Correcting surface winds by assimilating High-Frequency Radar surface currents in the German Bight

Alexander Barth[1,2], Aida Alvera-Azcárate[1,2],
Jean-Marie Beckers[1,2], Joanna Staneva[3]
Emil V. Stanev[3] Johannes Schulz-Stellenfleth[3]
[1]GeoHydrodynamics and Environment Research (GHER),
MARE, Liège, Belgium
[2]National Fund for Scientific Research, Belgium
[3]GKSS Research Centre, Geesthacht, Germany

February 1, 2011

Abstract

Surface winds are crucial for accurately modeling the surface circulation in the coastal ocean. In the present work, high-frequency (HF) radar surface currents are assimilated using an ensemble scheme which aims to obtain improved surface winds taking into account ECMWF (European Centre for Medium-Range Weather Forecasts) winds as a first guess and surface current measurements. The objective of this study is to show that wind forcing can be improved using an approach similar to parameter estimation in ensemble data assimilation. Like variational assimilation schemes, the method provides an improved wind field based on surface current measurements. However, the technique does not require an adjoint and it is thus easier to implement. In addition, it does not rely on a linearization of the model dynamics. The method is validated directly by comparing the analyzed wind speed to independent *in situ* measurements and indirectly by assessing the impact of the corrected winds on model sea surface temperature (SST) relative to satellite SST.

1

1 Introduction

During the last decade, an increasing number of high-frequency (HF) radar systems have been installed around the world (see Barrick et al (1977) for early accounts of the technique). These installations provide surface current measurements over a relatively large area (with a typical range of several hundreds of kilometers) of the coastal ocean. HF radar measurements are a very valuable data set to constrain regional and coastal models by data assimilation. Most studies use a sequential scheme to assimilate these surface currents. The covariances are either based on ensembles (Breivik and Satra, 2001; Oke et al, 2002; Barth et al, 2008) or use explicit statistical parameterizations (Lewis et al, 1998; Paduan and Shulman, 2004; Wilkin et al, 2005; Shulman and Paduan, 2009). It has also been shown that HF radar currents can be assimilated using adjoint-based assimilation schemes (Kurapov et al, 2003; Hoteit et al, 2009).

Observations measured at a high temporal resolution constitute however a challenge for sequential data assimilation schemes. The analysis steps produces often unrealistic transient processes (*e.g.* Malanotte-Rizzoli et al, 1989) which are not dissipated when the length of the assimilation cycle is short. Too frequent assimilation of observations can thus degrade the model results (Talagrand, 1972). To avoid this problem, it is thus preferable to correct the source of the model error (if this is possible) rather than the model state.

This problem is addressed in the context of HF radar assimilation by realizing that, in coastal zones, a large part of the model error in surface currents can be attributed to errors in surface winds (He et al, 2004; Barth et al, 2008). The surface current observations are thus used in this study to improve the wind forcing instead of modifying the model state vector. The optimization of forcing fields is often achieved by a 4D-Var assimilation scheme using the surface winds as control variable (*e.g.* Hoteit et al, 2009).

In this work, we use an ensemble-based assimilation scheme in a realistic context to estimate the covariance between observed surface currents and wind fields. The problem is similar to parameter estimation methods which have been applied to the Kalman filter (*e.g.* Gelb, 1974; Ljung, 1979), to the ensemble-Kalman filter in particular (Anderson, 2001; Annan et al, 2005; Aksoy et al, 2006; Evensen, 2007) and to the particle filter (Losa et al,

2003). The approach has also been applied to the estimation of forcing fields of an ocean model in idealized configurations (Skachko et al, 2009; Skandrani et al, 2009). It is also related to parameter estimation using Green's functions (Menemenlis et al, 1997). To our knowledge, this is the first time that HF radar surface currents are used to correct wind fields using an ensemble data assimilation approach in a realistic application, where the results are validated with independent in-situ wind measurements.

The characteristics of the hydrodynamical model used for this experiment are briefly presented in section 2. The assimilated HF radar observations to be assimilated and the observations for validation are described in section 3. Section 4 presents the data assimilation procedure. The results are discussed in section 5. Finally, the conclusions of this study are presented in section 6.

2 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. Prognostic variables include surface elevation, horizontal velocity, temperature and salinity. In the vertical, this configuration uses 21 σ levels. It covers the German Bight (figure 1) with a horizontal resolution of about 0.9 km. Its boundary conditions are extracted from a 5 km resolution North Sea-Baltic Sea model. Atmospheric fluxes are estimated by the bulk formula using 6-hourly ECMWF (European Centre for Medium-Range Weather Forecasts) re-analysis data at 1/2 degree resolution. The model is also forced by hourly river run-off data provided by the BSH (Bundesamt für Seeschifffahrt und Hydrographie, Germany) operational model. Details about the large-scale and nested models can be found in Staneva et al (2009).

3 Observations

HF radar surface currents measurements were carried out from August 1991 to February 1992 by the University of Hamburg in the frame of the PRISMA project (PRISMA, 1994). One system was installed on Helgoland and one near the town of St. Peter–Ording (figure 1). Both systems were based on

modified CODAR (Barrick et al, 1977) setups, operating at 29.85 MHz and consisting of a four–element array using direction–finding. The radial velocities of both sites were combined to derive zonal and meridional velocities together with estimates of measurement uncertainty on a 27×21 cartesian grid with 3 km horizontal resolution and 30 min temporal resolution (Barth et al, 2010).

For independent validation, *in situ* wind speed measurements at Helgoland and Sylt from the German Meteorological Service are used at a reference height of 10m. Wind direction measurements for those sites were unfortunately not available and only wind intensity was used for validation. The model surface temperature is also compared to daily nighttime AVHRR Pathfinder version 5 with a minimum quality flag of 4.

4 Data assimilation

The HF radar data set has been used previously to correct M2 tidal boundary conditions (Barth et al, 2010). In the present study, the focus is on optimizing the surface winds by data assimilation. In our ensemble simulation, the uncertainty in the driving wind field is represented by an ensemble of perturbed wind forcings. The spatial structure of the wind perturbations is obtained by using the Fourier decomposition of the ECMWF wind vectors **u** for September and October 1991:

$$\mathbf{u}(x, y, t) = \sum_k \mathbf{a}_k(x, y) \exp(i\omega_k t), \tag{1}$$

where ω_k is the k-th angular frequency (positive or negative) and $\mathbf{a}_k(x, y)$ are complex spatial vector fields corresponding to the Fourier coefficients of the angular frequency ω_k:

$$\omega_k = \frac{2\pi k}{\Delta t} \qquad k = -\frac{k_{max}}{2}, \ldots, \frac{k_{max}}{2} - 1 \tag{2}$$

where Δt is 6 hours and k_{max} is 244 (the number of six-hourly wind fields these two month). Wind vector perturbations $\mathbf{u}_p(x, y, t)$ are obtained by:

$$\mathbf{u}_p(x, y, t) = \alpha \operatorname{Re}\left(\sum_k \mathbf{a}_k(x, y) z_k(t)\right), \tag{3}$$

4

where z_k is a complex random time series with a temporal correlation scale of $T_k = 2\pi/|\omega_k|$, zero mean and unit variance. The real and imaginary parts of the random time series have the following covariance $C_T(t, t')$:

$$C_T(t, t') = e^{-\frac{(t-t')^2}{T_k^2}} \tag{4}$$

The random time series are constructed by multiplying a normal distributed random vector with the square root of matrix $\mathbf{C}_T$ (Evensen, 1994). By using such perturbations, we essentially assume that the wind errors have similar spatial and temporal scales than the ECMWF wind field. This approach is based on the assumption that the uncertainty of the ECMWF wind fields is mainly caused by phase and amplitude errors of the underlying processes rather than errors in the variability.

