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

First annual report

J.-M. Beckers A. Barth Y. Yan F. Laenen 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 practise 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 Year 1 achievements

The focus of year one was to setting up the scene for the project: For harmonisation we prepared the initial list of tools we want to share with the data assimilation (DA) community. This list includes already existing tools to be adapted to a common specifications and some new tools to be added. The list includes tools on diagnostics, perturbation generation, transformations and utilities. For the collaborative development, a platform on `sourceforge` was set up and all participants registered accordingly. One of the crucial points for the code-sharing aspects was the design of a data model and code-interfacing specifications, as the SANGOMA tools will need to comply with these specifications and existing toolbox must be able to exploit these tools with a minimum of adaptation effort on the toolbox side. The adopted approach is based on simple Fortran interfaces using arrays, with the possibility to use C-binding and call-back functions for manipulating more

complex and model-specific structures under the responsibility of the calling routine. For out-of-memory data manipulation the commonly used netCDF format is adopted in its CF compliant form. For testing DA tools, different benchmarks have been set up (small case, medium case and large case), with all necessary parameters for execution defined and access to models and model setups provided. First assimilation tests with the small and medium benchmarks have also been performed. Data necessary for the project execution are also available to each partner and the specific SANGOMA-MyOcean service level agreement allows for further access to reanalysis products. The project can now begin its second phase with the development of new data-assimilation techniques. This wrok will rely on the review of state-of-the art data assimilation also prepared during the first year. Highlights of year one are therefore

- Definition of a SANGOMA license that allows wide use of products and tools of SANGOMA.
- Definition of a data model and interfaces (in memory or files).
- Preparation of a first list of tools to be shared and preparation of these included in the first code release.
- First version of a document on state-of-the art data assimilation techniques and upcoming techniques.
- Setting up of the benchmarks and first full execution of benchmarks with EnKF.
- Preparation of a catalog of upcoming remote-sensing techniques and analysis of their specificities for data assimilation.
- Contacts with myOcean, GMES projects, private firms and operational centers to ensure data model and interfaces are pertinent; in particular strong links with myOcean have been put in place.
- Setting up the communication, sharing, dissemination and development platforms of the project.

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 first year

2.1 WP1: Harmonization

This workpackage aims at standardising assimilation tools and file formats for easier exchange and use both by project partners and other data assimilation experts. The first year has been focusing on

- analysing the different toolboxes currently in use (content, data model, calling interfaces, etc.),
- defining an initial list of exiting tools from those toolboxes to be shared,
- defining a data model and calling interface for SANGOMA tools.

A crucial point was the definition of data models and calling interfaces. Here, the adopted solution was to keep them simple in order to allow for rapid use of the tools by non specialists in advanced programming methods. Furthermore, the broad range of interfaces and programming languages used in the existing toolboxes made necessary a common denominator, leading to the following:

- for data exchange via files: use of netCDF files in CF compliant form. As the format allows for some variants, the general strategy is to "keep it simple" and to provide output files in a similar form than input files (even if not perfectly fitting CF conditions as is the case for some model outputs). Also version 3 features of netCDF are to be preferred over version 4 to enhance backward compatibility. Finally, ensembles will be treated by working on a collection of files instead of a single big file.
- for data exchange in memory (subroutine call): Use of basic FORTRAN structure `arrays`. No derived types are allowed (too much programming overhead in filling or adapting data types). For more complex interfacing or data structures the `call-back` approach has to be used. Example: if the observational error variance matrix $\mathbf{R}$ is non-diagonal and a tool needs to evaluate $\mathbf{R}\,\mathbf{y}$, the interface of the tool must include a call-back function which when called with argument $\mathbf{y}$ returns the product $\mathbf{R}\,\mathbf{y}$. In this case, in the call-back program more complex (system specific) structures can be

used without the need to define complicated interfacing in the SANGOMA tools. C-binding specifications are also provided.

```
module sangoma_callback

    use, intrinsic :: ISO_C_BINDING
    use sangoma_base, only:REALPREC, INTPREC
    implicit none

contains

   subroutine some_operation(x, n, f_callback) &
              bind(C,name="callback_some_operation")

      use, intrinsic :: ISO_C_BINDING
      implicit none

      integer(INTPREC), value, intent(in) :: n
      real(REALPREC),          intent(in) :: x(n)

      interface
         subroutine f_callback(x,n) bind(C)
            use, intrinsic :: ISO_C_BINDING
            use sangoma_base, only:REALPREC, INTPREC

            integer(INTPREC), value, intent(in) :: n
            real(REALPREC),          intent(in) :: x(n)
         end subroutine
      end interface
```

Advances, deviations or delays of WP1: This workpage has evolved as expected, without deviations and no delays in deliverables or milestones. The specifications found a large positive feedback in the DA community as well as the advisory board. Major advances were the agreement on these specification between partners with rather different approaches and subsequent positive feedback from other groups.

2.2 WP2: Sharing and collaborative development

This workpackage aims at sharing tools in a collaborative way using standards set up in WP1.

Technically, the sharing is done via a server with version control. The SANGOMA project was registered on http://sourceforge.net/projects/sangoma/ as it allows not only sharing of files but also easy setup of discussion forums, download procedures etc. All SANGOMA scientists have registered and are allowed to upload changes (commit of software changes). The site not only contains assimilation-related software but also `html` files for the web server and document templates so that each partner can easily contribute.

