

**SANGOMA: Stochastic Assimilation for the
Next Generation Ocean Model Applications
EU FP7 SPACE-2011-1 project 283580**

**Deliverable 5.6:
Regional model V1 analysis
Due date: 31/10/2015
Delivery date: 03/11/2015
Delivery type: Report , confidential**



Luc Vandembulcke Jean-Marie Beckers
Alexander Barth
University of Liège, BELGIUM

Peter Jan Van Leeuwen Sanita Vetra-Carvalho
University of Reading, UK

Lars Nerger
Alfred-Wegener-Institut, GERMANY

Arnold Heemink Nils van Velzen
Martin Verlaan
Delft University of Technology, NETHERLANDS

Pierre Brasseur Jean-Michel Brankart
CNRS-LEGI, FRANCE

Pierre de Mey
CNRS-LEGOS, FRANCE

Laurent Bertino
NERSC, NORWAY

Chapter 1

Introduction

Imperfect physical models (including imperfectly known parameters) and numerical implementations, limited horizontal and vertical resolution, and errors on the model initial condition, open-sea boundary conditions and atmospheric forcing fields all lead to errors in hydrodynamical model forecasts.

Therefore, modern ocean forecasting systems consist of a hydrodynamic model, and observational networks, both being fused through data assimilation methods in order to provide the best ocean estimates and forecasts. Thanks to progress in modelling and data assimilation, ever-increasing computational power, and more and better designed observational instruments, the quality of these forecasting systems has improved steadily over the last decades.

Recently, much efforts have been targetted at better specifying the uncertainty affecting model forecasts, leading to more accurate data assimilation updates. Building on the theory of the Ensemble Kalman Filter (EnKF) [Evensen, 2003], these efforts often rely on stochastic modelling of uncertainty sources in the models. Recent examples are gives in Lamouroux [2006], Vandenbulcke et al. [2008], Moure et al. [2008], Counillon et al. [2009], Lellouche et al. [2013], Quattrocchi et al. [2014], Ayoub et al. [2015]; operational applications are shown in Sakov et al. [2012], Vandenbulcke and Barth [2015]. Other research efforts in the field of data assimilation are aiming at new data assimilation methods (e.g. non-Gaussian methods), harmonization of tools and benchmarks, and the assimilation of new types of observations.

Forecasting of currents (and waves) in regional and coastal models is important for coastal management and security. One relatively new type of observation, that can help improving current forecasts, are observations by high-frequency (HF) radars. Radar remotely sensed current observations have been developed for about 30 years, and their inclusion in data assimilative systems has started about 15 years ago. Sometimes, when large local variability is present (mesoscale eddies, fronts etc), coastal HF radars can be the only way to provide reliable synoptic views of the current field.

In the present paper, the impact of assimilation of HF radar observations on model forecasts is studied. The analysis is performed in the particular case of an ensemble of 100 models of the Ligurian Sea, with current data from 2 radar

stations.

The members of the ensemble are obtained by perturbing 3 model parameters: the wind forcing field, which is a prominent source of errors [Burrillo et al., 2002, Auclair et al., 2003, Lamouroux, 2006, Jorda and De Mey, 2010, Vervatis et al., 2015], the open-sea boundary condition, and a supplementary, stochastic term in the momentum equation. The ensemble then serves to estimate the covariance matrix needed to assimilate observations.

The data assimilation state vector comprises the projection of the model current onto the radial directions towards the radar stations. Sensitivity experiments are conducted to assess the relative performance of the data assimilation filter with respect to various parameters (such as the length of the time window) and methods (such as the correction of only model variables, or of forcing fields as well).

Chapter 2 briefly reviews the problem of assimilating radar currents. In chapter 3, the ocean model, the method for generating the ensemble and the data assimilation tool are described, as well as the observation operator linking the data and model vectors. In chapter 4, the impact of the velocity observations is analyzed in space and in time. Their impact on other (non-observed) variables is assessed as well. Some conclusions are given in section 5.

Chapter 2

History of assimilation of HF radar data

Current observations by radar stations are available for about 30 years. They provide synoptic estimates of surface currents over a relatively large area, hence they are very useful to constrain regional or coastal models. An initial paper by Barrick et al (1977, 1978) already presented assimilation of radar currents in models. However, since, the quality of the observations has improved, as well as the models and data assimilation schemes. Let's mention:

- [Verlaan \[1998\]](#) assimilated current measurements by 2 HF radar stations, as well as water levels, into a regional model of the coast of Holland, using the reduced-rank Kalman filter variant called RRSQRT. The DA was shown to improve the model, although the author reckons that water level measurement probably have a larger impact. Retrospectively, the quality of the radar measurements may have been the cause for that (Verlaan, private communication).
- [Lewis et al. \[1998\]](#) use nudging by adding a pseudo-layer above the surface and impose a shear stress on top of the one generated by the wind. At the time, they also noted that the radar data accuracy was rather poor.
- [Breivik and Saetra \[2001\]](#): During the EuroROSE project, they performed operational data assimilation in a nested model along the Norwegian coast. Data is discarded when the velocity difference between model and observations is more than 0.5m/s, or the direction difference is more than 45°. The analysis is smoothed with a second-order Shapiro filter. Some scatter plots of the kinetic energy in analysis versus observations (1 plot per lead time) are presented, including the linear regression line.
- [Oke et al. \[2002\]](#) assimilates high-frequency radar observations off the Oregon coast. The data does not undergo a pre-selection (such as the one above), but the correction is applied in multiple steps, in order to avoid shocks in the model.
- [Kurapov et al. \[2003\]](#) presents a forecasting system off Oregon as well, which uses radar data to improve a model (including tides). Large uncer-

tainty enters the model solution through the open-sea boundary conditions; the radar currents allow to reduce model errors (compared to independent acoustic Doppler profiler data).

- [Wilkin et al. \[2005\]](#) assimilates radar data in a, operational forecasting system based on ROMS, off the New Jersey coast
- [Kaplan and Lekien \[2007\]](#) produce smooth two-dimensional fields from radar-based observations, though noting that assimilation of radial currents avoids the additional step and errors of creating the vector currents
- [Chao et al. \[2009\]](#) run a ROMS model nested in the Monterey bay. Available observations include 4 radars with a coverage of about 200km and with resolution of 1 to 3km, but this data is used for validation only. The hourly radar data is filtered with a 33-hour filter to remove tidal currents (diurnal and semi-diurnal time-scales). Regarding predictability, they note that glider velocity rms errors double in 48h, and conclude that it is necessary to assimilate current data.
- [Shulman and Paduan \[2009\]](#) present, in the Monterey bay experiment,
 - 1/ assimilation of low-pass filtered (33 hours) radar data into a non-tidal model, improving comparison with moored current observations
 - 2/ assimilation of unfiltered radar data into a model with tides yields the same level of improvement
 Thus, unfiltered radar data can be assimilated as long as the model represents the observed time-scales too.
- [Hoteit et al. \[2009\]](#) use HF radar data with a 4DVar filter to improve initial condition, open boundary conditions and atmospheric forcings of a MITgcm implementation along the coast of San Diego.
- [Barth et al. \[2010\]](#) use an ensemble Kalman filter to assimilate HF radar data and correct tides. The ensemble members are perturbed using the WCE algorithm.
- [Zhang et al. \[2010\]](#) use a 4DVar filter to improve a ROMS model of the New York Bight. Different datasets are assimilated. It is noteworthy that HF radar data degrades sub-surface temperature.
- [Barth et al. \[2011\]](#) implement an EnKF to correct the wind forcing field; multiple time instants are grouped in the state vector which is subsequently called the estimation vector.
- [Yu et al. \[2012\]](#) use a 4DVar filter to assimilate radar data, in order to improve the geometry of the upwelling SST front, and SSH. The state vector comprises (U,V,SSH,T,S). The error covariance is function only of the model-observations misfit. The model error covariance is diagonal and dependent on distance.
- [Gopalakrishnan and Blumberg \[2012\]](#) use radar data to nudge the "NY-HOPS" 3D estuarine/coastal model of the Raritan Bay and coastal waters

off New York and New Jersey, and show improvement of the model with respect to ADCP data.

