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

Second progress report

J.-M. Beckers A. Barth Y. Yan M. Canter
University of Liège, BELGIUM

P.-J. Van Leeuwen S. Vetra-Carvalho
University of Reading, UK

L. Nerger P. Kirchgessner
Alfred-Wegener-Institut, GERMANY

A. Heemink N. van Velzen M. Verlaan U. Altaf
Delft University of Technology, NETHERLANDS

P. Brasseur J.-M. Brankart G. Candille S. Metref
F. Garnier
CNRS-LEGI, FRANCE

P. De Mey
CNRS-LEGOS, FRANCE

L. Bertino F. Counillon
NERSC, NORWAY

Chapter 1

Summary

1.1 Project objectives

SANGOMA will provide new developments in data assimilation to ensure that future operational systems make use of state-of-the-art data-assimilation and related analysis tools. We are a European network of expert teams in advanced data assimilation. In the project we will extend existing modular data assimilation systems that have high flexibility in type of ocean model and assimilation method. Following specific design rules, new modules can be used in different modular systems. The systems will allow for efficient operational testing of the latest data assimilation methods, and quick comparison of assimilation methods for operational use. Furthermore, we will develop and implement modules that objectively determine the impact of existing and new observation types. This dedicated web portal will provide access to validated products, including documented performances on a variety of test cases. Consolidated versions will be made available to the science community and Marine Forecasting Centres with indications on best practice implementation. Workshops and summer schools on advanced assimilation methods and modular systems will ensure fast and efficient training of next generation oceanographers, ensuring world-leading operational oceanographic products for costumers and decision makers.

1.2 Second period achievements

Coherently with the project plan, the second period saw the highest activities in implementing and harmonizing tools (WP2), development of new data-assimilation techniques (WP3) and implementations of benchmarks of different complexity (WP4). In WP1 necessary adjustments to the specifications were prepared. WP2 made the initial list of tools fully compliant with the SANGOMA specifications, added examples to each tool and prepared the list of candidate tools to be added. WP3 provided some new ideas among which some ways to deal with non-linearities in advanced filter methods (Equivalent weight particle filters and Multivariate rank histogram filters) and stochastic perturbations via a state equation. WP4 fully described and implemented the different benchmarks and now comparisons between the different assimilation techniques can start. WP5 implemented all planned models and can now assess the observing systems in these applications.

1.3 Expected impact of project

The developments of SANGOMA will also serve costumers of MyOcean products, which is the first European project dedicated to the implementation of the GMES Marine Core Service for ocean monitoring and forecasting. For this purpose, we will concentrate on data-assimilation methods that deliver probabilistic information on the products. To this end, existing ensemble methods will be included and new methods that allow for nonlinear and non-Gaussian systems will be developed.

Chapter 2

Objectives and achievements of the second period

Coherently with the project plan, the second period saw the highest activities in implementing and harmonizing tools (WP2), development of new data-assimilation techniques (WP3) and implementations of benchmarks of different complexity (WP4).

2.1 WP1: Harmonization

2.1.1 Executive summary

The specification part was essential in the first reporting period and for the second reporting period only adaptations to the specifications were planned based on feedbacks and the additional tools to be added.

2.1.2 Scientific and technical highlights

The proposed formats and interfaces defined in the first reporting period were well accepted (internally and externally to the project, as proved by the survey) and implementation started. Only minor details regarding the definitions remained to be clarified and were discussed at the second progress meeting:

F90/F77 : As legacy of F77 is still important it was decided not to force people to use .F90 interfaces and keep things as simple as possible. For Fortran 90 users, it would be useful to provide a SANGOMA module, so that the compiler can automatically check if all mandatory parameters are present and have the correct type. Partners will investigate a way to provide a Fortran 77 library and a Fortran 90 module without duplicating the code base.

Single/double : some special handling is needed when interfacing with BLAS/LAPACK libraries when different precisions are used. As we cannot serve all situations and as users will generally optimize anyway the codes (memory/speed) when necessary it is decided to focus on the readability of the codes and examples provided with the codes rather than on fully flexible codes with selectable precision. Therefore double precision is retained as reference for SANGOMA.

Storage ordering : The memory storage of arrays (as specified in the common data specification) is column-major (which is used in Fortran and Matlab/Octave). Row-major languages such as C and Java need to change the order of the indices. For performance reason, no explicit memory reordering is performed. The adopted approach is equivalent to the one used in the NetCDF library where the order of the dimensions in a C program is reversed compared to the order of dimensions in a Fortran program.

2.1.3 Deviations from DOW

The WP advances as planned, deliverable 5.1 (living document) will need to be updated based on the adjustments discussed during the progress meeting.

2.1.4 Outreach

The specifications have been presented in the MyOcean Science days and the on-line survey `http://www.surveymonkey.com/s/ZX3P9D8` which is still open.

2.2 WP2: Sharing and collaborative development

2.2.1 Executive summary

The goal of this work package is to develop and share tools for data assimilation in a collaborate way. The tools are categorized according to four categories: diagnostic tools, transformation tools, perturbation tools, and utilities. The tools are implemented following the standards set up in WP1. Both, tools coded in Fortran and as scripts for use with Matlab or Octave are provided. The tools are shared via a revision server as a project at `http://sourceforge.net/projects/sangoma/`. In addition, bundled release versions of the tools are provided.

2.2.2 Scientific and technical highlights

During the second period, the set of SANGOMA tools was extended. The tools are coded in either Fortran or as scripts for Matlab, compatible with Octave. All tools that were part of the first software release (Version 0) were adapted to the data model standard of WP1. I.e. now all tools coded in Fortran use a C-bind interface to ensure the compatibility with programs coded in C. Several additional tools have been implemented.

The focus of the work was on the diagnostic tools. Here, several new tools to compute diagnostics of the fidelity of the ensembles used in the assimilation methods were implemented. In particular, the computation of advanced metrics like the continuous ranked probability score (CRPS), the Brier score and the random centered reduced variable (RCRV) is now provided by the tools (for the definitions of the metrics see Deliverable 4.1). For particle filters, metrics like the relative entropy, the mutual information and the sensitivity matrix can be computed. In addition, tools to compute histograms as well as higher-order ensemble statistics (skewness, kurtosis) are provided. With these tools, the performance diagnostics for the benchmarks in WP4 can be computed.

For the generation of ensemble perturbations, now a tool to compute pseudo-random fields in the spectral space with specified correlation lengths is provided, next to a tool to generate the perturbations in the form of empirical orthogonal functions from a singular value decomposition. Further, ensembles can be generated that fulfill weak linear constraints.

