Ocean Sci. Discuss., 6, 2423–2460, 2009
www.ocean-sci-discuss.net/6/2423/2009/

Ocean Science Discussions

Papers published in *Ocean Science Discussions* are under
open-access review for the journal *Ocean Science*

Ensemble smoother for optimizing tidal boundary conditions by assimilation of high-frequency radar surface currents – application to the German Bight

A. Barth[1,2], A. Alvera-Azcárate[1,2], K.-W. Gurgel[3], J. Staneva[4], A. Port[5], J.-M. Beckers[1,2], and E. Stanev[4]

[1]GeoHydrodynamics and Environment Research (GHER), MARE, Allée du 6-Août 17, University of Liège, Sart-Tilman B5, 4000 Liège, Belgium
[2]National Fund for Scientific Research, Belgium
[3]Institute of Oceanography, University of Hamburg, Germany
[4]GKSS Research Center, Geesthacht, Germany
[5]Institute for Chemistry and Biology of the Marine Environment (ICBM), University of Oldenburg, Germany

Received: 7 October 2009 – Accepted: 13 October 2009 – Published: 26 October 2009

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

High-Frequency (HF) radars measure the ocean 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 perturbation. This cost function ensures that the 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-dimension General Estuarine Ocean Model (GETM). Optimized boundary values are obtained using all observations within the assimilation period 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 (Singular Evolutive Interpolated Kalman, Pham, 2001) are sequential: the ensemble is updated using observations at the time they are measured.

OSD

6, 2423–2460, 2009

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). 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 haven 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 which reduces the generation of spurious gravity waves. It has been used with large-scale ocean models (e.g. Keppenne 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 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 missing or misrepresented in the model solution. As a first step for the assimilation HF radar data, we concentrate on the M2 tidal signal.

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 case of data assimilation 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

OSD
6, 2423–2460, 2009

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 was 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-Ording (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 m long ocean waves and thus provide surface currents averaged over the top 0.5 m of the sea surface. The radar measures the radial component of the surface current by analyzing the additional Doppler shift caused to the Bragg-resonant waves by the underlying current field.

The modified CODAR system used a transmit pulse length of 16 μs which results in a range resolution (range cell depth) of 2.4 km. However, the range cells were sampled every 8 μs 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 0.5λ diagonal spacing (5.02 m). The angle of arrival of the backscattered signal was determined by direction–finding in 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).

OSD
6, 2423–2460, 2009

2.1.1 Error statistics of the current velocity

After a fast Fourier transform, the backscatter spectra from the four antennas consist of N_{FFT}=4096 lines each and the additional Doppler shift due to 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 SNR≥6 dB and the respective radial current velocity $u_r(i)$ as well as the associated SNR(i) is then sorted into "direction–boxes" at 1° increments. This calculation is done for all range cells.

In a next step, a 27×21 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 SNR(i) in a "grid–box". From these K samples, the radial current velocity $\overline{u_r}$ is calculated as

$$\overline{u_r} = \frac{\sum_{i=1}^{K} u_r(i)\mathrm{SNR}(i)}{\sum_{i=1}^{K} \mathrm{SNR}(i)} \tag{1}$$

The variance of the radial current velocity σ_r^2 is calculated as

$$\sigma_r^2 = \frac{\sum_{i=1}^{K} u_r(i)^2\mathrm{SNR}(i)}{\sum_{i=1}^{K} \mathrm{SNR}(i)} - \overline{u_r}^2 \tag{2}$$

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 Acc_r is calculated as

$$\mathrm{Acc}_r = \frac{\sigma_r}{\sqrt{K}} \tag{3}$$

OSD
6, 2423–2460, 2009

Because the value of K is not very large, a correction based on the Student's t distribution is applied. The radial current velocity $\overline{u_r}$ as well as it's variance σ_r^2 and accuracy 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 $\boldsymbol{u}$. 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 $\boldsymbol{u}$. 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).

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 SNR 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. 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. 1) 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. 2. 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.

Only the 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

OSD
6, 2423–2460, 2009

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 $\mathbf{R}''_{\mathrm{HF}}$ is thus diagonal (but its diagonal elements depend on space). The representativity 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 representativity error S^2_{HF} is assumed proportional to the identity matrix and must be added to this error covariance:

$$\mathbf{R}_{\mathrm{HF}} = \mathbf{R}'_{\mathrm{HF}} + S^2_{\mathrm{HF}}\,\mathbf{I} \tag{4}$$

Figure 3 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. 10).

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. 4). 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 $\mathbf{R}_{\mathrm{EOT}}$ is assumed to be diagonal with a constant value S^2_{EOT} on its diagonal:

$$\mathbf{R}_{\mathrm{EOT}} = S^2_{\mathrm{EOT}}\,\mathbf{I} \tag{5}$$

3 Model

The model used is the General Estuarine Ocean Model (GETM Burchard and Bolding, 2002). It solves the 3-D primitive equations on an Arakawa C-grid (Arakawa and Lamb, 1977). In the vertical, the present configuration uses 21σ 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). Details about the large-scale model and the model nesting can be found in Staneva et al. (2009). 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 (Bundesamt für Seeschifffahrt und Hydrographie, Germany) operational model.

The amplitude and phase of the M2 tidal velocity (Fig. 5) 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. 3. The overall structure of u-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 however are largest in the southeastern part of the region covered by the HF radar observations.

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

OSD
6, 2423–2460, 2009

approach. The probability of a perturbation $\boldsymbol{x}$ is assumed to be Gaussian distributed:

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

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

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

$$= \boldsymbol{x}^T \mathbf{B}^{-1} \boldsymbol{x} \tag{8}$$

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, but it does not need to be formed explicitly. 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 $\boldsymbol{x}$ 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 $\boldsymbol{x}$ in this case is composed by the complex M2 tidal parameters of elevation ζ', and the depth-averaged currents u' and v' of the entire model domain:

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

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

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

where ω is the M2 tidal frequency. 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.

