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

Deliverable 5.4:
Assimilation experiment of new observation types in large-scale ocean models

Due date: 31/10/2015

Delivery date: 31/10/2015

Delivery type: Report , public

Jean-Marie Beckers Alexander Barth
University of Liège, BELGIUM

Peter Jan Van Leeuwen
University of Reading, UK

Lars Nerger
Alfred-Wegener-Institut, GERMANY

Arnold Heemink Martin Verlaan
Delft University of Technology, NETHERLANDS

Pierre Brasseur Jean-Michel Brankart Emmanuel Cosme
Pierre-Antoine Bouttier Guillem Candille
CNRS-LEGI, FRANCE

Pierre de Mey
CNRS-LEGOS, FRANCE

Laurent Bertino
NERSC, NORWAY

Chapter 1

Introduction

This deliverable is an update of deliverable 5.3, which is thus obsolete.

The purpose of this deliverable is to present assimilation experiments that have been conducted to assess new altimetric observations (SANGOMA task 5.3). As mentionned in the DOW, special studies will be designed to see how the spatial and temporal characteristics of the observations can be taken into account (see chapter 3 below).

This deliverable is based on the combination of two studies:

- the first one deals with the assimilation of the new SARAL-AltiKa altimetric data,
- the second one deals with the assimilation of the SWOT wide-swath altimetric data.

The first one is a system-oriented study dealing with the assimilation of real data in a realistic ocean model configuration (SANGOMA large case benchmark). It is a direct application of the method proposed in SANGOMA deliverable 4.5 on the problem of assimilating observations from SARAL-AltiKa. The point is thus to use ensemble methods to investigate the potential benefits of this new data set.

The second one is a method-oriented study dealing with the assimilation of synthetic data in an idealized ocean model configuration (SANGOMA medium case benchmark). The point is here that observation error correlation is a very imoprtant feature of the SWOT observations, and to propose an algorithm to properly account for this 2D correlation algorithm in ensemble data assimilation system.

The plan of the deliverable is as follows: chapter 2 describes ensemble assimilation experiments of SARAL/AltiKa altimetric observations (real data), and chapter 3 describes ensemble assimilation experiments of the future SWOT wide-swath altimetric mission (twin experiments).

Chapter 2

Ensemble assimilation of AltiKa altimetric data

The SARAL-AltiKa satellite mission is an India-France ISRO-CNES joint project. The satellite has been launched from Sriharikota, the main ISRO launch base, on Feb. 25, 2013. The SARAL (Satellite for ARgos and ALtika) payload consists of an ARGOS instrument, and an altimetry payload including the AltiKa radiometer-altimeter. SARAL/AltiKa is flying on the same orbit as Envisat.

The special feature of SARAL/AltiKa is that it is based on a wideband Ka-band altimeter (35.75 GHz, 500 MHz), which will be the first satellite altimeter dedicated to oceanography to operate at such a high frequency. Due to the single frequency Ka-band altimeter, the enhanced bandwidth leads to a better vertical resolution. The spatial resolution is also improved, thanks to the Ka-band smaller footprint. Also, the quality of SARAL/AltiKa data in terms of accuracy, data latency and availability has allowed to rapidly make the data available, leading especially to an efficient integration in several operational systems. First results show a lesser data loss/degradation than expected due to Ka-band sensitivity to rainy and cloudy conditions.

The objective of this study is to evaluate the impact of SARAL/AltiKa data as a complement to other missions using the **large-case SANGOMA benchmark**. This is achieved by assimilating real SARAL/AltiKa observations (i.e. not with a twin experiment approach) for the period July to December 2013. The assimilation approach is exactly the same as the one that has been proposed in deliverable 4.5 to assimilate Jason-1 and Envisat observations for the period 2005-2006, so that the results should be directly comparable.

2.1 Setup of the experiments

Model configuration. The definition of the large case SANGOMA benchmark has been provided in deliverable 4.1 (also described in Candille et al., 2015).

Here is a short summary of the model configuration. The large case benchmark is based on a realistic configuration of the NEMO ocean model, for the North Atlantic Ocean, at a $1/4^\circ$ resolution, figuring the operational MyOcean systems. This model was selected because it has been used in numerous assimilation

studies before and because it is based on the NEMO ocean model which is used by most MyOcean Monitoring and Forecasting Centres.

Ensemble simulations. The 96-member ensemble is initialized in the same way as in deliverable 4.5, by introducing perturbations in the equation of state (CNRS/LGGE approach in DL4.5). These perturbations are a direct application of the method developed in Brankart (2013). The method os based on the idea that unresolved scales represent a major source of uncertainties in the computation of the large scale horizontal density gradient (from T/S large scale fields), and that the impact of these uncertainties can be simulated by random processes representing unresolved T/S fluctuations.

As in deliverable 4.5, the reliability of the ensemble simulations produced with this method is checked using rank histograms. Since the same method is used to generate the ensemble, the same conclusions applies: the ensemble is reliable in the Gulf Stream region (for which it was designed), and underdispersive everywhere else. The behaviour of data assimilation will thus only be evaluated in this region (as for the experiment described in deliverable 4.5).

Assimilation system. The same ensemble assimilation method as in deliverable 4.5 is also used: ensemble Kalman filter, with square root analysis scheme, with 10-day assimilation cycles (to fit with the 10-day cycle of Jason data), and an Incremental Analysis Update (IAU) algorithm to progressively introduce the assimilation increment in the simulation.

In addition, to avoid the spurious effect of inaccurate long-range correlations the update is also performed with a localisation process: the local assimilation areas are limited by a radius of 4.5° (≈ 450km at 30N) and the observation influence is defined by Gaussian functions with standard deviation of 1.5° (≈ 150km at 30N).