The approach is similar to perturbing atmospheric forcings according to spatial empirical orthogonal functions (*e.g.* Barth et al, 2007; Vandenbulcke et al, 2008; Lucas et al, 2008; Béal et al, 2010). In both cases, the time variability of the forcing field is used to estimate the spatial patterns of the perturbation. However, the present approach based on the Fourier decomposition ensures that the spatial structure of the variability associated to, *e.g.*, the diurnal cycle is multiplied by a time series with a compatible temporal scale. The factor α takes into account that the expected wind error is in general smaller than the temporal variability. We assume that the expected wind error standard deviation is 30% of the temporal variability ($\alpha = 0.3$).

An ensemble of 100 wind perturbations is created according to this procedure and added to the ECMWF wind fields. The GETM model is run for each wind forcing for 30 days starting on the 1 September 1991. From each model run, the surface currents are extracted. A diagram summarizing the assimilation experiment in given in Figure 2.

Sequential assimilation is often implemented to derive the optimal state of the ocean, in which case, the state vector is composed by all prognostic model variables at all model grid points. In the present study, we want to first obtain improved wind forcing using the observations. Therefore we use here the terminology "estimation vector" instead of "state vector". The estimation vector $\mathbf{x}$ contains thus the u and v components of the wind field over the model domain (Figure 2).

In most sequential assimilation schemes, the "observation operator" $h(\cdot)$ represents a spatial interpolation operator which extracts the observed variables from the model state. But it can also be a general operator which links observations to the parameters for which one seeks to obtain a better estimate (*e.g.* Greenwald et al, 2002). Here the operator $h(\cdot)$ links surface currents and wind forcings. The observation operator applied to the vector $\mathbf{x}$, $h(\mathbf{x})$, represents the surface currents (within a 30-day integration period) of the model run using the perturbed wind forcing $\mathbf{x}$. The observation vector $\mathbf{y}^o$ includes all HF radar observations within the considered time interval and $\mathbf{R}$ is its corresponding error covariance. The observation error covariance matrix $\mathbf{R}$ is assumed diagonal for simplicity. It is constructed by adding a constant S^2_{HF} (accounting for the representativity error) to the error variance of the HF radar data $\mathbf{R}'_{\mathrm{HF}}$ (based on the noise variance in the averaging procedure and the geometric dilution of precision).

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

where $\mathbf{I}$ is the identify matrix. The value of this constant S^2_{HF} will be determined later (section 5.2).

From the ensemble simulation, we derive the following matrices whose columns represent the deviation of the ensemble members around the ensemble mean of the surface winds ($\mathbf{S}$) and the surface currents ($\mathbf{E}$):

$$
\begin{aligned}
(\mathbf{S})_l &= (N-1)^{-\frac{1}{2}}\left(\mathbf{x}^{(l)} - \langle\mathbf{x}\rangle\right) & (6) \\
(\mathbf{E})_l &= (N-1)^{-\frac{1}{2}}\left(h\left(\mathbf{x}^{(l)}\right) - \langle h\left(\mathbf{x}\right)\rangle\right) & (7)
\end{aligned}
$$

where N is the number of ensemble members, the index l refers to the ensemble member ($l = 1, \ldots, N$) and $\langle\cdot\rangle$ is the ensemble average. These matrices are scaled such that $\mathbf{S}\mathbf{E}^T$ and $\mathbf{E}\mathbf{E}^T$ represent the following covariance matrices:

$$
\begin{aligned}
\mathbf{S}\mathbf{E}^T &= \mathrm{cov}(\mathbf{x}^b, h(\mathbf{x}^b)) & (8) \\
\mathbf{E}\mathbf{E}^T &= \mathrm{cov}(h(\mathbf{x}^b), h(\mathbf{x}^b)) & (9)
\end{aligned}
$$

The analysis wind fields are computed using the analysis step of the Kalman filter with a non-linear observation operator (Chen and Snyder, 2007):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{S}\mathbf{E}^T \left(\mathbf{E}\mathbf{E}^T + \mathbf{R}\right)^{-1} \left(\mathbf{y}^o - h(\mathbf{x}^b)\right) \tag{10}$$

The superscripts a and b refer to the analysis and background estimates. As in the Kalman filter, we assume that the background estimate (here the ECMWF wind forcing) and the HF radar observations are unbiased. If the observation operator is strongly non-linear, those assumptions are not sufficient to guarantee that the analysis $\mathbf{x}^a$ is unbiased. However, one can show (see appendix A) that the analysis is unbiased if,

$$\left\langle h(\mathbf{x}^b) \right\rangle = h\left(\left\langle \mathbf{x}^b \right\rangle\right) \tag{11}$$

This requirement is obviously satisfied for any linear observation operator. Since the observation operator used in this study is non-linear, we verified this equation *a posteriori* using the ensemble simulation. Both sides of equation (11) represent surface currents. Their RMS difference is 0.01 m/s and can thus be considered small compared to other errors and biases. For a strongly non-linear operator $h(\cdot)$, the resulting ensemble might not be Gaussian distributed. Formally, the analysis can still be derived as the state which minimizes the error variance (instead of maximizing the likelihood), but the analyzed state might have a low probability. It is thus important to validate the analyzed wind field with independent data to verify the realism of the analysis.

Since the error covariances in equation (10) have a reduced rank, the Sherman–Morrison–Woodbury formula is used to perform the matrix inversions in the error subspace spanned by the ensemble members (*e.g.* Pham et al, 1998):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{S}\left(\mathbf{E}^T\mathbf{R}^{-1}\mathbf{E} + \mathbf{1}\right)^{-1}\mathbf{E}^T\mathbf{R}^{-1}\left(\mathbf{y}^o - h(\mathbf{x}^b)\right) \tag{12}$$

This assimilation procedure differs from most implementations of the Kalman filter and smoothers since the optimal perturbation (here of the

wind forcing) is derived instead of the optimal model state. The analyzed model state must thus be obtained in a final model integration using the analyzed wind forcings. This procedure ensures that the final model solution satisfies exactly the model equations. This filter has been applied in Barth et al (2010) to estimate tidal boundary conditions and is closely related to the Ensemble Smoother (van Leeuwen, 2001), 4DEnKF (Hunt et al, 2004, 2007) and to the asynchronous EnKF (Sakov et al, 2010) where the time dimension is also embedded in the model state and observation vector.

