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

Deliverable 4.2: Benchmark implementations

Due date: 31/10/2013

Delivery date: 15/04/2014

Delivery type: Report , public

Jean-Marie Beckers Alexander Barth
University of Liège, BELGIUM

Peter Jan Van Leeuwen
University of Reading, UK

Lars Nerger
Alfred-Wegener-Institut, GERMANY

Arnold Heemink Nils van Velzen
Martin Verlaan
Delft University of Technology, NETHERLANDS

Pierre Brasseur Jean-Michel Brankart Guillem Candille
Sammy Metref Florent Garnier
CNRS-LGGE, FRANCE

Pierre de Mey
CNRS-LEGOS, FRANCE

Laurent Bertino
NERSC, NORWAY

Chapter 1

Introduction

The main objective of SANGOMA is to advance the status of probabilistic assimilation methods and their applicability to reasaerch and operational MyOcean systems. A hierarchy of benchmarks of increasing complexity is then defined, and data assimilation tools are developed and implemented on these benchmarks. We remind that the specifications of each benchmark is described in deliverable 4.1.
The main goal of the present report is to list the different assimilation methods that have been implemented by SNGOMA partners on each benchmark. This catalog is presented as follows:
- model specifications (if different from the benchmark specifications in deliverable 4.1).
- perturbations generation methodology.
- observation specifications.
- implemented assimilation schemes.

Based on this inventory of benchmark implementations, a set of consistent assimilation experiments (*e.g.* performed with the medium case) will be derived and assembled with the aim of performance analysis and intercomparison studies.

Chapter 2

Implementations

For the three benchmarks, the descriptions and the general experiment specifications are detailed in the deliverable 4.1.

2.1 Small case benchmark

L96: Lorenz-96 model with 40 variables.

Partner: MEOM-LGGE (S. Metref)

- Particular model specifications:
  - 40 variables, parameter $F = 8$, time step $\delta t = 0.05$
  - initial condition : $x_i = 8, \forall i = 1, ..., 40$ except for $x_{20} = 8.01$ (van Leeuwen, 2010)
- Perturbations generation:
  - initial condition perturbed by an added Gaussian noise $\mathcal{N}(0, 0.1)$
  - tested for ensemble sizes between 20 and 100
- Particular observation specifications:
  - observational error variance $R = I_d$
  - various observation grids tested
  - various observation time frequencies tested
- Analysis method:
  - methods evaluated : EnKF, MRHF (Metref et al., submitted), RHF (Anderson, 2010), PF
  - localization of 19 points each side (Gaspari-Cohn, Eq.(4.10))

Partner: GHER (F. Laenen)

- Particular model specifications:
  - 40 variables, parameter $F = 8$, time step $\delta t = 0.05$
  - initial condition : $x_i = 8, \forall i = 1, ..., 40$ except for $x_{20} = 8.01$ (van Leeuwen, 2010)
  - spinup time: 150 time steps

- Perturbations generation:
  - initial condition perturbed by an added Gaussian noise $\mathcal{N}(0, 3e-5)$
  - tested for ensemble sizes of 25 and 80

- Particular observation specifications:
  - observational error variance $R = I_d$
  - various observation grids tested
  - various observation time frequencies tested

- Analysis method:
  - method evaluated: square root analysis scheme of the Ensemble Kalman Filter within Ocean Assimilation Kit (Vandenbulcke et al., 2006, Barth et al., 2007, Barth et al., 2008)
  - Anamorphosis transformation of model variables tested

Partner: AWI (P. Kirchgessner/L. Nerger)

- implementation of benchmark in PDAF

- Particular model specifications:
  - 40 variables, parameter $F = 8$, time step $\delta t = 0.05$
  - initial condition : $x_i = 8, \forall i = 1, ..., 40$ except for $x_{20} = 8.008$ and spin-up over 100 time steps.