Observation system. As in deliverable 4.5 (CNRS/LGGE contribution), the experiment focuses on the assimilation of altimetric observations, with the aim of controlling the eddy dynamics in the Gulf Stream region. The observations are along-track data coming from two different satellites: 10-day cycle Jason-2 (long inter-tracks) and 35-day cycle AltiKa (short inter-tracks).

As compared to the experiment presented in deliverable 4.5, Jason-1 data are thus replaced by Jason-2 data (which display the same characteristics), and Envisat data are replaced by AltiKa data. With this new experiment, it is thus possible to evaluate the benefit brought by AltiKa (Ka band altimetry) as compared to Envisat (Ku band altimetry).

2.2 Results of the experiments

As a first illustration of the assimilation process described in the previous chapter, local time evolutions of the 96-member ensemble is shown in Figure 2.1 for the 6

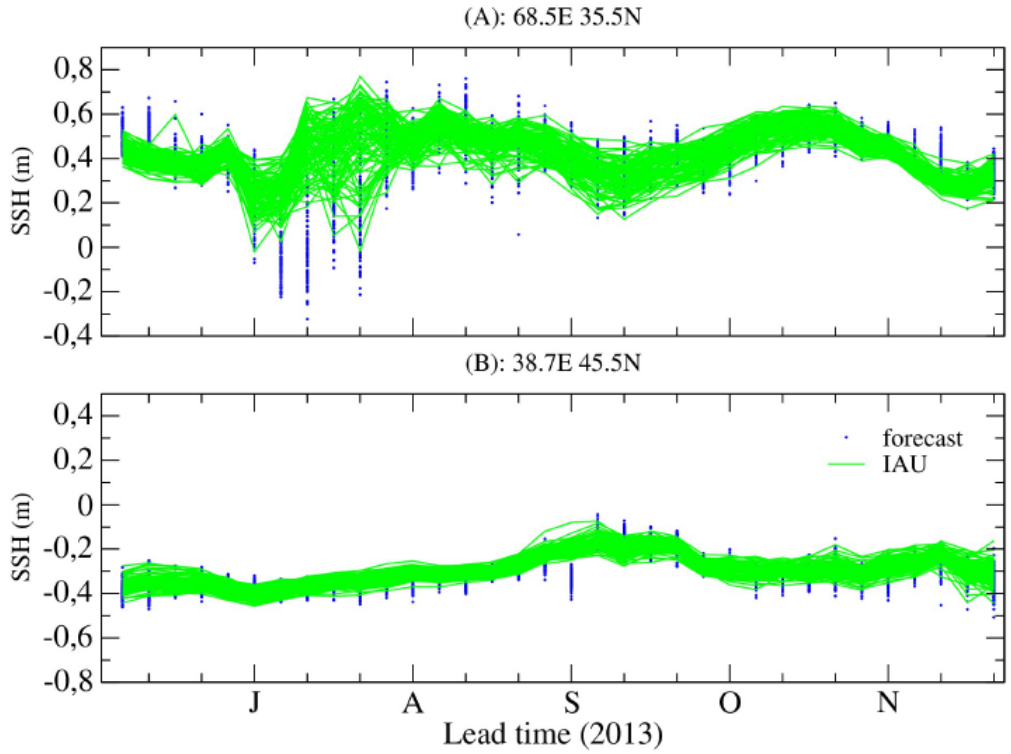

Figure 2.1: SSH time series of the ensembles (forecast and IAU).

months of the assimilation experiment. As in deliverable 4.5, the updated ensemble (blue dots for the forecast ensemble and green curves for the IAU ensemble) presents a noticeable spread reduction as compared to the free run.

As a diagnostic of the impact of data assimilation on the observed variable (SSH), Figure 2.2 shows the time series of the ensembles standard deviation averaged over Gulf Stream. The main point here is the reduction of the standard deviation reduction with the introduction of the altimetric corrections. This reduction is effective from the first assimilation cycle. After that, the averaged standard deviations are globally stabilized in the subsequent assimilation cycles. There is also a clear reduction of the standard deviations of the IAU ensembles as compared to the forecast, meaning that the stochastic perturbations are strong enough to produce a significant spread in 10 days and thus to avoid the ensemble collapse without introducing any inflating factor in the ensemble data assimilation system.

As a diagnostic of the impact of data assimilation on independent variables, Figure 2.3, shows a comparison of the forecast and updated ensemble with ARGO temperature and salinity profiles. Model equivalent of the ARGO profiles have been computed by a 4D interpolation at the exact ARGO time and locations. These results illustrate the benefit of AltiKA altimetric observations to reconstruct the 3D thermohaline structure of the ocean (which may be linked to the better accuracy and resolution of AltiKa measurements), and show the benefit that can be expect from simulating uncertainty (using ensemble) on the comparison between