Formally one can also derive the assimilation scheme by defining the estimation vector as the model trajectory augmented by the wind forcings, as this is usually done in parameter estimation schemes (Gelb, 1974). The analyzed wind field would be exactly the same as in the approach explained above. However, with the augmented state vector, one would obtain also directly an updated model trajectory which in general can be different than with the procedure explained above since the model is non-linear. Therefore, we prefer to present the assimilation procedure differently in this manuscript: the estimation vector contains only the wind field since the updated trajectory is obtained in a second step by rerunning the model.

Most implementations of the Ensemble Kalman Filter require spatial localization which amounts to an analysis performed zone by zone (e.g. water columns) to increase the effective dimension of the error subspace (*e.g.* Brankart et al, 2003; Evensen, 2003). Here a similar approach is used but in the time dimension not in space. The analysis is performed day by day, *i.e.* every day there are 48 time instances of HF radar observations used to derive 4 time instances of the wind fields. If a smaller time window is used, then model errors at tidal frequencies could produce a spurious correction of the wind fields. However, this might produce discontinuities in the wind forcing from one assimilation time interval to the next. This issue is examined in section 5.4.

8

5 Results

5.1 Error covariances

From the ensemble simulations, one can compute the error covariance (in space and time) between the ocean currents and wind forcings. These error covariances are essential in data assimilation, because they relate measured variables to the uncertain parameters that need to be constrained by the observations (here the wind forcing). Panel (a) of Figure (3) shows the correlation between the u-component of the surface currents at the location of the white marker and the u-component of the wind fields over the model domain. Considering that i is the grid index of the marked location (in observation space), then this figure shows the ith column of the of matrix $\mathbf{SE}^T$ (8):

$$
\begin{aligned}
\mathbf{SE}_i^T &= \mathrm{cov}(\mathbf{x}^b, h_i(\mathbf{x}^b)) \\
&= \left\langle (\mathbf{x} - \langle \mathbf{x} \rangle)\,(h_i(\mathbf{x}) - \langle h_i(\mathbf{x}) \rangle)^T \right\rangle
\end{aligned}
$$

where the ensemble average $\langle \cdot \rangle$ is calculated using the ensemble of wind fields and the resulting ensemble of ocean currents.

The same analysis is shown in panel (b) for the v components. Such diagnostics are instructive because a current measurement near the marker would produce a wind field correction proportional to this field (multiplied by the standard deviation of the wind field since the correlation is shown and not the covariance). There is no *a priori* length-scale chosen in the ensemble scheme, and the structures in covariance are only a product of the model dynamics and the perturbation scheme of the wind forcings.

The error correlation between zonal currents at the marker and zonal winds shows essentially a local structure where the currents correlates more strongly with local winds than with remote winds. In this case one can assume a direct momentum transfer between the atmosphere to the ocean responsible for this pattern. It is also interesting to note the relatively high correlation between ocean currents and wind over land. Measurements of ocean currents have thus the potential to improve also wind field over land.

Overall, the correlation between the meridional currents and the meridional winds is smaller and the structure is essentially non-local. This remote

9

interaction could be seen in different covariance diagnostics (not shown), and can be attributed to the fact that wind-driven currents can be generated offshore and then influence the flow closer to the shore. Near the shore the constraints from the coastline are also increasingly important which can also limit the effect of local winds.

It should be reminded here that the wind field is the only parameter that is considered as uncertain in this assimilation experiment. If other error sources were taken into account, the overall magnitude of the error correlation would likely be smaller.

5.2 Comparison to in situ wind

The RMS differences between *in situ* wind and the ECMWF wind fields RMS_f and analysis winds RMS_a for different values of the representativity error S_{HF} are shown in table 1. The RMS errors are computed between hourly wind measurements and 6-hourly (linearly interpolated) wind fields. Part of the RMS error is thus due to wind variability on a timescale shorter than 6 hours. The mean square skill score measures the relative improvement of the mean square error relative to the model simulation with the original ECMWF wind (called "free" model run).

$$\text{skill score} = 1 - \frac{\mathrm{RMS}_a^2}{\mathrm{RMS}_f^2} \tag{13}$$

Similar statistics are computed for model surface temperature (at 00:00 UTC) compared to daily satellite SST (table 2). The aim is to choose a value for S_{HF} to maximize the skill for both comparisons. We choose for S_{HF} 2 m/s as a compromise between the *in situ* wind and satellite SST validation (next section). This high value of the representativity error can be explained by the fact that the temporal (and spatial) correlation of the observations is not explicitly considered in the observation error covariance. Since all 48 observations are assimilated every day, the redundancy between the individual time instances is high. Also, it should be reminded that the difference between observed currents and model currents is not only due to errors in the wind field. Errors in *e.g.* the representation of the tidal signal, errors due to inappropriate parameterizations, wrong density structures or due to lack of resolution cannot be corrected by adjusting the wind field alone. All

this contributes to the parameter S_{HF}. A too low value for S_{HF}, could result in that *e.g.* the error in the density structure would be compensated by modifying the wind field or that information in the region with observations is extrapolated in an unrealistic way only to match observations closely.

Traditionally, the representativity error includes processes and scales in the observation which are not included in the model. Here this definition has to be extended to include processes which cannot be corrected by the proposed assimilation setup (since only the wind forcing is corrected).

A vector correlation analysis (Kundu, 1976) has been performed to characterize the difference between the original ECMWF winds and the analyzed winds. The magnitude of the vector correlation is 0.98 and the phase is -0.6 degrees showing that ECMWF captures well the timing of the wind events and that its direction is not biased in average. The RMS difference between the original ECMWF winds and the analyzed winds fields is 2.5 m/s and the regression coefficient is 1.35. The analyzed winds are thus in average 35% higher than the original ECMWF winds. The RMS difference between the direction of the ECMWF winds and the direction of the analyzed winds (for winds speeds higher than 1 m/s) is 8.5 degrees. This analysis shows therefore that the assimilation corrects essentially the wind speed and to a much lesser degree the wind direction.

The analyzed wind speed is compared to independent *in situ* wind speed measurements at Helgoland and Sylt (see figure 1 for their location). The assimilation increases the wind speed during storm events (such as on day 6 and 11) to a level which agrees with observations (figure 4). However during calmer periods (days 16 to 21), the ECMWF wind intensity is similar to the observed wind speed and the assimilation maintains this good agreement. Winds are thus only increased when the ECMWF underestimates the intensity of a storm and the assimilation does not degrade the results when there is already a good agreement between observations and in situ winds.

Figure 5 shows the spatial extent of the wind corrections. The correction is largest off-shore where the wind variability and presumably the wind uncertainty are highest (left panel). The wind correction scaled by the wind standard deviation (right panel) shows that the wind correction in relative terms is mostly corrected in the interior of the domain near the area covered

by the HF radar systems, as it can be expected. Over almost the entire model domain, the relative correction of the wind field exceeds 30 %. This is consistent with the spatial scales of figure 3. Additional sites with *in situ* winds (especially in deeper water) would allow for a more comprehensive validation, but the fact that the domain is relatively small and the winds at Helgoland and Sylt are quite similar gives confidence that the derived wind corrections improve the wind fields overall.

