Die Küste, 81 (2014), 133-151

Methods of Data Assimilation

Emil Stanev and Johannes Schulz-Stellenfleth

Summary

An overview of ocean forecasting techniques amalgamating numerical models, observations and data assimilation methods is presented. The basics of data assimilation as an application of estimation theory or control theory is described and the corresponding statistical and numerical methods are introduced. Classical approaches like Kalman filter or optimal interpolation are explained as well as state of the art methods like reduced rank filters and smoother approaches. Problems and challenges of coastal ocean forecasting are identified, which are associated with the specific variables of interest for coastal applications, such as: complex physics complicating the assimilation of data; characteristic time scales; vigorous adjustment process arising in sequential data assimilation, when models are restarted; specific data and observational platforms in coastal ocean and maximising the outcome of synergies between different data types; model and observation error specification; coupling coastal and deep ocean models and seamless transition between coastal and open-ocean scales. Illustrations of some of the above challenges and their treatment in the area of the German Bight are given by describing a pre-operational HF radar data assimilation system using three WERA stations, as well as an assimilation system using FerryBox surface temperate and salinity measurements.

Keywords

Kalman filter, variational data assimilation, smoothers, coastal ocean forecasting, model and observation errors, German Bight

Zusammenfassung

Es wird eine Übersicht über Vorhersagemethoden in der Ozeanographie, die numerische Modelle, Beobachtungen und Datenassimilation verbinden, gegeben. Die Grundlagen der Datenassimilation als eine Anwendung der Schätz- und Kontrolltheorie werden beschrieben und die zugehörigen statistischen und numerischen Methoden eingeführt. Klassische Verfahren wie der Kalman Filter oder Optimale Interpolation werden ebenso wie neuartige Ansätze wie der „reduced rank filter" oder „smoother" erläutert. Probleme und Herausforderungen werden angesprochen, die charakteristisch sind für Vorhersagen im Küstenbereich. Dazu gehören die folgenden Punkte: komplexe Physik, die die Datenassimilation erschwert; charakteristische Zeitskalen; starke Schockeffekte bei der sequentiellen Datenassimilation, wenn Modelle neu gestartet werden; spezielle Daten- und Beobachtungsplattformen im Küstenbereich und optimale Nutzung von Synergien zwischen verschiedenen Daten; schwierige Spezifizierung von Modell- und Beobachtungsfehlern; Kopplung von Modellen für den Küsten- und den Tiefwasserbereich und der nahtlose Übergang zwischen den verschiedenen räumlichen Skalen. Einige dieser Herausforderungen und ihre Behandlung werden für den Bereich der Deutschen Bucht veranschaulicht durch die Beschreibung eines präoperationellen HF Radar Datenassimilationssystems, das drei WERA Stationen verwendet, sowie ein

Assimilationssystem, das FerryBox Messungen von Temperatur und Salzgehalt an der Meeresoberfläche benutzt.

Schlagwörter

Kalman-Filter, 4D-VAR, Smoother, Vorhersagen im Küstenbereich, Modell- und Beobachtungsfehler, Deutsche Bucht

Contents

1 Introduction

Ocean data assimilation combines observations and models with the aim to (1) improve our understanding of ocean circulation on all relevant temporal and spatial scales, (2) monitor and (3) predict the ocean state. This combination of observations and numerical simulations enables efficient, accurate and realistic state estimations. Furthermore, for sampling rates, spatial domains and time intervals of practical interest, state estimates might not otherwise be feasible, because data acquisition in the ocean is difficult and costly in particular if done on a sustained basis. Thus, data assimilation can provide four-dimensional time series of dynamically adjusted fields, which can serve as high resolution and complete data sets (mix between observations and modelling) for dynamical studies. This goes beyond the classical methodology, where observations and theory were developing independently, or in the best case, when used in parallel, their synergy was not well exploited.

The challenge in data assimilation is to extract the most important information from relatively sparse and noisy observations, and to feed this information in an optimal way into numerical forecast models. The observation errors are due to instrumental noise, environmental noise, sampling, and possible misinterpretation of measurements. Numerical ocean models are obviously also not error-free; errors are originating from incomplete (non-perfect) model physics, insufficient grid resolution, problems in open boundary conditions, atmospheric or hydrological forcing. Even "perfect" ocean models will drift away from reality, which is known as loss of predictability beyond the predictability limit. This limit depends upon the type of geophysical fluids and dominating processes. For synoptic processes in the ocean it is of the order of weeks to months, for the coastal ocean it is of the order of hours and days. The loss of predictability is associated with the nonlinear transfer and growth of errors.

The above fundamental characteristics of geophysical systems inhibits accurate forecasts and calls for periodically correcting their evolution using observations with appropriate spatial coverage and at intervals less than the predictability limit. This correction process necessitates knowledge of multivariate aspects of model and observational data to be combined. A fundamental problem in data assimilation is the specification of the background and observation error covariance matrices, which determines how information is passed from the observations to the model. If this information is complete and used correctly, observations of one model variable can produce dynamically consistent corrections in other model variables or areas where no observations exist.

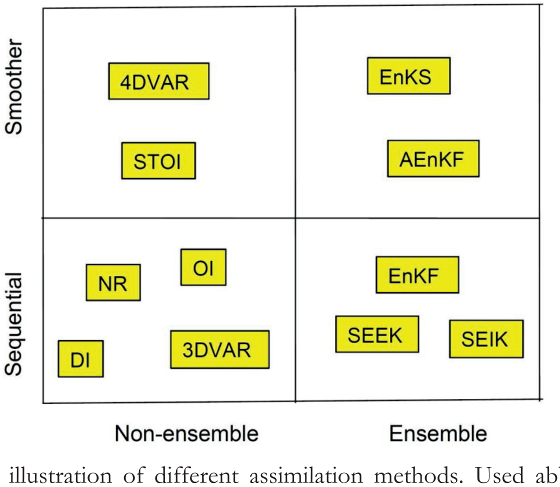

Figure 1: Schematic illustration of different assimilation methods. Used abbreviations are explained in Section 2. Direct insertion (DI) and Newtonean Relaxation (NR) methods are added to the bottom-left box although they are not typical representatives of sequential methods used nowadays.

