SANGOMA: Stochastic Assimilation for the Next Generation Ocean Model Applications SPA.2011.1.5-03 call, project 283580

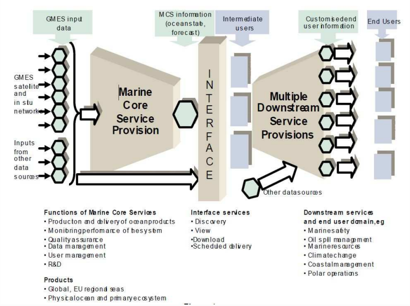
J.-M. Beckers and Sangoma consortium

November 24-25, 2011

1. Introduction
2. State of the art
3. Project
   - Workpackages
   - Groups involved
   - Budget

Introduction and objectives

MyOcean is the first E.U. project dedicated to the implementation of the GMES Marine Core Service (MCS) for ocean monitoring and forecasting.



MyOcean MCS does NOT foresee research in new Data Assimilation (DA) techniques, except short term implementation tasks.

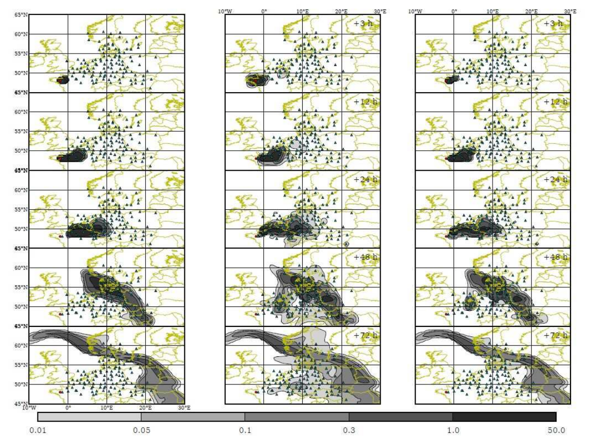
MyOcean Data Assimilation

- Optimal Interpolation
- Limited information on expected accuracy
- Deterministic approach
- Model specific
- Fragmented DA research at EU level hindering uptake of new techniques and new data types into operational mode

Sangoma objectives

- **networking** of expert teams at EU level in advanced data assimilation
- advance of **probabilistic assimilation methods** in high-resolution ocean models
- **harmonization** of existing ensemble assimilation concepts, algorithms and software
- convergence to a common data format in the DA (data-assimilation) framework
- **access** to validated tools, including benchmarks to the science community and operational centers
- **outreach and education** in advanced DA techniques
- **new products** in the form of improved error estimates of standard products
- investigation of the impact of **new data types** by exploring existing and new nonlinear measures for these impacts

State of the art and beyond



Boucquet et al 2010; Non-Gaussian approach versus Gaussian

DA techniques

1. Optimal Interpolation (prior covariances)
2. Extended Kalman Filters (propagation of covariances)
3. 3D and 4DVar methods (minimisation approach with adjoint model)
4. Ensemble methods (covariances by statistics on Monte-Carlo model realisations)
5. Particle filters (implementation of Bayes theorem)

Ensemble methods operational in meteorology, method for researchers in oceanography. Nonlinearities lead to major problems in 1-4.

DA toolboxes

- PDAF http://pdaf.awi.de/
- openDA http://www.openda.org
- Beluga/Sequoia
  http://sirocco.omp.obs-mip.fr/outils/Sequoia/Accueil/SequoiaAccueil.htm
- SESAM http://www-meom.hmg.inpg.fr/SESAM
- NERSC repository http://enkf.nersc.no
- DART http://www.image.ucar.edu/DAReS/DART
- OAK http://modb.oce.ulg.ac.be/mediawiki/index.php/Ocean_Assimilation_Kit

Implementing often similar schemes, preprocessing, postprocessing and perturbation tools, but with different optimisations, programming languages and specific ocean model support.