In the Utilities category, tools have been added that compute a proper orthogonal decomposition (POD) of a matrix of ensemble snapshots. A second tool utilizes the POD modes to compute a cost function and its gradient vector along the POD-directions.

Next to the tools, simple example test cases are provided. They allow a user to experiment with the tools. In addition, the examples demonstrate how a tool can be used from within an assimilation framework.

All tools are documented in Deliverable 2.4, which also explains how to use and compile the tools. Apart from this document a README file in HTML-format is provided in the bundled release version that describes the usage of the tools.

The following tools are included in the release V1 of the SANGOMA tools:

Diagnostic Tools

Fortran

sangoma_ComputeHistogram	Compute ensemble rank histograms
sangoma_ComputeEnsStats	Compute ensemble statistics
sangoma_computeRCRV	Compute the bias & the dispersion of the RCRV
sangoma_computeCRPS	Compute the CRPS and its decomposition
sangoma_computeBrier	Compute the Brier skill score and its decomposition, and the entropy
sangoma_ComputeSensitivity	Calculate the sensitivity matrix with **H** as matrix
sangoma_ComputeSensitivity_op	Calculate the sensitivity matrix with **H** as operator
sangoma_computeMI	Calculate the mutual information
sangoma_computeRE	Calculate the relative entropy

Matlab/Octave

mutual_information	Compute mutual information in a particle filter
relative_entropy	Compute relative entropy in a particle filter
sensitivity	Compute sensitivity of posterior mean to observations in a particle filter

Perturbation Tools

Fortran

sangoma_pseudornd2D	Generate random fields with given correlation length
sangoma_MVNormalize	Perform multivariate normalization
sangoma_EOFCovar	Initialize covariance matrix from EOF decomposition

Matlab/Octave

Weakly constrained ensemble perturbations	Create ensemble perturbations that have to satisfy an a priori linear constraint

Transformation Tools

Matlab/Octave

Empirical Gaussian Anamorphosis	Determine the empirical transformation function such that a transformed variable follows a Gaussian distribution

Utilities

Fortran

sangoma_computepod	Computes dominant POD modes from an ensemble of snapshots
sangoma_costgrad	Computes the values of objective function and gradient using reduced state dimensions

Matlab/Octave

hfradar_extractf	Observation operator for HF radar surface currents

2.2.3 Deviations from DOW

Overall, the work package develops as planned. The Deliverables are submitted and Milestones are reached on time.

In the project proposal the possible use of the assimilation system DART by the project partner at University of Reading was described. As currently the DART system will undergo a substantial modification, without the involvement of the project members, it was decided not to use the DART system. Instead, an alternative coupling method of model and data assimilation basing on the parallelization standard MPI will be used, which will also be included in the PDAF-framework.

Following the recommendations of the reviewer and the advisors, the focus of the tools will be set on the diagnostic tools. As mentioned in the project proposal, routines to compute the analysis steps of the filters are not included in the tool set. The reason for this is that the analysis steps are usually much closer connected to the assimilation frameworks used by the partners as the other tools are. As such, while a generic implementation of an analysis step would ensure independence from the frameworks, it would suffer from being computationally less efficient.

2.2.4 Outreach

Publications:

Nerger and Hiller (2013)

Presentations:

Nerger et al. (2013)

2.3 WP3: Innovative Data Assimilation techniques

2.3.1 Executive summary

This WP is focusing on stochastic DA methods, development a new DA method (including non linear observation operators), comparisons of non-Gaussian assumptions and developing algorithms for assessing observing systems. This WP has the most academic freedom of research leading to interesting discussions within the project and new lines of research.

Each group made significant progress in advanced data assimilation techniques which will pave the way for additional features in future operational data assimilation.

2.3.2 Scientific and technical highlights

Classical particle filter methods are subjected to the curse of dimensionality, demanding the ensemble size to grow exponentially with the problem

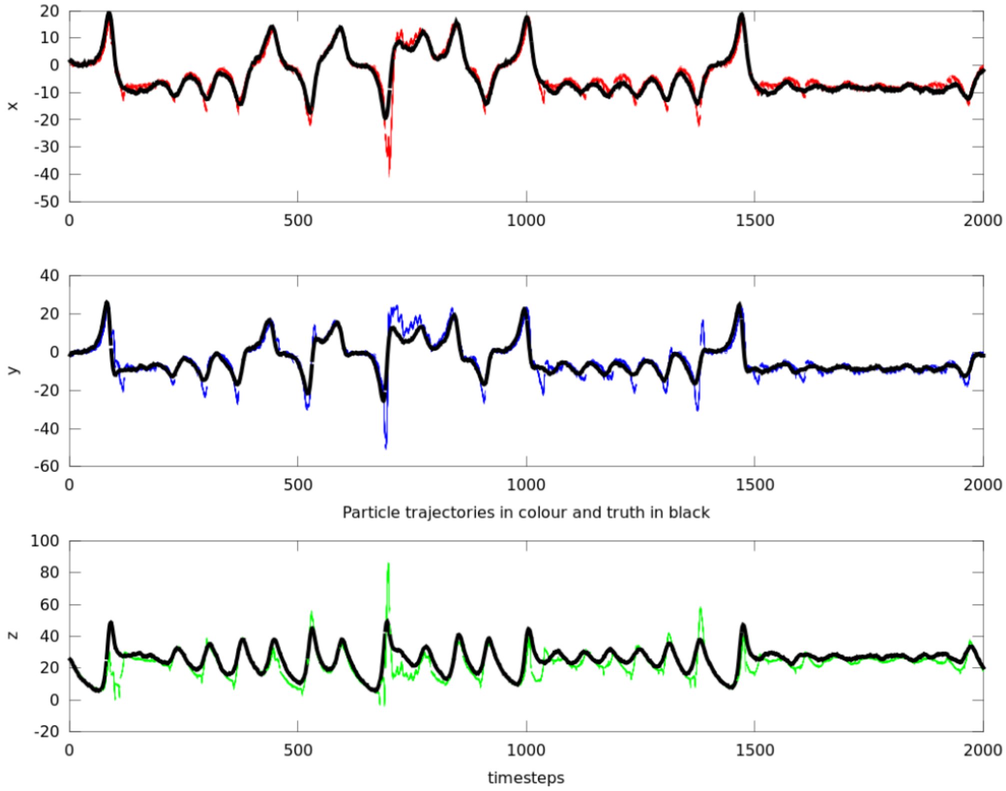

Figure 2.1: Results of EWPF application to the Lorenz 63 system when observing yz.