During the last years ocean data assimilation and forecasting has reached an impressive level of maturity (CHASIGNET and VERRON 2006). One good example is the Global Ocean Data Assimilation Experiment (GODAE), where several systems have been developed and are operated by the Australian Bureau of Meteorology (BLUElink Ocean Data Assimilation System, BODAS), the Jet Propulsion Laboratory (Estimating the Circulation and Climate of the Ocean, ECCO), the UK Met Office (Forecast Ocean Assimilation Model, FOAM), the MyOcean system operated in several EU countries (Nucleus for European Modeling of the Ocean VARiational data assimilation, NEMOVAR), and some others (CUMMINGS et al. 2009). These GODAE systems assimilate various measurements such as sea level anomaly data provided by satellite altimeters, subsurface temperature and salinity data from Argo floats, moored and drifting buoys, expendable bathythermograph (XBT), and conductivity-temperature-depth (CTD) recorders, in situ and satellite sea surface temperature data, as well as satellite derived sea ice concentration and drift data.

So far ocean data assimilation techniques have been applied for operational forecasts, error analysis, parameter optimization, for ocean process studies, and observational network design. The latter, known as Observation System Simulation Experiments (OSSEs), enables the optimization of future experimental or operational monitoring networks. Compared to the methodologies used in meteorology and global oceanography, coastal forecasting techniques are still at a relatively early stage of development. This is because the specific problems of ocean data assimilation in coastal ocean are full of challenges, which are not sufficiently addressed in global or regional ocean data assimilation. This motivates us to present in the present study the basics of data assimilation, identify specific coastal problems and illustrate solutions of some of them. The area of applications used to illustrate different approaches is the German Bight, which is a shallow, tidally driven part of the southern North Sea.

The present paper describes in section 2 the theoretical background of ocean data assimilation. In Section 3 examples addressing applications for the German Bight area are demonstrated followed by short conclusions in section 4.

2 Methods

2.1 Basic Concepts

Ocean data assimilation deals with the spatial distribution and temporal evolution of state variables (e.g. velocity, pressure, density, temperature and salinity), that is with the state estimation in three dimensions and time. Ocean forecasting systems provide the future ocean state given its state at an initial time. The tool for such predictions is a dynamical (circulation) model. This is *the first component of prediction systems,* which numerically approximates a set of prognostic field equations for state variables. These equations include parameters (e.g., associated with unresolved physical processes or fundamental constants) and are initialized and forced by given initial and boundary condition data, respectively. Dynamical models used in data assimilation have to correctly represent internal dynamics and the ocean response to external forcing. i.e., their performance has to statistically agree with the statistics of real processes.

The second component of prediction systems is a data assimilation tool used to link the state variables of the dynamical model to the observations. Such a tool has to work as dynamical interpolation and extrapolation of the data and to combine observations and simulations with weights inversely proportional to their relative errors. Data assimilation uses statistical estimation theory or control theory (Fig. 1). One classical approach belonging to the first class is the Kalman Filter (KF), which is based on a probability maximization. The control theory deals with the behavior of dynamical systems where one or more output variables of a system need to follow a certain reference over time, therefore a manipulation is applied to the inputs to obtain the desired effect on the output. The 4DVAR technique is the most established method in this category. *The third component of prediction systems* consists of the used observational networks, which have to be adequate to capture dominating processes and to possess the required accuracies. In this paper we will concentrate on the second component of prediction systems.

We will denote in the following the state variables at time step k by a vector x_k of dimension n. The observations of dimension m at time k are denoted by y_k^O. We will assume that the evolution of the state variables is described through a dynamical (forecasting) model

$$x_{k+1}^f = Mx_k^a + \eta_k \tag{1}$$

where the index "f" stays for forecasting, M is a $n \times n$ matrix, which corresponds to the discrete scheme associated with a given numerical model, and η_k is Gaussian noise with covariance matrix Q_k.

The relation between the state variables and the measurements are described using a linear model

$$y_k^O = Hx_k + \in_k \tag{2}$$

where H is the observation operator and $\in_k$ is Gaussian noise with covariance R, which is often assumed as white. Observation errors consist of instrument noise and so called representation errors, which are model dependent. The magnitude of the observation error, in combination with the numerical model error, determines the relative weight given to the observation. The model errors are often specified by the error covariance matrix P.

In the following different approaches are described to find a so called analyzed state x^a, which represents an optimal combination of a first guess model state x^f and measurements y^O. Most of the approaches (Fig. 1) can be formulated as a cost function minimization problem, where on the one hand the mismatch between model and observations (the so called innovation) is reduced and on the other the departure from the first guess is not too big.

2.2 Direct Insertion and Newtonian Relaxation

The simplest assimilation approach, called Direct Insertion (DI), replaces in the observation points the forecast values by the observed ones. Because this can only be done if the data are consistent, which is hardly ever the case, this method is sub-optimal. In comparison, the Nudging or Newtonian Relaxation (NR) scheme introduces in the prognostic equations terms proportional to the difference between the data and state variables (i.e., data residuals). This method “relaxes” the model towards the observations. The relaxation times should be consistent with the dominating time-scales, but cannot be too small to avoid model disruptions.

2.3 Sequential Approaches

In sequential data assimilation methods (BRASSEUR 2006) repeated forecast analysis cycles are performed, where at each analysis time step a new initial model state for the next forecast is computed based on the model state and the observations available at that time. In the following different approaches falling into that category are described.

2.3.1 Optimal Interpolation

Optimal interpolation (OI) (GANDIN 1963; LORENC 1981; DALEY 1991) is a simplification of the KF, where the forecast error covariance is replaced by the background error covariance. In simpler implementations the weights used in the filter are empirically assigned. The fundamental hypothesis of this method is that *for each model variable, only a few observations are important to determining the analysis increment.* P-matrix specification usually relies on the design of empirical auto-correlation functions (e.g. Gaussian), which are considered as time-independent. In the OI method significant weight is assigned to those observations which have significant background error covariance. In practice, the correlation radius limits the geometrical domain around model variables, which need to be considered.