- [Supulveda et al. \[2013\]](#) perform a twin-experiment using a 3DVar filter to assimilate data off northern Chile.
- [Paduan and Washburn \[2013\]](#) present a review article of radar current data utilization.
- [Kurapov \[2014\]](#) Assimilates sea surface height, temperature and radar data in the presence of a river plume. Let's note that the authors recently switched to an ensemble-based covariance matrix.
- [Mermain et al. \[2014\]](#) assimilate HF radar data around Toulon with the aim to correct wind and open sea boundary conditions, which will in turn improve the model surface currents. Root mean square (rms) errors on radial velocity and surface currents are about 0.2m/s, and improved very slightly; when errors are larger (around 0.3m/s), the improvement is larger (0.1 m/s). Assimilation of radar data brings no significant changes on T-S profiles.
- [Sperrevik et al. \[2015\]](#) use ROMS-4DVar in an experiment off Norway, assimilating HF-radar and CTD data. The free simulation has relatively bad velocities (compared to drifters). Assimilating radar velocities improves the current field. Assimilating also CTD T-S profiles does not improve surface velocity, but improves the density field.

In general, older studies blame the radar observations for the relatively small improvement that data-assimilation brings to the forecasts. As with other types of observations, over time, the assimilation schemes evolved from nudging schemes (direct or via a pseudo-shear stress) to more elaborate schemes such as 3DVar and 4DVar, and ensemble methods.

Recent studies usually assimilate directly the radial currents (with respect to the radar position), or if orthogonal currents are still assimilated, the authors recommended to switch to the original, radial currents in the future. Indeed, if observed (radial) currents are interpolated on an orthogonal grid, and subsequently interpolated again on (another) orthogonal model grid, more interpolation errors are to be expected.

Let's also note that most authors recommend a 24h window to assimilate radar data.

Chapter 3

System description

3.1 Hydrodynamic model and ensemble generation

The hydrodynamic model is the broadly-used ROMS model, described in detail in [Shchepetkin and McWilliams \[1998, 2003, 2005\]](#).

The present implementation has $1/60^\circ$ horizontal resolution and 32 vertical layers, and covers the domain $8-11.2^\circ$ longitude, and $42.5-44.5^\circ$ latitude. The model domain is shown in Fig. 3.1, as well as simulated temperature and velocity on 06/07/2010.

Open boundary conditions are obtained from the Mediterranean Forecasting System (MFS). Atmospheric forcing fields are from the COSMO forecasts. The model run coincides with the Recognized Environmental Picture campaign that took place during summer 2010 (REP'10).

The Ligurian sea is conditioned by the Liguro-Provencal Current, which is created when the Eastern and Western Corsican Currents join North of Corsica. The region is also the siege of large mesoscale activity; as well as inertial oscillations with a period of approximately 17 hours.

3.2 Ensemble generation

An ensemble of models is generated by perturbing the open-sea boundary conditions, the wind field, and by adding a stochastic term to the momentum equation. The wind field is perturbed by computing its Fourier transform over a certain period, then using the Fourier modes to generate a random perturbation that has similar spatial and temporal scales as the original wind field. The underlying assumption is that wind field errors are mainly phase and amplitude errors, rather than errors in the variability (as would be obtained with a perturbation based on the original field EOFs). The intensity of the perturbation is chosen so that its standard deviation is 30% of its temporal variability. The method is detailed in [Barth et al. \[2011\]](#). The open-sea boundary condition coming from the MFS model is also perturbed following the same method, for the variables of temperature, salinity, surface elevation and velocity.

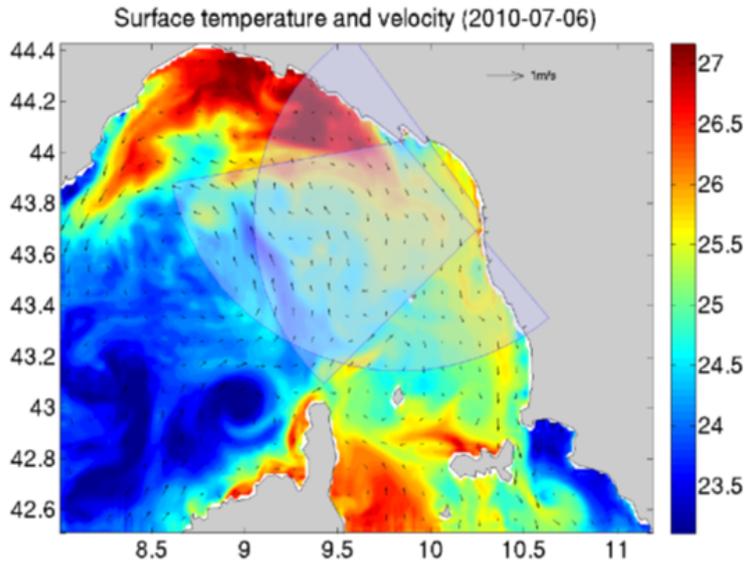


Figure 3.1: Model domain, and temperature and velocity forecast for 06/07/2010. The coverage of radar observations is shown as well.

The stochastic term in the momentum equation is represented by the last term in the right-hand side of:

$$\frac{d\mathbf{u}}{dt} + \Omega \wedge \mathbf{u} = -\frac{1}{\rho_0} \nabla_h p + \frac{1}{\rho_0} \nabla \cdot \mathbf{F}^u + \nabla_h \wedge \epsilon \mathbf{e}_z \quad (3.1)$$

This does not create horizontal divergences or convergences, and can create (absent or misplaced) mesoscale features in the flow.

Three test ensembles are run with the respective perturbations; the perturbation intensities are then scaled so that the 3 kinds of perturbations generate the same magnitude of spread. Finally, the ensemble is spun up for 1 week with all the perturbations together, in order for the members to separate and create different mesoscale circulation features. Hence during the subsequent actual experiment, all members have different initial conditions and boundary conditions.

The model error covariance matrix is subsequently estimated by the ensemble covariance matrix. For example, after the 1-week spin-up, the ensemble surface velocity spread is about 10 cm/s. The spatial correlation is about 50km (for temperature) and 10 km (for velocity). The obtained ensemble should represent the variability at all spatial and temporal scales represented in the simulation.

The ensemble spread at the end of the spin-up is shown in Figure 3.2. The spatial mean of the spread is 1cm for surface elevation, and 8cm.s⁻¹ for surface velocity. Larger values appear along the (perturbed) open boundary, but also in areas with strong currents or eddies. The largest spread occurs at the surface, southward of the area covered by the HF radar observations. The mean temperature spread is 0.6°C; the plot (not shown) does not place the largest spread at the southern boundary, but rather along the coastlines. A supplementary perturbation of the surface heat flux may further increase the spread in the open sea.