DA benchmarks

- Toy examples (Lorenz and its variants)
- Schematic situations (QG models in rectangular basins)
- Realistic situations (reasonable resolution models with controlled data)
- Operational situations (very high resolution and operational data flow)

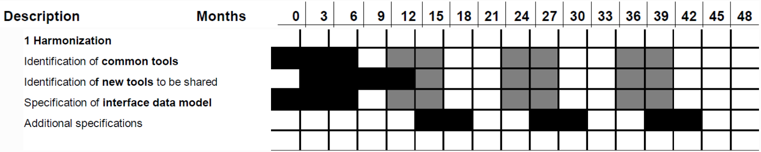
With different models and file formats, variable IP rights in implementations and outputs, diverse computing environments and diagnostics.

Beyond state of the art

- Ease up interchangeability of tools, formats and benchmarks
- Development of new DA techniques including for strongly non-linear problems
- Preparation for and evaluation of new data types (SMOS, geostationnary satellites, HF radars, ...)

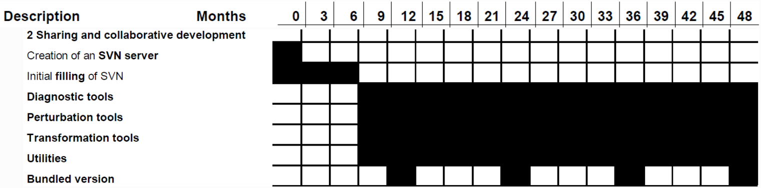
Structured into diagnostic components, perturbation-generation and stochastic methods, transformation tools, analysis steps and utilities.

WP1: Harmonization of assimilation tools (TUD)



Critical part: data-model sufficiently general yet not too complicated (at minimum compatible with models used in MyOcean), leading to specifications of interfaces and tools. Continous feedback and adaptation.

WP2: Sharing and collaborative development (AWI)



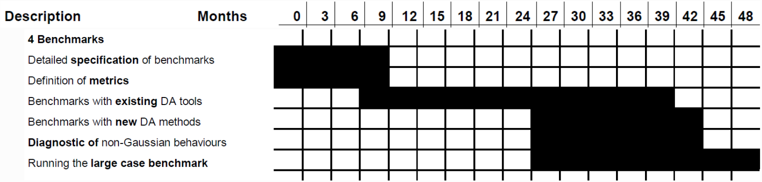
Complying with specifications of WP1 and inclusion of simple test routines with documentation. (.F95 or .m depending on use).

WP3: Innovative DA techniques (UREAD)

Description	Months 0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
3 Innovative Data Assimilation techniques																	
Non-linear data-assimilation methods	■	■	■														
Develop new non-linear DA methods				■	■	■	■	■	■	■	■	■	■				
Non-Gaussian assumptions						■	■	■	■	■	■	■	■	■			
Standard **implementation**										■	■	■	■	■	■		

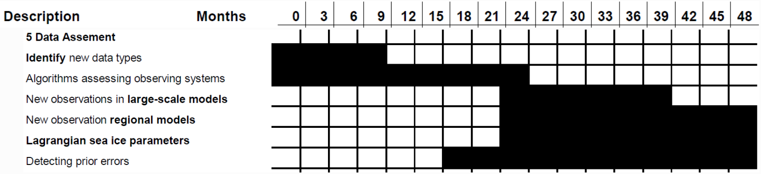
Most "explorative" WP on new methodologies (excluding methods requiring adjoint models). Must include new objective comparison techniques.

