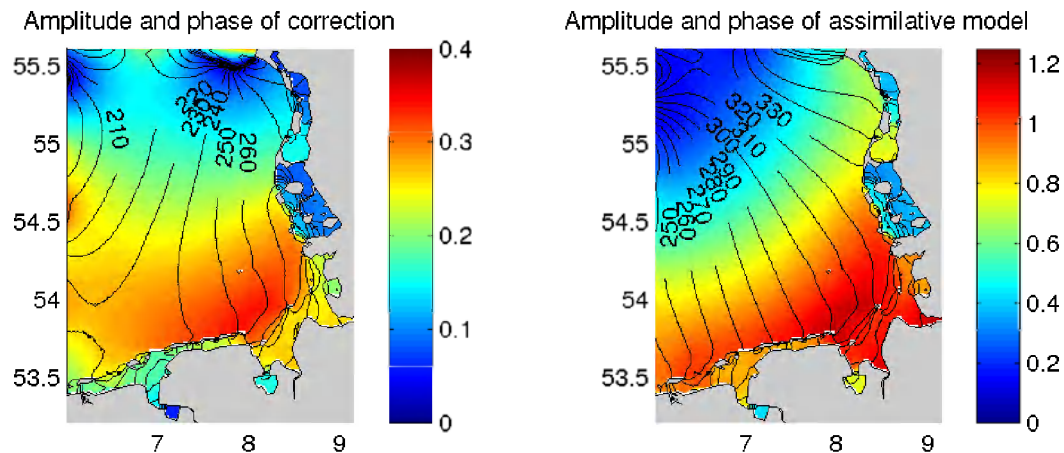
In a last series of sensitivity experiments, we assess the benefit of assimilating only EOT data, and leaving the HF radar observations as an independent validation data set (Fig. 10). In this case, we notice that for all values of  $S_{EOT}$ , the assimilation improves the results compared to the free model run for both data sets. However, as expected, the improvement relative to the HF radar data is smaller. In all

following experiment,  $S_{HF}$  and  $S_{EOT}$  are set to 0.2 m/s and 0.01 m respectively.

Figure 11 shows the elevation correction from the assimilation scheme and the elevation M2 tidal amplitude and phase diagnosed from the assimilative run. Note that the scheme provides the correction over the whole domain but only the corrections at the boundary are actually used by the GETM



**Fig. 10.** RMS error relative to EOT analysis (in m) and HF radar (in m/s) for different values for  $S_{\text{EOT}}$  (in m) without assimilation of HF radar observations. Free run is in red and run with data assimilation in blue.



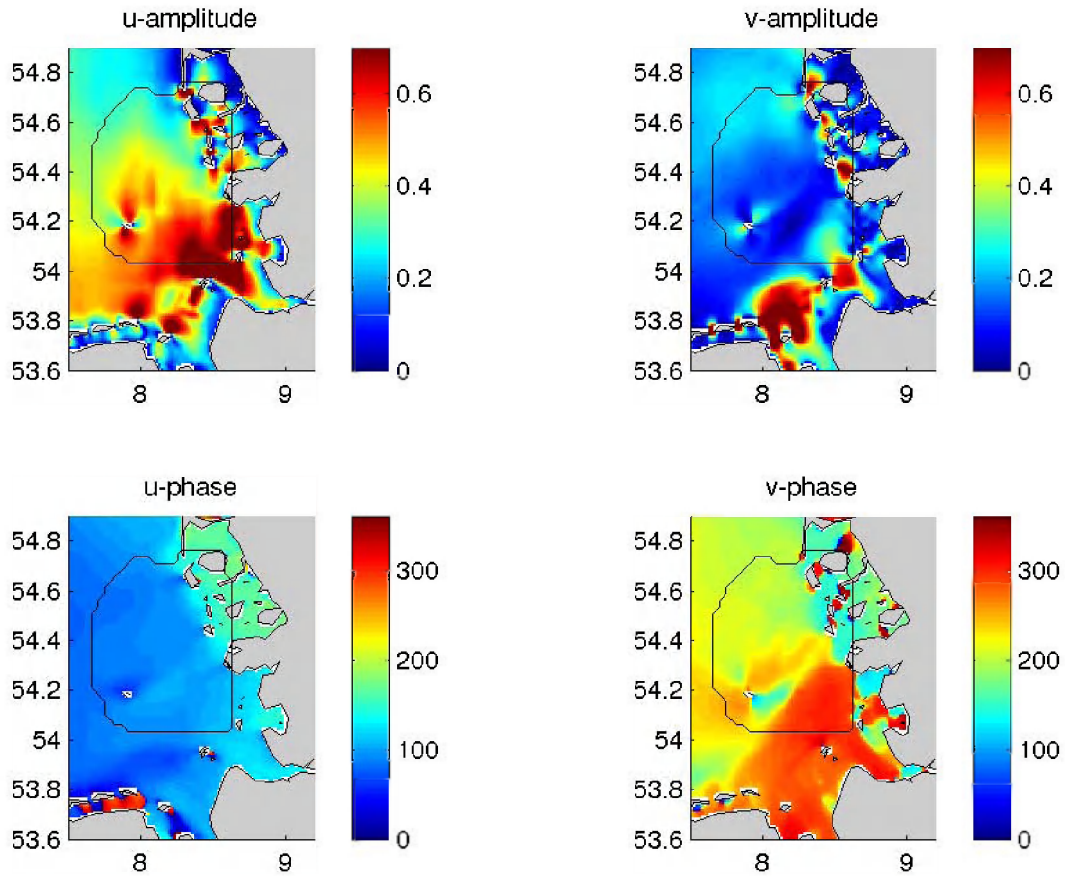
**Fig. 11.** Amplitude and phase of the correction of the M2 tidal elevation (in m, left panel) and of the assimilative model run (right panel). The increment of the phase contour lines is 10.