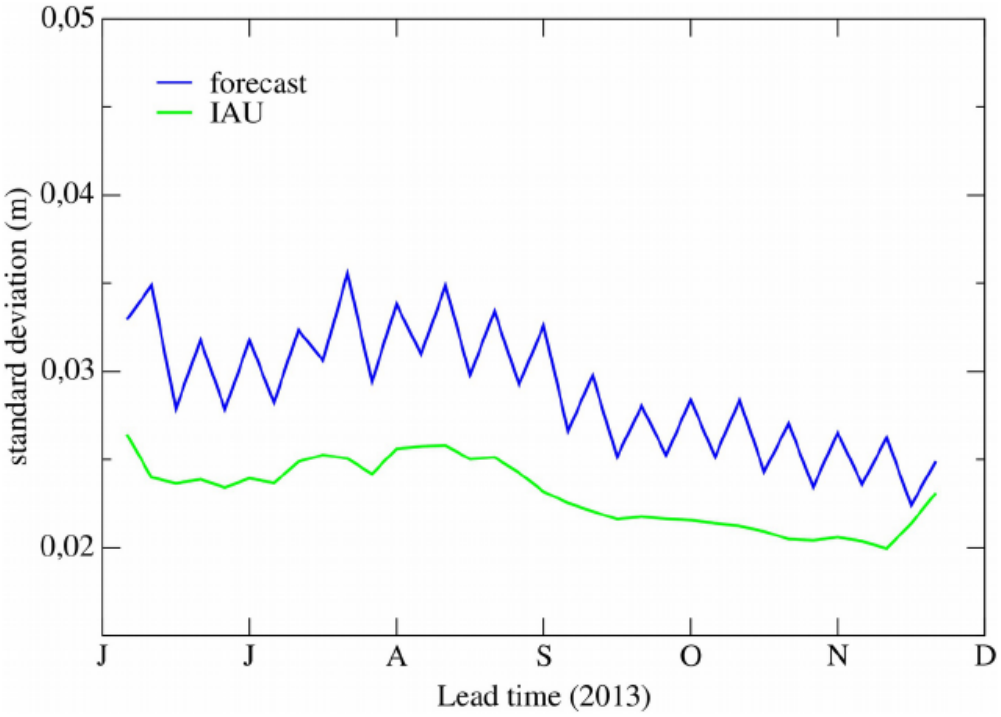

Figure 2.2: SSH ensemble spread (forecast and IAU).

simulation and observations.

2.3 Conclusion

Jason-2 and SARAL/AltikA observations have been assimilated in a 1/4° model of the North Atlantic (NATL025) for the period July to December 2013. For this experiment, the method proposed in SANGOMA WP4 (deliverable 4.5) has been applied. The core ingredients of this method are: stochastic perturbations of T/S in the equation of state to simulate the effect of unresolved scales, 4D update of the forecast ensemble over 10-day assimilation windows (using equivalent of all ensemble members at appropriate time), and objective probabilistic validation. The results obtained in 2013 with Jason-2 and SARAL/AltikA are quite comparable with those obtained in 2005–2006 with Jason-1 and Envisat. A general improvement is observed regarding the comparison to the ARGO temperature and salinity profiles. It is not clear however that this improvement can be directly attributed to the use of the AltiKA observation dataset.

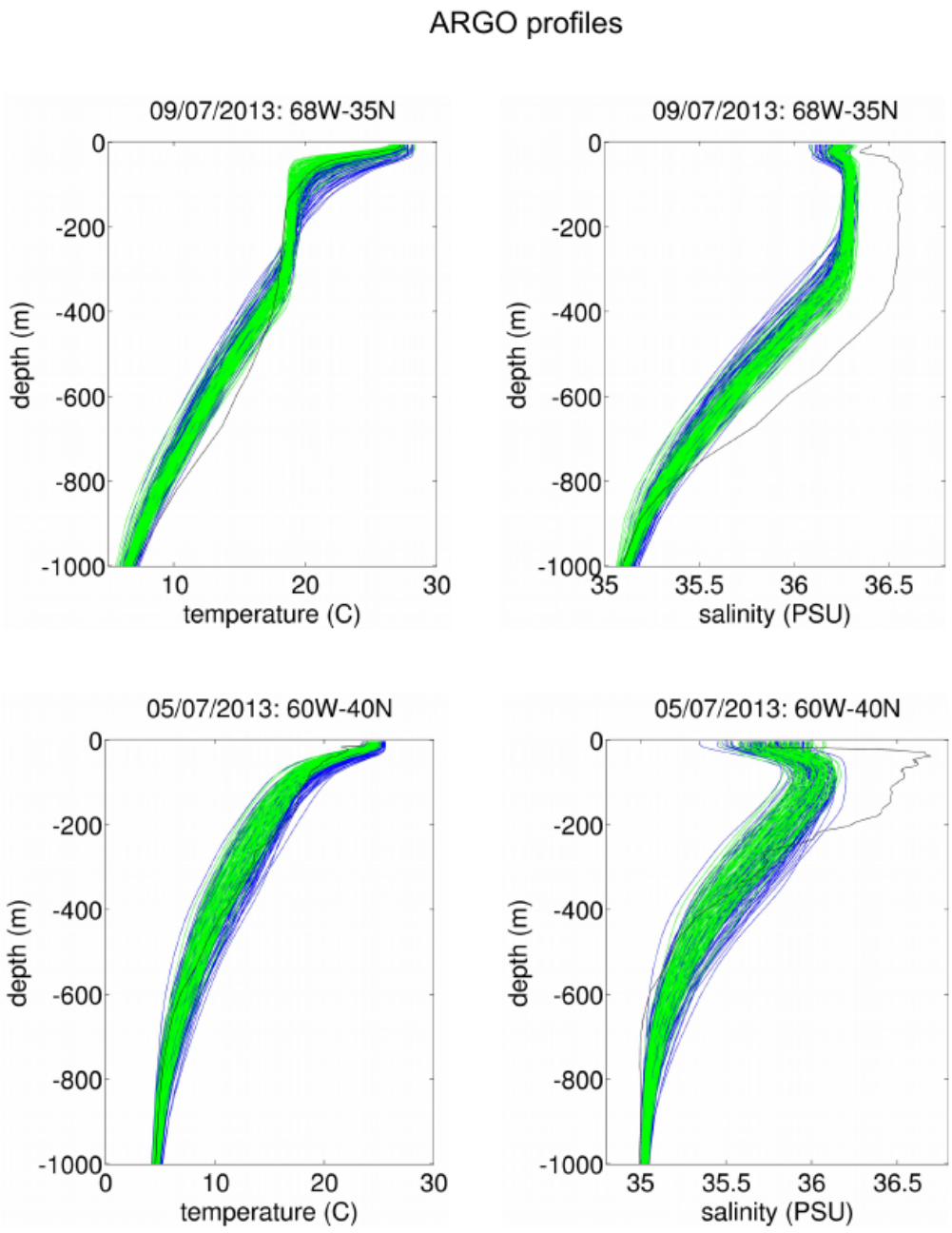

Figure 2.3: Comparison to ARGO profile data (forecast and IAU).

