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

November 24-25, 2011, Liège

Participants

- Alexander Barth, Jean-Marie Beckers (ULg)
- Arnold Heemink, Nils van Velzen, Martin Verlaan (TUD-DELTARES)
- Peter Jan Vanleeuwen (UREAD)
- Lars Nerger (AWI)
- Laurent Bertino (NERSC)
- Jean-Michel Brankart (CNRS-LEGI)
- Pierre De Mey (CNRS-LEGOS)

+

- Eric Dombrowsky (MyOcean)
- Lars Isaksen (ECMWF)

+

- Dominique Obaton (MyOcean products)
- Pierre-Philippe Mathieu (ESA-FRASCATI)

+

- Paola Chiarini (EU officer)

+

- Charlotte Peelen (Secretariat)

Agenda: day 1 (24/11) : contractual matters, steering committee meeting and data access

9:00- 9:30 :	Welcome and logistics
9:30-10:00 :	National contact point and coordinator's briefing on contractual matters
10:00-10:15 :	Coffee break
10:15-12:15 :	WP1-WP4: Review of objectives and tasks by WP leaders
12:15-13:45 :	Lunch
13:45-15:15 :	WP5-WP7: Review of objectives and tasks by WP leaders
15:15-15:45 :	EU officier P. Chiarini : advice from EU
15:45-16:00 :	Coffee break
16:00-17:00 :	P.-P. Mathieu : Data products available at ESA
17:00-18:00 :	D . Obaton : Data products available at MyOcean
20:00-00:00 :	Dinner

Agenda: day 2 (25/11) : Scientific session

9:00- 9:45 : L. Isaksen : Operational data assimimilation at ECMWF and medium-term plans
9:45-10:30 : E. Dombrowski : Operational data assimilation within MyOcean and medium-term plans
10:30-11:00 : Coffee break
11:00-11:35 : J.-M. Brankart: Data Assimilation at LEGI
- 11:35-12:10 : L. Bertino: Data Assimilation at NERSC
12:10-13:30 : Lunch
13:30-14:05 : P. de Mey: Data Assimilation at LEGOS
14:05-14:40 : L. Nerger: Data Assimilation at AWI
14:40-15:15 : A. Heemink: Data Assimilation at TUD
15:15-15:30 : Coffee break
15:30-16:05 : P. J. Van Leeuwen: Data Assimilation at UREAD
16:05-16:40 : A. Barth: Data Assimilation at ULg
16:40-17:30 : Discussion and feedback from advisors

Logistics

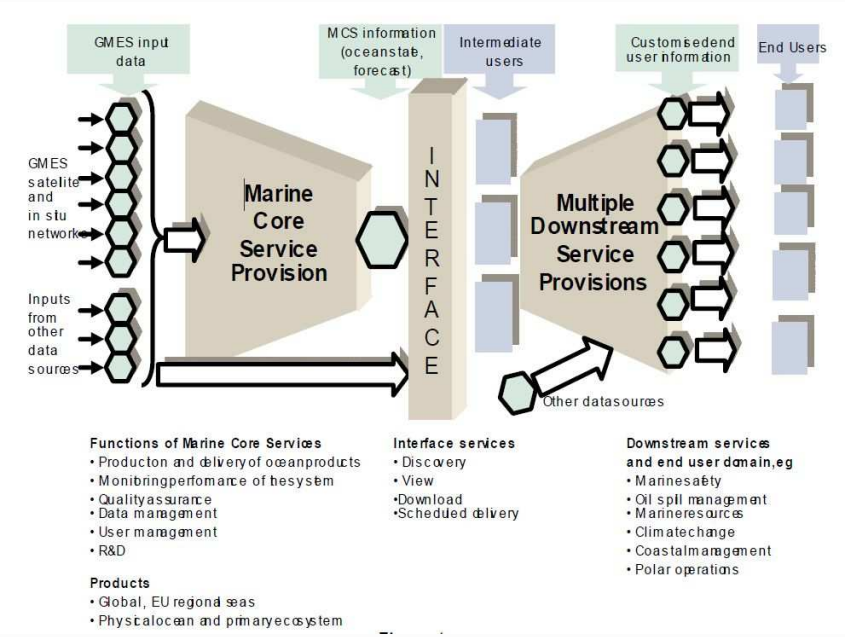
- Bus: some tickets are available if you plan to move in groups
- Wifi: use ULg-Open with `userid/password` provided
- Lunch: check menu for special needs as vegetarian food
- Dinner: La Capitanerie `www.lacapitaineriedeliege.be/` at 20:00
- Wrap-up, copy of presentations: leave copy on JMB's computer
- Printer available; for larger files send email to `oceanphy@ulg.ac.be`
- Questions ?