- Perturbations generation:
  - initial condition perturbed by an added Gaussian noise with standard deviation 0.1, 0.5, or 1.0
  - tested for ensemble sizes between 8 and 40

- Particular observation specifications:
  - observational error variance $R = I_d$ and incomplete observations
  - different observation time frequencies or to 5 time steps tested

- Analysis method:
  - methods evaluated : (L)ESTKF (Nerger et al., 2012), (L)ETKF (Kirchgessner et al., 2014), (L)ESTKF with smoother extension (Nerger et al., 2013)
  - different localization radii with weighting by 5th-order polynomial (Gaspari-Cohn, 1999, Eq.(4.10)) - The EWPF (Van Leeuwen, 2010) is compared to the other ensemble methods.

2.2 Medium case benchmark

SQB: idealized configuration of the NEMO primitive equation ocean model (Cosme et al. 2010), *i.e.* a square and 5000-meter deep flat bottom ocean at mid latitudes.

Partner: MEOM-LGGE (P.-A. Bouttier)

- Particular model specifications:
  - 80 years spin-up.
  - 1 year to generate truth trajectory and observations.
  - background taken at 82 years of simulation.
- Perturbations generation:
  - none (variational method).
- Particular observation specifications:
  - assimilated data: sea surface height data from the model simulation sampled at the Jason-1 and SARAL/AltiKA tracks with observational error variance $R = 0.03I_d$.
- Analysis method:
  - incremental 4DVAR and 3DFGAT algorithms, provided by NEMOVAR framework; quasi-static strategy was also performed.
  - B is parametrized according to the model variability and correlation scales.
  - DA experiments chains 12 1-month cycles; for each cycle, 6 outer loops and 10 inner loops are performed.

Partner: MEOM-LGGE (G. Ruggiero/E. Cosme)

- Particular model specifications:
  - 60 years spin-up.
  - 10 years to generate the model perturbations.
  - 2 years to generate the observations
- Perturbations generation:
  - an EOF was calculated using the 10 years simulation; fields were taken every 5 days.
  - ensemble sizes between 20 and 100.
- Particular observation specifications:
  - assimilated data: sea surface height data from the model simulation sampled at the Jason-1 track with observational error variance $R = 0.064I_d$.
  - temperature profiles from the model simulation with observational error variance $R = 0.04I_d$ (similar to Cosme et al. 2010).
- Analysis method:
  - SEEK filter (Brasseur and Verron 2006) and SEEK smoother (Cosme et al. 2010); square root algorithms not requiring observation perturbations.
  - iterative SEEK smoother (Cosme et al. 2010, Kalnay and Yang 2010).

- Backward Smoother: the SEEK filter analysis at the end of the data assimilation window is propagated backwards in time by using the non-linear backward model to produce a smoothed trajectory.
- iterative Backward Smoother: iterative version of the Backward Smoother.
- Back and Forth Kalman Filter: iterative smoother based on the Back and Forth Nudging (Auroux and Blum 2005) and on the SEEK filter.
- all methods use the Domain Localization (radius ≈ 400 km) and the Observation Localization (inflates the observation error covariance by multiplying it by a Gaussian function with standard deviation equal to 170 km).
- all methods use the background error covariance inflation (ρ varies between 0.95 and 1, see Cosme et al. 2010 for details).

Partner: GHER (Y. Yan)

- Particular model specification:
  - 40 years spin-up.
  - 10 years to generate the model perturbations.
  - 1 year to generate the observations.
- Perturbations generation:
  - model output every 30 days during 10 years simultion for the 100-member ensemble; 39 modes from the 100-member ensemble by singular value decomposition for the 40-member ensemble.
  - ensemble size: 40 and 100.
- Particular observation specifications:
  assimilated data:
  - sea surface height data from model simulation sampled at the Jason-1 and Envisat tracks with observational error 6 cm.
  - temperature profiles from the model simulation with observational error 0.3°.
- Analysis method:
  - square root analysis scheme of the Ensemble Kalman Filter within Ocean Assimilation Kit (Vandenbulcke et al., 2006, Barth et al., 2007, Barth et al., 2008)
  - observation localisation (radius: ≈ 250 km and ≈ 400 km) with a Gaussian function.
  - background error covariance inflation (1.05 and 1.1, see medium benchmark report by Y. Yan for details)

Partner: AWI (P. Kirchgessner/L. Nerger)

- Particular model specifications:
  - 75 years spin-up.
  - 1 year to generate truth trajectory and observations.
  - Additive gaussian random noise is added at each time step for assimilation with EWPF.
  - The covariance matrix for the model error is generated to get a balanced state.