model. The assimilation increases the amplitude of the incoming tidal wave (in the south of the western boundary) and decreases its phase.

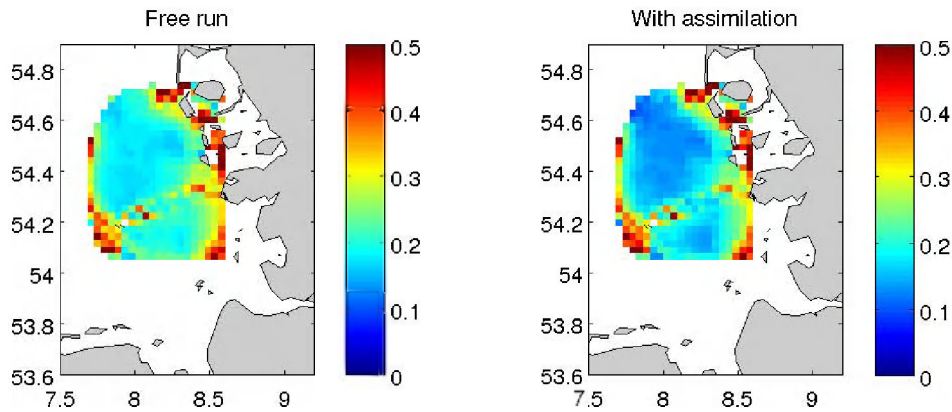
## 5.2 Error analysis

The amplitude and phase of the M2 tidal constituent are computed over the first 40 days of simulation (starting from 1 September 1991) by a tidal analysis (Fig. 12). The largest improvement obtained by the assimilation is the amelioration of the  $u$ -component phase relative to the free model. The assimilation did not damage the  $u$ -component amplitude which is close to the observations as it is the case for

the free model run. The free model run underestimates the  $v$ -component amplitude in the northern part of the zone covered by the HF radar. This is improved by the assimilation, but the model  $v$ -component amplitude is still lower than the  $v$ -component amplitude obtained from the observations. The  $v$ -component phase is in general also closer to the observations in the model run with assimilation than in the free run.



**Fig. 12.** M2 amplitude (in m/s) and phase (in degrees) of zonal ( $u$ ) and meridional ( $v$ ) surface currents of the model run with assimilation.



**Fig. 13.** RMS difference (in m/s) between surface current observations and model results without (left panel) and with assimilation (right panel).

To obtain an overall view of the impact of the assimilation, RMS maps of surface currents are computed between the observations and the free and assimilative model runs (Fig. 13).

$$\text{RMS}^2 = \frac{1}{2N} \sum_{i=1}^N (u_i^o - u_i^m)^2 + (v_i^o - v_i^m)^2 \quad (29)$$

where the subscripts  $o$  and  $m$  correspond to the observations and the model (free or with assimilation) respectively and  $N$  is the number of observations. It is important to note that the structure seen in this figure can be either due to errors in the model or in the HF radar currents. The higher errors near the line joining both antennas is probably due to the geometric dilution of precision. Since the signal strength decreases



away from the antenna a higher error is also expected near the edges of the covered region, in particular near the western and southern edge of the covered zone. In the northeastern part we have probably a combination of this effect and model errors due to the complex topography. Indeed, these zones show a higher expected error standard deviation of the observations (left panel of Fig. 5).

In the interior of the covered zone, the free model run has an RMS error of about 0.2 m/s. This value is reduced to about 0.15 m/s in the model run with assimilation. Since the RMS error is computed directly based on the surface currents, it is the combination from several errors. The RMS error includes the errors in M2 tides generated by the boundary conditions (which the assimilation aims to reduce), but among others also errors in M2 tides generated by errors in the bathymetry and bottom roughness for example, errors due to other tidal constituents, errors in wind forcings and lack of resolution in general.