5.3   Assessment of model SST

In data assimilation, there is always the risk that variables not directly measured are deteriorated by the assimilation due to poorly estimated error covariance between them. This is especially the case for variables and parameters weakly correlated to the observed variables (Hamill et al, 2001).

The impact of the corrected wind fields on the model SST was therefore assessed. The model SST is not relaxed towards satellite SST. Its evolution is thus only influenced by air-sea interactions and ocean dynamics. In order to expect an improvement of model SST, the link between surface winds and surface temperature must be sufficiently strong. Wind velocity is however an important parameter to control the surface heat flux (Mourre et al, 2008). The correlation between the ECMWF winds over the German Bight and the surface temperature used by this atmospheric model was computed to asses this relationship in the present model simulations. For daily averaged values, a correlation of -0.15 between zonal winds and SST and a correlation of -0.23 between meridional winds and SST was obtained. Those results show that a (small) improvement in the model SST is possible when the model simulation is carried out with a more accurate wind forcing.

The RMS error between the model SST and satellite SST averaged over the integration period of the model (figure 6) reveals that the free-running model has large errors near the western boundary. The temperature imposed at the open boundary is too low and this issue propagates into the model domain. Panel (c) of figure 6 shows the model mean currents (averaged in time and over 15×15 grid cells) and the model SST bias of the free run which is largely responsible for the RMS error. The RMS error is also relatively high north of 55°N and south-west of Helgoland. In this area the model temperature is generally too high compared to the satellite SST.

12

With sea surface winds increased to a more realistic strength by the assimilation, the surface waters are cooled in the simulation with corrected winds and are thus in a better agreement with the observations. Overall, the RMS error of the model SST is reduced from 1.21 °C to 1.11 °C (for $S_{HF} = 2$m/s) by the assimilation. Since all forcing parameters except surface winds are identical in these simulations, this reduction of RMS error in SST is thus attributed to the surface winds. The improvement in model SST is essentially due to the reduction of the SST bias (panel (d) of figure 6): the temperature is lowered in areas where the free run is too warm. The assimilation does not affect the temperature at the eastern boundary where the RMS error and bias unchanged. Those errors are attributed to problems in the boundary conditions which where not changed in this assimilation experiment.

5.4 Sensitivity to the length of the assimilation cycle

As mentioned previously, the assimilation has not been performed using all observations in a single assimilation step as in Barth et al (2010), but in several analysis over time periods of one day. The impact of the length of this assimilation window is now examined.

Table 3 summarizes the error statistics of the improved wind fields for assimilation window lengths of 0.5, 1, 2 and 4 days for $S_{HF} = 2$ m/s. While all these experiments produce a reduced RMS error compared to the initial ECMWF wind fields, the improvement decreases for longer assimilation windows. The finite ensemble size limits the ability of the assimilation scheme to produce an analysis close to the observations since the correction of the analysis belongs to a sub-space formed by the ensemble members. The time localization effectively increases substantially the degrees of freedom of the correction.

This localization is simply implemented here by performing the analysis independently over different sub-intervals. However, this can produce time discontinuities between different assimilation intervals. σ_a is defined as the average RMS difference of the wind fields between two consecutive assimilation intervals and σ_s is the average RMS difference of the two consecutive wind fields within an assimilation cycle. Ideally, those quantities should be similar and represent the time variability over 6 hours. However, the

assimilation produces for all tests discontinuities between cycles and these discontinuities tend to increase with an increasing length of the assimilation cycle.

A trade-off has to be made between the RMS reduction relative to *in situ* winds, the size of the discontinuity and its frequency. Taking those elements into account, an assimilation window of 1 day seems to be justified. More sophisticated localization schemes such as the approach proposed by Brankart et al (2003) could be implemented in future studies to reduce the size of the temporal discontinuities of the analyzed wind fields.

6 Conclusions

An ensemble scheme has been implemented to derive atmospheric forcing fields from ocean observations. The application of this ensemble scheme demonstrates the feasibility to use ensemble data assimilation to estimate wind forcings in a similar way to variational data assimilation. The present method does however not require an adjoint and can be expected to be more robust in representing the non-linear saturation of error growth.

However, there are also difficulties associated to the present approach. As all ensemble approaches, the ensemble size must be relatively large to provide a realistic correction. The approach presented here can also reduce the model error where the error source is clearly identified, but it cannot reduce the model error due to non-deterministic model behavior, for example. Such error must still be addressed for example with a Kalman Filter correcting directly the model state.

High-Frequency radar surface current measurements (from 1 September to 30 September 1991) have been used to correct wind fields which have been validated independently with *in situ* wind speed measurements at Helgoland and Sylt. At both sites the analyzed wind speed is 40% closer in terms of mean square error (18% in terms of RMS error) to the observed winds. The correction is largest during storms whose intensity is generally underestimated in the ECMWF fields used.

The improved wind fields have also a positive impact on the realism of

14

the sea surface temperature (SST). The analyzed winds reduce the SST error by 16% in terms mean square error (8% in terms of RMS error).

Only relatively coarse wind forcings at a resolution of 0.5 degrees were available for the model simulation. For more recent years, higher resolution wind fields have been produced. The impact of assimilation for those wind products remains to be evaluated.

The analyzed wind forcing has the same spatial resolution as the original ECMWF wind. It would be also interesting to investigate in future studies if the effective resolution of the analyzed winds can be increased, by applying perturbations at smaller scales than the original resolution of the ECMWF winds.

Acknowledgments

A   Analysis update for non-linear observation operators

The background estimate $\mathbf{x}^b$ and the observations $\mathbf{y}^o$ (involving a possibly non-linear observation operator $h(\cdot)$) are assumed to be an unbiased estimate of the true state $\mathbf{x}^t$.

$$E\left[\mathbf{x}^b\right] = \mathbf{x}^t \tag{14}$$

$$E\left[\mathbf{y}^o\right] = h(\mathbf{x}^t) \tag{15}$$

The error of the background estimate $\mathbf{x}^b$ and the observations $\mathbf{y}^o$ are assumed to be independent. One seeks an analysis scheme of the following form,

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - h(\mathbf{x}^b)) \tag{16}$$

where $\mathbf{K}$ is a matrix to be defined by requiring that the analysis has a minimal error. On average, the analysis should be unbiased.

$$E\left[\mathbf{x}^a\right] = \mathbf{x}^t + \mathbf{K}(h(\mathbf{x}^t) - E\left[h(\mathbf{x}^b)\right]) \tag{17}$$

This is the case if:

$$E\left[h(\mathbf{x}^b)\right] = h(\mathbf{x}^t) \tag{18}$$

The error of the analysis is given by:

$$\mathbf{x}^a - \mathbf{x}^t = \mathbf{x}^b - \mathbf{x}^t + \mathbf{K}(\mathbf{y}^o - h(\mathbf{x}^t)) - \mathbf{K}(h(\mathbf{x}^t) - h(\mathbf{x}^b)) \tag{19}$$