National Contact Point presentation and coordination issues

NCPSANGOMA.pdf

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.

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/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)

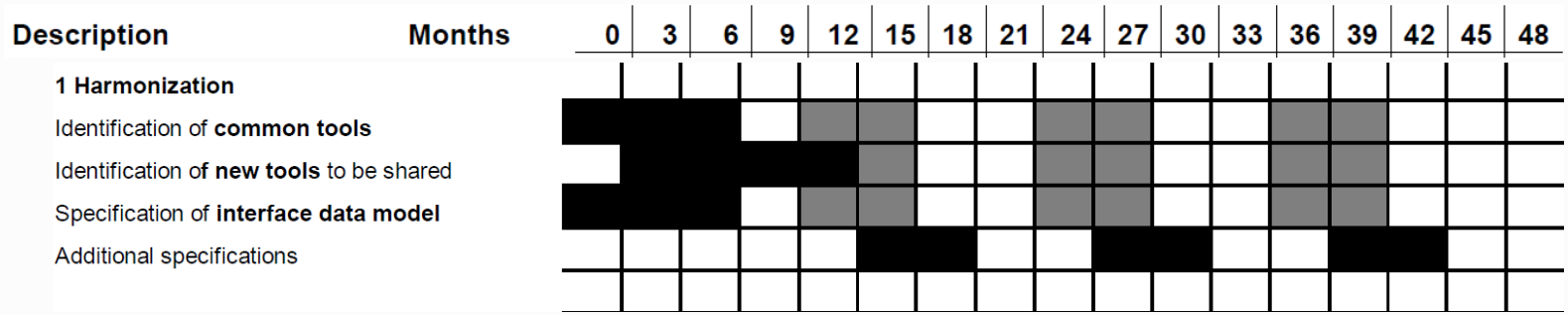
With different models and file formats, variable IP rights in implementations and outputs, diverse computing environments and diagnostics.

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, ...)

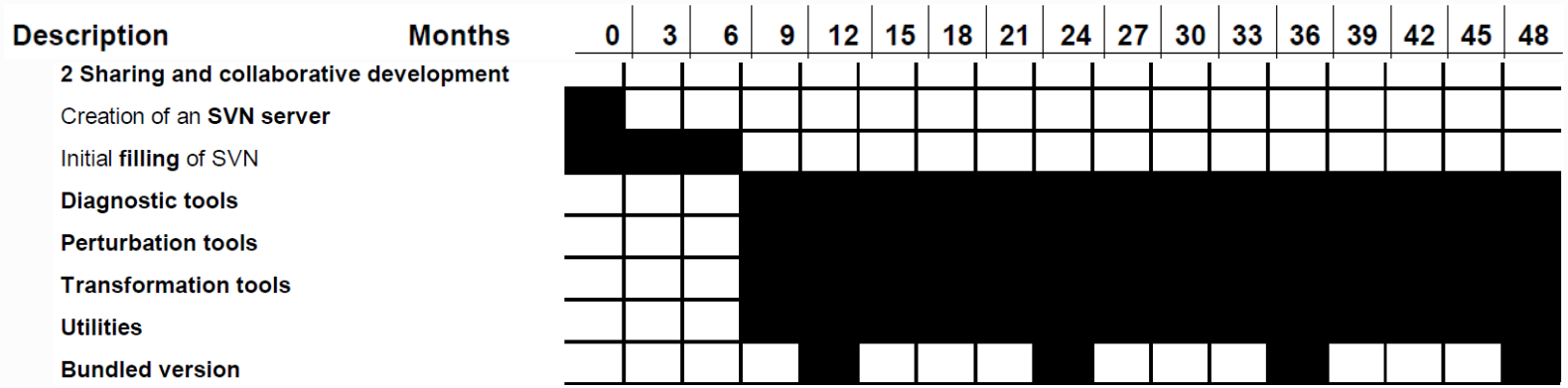
Structured into diagnostic components, perturbation-generation and stochastic methods, transformation tools, analysis steps and utilities.