The first collection of tools (V0) found on the server http://sourceforge.net/projects/sangoma/files/latest/download are those specified in WP1 and as detailed in the DOW, not yet adapted to the newly defined interfacing rules. The following tools are included:

Diagnostic Tools

sangoma_ComputeHistogram	Compute ensemble rank histograms
sangoma_ComputeEnsStats	Compute ensemble statistics
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

sangoma_MVNormalize	Perform multivariate normalization
sangoma_EOFCovar	Initialize covariance matrix from EOF decomposition
Weakly constrained ensemble perturbations	Create ensemble perturbations that have to satisfy an a priori linear constraint

Transformation Tools

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

Utilities

hfradar_extractf	Observation operator for HF radar surface currents
PodCalibrate	Calibration tool for estimating uncertain model parameters
EnKF	Ensemble Kalman filter as introduced by Evensen and Burgers

All tools include a full documentation on how to use and compile them in their present form, but the adaptation of interfaces to the SANGOMA standard will now start and additional tools be included.

```
make

gfortran -O3 -fdefault-real-8  -c
diagnostics/sangoma_ComputeEnsStats.F90 -o
diagnostics/sangoma_ComputeEnsStats.o

gfortran -O3 -fdefault-real-8  -c
diagnostics/sangoma_ComputeHistogram.F90 -o
diagnostics/sangoma_ComputeHistogram.o

gfortran -O3 -fdefault-real-8  -c
perturbations/sangoma_EOFCovar.F90 -o
perturbations/sangoma_EOFCovar.o

gfortran -O3 -fdefault-real-8  -c
perturbations/sangoma_MVNormalize.F90 -o
perturbations/sangoma_MVNormalize.o

ar -r  libsangoma_tools.a  diagnostics/sangoma_ComputeEnsStats.o
diagnostics/sangoma_ComputeHistogram.o
perturbations/sangoma_EOFCovar.o
perturbations/sangoma_MVNormalize.o

ar: creating libsangoma_tools.a

ranlib  libsangoma_tools.a
```

Advances, deviations or delays of WP2: Good progress was made in defin-

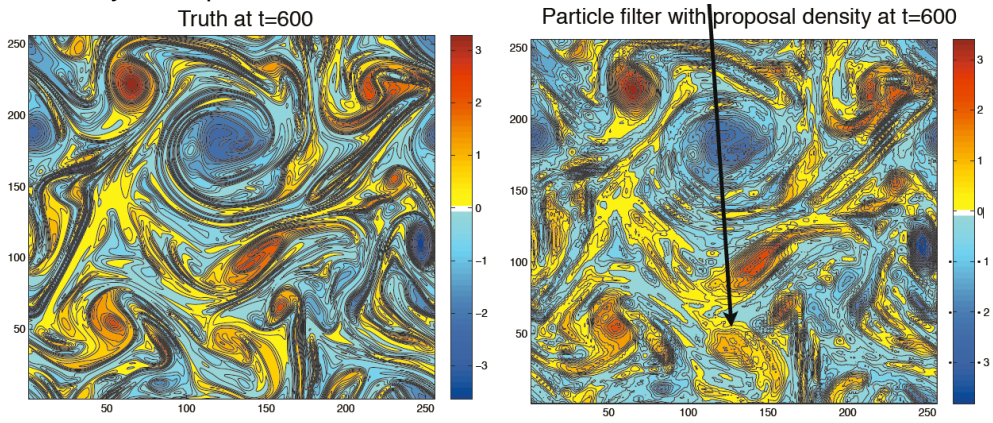
ing the first list of tools to be shared and preparing the relevant documentation, the package being available on the SANGOMA web site and `sourceforge`. No delay in deliverables or milestones were encountered, but some tools from partner CNRS of the initial list need to be added to the `sourceforge` repository.

2.3 WP3: Innovative Data Assimilation techniques

This workpackage is the most exploratory and uncertain as it will test advanced data-assimilation techniques (such as particle filters) in the context of ocean models. The first year was basically used to provide some review reports both for SANGOMA partners and external scientists. These reports contain

- DL3.1: Initial report on non-linear data assimilation
- DL3.2: Living document starting from the initial report enriched over the project lifetime with description of new techniques. Presently, the use of Proper Orthogonal Decomposition to calculate model gradients without needing to formulate the adjoint of the model is described on top of DL3.1.

The groups made interesting progress in using particle filters in quasi-geostrophic models by adding stochastic terms designed to have a proposal density leading to a more uniformely distributed particle weight. New techniques as multivariate rank histogram filters have also been tested. Previous work on anamorphosis effects in ecosystem data assimilation and finalised within the SANGOMA project has already been published.

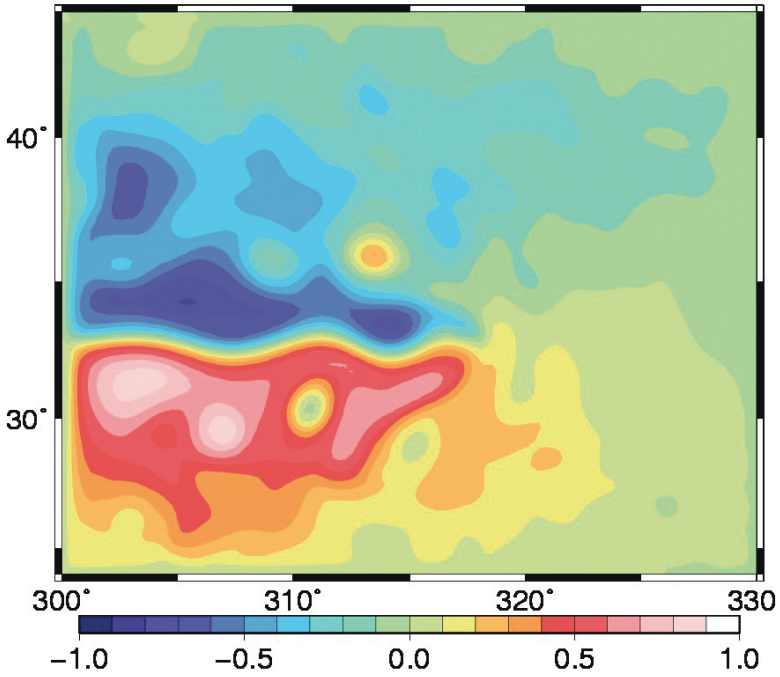


Exploration of particle filters in a quasi-geostrophic model.