Since the background estimate is independent from the observations, $h(\mathbf{x}^b)$ is also independent from the observations $(\mathrm{cov}(\mathbf{y}^o, h(\mathbf{x}^b)) = 0)$. The error covariance of the analysis yields:

$$\begin{aligned}
\mathbf{P}^a &= \mathbf{P}^b - \mathbf{K}\mathrm{cov}(h(\mathbf{x}^b), \mathbf{x}^b) - \mathrm{cov}(\mathbf{x}^b, h(\mathbf{x}^b))\mathbf{K}^T \\
&\quad + \mathbf{K}\mathrm{cov}(h(\mathbf{x}^b), h(\mathbf{x}^b))\mathbf{E}^T + \mathbf{K}\mathbf{R}\mathbf{K}^T
\end{aligned} \tag{20}$$

Covariances are approximated by the ensemble covariances:

$$\mathbf{P}^a = \mathbf{P}^b - \mathbf{K}\mathbf{E}\mathbf{S}^T - \mathbf{S}\mathbf{E}^T\mathbf{K} + \mathbf{K}\left(\mathbf{E}\mathbf{E}^T + \mathbf{R}\right)\mathbf{K}^T \tag{21}$$

16

The optimal Kalman gain $\mathbf{K}$ is obtained by minimizing the total error $\mathrm{tr}(\mathbf{WP}^a)$ where $\mathbf{W}$ is an arbitrary weighting matrix.

$$J(\mathbf{K}) = \mathrm{tr}(\mathbf{WP}^a) \tag{22}$$

The minimum is obtained by:

$$
\begin{aligned}
\delta J &= J(\mathbf{K} + \delta\mathbf{K}) - J(\mathbf{K}) \\
&= -2\mathrm{tr}(\mathbf{W}\delta\mathbf{K}\mathbf{E}\mathbf{S}^T) + 2\mathrm{tr}(\mathbf{W}\delta\mathbf{K}(\mathbf{E}\mathbf{E}^T + \mathbf{R})\mathbf{K}^T) = 0
\end{aligned}
\tag{23}
$$

δJ is zero for any $\delta\mathbf{K}$ if,

$$\mathbf{K} = \mathbf{S}\mathbf{E}^T(\mathbf{E}\mathbf{E}^T + \mathbf{R})^{-1} \tag{24}$$

In summary, the analysis update with a non-linear observation operator is not guaranteed to be unbiased. The condition of equation (18) has thus to be verified separately. The Kalman gain can then be derived without additional assumptions (compared to the case of a linear observation operator).

In the context of the Extended Kalman Filter, a non-linear observation operator $h(\cdot)$ is linearized around the model forecast. Such linearization is not necessary here because the covariance matrices in the Extended Kalman Filter $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ and $\mathbf{P}^b\mathbf{H}^T$ are directly derived from the ensemble $(\mathrm{cov}(h(\mathbf{x}^b), h(\mathbf{x}^b))$ and $\mathrm{cov}(\mathbf{x}^b, h(\mathbf{x}^b)))$.

References

Aksoy A, Zhang F, Nielsen-Gammon JW (2006) Ensemble-based simultaneous state and parameter estimation with MM5. Geophysical Research Letters 33:L12,801

Anderson JL (2001) An Ensemble Adjustment Filter for Data Assimilation. Monthly Weather Review 129:2884–2903

Annan J, Lunt D, Hargreaves J, Valdes P (2005) Parameter estimation in an atmospheric GCM using the ensemble Kalman filter. Nonlinear Processes Geophysics 12:363–371

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

Barth A, Alvera-Azcárate A, Beckers JM, Rixen M, Vandenbulcke L (2007) Multigrid state vector for data assimilation in a two-way nested model of the Ligurian Sea. Journal of Marine Systems 65(1-4):41–59, DOI 10.1016/j.jmarsys.2005.07.006, URL http://hdl.handle.net/2268/4260

Barth A, Alvera-Azcárate A, Weisberg RH (2008) Assimilation of high-frequency radar currents in a nested model of the West Florida Shelf. Journal of Geophysical Research 113:C08,033, DOI 10.1029/2007JC004585, URL http://hdl.handle.net/2268/26171

Barth A, Alvera-Azcárate A, Gurgel KW, Staneva J, Port A, Beckers JM, Stanev EV (2010) Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of High-Frequency radar surface currents - application to the German Bight. Ocean Science 6(1):161–178, DOI 10.5194/os-6-161-2010, URL http://www.ocean-sci.net/6/161/2010/

Béal D, Brasseur P, Brankart JM, Ourmières Y, Verron J (2010) Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. Ocean Science 6:247–262, DOI 10.5194/os-6-247-2010

Brankart JM, Testut CE, Brasseur P, Verron J (2003) Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models: application to a 1993-96 hindcast of the North Atlantic Ocean circulation. Journal of Geophysical Research 108(C3):3074, DOI 10.1029/2001JC001198

Breivik O, Satra O (2001) Real time assimilation of HF radar currents into a coastal ocean model . Journal of Marine Systems 3-4:161–182

Burchard H, Bolding K (2002) GETM – a general estuarine transport model. Scientific documentation. Tech. Rep. EUR 20253 EN, European Commission

Chen Y, Snyder C (2007) Assimilating Vortex Position with an Ensemble Kalman Filter. Monthly Weather Review 135:1828–1845

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research 99(C5):10,143–10,162

Evensen G (2003) The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics 53:343–367, DOI 10.1007/s10236-003-0036-9

Evensen G (2007) Data assimilation: the Ensemble Kalman Filter. Springer, 279pp

Gelb A (1974) Applied optimal estimation. MIT Press, Cambridge, MA, 374 pp.

Greenwald TJ, Hertenstein R, Vukićević T (2002) An all-weather observational operator for radiance data assimilation with mesoscale forecast models. Monthly Weather Review 130:1882–1897

Hamill TM, Whitaker JS, Snyder C (2001) Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthy Weather Review 129:2776–2790

He R, Liu Y, Weisberg RH (2004) Coastal ocean wind fields gauged against the performance of an ocean circulation model. Geophysical Research Letters 31:L14,303, doi:10.1029/2003GL019261

Hoteit I, Cornuelle BD, Kim SY, Forget G, Köhl A, Terrill EJ (2009) 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.2008.11.005

Hunt BR, Kalnay E, Kostelich EJ, Ott E, Patil DJ, Sauer T, Szunyogh I, Yorke JA, Zimin AV (2004) Four-dimensional ensemble Kalman filtering. Tellus 56A:273–277

Hunt BR, Kostelich EJ, Szunyogh I (2007) Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D 230:112–126

Kundu PK (1976) Ekman veering observed near the ocean bottom. Journal of Physical Oceanography 6:238–242

19

Kurapov AL, Egbert GD, Allen JS, Miller RN, Erofeeva SY, Kosro P (2003) The M2 internal tide off Oregon: inferences from data assimilation. Journal of Physical Oceanography pp 1733–1757

van Leeuwen PJ (2001) An Ensemble Smoother with Error Estimates. Monthly Weather Review 129:709–728