Chapter 3

Ensemble assimilation of SWOT altimetric data

The arrival of wide-swath sea surface height measurements, planned for 2020 [e.g., the Surface Ocean and Water Topography (SWOT) Mission] promises to be a cornerstone for oceanography. It is intended to provide high resolution ($1km$) bi-dimensional measurements of the sea surface topography with an estimated error of $0.09cm$. Such observations are of great interest especially for operational oceanography centers, since they might permit to properly constrain, through Data Assimilation (DA) techniques, a wide range of processes, from submesoscale ($O(10km)$) to mesoscale ($O(100km)$) processes.

Data assimilation methods such as Kalman Filters and variational methods are widely used by forecast centers to produce forecast initial conditions. These methods are usually very efficient as long as the matrix $\mathbf{R}$, parameterizing the observation error covariance structure is diagonal, i.e. when the observation errors are not correlated.

The assimilation of SWOT measurements may be a challenge for the data assimilation algorithms currently used by the prediction centers to assimilate altimetric observation since the errors associated with these measurements are expected to be highly correlated in space as well as the number of available observations will significantly increase. Therefore, it is desirable to still relying on the algorithm for which the complexity is linear in the number of observations and at the same time it is necessary to take into account for the error correlations to extract the most of information from the observations. The problem of dealing with correlated observation errors in DA methods is not new but the prohibitive costs of explicitly taking into account for the error correlations has lead to the use of ad-hoc solutions, must of time far from optimal. Two common solutions are data thinning, which consists in discarding observations that are too close to each other in order to assimilate only uncorrelated observations, and the use of a diagonal matrix containing the diagonal entries of $\mathbf{R}$ inflated by a factor greater than one, which in practice consists in giving less weight to each observation since they have correlated errors.

In this study (described in more details in Ruggiero et al., 2015), we develop and apply an algorithm to account for the correlations of the errors expected in the SWOT measurements while keeping using a diagonal $\mathbf{R}$. This is done by

exploiting the idea of transforming the observation vector so that the observation error covariance matrix is diagonal. To be efficient, the transformation must be very simple. In the followinf, we will use the first and second order spatial derivatives of the original observation. This has a double benefit: first, it is a local transformation, and is easily taken into account when spatial localization is used in the analysis step, and second, it takes into account the leading order dynamical property of the physical system, namely the geostrophic dynamics.

This idea of a transformed observation space is not new and is presented in the work of Brankart et al., (2009). The novelty of this study is the fitting of a parametric covariance matrix to the "observed" SWOT covariance matrix by solving an optimization problem. The objectives are then:

1. to show how to account for the spatial covariances of the SWOT observation error, and;

2. to test and illustrate the method for the estimation of the ocean state.

3.1 Simulation of SWOT measurement errors

SWOT measurement errors are modeled using the "SWOT simulator" software developed at the JPL laboratory. It produces fields of SWOT-like errors and noise according to the technical requirements determined by the SWOT scientific team.

The SWOT simulator models six type of errors: Ka-band Radar Interferometer-KaRIn- error (due to thermal noise in the interferometer channel), roll error (due to oscillations of satellite platform), timing error (due to the precision of the radar timing system), phase error (due to roughness of the sea surface at the scale of the radar pulse), baseline dilatation error (due to variation of the length of the baseline) and wet troposphere error (due to the path delay of the radar pulse due the humidity contained in the troposphere).

Figure 3.1 illustrates one sample of each component of the error listed above at 9km resolution. The most important sources of error are the wet troposphere and roll errors, while the least important sources are the timing and phase errors. The spatial distribution of the error amplitude is heterogeneous, with larger errors at the outer boundary of each swath.

Figure 3.1 also suggests that timing and baseline errors are spatially correlated in the cross-swath and along-swath directions and roll and phase errors are anti-correlated in the cross-swath direction and correlated in the along-swath direction. To quantify these statistics and get an estimate of the SWOT error covariance matrix, 5000 realizations of the error are computed with the simulator at a spatial resolution of 9km.

3.2 Parameterization of the covariance matrix

Theoretical aspects. To evaluate the observation term in a transformed space, it is first needed to define a linear transformation $\mathbf{T} : \vec{y^+} = \mathbf{T}\vec{y}$, with $\vec{y^+} \in \mathbb{R}^z$ for $z \geq m$, and an appropriated diagonal covariance matrix $\mathbf{R}^+ \in \mathbb{R}^{z \times z}$, which is