WP1: Harmonization of assimilation tools (TUD)



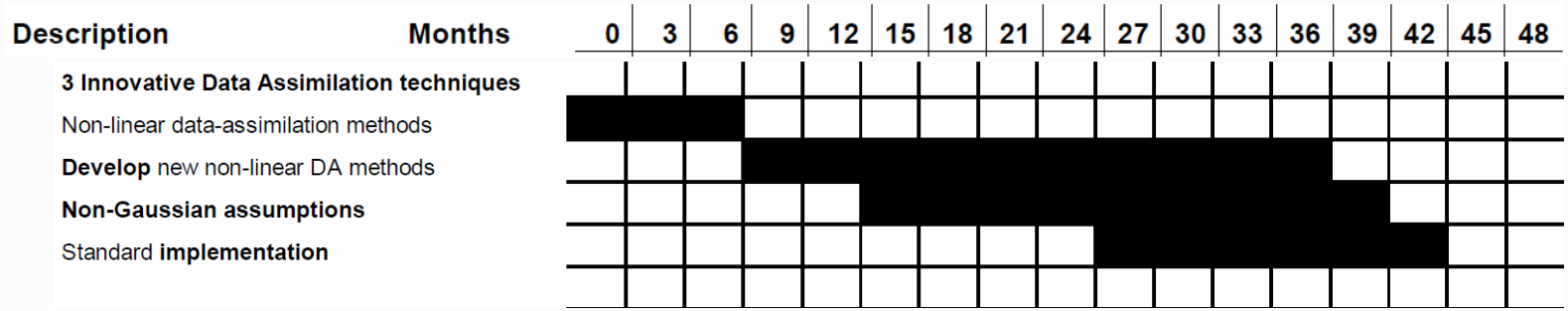
Critical part: data-model sufficiently general yet not too complicated (at minimum compatible with models used in MyOcean), leading to specifications of interfaces and tools. Continous feedback and adaptation.

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)



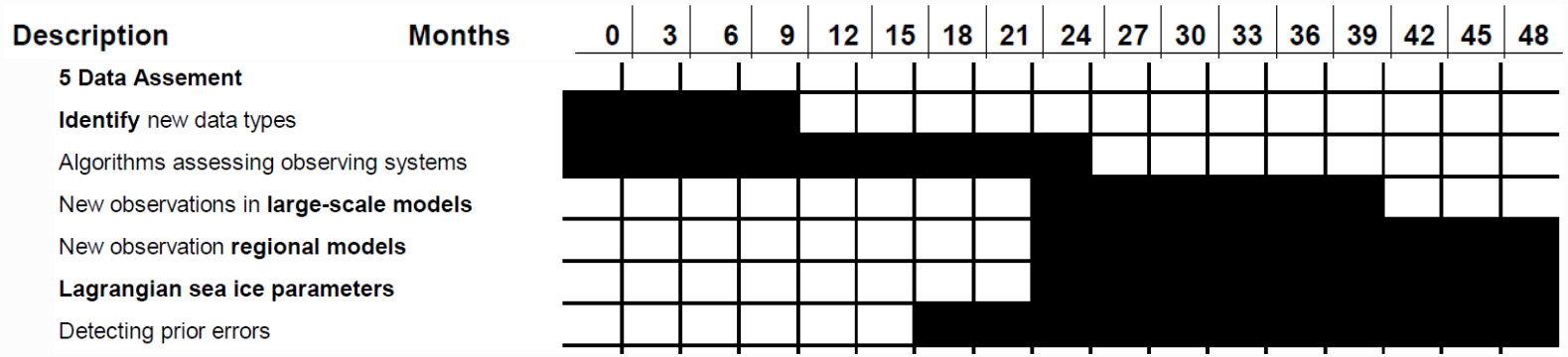
Most "explorative" WP on new methodologies (excluding methods requiring adjoint models). Must include new objective comparison techniques.