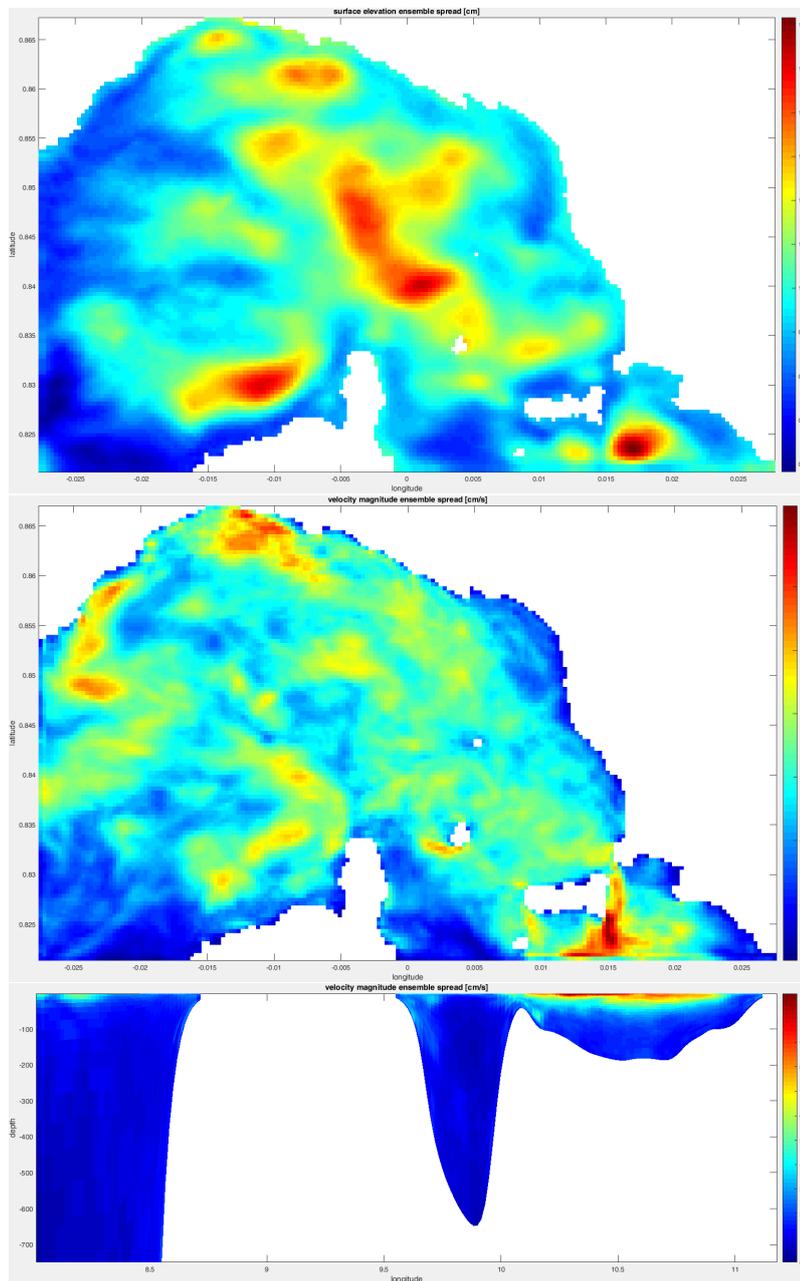


Figure 3.2: Ensemble spread at the end of the spinup, (a) surface elevation, (b) velocity, (c) velocity at the southern boundary (limited to 700m depth)

3.3 Data

During the REP'10 experiment, HF-radar currents, satellite SST images, and glider data were obtained. Two WERA radars were operated by the NATO Undersea Research Center (NURC, now CMRE) in the Insula Palmaria and San Rossore locations, operating at a frequency $\nu=12.359\text{MHz}$ and coupled to a wavelength of $\lambda_b=12.13\text{m}$.

They provide radial current fields with an azimuthal resolution of 6 degrees. These are smoothed in the azimuthal direction as a function of distance to the radar to account for loss of precision at larger distances.

We average the observed currents over 1 hour. An example of radar coverage is shown in Fig. 3.1.

3.4 Data assimilation scheme

Data assimilation is performed by the Ocean Assimilation Kit (OAK) package implementing the SEEK and Ensemble Kalman Filter (EnKF) [Barth et al., 2008]. In this case, the EnKF code is used, albeit with some particular choices detailed hereunder.

The state vector usually contains the model trajectory, i.e. the prognostic model variables at all gridpoints, at the time observations are available. However, it may also contain different variables (e.g. forcing variables), in which case it is called the estimation vector [Barth et al., 2011]. In our experiment, on top of the prognostic model variables, the estimation vector may also contain uncertain (and perturbed) forcing fields such as the wind field, the initial condition, the open boundary condition, and the stochastic error term. In theory, by assimilating enough observations, all these forcing fields can then be optimized.

Radar observations have high temporal resolution, but the modification of model fields by data assimilation may generate unrealistic transient features that do not have time to dissipate if the assimilation frequency is too high. It is well known that too frequent assimilation can degrade the forecast [Talagrand, 1972]. On the other hand, if one wishes to correct rapid processes (e.g. the phase of inertial oscillations), assimilating with lower frequency (e.g. a single current field per day) is not enough, certainly if the observations are time-averaged.

We allow the estimation vector to contain multiple instances of the variables, taken at different model time-steps. In our setup, if we define a 2-day assimilation window, each member of the ensemble saves 48 1-hour-averaged outputs which are all assembled in the estimation vector. The filter is then closely related to the Ensemble Kalman Smoother (EnKS, see van Leeuwen [2001]), the 4D-EnKF [Hunt et al., 2004] and the Asynchronous Ensemble Kalman Filter (AEnKF) [Sakov et al., 2010].

As noted in section 2, it is recommended to use directly the observed, radial currents. In our experiment, the (orthogonal) model currents are first projected

onto the radial axis, and these radial currents are then assembled in the state vector for data assimilation. The observation operator is described in the next section.

Sperrevik et al. [2015] reports that introducing correlations between the u and v components of the (orthogonal, regridded) radar observations only brings marginal improvements to the forecasts. However, even when considering a single component, or radial observations from a single radar, the points in the dense field of radar velocity observations are not spatially uncorrelated.

However, in the current implementation of the EnKF, the observation error covariance matrix \mathbf{R} must be diagonal. Hence, to compensate for the missing non-diagonal elements in \mathbf{R} , we artificially strongly increase the representativity error component in (the diagonal of) \mathbf{R} :

$$\mathbf{R} = \mathbf{R}_{\text{instrument}} + \mathbf{R}_{\text{repr}} \quad (3.2)$$

where $\mathbf{R}_{\text{instrument}}$ is the instrumental (measurement) error and \mathbf{R}_{repr} needs to be determined.

In the present tests, considering a 2-day assimilation window, the estimation vector \mathbf{x} may contain the 48 1-hourly-averaged radial velocities, the model prognostic variables (temperature, salinity, barotropic and baroclinic velocity, and sea surface elevation) at the end of the 2-day window, and also the wind forcing. During the EnKF step, all these variables need to be loaded in memory for all the ensemble members. When observations are available, the filter then updates all the variables in the estimation vector; but the size of the latter in the computer memory can be prohibitive.

Hence, in practice, the estimation vector does not contain all the other model variables, but only the variables directly required to compute $h(\mathbf{x}^f)$, where h is the observation operator and \mathbf{x}^f is the first guess of the estimation vector. Then, the data assimilation procedure saves the local coefficients α_i of the linear combination of ensemble members (forecasts), that will point-wise yield the analysis and the analyzed members:

$$\mathbf{x}_a = \mathbf{x}_f + \sum \alpha_i \mathbf{x}_i \quad (3.3)$$

Separately, the same linear combination can then be applied to any other model variables or forcing fields such as the wind forcing, and the analyzed variable or field is obtained. The procedure is illustrated in Figure 3.3. The model saves 48 averaged fields, and a restart file. The observation operator described below is applied to the averaged velocity fields, which are then assembled in the estimation vector together with (in this example) the restart-file temperature. The EnKF filter saves the amplitudes. All model restart variables are then updated (for each ensemble member, one at a time) and new restart files are generated.

Finally, a localization procedure removes spurious long-range correlations by multiplying the correction generated by each observation, with a radial Gaussian function centered on the observation. The Gaussian's radius, as well as a cut-off length after which the correction is set to zero completely, must be determined.

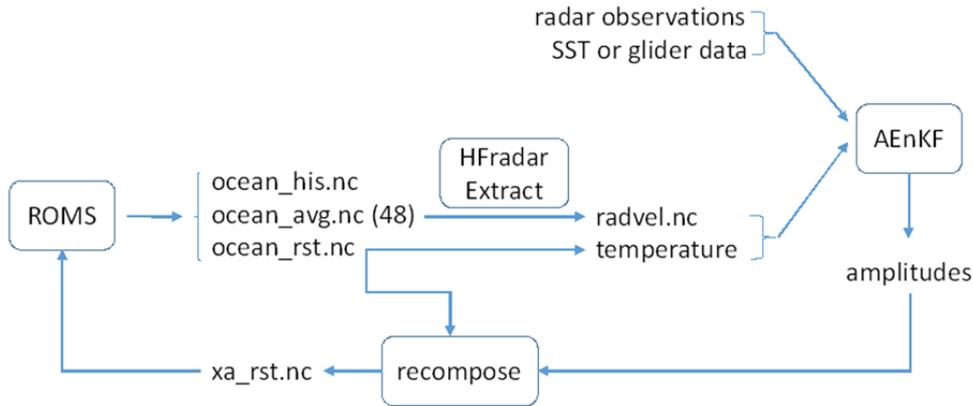


Figure 3.3: Succession of model integration, observation operator, data assimilation scheme, and recomposition.

3.5 Radar observations observation operator

The non-linear transformation of orthogonal (model) currents into radial currents pointing to (or away from) the radar station, is performed by the radar observation operator publically available as part of the Sangoma tools, see <http://www.data-assimilation.net>.

The transformation is performed according to:

$$u_{HF} = \frac{k_b}{1 - \exp(-k_b h)} \int_{-h}^0 \mathbf{u}(z) \cdot (e)_r \exp k_b z dz \quad (3.4)$$

where $k_b = \frac{2\pi}{\lambda_b}$, and e_r is the unit vector pointing in the opposite direction to the location of the radar. Positive values hence represent currents away from the system. The operator essentially represents an average over the upper meters.