In an attempt to decompose the model errors we computed the RMS error only due M2 tides ( $\text{RMS}_{\text{M2}}^2$ ) and the remaining errors sources ( $\text{RMS}^2$ ):

$$\text{RMS}^2 = \text{RMS}_{\text{M2}}^2 + \text{RMS}^2 \quad (30)$$

The contribution of the M2 tides  $\text{RMS}_{\text{M2}}^2$  is computed by a tidal analysis of the observed and model velocity. The RMS error is then computed according to appendix A. This is equivalent of computing the RMS of the error projected onto the subspace spanned by the harmonic functions at the M2 tidal frequency. Such decomposition of the model error in a orthogonal vector space is a useful approach for model validation (e.g. Alvera-Azcárate et al., 2007). Figure 14 shows the RMS difference between observations and models only due to errors in the M2 tides. As expected, the relative error reduction is largest in this analysis: in the interior of the zone covered by the HF radar, the error is reduced from 0.14 m/s to about 0.08 m/s. To reduce the model error in the very shallow eastern part, a higher resolution would probably be required.

All RMS errors in Figs. 13 and 14 are computed based on the 40 days of HF radar velocity used in the assimilation. Those error analyses were repeated for the following 20 days of model simulation. The error reduction for this analysis is very similar to the one obtained by using the assimilated observations (Fig. 15). This indicates that the boundary conditions were not adjusted for a particular month and that they can also be used for simulating a different time period. However, the RMS error at the edges of the covered zone is higher when the model is compared to the data from 10 October 1991 to 30 October 1991. The fact that this is observed in both simulations (free and with assimilation) and also in the averaged expected error standard deviation of the observations indicate that the observations were inherently less accurate in the second time period (Fig. 5).

### 5.3 Tide gauge station

The model results are also compared to tide gauge data from Helgoland (54.18° N, 7.88° W) and from Cuxhaven (53.87° N, 8.72° W). The tide gauge station at Helgoland is covered by the HF radar observations but this is not the case for the tide gauge at Cuxhaven. For both sites, hourly sea level data from 21 January 2006 to 7 December 2006 are available.

Since the model time frame differs from the time period where sea level is available, a direct comparison of the sea level is not possible. However, the tidal amplitude and phase can be compared (after applying nodal corrections). Table 1 shows the M2 amplitude and phase computed by tidal analysis from the observations, the free model run and the model run with assimilation. At Helgoland, the phase difference is reduced from 14° to −2° and also the relative amplitude error is improved from 28% to 14%. At Cuxhaven, the assimilation reduced the phase error from −41° to −28° and the relative amplitude error from 30% to 21 %. Table 1 also contains the RMS error as computed by Eq. (A2) of Appendix A. Those RMS errors combine the contribution of the amplitude and phase improvements. The assimilation reduces the RMS error by a factor of 2 for Helgoland and by a factor of 1.4 for Cuxhaven.

### 5.4 Optimizing the model trajectory instead of tidal boundary conditions

Instead of optimizing the tidal boundary conditions one could directly try to estimate the best model trajectory (van Leeuwen, 2001; Hunt et al., 2004, 2007; Sakov et al., 2010). Whether one approach is preferred over the other depends on the application. In the present case, the amplitude and phase of the M2 tidal signal is a time invariant field. If tidal amplitude and phase at the boundary are corrected, the model can be rerun for any time period. This is not the case if the model trajectory is corrected. However, in the present approach one can easily try both methods without re-computing the ensemble members.

The vector  $\tilde{\mathbf{x}}^{(k)}$  represents the model trajectory (space and time) and the observation operator  $\tilde{h}(\cdot)$  extracts the observed surface currents and elevation tidal parameters from the model trajectory. The rows of the matrix  $\tilde{\mathbf{S}}$  are defined in a similar way as previously:

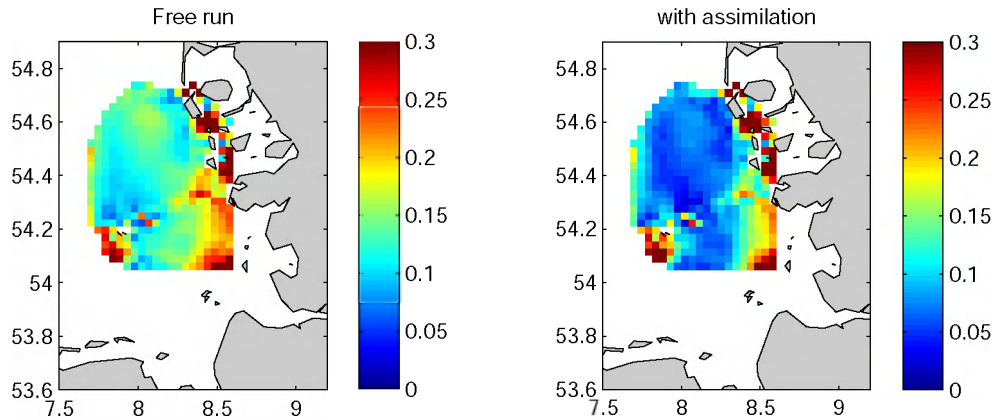
$$(\tilde{\mathbf{S}})_k = (N-1)^{-\frac{1}{2}} (\tilde{\mathbf{x}}^{(k)} - \langle \tilde{\mathbf{x}} \rangle) \quad (31)$$