Advances, deviations or delays of WP3: The workpackage leader had some difficulties to find an adequate postdoc and was himself overloaded during 2011 so that the review paper (DL3.1) on advanced data assimilation techniques planned for month 6 was only delivered in draft version at month 12. It is however an excellent base for the living document which will be one the highlights of SANGOMA in the science aspects. The delay is not critical and in addition during the annual meeting very promising new ideas have come up from several partners.

2.4 WP4: Benchmarks

This workpackage aims to provide some well documented and representative test cases for comparing data assimilation techniques. It also will focus on adequate metrics to assess quality in non-linear regimes. The first year was concentrating in defining the details of the setup of the benchmark models. As planned, a small size (Lorenz 96), medium size (double gyre) and large size problem (Atlantic Ocean) was formulated. The setup of the benchmark models is completely specified and documented and some basic metrics are defined. A list of metrics not relying on the Gaussian-distribution assumption have also been prepared and discussed within the group and will serve as a basis for future assessment. It is also worth noting that the benchmark setup and use within an assimilation toolbox seems quite well prepared, as internal tests showed that a newcomer (both for NEMO and the toolbox) could reach a fully working ensemble assimilation within three month. Also an experienced external private company was able to install models for the medium benchmarks within a few days and is now trying to implement the assimilation.



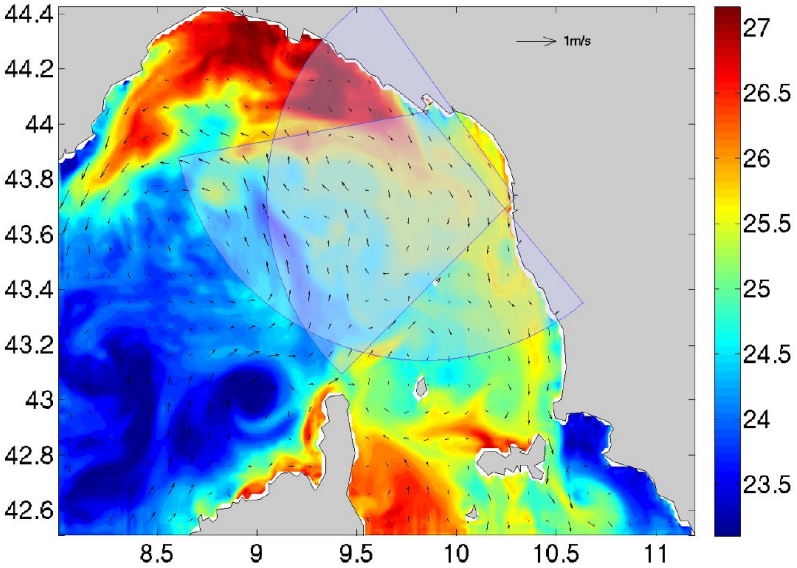
Double gyre setup for benchmarking.

Advances, deviations or delays of WP4: Excellent progress was made in this WP and the benchmark approach has been identified by advisors and myOcean Science days as one of the best ways to enhance collaborations between data assimilation groups, including those relying on variational approaches. No deviations from the DOW or delays have been observed and the WP advances as it should.

2.5 WP5: Data Assessment

This WP is at an early stage as assimilation tests are planned later in the project and most of the exercises rely on simple setups or existing data sets. Nevertheless, an agreement with myOcean was set up to easily assess data whenever needed.

Furthermore, a living document on new data types of possible interest for data assimilation is to be published within the next month.



Assimilation experiment in the Ligurian Sea using HF radar velocity estimates

Advances, deviations or delays of WP5: Some delay on the report on new possibilities in DA from remote sensing was encountered (DL5.1) because of delayed hiring of personel. It is however uncritical as most partners already had access to the data needed. The report itself includes information on specificities of the new data types related to data assimilation and is therefore of interest to the data assimilation community.

2.6 WP6: Knowledge transfer

This WP has to ensure that knowledge gained from the project and tools developed in it will be exploited outside of the SANGOMA consortium. Among the first-year activities we find the release of V0 of the code, publication of papers and presentation of the project partners at different meetings (for a detailed list of activities see Chapter 5).

An important point is that the tools developed by SANGOMA are useful to the scientific and operational community:

- For scientific groups, the open sharing of tools and benchmarks will allow other scientist to test their new ideas and their effect by comparing them to results from SANGOMA. Also reports on DA techniques and new data

types will speed up uptake of new techniques by scientist using data assimilation. We will reach this community by standard scientific publications and communications but also via the planned Ph.D workshops, schools and student exchanges.

- myOcean: a strong involvement of myOcean representatives in SANGOMA was planned and implemented (two partners are also partners in myOcean, one as MFC; the steering committee of SANGOMA includes E. Dombrowsky as representative of myOcean; already two of the variable advisory-board members emerged from myOcean: D. Obaton and C.E. Testut; a specific consortium agreement between SANGOMA and myOcean was set up; myOcean science days workshop on data assimilation was organised with SANGOMA.)