Lewis JK, Shulman I, Blumberg AF (1998) Assimilation of CODAR observations into ocean models. Continental Shelf Research 18:541–559

Ljung L (1979) Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. IEEE Trans Automat Control 24:36–50

Losa SN, Kivman GA, Schroter J, Wenzel M (2003) Sequential weak constraint parameter estimation in an ecosystem model. Journal of Marine Systems 43:31–49

Lucas MA, Ayoub N, Barnier B, Penduff T, de Mey P (2008) Stochastic study of the temperature response of the upper ocean to uncertainties in the atmospheric forcing in an Atlantic OGCM. Ocean Modelling 20:90–113, DOI 10.1016/j.ocemod.2007.07.006

Malanotte-Rizzoli P, Young RE, Haidvogel DB (1989) Initialization and data assimilation experiments with a primitive equation model. Dynamics of Atmospheres and Oceans 13:349–378

Menemenlis D, Fieguth P, Wunsch C, Willsky A (1997) Adaptation of a fast optimal interpolation algorithm to the mapping of oceanographic data. Journal of Geophysical Research 102(C5):10,573–10,584

Mourre B, Ballabrera-Poy J, Garcia-Ladona E, Font J (2008) Surface salinity response to changes in the model parameters and forcings in a climatological simulation of the eastern North-Atlantic Ocean. Ocean Modelling 23:21–32

Oke PR, Allen JS, Miller RN, Egbert GD, Kosro PM (2002) Assimilation of surface velocity data into a primitive equation coastal ocean model. Journal of Geophysical Research 107(C9):5–1–5–25

Paduan JD, Shulman I (2004) HF radar data assimilation in the Monterey Bay area. Journal of Geophysical Research 109:C07S09

Pham DT, Verron J, Roubaud MC (1998) A singular evolutive extended Kalman filter for data assimilation in oceanography. Journal of Marine Systems 16(3–4):323–340

PRISMA (1994) Prozesse im Schadstoffkreislauf Meer-Atmosphäre: Ökosystem Deutsche Bucht. BMFT-Projekt 03F0558A1 (1.1.1990 - 31.10.1993). Abschlussbericht, ZMK-Universität Hamburg

Sakov P, Evensen G, Bertino L (2010) Asynchronous data assimilation with the EnKF. Tellus 62A:24–29

Shulman I, Paduan JD (2009) Assimilation of HF radar-derived radials and total currents in the Monterey Bay area. Deep Sea Research Part II: Topical Studies in Oceanography 56(3-5):149–160, DOI 10.1016/j.dsr2.2008.08.004

Skachko S, Brankart JM, Castruccio F, Brasseur P, Verron J (2009) Improved turbulent air-sea flux bulk parameterization for controlling the response of the ocean mixed layer: a sequential assimilation approach. Journal of Atmospheric and Oceanic Technology 26:538–555

Skandrani C, Brankart JM, Ferry N, Verron J, Brasseur P, Barnier B (2009) Controlling atmospheric forcing parameters of global ocean models: sequential assimilation of sea surface Mercator-Ocean reanalysis data. Ocean Science 5:1–15

Staneva J, Stanev EV, Wolff JO, Badewien TH, Reuter R, Flemming B, Bartholomä A, Bolding K (2009) Hydrodynamics and sediment dynamics in the German Bight. A focus on observations and numerical modelling in the East Frisian Wadden Sea. Continental Shelf Research 29:302–319

Talagrand O (1972) On the damping of high-frequency motions in four-dimensional assimilation of meteorological data. Journal of the Atmospheric Sciences pp 1571–1547

Vandenbulcke L, Rixen M, Beckers JM, Alvera-Azcárate A, Barth A (2008) An analysis of the error space of a high-resolution implementation of the GHER hydrodynamic model in the Mediterranean Sea.

Ocean Modelling 24(1-2):46–64, DOI 10.1016/j.ocemod.2008.05.007, URL http://hdl.handle.net/2268/25529

Wilkin JL, Arango HG, Haidvogel DB, Lichtenwalner CS, Glenn SM, Hedström KS (2005) A regional ocean modeling system for the Long-term Ecosystem Observatory. Journal of Geophysical Research 110:C06S91
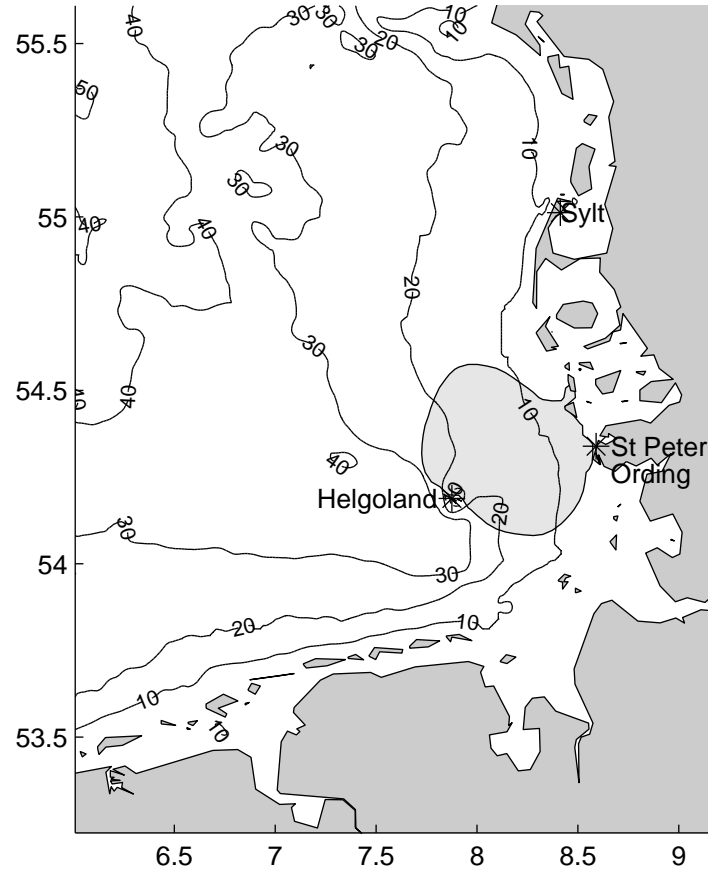
Figure 1: Model domain and bathymetry. The light gray shading area corresponds to the zone where HF radar surface currents are available more than 50 % of the time.

Table 1: Comparison with *in situ* wind measurements. RMS differences and S_{HF} are expressed in m/s. The mean square error skill score is defined by equation (13)

	S_{HF}	Helgoland		Sylt	
		RMS	skill score	RMS	skill score
Free	–	2.40	0.00	1.98	0.00
Analysis	0.5	2.14	0.21	1.96	0.03
	1.0	2.02	0.29	1.82	0.15
	1.5	1.93	0.35	1.64	0.31
	2.0	1.89	0.38	1.50	0.43
	2.5	1.88	0.39	1.43	0.48
	5.0	2.05	0.27	1.54	0.39

.

Table 2: Comparison with satellite SST. RMS difference is expressed in °C and S_{HF} in m/s.