2.3.2 Kalman Filter

The estimation of state vector can be formulated as the maximization of an *a posteriori* probability of the system state for given observations y_k^0 and for a given first guess model state x_k^f. Estimation theory (e.g., GELB 1974) states that the so called analysis x_k^a, which is the optimal combination of the model and observation is given by

$$x_k^a = x_k^f + K_k\left(y_k^o - Hx_k^f\right) \tag{3}$$

where

$$K_k = P_k^f H^T\left(HP_k^f H^T + \mathrm{R}\right)^{-1} \tag{4}$$

is the Kalman gain matrix. Thus, the filter can be considered as a two step process: (i) the forecast of the state vector and of its error covariance are computed as

$$x_k^f = Mx_{k-1}^a \tag{5}$$

$$P_k^f = MP_{k-1}^a M^T + Q_k \tag{6}$$

(ii) The analysis is derived from equations (3) and (4) and finally (iii) the a posteriori covariance is computed as

$$P_a^k = \left(I - K_k H\right) P_k^f \tag{7}$$

The analysis step is a linear combination of the dynamical forecast x_k^f with the difference between the data and model predictions $y_k^a - Hx_k^f$, which is called data residual.

In conclusion, the KF (KALMAN 1960) is a simplification of Bayesian estimation for the case of linear systems, i.e, it is only optimal for linear models. Linearization of the model around the state estimate leads to the so-called extended Kalman (EK) filter (JAZWINSKI 1970).

A KF analysis is obtained when errors associated with forecasts and observations are known and accurately specified. Since these statistics are not generally available, actual implementations of assimilation algorithms are always sub-optimal (DEE and DA SILVA 1998).

2.3.3 Ensemble techniques

A further variant of the Kalman filter based on ensemble techniques and Monte-Carlo methods was developed to avoid the linearization of the model required in the KF. This method known as ensemble KF (EnKF; EVENSEN 1994; BURGERS et al. 1998) uses an ensemble of model states to represent the error statistics given in the EK filter (the covariance is approximated by sample covariances). The estimation of the flow-dependent background error covariance makes this method a good alternative to the variational approaches (see section 2.3.4).

2.3.4 State and error sub-spaces reduction

A full KF cannot be implemented into realistic ocean models, because error forecast and analysis equations are too demanding (CPU and memory requirements are too high). On the other side OI over-simplifies the propagation of errors. In the filter proposed by CANE et al. (1996) the state space is reduced through the projection onto a linear subspace of basic functions using a limited number of empirical orthogonal functions (EOF).

Another approach to reduce the computational costs is based on a low-rank approximation of the state covariance matrix. Examples of low-rank filters are the reduced rank square-root (RRSQRT) algorithm (VERLAAN and HEEMINK 1995) and the singular evolutive extended Kalman (SEEK) filter (PHAM et al. 1998). In the SEEK filter the error covariance matrix is approximated by a singular low rank matrix. In practice the SEEK filter corrects the forecast in the directions for which the error is not sufficiently attenuated by the dynamics of the model. These 'directions of correction' evolve with time according to the model evolution. For an improved treatment of non-linear error evolution, the singular evolutive interpolated Kalman (SEIK) filter (PHAM et al. 1998) was introduced as a variant of the SEEK filter. It combines the low-rank approximation with an ensemble representation of the covariance matrix.

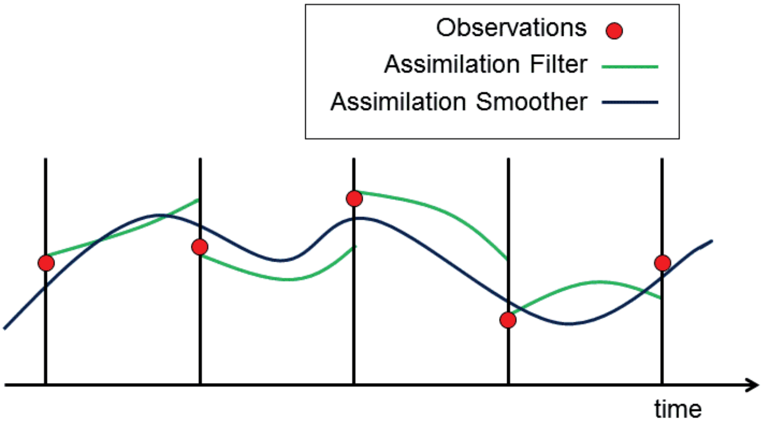

Figure 2: Trajectories of the model state during the data assimilation process in variational methods (blue line) and sequential methods (green lines). Red dots symbolize observations.

2.3.5 The three-dimensional analysis (3D-VAR)

The 3D-VAR was developed for the first time by SASAKI (1958). It was introduced into operational global numerical weather prediction by LORENC et al. (2000); recent applications in ocean forecasting are described by DOBRICIC and PINARDI (2008). Like in the 4D-VAR approach introduced in section 2.4.2 the computation of the gain matrix K (Eq. 4) is avoided. However, unlike to the 4D-VAR case, all observations (in the time-window around the analysis time) are assigned to analysis time. The solution is sought iteratively by performing several evaluations of the cost function

$$J(x) = \left(x - x_k^f\right)^T B^{-1}\left(x - x_k^f\right) + \left(y_k^o - H[x]\right)^T R^{-1}\left(y_k^o - H[x]\right) \tag{8}$$

and of its gradient

$$\nabla J(x) = 2B^{-1}\left(x - x_k^f\right) - 2H^T R^{-1}\left(y_k^o - H[x]\right) \tag{9}$$

which is needed to find the minimum using a suitable descent algorithm. The estimated state x^a minimizing the cost function is then used as the initial state for the next forecast.

2.4 Smoother Approaches

2.4.1 General presentation

Important optimal control theory schemes (GHIL and MALANOTTE-RIZZOLI 1991) include the generalized inverse and adjoint methods. These methods, which are known as variational methods, seek to minimize the misfit between data and model trajectory over a given period. They were first developed by MARCHUK (1974) and further made popular in the environmental modelling by TALAGRAND and COURTIER (1987) and MOORE et al. (2011).