Furthermore, the radial currents are smoothed in the azimuthal direction using a diffusion operator. Hence, currents farther away from the radar station will naturally be more smoothed, similar to what occurs with observed currents.

The analysis update with a non-linear observation operator is not guaranteed to be unbiased; the standard EnKF algorithm can still be used but one needs to verify separately that the following relation holds:

$$\langle h(\mathbf{x}^f) \rangle = h(\langle \mathbf{x}^f \rangle) \quad (3.5)$$

where $\langle . \rangle$ represents the ensemble average [Barth et al., 2011].

Chapter 4

Results

4.1 Assimilation of HF radar currents

Our base experiment has a 2-day assimilation window, leading to 19 windows during the experiment. Some of these windows do not have data at the San Rossore radar station. In these windows, data assimilation is limited to one velocity component (Palmaria station radial velocities). Assimilation of orthogonal velocities would have been impossible for these windows.

The (radial) velocity rms error between the free run and the radar observations is shown by the black curve in Fig. 4.1. Concerning the Palmaria radar, errors are between 10 and 15 cm s^{-1} except during the last week of July 2010, when they are over 20 cm s^{-1} . Thus, the ensemble spread (10 cm s^{-1} , see previous section) is of the same order of magnitude as the model-observations discrepancy. Compared to most studies presented in chapter 2, these errors are already relatively low, even without data assimilation. Let's note that in another recent study [Mermain et al., 2014], the authors found that when the model-observations discrepancy is small (around 10 cm s^{-1}), assimilation sometimes degrades the simulation. Concerning the San Rossore radar, errors are also around 10 cm s^{-1} during the first week; but then data becomes unavailable. Later during the experiment, the errors are much higher, around 20 cm s^{-1} .

The representativity error component of the observation error variance is set to 250 cm s^{-1} . This value may appear large, but the radar observations are very dense in space, and the representativity error needs to account for the processes represented (or not) in the model compared to the observations. Mermain et al. [2014] used observation errors of 100 and 200 cm s^{-1} .

The localization radius is set to 30km, with the cut-off length at 2000km. The procedure updates the model's (orthogonal) velocity, sea surface elevation, temperature, and salinity at the restart time (i.e. the end of the assimilation window). In between data assimilation cycles, the full model forwards the 100 ensemble members with perturbed forcing fields. We also run one free member without perturbations nor data assimilation.

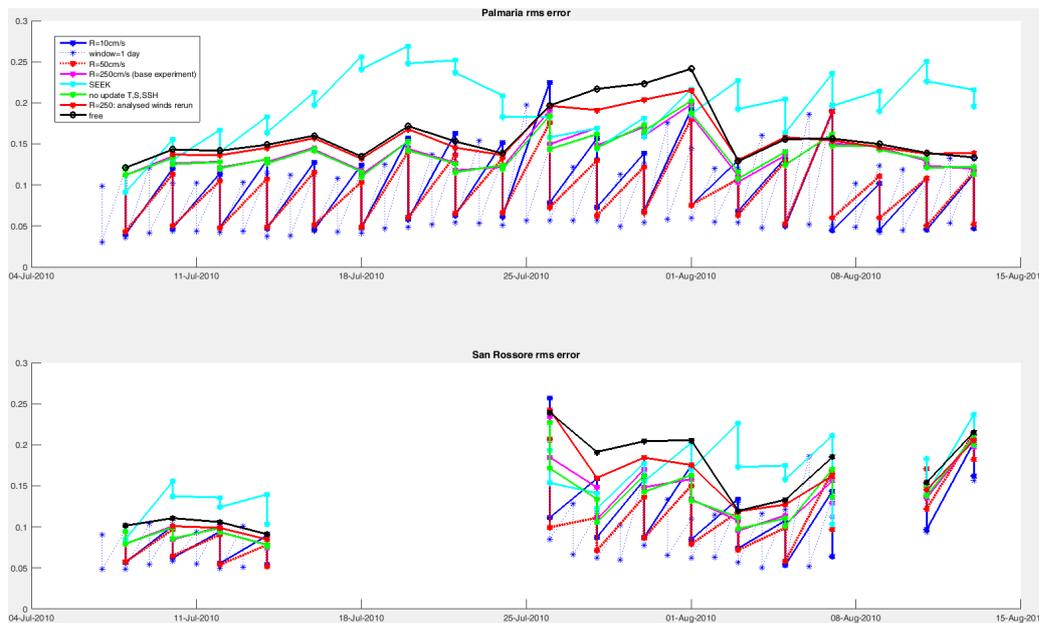


Figure 4.1: Root mean square difference [$\text{m}\cdot\text{s}^{-1}$] between model (different cases) and observation radial currents for (a) the Palmaria station, (b) the San Rossore station.

The rms errors of the data assimilative run are represented by the pink curve in Fig. 4.1. Errors are somewhat reduced by the data assimilation procedures. Only when the error is large (20cm s^{-1} and more), the corrections also become more important and the improvement compared to the free run is about 5 cm s^{-1} . In any case, the errors are smaller than the free run errors even after the subsequent 2-day model integrations.

In order to further decrease the errors, one may try to reduce the representativity error. Hence, another experiment with strongly decreased values of (the representativity component of) the observation error variance matrix was carried out; results are shown by the dotted red curve and the blue curve, using representativity errors of respectively 50 and 5 cm s^{-1} . As expected, the corrections are much larger; but the model creates new errors in between assimilation cycles much more strongly than before. Only during the last decade of July, the period when the free model errors were large (about 20 cm s^{-1}), data assimilation managed to significantly reduce them without the model generating them again. In general, out of the 3 simulations with different representativity errors, there is not one single simulation that has consistently smaller errors during the whole experiment.

For the model not to re-create errors so strongly, one could try to shorten the assimilation window. Other experiments were carried out, similar to the last experiment (blue curve) except for the assimilation window length, chosen as 24h (dotted blue curve in Fig 4.1) and 72h (not shown). In general, similar to findings

in [Sperrevik et al. \[2015\]](#), it appears that the assimilation window length is not critical.

Other experiments used a localization radius of 60km instead of 30. It appeared that, as the radar observations cover a large area, the radius of influence of an observation is not critical because there are always plenty of observations in close proximity (or none at all). This is unlike assimilation of e.g. ARGO data.

Another experiment consisted in reproducing the simulation with the largest representativity error (pink curve), but updating only the model velocity (u, v) without changing the other model variables (temperature, salinity and surface elevation). Errors are shown in the green curve, and are almost indistinguishable from the pink curve. This means that the model physics do not have time, in between assimilation cycles, to propagate (small) density changes to the model velocity.

Yet another experiment, based on the basis experiment (pink curve), includes the wind forcing in the estimation vector. In practice, after the ensemble is integrated in time, the data assimilation is performed on the radial velocities (as before), and the wind is optimized afterwards, during the recomposition step. The model is then run one more time using the analyzed winds. This leads to the red error curve in Fig. 4.1, which is also lower than the errors of the free run. However, in this experiment, analyzing the wind rather than directly changing the model variables did not lead to smaller errors, indicating that maybe another one of the perturbations is more critical, or at least that the wind optimization alone does not improve the model enough. Actually, in a study in the neighboring North-Western Mediterranean Sea (just to the west of our study domain), [Mermain et al. \[2014\]](#) noted that optimizing the wind forcing did not allow to correct the phase of inertial oscillations, which are an important component of the dynamics in both domains.

None of these experiments could decrease the rms errors further (not shown), and indeed differences between the experiment results were marginal.

Given the difficulty to further improve the model velocities with the EnKF / EnKS proposed here, a fixed-based SEEK filter [[Pham et al., 1998](#)] was also tried out. The fixed directions of the error space are provided as the (temporal) EOFs of a 1-month run immediately preceding the experiment. Results are shown by the light blue curve in 4.1. It can be seen, in our particular experiment, that using a static error space approximated by the model's temporal variability does not allow to improve the model, rather degrading the results.