- Other GMES service evolution projects showed interest (OSS2015 and myWave for data assimilation techniques) but no particular action has yet emerged. SANGOMA is particularly interested in learning from their experience in using data assimilation in strongly nonlinear regimes. Also with OSS2015 the Ligurian sea experiment is of interest for both, but no practical action has been taken yet. We will make sure information on our project outcomes will be communicated to these projects and possibly invite them to our next general assembly.

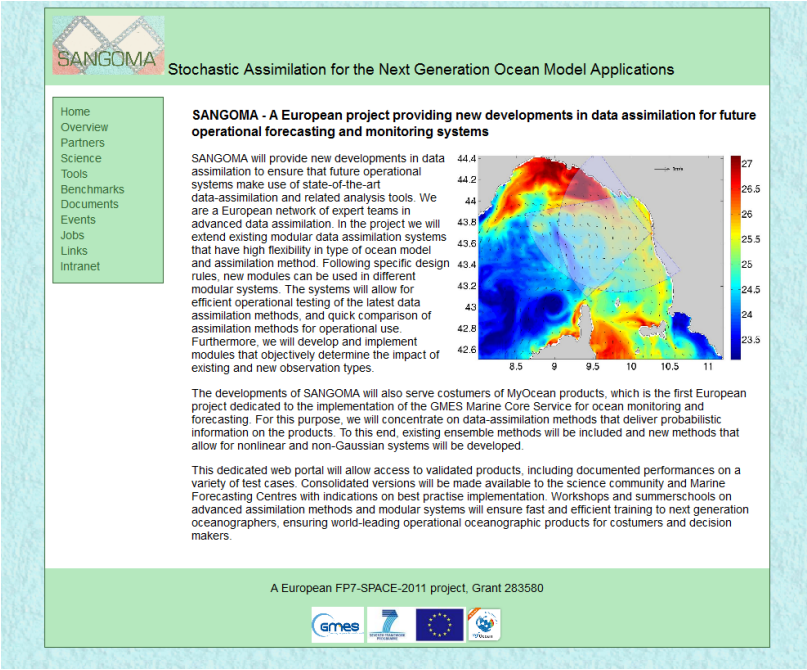
- Operational centers (within myOcean and outside) should find in the SANGOMA toolboxes and benchmarks easy ways to test new techniques before deciding to optimise them for their operational implementation. Diagnostic tools should also be of direct interest. To make sure our work is relevant we plan to submit the next list of tools to be developed to myOcean. We will further propose to interested MFCs to send one of the SANGOMA scientists to these centers to help them trying out the SANGOMA tools and benchmarks.

- Private firms have shown their interest in using tested tools as they generally have less time for reimplementing documented tools. For these firms (presently three), we will keep them informed and certainly invite them to the last meeting with operational users.

- Contacts with ESA (ESFRI) through the advisory board participation at the kick-off meeting is kept alive, notably to prepare the ESA summerschool contribution from SANGOMA. NERSC is also involved in ESA's Ocean Colour CCI in which he uses data assimilation techniques further developed within SANGOMA.

Furthermore, a workshop with operational users was planned at month 12. During GMES marine projects coordination meeting on 24/05/2012 in Brussels we discussed this matter with the myOcean group and it was agreed that we should reach this community together during the myOcean Science Days. Therefore, invitations were send via the myOcean contact list but also directly from the SANGOMA side to a dozen of additional institutes interested in operational modelling (like Actimar and MUMM).

During the myOcean Science Days, 60-70 participants attended the workshop on data assimilation. Four oral presentations and four posters showed the SANGOMA approach. A call for feedback was also launched and a quick survey is ongoing http://www.surveymonkey.com/s/ZX3P9D8, with first results indicating very positive views on our proposed data model and interfaces.

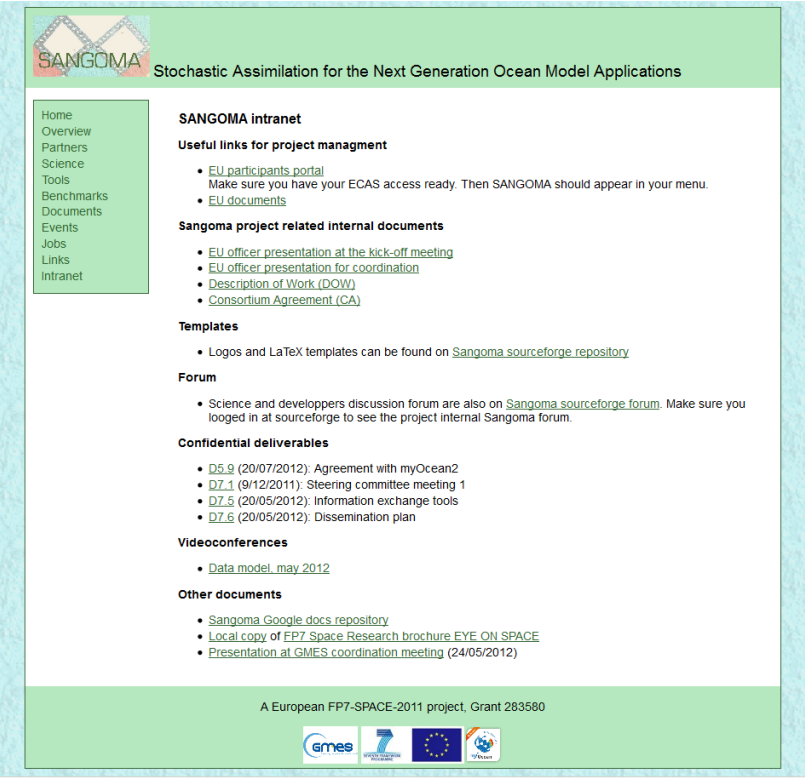


SANGOMA web site http://www.data-assimilation.net.