WP4: Benchmarks (CNRS-LEGI)

Description	Months	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
4 Benchmarks																		
Detailed **specification** of benchmarks		■	■	■	■													
Definition of **metrics**		■	■	■	■													
Benchmarks with **existing** DA tools				■	■	■	■	■	■	■	■	■	■	■	■			
Benchmarks with **new** DA methods											■	■	■	■	■	■		
Diagnostic of non-Gaussian behaviours											■	■	■	■	■	■		
Running the **large case benchmark**											■	■	■	■	■	■	■	■

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)



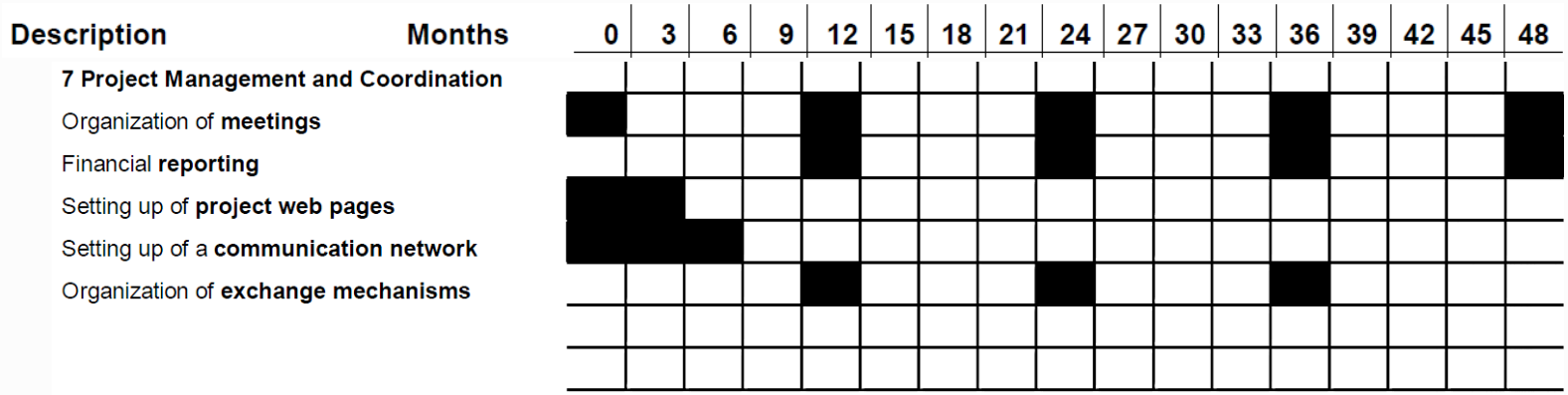
New data: SST from geostationnary satellites and SSS from SMOS (large scale), coastal altimetry, HF radars and gliders (regional models). WP will include development of specific observation operators and new measures of impact of observing systems in non-Gaussian context.

WP6: Knowledge transfer (ULg)