dimension size. A particle filter method trying to avoid the curse of dimensionality was thus developed further: The equivalent weight particle filter (EWPF) tries to avoid the collapsing of the ensembles by keeping a relatively small number of particles with equal weights at analysis time. For this, simple nudging terms are added to the equation to provide a proposal density, which modify the particle weights, and a resampling technique exploiting the maximum weight each particle can achieve. The calculation of this maximum weight is possible analytically only if the observing operator is linear. The method was successfully applied to a 2D case with a nonlinear barotropic flow. For cases with nonlinear observing operators an iterative solver to find the maximum weights was implemented and tested on the chaotic Lorenz 63 system. The most challenging nonlinear observations in this case are those which can lead to two different states (of different signs) because the observation of the product for example does not allow to distinguish which combination to retain. For details see `http://www.data-assimilation.net/Events/Year3/WP3.pdf` (Presentation at progress meeting) and `http://www.data-assimilation.net/Events/Year3/SANGOMASanita.pdf` (Presentation on nonlinear H)

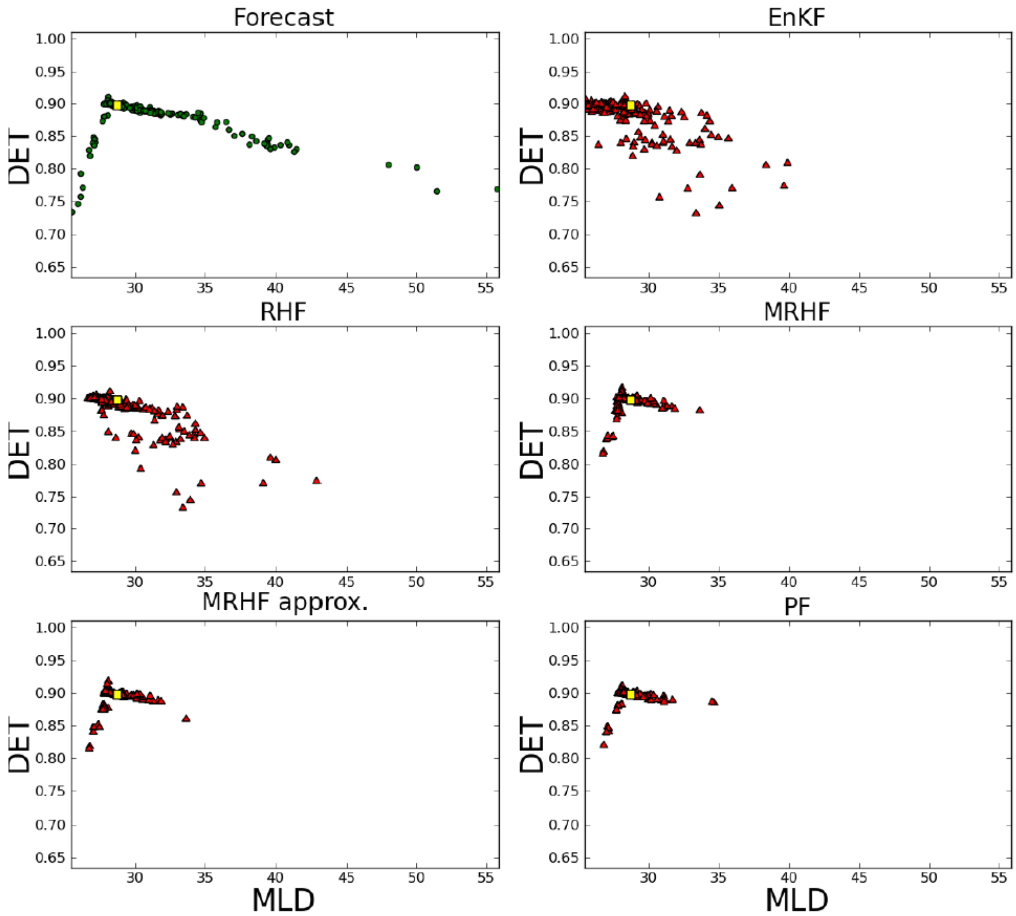

Figure 2.2: Application of the MRHF to the NATL025 implementation with bio-chemical model. Chl is observed and mixed layer depth and detritus are analysed. The members of the MRHF with or without the approximation in the joint probabilities work very well in this case.

A Multivariate Rank Histogram Filter (MRHF) exploits joint probability density functions to infer unobserved variables from observed ones. When working in a low dimensional system, the approach performs very well for non-gaussian problems, but in higher dimensions some strong hypotheses need to be formulated. By neglecting unobserved variables in some of the conditional statements a feasible MRHF is within reach but good insight in the effect of the hypotheses is needed. For details see http://www.data-assimilation.net/Events/Year3/MRHF.pdf (Presentation at progress meeting) and http://das6.umd.edu/program/Daily/slides/2.2-Metref_Sammy.pdf (Poster)

An original method to include stochastic perturbations into a model was developed. By using stochastic perturbations applied to the state equation one retrieves perturbations in which heat and salt are conserved. Also

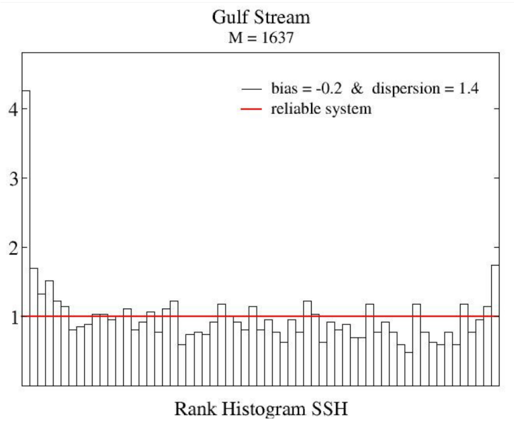

Figure 2.3: Application of stochastic perturbations on the state equation of the NATL025 implementation. SSH is observed and the rank histogram exhibits the nice ensemble spread.

the perturbations lead to nice and stable ensemble spread. For details see http://www.data-assimilation.net/Events/Year3/brankartGC.pdf (Presentation at progress meeting) and http://dx.doi.org/10.1016/j.ocemod.2013.02.004 (publication Brankart (2013))