Advances, deviations or delays of WP6: The workpackage advanced very nicely and SANGOMA has created a strong link with myOcean. A deviation from the original DOW is the joint organisation of the workshop with myOcean instead of a stand-alone one for SANGOMA to avoid unnecessary traveling and meeting organisation. The synergy certainly allowed SANGOMA to reach the majority of groups interested in data assimilation from a scientific, operational and user perspective. Because of the later organisation of this workshop with respect to the DOW, DL6.7 was delayed, which is however not critical.

2.7 WP7: Management

The management task has been simple during the first year, mostly dealing with organising the kick-off and first year meetings, setting up the agreement with myOcean, preparing the dissemination plan and setting up the information exchange platform (email lists, web server and virtual meeting configuration). It also included the following of project evolution and checking that deliverables were produced on time. As explained in Section 4, delayed DL5.1 and DL3.1 are available as draft and will only be submitted in January 2013 whereas delayed DL6.7 and DL7.2 will be submitted by the end of December 2012.



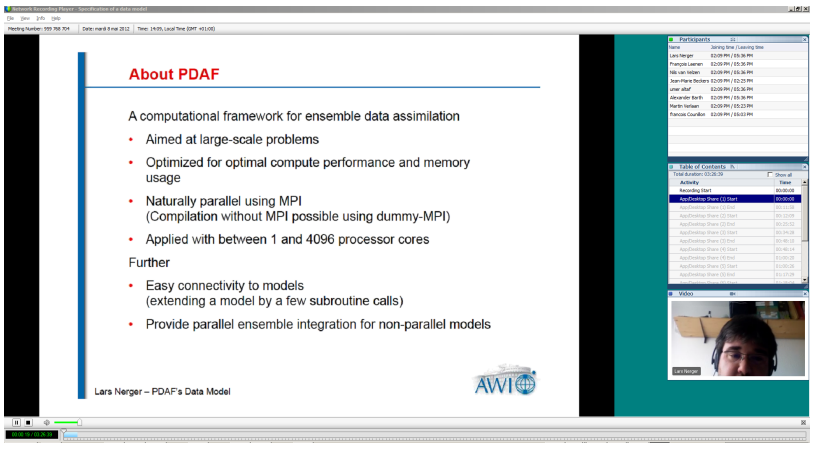
SANGOMA intranet.

Advances, deviations or delays of WP7: The management has fulfilled all necessary steps of year 1, even if the setting up of the consortium agreement with myOcean took more time than planned due to administrative reasons. This delay has however no impact as it concerns data not planned to be used before the second year of the project. All communication tools were set up without difficulties and each partner knows how to use them. The annual meeting could only be held at month 12.5 due to the difficulty of finding a date suitable for all partners and advisors. Is also explained the delay in DL7.2 which is simply due to the organisation of the annual meeting after the first reporting period. Both delays are minor and not critical for the project evolution.

Chapter 3

Organisational aspects of interest

- As mentioned in the proposal, finding experts in modelling and data assimilation is not easy and some partners had delays in hiring personnel. We published the open positions on our own web site, via met-jobs (`http://www.lists.rdg.ac.uk/mailman/listinfo/met-jobs`) and other channels. Despite this delay, work in SANGOMA has progressed.
- The advisory board (two permanent members + two topic related experts) provides interesting feedback and the flexibility in nominating two different advisors for each meeting allows to profit from a wide range of expertise.
- The coordination with myOcean clearly facilitates interaction with operational users.
- The virtual meeting via WEBEX for defining data interface was quite efficient (a film of meeting is available on our intranet)



Webex meeting snapshot.

Chapter 4

Deliverables and milestones

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

Submitted deliverables

WP n°	Deliverable N°	Title	Version	Lead beneficiary	Person months	Nature	Dissemination level	Document type	Delivery date from Annex I (proj month)	Actual date	Status	Status Date	Action	Ares/Hermes Id
1	1	List of existing tools	1.0	TECHNISCHE UNIVERSITEIT DELFT	2.0	Report	PU		31/01/2012 (3 months)	03/02/2012	Received	03/02/2012		
1	2	List of tools to be adapted	1.0	TECHNISCHE UNIVERSITEIT DELFT	2.0	Report	PU		30/04/2012 (6 months)	22/05/2012	Received	22/05/2012		Ares(2012)608065
1	3	Specifications of data model	1.0	ALFRED-WEGENER-INSTITUT FUER POLAR- UND MEERESFORSCHUNG	4.0	Report	PU		30/04/2012 (6 months)	10/07/2012	Received	10/07/2012		Ares(2012)840477
1	4	Augmented list of common and new tools	1.0	TECHNISCHE UNIVERSITEIT DELFT	5.0	Report	PU		31/10/2012 (12 months)	20/11/2012	Received	20/11/2012		Ares(2012)1365231
1	5	Documentation of specifications	1.0	TECHNISCHE UNIVERSITEIT DELFT	2.0	Report	PU		31/10/2012 (12 months)	20/11/2012	Received	20/11/2012		Ares(2012)1365234
2	1	SVN server description for collaborative development	1.0	UNIVERSITE DE LIEGE	0.5	Report	PU		30/11/2011 (1 months)	15/12/2011	Received	15/12/2011		
2	3	Software V0 report	1.0	ALFRED-WEGENER-INSTITUT FUER POLAR- UND MEERESFORSCHUNG	0.5	Report	PU		31/10/2012 (12 months)	20/11/2012	Received	20/11/2012		Ares(2012)1365248
3	2	Living document	1.0	TECHNISCHE UNIVERSITEIT DELFT	1.0	Report	PU		30/04/2012 (6 months)	22/05/2012	Received	22/05/2012		Ares(2012)608072
4	1	Benchmark definitions	1.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	6.0	Report	PU		31/10/2012 (12 months)	20/11/2012	Received	20/11/2012		Ares(2012)1365228
5	9	Agreement with MyOcean	1.0	UNIVERSITE DE LIEGE	0.5	Report	RE		30/11/2011 (1 months)	22/08/2012	Received	22/08/2012		Ares(2012)991963
6	1	Web pages V1 report	1.0	UNIVERSITE DE LIEGE	0.5	Report	PU		31/10/2012 (12 months)	20/11/2012	Received	20/11/2012		Ares(2012)1365250
6	12	Code release 1 documentation	1.0	UNIVERSITE DE LIEGE	1.0	Report	PU		31/10/2012 (12 months)	20/11/2012	Received	20/11/2012		Ares(2012)1365244
7	1	Steering committe 1	1.0	UNIVERSITE DE LIEGE	0.1	Report	CO		30/11/2011 (1 months)	15/12/2011	Received	15/12/2011		
7	5	Report on project information exchange tool	1.0	UNIVERSITE DE LIEGE	2.6	Other	RE		30/04/2012 (6 months)	22/05/2012	Received	22/05/2012		Ares(2012)608086
7	6	Dissemination plan	1.0	UNIVERSITE DE LIEGE	2.0	Report	CO		31/01/2012 (3 months)	22/05/2012	Received	22/05/2012		Ares(2012)608061