WP4: Benchmarks (CNRS-LEGI)



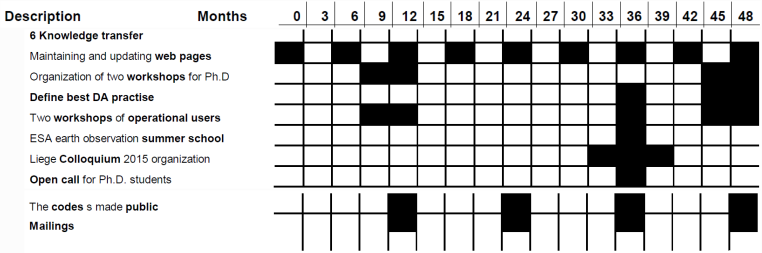
Benchmarks will include small (Lorenz), medium (double gyre with NEMO) and large cases (North Atlantic 1/4°). Benchmarks will include metrics to compare effect of different DA techniques. Will also later test new non-Gaussian criteria of WP3.

WP5: Data Assessment (NERSC)



New data: SST from geostationnary satellites and SSS from SMOS (large scale), coastal altimetry, HF radars and gliders (regional models). WP will include development of specific observation operators and new measures of impact of observing systems in non-Gaussian context.

WP6: Knowledge transfer (ULg)

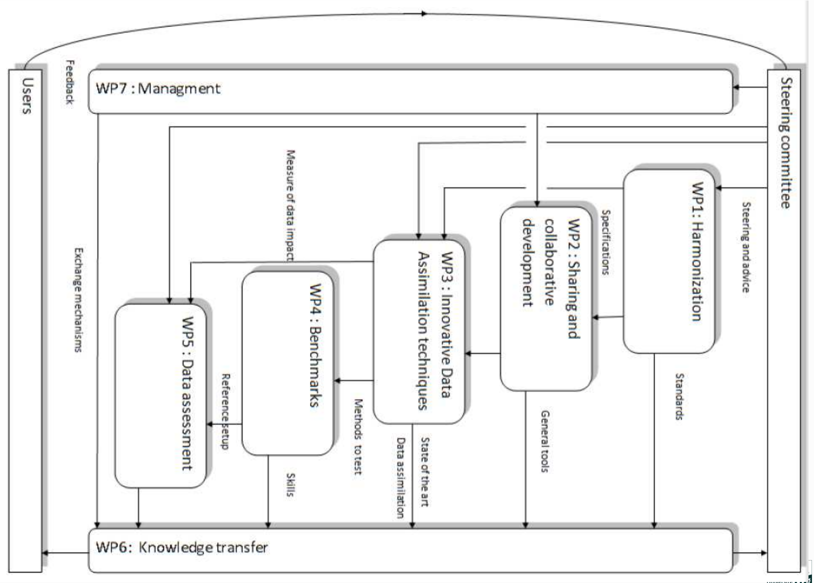


Important effort including workshops, best practise recommendation for operational models and final report.

WP7: Managment (ULg)

Description	Months	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
7 Project Management and Coordination																		
Organization of **meetings**		■				■				■				■				■
Financial **reporting**						■				■				■				■
Setting up of **project web pages**		■	■															
Setting up of a **communication network**		■	■	■														
Organization of **exchange mechanisms**						■				■				■				

Communication via several channels, exploiting developer platforms (forum and mailing lists).


Users
Feedback
WP7 : Managment
Steering committee
Measure of data impact
WP1: Harmonization
Steering and advice
Specifications
WP2 : Sharing and collaborative development
WP3 : Innovative Data Assimilation techniques
WP4 : Benchmarks
WP5 : Data assessment
Exchange mechanisms
Reference setup
Methods to test
Skills
Data assimilation
State of the art
General tools
Standards
WP6: Knowledge transfer

Partners

- P1-University of Liège: Jean-Marie Beckers and Alexander Barth. DA in regional models and perturbation generation.
- P2-University of Reading: Peter Jan van Leeuwen. Advanced innovative DA schemes.
- P3-Alfred Wegener Institute: Lars Nerger. DA expertise and scientific computing.
- P4-Delft University of Technology: Arnold Heemink and Martin Verlaan. DA in coastal seas with commercial software development and specifications.
- P5-CNRS-LEGI: Pierre Brasseur, Jean-Michel Brankart and Jacques Verron. DA at large scale, MyOcean.
- P5-CNRS-LEGOS: Pierre de Mey and Nadia Ayoub. DA expert with focus on objective observation-array design.
- P6-NERSC: Laurent Bertino, Geir Evensen, Pavel Sakov, François Counillon. Reference group in DA with strong involvment in operational aspects of MyOcean.