Ensemble methods heavily exploit the reduced rank of the covariance matrix for efficient matrix inversions via the famous Woodbury formula. A new method combining such reduced rank matrices with local parametric covariances was developed to be able to combine large scale processes well modeled ensemble covariances the with smaller scale processes generally well modeled by parametric covariance functions. The combination normally does not lend itself to efficient inversion but with an iterative approach it still leads to efficient computations. In addition it allows for scale separations in the analysis. For details see http://www.ocean-sci-discuss.net/11/895/2014/osd-11-895-2014.pdf (Publication Beckers et al. (2014))

Anamorphosis (data based change of variables) is a way to deal with data having a non-gaussian distribution, by transforming the data such that the transformed variables have a gaussian distribution. The transformation was tested in a coupled physical-biochemical model and indicates better spatial correlations when using the transformation. Therefore better corrections during assimilations can be expected. For details see http://www.ocean-sci.net/8/121/2012/os-8-121-2012.pdf (Publication Brankart et al. (2012))

For operational systems, error calculations can be quite time consuming when Optimal interpolation or 3DVar approaches are used. Simplified and approximate methods can alleviate the problem as shown in Beckers et al. (2014).

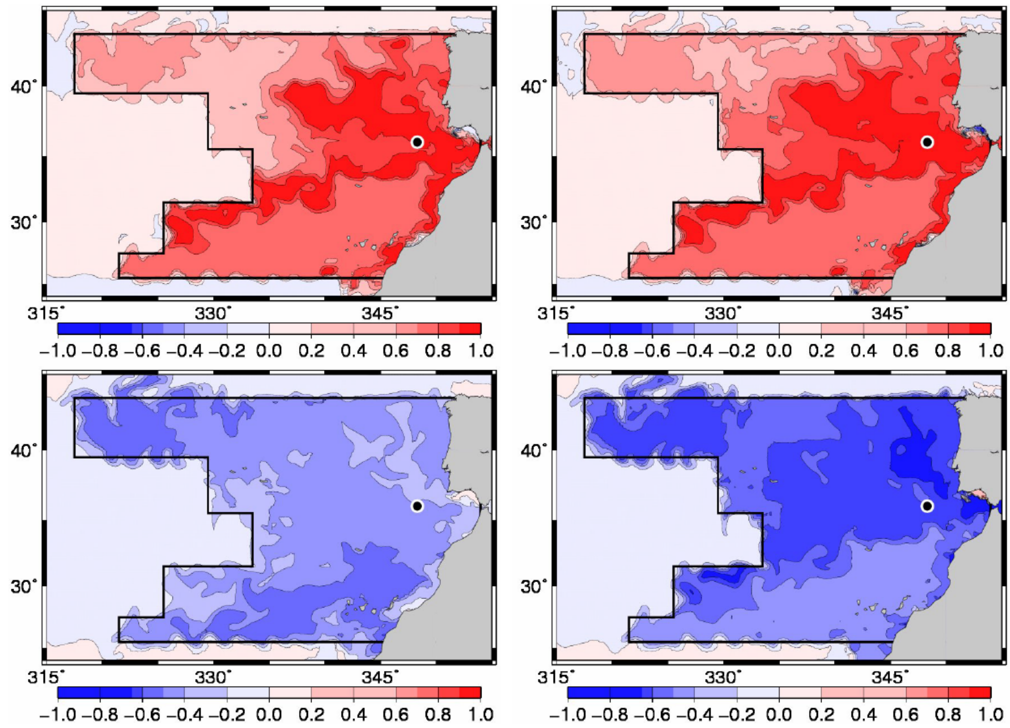

Figure 2.4: Phytoplankton (top panels) and nitrate (bottom panels) horizontal correlation structure with respect to phytoplankton without anamorphosis (left panels), and after local anamorphosis transformations (right panels).

2.3.3 Deviations from DOW

The research advances very well but need to be reflected in the living documents. DL3.1 was not submitted yet and DL3.2 which is the living document is frozen on the EU portal but needs to be updated in our web site.

2.3.4 Outreach

Several papers have already been published but others need to be submitted.

Publications:

Brankart (2013), Brankart et al. (2012), Beckers et al. (2014), Beckers et al. (2014)

Presentations:

Yan et al. (2013), Kirchgessner and Nerger (2014), Kirchgessner et al. (2013)

2.4 WP4: Benchmarks

2.4.1 Executive summary

All benchmarks have been fully documented and implemented and executed by at least two partners. Details on which perturbations and ensemble members to generate were discussed and specified.

2.4.2 Scientific and technical highlights

In the previous reporting period, the 3 SANGOMA benchmarks (small, medium and large case) had been defined, together with statistical metrics to evaluate stochastic assimilation methods (deliverable 4.1).

Since then, these 3 benchmarks have been implemented by the SANGOMA partners to evaluate various kinds of assimilation methods.

- Deliverable 4.2 lists the benchmarks that are being implemented by every SANGOMA partner, and provides the list of assimilation methods that wil be evaluated by each partner.
- Deliverable 4.3 (in draft version) provides detailed information on how the probabilistic metrics are applied on small and medium benchmarks.

At the SANGOMA progress meeting of April 2014, presentations of the various benchmark implementations have shown that valuable results have already been obtained. In particular :