Variational methods assume that the analysis is at the initial time, and the individual observations at any given observation time t_i are distributed among n time steps over a given time interval. Like in the presentation of sequential methods we denote with y_i, x_i and x_{ti} the observations, the model and the true states at time t_i. The error covariance matrix for the observation errors $y_i - H(x_{ti}) - H(x_{ti})$ is R_i.

2.4.2 The four-dimensional analysis (4D-VAR)

The 4D-VAR minimizes the following cost function:

$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_{i=1}^{n} \left(y_i - H_i[x_i]\right)^T R_i^{-1} \left(y_i - H_i[x_i]\right) \tag{10}$$

This cost function, which is the sum of the of the squared deviations of the forecast fields and the analyzed fields weighted by the accuracy of the forecast (first term in the right hand side of 10) plus the sum of the squared deviations of the analysis values from the observations weighted by the accuracy of the observations (second term in the right hand side of 10), ensures that the analysis does not drift too far away from observations and forecasts. The significant advantage of the variational approaches is that the minimization problem is subject to the constraint that the sequence of model states x_i must be a solution of the model equations. Usually the gradient of the cost function is estimated using an adjoint model, which is quite demanding concerning the implementation and maintenance of the computer code. Furthermore, in a real-time forecasting the assimilation has “to wait” for the observations over the whole 4D-VAR time interval to be available. In sequential systems observations are used shortly after they become available.

2.4.3 Ensemble smoother: optimization of boundary conditions and meteorological forcing

A modification of the EnKF called ensemble Kalman smoother (EnKS) solves a smoothing problem (VAN LEEUWEN and EVENSEN 1996; EVENSEN and VAN LEEUWEN 2000; VAN LEEUWEN 2001; SAKOV et al. 2010). In order to estimate the ocean state at a certain time, it uses the data available before and after that time. Thus by propagating the future data information backward in time EnKS smooths the dynamical state ensuring a longer assimilation window, assimilating all collected observations in a single update and taking into account the evolution of the state and state error covariance over the length of the window.

BARTH et al. (2010; 2011) demonstrated the use of ensemble perturbation smoothers for optimizing tidal boundary conditions and correcting surface winds, respectively. In both cases assimilated data were HF radar surface currents in the German Bight. As stated in the previous section the situation is different from the open-ocean case because tides in coastal models 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 used smoother assumes that all observations within the model integration period are grouped into a single observation vector (y^o) with their corresponding error covariance (R). To create an ensemble of dynamically realistic boundary conditions (BARTH et al. 2009), a cost function is formulated, which is directly related to the probability of each boundary condition perturbation. This cost function ensures that the perturbations have a finite energy, are smooth and satisfy a linear constraint. The approach is closely related to the asynchronous EnKF (AEnKF, SAKOV et al. 2010), where model trajectories (i.e. the model results in space and time), instead of model states, are optimized. For an increasing number of ensemble members, the Ensemble Smoother does converge to the AEnKF.

3 Data assimilation in coastal ocean: German Bight examples

3.1 Specific problems of data assimilation in coastal ocean

Complexity of data assimilation in the coastal ocean increases with the vast range of phenomena and the multitude of interactive scales in space and time (DE MEY et al. 2009; DE MEY and PROCTOR 2009). The spatial and temporal resolution required for realistic coastal predictions is much higher than the resolution required for the deep ocean. Processes, which are sometimes disregarded in the open-ocean data assimilation such as tides and the high-frequency barotropic response to atmospheric forcing, become dominant in the coastal ocean. The small temporal scales (hours) and horizontal scales (hundreds of meters) are both computationally and scientifically challenging.

Most methods described in section 2 are presently used to assimilate data in coastal models. Their diversity reflects the complexity of coastal processes and the status of forecasting systems, which are still dealing with research issues. Efforts are however underway to test and improve quality of data assimilation in operational practices (STANEV et al. 2011).

Several problems associated with coastal data assimilation are addressed below.

1. *The variables of interest* for coastal applications include the same physical properties as in the open-ocean models, but also near bottom currents, which are important for sediment transport and a large number of biogeochemical properties. This greatly increases the number of variables and the complexity of models and the assimilation schemes. Short time scales (e.g., minutes to hours for tides) increase not only the demand for high-quality observations, but also for specific data assimilation schemes.
2. *Vigorous adjustment process* arises in sequential data assimilation, when they are 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). One approach to overcome this problem will be illustrated in following sections.
3. *Data and observational platforms* differ from the ones in the open ocean. For example, satellite altimetry does not fully resolve all important coastal-ocean scales; data from profiling floats are not available in the shelf seas. However, data from high frequency (HF) radars and ADCP, sea level from coastal tide gauges and bottom

pressure gauges, water properties from fixed data stations and ferries (Fig. 3), gliders, and AUVs gives new perspectives and challenges. In particular, the assimilation of altimetry must also account for the aliasing of the tidal signal, which can well be compensated by using the synergy between altimeter, tide gauge and HF radar data.

4. *Complex physics* in the coastal zone complicates the assimilation of data and necessitates resolving the whole spectrum of free-surface variations (tides, storm surges), multiple scales, friction and mixing effects and associated tidal straining and fronts, dependency of solution of small-scale bathymetric channels and variations of bathymetry (which is not well known), control of straits for the inter-basin exchange and inlets for the exchange between tidal flats and open ocean, drying and flooding. The situation is further complicated by complex nonlinear processes (e.g. creating of over-tides) and other complex coupling of the variability at different frequencies.