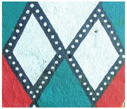
Consortium

- ULg for management and dissemination activities. Scientifically, ULg will bring expertise in perturbation generation, radar-data assimilation into regional models and parameter estimations.
- UREAD will be in charge of coordinating the innovative DA developments within Sangoma.
- AWI has a special interest in computing aspects and will naturally be in charge of the collaborative developments.
- TUD is well experienced in commercial software development and takes care of harmonization issues.
- CNRS has a broad experience in using NEMO in DA exercises and will supervise the benchmarkings, most of them using this model.
- NERSC experience of the TOPAZ implementation for operational purposes. In charge of data assessment work package, of particular interest to operational centers.

Budget

Project accepted as proposed: 14.5/15 in review process: EXPECTATIONS ARE HIGH

Why SANGOMA?



Logo choice

Some LaTeX beamer style files on

http://sangoma.svn.sourceforge.net/viewvc/sangoma/LaTeXtemplates/

Coffee time


sangoma retreat
Michael Podles
Derek Butler
PO Box 3075
Grose Vale NSW 2753
Tel: 02 4572 2592
MP: 0411 368 936
DB: 0411 500 022
info@sangoma.net.au
70 Grandview Lane
Bowen Mountain NSW 2753

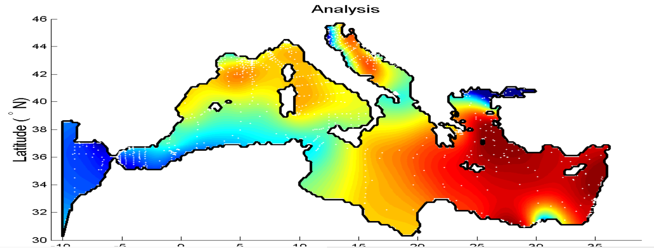
Optimal Interpolation

Combination of forecast $\mathbf{x}^f$ and observations $\mathbf{y}$

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathrm{T} + \mathbf{R}\right)^{-1}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right). \tag{1}$$

with $\mathbf{P}^f$ the forecast-error covariance matrix (reduced rank), **P** the observational error covariance and **H** the observation operator.

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^f = \mathbf{P}^f \; - \; \mathbf{P}^f\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathrm{T} + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{P}^f \tag{2}$$



Extended Kalman Filter

Initialization:
$$\begin{aligned} \mathbf{x}_0^a &= \mathbf{x} \\ \mathbf{P}_0^a &= \mathbf{P} \end{aligned}$$

Forecast:
$$\begin{aligned} \mathbf{x}_{n+1}^f &= \mathcal{M}(\mathbf{x}_n^a) \\ \mathbf{P}_{n+1}^f &= \mathbf{M}_n \mathbf{P}_n^a {\mathbf{M}_n}^T + \mathbf{Q}_n \end{aligned}$$

Analysis:
$$\begin{aligned} \mathbf{x}_{n+1}^a &= \mathbf{x}_{n+1}^f + \mathbf{K}_{n+1}\left(\mathbf{y}_{n+1} - \mathbf{H}_{n+1}\mathbf{x}_{n+1}^f\right) \\ \mathbf{K}_{n+1} &= \mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T \left(\mathbf{H}_{n+1}\mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T + \mathbf{R}_{n+1}\right)^{-1} \\ \mathbf{P}_{n+1}^a &= \mathbf{P}_{n+1}^f - \mathbf{K}_{n+1}\mathbf{H}_{n+1}\mathbf{P}_{n+1}^f \end{aligned}$$