Deliverables in progress

WP n°	Deliverable N°	Title	Version	Lead beneficiary	Person months	Nature	Dissemination level	Document type	Delivery date from Annex I (proj month)	Forecast date	Status	Status Date	Action	Ares/Hermes Id
2	2	Software codes on SVN (continually)	0.0	ALFRED-WEGENER-INSTITUT FUER POLAR- UND MEERESFORSCHUNG	17.5	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
2	4	Software V1 report	0.0	ALFRED-WEGENER-INSTITUT FUER POLAR- UND MEERESFORSCHUNG	0.5	Report	PU		30/04/2014 (30 months)	30/04/2014	Pending		Upload	
2	5	Software V2 report	0.0	ALFRED-WEGENER-INSTITUT FUER POLAR- UND MEERESFORSCHUNG	0.5	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
3	1	State of the art non-linear data-assimilation	0.0	THE UNIVERSITY OF READING	1.0	Report	PU		30/04/2012 (6 months)	30/04/2012	Pending		Upload	
3	3	Report on implementation	0.0	ALFRED-WEGENER-INSTITUT FUER POLAR- UND MEERESFORSCHUNG	25.0	Report	PU		30/04/2014 (30 months)	30/04/2014	Pending		Upload	
3	4	Report on new methods	0.0	THE UNIVERSITY OF READING	25.0	Report	PU		30/04/2015 (42 months)	30/04/2015	Pending		Upload	
3	5	Report on new non-linear methods	0.0	THE UNIVERSITY OF READING	1.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
4	2	Report on benchmark implementations	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	15.0	Report	PU		31/10/2013 (24 months)	31/10/2013	Pending		Upload	
4	3	Report on metric analysis	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	14.0	Report	PU		30/04/2014 (30 months)	30/04/2014	Pending		Upload	
4	4	Report on metrics with new tools	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	14.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
4	5	Report on metrics in large cases	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	14.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
5	1	List of remote sensed variables	0.0	UNIVERSITE DE LIEGE	2.5	Report	PU		31/10/2012 (12 months)	31/10/2012	Pending		Upload	
5	2	Ecosystem data report	0.0	THE UNIVERSITY OF READING	3.0	Report	PU		31/10/2014 (36 months)	31/10/2014	Pending		Upload	
5	3	Large scale model V1 analysis	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	9.0	Report	PU		31/10/2014 (36 months)	31/10/2014	Pending		Upload	
5	4	Large scale model V2 analysis	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	9.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
5	5	Regional model V1 analysis	0.0	UNIVERSITE DE LIEGE	9.0	Report	PU		31/10/2014 (36 months)	31/10/2014	Pending		Upload	
5	6	Regional model V2 analysis	0.0	UNIVERSITE DE LIEGE	9.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
5	7	Lagrangian ice data analysis	0.0	STIFTELSEN NANSEN SENTER FOR FJERNMAALING	7.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
5	8	Array analysis	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	6.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
6	2	Web pages V2 report	0.0	UNIVERSITE DE LIEGE	0.5	Report	PU		30/04/2014 (30 months)	30/04/2014	Pending		Upload	
6	3	Web pages V3 report	0.0	UNIVERSITE DE LIEGE	0.5	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
6	4	Ph.D workshop 1 report	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	1.0	Report	PU		31/10/2013 (24 months)	31/10/2013	Pending		Upload	
6	5	Ph.D workshop 2 report	0.0	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	1.0	Report	PU		30/09/2015 (47 months)	30/09/2015	Pending		Upload	
6	6	Report on best DA for operational purposes	0.0	UNIVERSITE DE LIEGE	2.0	Other	PU		31/10/2014 (36 months)	31/10/2014	Pending		Upload	
6	7	Report on workshop 1 for operational users	0.0	UNIVERSITE DE LIEGE	1.5	Report	PU		31/10/2012 (12 months)	31/10/2012	Pending		Upload	
6	8	Report on workshop 2 for operational users	0.0	UNIVERSITE DE LIEGE	1.5	Report	PU		30/09/2015 (47 months)	30/09/2015	Pending		Upload	
6	9	Documentation of exercise sets for summerschools	0.0	UNIVERSITE DE LIEGE	3.0	Report	PU		30/04/2014 (30 months)	30/04/2014	Pending		Upload	
6	10	Report on International Colloquium	0.0	UNIVERSITE DE LIEGE	3.5	Report	PU		28/02/2015 (40 months)	28/02/2015	Pending		Upload	
6	11	Publication	0.0	UNIVERSITE DE LIEGE	2.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
6	13	Code release 2 documentation	0.0	UNIVERSITE DE LIEGE	1.0	Report	PU		30/04/2014 (30 months)	30/04/2014	Pending		Upload	
6	14	Code release 3 documentation	0.0	UNIVERSITE DE LIEGE	2.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
6	15	Final public report with synthesis	0.0	UNIVERSITE DE LIEGE	2.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending		Upload	
7	2	Steering committe 2	0.0	UNIVERSITE DE LIEGE	0.1	Report	CO		31/10/2012 (12 months)	31/10/2012	Pending		Upload	
7	3	Steering committe 3	0.0	UNIVERSITE DE LIEGE	0.1	Report	CO		30/04/2014 (30 months)	30/04/2014	Pending		Upload	
7	4	Steering committe 4	0.0	UNIVERSITE DE LIEGE	0.1	Report	CO		31/10/2015 (48 months)	31/10/2015	Pending		Upload	