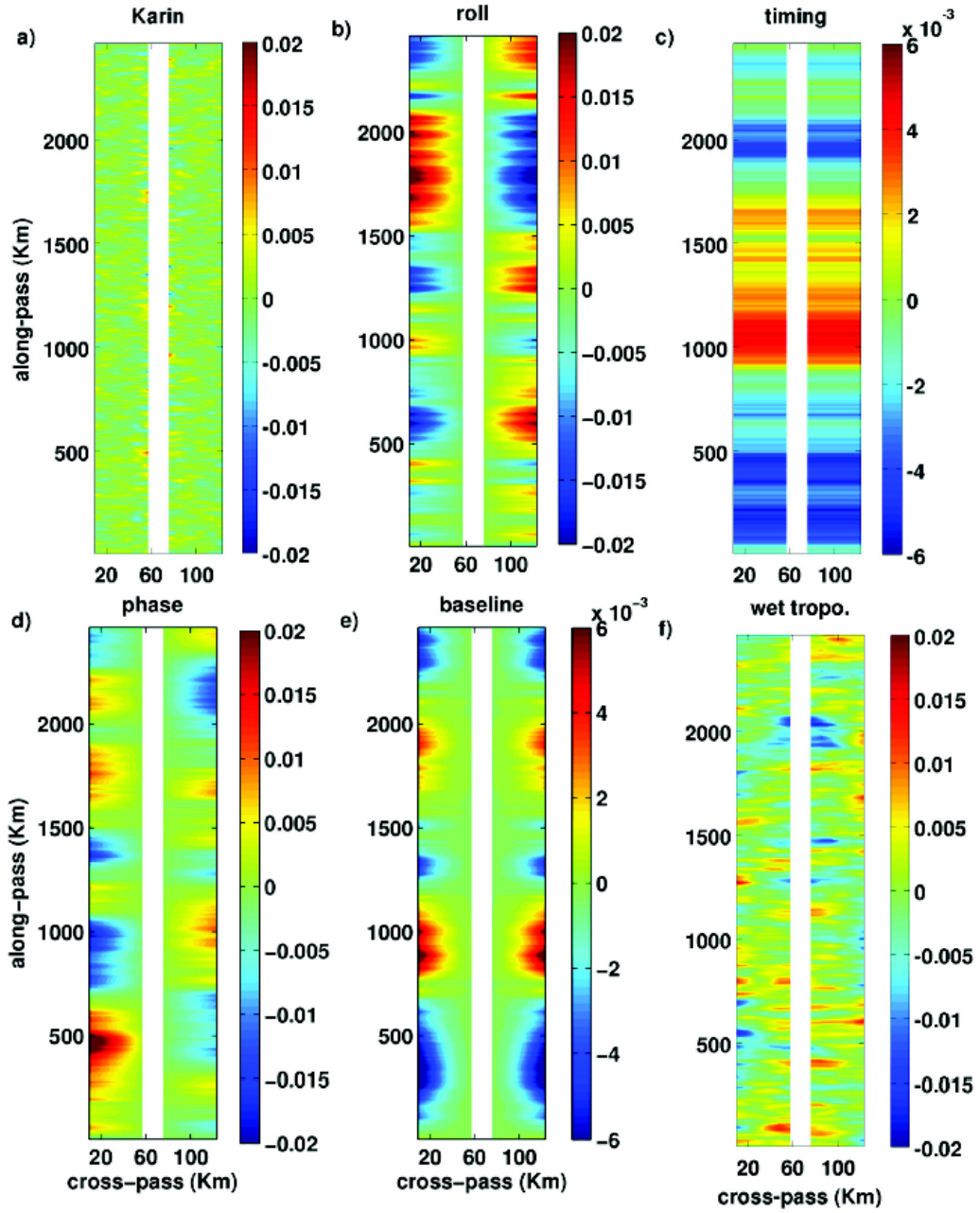

Figure 3.1: Errors as modeled by the SWOT simulator: a) Karin error, b) roll error, c) timing error, d) phase error, e) baseline dilation error and f) wet troposphere error.

associated to the transformed observation vector $\vec{y}^{+}$. Then, one may see that the observation term in the original space is equal to the observation term in the transformed space:

$$(\vec{y} - \mathbf{H}\vec{x})^T \mathbf{R}^{-1} (\vec{y} - \mathbf{H}\vec{x}) = (\mathbf{T}\vec{y} - \mathbf{T}\mathbf{H}\vec{x})^T \mathbf{R}^{+^{-1}} (\mathbf{T}\vec{y} - \mathbf{T}\mathbf{H}\vec{x}) \tag{3.1}$$

if

$$\mathbf{R}^{-1} = \mathbf{T}^T \mathbf{R}^{+^{-1}} \mathbf{T}. \tag{3.2}$$

Therefore, our problem is to find an appropriate diagonal matrix $\mathbf{R}^{+}$ once we are given a covariance matrix $\mathbf{R}$ and a linear transformation $\mathbf{T}$.

The principal axis theorem guarantees the existence of a transformation $\mathbf{T}$ which applied to $\mathbf{R}$ produces a diagonal matrix $\mathbf{R}^{+}$. However this is obtained through the eigendecomposition of the matrix $\mathbf{R}$, which has a computational complexity of order $O(m^3)$. In addition, in this case the matrix $\mathbf{T}$ has a complex structure and each new observation $\vec{y}_i^{+}$ is a linear combination of all original observations $\vec{y}_j, \forall j = 1, ..., m$. This is an important point since in the cases in which domain localization is used in the assimilation procedure, it requires the computation of the eigendecomposition of the associated local covariance matrix at each analysis step.

The main interest of the presented method is to use simple local linear transformations for which the cost of calculating $\mathbf{T}\vec{y}$ remains linear on the number of observations. As it is shown in Brankart et al. (2009), when $\mathbf{R}$ has some specific form, such as isotropic exponential correlations for instance, or is characteristic of a random function governed by a difference equation, there exists a transformation $\mathbf{T}$ and a diagonal matrix $\mathbf{R}^{+}$ such that $\mathbf{T}^T \mathbf{R}^{+^{-1}} \mathbf{T}$ exactly matches $\mathbf{R}^{-1}$.