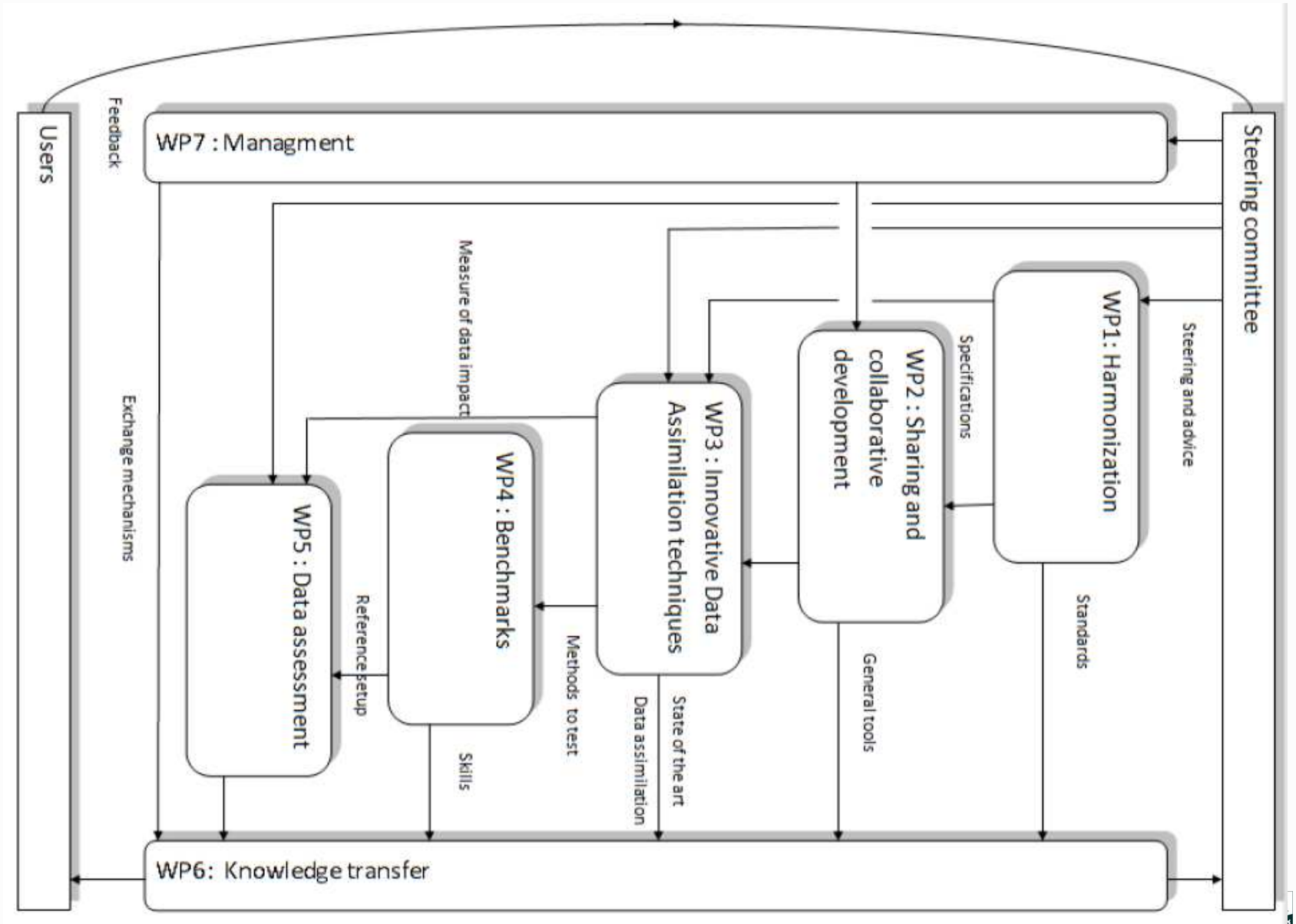
Description	Months	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
6 Knowledge transfer																		
Maintaining and updating **web pages**		■		■		■		■		■		■		■		■		■
Organization of two **workshops** for Ph.D					■	■											■	■
Define best DA practise														■			■	■
Two **workshops** of **operational users**					■	■								■			■	■
ESA earth observation **summer school**														■				
Liege **Colloquium** 2015 organization													■	■	■			
Open call for Ph.D. students														■				
The **codes** s made **public**					■				■					■				■
Mailings					■				■					■				■

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

WP7: Managment (ULg)



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

Science WP1-WP5 details by leaders

See `Linktopresentations`

WP 6: Knowledge transfers details

ULg responsible unless stated differently, but with input from all partners

- Maintaining and updating web pages: `http://www.data-assimilation.net/` is a placeholder for the moment, first report **M12**
- Organization of two workshops for Ph.D M24 and M48 (CNRS-LEGI and ULg)
- Define best DA practise (M36)
- Two workshops of operational users: **M12** to learn needs from users (particularly MyOcean); second together with PhD workshop M48
- Contribution to ESA earth observation summer school 2014
- Liege Colloquium organization May 2015 (now M42)
- Foster exchanges between partners
- Open call for Ph.D. students (attract non Sangoma Ph.D studends into partner locations) call in M32, selection in M36
- The codes (with automated checking) made public (AWI, first release **M12**)
- Mailings
- Final report