The matrix containing the scaled ensemble anomalies of the observed part of the model trajectory  $\mathbf{E}$  is the same as before. According to the analysis step of the Kalman filter, the optimal trajectory  $\tilde{\mathbf{x}}^a$  is given by:

$$\tilde{\mathbf{x}}^a = \tilde{\mathbf{x}}^b + \tilde{\mathbf{S}} (\mathbf{E}^T \mathbf{R}^{-1} \mathbf{E} + \mathbf{1})^{-1} \mathbf{E}^T \mathbf{R}^{-1} (\mathbf{y}^o - \tilde{h}(\tilde{\mathbf{x}}^b)) \quad (32)$$

**Table 1.** Comparison with tide gauge observations. Amplitude and RMS are in m and phase in degrees.

	Helgoland			Cuxhaven		
	amplitude	phase	RMS	amplitude	phase	RMS
Observations	1.13	304		1.36	334	
Free run	0.81	318	0.28	0.95	15	0.63
Assimilative run	0.97	302	0.12	1.08	2	0.46

**Fig. 14.** RMS difference (in m/s) between surface current observations due to the M2 tides and the corresponding model results without (left panel) and with assimilation (right panel).

The correction to the trajectory  $\tilde{\mathbf{x}}^a$  is thus a linear combination of the rows of  $\tilde{\mathbf{S}}$ . The coefficients of this linear combination are the same as the ones obtained for optimizing the tidal boundary conditions Eq. (22) since:

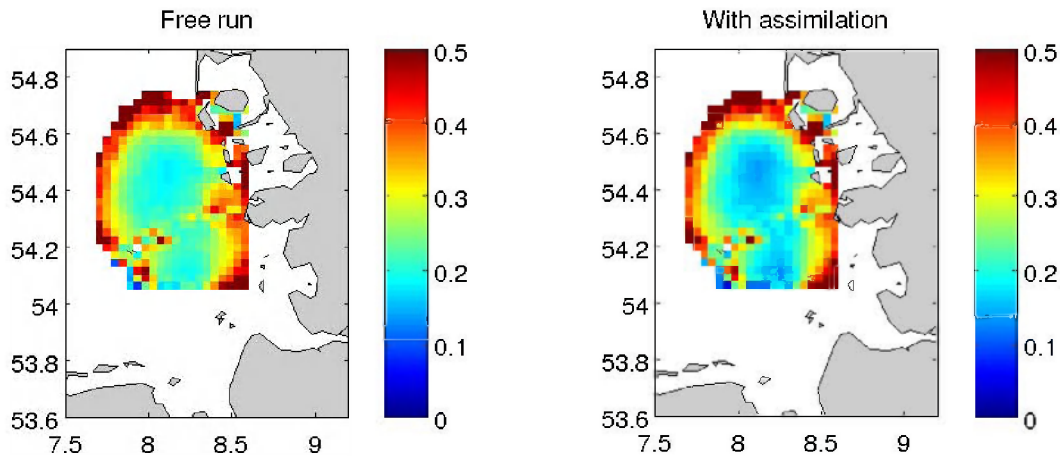
$$\tilde{h}(\tilde{\mathbf{x}}^b) = h(\mathbf{x}^b) \quad (33)$$

Therefore with almost no additional effort, one can compute the optimal trajectory  $\tilde{\mathbf{x}}^a$ . However, unlike the optimization of the boundary conditions, one cannot compute the model results obtained for a different time period and compare them to the corresponding HF radar observations. We choose thus to compare the model results to the tide gauge data to compare both approaches (as in Sect. 5.3). By correcting the model trajectory, one obtains a RMS error 0.46410 m and 0.11976 m for Cuxhaven and Helgoland respectively. These results are very similar to the RMS error obtained by optimizing the tidal boundary condition (0.46431 m for Cuxhaven and 0.11990 m for Helgoland). One notices that the correction to the model state leads to slightly better results than correcting the boundary conditions but essentially both approaches provide virtually the same results. One could expect exactly the same results if the model dynamics would be linear. The fact that the RMS errors differ only on the order of tenths of millimeters shows a posteriori that the tidal propagation is mostly linear. The advantage however for correcting directly the tidal signal is that the new tidal parameters can be used for simulating different time periods.

## 6 Conclusions

A new ensemble generation scheme to create dynamically constrained perturbations is tested in a realistic assimilative model setup. Since tidal boundary conditions are determined by an ensemble assimilation scheme, the ensemble perturbations are required to satisfy approximately the harmonic shallow water equations. An ensemble based on this method was successfully used to improve M2 boundary conditions by assimilation of HF radar and tidal data derived from altimetry.

