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

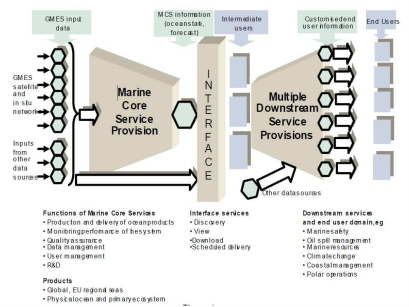
J.-M. Beckers and SANGOMA consortium

GeoHydrodynamics and Environment Research, MARE, University of Liège,
JM.Beckers@ulg.ac.be
www.data-assimilation.net

November 19-21, 2012, Geesthacht

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 is not focused on research in new Data Assimilation (DA) techniques, mostly short term (1 year) implementation tasks or performance issues.

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

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/OAK

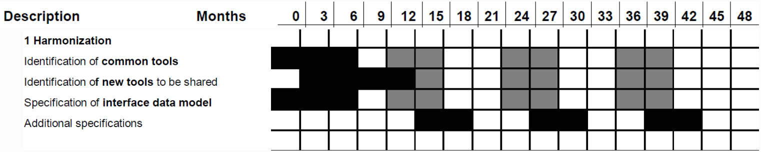
Implementing often similar schemes, preprocessing, postprocessing and perturbation tools, but with different optimisations, programming languages, specific ocean model support or coupling with models.

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, ...)

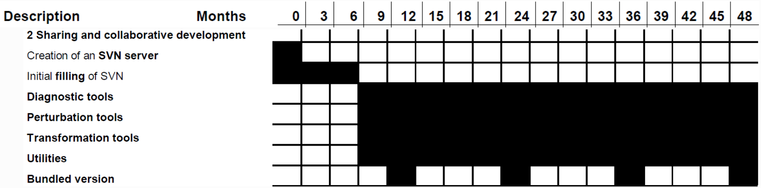
Structured into diagnostic components, perturbation-generation and stochastic methods, transformation tools, analysis steps and utilities.

WP1: Harmonization of assimilation tools (TUD)



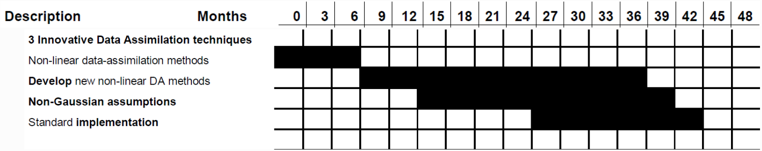
Critical part: data-model sufficiently general yet not too complicated (at minimum compatible with models used in MyOcean), leading to specifications of interfaces and tools. Continous feedback and adaptation.

WP2: Sharing and collaborative development (AWI)



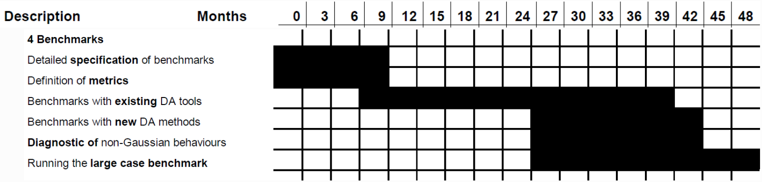
Complying with specifications of WP1 and inclusion of simple test routines with documentation. (.F95 or .m depending on use).

WP3: Innovative DA techniques (UREAD)



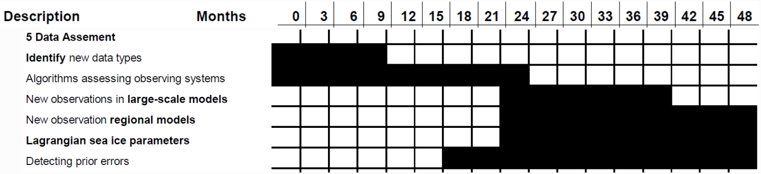
Most "explorative" WP on new methodologies (excluding methods requiring adjoint models). Must include new objective comparison techniques.

WP4: Benchmarks (CNRS-LEGI)



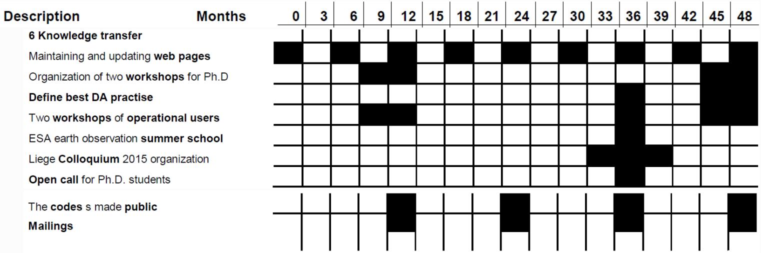
Benchmarks will include small (Lorenz), medium (double gyre with NEMO) and large cases (North Atlantic 1/4°). Benchmarks will include metrics to compare effect of different DA techniques. Will also later test new non-Gaussian criteria of WP3.