OSD
6, 2423–2460, 2009

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:

$$\boldsymbol{x}^T \mathbf{W}_E \boldsymbol{x} = \frac{\alpha}{2} \int_{\partial S} g\|\zeta'\|^2 + h\|u'\|^2 + h\|v'\|^2 \mathrm{d}x \tag{12}$$

where g is the acceleration due to gravity and h is the water depth. The parameter α determines how strong this constraint has to be enforced. In continuous form, the smoothness constraint is written as:

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

t 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 boundary. The dynamical constraint is active over the whole domain and can be written as:

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

where ε_ζ, ε_u and ε_v are given by:

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

$$i\omega u' - fv' + g\frac{\partial \zeta'}{\partial x} + c_u u' = \omega\varepsilon_u \tag{16}$$

$$i\omega v' + fu' + g\frac{\partial \zeta'}{\partial y} + c_v v' = \omega\varepsilon_v \tag{17}$$

Unlike the approach presented in Barth et al. (2009), the energy constraints and the smoothness constraints are applied only at the boundary. 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 ($\boldsymbol{y}^o$) with their corresponding error covariance (matrix $\mathbf{R}$).

The vector $\boldsymbol{x}^{(k)}$, 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 an "observation operator" $\mathbf{H}$ applied to the ensemble perturbation $\boldsymbol{x}^{(k)}$ 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 $\mathbf{H}\boldsymbol{x}^{(k)}$ can thus be directly compared to its corresponding element in the observation vector $\boldsymbol{y}^o$.

The classical analysis equation provides the optimal perturbation minimizing its expected uncertainty:

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}(\boldsymbol{y}^o - \mathbf{H}\boldsymbol{x}^b) \quad (18)$$

OSD
6, 2423–2460, 2009

for a given background estimate $\boldsymbol{x}^b$ and a given background error covariance $\mathbf{P}^b$. The analysis $\boldsymbol{x}^a$ 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 in choosing the optimal combination of the perturbations. All covariances in Eq. (18) are estimated from the ensemble:

$$\mathbf{P}^b\mathbf{H}^T = \left\langle (\boldsymbol{x} - \langle \boldsymbol{x} \rangle)(\mathbf{H}\boldsymbol{x} - \langle \mathbf{H}\boldsymbol{x} \rangle)^T \right\rangle \tag{19}$$

$$\mathbf{H}\mathbf{P}^b\mathbf{H}^T = \left\langle (\mathbf{H}\boldsymbol{x} - \langle \mathbf{H}\boldsymbol{x} \rangle)(\mathbf{H}\boldsymbol{x} - \langle \mathbf{H}\boldsymbol{x} \rangle)^T \right\rangle \tag{20}$$

where $\langle \cdot \rangle$ is the ensemble average. But 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).

For a linear model and an infinite large ensemble, Eq. (18) minimizes,

$$J(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}^b)^T {\mathbf{P}^b}^{-1} (\boldsymbol{x} - \boldsymbol{x}^b) + (\boldsymbol{y}^o - \mathbf{H}\boldsymbol{x})^T \mathbf{R}^{-1} (\boldsymbol{y}^o - \mathbf{H}\boldsymbol{x}) \tag{21}$$

$$= (\boldsymbol{x} - \boldsymbol{x}^b)^T {\mathbf{P}^b}^{-1} (\boldsymbol{x} - \boldsymbol{x}^b) + \sum_n (\boldsymbol{y}_n^o - (\mathbf{H}\boldsymbol{x})_n)^T {\mathbf{R}_n}^{-1} (\boldsymbol{y}_n^o - (\mathbf{H}\boldsymbol{x})_n) \tag{22}$$

where n references to the indexed quantities at time n. This is the cost function from which 4-D-Var and Kalman Smoother can be derived.

A similar approach is used in the Ensemble Smoother (van Leeuwen, 2001) where model trajectories, instead of model states, are optimized. For the method presented here, the observations do not need to be perturbed as in the standard Ensemble Kalman Filter (Burgers et al., 1998) and Ensemble Smoother. Ensemble Smoother is generally implemented such that the vector $\boldsymbol{x}$ represents a model trajectory. Here the vector estimated by assimilation procedure are the perturbed forcings. The optimal model trajectory is then obtained by rerunning the model with the updated forcings. For a linear model, both approaches provide the same results. However, for a non-linear model both approaches can provide different results. The optimal solution from the Ensemble Smoother 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

The Hessian matrix of the cost function ($\mathbf{B}^{-1}$) is decomposed into eigenvectors and eigenvalues:

$$\mathbf{B}^{-1} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{T} \tag{23}$$

Only the 50 eigenvectors with the smallest eigenvalues are retained. Those eigenvectors correspond to the dominant error modes defined by the covariance matrix **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):

$$\boldsymbol{x}^{(k)} = \mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{A}(\mathbf{\Omega})_k \quad \text{for} \quad k = 1, \ldots, N+1 \tag{24}$$

where N is the number of eigenvectors retained, **A** is a $N+1 \times N$ matrix whose column vectors form an orthonormal basis perpendicular to the vector $\mathbf{1}_{N+1\times 1}$ and $(\mathbf{\Omega})_k$ is the kth column of a $N \times N$ random orthogonal matrix **Ω**. This ensemble will have exactly a zero mean and its ensemble covariance is the covariance matrix **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. 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. (18) are obtained, the model is rerun with the corrected boundary values for 60 days.

OSD
6, 2423–2460, 2009

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 EOT data S_{EOT} is fixed at 0.01 m (the impact of this value will be discussed later) and 10 different values for S_{HF} are used ranging from 0.002 to 1 m/s. Each blue circle point in Fig. 6 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) is also included in Fig. 6 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_{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_{HF}=0.2 m/s for the following experiment.

In a second series of experiment, S_{HF} is thus fixed to 0.2 m/s and S_{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. As long as S_{EOT} is smaller than 0.017 m, an improvement is observed compared to both data sets (Fig. 7). This justifies a posteriori the value of 0.01 m chosen previously.