3DVar

Minimization approach in 3D

$$J(\mathrm{x}) = \frac{1}{2}(\mathrm{x} - \mathrm{x}^f)^\mathsf{T} \mathbf{P}^{f\,-1} (\mathrm{x} - \mathrm{x}^f) + \frac{1}{2}(\mathbf{H}\mathrm{x} - \mathrm{y})^\mathsf{T} \mathbf{R}^{-1} (\mathbf{H}\mathrm{x} - \mathrm{y}) \quad (3)$$

or 4D

$$\begin{aligned} J(\mathbf{x}_0) &= \left(\mathbf{x}_0 - \mathbf{x}^i\right)^T \mathbf{P}^{i\,-1} \left(\mathbf{x}_0 - \mathbf{x}^i\right) \\ &+ \sum_{n=1}^{N} (\mathbf{y}_n^o - h_n(\mathbf{x}_n))^\mathsf{T} \mathbf{R}_n^{-1} (\mathbf{y}_n^o - h_n(\mathbf{x}_n)) \end{aligned}$$

with $\mathbf{x}_{n+1} = \mathcal{M}(\mathbf{x}_n)$.

Ensemble Kalman Filter

- In an ensemble simulation, a model is run a large number of times with different forcings, initial condition, parametrization,... within the uncertainty limit of the perturbed variable
- The spread of the ensemble reflects the resulting uncertainty in the model results
- Statistics such as mean and covariance can be computed from the ensemble

Ensemble representation: $\mathbf{x}^{(r)}, r = 1, \ldots, K$

$\mathbf{P} =< (\mathbf{x} - < \mathbf{x} >)(\mathbf{x} - < \mathbf{x} >)^{\mathsf{T}} >= \mathbf{X}\mathbf{X}^{\mathsf{T}}$ $\quad <>=$ ensemble average

In general slower convergence ($K^{-1/2}$) if K increases.
$K \approx 100 - 500$.

Particle filter and Bayes theorem

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{p(\mathbf{y}^o|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y}^o)} \tag{4}$$

- $p(\mathbf{x}|\mathbf{y}^o)$: a posteriori pdf, pdf of the model state $\mathbf{x}$ given the observations $\mathbf{y}^o$.
- $p(\mathbf{x})$: a priori pdf, pdf of the model state $\mathbf{x}$ before knowing the observations $\mathbf{y}^o$.
- $p(\mathbf{y}^o|\mathbf{x})$: probability of a measurement $\mathbf{y}^o$ if the system is in the state $\mathbf{x}$. For Gaussian observations errors:

$$p(\mathbf{y}^o|\mathbf{x}) = A \exp\left(\left(\mathbf{y}^o - h(\mathbf{x})\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{y}^o - h(\mathbf{x})\right)\right) \tag{5}$$

- $p(\mathbf{y}^o)$: The denominator is just a normalization to ensure that the pdf integrates to one.

The model pdf is represented by an ensemble (or by particles) $\mathbf{x}^{(r)}$ $(r = 1, \ldots, K)$:

$$p(\mathbf{x}) = \frac{1}{K} \sum_{r=1}^{K} \delta(\mathbf{x} - \mathbf{x}^{(r)}) \tag{6}$$

Initially all particles are equally probable, but by comparison to the observations, the particles who are closer to the observations are more likely than the particles who a farther away from the observations.

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{1}{K} \sum_{r=1}^{K} w_r \delta(\mathbf{x} - \mathbf{x}^{(r)}) \tag{7}$$

where the weights are given by:

$$w_r = \frac{p(\mathbf{y}^o|\mathbf{x}^{(r)})}{\sum_{r=1}^{K} p(\mathbf{y}^o|\mathbf{x}^{(r)})} \tag{8}$$

Problems

- Re-sampling: Particles with very low probability are ignored and particles with high probability are duplicated.
- No Gaussian assumption of the model error is necessary.
- Curse of dimensionality: Large number of particles are needed for high-dimensional problems.