The feasibility of a non-sequential assimilation scheme to derive the optimal correction as a combination from ensemble perturbations is shown. Since the perturbations of uncertain forcings fields (here tidal boundary conditions) are analyzed by using all observations within the model integration period, the assimilation scheme acts similar as an Kalman Smoother. For a linear model and infinite ensemble, it would provide the same solution as 4D-Var but without requiring the formulation of an adjoint. The method aims to derive the optimal perturbation and not the optimal state. Therefore, the final solution is obtained by rerunning the model. The analyzed model solution satisfies thus the model equations exactly. No spurious, transient motion such as gravity waves, are generated during the model run with this procedure as it is often the case in sequential assimilation schemes. All quantities which are conserved by a free-running model (taking the fluxes through boundaries of the domain into account) will also be conserved by the analyzed solution.



**Fig. 15.** RMS difference (in m/s) between surface current observation and model results without (left panel) and with assimilation (right panel) compared to independent data.

The applicability of this approach for different degrees of non-linearity and for uncertain time-dependent forcings remain to be shown. In the latter case, one could apply a time-localization similar to the space localization used in reduced-rank Kalman filters.

Besides the expected error reduction relative to the assimilated HF radar observations and the EOT08a data set (Empirical Ocean Tides), the improved boundary conditions also reduce the error relative to HF radar velocities set aside for validation and the M2 amplitude and phase of tide gauge observations at Helgoland and Cuxhaven.

Instead of optimizing the M2 tidal boundary conditions, one can with little effort also obtain the solution of the smoother problem where the trajectory is directly optimized (van Leeuwen, 2001; Hunt et al., 2004, 2007; Sakov et al., 2010). If the system is linear, both approaches should provide the same result. In the present case, both methods lead indeed to almost the same result indicating a posteriori that the non-linear effects in the M2 tidal propagation are indeed small. The main advantage however in optimizing the tidal parameters is that the correction can be used to simulate a different time period.

Several points can also be improved concerning the way the HF radar data are assimilated. Instead of working with the estimated  $u$  and  $v$  components of the surface currents, it would be preferable to directly assimilate the radial velocities. The  $u$  and  $v$  components of the surface currents can only be estimated where at least two radial measurements are available. Not all data can thus be used when the  $u$  and  $v$  components are assimilated.

Another future area of research is the handling of the Stokes drift which is included in the data but not in the model. For the analysis scheme it is thus seen as a “spurious” variability in the observations. The testing of various observation error variances and the comparison of the analyzed solution

to the observations provide an indirect safeguard preventing to fit the Stokes drift by adjusting the tidal boundary conditions. However, an explicit approach to remove the Stokes drift from the data (or adding it to the model) is preferable.

Only errors due to uncertainties in the M2 tidal boundary conditions are considered in the manuscript. As a next step, this procedure could be extended to other tidal constituents. This would require a longer integration time period to distinguish constituents with similar frequency. However, the tidal signal is also affected by error bathymetry and friction. To further improve the representation of tides, those parameters can also be perturbed and their uncertainties reduced by data assimilation (see Mourre et al. (2004) for an application where errors in the bathymetry are taken into account). The proposed scheme is particularly well suited to optimize tidal boundary conditions since the tidal amplitude and phase are essentially time independent. Beyond the improvement of tides, it would be interesting to study if the scheme is also applicable for time dependent fields such as wind forcings and heat flux (by assimilating for example satellite SST).

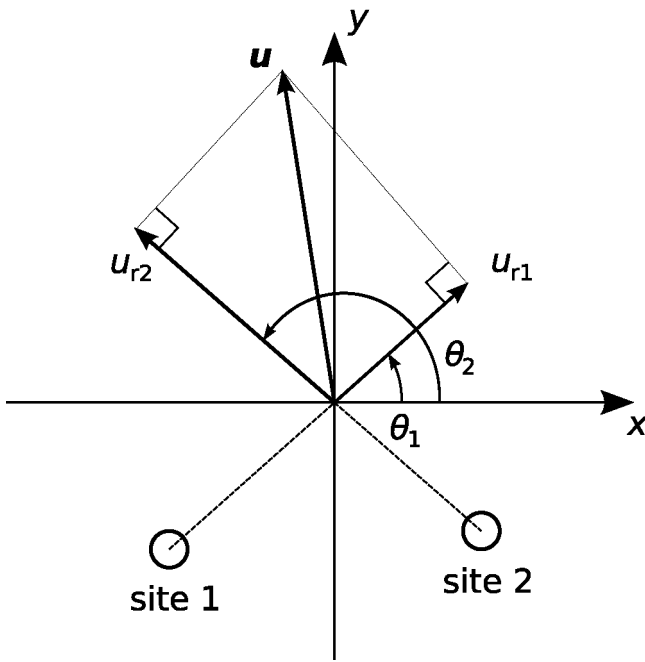
The Source code of the Ensemble Perturbation Smoother is freely available at <http://modb.oce.ulg.ac.be/mediawiki> or from the author. The assimilation code runs on Matlab and GNU Octave.