OSD
6, 2423–2460, 2009

In a last series of sensitivity experiments, we assess the benefit of assimilating only EOT data, and leaving the HF radar observation as an independent validation data set (Fig. 8). In this case, we notice that for all values S_{EOT}, the assimilation improves the results compared to the free model run for both data sets. However, as expected, the improvement relative of 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.

5.2 Error analysis

The amplitude and phase of the M2 tidal constituent are computed over the first 40 days of simulation by a tidal analysis (Fig. 9). The largest improvement obtained by assimilation is the amelioration of the u-phase relative to the free model. The assimilation did not damage the u-amplitude which is close to the observations as it is the case for the the free model run. The free model run underestimates the v-amplitude in the northern part of the zone covered by the HF radar. This is improved by the assimilation, but the model v-amplitude is still lower than the v-amplitude obtained from the observations. The v-phase is in general also closer to the observations in the model run with assimilation than in the free run.

To obtain an overall view of the impact of the assimilation, RMS maps are computed between the observations and the free and assimilative model runs (Fig. 11).

$$\mathrm{RMS}^2 = \frac{1}{2N} \sum_{i=1}^{N} \left(u_i^o - u_i^m\right)^2 + \left(v_i^o - v_i^m\right)^2 \tag{25}$$

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

OSD
6, 2423–2460, 2009

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 an higher expected error standard deviation of the observations (left panel Fig. 10).

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 (RMS^2_{M2}) and the remaining errors sources (RMS'^2):

$$RMS^2 = RMS^2_{M2} + RMS'^2 \quad (26)$$

The contribution of the M2 tides RMS^2_{M2} 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 12 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 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 those RMS errors in Figs. 11 and 12 are computed based on the 40 days of HF radar velocity used in the assimilation. Those error analyses where repeated for the following 20 days of model simulation. The error reduction for this analysis is very sim-

OSD
6, 2423–2460, 2009

ilar to the one obtained by using the assimilated observations (Fig. 13). This indicates that the boundary conditions where 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. 10).

5.3 Tide gage station

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

OSD
6, 2423–2460, 2009

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 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 it acts similar as an Kalman Smoother. For a linear model and infinite ensemble, it would provide the same solution as 4-D-Var but without requiring the formulation of an adjoint. Since the method aims to derive the optimal perturbation and not the optimal state, the final solution is obtained by re-running the model. The analyzed model solution satisfies thus the model equations exactly.

The applicability of this approach for different degrees of non-linearity and for uncertain time-dependent forcings remain to be shown. In the later 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 unassimilated HF radar velocities and the M2 amplitude and phase of tide gage observations at Helgoland and Cuxhaven.

OSD
6, 2423–2460, 2009

Appendix A

Root mean square difference between tidal signals

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

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

$$= \frac{A^2 + A'^2}{2} - A\,A'\cos(\phi - \phi') \tag{A2}$$

This measure combines a difference in amplitude and phase and is thus useful to asses 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. 14. The meridional (u) and zonal (v) components of the surface current $\boldsymbol{u}$ can be calculated by solving the following equation:

$$\begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ \vdots & \vdots \\ \cos\theta_n & \sin\theta_n \end{pmatrix} \underbrace{\begin{pmatrix} u \\ v \end{pmatrix}}_{\boldsymbol{u}} = \begin{pmatrix} u_{r1} \\ \vdots \\ u_{rn} \end{pmatrix} \tag{B1}$$

For more than two radial components, the system is overdetermined. In this case, the system can be solved by minimizing the total error ε^2 using all n radial components

OSD
6, 2423–2460, 2009

$\boldsymbol{u}_{rj}$ (j=1, . . . , n) and their error variance, σ^2_{rj}:

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

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

$$\varepsilon^2 = |\, \mathbf{A}\boldsymbol{u} - \mathbf{b}\,|^2 \tag{B3}$$

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

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

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

$$(\mathbf{A}^T\mathbf{A})\boldsymbol{u} = \mathbf{A}^T\mathbf{b} \tag{B5}$$

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 $\mathbf{C}$:

$$\mathbf{C} = (\mathbf{A}^T\,\mathbf{A}\,)^{-1} \tag{B6}$$

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

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

Acknowledgements. The DGFI (Deutsches Geodätisches Forschungsinstitut) is acknowledged for providing for the EOT08a data. J. 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 A. Barth and A. Alvera Azcárate. This is a MARE publication.

References

Alvera-Azcárate, A., Barth, A., Bouallègue, Z. B., Rixen, M., and Beckers, J.-M.: Forecast verification of a 3-D model of the Ligurian Sea. The use of discrete wavelet transforms in the skill assessment of spatial forecasts, J. Mar. Syst., 65, 460–483, doi:10.1016/j.jmarsys.2005.09.015, 2007. 2438

Arakawa, A. and Lamb, V.: Computational design of the basic dynamical process of the UCLA general circulation model, in: Methods Computational Physics, Academic Press, New York, 173–265, 1977. 2430

Barrick, D., Evans, M., and Weber, B.: Ocean surface currents mapped by radar, Science, 198, 138–144, 1977. 2426

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 Modell., 19, 204–218, doi:10.1016/j.ocemod.2007.06.007, 2007. 2425

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, 2008. 2425

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/. 2425, 2430, 2432

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. 2425

Burchard, H. and Bolding, K.: GETM – a general estuarine transport model. Scientific Documentation, Tech. Rep. EUR 20253 EN, European Commission, 2002. 2430