- Perturbations generation:
  - The background ensemble is generated from a free run simulation (years 41-50), where output at every 30 is used to initialise the ensemble.
  - The ensemble is generated through a SVD decomposition using the sample from the background run.

- Particular observation specifications:
  The observation specification and observation files are provided by Yajing Yan from University of Liège, Belgium
  Assimilated data:
  -sea surface height data from the model simulation sampled at the Jason-1 and ENVISAT tracks with observational error variance $R = 0.03 I_d [m]$ is generated.
  -For temperature data, spatially uncorrelated gaussian noise with standard deviation of $0.3°$ is added to the true temperature field. The resolution of the observational grid is $3° \times 3°$. It is generated to mimic ARGO observations.

- Analysis method:
  - methods evaluated : (L)ESTKF (Nerger et al., 2012) and Equivalent Weights Particle Filter (EWPF, Van Leeuwen, 2010).
  - As observation operator two or three dimensional interpolation is used.

Partner: TUDelft

- Particular model specifications:
  - 50 years spin-up.
  - 10 years to generate the model perturbations.
  - 1 year to generate the observations

- Perturbations generation:
  - initial ensemble generated with 10 years interval of free model simulation with output every 30 days.
  - ensemble sizes between 20 and 100 by taking equal time interval from initial ensemble.

- Particular observation specifications:
  - Taking model output at 75th year as truth.
  - assimilated data: sea surface height data and temperature profiles from the model simulation (similar to Y.Yan 2013).
  - analysis frequency: 2 days

- Analysis method:
  - EnKF filter (Evensen and Burgers 1998) and DEnKF filter (Sakov and Oke 2008); which are part of OpenDA toolbox to perform the assimilation.

2.3 Large case benchmark

NATL025: realistic configuration of the NEMO ocean model, for the North Atlantic Ocean, at a $1/4°$ resolution (Barnier et al, 2006).

Partner: MEOM-LGGE (G. Candille)

- Particular model specifications:
  - only dynamical part of the model is used here.
  - spinup from January 1st 1989 (Levitus) to January 1st 2005.
- Perturbations generation:
  - using a set of Temperature/Salinity perturbations to simulate the unresolved T/S fluctuations in the equation of state (Brankart 2013); computation of the random perturbations as a scalar product of the local gradient and a random walk generated by an autoregressive process (first order).
  - growing perturbations during 6 months (from January 1st to June 29th 2005).
  - ensemble size N=96.
- Particular observation specifications:
  - assimilated data: altimetric data from Jason-1 and Envisat missions (Sea Level Anomaly) with observational error variance $R = 0.1I_d$.
  - verification data: altimetric data (SLA) and T/S ARGO-profiles.
- Analysis method:
  - start at June 29th 2005.
  - 10-day assimilation window (Jason-1 cycle).
  - ensemble mean and anomalies updated with SEEK algorithm (Brasseur and Verron 2006); square root algorithm not requiring observation perturbations; update at the center of the assimilation window.
  - localization parameters: domain localization with radius $4.5°$ (≈ 433 km at $30°$ N) and observation influence defined by a Gaussian function with standard deviation equal to $1.5°$ (≈ 150 km at $30°$ N).
  - observation equivalents of ensemble mean and anomalies at appropiate observation time (4D) with NEMO-obs module (H operator).
  - building increments for each ensemble anomalies and performing Incremental Analysis Updates (IAU) integrations over the 10-day window (resulting in continuous ensemble of trajectories).

Partner: GHER (Y. Yan)

- Particular model specifications:
  - only dynamical part of the model is used.
  - spinup from January 1st 1989 (Levitus) to January 1st 2005.
- Perturbations generation:
  - add realistic noise in the forcing variables (wind velocities, air temperature, long wave and short wave radiation flux)

- growing perturbations during 6 months (from January 1st to June 29th 2005).
- ensemble size N=60

- Particular observation specifications:
  - assimilated data: altimetric data from Jason-1 (Sea Level Anomaly) with observational error of 5 cm, sea surface temperature from AVHRR with observational error map, ARGO temperature profiles with observational error 0.3°.
  - verification data: altimetric data from Envisat (Sea Level Anomaly), Mercator sea surface temperature reanalysis, ARGO salinity profiles.