Implementation for SWOT observations. Since the covariance matrix $\mathbf{R}$ from the SWOT simulator does not have any analytic form and is not a solution of a known differential equation, our approach is to choose a simple linear transformation and then find a $\mathbf{R}^{+}$ that minimizes the square of the residual r measured by the Frobenius norm:

$$||r||_F^2 = ||\mathbf{R}(\mathbf{T}^T \mathbf{R}^{+^{-1}} \mathbf{T}) - \mathbf{I}||_F^2 \tag{3.3}$$

where $\mathbf{I}$ is the identity matrix. We recall that the Frobenius norm is induced by the inner product of two matrices $<,>_F$ from which we may write the norm as $||\mathbf{A}||_F = \sqrt{(Tr(\mathbf{A}^T\mathbf{A}))}$. The advantage of using a norm induced from an inner product is the possibility of using orthogonal projections to solve our problem. Moreover, the chosen cost function has the advantage of avoiding the computation of the inverse of $\mathbf{R}$.

Given the nature of the observations this study is dealing with, the chosen transformation operator is composed by the identity and the first two derivatives of the observations

$$\vec{y}^{+} = \mathbf{T}\vec{y} = \begin{pmatrix} \mathbf{I} \\ \delta_a \\ \delta_c \\ \delta_a^2 \\ \delta_c^2 \end{pmatrix} \vec{y} \tag{3.4}$$

where $\delta_{[a,c]}$ is the first order spatial derivative in the along and cross swath directions and $\delta^2_{[a,c]}$ is the second order spatial derivative in the along and cross swath directions. Physically speaking, the first derivative of the SSH is related to the geostrophic velocity and the second derivative of the SSH is related to the geostrophic vorticity.

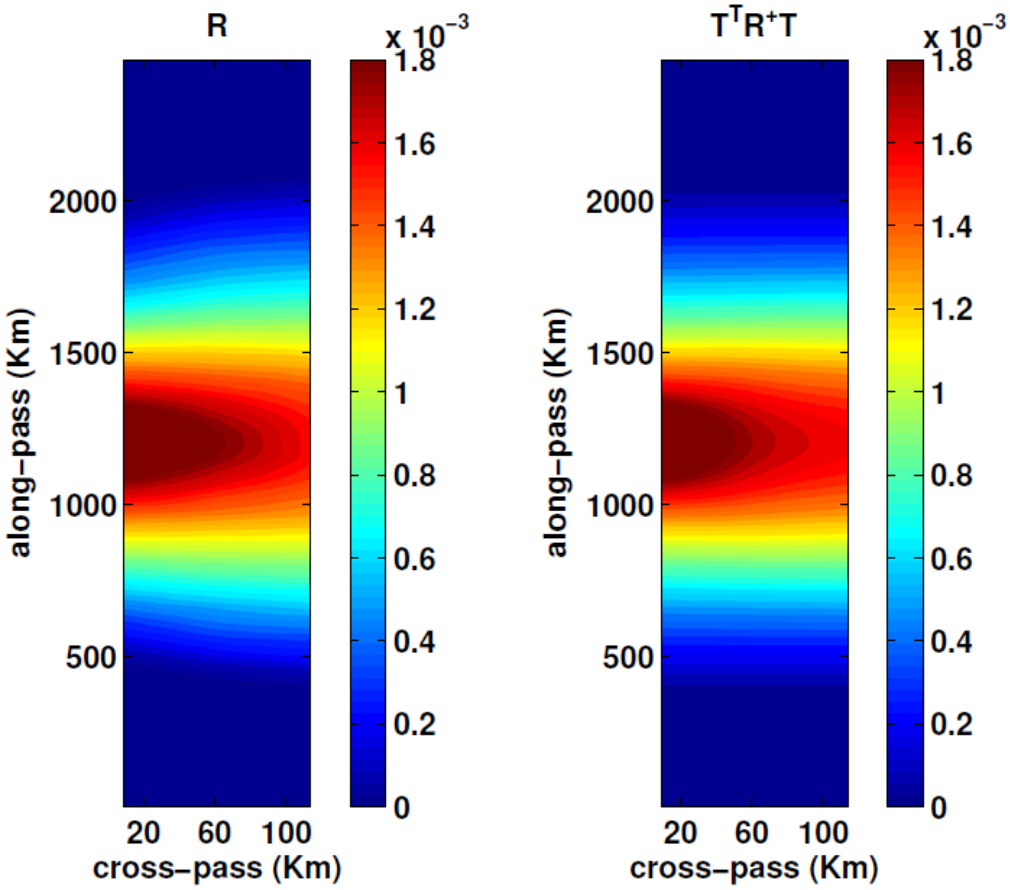

Figure 3.2: Covariance calculated for a point in the outer boundary of the left swath: (left) from the SWOT simulator $\mathbf{R}$ and (right) the identified covariance matrix $\widetilde{\mathbf{R}^{+}}$.

Figure 3.2 shows the covariance fields from the reference SWOT simulator-based covariance matrix, and from the identified covariance matrix, with respect to the same grid point. It is seen that the identified matrix is a reliable approximation of the SWOT error covariance matrix especially for the near field. In addition, the spectral characteristics of the covariance matrices has been analyzed, show-

ing that the covariance matrix coming from the SWOT simulator and the matrix identified by our method have very similar. The spectra are very close for the first 2440 singular values, which is quite a good result since the estimated rank of the matrix $\mathbf{R}$ is 3015.

3.3 Assimilation experiments

The SANGOMA medium case benchmark at $1/12^o$ resolution is used to perform the experiment.

Method. The true state $\vec{x}^t$ is drawn from a free run simulation. The observations are obtained by interpolating the sea surface height field from the true state onto the SWOT swath and then by adding to it one realization of the error from the SWOT simulator.