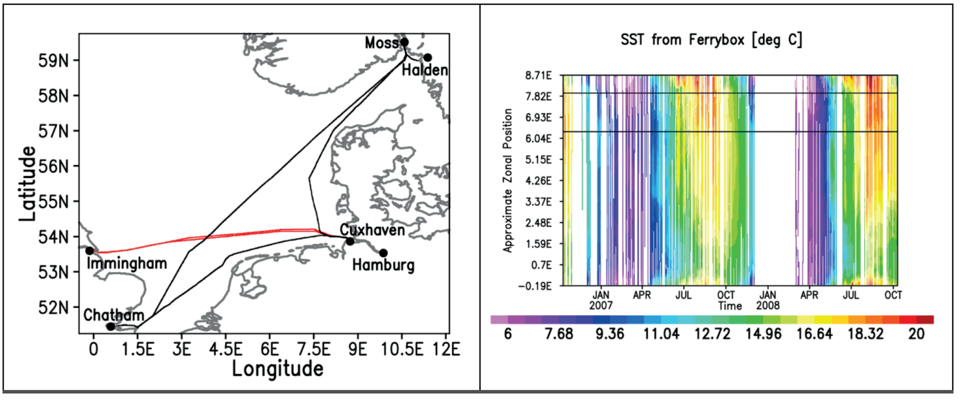

Figure 3: FerryBox routes in the North Sea with Cuxhaven – Immingham track in red (left) and FerryBox SST measurements (right) taken in 2007 and 2008 (from GRAYEK et al. 2011).

5. *Error specification* is extremely challenging in the coastal zone. Strong nonlinearities that couple variability at different frequencies (e.g., M2 tidal frequencies and lower frequency processes such as variation of stratification between neap and spring tides) necessitate using dynamically-consistent error prediction schemes. Furthermore, most existing assimilation schemes assume unbiased observations with Gaussian noise, which is often unrealistic. For many coastal observational platforms the determination of errors is also difficult, some platforms, e.g., satellite altimeters show larger errors in the coastal zone.

6. *Coupling coastal and deep ocean models* is still not a well-solved problem. Most coastal models are one-way nested; the model solution is strongly controlled by boundary forcing originating from larger-scale models. Two-way nested models enable that (assimilated) information from coastal observations, which is usually not assimilated by the larger-scale forecasting systems, is propagated out of the coastal region. These upscaling capabilities could become beneficial for regional models. One effective way to achieve a seamless transition is using unstructured-grid models representing much better the transition between coastal and open-ocean scales. In these models observations in the shelf sea would correct the deep ocean state and the deep ocean data would at the same time correct the coastal state. A good example of the potential of such capabilities tested in structured grid model is demonstrated

by STANEV et al. (2014). Available FerryBox and HF radar observations are used in the following sections as an enhancement of predictive modelling and to demonstrate solutions of some of problems described above.

3.2 FerryBox Sea Surface Temperature and Salinity Assimilation

In this section we will describe the assimilation of sea surface salinity (SSS) and surface temperature (SST) data acquired by a FerryBox system in a coastal area using a sequential filter approach. A FerryBox is an autonomous measurement, data logging and transmission system, which operates continuously while the carrying ship is on its way (PETERSEN et al. 2006). Measurements are made using devices, which are either in direct contact with or sample from a continuous flow of seawater taken at a water depth of 4–6 m. The basic sensors used in this study measure temperature, salinity, turbidity, and chlorophyll-a fluorescence.

The North Sea routes so far equipped with FerryBox systems are the ones between Büsum and Helgoland, Cuxhaven and Harwich, Cuxhaven and Immingham and recently between Hamburg, Cuxhaven, Chatham, Moss and Halden (Fig. 3 left). Depending on the travel distance, the routes provide the following revisit times: Büsum–Helgoland, daily, Cuxhaven–Immingham, less than 36 h, Hamburg–Cuxhaven–Chatham–Moss–Halden about 8 days. The route analyzed here is the one from Cuxhaven to Immingham (see the red line in Fig. 3 left) for the period 2007-2008.

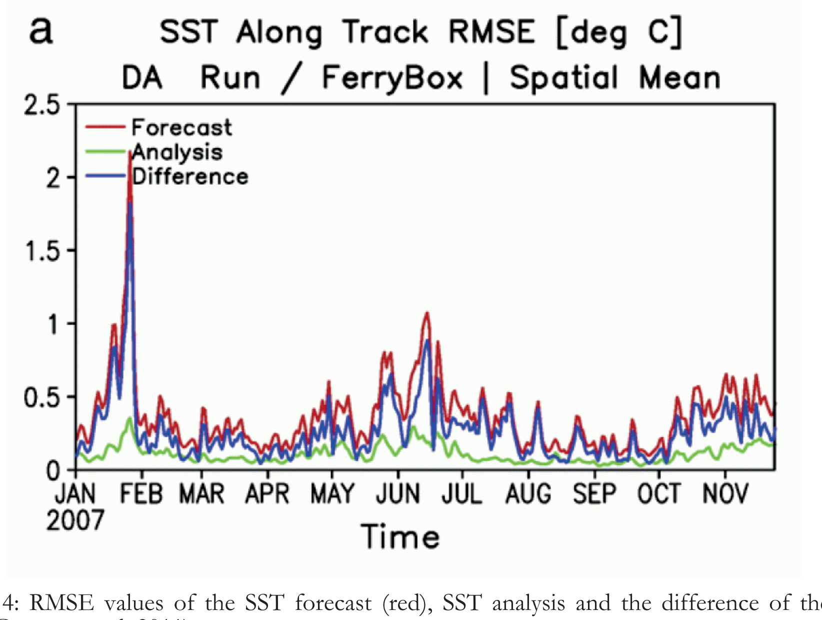

Figure 4: RMSE values of the SST forecast (red), SST analysis and the difference of the two (from GRAYEK et al. 2011).

The model used here is a 3-D primitive equation numerical model (BURCHARD and BOLDING 2002), in which the equations for the three velocity components u, v and w, and sea surface height ζ, as well as the equations for the turbulent kinetic energy and the eddy dissipation rate due to viscosity are solved. The application of the model to the German Bight (Fig. 3) is described in STANEVA et al. (2009).

The potential of FerryBox data for forecast skill improvements was analyzed in GRAYEK et al. (2011). As already pointed out there are many assimilation methods described in literature, which are applicable for this problem (e.g., EVENSEN 2003; BRASSEUR 2006; NERGER et al. 2006; DOBRICIC and PINARDI 2008). For this study we decided to use a relatively simple assimilation approach based on optimal interpolation (OI) filter, because we wanted to make statements about the potential of FerryBox data for forecast skill improvements in general, rather than statements about specific assimilation techniques.