- Analysis method:
  - start at June 29th 2005.
  - 10-day assimilation window (Jason-1 cycle).
  - square root analysis scheme of the Ensemble Kalman Filter within Ocean Assimilation Kit (Vandenbulcke et al., 2006, Barth et al., 2007, Barth et al., 2008)
  - observation localisation (radius ≈ 300 km) with Gaussian function.
  - Incremental Analysis Updates (IAU), scheme 0 (see Yan et al. 2014 for detail), building T/S increments

Chapter 3

References

Anderson J. L., 2010. A non-gaussian ensemble filter update for data assimilation. *Mon. Wea. Review*, **138**, pp 4186–4198.

Auroux D. and J. Blum, 2008. A Nudging-based data assimilation method for oceanographic problems: the Back and Forth Nudging (BFN) algorithm. *NPG*, **15**, pp 305–319.

Barnier B., G. Madec, T. Penduff, J.-M. Molines, A.-M. Treguier, J. Le Sommer, A. Beckmann, A. Biastoch, C. Böning, J. Dengg, C. Derval, E. Durand, S. Gulev, E. Remy, C. Talandier, S. Theetten, M. Maltrud, J. McClean, and B. DeCuevas. 2006. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy permitting resolution. *Ocean Dynamics*, **56**, pp 543–567.

Barth A., A. Alvera-Azcárate, J. Beckers, M. Rixen and L. Vandenbulcke, 2007. Multigrid state vector for data assimilation in a two-way nested model of the ligurian sea. *Journal of Marine System*, **65**, pp 41–59.

Barth A., A. Alvera-Azcárate and R. Weisberg, 2008. Assimilation of highfrequency radar currents in a nested model of the west florida shelf. *Journal of Geophysical Research - Oceans*, **113**, pp 1–15.

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

Brasseur P. and J. Verron, 2006. The SEEK filter method for data assimilation in oceanography : a synthesis. *Ocean Dynamics*, **56**, pp 650–661.

Cosme E., J.-M. Brankart, J. Verron, P. Brasseur, and M. Krysta, 2010. Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation. *Ocean Modeling*, **33**, pp 87–100.

Kalnay E. and S-C. Yang, 2010. Accelerating the spin-up of ensemble Kalman filtering. *QJRMS*, **136**, pp 1644–1651.

Kirchgessner P., L. Nerger and A. Bunse-Gerstner, 2014. On the choice of an optimal localization radius in ensemble Kalman filter methods. *Monthly Weather Review*, accepted.

Metref S., E. Cosme, C. Snyder and P. Brasseur, 2014. A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. *Submitted to NPG*.

Nerger, L., T. Janjic, J. Schröter and W. Hiller, 2012. A unification of ensemble square root Kalman filters. *Mon. Wea. Rev.*, **140**, pp 2335–2345

Nerger, L., S. Schulte and A. Bunse-Gerstner, 2014. On the influence of model nonlinearity and localization on ensemble Kalman smoothing. *Q. J. Roy. Meteorol. Soc.*, in press.

Sakov P. and P. Oke, 2008. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus* **60A**, pp 361–371.

Van Leeuwen, P., 2010. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Q.J.R.M.S.*, **136**, pp 1991–1999.

Vandenbulcke L., A. Barth, M. Rixen, A. Alvera-Azcárate, Z. B. Bouallegu and J. Beckers, 2006. Study of the combined effects of data assimilation and grid nesting in ocean models: Application to the Gulf of Lions. *Ocean Science*, **2**, pp 213–222.

Yan Y., 2013. SANGOMA: Stochastic Assimilation for the Next Generation Ocean Model Applications - Medium case benchmark report.

Yan Y., A. Barth and J. Beckers, 2014. Comparison of different assimilation schemes in a sequential Kalman filter assimilation system. *Ocean Modelling*, **73**, pp. 123–137, *DOI: 10.1016/j.ocemod.2013.11.002*.