## Appendix A

### Root mean square difference between tidal signals

For two harmonic time series at the same frequency  $\omega$  characterized by an amplitude and a phase  $A, \phi$  and  $A', \phi'$  respectively, one can compute its time averaged RMS difference by:

$$\text{RMS}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (A \cos(\omega t - \phi) - A' \cos(\omega t - \phi'))^2 dt \quad (\text{A1})$$



**Fig. A1.** Calculation of the surface current  $\mathbf{u}$  from two radial components  $u_{r1}$  and  $u_{r2}$ .

$$= \frac{A^2 + A'^2}{2} - AA' \cos(\phi - \phi') \quad (\text{A2})$$

This measure combines a difference in amplitude and phase and is thus useful to assess the overall difference between the two periodic time series.

## Appendix B

### Calculating the surface current vector from two or more radial components

The geometry to combine two radial components  $u_{r1}$  and  $u_{r2}$  is shown in Fig. A1. The meridional ( $u$ ) and zonal ( $v$ ) components of the surface current  $\mathbf{u}$  can be calculated by solving the following equation:

$$\begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ \vdots & \vdots \\ \cos\theta_n & \sin\theta_n \end{pmatrix} \underbrace{\begin{pmatrix} u \\ v \end{pmatrix}}_{\mathbf{u}} = \begin{pmatrix} u_{r1} \\ \vdots \\ u_{rn} \end{pmatrix} \quad (\text{B1})$$

For more than two radial components, the system is overdetermined. In this case, the system can be solved by minimizing the total error  $\varepsilon^2$  using all  $n$  radial components  $u_{rj}$  ( $j = 1, \dots, n$ ) and their error variance,  $\sigma_{rj}^2$ :

$$\varepsilon^2 = \sum_{j=1}^n \frac{1}{\sigma_{rj}^2} (\cos\theta_j u + \sin\theta_j v - u_{rj})^2 \quad (\text{B2})$$

In matrix form, this cost function can be written as:

$$\varepsilon^2 = |\mathbf{A}\mathbf{u} - \mathbf{b}|^2 \quad (\text{B3})$$

where vector  $\mathbf{b}$  contains all radial velocities scaled by their error standard deviation and the matrix  $\mathbf{A}$  are given by:

$$\mathbf{A} = \begin{pmatrix} \frac{\cos\theta_1}{\sigma_{r1}} & \frac{\sin\theta_1}{\sigma_{r1}} \\ \vdots & \vdots \\ \frac{\cos\theta_n}{\sigma_{rn}} & \frac{\sin\theta_n}{\sigma_{rn}} \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \frac{u_{r1}}{\sigma_{r1}} \\ \vdots \\ \frac{u_{rn}}{\sigma_{rn}} \end{pmatrix} \quad (\text{B4})$$

The solution  $\mathbf{u}$  is found by solving the following equation:

$$(\mathbf{A}^T \mathbf{A}) \mathbf{u} = \mathbf{A}^T \mathbf{b} \quad (\text{B5})$$

The variances of the meridional  $\sigma^2(v)$  and zonal  $\sigma^2(u)$  components of the surface current is given by the diagonal elements of the covariance matrix  $\mathbf{C}$ :

$$\mathbf{C} = (\mathbf{A}^T \mathbf{A})^{-1} \quad (\text{B6})$$

$$\sigma^2(u) = c_{11}; \quad \sigma^2(v) = c_{22} \quad (\text{B7})$$

This algorithm can also be used to combine two radial components only. In contrast to just solving equation (B1), it also provides the variances of the surface current  $\mathbf{u}$ .

**Acknowledgements.** The DGF (Deutsches Geodätisches Forschungsinstitut) is acknowledged for providing the EOT08a data. Johannes Schulz-Stellenfleth from GKSS is thanked for his help with the HF radar data and helpful discussions. This work was supported by the ECOOP project (European COastal-shelf sea Operational observing and forecasting system) of the European Union. The National Fund for Scientific Research, Belgium is acknowledged for funding the post-doctoral positions of Alexander Barth and Aida Alvera-Azcárate. We would like also to thank Michel Rixen, François Counillon and a third anonymous reviewer for their careful reading of this manuscript and constructive criticism. This is a MARE publication.