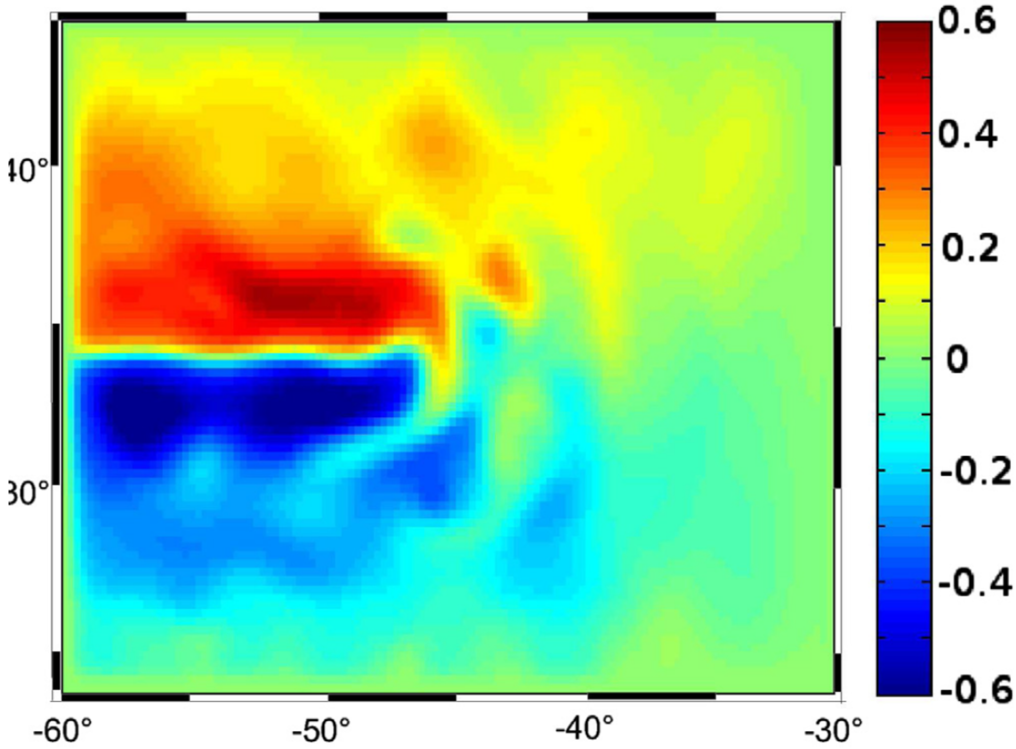

Figure 2.5: Snapshoot of SSH in the medium-size benchmark.

- Five implementations of the medium case benchmark will be available to intercompare different assimilation methods.

- Two implementations of the large scale benchmark (by GHER and LGGE as planned in the project) have been tested with a 6-month assimilation experiment (assimilating real-world ocean observations). A first evaluation of these experiments has been performed, by applying more physical metrics at GHER and more probabilistic metrics at LGGE. Also the two implementations use complementary perturbations (density or external forcing).

This WP is thus on schedule. The next steps are : (i) to finalize the experiments with the medium case benchmark and start the intercomparison exercise, and (ii) to converge on a set of physical and probabilistic metrics to evaluate the experiments made with the large-scale benchmark.

Among the highlights we can mention that the medium size benchmark was successfully implemented by a post-doc new to the assimilation toolbox and the NEMO model within 3 month. Adding then different assimilation strategies (difference incremental assimilation updates IAU with different levels of implicitness) allowed to show the benefits of using IAU(0) compared to standard intermittent updates. For details see http://www.

sciencedirect.com/science/article/pii/S1463500313002035 (Publication Yan et al. (2014))

2.4.3 Deviations from DOW

The WP is fully on track and milestones (particularly M11) reached.

2.4.4 Outreach

Publications:

Yan et al. (2014)

Presentations:

Presentation at GODAE http://www.sciencedirect.com/science/article/pii/S1463500313002035

Presentation at EGU2014 http://orbi.ulg.ac.be/bitstream/2268/165348/1/EGU2014-2576.pdf

2.5 WP5: Data Assessment

2.5.1 Executive summary

The WP tries to assess the impact of new remote sensed ocean data on the model state estimations and their potential in a data assimilation setup in a future operational context.

The new data types have beed identified and characterized. Several model implementations have been achieved for assessing observing systems and impact of new data types.

2.5.2 Scientific and technical highlights

Task 5.1 Identify new data types This taks has been completed by the Delivery of D5.1 List of remote-sensed variables with their associated characteristics (Completed, M12, all partners)

Task 5.2 Assessing observing systems Some progress has been made as part of the work from G. Candille, CNRS-LEGI, evaluating the relative impact of different altimeters constellations (Envisat and/or Jason 1) for the reduction of uncertainties and the improvement of model skills for unobserved temperature and salinity profiles. The work has used the large-scale NEMO benchmark from WP4. D5.2 Report on the impact of new ecosystem data (M36, CNRS-LEGI) is therefore on track although the contents will be related to ocean physics more than ecosystem, consistently with the work performed in WP4.

Task 5.3 Expts. Large-scale models This tasks builds up on the synthetic data experiments carried out by G. Candille (CNRS-LEGI) in WP4. The initial results on assimilating real along-track altimeter data looks promising. D5.3, D5.4 Results of a data assimilation experiment with a large-scale ocean model (ongoing V1 at M36, V2 at M48, CNRS-LEGI) are therefore on track.

Task 5.4 Exp in regional scale models The work has already advanced quite well with a 1/60th deg ROMS model of the Ligurian Sea, using realistic forcings and real observations from two WERA HF radar systems operating during 2009 and 2010. The initial work has focused on the representation of model errors and covariance localization using as success criteria both the statistical robustness of increments and the expected error reduction (as in Task 5.2)

Localization is often applied for ensemble methods to avoid long range corrections introduced by artificial correlations found when undersampling due to the number of members of the ensemble. Two problems which occur in this context can now be tackled: how to build the localization function objectively and how to maintain global conservation constraints when adding localization. For the limitation in space of the corrections based on an observation, the criterium is that the expected improvement brought by the correction should be larger than the error introduced by an incorrect specification of covariances. The latter effect is measured by a bootstrap method so that the localization function can be calculated. For maintaining conservation properties when applying the localisation approach, one has to modify the ensemble covariance such that the error variance on the conserved property is zero. This can be formulated mathematically rather easily for the forecast error but requires special care when updating the analysis error covariance. For details see `http://www.data-assimilation.net/Events/Year3/assimlocens.pdf` (Presentation at progress meeting) D5.5, D5.6 Results of a data assimilation experiment with a regional-scale ocean model (ongoing V1 at M36, V2 at M48, ULg) are both well on track to delivery.

Task 5.5 Lagrangian sea ice parameters The work has been initiated in an external collaboration with F. Massonnet at Université Catholique de Louvain la Neuve. Assimilation of satellite sea ice drift data in a NEMO-LIM model has been performed with the EnKF and global parameters of sea ice drag coefficients and sea ice strength (known as P* in the (Elastic)-Viscous-Plastic sea ice rheology) have been estimated by a technique of state augmentation. The results show strong improvements for the estimation of two parameters (one drag coefficient and P*) but the only the ratio of air and ocean drag coefficient can be estimated correctly. The results are being compiled in a publication (in review) and will start as a first step for the estimation of Lagrangian parameters in TOPAZ. The latter task has not yet started.