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.

Partners

- P1-University of Liège: Jean-Marie Beckers, Alexander Barth, Yajing Yan, François Laenen, Martin Canter. DA in regional models and perturbation generation.
- P2-University of Reading: Peter Jan van Leeuwen, Sanita Vetra-Carvalho. Advanced innovative DA schemes.
- P3-Alfred Wegener Institute: Lars Nerger, Paul Kirchgessner. DA expertise and scientific computing.
- P4-Delft University of Technology: Arnold Heemink, Martin Verlaan, Nils van Velzen, Umer Atlaf. DA in coastal seas with commercial software development and specifications.
- P5-CNRS-LEGI: Pierre Brasseur, Jean-Michel Brankart, Lucie Iskandar, Guillem Candille, Sammy Metref. 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, François Counillon. Reference group in DA with strong involvment in operational aspects of MyOcean.

Data model and interfacing

"Keep it simple" and need for common denominator between toolboxes:

- for data exchange via files:
  - use of netCDF file in CF compliant form.
  - provide output files in a similar form than input files (even if not perfectly fitting CF conditions).
  - when reasonable use version 3 features to enhance backward compatibility.
  - ensembles will be treated by working on a collection of files instead of a single big file.

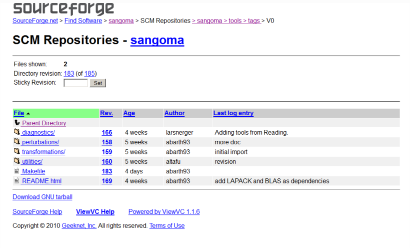
- for data exchange in memory (subroutine call):
  - use of basic FORTRAN structure `arrays`.
  - no derived types allowed (too much programming overhead in filling or adapting data types)
  - for more complex interfacing or data structures: use of `call-back` approach. Ex: to evaluate **Ry**, include a call-back function which when called with argument **y** returns the product **Ry**. The call-back program internal can be more complex but 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)
```

Tools

The first collection of tools on sourceforge (or `www.data-assimilation.net/Tools/`) .

- Full documentation on how to use and compile them in their present form.
- Adaptation of interfaces to the SANGOMA standard will now start and additional tools be included.



sourceforge

SourceForge.net > Find Software > sangoma > SCM Repositories > sangoma > tools > tags > V0

SCM Repositories - sangoma

Files shown: 2
Directory revision: 183 (of 185)
Sticky Revision: Set

File	Rev.	Age	Author	Last log entry
Parent Directory				
diagnostics/	166	4 weeks	larsnerger	Adding tools from Reading.
perturbations/	158	5 weeks	abarth93	more doc
transformations/	159	5 weeks	abarth93	initial import
utilities/	160	5 weeks	altafu	revision
Makefile	183	4 days	abarth93	
README.html	169	4 weeks	abarth93	add LAPACK and BLAS as dependencies

Download GNU tarball

SourceForge Help ViewVC Help Powered by ViewVC 1.1.6

Copyright © 2010 Geeknet, Inc. All rights reserved. Terms of Use

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

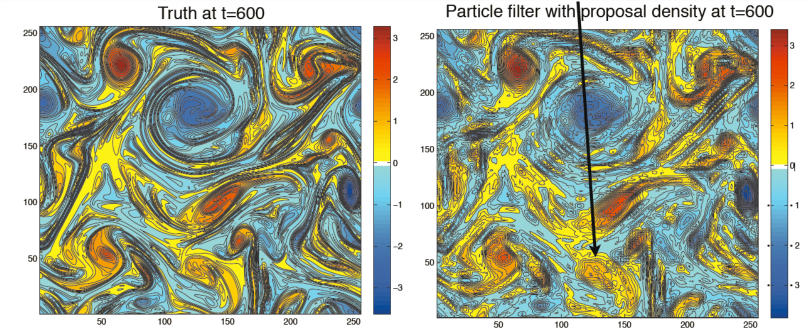
Empirical Gaussian Anamorphosis	Determine the empirical transformation function such that a transformed variable follows a Gaussian distribution

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

New DA techniques

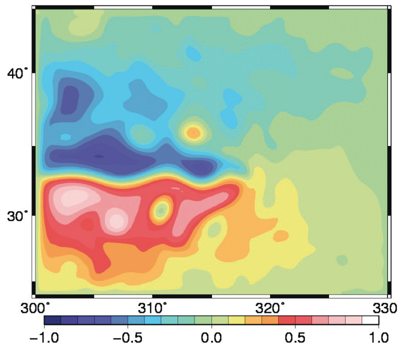
See previous talks and posters