Finally, as different authors mentioned (see chapter 2), a possible cause for the difficulty to further improve the simulations may be the quality of (some) observations. Artifacts may be present at the extremities of the observed field, both at the azimuthal edges and the maximum distance edge [[Gomez et al., 2015](#)]. An azimuthal "smear" of the observations may also be observed, particularly far away from the radar station. Therefore, we carried out new, restrictive experiments where all radar observations are eliminated, that are associated with a quality-control error estimate larger than 1.5 cm s^{-1} , or that are located at the azimuthal edges of the field, or further away than 50km from the radar station. The experiments were carried out with representativity errors of 10, 25 and 50 cm s^{-1} . Rms

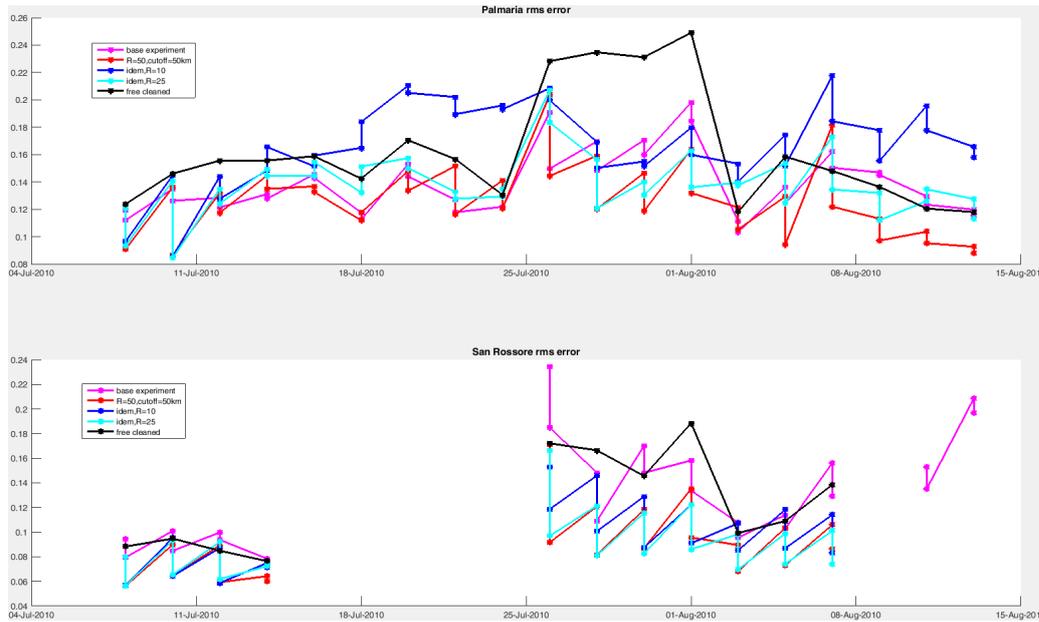


Figure 4.2: Root mean square difference [$\text{m}\cdot\text{s}^{-1}$] between model and "cleaned up" radial current observations for (a) the Palmaria station, (b) the San Rossore station.

error timeseries are shown in Figure 4.2, respectively the dark blue, light blue and red lines. The base experiment (with representativity error of 50 cm s^{-1}) is included for comparison (pink curve), the only difference compared with the red curve being the considered observations. The free run error curve has been recalculated as well and is drawn in black.

Results are somewhat better, particularly in the second half of the experiment. It also appears again that too small representativity errors are yielding bad results; a good value for the considered experiment is probably in the range $50\text{-}100 \text{ cm s}^{-1}$.

4.2 Spatial analysis

The rms plots above do not indicate how the correction is spatially distributed. As an example, the first correction for the base experiment is shown in Fig. 4.3. As expected from the relatively low rms errors shown before, the forecast and observations are already in good agreement. One eddy north of Corsica, visible in the observations, is either absent or placed too much to the East in the model. One can also observe the azimuthal smear mentioned above, which is more accentuated further away from the radar station.

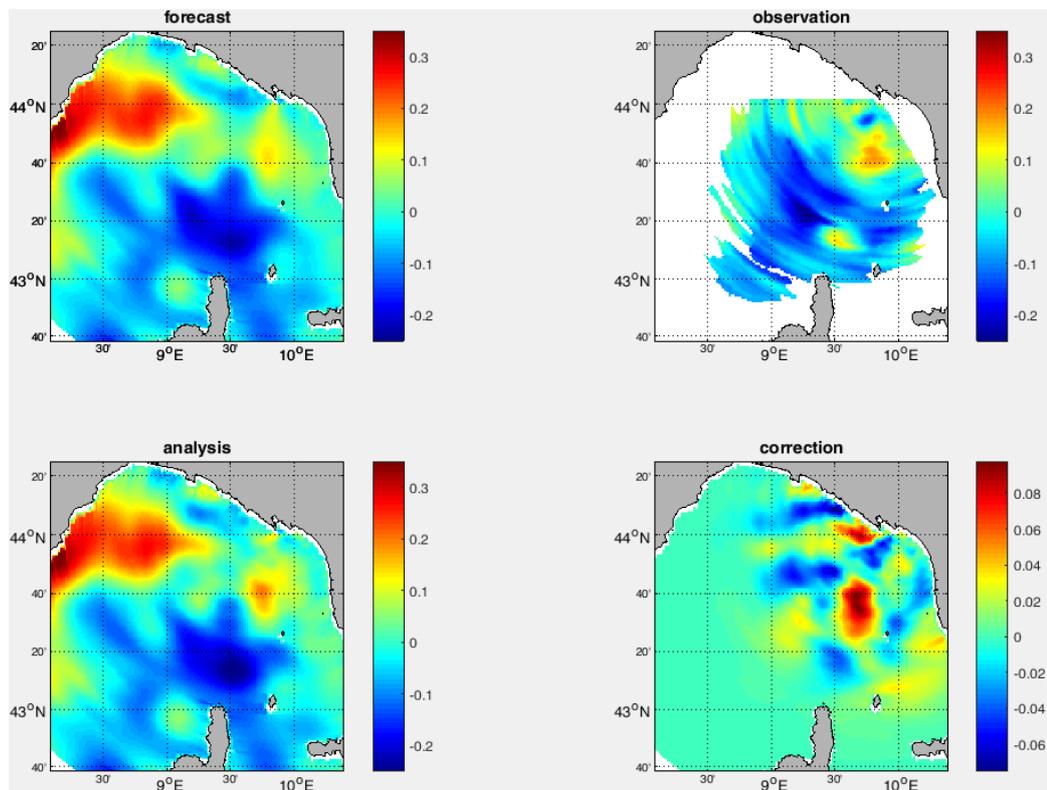


Figure 4.3: For the base experiment, for the second hour of the first assimilation window, (a) forecasted radial currents away from the Palmaria station, (b) observations, (c) analysis, (d) correction.

4.3 Temporal analysis

An important question is how long an impact a radar observation will have on the model, or alternatively, how frequently current observations are required to best constrain the model.

Therefore, we study the impact of the 2 radar velocities in a single point, at the exact location where the radial currents toward the 2 radars are orthogonal. We first consider a single observation per assimilation window, the second hourly-averaged. The correction (also called increment) to hourly-averaged velocities, obtained at that particular point, is represented by the blue curve in Fig. 4.4. The coefficients of the analysis change every 48 hours, which explain the discontinuities. As (at least some of) the ensemble anomalies must present dephased inertial oscillations, the correction also presents the typical inertial cycle with a period of 17 hours, although just one observation is provided.

The forecast, observation and analysis (for the second 48-hour period) are shown in Fig. 4.5. The single assimilated observation is represented by a red arrow; the other observations are plotted for reference only but are not assimilated. One can observe that although a single linear combination is built for the 48 hours of the assimilation window, the result of the assimilation is so, that the correction is larger mostly during the first 12 hours, i.e. close to the observation. During this period, the velocity is changed from north-eastward to south-eastward. Afterwards, the intensity of the velocity is changed less heavily, and the phase is almost in the forecast and the analysis.

When using 48 hourly-averaged velocities during the 48-hour window, but still only in 1 single point in space, the corresponding correction is shown by the red curve in Fig. 4.4. One can see that the inertial oscillation correction is much stronger. This shows the beneficial impact of having very frequent observations (i.e. every hour).

During the REP'10 campaign, surface drifters were also launched. However, very large rms velocity errors were obtained between model and drifters ($27 \text{ cm}\cdot\text{s}^{-1}$), but also between radar and (projected) drifter velocities ($25 \text{ cm}\cdot\text{s}^{-1}$). This may be due to the fact that hourly-averaged model and radar velocities do not correspond to the mean velocity measured by the drifter between 2 transmissions (usually larger than 6 hours, i.e. one third of the inertial oscillation period). It may also be caused by outliers in the drifter data, with huge errors (up to $70 \text{ cm}\cdot\text{s}^{-1}$).