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, 1998. 2434

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/. 2425

Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, 2003. 2424

OSD

6, 2423–2460, 2009

Evensen, G.: Sampling strategies and square root analysis schemes for the EnKF, Ocean Dynam., 54, 539–560, 2004. 2424
Gurgel, K.-W.: Shipborne measurement of surface current fields by HF radar (extended version), L'Onde Electrique, 74, 54–59, 1994. 2428
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, Nonlin. Processes Geophys., 12, 491–503, 2005, http://www.nonlin-processes-geophys.net/12/491/2005/. 2425
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. 2424
Lettellier, T., Lyard, F., and Lefebre, F.: The new global tidal solution: FES2004, in: Ocean Surface Topography Science Team Meeting, St. Petersburg, Florida, 2004. 2429
Lyard, F., Lefevre, 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. 2429
Malanotte-Rizzoli, P., Young, R. E., and Haidvogel, D. B.: Initialization and data assimilation experiments with a primitive equation model, Dynam. Atmos. Ocean., 13, 349–378, 1989. 2425
Nerger, L., Hiller, W., and Schrter, J.: A Comparison of Error Subspace Kalman Filters, Tellus A, 57A, 715–735, doi:10.1111/j.1600-0870.2005.00141.x, 2005. 2434
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. 2425
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. 2435
Pham, D. T.: Stochastic methods for sequential data assimilation in strongly nonlinear systems, Mon. Weather Rev., 129, 1194–1207, 2001. 2424, 2435
PRISMA: Prozesse im Schadstoffkreislauf Meer-Atmosphäre: Ökosystem Deutsche Bucht. BMFT-Projekt 03F0558A1 (1 January 1990–31 October 1993), Abschlussbericht, ZMK-Universität Hamburg, 1994. 2426
Savcenko, R. and Bosch, W.: EOT08a – empirical ocean tide model from multi-mission satellite altimetry, Tech. rep., Deutsches Geodätisches Forschungsinstitut, online available at: ftp:

OSD
6, 2423–2460, 2009

//ftp.dgfi.badw.de/pub/EOT08a/doc/EOTO8a.pdf, 2008. 2429
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, 271–289, Springer Verlag, 1994. 2431
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. 2430
Talagrand, O.: On the damping of high-frequency motions in four-dimensional assimilation of meteorological data, J. Atmos. Sci., 1571–1547, 1972. 2425
van Leeuwen, P. J.: An ensemble smoother with error estimates, Mon. Weather Rev., 129, 709–728, 2001. 2434

Table 1. Comparison with tide gage observations. Amplitude is in m and phase in degrees.

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

OSD

6, 2423–2460, 2009

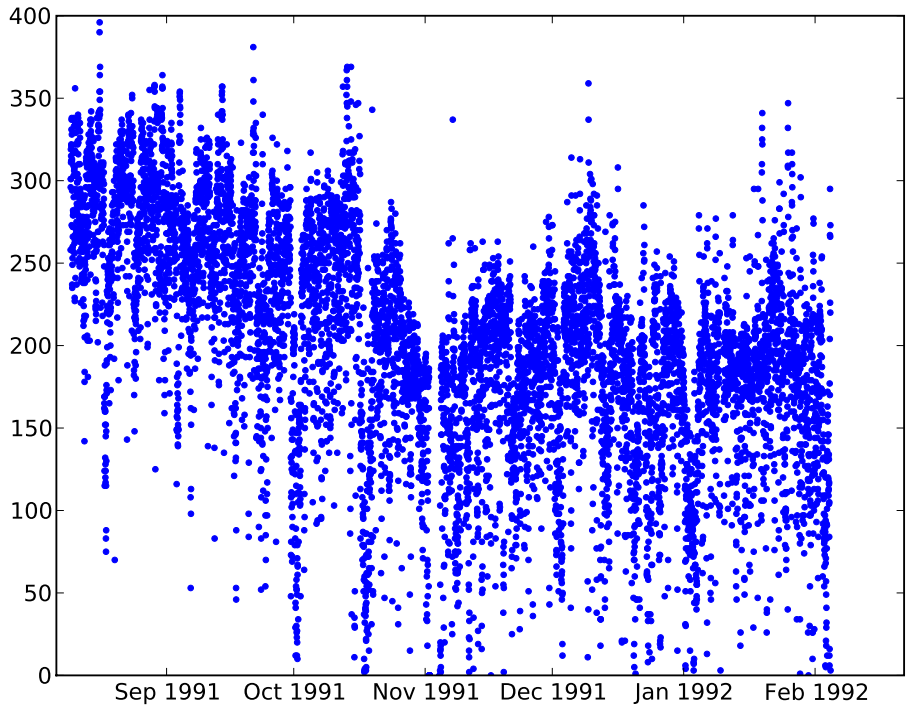

Fig. 1. Temporal coverage of the HF radar zonal and meridional surface velocity observations: the number of observation grid points with valid data available is shown for each sample time. The Cartesian observation grid has 567 points, but due to land coverage and distance from the antennas, the maximum spatial coverage is 396 grid cells.