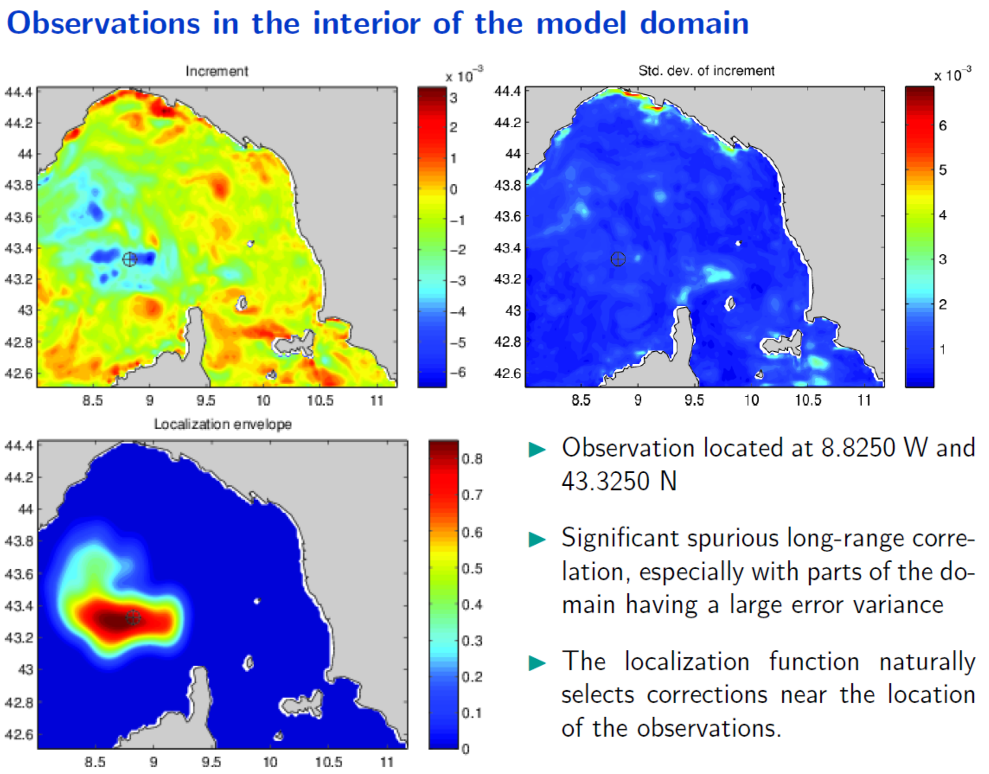

Figure 2.6: Application of the objective localisation determination in the ROMS Ligurian sea implementation. Velocity is observed by radar and the original increment shows spurious long range correlations particularly in regions where the error covariance is uncertain. With the objective localisation detection increments will be applied only in relevant points.