Delays in the delivery date with respect to the planned date are not critical and no domino effect is to be feared. The deliverables under delay on 14/12/2012 are

- DL5.1: A draft is ready but we prefer to circulate it internally and discussed it at the annual meeting before publishing it.
- DL3.1: idem
- DL6.7: The meeting was held on November 20th and the report was validated at the annual meeting. The report will be submitted before the end of December 2012.
- DL7.2: The steering committee met at the annual meeting and the report circulated soon after the meeting end. The report will be submitted before the end of December 2012.

The milestones relevant for year 1 are summarised in the following table, with

a self assessment which show that the project is on track.

MS	Description	Due Date	Fullfillment	Critical input	Output to
MS1	Web pages and communication tool	1/05/2012	1/05/2012	Partner contacts coordinates	Parners, users and developpers Needed in WP2, WP3,WP4 and WP6
MS2	Identification of new tools to be shared and specification of the interfaces	1/11/2012	1/11/2012	List of tools from partners Agreement on data model	WP2
MS5	Software release V0	1/11/2012	1/11/2012	Codes from partners	Open source community
MS13	Signature of agreement with myOcean2	1/01/2012	20/07/2012	Service level agreement	Internal document
MS14	Analysis of new data and its access	1/01/2012	1/01/2013	D1.5	Partners

MS	WPs involved	How to assess completion	Self assessment	Explanation of differences/delays
MS1	All	Visit of http://www.data-assimilation.net/ http://sourceforge.net/projects/sangoma/	Fullfilled as expected	
MS2	WP1,WP2,WP3	Monitoring of living document D1.1 and D1.2 Completness of data model (interfacing) specification D1.3	More additional tools will appear later Excellent progress on critical point	
MS5	WP1,WP2,WP3,WP6	http://sourceforge.net/projects/sangoma/files/latest/download	SVN server exploitet correctly collaborative approach well functioning Contributions from partners not yet balanced	
MS13	WP4, WP5	D5.9	Agreement concluded Includes even some reanalysis products normall not distributed	Administrative delays on myOcean side Not critical as data used only later
MS14	WP4, WP5	D1.5 and description of data access	We prefer to make the description more exhaustive (larger use)	Not critical as data used only later

Chapter 5

Dissemination activies

The following dissemination actions have been taken

- Setting up of web pages
- Publication of two scientific papers with acknowledgments to the SANGOMA projects
- Poster presentations (at least four)
- Presentations at other project meetings (at least six)
- Presentation at GMES marine projects coordination meeting on 24/05/2012 in Brussels
- Contribution to the International Data Assimilation Workshop in Chongqing, P.R. China from 25th June to 7th July 2012, organized by the Institute of Atmospheric Physics, Chinese Academy of Science.

Chapter 6

Contributions from Partners

6.1 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 first software release of WP2 and contributed several tools, including their documentation to the package. Feedback was provided on draft DL3.1. In WP4, ULg was the first partner to actually implement completely the medium benchmark including its assimilation part and proved the feasibility of proposing such benchmarks also to non SANGOMA partners. Also some work on the Lorenz 96 benchmark was performed and feedback provided to CNRS on the benchmark definitions. For WP5, ULg prepared the first draft version of DL5.1 with new data types explained, later updated by partners. WP6 and preparation of associated deliverables was basically executed by ULg, as well as tasks and deliverables of WP7.

6.2 AWI

During the first project year, AWI contributed to the work packages 1 to 4: In WP1, AWI provided for all deliverables detailed information about existing or desired new tools as well about the structure of our tools. An extensive amount of time required the preparation of DL1.3 ("Specifications of data model"). AWI was leading the preparation of this deliverable. As the data model is of critical importance, it had to be discussed thoroughly (e.g. during the kick-off meeting, a video conference, and by mail). One of AWI (L. Nerger) traveled to the Partner location at TU Delft to work on details of the data model. For WP2, AWI coordinated the first software release (Milestone MS5) together with A. Barth from ULg. For DL2.2 AWI initiated the uploading of source codes to the SVN-Server. AWI coordinated the preparation of DL2.3, which provides a detailed description of the tools included in the first software release. In WP3 AWI contributed to DL3.1 with input on different Kalman filter algorithms. In this work package, AWI also started the research work. A first publication considers the parallel performance of filter algorithms. A second publication that examines the influence of nonlinearity on smoother algorithms is currently in preparation. For WP4 AWI provided input on

the small-case benchmark for DL4.1 and worked on the implementation of this benchmark in our assimilation system.