Liege colloquium May 2015 (slight unavaidable shift in deliverable D6.10 date M40->M43)

Possible shift of D6.9 (documentation for ESA summer school) from M30 to M32 (July 2014)

Deliverables

6	1	Web pages V1 report	0.0	UNIVERSITE DE LIEGE	0.5	Report	PU		31/10/2012 (12 months)	31/10/2012	Pending
6	2	Web pages V2 report	0.0	UNIVERSITE DE LIEGE	0.5	Report	PU		30/04/2014 (30 months)	30/04/2014	Pending
6	3	Web pages V3 report	0.0	UNIVERSITE DE LIEGE	0.5	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending
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
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
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
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
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
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
6	10	Report on International Colloquium	0.0	UNIVERSITE DE LIEGE	3.5	Report	PU		28/02/2015 (40 months)	28/02/2015	Pending
6	11	Publication	0.0	UNIVERSITE DE LIEGE	2.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending
6	12	Code release 1 documentation	0.0	UNIVERSITE DE LIEGE	1.0	Report	PU		31/10/2012 (12 months)	31/10/2012	Pending
6	13	Code release 2 documentation	0.0	UNIVERSITE DE LIEGE	1.0	Report	PU		30/04/2014 (30 months)	30/04/2014	Pending
6	14	Code release 3 documentation	0.0	UNIVERSITE DE LIEGE	2.0	Report	PU		31/10/2015 (48 months)	31/10/2015	Pending
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

Milestones

MS3	Collection of DA exercises	1	30	Collection of exercises for summer or winter schools
MS4	Collection of workshop presentations	1	24	Collection of workshop presentations (V1)
MS6	Best practise recommendation	1	36	Based on results of WP3 to WP5
MS7	Feedback from Ph.D exchanges	1	30	An assement of the Ph.D exchanges to this date will allow to redirect efforts if necessary
MS8	Collection of workshop presentations	1	48	Collection of workshop presentations (V2)

WP 7: Managment

- Standard tasks of meeting organization, reporting, CA maintenace, six-monthly informal reports
- Setting up of communication networks:
  - Need for more mailing lists (per WP, or for pure science questions) ?
  - Virtual conference via SKYPE (up to 25 participants free if all are on SKYPE). Any experience (quality/ease of use)?

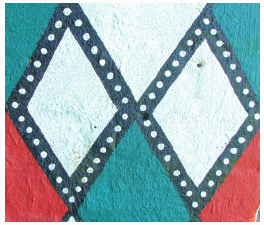
Meetings

- Kick-off meeting M1 (11/2011)
  - Virtual conference M6 (5/2012)
- First year meeting **M12** (11/2012)
  - Virtual conference M18 (5/2013)
  - Virtual conference M24 (11/2013)
- Intermediate meeting M30 (5/2014)
  - Virtual conference M36 (11/2014)
  - Virtual conference M42 (5/2015)
- Final meeting M48 (11/2015)
- Review meeting in Brussels M12, M30 and M48

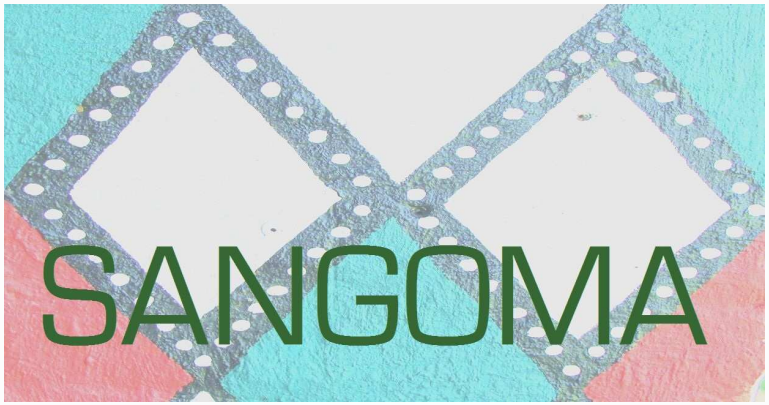
Where next?