4.4 Assimilation of radar and temperature observations

The ensemble temperature spread (about 0.5°C) has the same order of magnitude as the model-observation difference (about 1.5°C). However, assimilating radar currents leads to a (small) deterioration of the rms error between model and observed (satellite) sea surface temperature (SST). The errors are represented in Fig. 4.6 for all the same experiments as described above; they are all larger

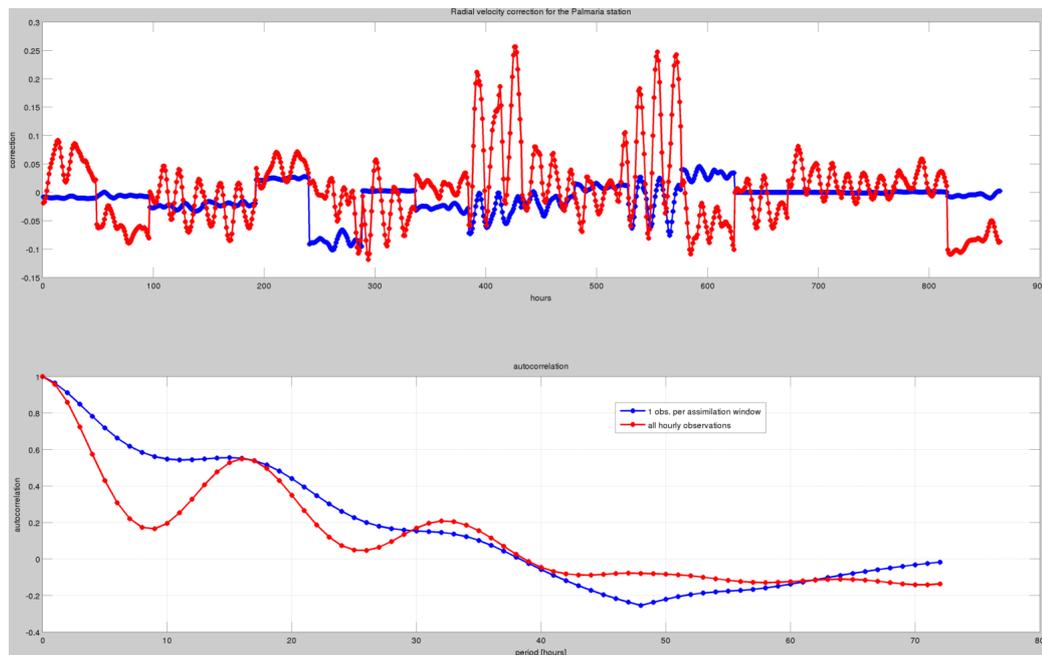


Figure 4.4: Upper panel: velocity increment when assimilating velocity in a single point. The blue curve corresponds to the case when assimilating once per time-window, the red curve to the case when assimilating each hourly-averaged observation. Lower panel: autocorrelation of the timeseries in the upper panel.

than the for the free run. Unfortunately, the inability to improve temperature (and salinity) fields by observing radar currents was also noted in other, recent studies [Zhang et al., 2010, Mermain et al., 2014, Sperrevik et al., 2015].

Therefore, another experiment was conducted where we assimilated satellite SST measurements as well. The SST data assimilation procedure itself is exactly the same as for radar data; R is chosen as 0.75°C .

Three experiments are carried out. In the first one, the whole estimation vector (i.e. model velocity, temperature, salinity and surface elevation) is updated by the two observations. In the second experiment, radar observations update only model velocities, and SST observations update only temperature, salinity and surface elevation. In the last experiment, salinity and surface elevation are not updated at all. The 3 experiments correspond respectively to the red, dark blue and light blue curves in the bottom panel of Figure 4.6.

One can observe that by assimilating both radar and SST data, the temperature errors are now consistently lower than the free run, instead of being slightly degraded by assimilation of radar data alone. The lack of difference in between the 3 experiments seems to confirm the short term decoupling between the velocity and density updates by the data assimilation, mentioned previously. Furthermore, plots of the velocity rms error when assimilating both radar and SST data (not shown) are very similar to the base experiment (without assimilating SST), indicating that SST observations neither improve nor degrade the velocity field.

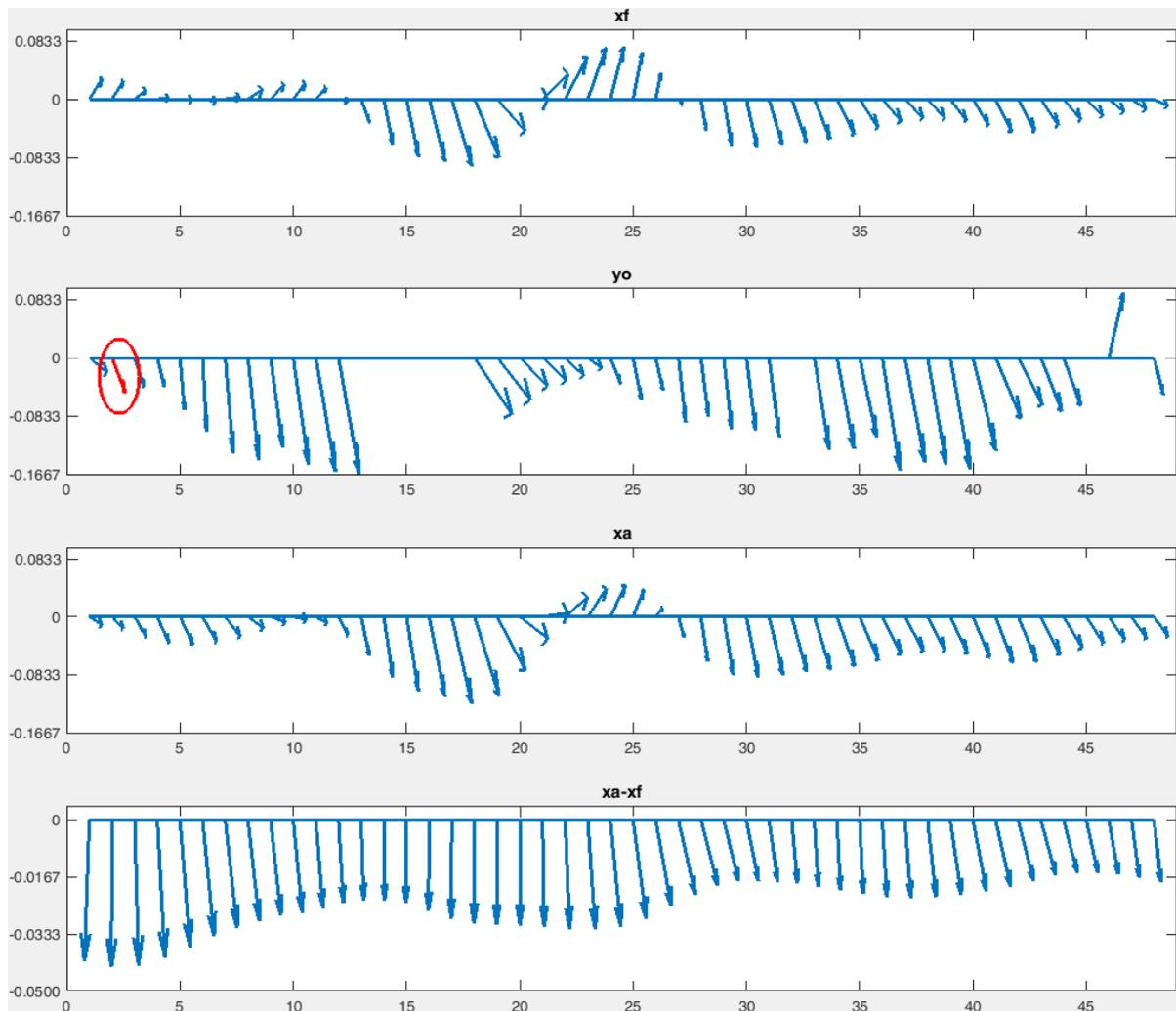


Figure 4.5: Feather plot of the velocity at the single point where data is assimilated, during the 48 hours of the second assimilation window. First panel: model forecast. Second panel: observation; the arrow corresponding to the assimilated observation is represented in red, while the other arrows are drawn for illustration only. Third panel: analysis. Fourth panel: data assimilation increment. The axis in the feather plots are orthonormal; the velocity at the chosen location just happens to be oriented mainly north-south. The fourth panel has different orthonormal axis.