In the standard Kalman filter the forecast covariance matrix P has to be updated in each analysis step using either linearizations of the model operator or ensemble techniques. The OI method as used in this study avoids this complication by assuming that the forecast error statistics is stationary. Furthermore, the OI assimilation scheme uses a distance-dependent localization, which filters out long-range correlations in the background covariance matrix P. A Gaussian function with a width of 30 km is used for this purpose.

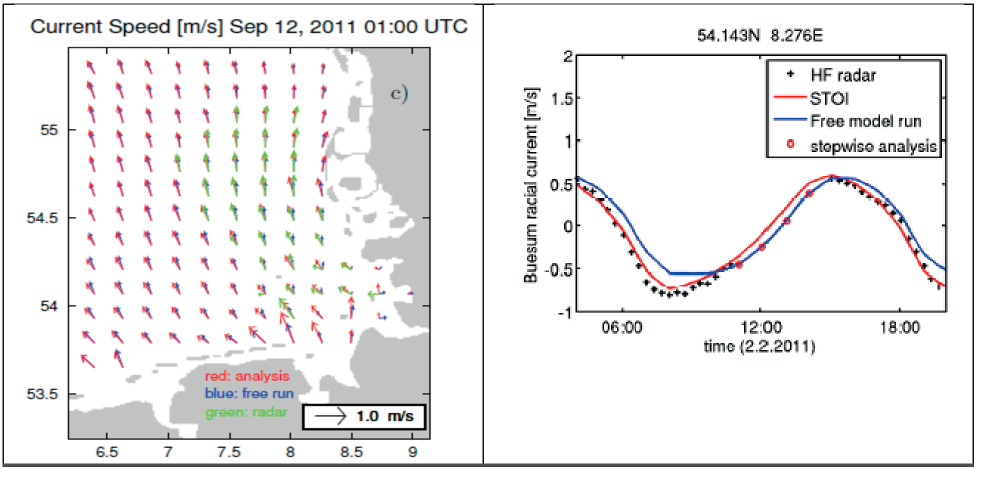

Figure 5: (Left) Illustration of differences between free run (blue), analysis (red) and HF-radar observations (green). (Right) Radial current velocity from HF radar (black crosses), the free model run (blue line), and the STOI analysis (red line) as well as an analysis obtained with a sequential approach (red circles).

Based on the OI scheme an assimilation system was implemented, which performs an analysis every day at 12:00. Fig. 4 shows a comparison of the free model run (red) with the analysis (green) in terms of spatially averaged SST RMSE errors. In this case satellite based SST measurements (OSTIA) were used as a reference. As one can see the analysis is able to decrease the errors most of the time.

3.3 Spatio-temporal OI (STOI): A step towards "best surface current estimate"

Here, we give an illustration of how to maximize the operational outcome of observations in the frame of the Coastal Observing SYstem for Northern and Arctic Seas (COSYNA), which was recently deployed in the area of the German Bight. This system uses three WERA stations and data from stationary platforms. A major challenge of HF radar assimilation is the treatment of tides. This issue is obviously of particular concern in areas like the German Bight, which are dominated by tidal currents. For HF radar data, which are typically acquired several times per hour, analysis and model restart each time new observations become available is not advisable, because models cannot reach equilibrium between two analysis time steps. The method called spatio temporal optimal interpolation (STOI, STANEV et al. 2014) uses elements of both classical assimilation filters and techniques, which use observations alone (FROLOV et al. 2012; WAHLE and STANEV 2011). The proposed data assimilation approach (see Fig. 1 which summarizes different methodologies) has similarities with the methods described in BARTH et al. (2010) and SAKOV (2010). However, it uses a simpler formulation of the model error covariance matrix, but at the same time addresses the forecast capability.

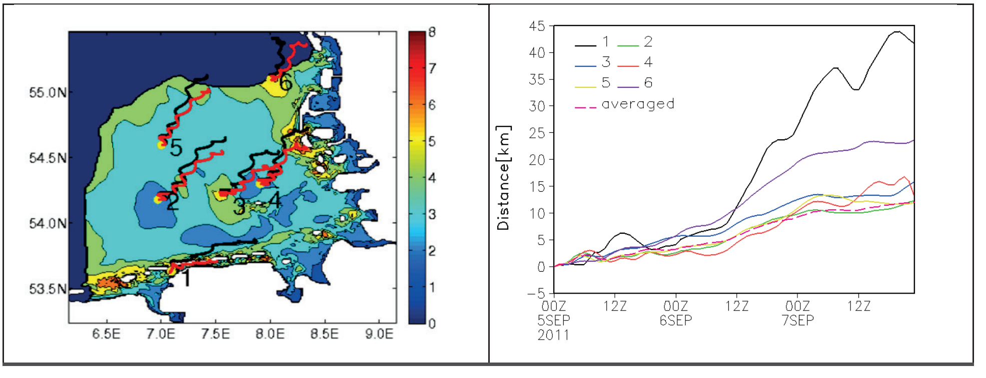

Figure 6: (Left) Using observations from COSYNA HF radars enables preparing useful data products applicable, e.g. for search and rescue operations. The figure on the left exemplifies the displacement of floating objects (in this case Lagrangean particles). Black lines give the results from the model free run, red trajectories visualize the results in the data assimilation run. Color coding gives the mean distance in km between position of Lagrangean particles in the analysis and free run for September 2011 after 24 h of integration. (Right) Distance of drifting particles in the assimilation run and the free run.

Surface currents are analyzed simultaneously using an analysis window of 13 or 24 hours. Using this approach a continuous surface current trajectory over one or two *M*2 tidal cycles is obtained. This block-wise analysis avoids the problem with HF-radar data, which are taken at short intervals (20 minutes for the radar system used in this study). Compared to a classical filter approach the method also has advantages concerning observation data gaps as illustrated in Fig. 5 (right). In this case the sequential method follows the free run when no observations are available, while the STOI method is able to correct a phase error also during this period. To increase the area with available measurements and to avoid any issues related to the processing of two dimensional current vectors from HF-radar data, radial components are used as input for the analysis instead of zonal and meridional components. The main difference from implementations using ensemble Kalman smoother (Barth et al. 2010; Barth et al. 2011) and the technique described in Sakov et al. (2010) is that the model error covariance matrix is estimated from the model background statistics, and not from an ensemble of model runs.