	S_{HF}	RMS	skill score
Free	–	1.21	0.00
Analysis	0.5	1.09	0.19
	1.0	1.09	0.19
	1.5	1.10	0.18
	2.0	1.11	0.16
	2.5	1.12	0.14
	5.0	1.16	0.08

Table 3: Results of the assimilation experiment of different assimilation window lengths.

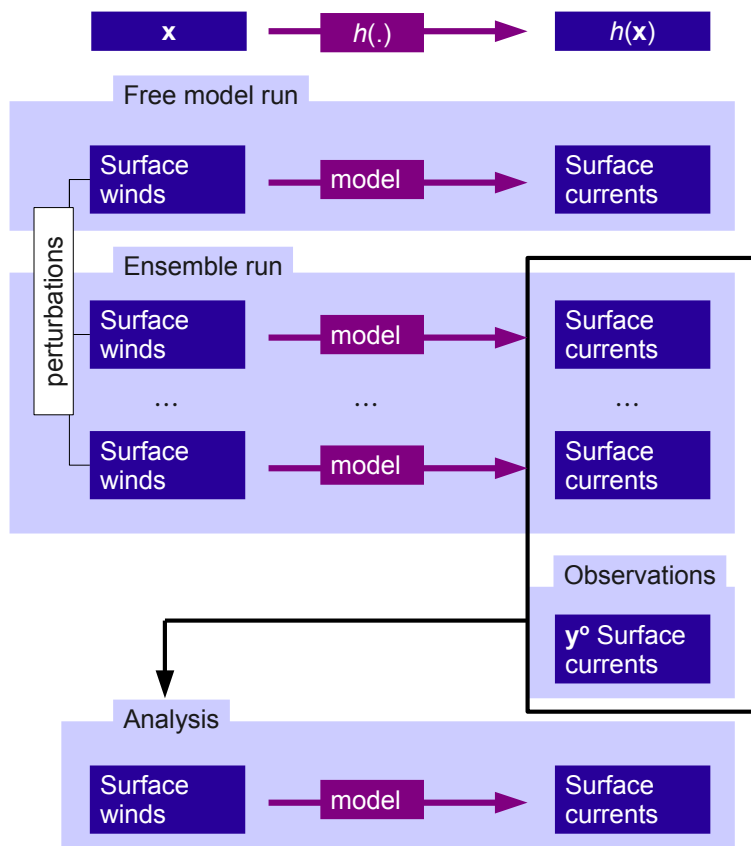| Length	σ_w	σ_a	RMS at Helgoland	RMS at Sylt
0.5	3.04	3.59	1.87	1.40
1	2.92	4.09	1.89	1.50
2	2.87	4.38	2.12	1.71
4	2.87	4.78	2.16	1.66
ECMWF	–	–	2.40	1.98

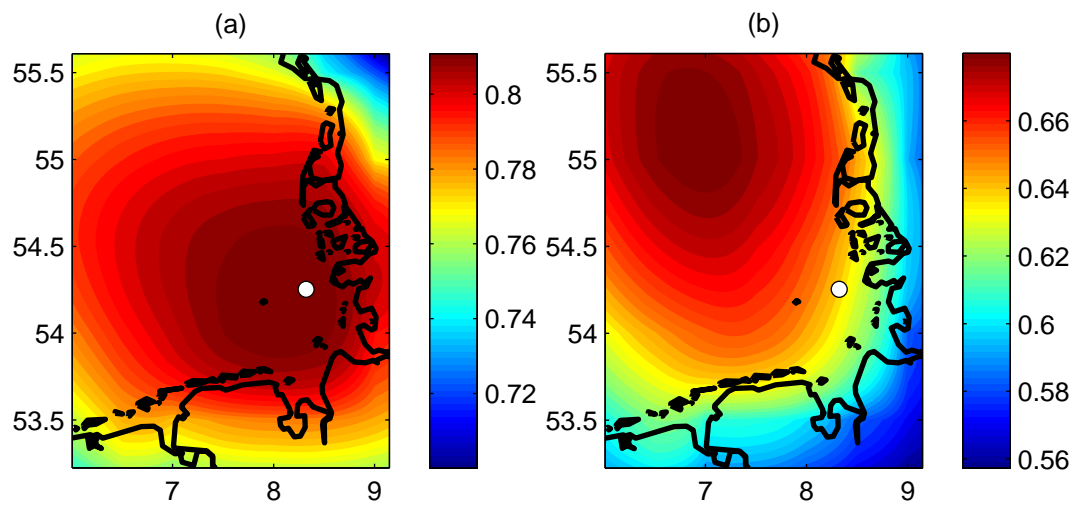Figure 2: Diagram of the assimilation scheme.