The background state $\vec{x}^b$ is drawn from a Gaussian distribution $\mathcal{N}(\vec{x}^t, \alpha\mathbf{P}^b)$, where $\vec{x}^t$ is the true system state and α is a factor used to reduce the spread of the background perturbations. The α parameter is used to make the variance level of the background error closer to the variance of the observation error. This situation is expected in an assimilation system after it has reached its asymptotic error level and for which the model error is reasonably small compared to the uncertainties in the initial condition.

Due to the limited size of the ensemble used to construct the background covariance matrix, domain localization is used to cut-off unrealistic long range correlations due to sampling errors. In this method each model grid point is analyzed independently from the others and only observations that lie within a ball of radius 600km centered at the analyzed point is taken into account. In addition, the inverse of the observation error covariance matrix is Schur multiplied by a Gaussian function to ensure that the observations that are closer to the analyzed point have greater weights.

To give a statistical meaning to our experiments, an ensemble of 100 backgrounds states is drawn from the Gaussian distribution $\mathcal{N}(\vec{x}^t, \alpha\mathbf{P}^b)$ and 100 analyses are made accordingly. For each set of 100 background states three experiments using different approximations of $\mathbf{R}$ a re conducted: (i) using a diagonal matrix containing only the diagonal terms of the original covariance matrix $\mathbf{Rd} = diag(r_{11}, ..., r_{mm})$, and (ii) using our parametric matrix $\mathbf{R}^+$. We recall that in latter case the assimilated observations are sea surface height and its first and second derivatives.

The assimilation method is the Local Ensemble Transformed Kalman Filter (LETKF).

Results. Starting from the experiment using $\mathbf{Rd}$, Fig 3.3 shows the rms error and the square root of the diagonal of the posterior matrix calculated by the filter, $\sqrt{diag(\mathbf{P}^a)}$, for the SSH, which is the observed variable, meridional and zonal surface velocities. The filter produces larger rms error than the background rms

error for the SSH, and meridional velocity. Zonal velocity error is reduced, but there are some regions with larger error notably at the southern portion of the observed region. In terms of coherence, for the SSH variable the $\sqrt{diag(\mathbf{P}^a)}$ has collapsed, i.e has very low value. Therefore, we say there is no coherence between the statistics produced by the filter and the true statistics. For the velocities and temperature, the values of the standard deviation produced by the filter are not as low as for the SSH but there is no coherence. These results show that by neglecting the observation error covariance, too much confidence is given to the observation and this has a double effect: (i) the analysis field is strongly contaminated with the error from the measurements resulting in high rms error, and (ii) the analysis error covariance matrix does not represent the uncertainty of the system.

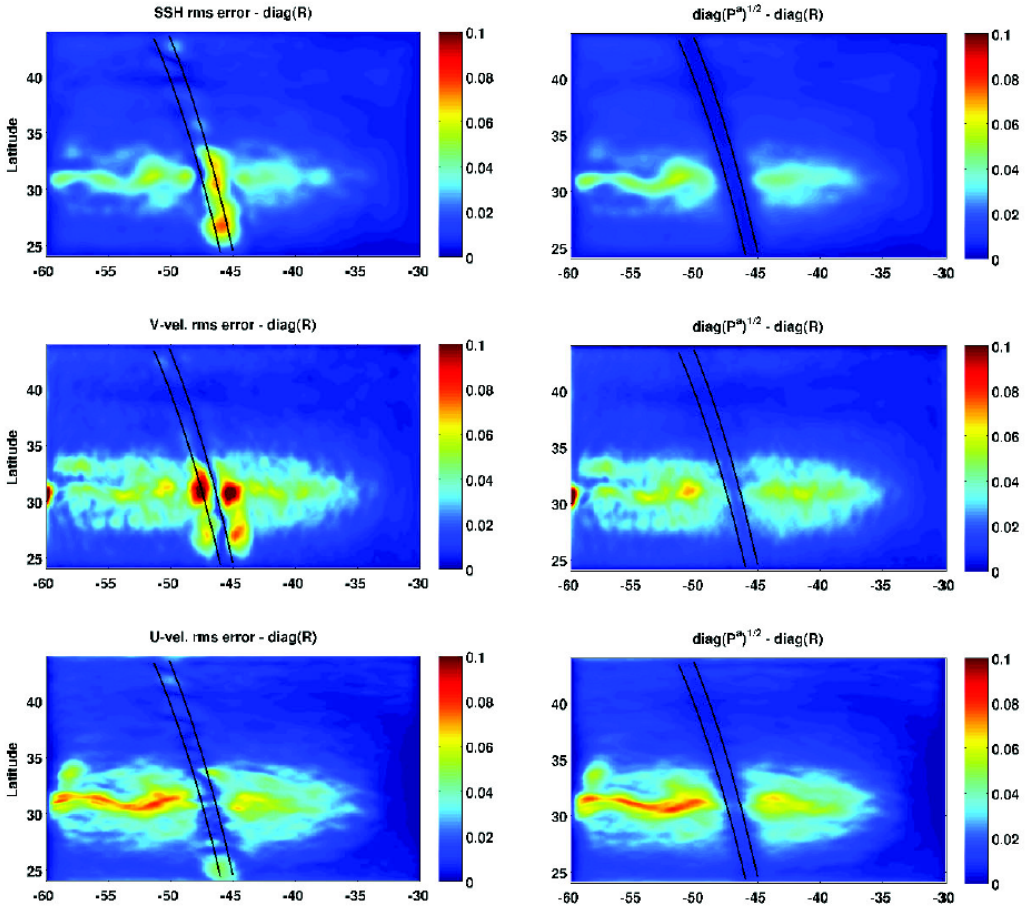

Figure 3.3: Rms error calculated from the ensemble of analyses (right panels) and error standard deviation calculated by the filter (left panels). Results obtained from the experiment that uses $\mathbf{Rd}$ as the observation error covariance matrix. From top to bottom: SSH, surface meridional velocity (V), and surface zonal velocity (U).