A reduced rank approximation of the state covariance matrix is estimated by EOF analysis of a period of model simulation (February-April 2011). The analysis window was progressed in hourly time steps. This method is different from the classical assimilation approach, where a model restart is performed for the forecast horizon using analyzed fields for initialization. The method also differs from classical statistical forecast methods (e.g. Wahle and Stanev 2011), where forecasts are computed based on measurements alone. In the STOI-forecast mode we use the original free run as a prior estimate for the forecasts.

3.4 Demonstration of the potential usefulness of operational surface current products

The validation of STOI in Stanev et al. (2014) demonstrated that the HF-radar data can be not only interpolated, but also 'extended' in space, which makes possible to generate homogeneous mapped data series. The key difference of the STOI method to other techniques, which provide extrapolated surface current fields based on HF radar data alone is that the state estimates are dynamically consistent with a numerical model run. This is an important aspect when considering the use of such methods in operational systems, many of which have numerical models as their core part.

The following two experiments have been carried out: 33746 Lagrangean particles (this number equals the number of wet model points) have been released every day starting from 00:00 on September 1, 2011 at the surface in the center of every grid cell and were 2-D tracked with a Lagrangean model. Trajectories were computed over three days using hourly model output from either analysis or free model run. The trajectory simulations for the same initial positions of particles have been restarted every day for the same integration time of three days. The Lagrangean model output consisted of 33746x30x24 individual positions. In Fig. 6 (left) the monthly averaged distance between positions of particles in the two runs 24 hours after the release is shown. Release locations from where particles reached the model boundary were excluded from the statistical analysis.

This map gives an idea about the expected success of search and rescue operation if data from HF radar are used or not used. In the latter case the positioning of a lost object would be 3-6 km wrong after one day. Errors could be particularly big if release is in the proximity of barrier islands or close to the northern model boundary. The complicated mesoscale currents around the Helgoland Island could pose problems in the model and observations and explain larger spatial variability of the error pattern. Superimposed on the same figure are the trajectories from the two runs in 6 exemplary locations during three days of integration starting on September 5. They give an idea about the dominating propagation patterns, as well as an illustration that the coherence of tidal oscillations is lost relatively soon after the release. This illustrates the need for intra-tidal information from measurements to correct model trajectories.

The temporal evolution of the distance between particles released at the same positions (Fig. 6 (right)) demonstrates the rapid increase of the distances between trajectories in the two runs. The averaged positioning error plotted by the dashed line gives an overall idea about accuracy in the search and rescue operations using output from the free run. In three days reduction of the error of positioning of an object due to the use of HF radar data is about 10 km on average. It is obvious that this is a relevant difference for applications like Search and Rescue.

4 Conclusions and future challenges

The economic implications of coastal forecasting provide a strong argument for enhancing coastal observing systems. Marine operations depend largely on the success of data assimilation methods, because applications like risk assessment require dynamically consistent data, which are close to available measurements. Regional ocean prediction and monitoring systems, designed by OSSEs, are about to be put in place, in which predictability error is an important part of products. Multi-model ensembles could be also very useful for practical applications.

The examples presented in this paper illustrate the application of data assimilation in the German coastal waters and give an idea of how to develop data assimilation methods appropriate for coastal areas dominated by tides. These methods address complex physics, and the respective characteristic time scales, the outcome of using specific coastal data (radial components of surface currents) from three WERA stations, as well as some pre-operational issues.

There is still a number of challenges in coastal ocean data assimilation. Diagnostics and metrics for assessing performance of the coastal assimilation models need further improvements. Coupling between coastal and open-ocean assimilation systems is still an open problem. Forecasting biogeochemistry state variables in the coastal ocean, although extremely important, is still in infancy. Treatment of river flows, mixing, bottom roughness and small-scale topography is still an issue. Non-homogeneity in space and time of model error statistics needs further consideration. Of particular importance is the optimal use of non-homogeneous data from different origins and platforms. Here, it is expected that 4D-VAR and EnKS could largely contribute to advancing practical applications.

5 Acknowledgements

We thank Alexander Barth, Sebastian Grayek, Joanna Staneva, and Xi Lu for inspiring discussions and support.

6 References

BARTH, A.; ALVERA-AZCÁRATEM , A.; Beckers, J.M.; Staneva, J.; STANEV, E.V. and SCHULZ-STELLENFLETH, J.: Correcting surface winds by assimilating high-frequency radar surface currents in the German Bight, Ocean Dyn., 61, 599-610, 2011.

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.

BARTH, A.; ALVERA-AZCÁRATE, A.; GURGEL, K.-W.; STANEVA, J.; PORT, A.; BECKERS, J.-M. and STANEV, E.: Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of High-Frequency radar surface currents – application to the German Bight, Ocean Sci., 6, 161-178, 2010.

BRASSEUR, P.: Ocean data assimilation using sequential methods base on the Kalman filter, in: CHASIGNET, E. and VERRON, J. (eds.): Ocean weather forecasting: An integrated view of oceanography, Springer, 271-316, 2006.

BURCHARD, H. and BOLDING, K.: GETM – a General Estuarine Transport Model, No EUR 20253 EN, printed in Italy, European Comission, 2002.

BURGERS, G.; VAN LEEUWEN, J. and EVENSEN, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126 (6), 1719-1724, 1998.

CANE, M.A.; KAPLAN, A.; MILLER, R.N.; TANG, B.; HACKERT, E.C. and BUSALACCHI, A.J.: Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter, J. Geophys. Res., 101 (C10), 22599-22617, 1996.

CHASIGNET, E. and VERRON, J.: Ocean Weather Forecasting: An Integrated View of Oceanography, Springer, Dordrecht, 2006.

CUMMINGS, J.; BERTINO, L.; BRASSEUR, P.; FUKUMORI, I.; KAMACHI, M.; MARTIN, M.; MOGENSEN, K.; OKE, P.; TESTUT, C.; VERRON, J. and others: Description of assimilation methods used in GODAE systems, Oceanography Magazine, 22, 96-109, 2009.