OSD

6, 2423–2460, 2009

Ensemble smoother for optimizing tidal boundary conditions

A. Barth et al.

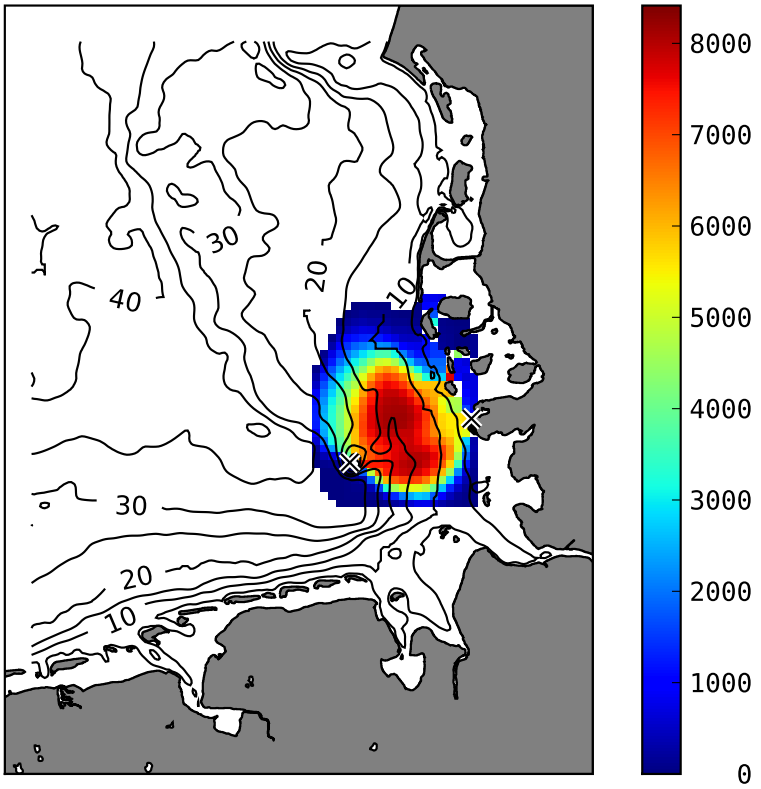

Fig. 2. Spatial coverage of the HF radar zonal and meridional surface velocity observations: the number of samples available at each observation grid point is color-coded according to the color-bar. The entire data set comprises 8414 samples in time. Contour lines show depth in meters as used in the hydrodynamic model's bathymetry. The crosses show the location of HF antennas.

OSD

6, 2423–2460, 2009

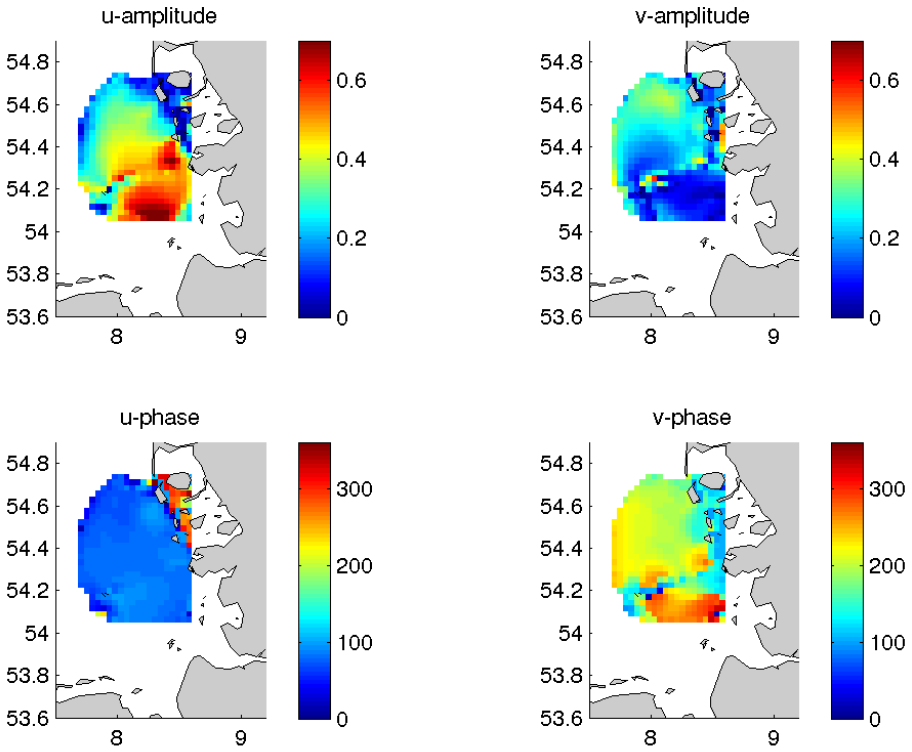

Fig. 3. M2 amplitude (in m/s) and phase (in degrees) of zonal (u) and meridional (v) surface currents of the observations.

OSD

6, 2423–2460, 2009

Ensemble smoother for optimizing tidal boundary conditions

A. Barth et al.

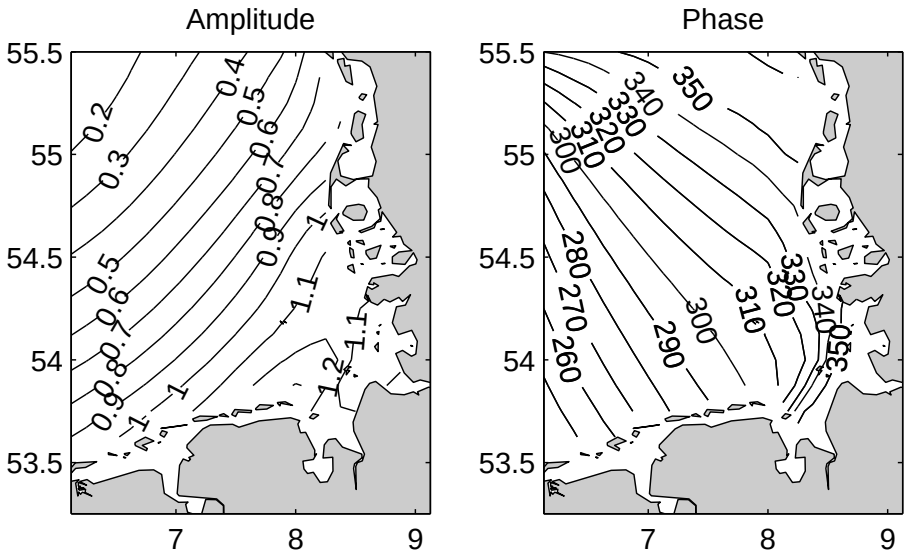

Fig. 4. M2 amplitude (in m) and phase (in degrees) of EOT08a for the German Bight.