6.3 NERSC

NERSC has affected little efforts in the first year as the qualified candidate in sea ice modeling has only been recruited from January 2013 (S. Bouillon). Meanwhile, NERSC has checked all deliverables related to common developments and contributed to DL5.1 with the description of new sea ice thickness measurements. NERSC tied links towards the MyOcean Arctic MFC (which it is leading), as well as ESA's Ocean Colour CCI, in which one of the methods supported by SANGOMA (The DEnKF with Gaussian Anamorphosis) is used to assimilate the OC CCI data and evaluate the impact of improved satellite data error estimates. Laurent Bertino and François Counillon have been teaching at the International Data Assimilation Workshop in Chongqing, P.R. China from 25th June to 7th July 2012, organized by the Institute of Atmospheric Physics, Chinese Academy of Science.

6.4 CNRS

CNRS/LEGI is responsible for WP4 (SANGOMA benchmarks). The first year of the project was thus mainly dedicated to the preparation of the benchmarks (model configurations, definition of the inverse problems, metrics for evaluating the system) and to their distribution to SANGOMA partners. The result of this activity is summarized in DL4.1. On the other hand, CNRS/LEGI also contributed to WP1 (inventory of existing tools in DL1.1) and to WP3 (nonlinear assimilation methods in DL3.1, for the anamorphosis method and the rank histogram filter). 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.

6.5 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 first software release.

6.6 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. During the first year of the SANGOMA program TUD worked on WP1, WP2 and WP3. The main tasks under WP1 are lead by TUD, which prepared an inventory document for DL1.1 that contains an initial list of common software components to be shared during SANGOMA. The document was circulated with SANGOMA partners and based on the feedback

an initial list of the software components was finalized. A deliverable document DL1.2 was then initiated by incorporating DL1.1. DL1.2 will serve as a living document for new software components. TUD worked closely with Lars Nerger (AWI) on the design of the interface. He came to TUD for two days during which we finalized the design. The deliverable document DL1.3 was then prepared by AWI and TUD and being circulated within SANGOMA for the feedback. Under task 1.4 of WP1 a document was initiated which gives a detailed specifications of the list of software components that are finalized in DL1.2. An Example specification for one of the tool (POD calibration tool) was prepared and circulated within SANGOMA for feedback. POD calibration tool is elaborated and detailed description of the physical interfaces is provided. This includes the sequence of input / output arguments and their types. Based on the feed a living document DL1.5 was updated. For an initial release of SANGOMA software under WP2 TUD provided two software components (POD and EnKF). A short document of these modules were also provided which gives inputs and output arguments of these software components. For WP3 a living document (DL3.2) is initiated which gives a detailed description of the DA methods that include uncertainty estimation and that can be implemented for large dimensional ocean models. This document will be updated throughout the project duration whenever a method is selected for a common DA toolbox.

Chapter 7

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:

7.1 ULG

The costs for the advisory board (under RTD) of the first year correspond to 15% of the total budget, which is less than the 25% on average for each of the four project meetings, because of reduced flight costs (notably from ESA to Belgium via a low-cost company).

The kick-off meeting was organised by ULg, which took 35% of the management budget dedicated to project meeting organization, which is more than the average of 25% but included the reception event.

Travel costs of ULg for dissemination were as planned.

As the operational user meeting was only held on 19-21 November (after the review period), associated costs will be charged to the next reporting period (and be lower than planned because of the joint organisation with myOcean).

Salary costs only covered 13% of the total cost (instead of the average 25%): Due to delays in setting up the internal accounting system to account for the coordinator's time, it was not charged to the project even though he spent a significant amount of his time on SANGOMA. Furthermore, as for all partners, finding appropriate data assimilation experts took some time and the post-doc position was filled only in July 2012.

7.2 AWI

The partner hired a PhD student, Paul Kirchgessner, to work on SANGOMA who started August 2012. As a PhD student, P. Kirchgessner is employed part time and 1.5 person months were charged to the project. 2 person months of Lars Nergers salary were charged to SANGOMA. This corresponds to 14% of the total salary cost instead of the average 25%. The travel costs were also slightly lower than planned.

7.3 NERSC

0.8 man month in RTD were charged, which amounts to 2% of the total salary cost, much lower than the average of 25%.

Travel costs were slightly lower than planned and 25% of the total dissemination costs were charged to the project.

7.4 CNRS

CNRS charged 4.55 MM which amounts to 5% of the total salary cost, lower than the average of 25%, because both CNRS groups had delays in hiring scientists.

7.5 UREAD

The partner hired a SANGOMA dedicated person from 2013 but in the meanwhile Peter Jan Leewens and Sanita Vetracarvalho contributed to the review paper and 1.5MM+3MM salary have been charged to the project which amounts to 15% of the total salary cost, lower than the average of 25%.

7.6 TUD

TUD charged 12% of the total salary cost, lower than the average of 25%.

Travel costs of TUD correspond to the planned costs.

Chapter 8

Plans

The project will continue as planned and particular effort in the next months will be focused on

- extending the list of tools,
- updating living documents,
- preparation of a newsletter,
- balancing the contribution of partners to the list of tools,
- adapting the tools to the interfacing specifications,
- development of new DA techniques and metrics for non-linear systems,
- using benchmarks and start documenting results of assimilation experiments.