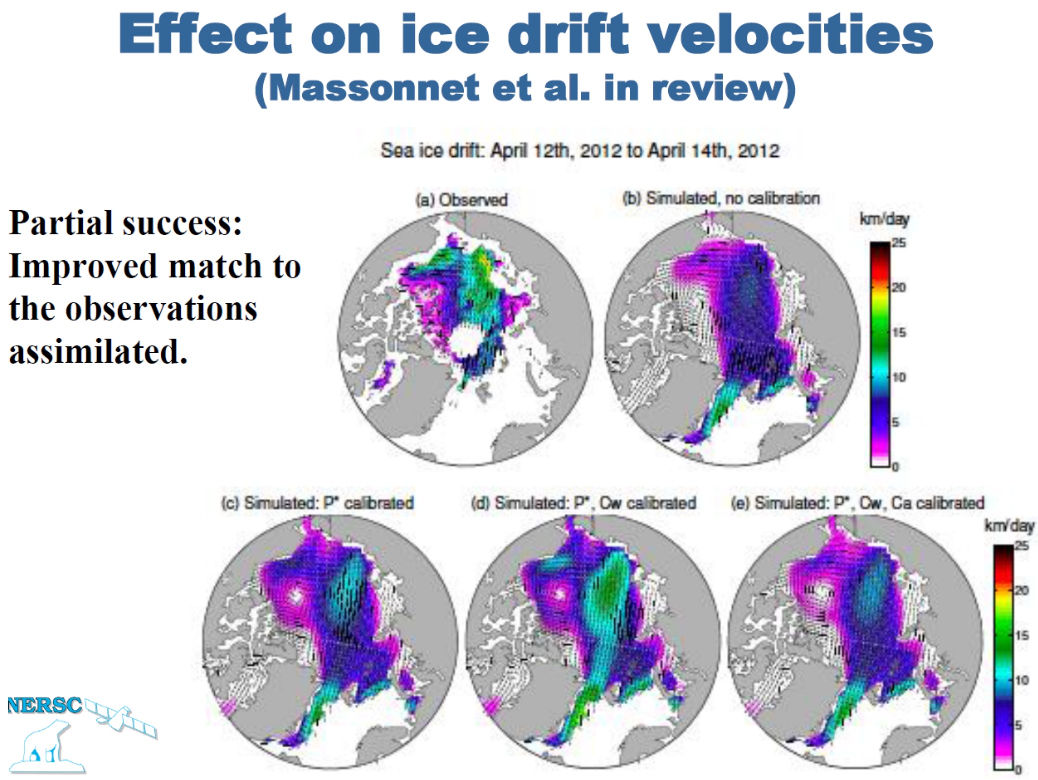

Figure 2.7: Assimilation of ice drift

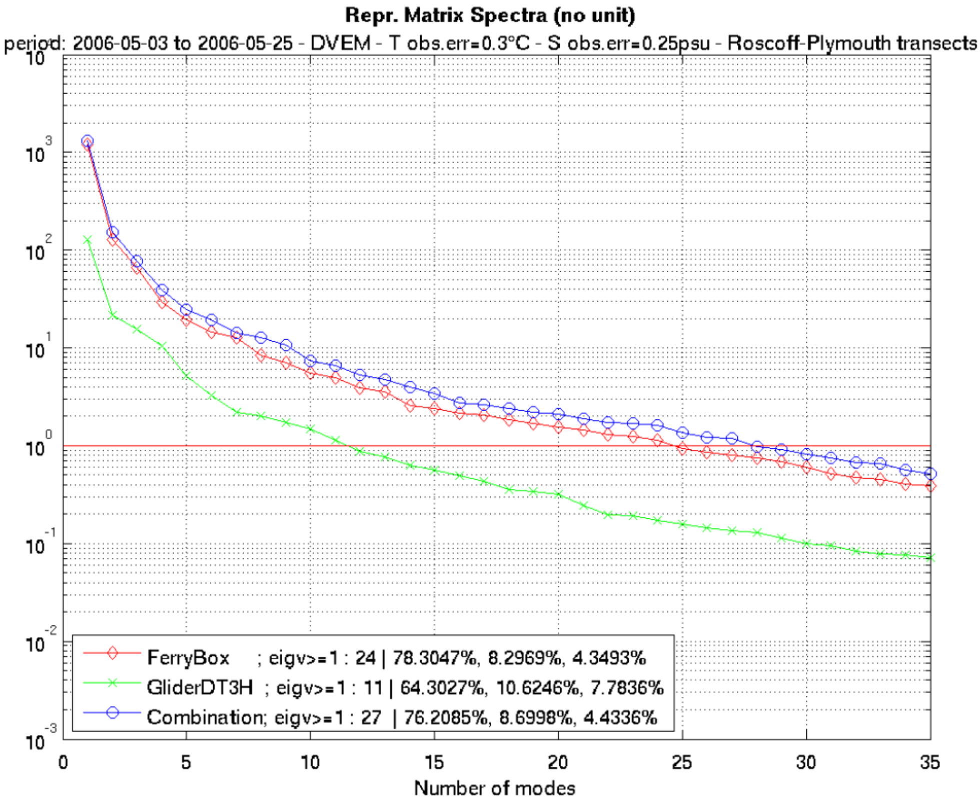

Figure 2.8: Eigenvalues of the representers from Ferrybox (red) remain larger than baseline from obervations (flat line) for a larger number of eigenvalues than the one deducted for gliders (green line). Hence they potentially have a larger range of detectable modes

D5.7: Result of the data assimilation experiment aiming to estimate Lagrangian sea ice parameters (M48, NERSC) is still on track.

Task 5.6 Prior errors detection by observational arrays A diagnostic code has been tested by CNRS-LEGOS on a coastal oceanographic application and shared with CNRS-LEGI. D5.8: RMSpectrum library and results of array performance analyses (M48, CNRS-LEGOS) is well on track.

Selecting among different locations and combinations of observing systems is simplified by characterizing the incremental information via the spectrum of the observed part of the error covariance of the forecast. This leads to selection criteria which allow to distinguish the added value of ferryboxes vs gliders for example For details see http://www.data-assimilation.net/Events/Year3/demey-sangoma2014.pdf (Presentation at progress meeting)

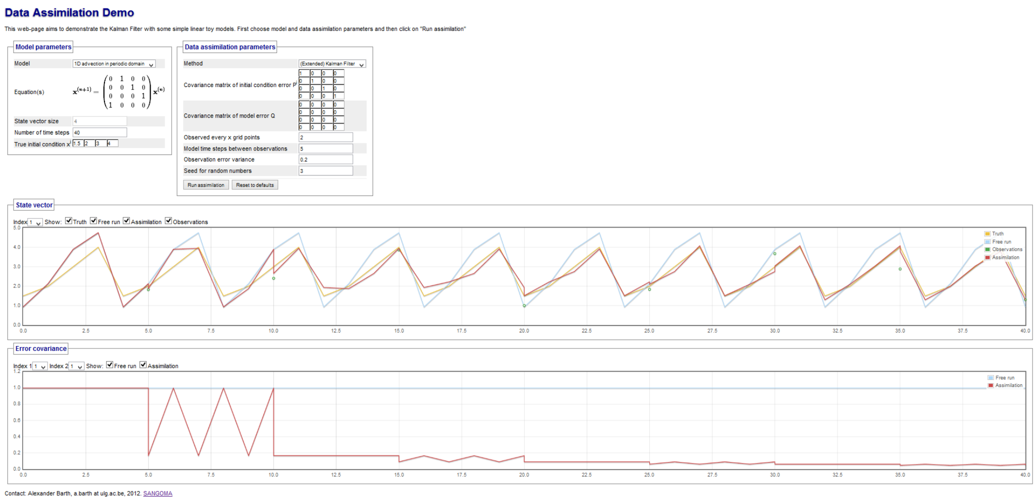

Figure 2.9: Web interface to data assimilation demonstration

2.5.3 Deviations from DOW

WP fully on track

2.5.4 Outreach

Presentations:

Barth et al. (2011)

2.6 WP6: Knowledge transfer

Web pages of SANGOMA have been updated regularly (see DL6.2) and include now scientific highlights.

The Liège colloquium 2015 will focus on DA and SANGOMA will take a leading role there (GHER will prepare draft of terms and reference and send around for advice and scientific committee nominations). ESA (after verification on 5/4) is still interested in contribution from SANGOMA for their summer school. For this, a lecture with exercises is to be prepared (DL6.9). It will exploit the web interface to some assimilation techniques to illustrate the concepts: `http://www.data-assimilation.net/Tools/AssimDemo/`

Publications of SANGOMA partners are already appearing, we now should focus on common papers and the special issue of the Liège colloquium. For DA workshop with Ph.D students, the first one was done during a cruise (DL6.4) , but a more open second school is to be scheduled by MEOM. Another important dissemination activity will be the participation in the MyOcean Science days in September 2014. Finally no newsletter was yet send out (because of only moderate interest in receiving unsolicited mails and the not yet harmonized software interfaces). Now with V1 coming out a newsletter will be send out.

The bundled V1 software was also released with the documentation (DL6.13)

2.6.1 Deviations from DOW

DL6.4 was delayed because of several other data assimilation schools already taking place.

2.7 WP7: Management

Meetings have been organised as planned and reports on the meeting provided (DL7.3)

The need for timely financial report was repeated. The different communication tools set up were reminded but the use of `sourceforge` forums for discussions is still weak, most of partners relying on email for discussions (the sangoma-science mailing list is public). For preparing Deliverables, it was stressed that whenever possible reference to reviewers suggestions should be included and in all cases deviations from the DoW clearly mentioned and explained. Also target audience should be included in the deliverable description to avoid overloading readers with unnecessary reports for them.

2.7.1 Deviations from DOW

On track

Chapter 3

Deliverables and milestones

The state of the deliverables is as follows (snapshot from the SESAM site).

Chapter 4

Dissemination activies

Chapter 5

Contributions from Partners

5.1 ULG

For WP1, ULg contributed to the specifications and list of tools by participating actively in the discussion and contribution to deliverables of WP1 and WP2. With AWI, ULg organised the second software release of WP2 and contributed several tools, including their documentation to the package. In WP3 ULg took actively part in the development of new DA techniques (scale separation, conservation in localisation techniques, comparison of IAU methods). In WP4, ULg implement all benchmarks including its assimilation part. Also its medium-size benchmark implementation provided the default setup and member generations. For WP5, the regional sea model was implemented and assimilation of radar data tested with novel localisation techniques included. WP6 and preparation of associated deliverables was basically executed by ULg, as well as tasks and deliverables of WP7.