OSD
6, 2423–2460, 2009

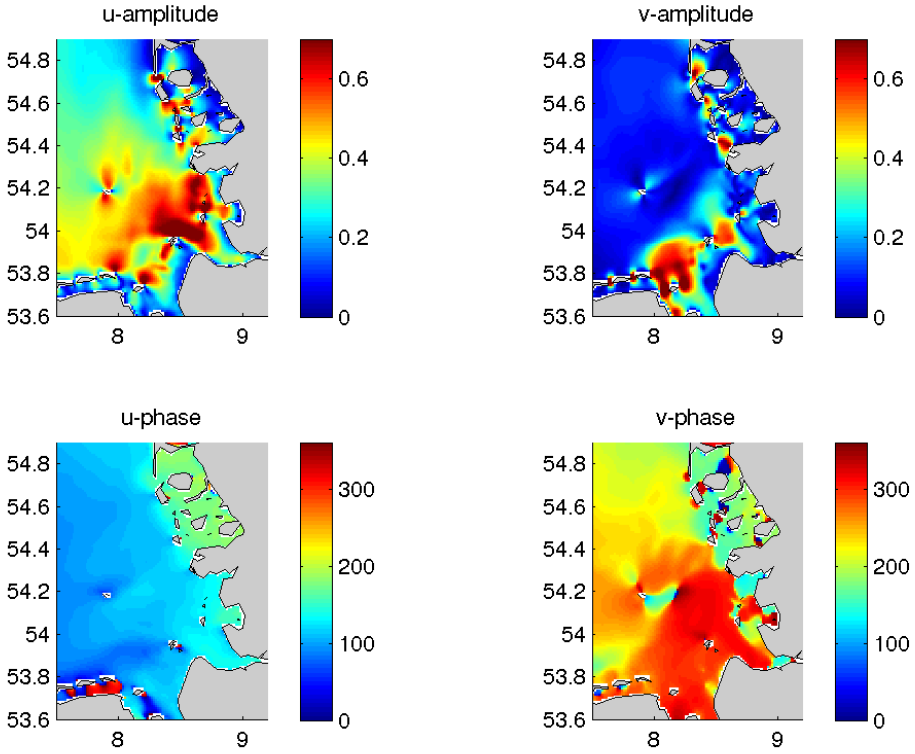

Fig. 5. M2 amplitude (in m/s) and phase (in degrees) of zonal (u) and meridional (v) surface currents of the free model run.

OSD

6, 2423–2460, 2009

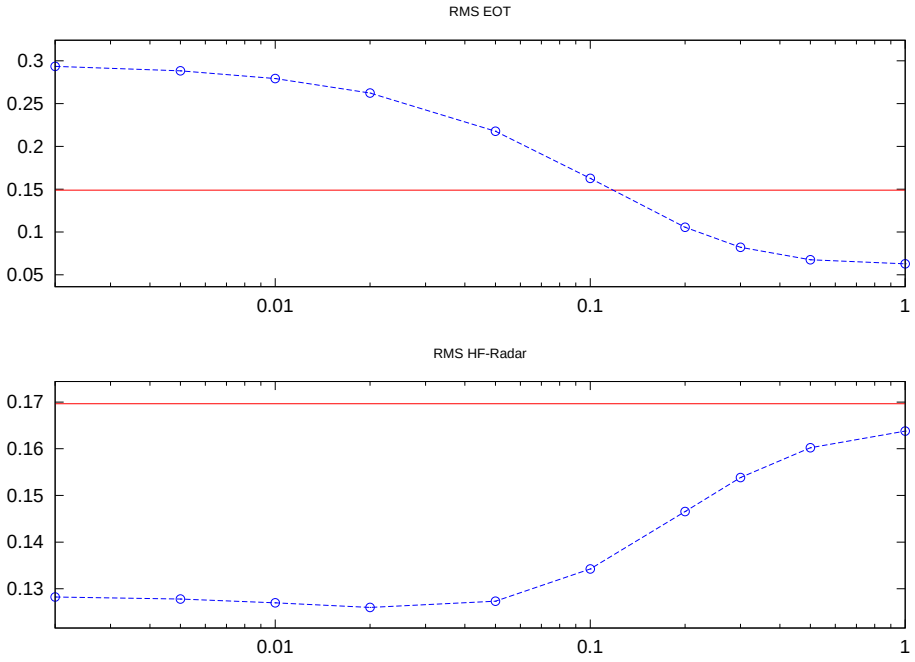

Fig. 6. RMS error relative to EOT analysis (in m) and HF radar observations (in m/s) for S_{EOT}=0.01 m and different values for S_{HF} (in m/s). Free run is in red and run with data assimilation in blue.

OSD
6, 2423–2460, 2009

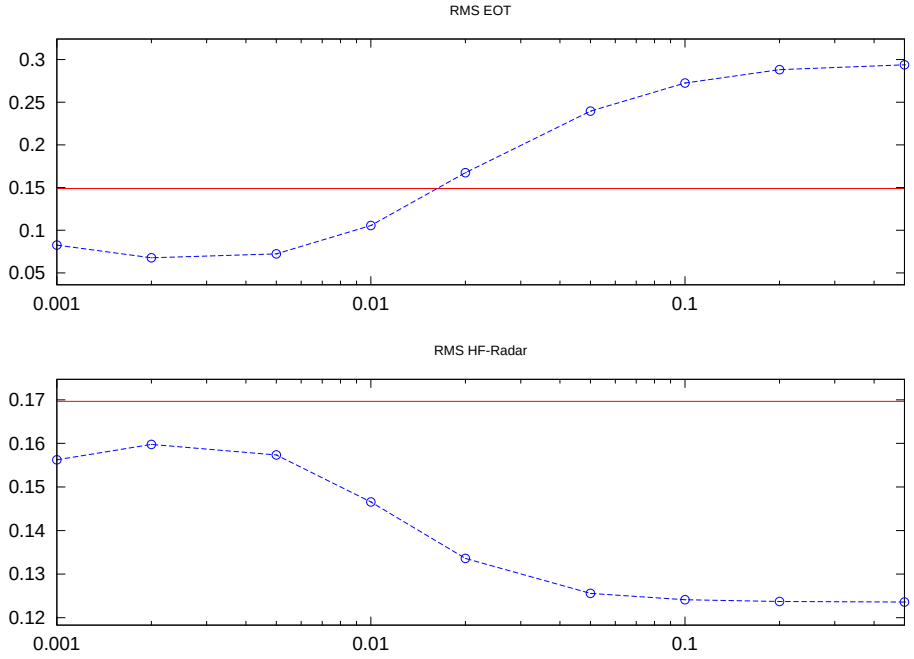

Fig. 7. RMS error relative to EOT analysis (in m) and HF radar observations (in m/s) for S_{HF}=0.2 m/s and different values for S_{EOT} (in m). Free run is in red and run with data assimilation in blue.