Figure 3.4 shows the results of the experiment using $\mathbf{R}^+$. Among the three conducted experiments, this one produces the smallest rms error for all variables

and a good coherence between rms errors and $\sqrt{diag(\mathbf{P}^a)}$. Definitely, including additional information about the first and second derivatives of the observation in the observation vector to parametrize the error spatial covariances has a strong and beneficial effect on the analyzed field, especially for the observed variable (SSH). In particular, it makes the filter capture the true uncertainty of the ocean state estimation.

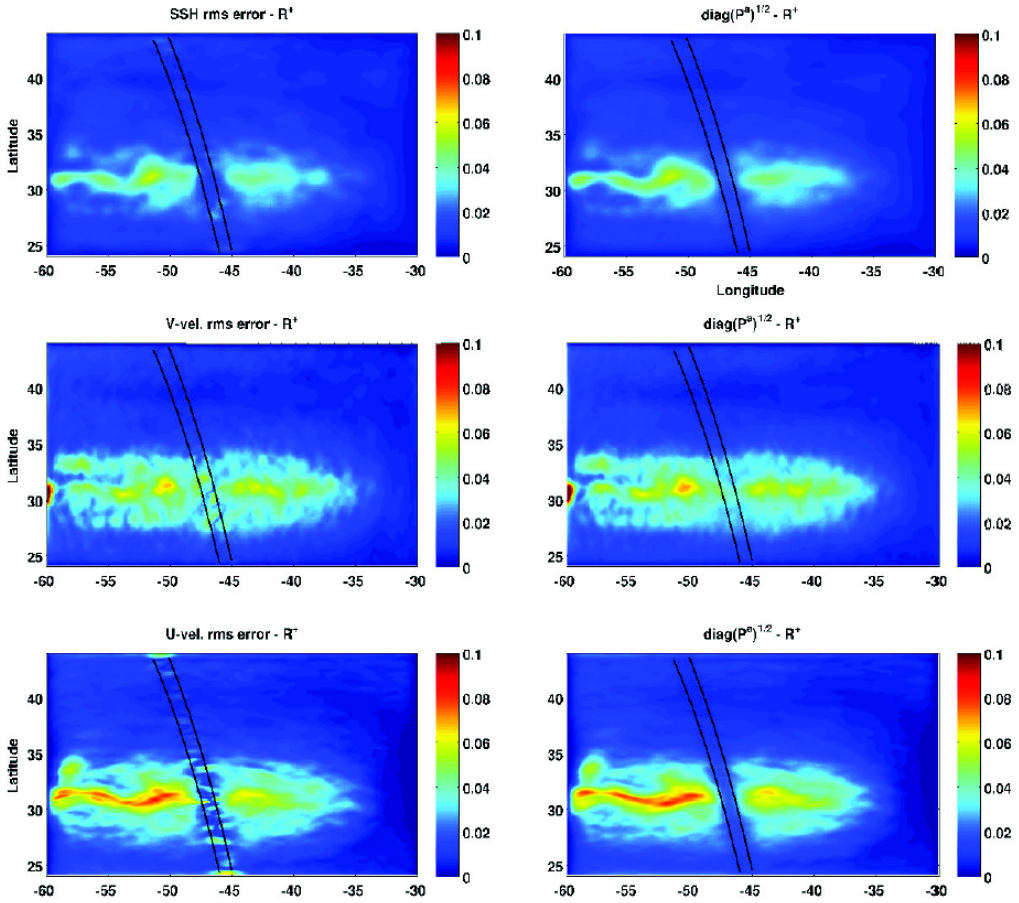

Figure 3.4: Same as for figure 3.3 but obtained from the experiment that uses $\mathbf{R}^+$ as the observation error covariance matrix.

3.4 Conclusion

This study has explored the idea of extending the observation vector with a simple local transformation of the original observation vector to simulate spatial correlations presented in the observation error. The transformation is designed by augmenting the observation vector with the first and second derivatives of the original observation. A diagonal covariance matrix, whose entries were obtained through the solution of an optimization problem, is then associated to this new set of observations. The results have proven that the used linear transformation

is appropriate for the assimilation of SWOT-like observations with 9km resolution, with smaller rms error and better coherence of the filter statistics. A unique answer to how important it is for an assimilation system to take into account for the observation error correlation is however difficult to obtain. First because least square based data assimilation systems rely on the prescription of the two matrices $\mathbf{P}$ and $\mathbf{R}$; and second because even if both are perfectly specified, it is the spectral content of each one that will determine the resulting weights given to observation and model.

Chapter 4

References

Brankart J.-M., Ubelmann C., Testut C.-E., Cosme E., Brasseur P., Verron J., 2009. Efficient parameterization of the observation error covariance matrix for square root or ensemble Kalman filters: application to ocean altimetry. Monthly Weather Review, 137(6), 1908–1927.

Brankart J.-M., 2013. Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling. *Ocean Modelling*, **66**, 64–76.

Candille G., Brankart J.-M., and Brasseur P., 2015. Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation. *Ocean Science*, **11**, 425-438.

Abdelnur Ruggiero G., Cosme E., Brankart J.-M., Le Sommer J., Ubelmann C., 2015 : An efficient way to account for error correlations in the assimilation of observations from the future SWOT High-Resolution altimeter mission. Journal of Atmospheric and Oceanic Technology, submitted.