5.2 AWI

During the second period, AWI continued to contribute to the work packages 1 to 4. New tools to be implemented were discussed in WP1, followed by their implementation in WP2. For WP2, AWI coordinates and takes part in the the implementations. AWI coordinated the second software release (Milestone MS9) together with A. Barth from ULg and lead the preparation of DL2.4, which documents the tools included in the second software release. For WP6, AWI supported the tools code release and presented software developments and progress on assimilation methods at international conferences. An extensive amount of time was spend for the research work in WPs 3 and 4. Here, AWI coupled the medium case benchmark to the assimilation framework PDAF. The PhD student at AWI (P. Kirchgessner) participated in the PhD student exchange. He worked from May to July 2013 in the group of the project partner at University of Reading. Here, he intensified his research work on equivalent weight particle filters (EWPFs) for PW3. For WP4, he implemented the filter in the small case benchmark and prepared the perturbation scheme for the NEMO model that is used with the EWPFs for the medium case benchmark. A publication on the influence of nonlinearity on smoother algorithms in the Quarterly Journal of the Royal Meteorological Society is currently in press as is a publication in the Monthly Weather Review on aspects of localization in ensemble Kalman filters.

5.3 NERSC

NERSC has contributed to WP2 the toolbox on random perturbations of forcing fields, initiated the work on estimation of sea ice parameters in WP5. As part of WP6, NERSC has applied the Gaussian anamorphosis tools to the ESA GlobColour project and participated to the organization of the forthcoming EnKF workshop in Os, Norway `http://enkf.iris.no/`.

5.4 CNRS

CNRS/LEGI is responsible for WP4 (SANGOMA benchmarks).

CNRS/LEGOS mainly contributed to the description of toolboxes of WP1 and provided a list of their tools that will be included in the SANGOMA repository.

5.5 UREAD

UREAD took the lead in preparing the review paper on state of the art data assimilation techniques which is circulating now for feeback and additions. UREAD also contributed matlab tools to the software release.

5.6 TUD

Delft University of Technology is a partner institute of SANGOMA program. The activities at TuDelft are headed by Prof. Arnold Heemink and involve Martin Verlaan, Nils van Velzen and M. Umer Altaf.

Chapter 6

Financial resources

As already mentioned in some of the work description, most partners had difficulties in hiring qualified Ph.D students or postdocs in data assimilation and modelling. Even with extensive use of web announcement, internal distribution in institutes and use of several email lists (among which the widely exploited http://www.lists.rdg.ac.uk/mailman/listinfo/met-jobs), it took time to find adequate personnel and then go through the administrative processes.

As the largest amount of the budget (90% of direct costs) is dedicated to personnel costs, almost uniformly distributed over the project lifetime, underspending during the first year occurred but have been resorbed in the second phase

6.1 ULG

6.2 AWI

As the financial reporting is not yet completed, only preliminary numbers on the financial resources can be reported: AWI charged about 9.5 person months to the project. This corresponds to about 45% of the total salary cost. The travel costs were rather high because P. Kirchgessner worked in the group of the partner at the University of Reading for three months from May to July 2013 as part of SANGOMA's PhD student exchange program. The funds for this stay will be shifted from the budget of ULg to AWI. Other travel costs were as planned.

6.3 NERSC

6.4 CNRS

6.5 UREAD

6.6 TUD

Chapter 7

Additional Plans

Some random specific points to act were identified

- The diagnostic tools should include a short documentation for beginners
- Presentation to the MyOcean science meeting (and others) should be done
- Increase exchange mechanism should be used in WP3
- diagnostic tools could also be very convenient for teaching
- As a coastal benchmark the ROMS implementation in the Ligurian sea could be made available
- When comparing assimilation methods we should show what ensemble approaches can add to standard methods
- Contributing to a Wikipedia article could make SANGOMA more widely known
- Newsletter, summerschools and joint workshops must be organized
- Participation to Nemo user meeting July 7

Bibliography

Barth, A., J. Chiggiato, A. Alvera Azcarate, B. Mourre, J.-M. Beckers, J. Horstmann, and M. Rixen (2011). Assimilation of high-frequency radar currents in the ligurian sea.

Beckers, J.-M., A. Barth, I. Tomazic, and A. Alvera-Azcárate (2014). Multi-scale optimal interpolation: application to DINEOF analysis spiced with a local optimal interpolation. *Ocean Science Discussions 11*(2), 895–941.

Beckers, J.-M., A. Barth, C. Troupin, and A. Alvera-Azcárate (2014). Approximate and efficient methods to assess error fields in spatial gridding with DIVA (data interpolating variational analysis). *Journal of Atmospheric and Oceanic Technology 31*, 515 – 530.

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

Brankart, J.-M., C.-E. Testut, D. Béal, M. Doron, C. Fontana, M. Meinvielle, P. Brasseur, and J. Verron (2012). Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations. *Ocean Science 8*, 121–142.

Kirchgessner, P. and L. Nerger (2014). A comparison of linear and non-linear data assimilation techniques using a toy model. In *International Symposium on Data Assimilation, Munich, Germany, February 24-27*.

Kirchgessner, P., L. Nerger, and A. Bunse-Gerstner (2013). Adaptive localization in ensemble kalman filter methods by controlling the observation space. In *6th WMO Symposium on Data Assimilation, College Park, MD, USA, October 7 - 11*.

Nerger, L. and W. Hiller (2013). Software for ensemble-based data assimilation systems: Implementation strategies and scalability. *Computers and Geosciences 55*, 110 – 118. Ensemble Kalman filter for data assimilation.

Nerger, L., W. Hiller, and J. Schröter (2013). Scalable sequential data assimilation with the parallel data assimilation framework pdaf. In WMO

(Ed.), *6th WMO Symposium on Data Assimilation, College Park, MD, USA, October 7 - 11, 2013*.

Yan, Y., A. Barth, and J.-M. Beckers (2014). Comparison of different assimilation schemes in a sequential Kalman filter assimilation system. *Ocean Modelling 73*, 123 – 137.

Yan, Y., A. Barth, Laenen, and J.-M. Beckers (2013). Assimilation of simulated satellite altimetric data and argo temperature data into a double-gyre nemo ocean model.