OSD
6, 2423–2460, 2009

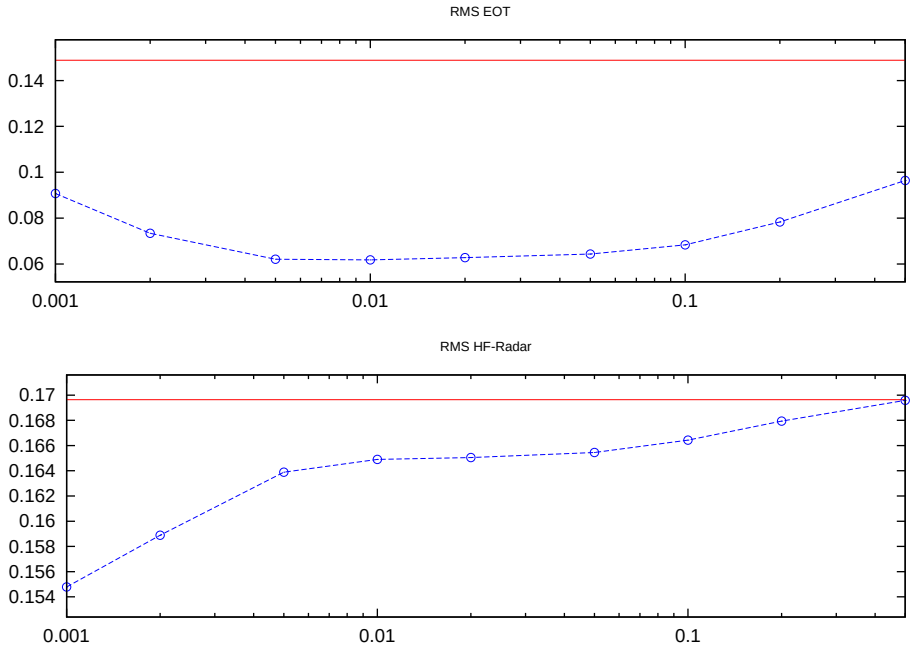

Fig. 8. RMS error relative to EOT analysis (in m) and HF radar (in m/s) for different values for S_{EOT} (in m) without assimilation of HF radar observations. Free run is in red and run with data assimilation in blue.

OSD
6, 2423–2460, 2009

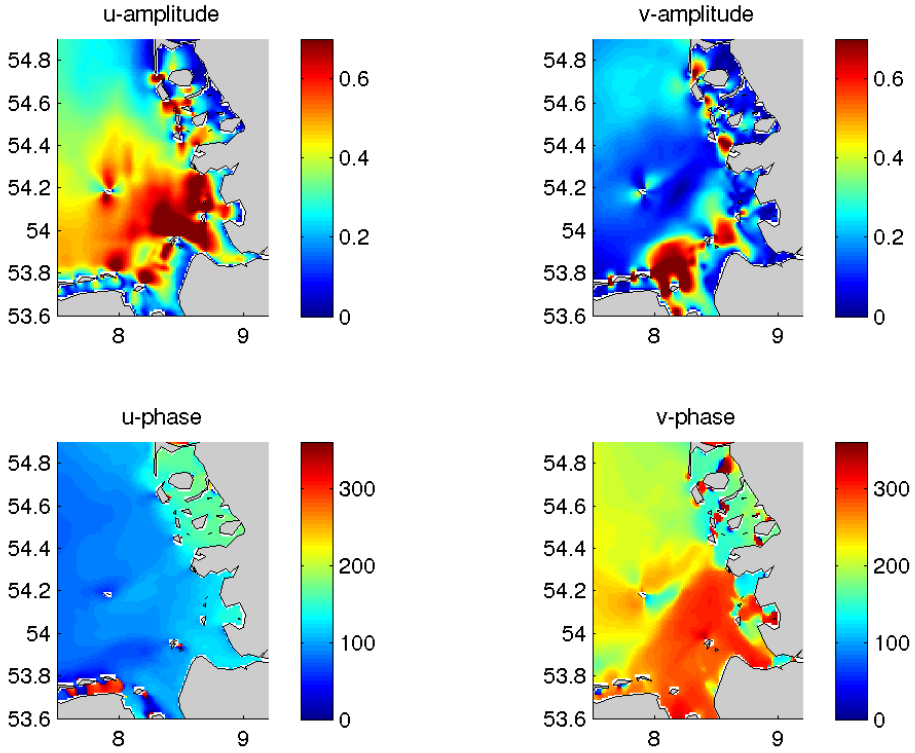

Fig. 9. M2 amplitude (in m/s) and phase (in degrees) of zonal (u) and meridional (v) surface currents of the model run with assimilation.

OSD
6, 2423–2460, 2009

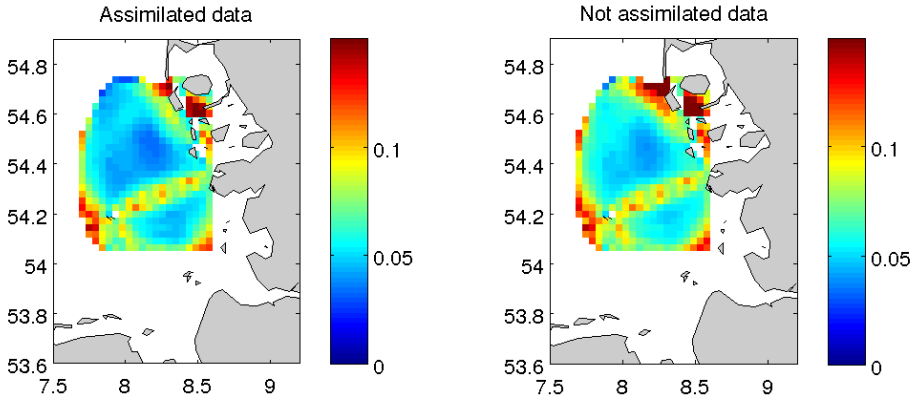

Fig. 10. The square root of the averaged expected observational error variance (without the representativity error) for the observations from 1 September 1991 to 10 October 1991 (assimilated; right panel) and from 10 October 1991 to 30 October 1991 (not assimilated; left panel). Units are m/s.