DALEY, R.: Atmospheric Data Analysis, Cambridge University Press, 1991.

DEE, D.P. and DA SILVA, A.M.: Data assimilation in the presence of forecast bias, Q. J. R. Meteorol. Soc., 124 (545), 269-295, 1998.

DE MEY, P.; CRAIG, P.; DAVIDSON, F.; EDWARDS, C.; ISHIKAWA, Y.; KINDLE, J.; PROCTOR, R.; THOMPSON, K.; ZHU, J.; the GODAE COASTAL and S.S.W.G.C. COMMUNITY: Applications in coastal modeling and forecasting, Oceanography, 22 (3), 198-205, 2009.

DE MEY, P. and PROCTOR, R.: Assessing the value of GODAE products in coastal and shelf seas, Ocean Dyn., 59, 1-2, 2009.

DOBRICIC, S. and PINARDI, N.: An oceanographic three-dimensional variational data assimilation scheme, Ocean Modelling, 22, 89-105, 2008.

EVENSEN, G.: The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dyn., 31, 961-981, 2003.

EVENSEN, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10, 143-10, 162, 1994.

EVENSEN, G. and VAN LEEUWEN, P.: An ensemble Kalman smoother for nonlinear dynamics., Mon. Weather Rev., 128, 1852-1867, 2000.

FROLOV, S.; PADUAN, J.; COOK, M. and BELLINGHAM, J.: Improved statistical prediction of surface currents based on historic HF-radar observations, Ocean Dyn., 62, 1111-1122, 2012.

GANDIN, L.: Objective Analysis of Meteorological Fields, Translated by Israel Program for Scientific Translations, Jerusalem, 242, 1963.

GELB, A.: Applied optimal estimation, The MIT press, 1974.

GHIL, M. and MALANOTTE-RIZZOLI, P.: Data assimilation in meteorology and oceanography, Adv. Geophys, 33, 141-266, 1991.

GRAYEK, S.; STANEVA, J.; SCHULZ-STELLENFLETH, J.; PETERSEN, W. and STANEV, E.: Use of FerryBox surface temperature and salinity measurements to improve model based state estimates for the German Bight, J. Mar. Syst., 88, 45-59, 2011.

JAZWINSKI, A.H.: Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

KALMAN, R.: A new approach to linear filtering and prediction problems, Journal of basic Engineering, 82 (1), 35-45, 1960.

LORENC, A.: A global three-dimensional multivariate statistical interpolation scheme, Mon. Weather Rev., 109 (4), 701-721, 1981.

LORENC, A.; BALLARD, S.; BELL, R.; INGLEBY, N.; ANDREWS, P.; BARKER, D.; BRAY, J.; CLAYTON, A.; DALBY, T.; LI, D. and others: The Met. Office global three-dimensional variational data assimilation scheme, Q. J. R. Meteorol. Soc., 126 (570), 2991-3012, 2000.

MALANOTTE-RIZZOLI, P., YOUNG, R.E. and HAIDVOGEL, D.B.: Initialization and data assimilation experiments with a primitive equation model, Dyn. Atmos. Oceans, 13 (3), 349-378, 1989.

MARCHUK, G.: Numerical methods in weather prediction, Academic Press, 1974.

MOORE, A.M.; ARANGO, H.G.; BROQUET, G.; POWELL, B.S.; WEAVER, A.T. and ZAVALA-GARAY, J.: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: part I–system overview and formulation, Progress in Oceanography, 91 (1), 34-49, 2011.

NERGER, L.; DANILOV, S. and HILLER, W.: Using sea-level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter, Ocean Dyn., 56, 634-649, 2006.

PETERSEN, W.: FerryBox: A mature system for operational monitoring, Sea Technology, 47, 53-57, 2006.

PHAM, D.; VERRON, J. and ROUBAUD, M.: A singular evolutive extended Kalman filter for data assimilation in oceanography, J. Mar. Syst., 16, 323-340, 1998.

Sakov, P.; EVENSEN, G. and BERTINO, L.: Asynchronous data assimilation with the EnKF, Tellus, 62A, 24-29, 2010.

SASAKI, T.: An objective analysis based on the variational analysis, J. Meteorol. Soc. Japan, 36, 738-742, 1958.

STANEV, E.; SCHULZ-STELLENFLETH, J.; STANEVA, J.; Grayek, S.; SEEMANN, J. and PETERSEN, W.: Coastal Observing and Forecasting System for the German Bight. Estimates of Hydrophysical States, Ocean Sci., 7, 1-15, 2011.

STANEV, E.; ZIEMER, F.; SCHULZ-STELLENFLETH, J.; SEEMANN, J.; STANEVA, J. and GURGEL, K.: Blending surface currents from HF radar observations and numerical modelling: Tidal hindcasts and forecasts, Submitted to J. Atmos. Ocean Techn., 2014.

STANEVA, J.; STANEV, E.; WOLFF, J.-O.; BADEWIEN, T.H.; REUTER, R.; FLEMMING, B.; BARTHOLOMAE, A. and BOLDING, K.: Hydrodynamics and sediment dynamics in the German Bight. A focus on observations and numerical modeling in the East Frisian Wadden Sea, Cont. Shelf Res., 29, 302-319, 2009.

TALAGRAND, O.: On the Damping of High-Frequency Motions in Four-Dimensional Assimilation of Meteorological Data, Journal of Atmospheric Sciences, 29, 1571-1571, 1972.

VAN LEEUWEN, P.: An Ensemble Smoother with Error Estimates, Mon. Weather Rev., 129, 709-728, 2001.

VAN LEEUWEN, P.J. and EVENSEN, G.: Data assimilation and inverse methods in terms of a probabilistic formulation, Monthly Weather Review, 124, 2898-2913, 1996.

VERLAAN, M. and HEEMINK, A.: Reduced rank square root filters for large scale data assimilation problems. Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, 1995.

WAHLE, K. and STANEV, E.: Consistency and Complementarity of Different Coastal Ocean Observations, A Neural Network-based Analysis for the German Bight, Geophys. Res. Lett., 38, 1-4, 2011.