- Sea and Sun `http://www.stareso.com/`,
- Relax `http://www.gite-roumaillac.fr`
- Hautes Fagnes by bicyle followed by a good restaurant `http://www.hotelzurpost.be/`
- Other ?

Why SANGOMA?



Logo choice



Some LaTeX beamer style files on

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

Deliverables and Milestones

7	1	Steering committe 1	0.0	UNIVERSITE DE LIEGE	0.1	Report	CO		30/11/2011 (1 months)	30/11/2011	Pending	
7	2	Steering committe 2	0.0	UNIVERSITE DE LIEGE	0.1	Report	CO		31/10/2012 (12 months)	31/10/2012	Pending	
7	3	Steering committe 3	0.0	UNIVERSITE DE LIEGE	0.1	Report	CO		30/04/2014 (30 months)	30/04/2014	Pending	
7	4	Steering committe 4	0.0	UNIVERSITE DE LIEGE	0.1	Report	CO		31/10/2015 (48 months)	31/10/2015	Pending	
7	5	Report on project information exchange tool	0.0	UNIVERSITE DE LIEGE	2.6	Other	RE		30/04/2012 (6 months)	30/04/2012	Pending	
7	6	Dissemination plan	0.0	UNIVERSITE DE LIEGE	2.0	Report	CO		31/01/2012 (3 months)	31/01/2012	Pending	

MS1	Web pages and communication tools	1	6	This platform is essential to the project

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

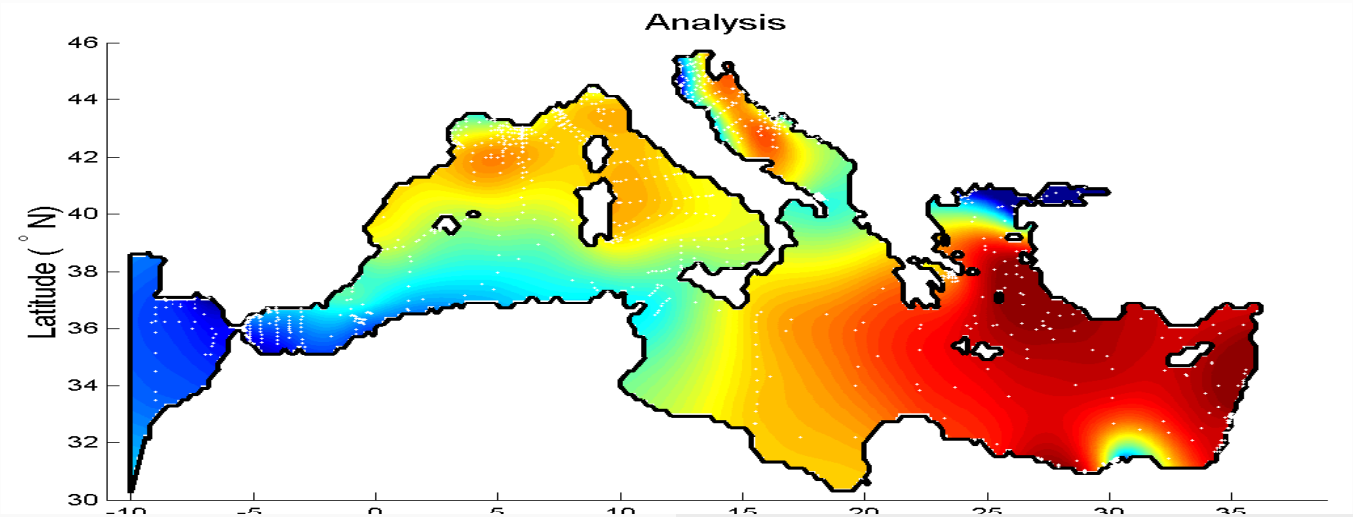
Optimal Interpolation

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

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

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

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



Extended Kalman Filter

Initialization:

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

Forecast:

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

Analysis:

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

3DVar

Minimization approach in 3D

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

or 4D

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

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

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

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:

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

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