Figure 3: Panel (a): error correlation between the u-component of the surface currents at the location of the white marker and the u-component of wind fields over the model domain (17 September 1991). Panel (b): item for the v-components.
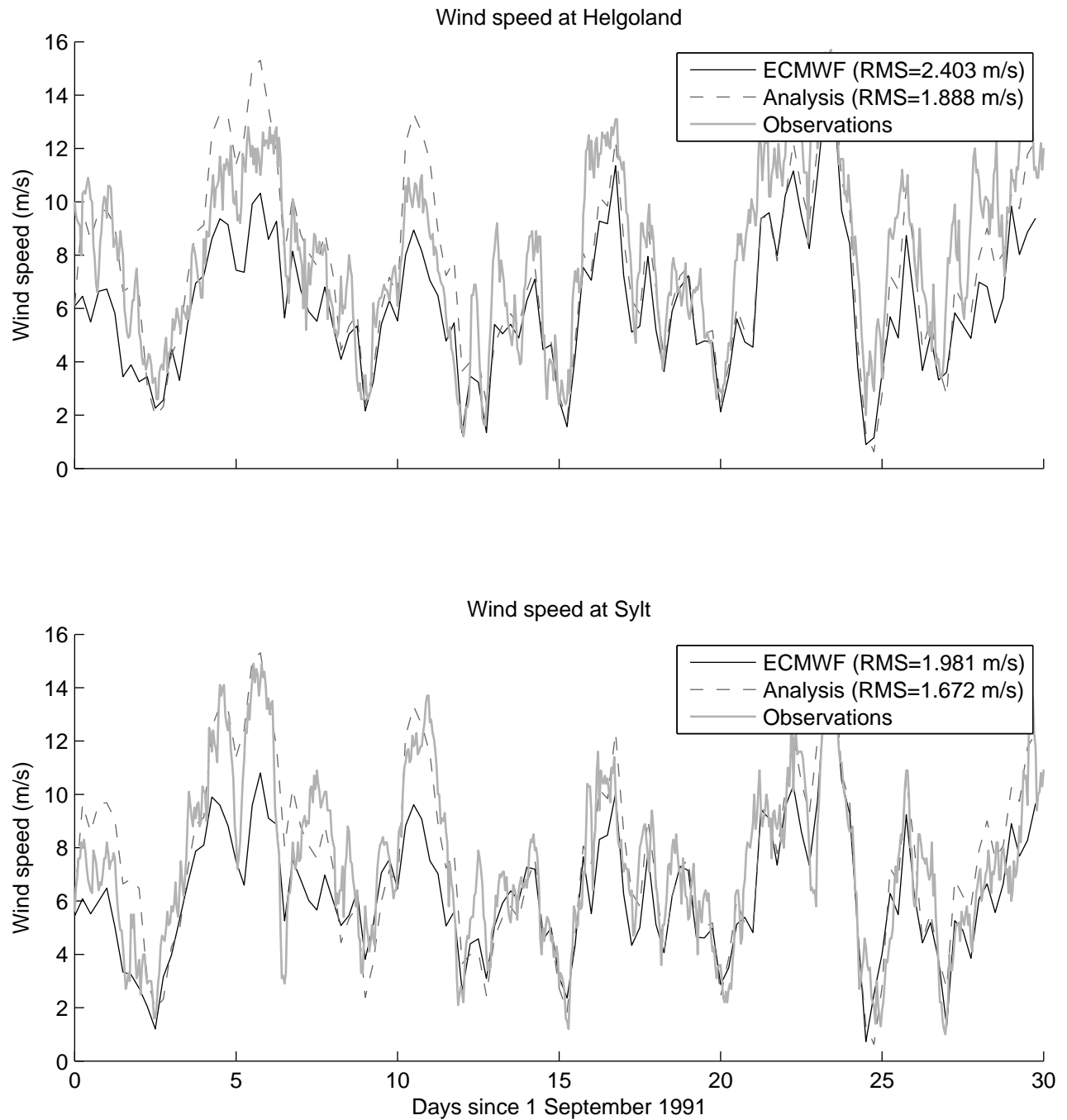
Figure 4: Measured wind speed, wind speed from ECMWF and analyzed wind speed at Helgoland and Sylt. Units are m/s.
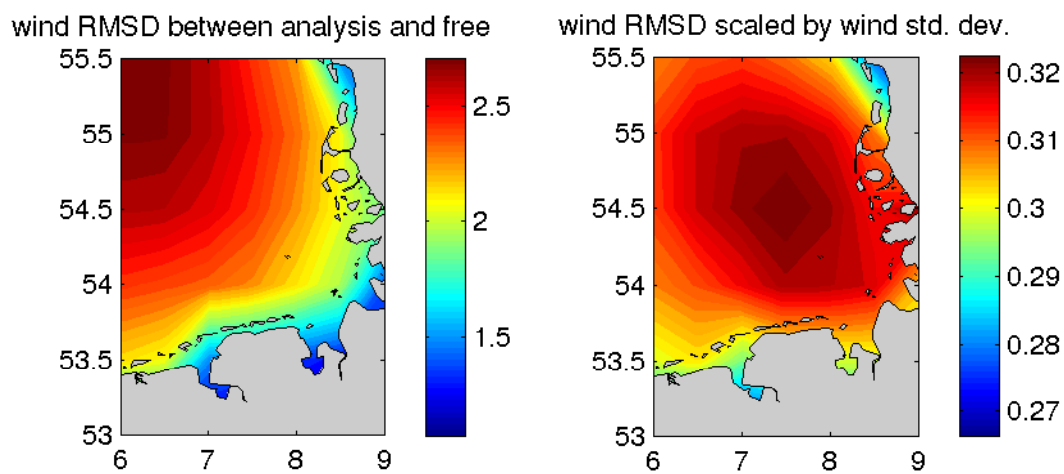
Figure 5: Left panel: RMS difference between analyzed winds and ECMWF winds (averaged over time). Right panel: RMS difference scaled by wind standard deviation
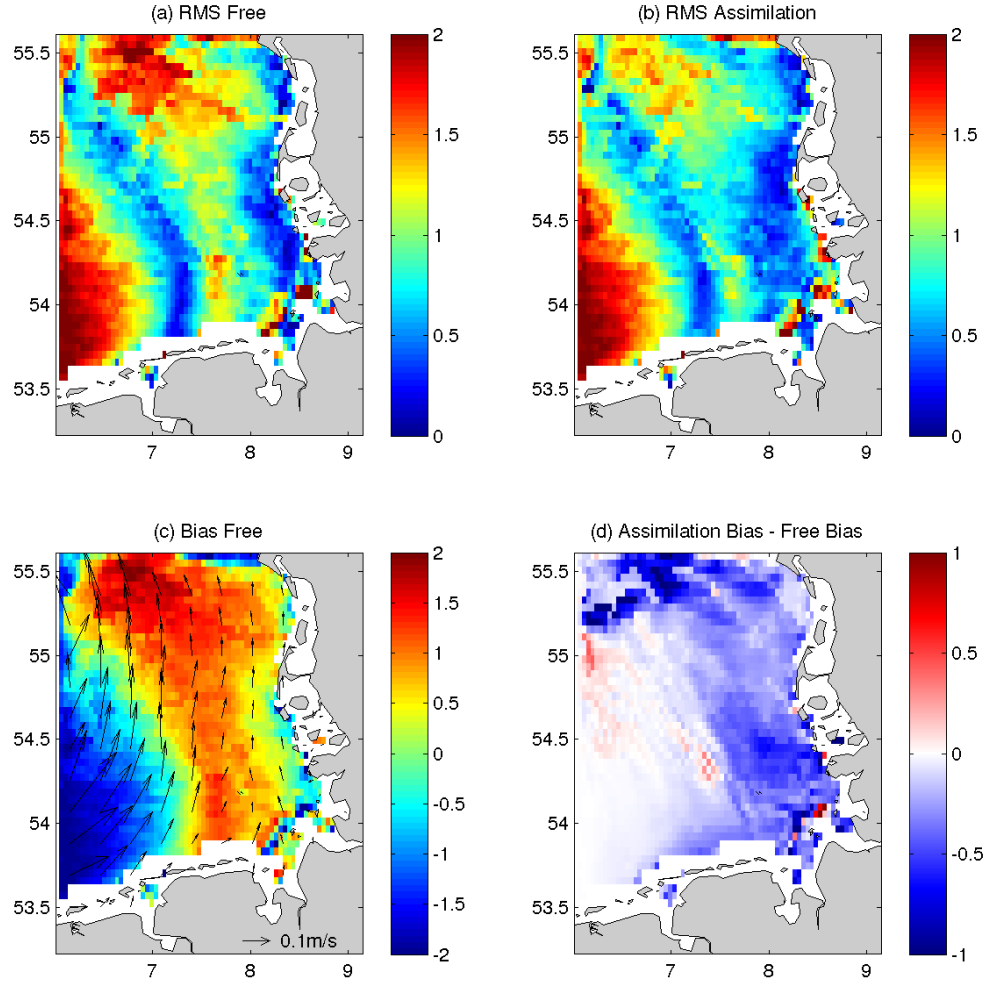
Figure 6: RMS difference between AVHRR SST and model SST without assimilation (panel a) and with assimilation (panel b), bias (averaged in time) of the model SST without assimilation and the averaged surface currents (panel c) and the change of model bias due to the assimilation (panel d)