## References

- Alvera-Azcárate, A., Barth, A., Bouallègue, Z. B., Rixen, M., and Beckers, J.-M.: Forecast Verification of a 3D model of the Ligurian Sea. The use of Discrete Wavelet Transforms in the skill assessment of spatial forecasts, *J. Marine Syst.*, 65, 460–483, doi: 10.1016/j.jmarsys.2005.09.015, <http://hdl.handle.net/2268/4264>, 2007.
- Amante, C. and Eakins, B. W.: ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis, Tech. rep., NOAA Technical Memorandum NESDIS NGDC-24, 19 pp., 2009.
- Arakawa, A. and Lamb, V.: Computational design of the basic dynamical process of the UCLA general circulation model, *Methods in Computational Physics*, Academic Press, New York, 173–265, 1977.
- Barrick, D., Evans, M., and Weber, B.: Ocean surface currents mapped by radar, *Science*, 198, 138–144, 1977.
- Barth, A., Alvera-Azcárate, A., Beckers, J.-M., and Rixen, M.: Coupling a two-way nested primitive equation model and a statistical SST predictor of the Ligurian Sea via data assimilation, *Ocean Modelling*, 13, 255–270, doi:10.1016/j.ocemod.2006.02.003, <http://hdl.handle.net/2268/4306>, 2006.



- Barth, A., Beckers, J.-M., Alvera-Azcárate, A., and Weisberg, R. H.: Filtering inertia-gravity waves from the initial conditions of the linear shallow water equations, *Ocean Modelling*, 19, 204–218, doi:10.1016/j.ocemod.2007.06.007, <http://hdl.handle.net/2268/4266>, 2007.
- Barth, A., Alvera-Azcárate, A., and Weisberg, R. H.: Assimilation of high-frequency radar currents in a nested model of the West Florida Shelf, *J. Geophys. Res.*, 113, C08033, doi:10.1029/2007JC004585, <http://hdl.handle.net/2268/26171>, 2008.
- Barth, A., Alvera-Azcárate, A., Beckers, J.-M., Weisberg, R. H., Vandenbulcke, L., Lenartz, F., and Rixen, M.: Dynamically constrained ensemble perturbations - application to tides on the West Florida Shelf, *Ocean Sci.*, 5, 259–270, 2009, <http://www.ocean-sci.net/5/259/2009/>.
- Bloom, S. C., Takacs, L. L., Silva, A. M. D., and Ledvina, D.: Data assimilation using incremental analysis updates, *Mon. Weather Rev.*, 124, 1256–1271, 1996.
- Burchard, H. and Bolding, K.: GETM – a general estuarine transport model. Scientific Documentation, Tech. Rep. EUR 20253 EN, European Commission, 2002.
- Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, 126, 1719–1724, 1998.
- Chen, Y. and Snyder, C.: Assimilating Vortex Position with an Ensemble Kalman Filter, *Mon. Weather Rev.*, 135, 1828–1845, 2007.
- Dick, S., Eckard, K., Müller-Navarra, S. H., Klein, H., and Komo, H.: The operational circulation model of BSH (BSHcmod) - model description and validation, Tech. Rep. 29, Berichte des Bundesamtes für Seeschifffahrt und Hydrographie (BSH), 49 pp., 2001.
- Dobricic, S., Pinardi, N., Adani, M., Tonani, M., Fratianni, C., Bonazzi, A., and Fernandez, V.: Daily oceanographic analyses by Mediterranean Forecasting System at the basin scale, *Ocean Sci.*, 3, 149–157, 2007, <http://www.ocean-sci.net/3/149/2007/>.
- Egbert, G. and Erofeeva, S.: Efficient inverse modeling of barotropic ocean tides, *J. Atmos. Ocean. Tech.*, 19, 183–204, 2002.
- Emery, W. J. and Thomson, R. E.: *Data Analysis Methods in Physical Oceanography*, Pergamon, Elsevier Science Inc., New York, 634 pp., 1998.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- Evensen, G.: Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynam.*, 54, 539–560, 2004.
- Greenwald, T. J., Hertenstein, R., and Vukićević, T.: An all-weather observational operator for radiance data assimilation with mesoscale forecast models, *Mon. Weather Rev.*, 130, 1882–1897, 2002.
- Gurgel, K.-W.: Shipborne measurement of surface current fields by HF radar (extended version), *L'Onde Electrique*, 74, 54–59, 1994.
- Gurgel, K.-W., Essen, H.-H., and Kingsley, S. P.: HF radars: Physical limitations and recent developments, *Coastal Engineering*, 37, 201–218, doi:10.1016/S0378-3839(99)00026-5, 1999.
- Hoteit, I., Cornuelle, B. D., Kim, S. Y., Forget, G., Köhl, A., and Terrill, E. J.: Assessing 4D-VAR for dynamical mapping of coastal high-frequency radar in San Diego, *Dynamics of Atmosphere and Oceans*, 48, 175–197, doi:10.1016/j.dynatmoce.2009.2009.
- Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., Sauer, T., Szunyogh, I., Yorke, J. A., and Zimin, A. V.: Four-dimensional ensemble Kalman filtering, *Tellus*, 56A, 273–277, 2004.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D*, 230, 112–126, 2007.
- Keppenne, C. L., Rienecker, M. M., Kurkowski, N. P., and Adamec, D. A.: Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction, *Nonlinear Proc. Geoph.*, 12, 491–503, 2005.
- Lermusiaux, P. F. J. and Robinson, A. R.: Data assimilation via error subspace statistical estimation, Part II, Middle Atlantic Bight shelfbreak front simulations and ESSE validation., *Mon. Weather Rev.*, 127, 1408–1432, 1999.
- Lettellier, T., Lyard, F., and Lefebvre, F.: The new global tidal solution: FES2004, in: *OceanSurface Topography Science Team Meeting*, St. Petersburg, Florida, 2004.
- Lyard, F., Lefebvre, F., Letellier, T., and Francis, O.: Modelling the global ocean tides: modern insights from FES2004, *Ocean Dynam.*, 56, 394–415, doi:10.1007/s10236-006-0086-x, 2006.
- Malanotte-Rizzoli, P., Young, R. E., and Haidvogel, D. B.: Initialization and data assimilation experiments with a primitive equation model, *Dynam. Atmos. Oceans*, 13, 349–378, 1989.
- Mourre, B., De Mey, P., Lyard, F., and Le Provost, C.: Assimilation of sea level data over continental shelves: an ensemble method for the exploration of model errors due to uncertainties in bathymetry, *Dynam. Atmos. Oceans*, 38, 93–121, doi:10.1016/j.dynatmoce.2004.09.001, 2004.
- Nerger, L., Hiller, W., and Schrter, J.: A Comparison of Error Subspace Kalman Filters, *Dynamic meteorology and oceanography*, *Tellus*, 57A, 715–735, doi:10.1111/j.1600-0870.2005.00141.x, 2005.
- Ourmières, Y., Brankart, J.-M., Berline, L., Brasseur, P., and Verron, J.: Incremental analysis update implementation into a sequential ocean data assimilation system, *J. Atmos. Ocean. Tech.*, 23, 1729–1744, 2006.
- Pawlowicz, R., Beardsley, B., and Lentz, S.: Classical Tidal Harmonic Analysis Including Error Estimates in MATLAB using T\_TIDE, *Comput. Geosci.*, 28, 929–937, 2002.
- Pham, D. T.: Stochastic methods for sequential data assimilation in strongly nonlinear systems, *Mon. Weather Rev.*, 129, 1194–1207, 2001.
- PRISMA: Prozesse im Schadstoffkreislauf Meer-Atmosphäre: Ökosystem Deutsche Bucht, BMFT-Projekt 03F0558A1 (1.1.1990–31.10.1993), Abschlussbericht, ZMK-Universität Hamburg, 1994.
- Rixen, M., Gac, J.-C. L., Hermand, J.-P., Peggion, G., and Coelho, E.: Super-ensemble forecasts and resulting acoustic sensitivities in shallow waters, *J. Marine Syst.*, 78, S290–S305, doi:10.1016/j.jmarsys.2009.01.013, 2009.
- Rémy, E., Gaillard, F., and Verron, J.: Variational assimilation of ocean tomographic data: twin experiments in a quasi-geostrophic model, *Quarterly Journal of the Royal Meteorological Society*, 128, 1739–1758, 2002.
- Sakov, P., Evensen, G., and Bertino, L.: Asynchronous data assimilation