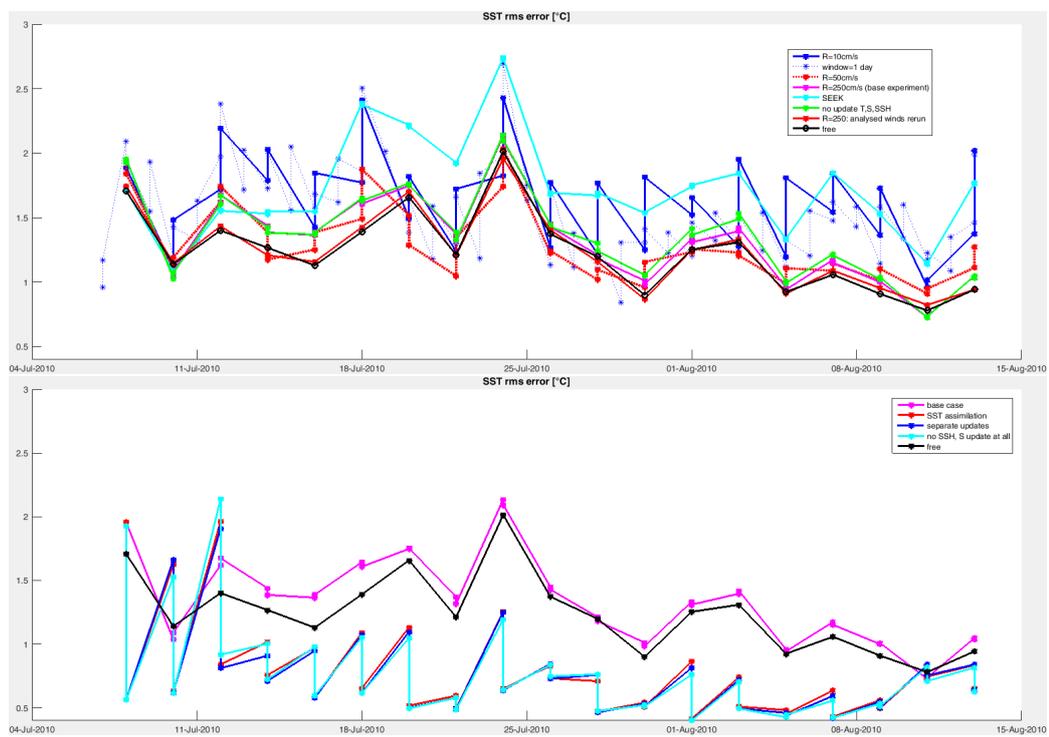


Figure 4.6: Rms SST errors. Upper panel: experiments with DA of radar currents only (same experiments as in Fig. 4.1). Lower panel: experiments with DA of radar and SST observations.

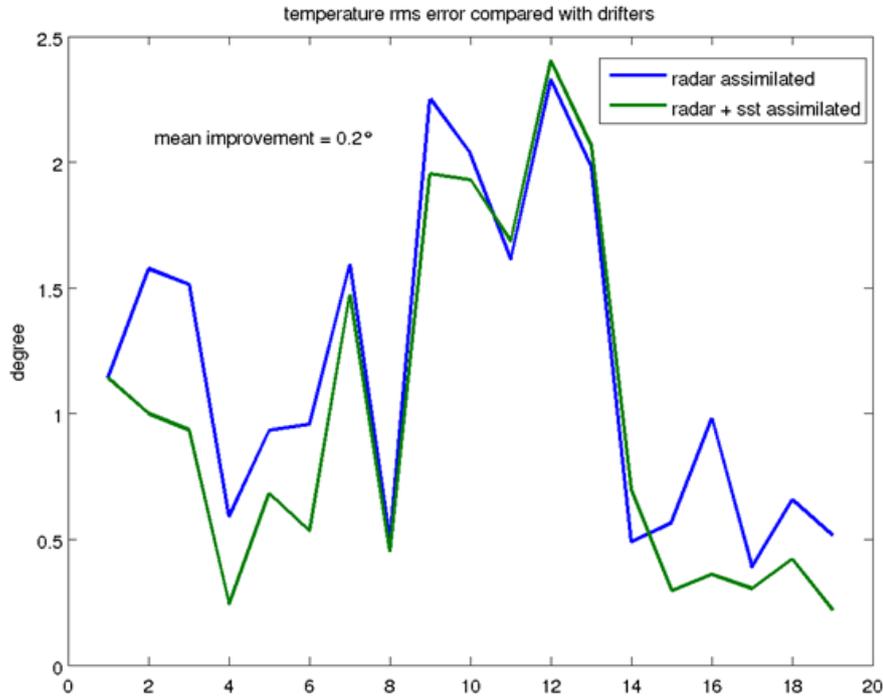


Figure 4.7: RMS temperature error between drifters and model forecasts when assimilating only radar velocities (blue curve) or when assimilating SST images as well (green curve).

Comparisons of independent temperature measurements realized by the drifters, with the “free” model run (i.e. assimilation of radar velocity only) and the SST-assimilating model are represented respectively by the blue and green curves in Fig. 4.7. One can observe a average reduction of 0.2°C of the rms errors in the latter case.

Chapter 5

Discussion and conclusion

In this report, we analyzed whether radar observations could improve a regional model of the Ligurian Sea, using an ensemble of 100 ROMS models. The ensemble is generated by perturbing the open sea boundary conditions and the wind forcing, and by adding a supplementary stochastic term to the momentum equations.

The estimation vector used by the Ensemble Kalman Filter (EnKF) contained multiple instances of hourly-averaged currents (48 instances in the case when the assimilation window is 2 days), making the EnKF closely related to the Asynchronous Ensemble Kalman Filter (AEnKF) and the Ensemble Kalman Smoother (EnKS).

The free run presented relatively small errors compared to the radar observations, of the order of 15 cm.s^{-1} . An experiment with a fixed-based SEEK filter based on "historical" EOFs did degrade the velocity field rather than improve it.

The ensemble spread is of the same order of magnitude as the free run - observations discrepancy. The EnKF data assimilation results showed that the radar data allows to somewhat further reduce the discrepancy between model velocity and radar observations, particularly when the model itself departs more strongly from observations ($> 20 \text{ cm.s}^{-1}$).

The results showed very little sensitivity to the assimilation window length, to the localization radius and cut-off length. The representativity component of the observation however must not be chosen too small, leading to over-fitting of the data and the model immediately creating large transients and new errors. An appropriate range, in the case of the current experiment, is 50 to 100 cm.s^{-1} .

Adding the wind forcing in the estimation vector so as to obtain an analyzed forcing field, and then re-running the model with this new forcing, did not yield better results than updating directly the model variables.

Observations have the largest impact during the next few hours. Furthermore, the ensemble was shown to adequately represents the uncertainty due to different phases and amplitudes of the inertial oscillations. Even a single observation allows the assimilation filter to modify the inertial oscillation phase (during a few hours). When observations are available with high temporal frequency of observations (hourly, in this case), the phase of the inertial oscillations is constantly modified in the model.

Surface drifter velocities could not be used, because they represent an average velocity over a much longer period, and/or due to the presence of outliers, and hence they presented very large discrepancies compared to both the model and the radar observations.

Assimilation of radar data does not allow to improve model SST; actually, model SST is very slightly degraded. In general, the radar data and the model density are not sufficiently correlated for one to influence the other through high-frequency data assimilation, as the model needs more time for a density change to reflect in the velocity field. However, assimilating both radar velocities and satellite SST allows to improve both the model velocity and the model SST. This is confirmed by comparisons with drifter temperature measurements.