OSD
6, 2423–2460, 2009

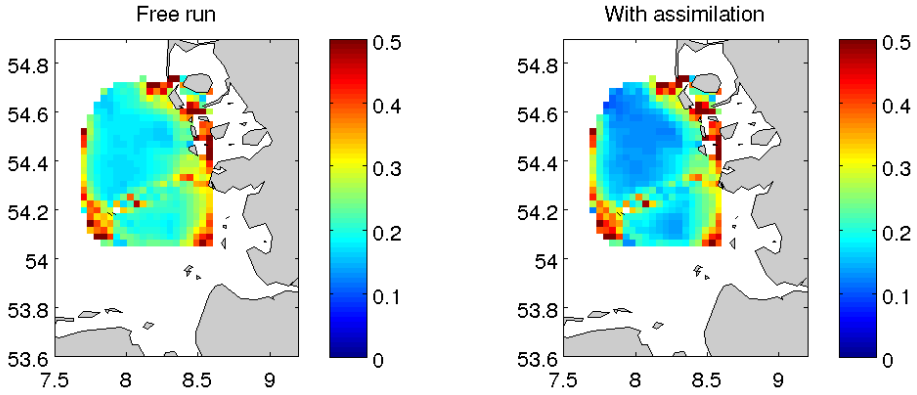

Fig. 11. RMS difference (in m/s) between surface current observations and model results without (left panel) and with assimilation (right panel).

OSD

6, 2423–2460, 2009

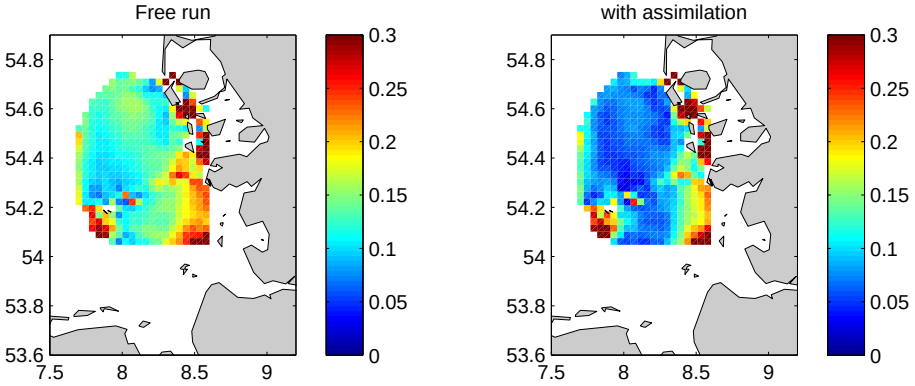

Fig. 12. RMS difference (in m/s) between surface current observations due to the M2 tides and the corresponding model results without (left panel) and with assimilation (right panel).

OSD
6, 2423–2460, 2009

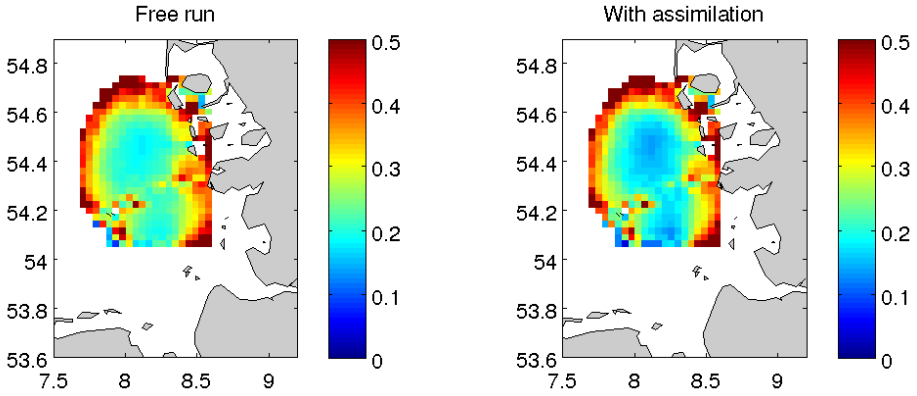

Fig. 13. RMS difference (in m/s) between surface current observation and model results without (left panel) and with assimilation (right panel) compared to independent data.

OSD

6, 2423–2460, 2009

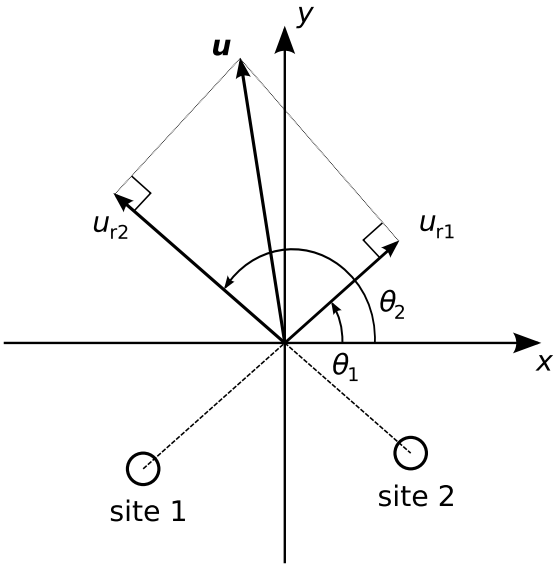

Fig. 14. Calculation of the surface current ***u*** from two radial components u_{r1} and u_{r2}.

OSD
6, 2423–2460, 2009