- lation with the EnKF, *Tellus*, 62A, 24–29, 2010.
- Savcenko, R. and Bosch, W.: EOT08a – empirical ocean tide model from multi-mission satellite altimetry, Tech. rep., Deutsches Geodätisches Forschungsinstitut, <ftp://ftp.dgfi.badw.de/pub/EOT08a/doc/EOT08a.pdf>, 2008.
- Schirmer, F., Essen, H.-H., Gurgel, K.-W., Schlick, T., and Hessner, K.: Circulation and contaminant fluxes in the North Sea, chap. Local variability of surface currents based on HF-radar measurements, Springer Verlag, 271–289, 1994.
- Spitz, Y. H., Moisan, J. R., Abbott, M. R., and Richman, J. G.: Data assimilation and a pelagic ecosystem model: parameterization using time series observations, *J. Marine Syst.*, 16, 51–68, 1998.
- Staneva, J., Stanev, E. V., Wolff, J.-O., Badewien, T. H., Reuter, R., Flemming, B., Bartholomä, A., and Bolding, K.: Hydrodynamics and sediment dynamics in the German Bight, A focus on observations and numerical modelling in the East Frisian Wadden Sea, *Cont. Shelf Res.*, 29, 302–319, 2009.
- Talagrand, O.: On the Damping of High-Frequency Motions in Four-Dimensional Assimilation of Meteorological Data, *J. Atmos. Sci.*, 1571–1547, 1972.
- van Leeuwen, P. J.: An Ensemble Smoother with Error Estimates, *Mon. Weather Rev.*, 129, 709–728, 2001.