Bibliography

- F. Auclair, P. Marsaleix, and P. De Mey. Space-time structure and dynamics of the forecast error in a coastal circulation model of the gulf of lions. *Dyn. Atmos. Oceans*, 36:309–346, 2003.
- N. Ayoub, M. Lucas, and P. De Mey. Estimating uncertainties on a gulf stream mixed-layer heat budget from stochastic modeling. *Journal of Marine Systems*, 150:66–79, 2015.
- A. Barth, A. Alvera-Azcarate, and R. H. Weisberg. Assimilation of high-frequency radar currents in a nested model of the West Florida shelf. *Journal of Geophysical Research*, 113, 2008.
- A. Barth, A. Alvera-Azcarate, K.-W. Gurgel, J. Staneva, A. Port, J.-M. Beckers, and E. Stanev. Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of high frequency radar surface currents - application to the German Bight. *Ocean Science*, 6:161–178, 2010.
- A. Barth, A. Alvera-Azcarate, J.-M. Beckers, J. Staneva, E. Stanev, and J. Schulz-Stellenfleth. Correcting surface winds by assimilating high-frequency radar surface currents in the German Bight. *Ocean Dynamics*, 61:599–610, 2011.
- O. Breivik and O. Saetra. Real time assimilation of HF currents into a coastal ocean model. *Journal of Marine Systems*, 28:161–182, 2001.
- A. Burrillo, G. Caniaux, M. Gavart, P. De Mey, and R Baraille. Assessing ocean-model sensitivity to wind forcing uncertainties. *Geophys. Res. Lett.*, 29, 2002.
- Y. Chao, Z. Li, J. Farrara, J. McWilliams, J. Bellingham, X. Capet, F. Chavez, J.-K. Choi, R. Davis, J. Doyle, D. Fratantoni, P. Li, P. Marchesiello, M. Moline, J. Paduan, and S. Ramp. Development, implementation and evaluation of a data-assimilative ocean forecasting system off the central California coast. *Deep-Sea Research II*, 56:100–126, 2009.
- F. Counillon, P. Sakov, and L. Bertino. Application of a hybrid enkf-oi to ocean forecasting. *Ocean Science*, 5:389–401, 2009.
- G. Evensen. The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.
- R. Gomez, T. Helzel, C.R. Merz, Yonggang L., R.H. Weisberg, and N Thomas. Improvements in ocean surface radar applications through real-time data quality-control. In *Current, Waves and Turbulence Measurement (CWTM), 2015 IEEE/OES Eleventh*, 2015.

- G. Gopalakrishnan and A. Blumberg. Assimilation of HF radar-derived surface currents on tidal timescales. *Journal of Operational Oceanography*, 5:75–87, 2012.
- I. Hoteit, B. Cornuelle, S.Y. Kim, G. Forget, A. Kohl, and E. Terrill. Assessing 4d-var for dynamical mapping of coastal high-frequency radar in san diego. *Dynamics of Atmospheres and Oceans*, 48:175–197, 2009.
- BR. Hunt, E. Kalnay, E.J. Kostelich, E. Ott, DJ. Patil, T. Sauer, I. Szunyogh, JA. Yorke, and AV. Zimin. Four-dimensional ensemble kalman filtering. *Tellus*, 56: 273–277, 2004.
- G. Jorda and P. De Mey. Characterization of error dynamics in a 3d coastal model of the catalan sea using stochastic modelling. *Cont. Shelf Res.*, 30:419–441, 2010.
- D. Kaplan and F. Lekien. Spatial interpolation and filtering of surface current data based on open-boundary modal analysis. *Journal of Geophysical Research - Oceans*, 112, 2007.
- A. Kurapov. Improvements in the Oregon coastal ocean forecast system: data assimilation in the presence of the Columbia River plume. *presentation*, 2014.
- A. Kurapov, G. Egbert, J. Allen, R. Miller, S. Erofeeva, and P. Kosro. The M2 internal tide off Oregon: Inferences from data assimilation. *Journal of Physical Oceanography*, 33, 2003.
- J. Lamouroux. *Erreur de prevision d'un modele oceanique barotrope du Golfe de Gascogne en reponse aux incertitudes sur les forcages atmospheriques: caracterisation et utilisation dans un schema d'assimilation de donnees a ordre reduit*. PhD thesis, Universite Paul Sabatier, Toulouse, 2006.
- J.M. Lellouche, O. Le Galloudec, M. Drevillon, C. Regnier, E. Greiner, G. Garric, N. Ferry, C. Desportes, C.-E. Testut, C. Bricaud, R. Bourdalle-Badie, B. Tranchant, M. Benkiran, Y. Drillet, A. Daudin, and C. De Nicola. Evaluation of global monitoring and forecasting systems at mercator ocean. *Ocean Science*, 9: 57–81, 2013.
- J. Lewis, I. Shulman, and A. Blumberg. Assimilation of Doppler radar current data into numerical ocean models. *Cont. Shelf Research*, 18:541–559, 1998.
- J. Mermain, A. Molcard, P. Forget, A. Barth, and Y. Ourmieres. Assimilation of HF radar surface currents to optimize forcing in the northwestern Mediterranean Sea. *Nonlin. Processes Geophys.*, 21:659–675, 2014.
- B. Moure, J. Ballabrera-Poy, E. Garcia-Ladona, and J. Font. 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, 2008.
- P. Oke, J. Allen, R. Miller, G. Egbert, and P. Kosro. Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research*, 107, 2002.

- J. Paduan and L. Washburn. Observations of ocean surface currents. *Annu. Rev. Mar. Sci.*, 5, 2013.
- D. Pham, J. Verron, and M. Roubaud. A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16:323–340, 1998.
- G. Quattrocchi, P. De Mey, N. Ayoub, V.D. Vervatis, C.-E. Testut, G. Reffray, J. Chanut, and Y. Drillet. Characterisation of errors of a regional model of the Bay of Biscay in response to wind uncertainties: a first step toward a data assimilation system suitable for coastal sea domains. *J. Oper. Oceanography*, 7:25–34, 2014.
- P. Sakov, G. Evensen, and L. Bertino. Asynchronous data assimilation with the enfk. *Tellus*, 62A:24–29, 2010.
- P. Sakov, F. Counillon, L. Bertino, O.R. Lisaeter, K.A. abd Oke, and A. Korablev. Topaz4: an ocean-sea ice data assimilation system for the North Atlantic and arctic. *Ocean Science*, 8:633–656, 2012.
- A. F. Shchepetkin and J. C. McWilliams. A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate. *J. Geophys. Res.*, 108(C3), 2003.
- A. F. Shchepetkin and J. C. McWilliams. The regional ocean modeling system: A split-explicit, free-surface, topography following coordinates ocean model. *Ocean Modelling*, 9:347–404, 2005.
- A.F. Shchepetkin and J.C. McWilliams. Quasi-monotone advection schemes based on explicit locally adaptive dissipation. *Monthly Weather Review*, 126: 1541–1580, 1998.
- I. Shulman and J. Paduan. Assimilation of HF radar-derived radials and total currents in the Monterey Bay area. *Deep-Sea Research II*, 56:149–160, 2009.
- A. K. Sperrevik, K.H. Christensen, and J. Rohrs. Constraining energetic slope currents through assimilation of high-frequency radar observations. *Ocean Science*, 11:237–249, 2015.
- H. Supulveda, P. Marchesiello, and Z. Li. Ocean data assimilation study in northern Chile: use of a 3DVAR method. *Lat. Am. J. Aquat. Res.*, 41:570–575, 2013.
- O. Talagrand. On the damping of high-frequency motions in four-dimensional assimilation of meteorological data/. *Journal of the Atmospheric Sciences*, pages 1571–1547, 1972.
- P.J. van Leeuwen. An ensemble smoother with error estimates. *Monthly Weather Review*, 129:709–728, 2001.
- L. Vandenbulcke and A. Barth. A stochastic operational forecasting system of the black sea: Technique and validation. *Ocean Modelling*, 93:7–21, 2015.

- L. Vandenbulcke, M. Rixen, J.-M. Beckers, A. Alvera-Azcarate, and A. Barth. An analysis of the error space of a high-resolution implementation of the gher hydrodynamic model in the mediterranean sea. *Ocean Modelling*, 24:46–64, 2008.
- M. Verlaan. *Efficient Kalman filtering algorithms for hydrodynamic models*. PhD thesis, TU Delft, The Netherlands, 1998.
- V. Vervatis, C.-E. Testut, P. De Mey, N. Ayoub, J. Chanut, and G. Quattrocchi. Data assimilative twin-experiment in a high-resolution bay of biscay configuration: 4d enoi based on stochastic modelling of the wind forcing. *Ocean Modelling*, 2015.
- J.L. Wilkin, H.G. Arango, D.B. Haidvogel, C.S. Lichtenwalner, S.M. Glenn, and K.S. Hedstrom. A regional ocean modeling system for the long-term ecosystem observatory. *Journal of Geophysical Research*, 110, 2005.
- P. Yu, A. Kurapov, G. Egbert, J. Allen, and P. Kosro. Variational assimilation of HF radar surface currents in a coastal ocean model off Oregon. *Ocean Modelling*, 49:86–104, 2012.
- W. Zhang, J. Wilkin, and H. Arango. Towards an integrated observation and modeling system in the New York Bight using variational methods. part I: 4dvar data assimilation. *Ocean Modelling*, 35:119–133, 2010.