Benchmarks

- small size: Lorenz 96
- medium size: Double gyre
- large size: Atlantic ocean

Fully detailed setup was formulated, see http://www.data-assimilation.net/



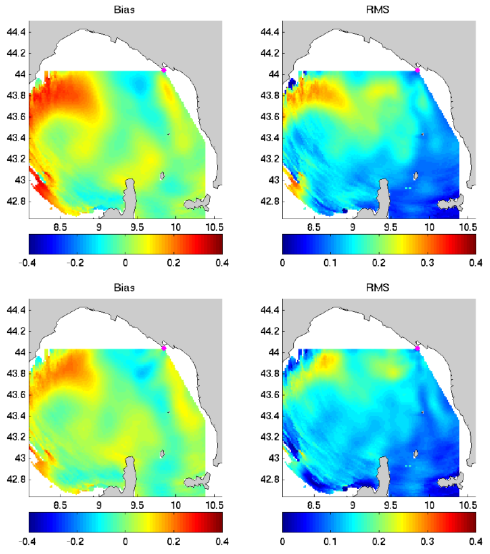
Example of other data types

Two WERA HF radar systems (Palmaria, San Rossore) by NATO Undersea Research Centre (NURC) from 2009 to 2010: provide velocity component directed towards (or away from) radar

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

where $k_b = \frac{2\pi}{\lambda_b}$

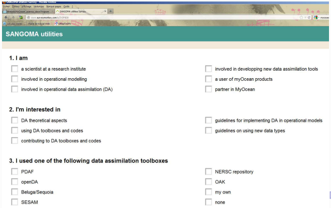
Assimilation with OAK



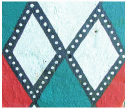
Wrap up: need your feedback

Survey: less than a minute of your time
http://www.data-assimilation.net/Events (google: SANGOMA data assimilation, then Events)

or directly on http://www.surveymonkey.com/s/ZX3P9D8



Why SANGOMA?



Logo choice

Poster time with drinks ?

Backup slides

Just in case some questions come up.

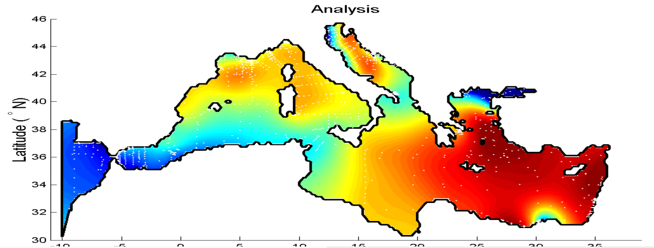
Optimal Interpolation

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

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right). \qquad (2)$$

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

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^f = \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{P}^f \qquad (3)$$



Extended Kalman Filter

Initialization:

$$\mathbf{x}_0^a = \mathbf{x}$$
$$\mathbf{P}_0^a = \mathbf{P}$$

Forecast:

$$\mathbf{x}_{n+1}^f = \mathcal{M}(\mathbf{x}_n^a)$$
$$\mathbf{P}_{n+1}^f = \mathbf{M}_n \mathbf{P}_n^a \mathbf{M}_n{}^{\mathrm{T}} + \mathbf{Q}_n$$

Analysis:

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

3DVar

Minimization approach in 3D

$$J(\mathrm{x}) = \frac{1}{2}(\mathrm{x}-\mathrm{x}^f)^{\mathsf{T}}\mathbf{P}^{f\,-1}(\mathrm{x}-\mathrm{x}^f)+\frac{1}{2}(\mathbf{H}\mathrm{x}-\mathrm{y})^{\mathsf{T}}\mathbf{R}^{-1}(\mathbf{H}\mathrm{x}-\mathrm{y}) \quad (4)$$

or 4D

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

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

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

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

Particle filter and Bayes theorem

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{p(\mathbf{y}^o|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y}^o)} \quad (5)$$

- $p(\mathbf{x}|\mathbf{y}^o)$: 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:

$$p(\mathbf{y}^o|\mathbf{x}) = A \exp\left((\mathbf{y}^o - h(\mathbf{x}))^{\mathsf{T}} \mathbf{R}^{-1} (\mathbf{y}^o - h(\mathbf{x}))\right) \quad (6)$$

- $p(\mathbf{y}^o)$: 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) $\mathbf{x}^{(r)}$ $(r = 1, \ldots, K)$:

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

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.

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

where the weights are given by:

$$w_r = \frac{p(\mathbf{y}^o|\mathbf{x}^{(r)})}{\sum_{r=1}^{K} p(\mathbf{y}^o|\mathbf{x}^{(r)})} \tag{9}$$

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.

Lorenz 96 model

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F \qquad (10)$$

cyclic conditions in i. Depending on value of F exhibits chaotic behavior with spatial structure.