NEMO data assimilation and PDAF cooperation

Lars Axell & Ye Liu, SMHI
Outline

• NEMO data assimilation
  – An introduction
  – NEMO DA
  – 3D/4D Var
  – 3D/4D EnVar

• PDAF
  – Different ensemble filters
  – NEMO-Nordic implementation
  – Cooperation
NEMO data assimilation: An introduction

• In general, to make good forecasts we need two things:
  – A good initial condition
  – A good model, used to integrate the initial condition forward in time (e.g. NEMO, HBM, GETM, … etc)

• A good initial condition is obtained from a combination of:
  – A first guess (normally a short forecast, e.g. +6H)
  – Observations
  – A data assimilation program
NEMO data assimilation: NEMO DA

- Two methods can be used:
  - Modify the restart files (this method was used with HIROMB, and HBM in MyOcean projects; used by e.g. Mercator)
  - Produce an increment file for NEMO to read (NetCDF format; “assim_background_increment.nc”; used operationally at SMHI)

- Increments are implemented in NEMO-3.6 for these variables:
  - S/T
  - U/V
  - SLA (sea level anomaly)
  - Ice concentration and ice thickness (implemented by SMHI)
NEMO data assimilation: 3D/4D Var

- Variational techniques, meaning a cost function $J$ is minimized:

\[
J(w) = \frac{1}{2} w^T w + \frac{1}{2} \sum_{i=1}^{I} (H M_i U_i w - d_i)^T R^{-1} (H M_i U_i w - d_i)
\]

\[
\nabla_w J = w + \sum_{i=1}^{I} U_i^T M_i^T H^T R^{-1} (H M_i U_i w - d_i)
\]

- NEMOVAR:
  - Developed by a European group; UK, France, ...?
  - Not part of the NEMO distribution
  - Not tested at SMHI yet

- Normally requires a parameterization of horizontal and vertical structure functions (correlations); could introduce problems in the e.g. the thermocline

- 4D-Var requires tangent linear forward and backard versions of the full nonlinear model (high cost, both for development and running)
NEMO data assimilation: 3D/4D EnVar

- "Ensemble Variational data assimilation"
- Based on Liu et al. (2008; 2009); developed for the ocean by SMHI 2010-2018 (Axell and Liu, 2016; 2017)
- Produces an increment file
- Needs an ensemble of model states
- Used in MyOcean/Copernicus reanalyses with HIROMB
- Code parallelized with OpenMP
- Still rather heavy to run for reanalysis projects, better to use ensemble filter methods for speed!
NEMO data assimilation:
3D/4D EnVar (cont.)

• Advantages of 4D EnVar (compared to standard 4D-Var):
  – Background error covariances are...
    • flow dependent (change with location and time)
    • updated daily
  – No tangent linear or adjoint model versions required!!!
  – Similar results, but much faster than standard 4D-Var
    (Gutafsson and Bojarova, 2014)

• Disadvantages, or pit falls:
  – Quality of results depends on ensemble used
  – The work of improving the ensemble never ends...
PDAF:
Applications at SMHI

- Recently used in Copernicus reanalysis
  - 1993-2016
  - V4
  - Assimilated variables:
    - S/T profiles and SST
    - Biogeochemical variables (in-situ)

- Ongoing project: Sea Level Anomalies

- Faster than 3D EnVar (used in V3 with HIROMB)
PDAF:
Different **ensemble** filters

- Parallell Data Assimilation Framework
  - developed at AWI
  - [http://pdaft.awi.de/trac/wiki](http://pdaft.awi.de/trac/wiki)
- PDAF: not a data assimilation method, rather a package of **ensemble**-based filters:
  - SEEK
  - (L)SEIK ← used by SMHI!
  - (L)EnKF
  - (L)ETKF
  - (L)ESTKF ← used by BSH and DMI! Also recommended by AWI
  - (L)NETF

- Localized often best for our applications

"L" means "localized"
PDAF: NEMO-Nordic implementation

- **NEMO-Nordic-PDAF:**
  - Localized Sequential Evolutive Interpolated Kalman Filter (LSEIK) algorithm.
  - Updated to the almost latest release of NEMO (V3.6)

- **Observation data:**
  The observation data is processed and quality controlled; super-observations applied (thinning in both horizontal and vertical direction) before the data assimilation

- **Ensemble model states generation:**
  Empirical orthogonal functions (EOFs) are applied to generate the Principal component analysis (PCA) for the ensemble model states (T, S, U, V and SSH). Currently, the PCA is based on a historical model run

- **PROS & CONS**
  Computationally cheap. One forward NEMO run, can assimilate most of the useful ocean data, Ensemble spans different seasons/years, No dynamical evolution of the error..
Observations

Moorings

ICES (http://ices.dmi.dk )

SHARK (http://sharkweb.smhi.se)

Buoy

CTD
Surface and bottom salinity (V4) VS. independent observation
Monthly-averaged bias and RMSE of temperature (left) and salinity (right).
Average (1993-2016) vertical profiles for temperature and salinity

Time series of temperature and salinity at top (5m) and bottom level of BMJP1
Map of the RMSE of SST from ASSIM run (left) and FREE run (right) calculated against IceMap SST in 2010, respectively.

The evolution of the simulation SST bias, error and number of IceMap SST in 2010.
PDAF: Cooperation

- Comparison on-line vs. off-line versions; different?
- Comparison of different ensemble filters:
  - LSEIK Filter
  - LESTK Filter
  - ...
- Development of ensemble generation techniques
- Technical implementation for NEMO-Nordic?
- Test implementations for sea ice, SLA, chlorophyll-a, ...
- Regular meetings within a data assimilation working group; Who? How? How often?
- ...?
The end...
About on-line version of PDAF
Online coupled NEMO with PDAF

Add coupled model & filter parallelization

Initialize ensemble

Perform filter analysis

Perform final processing & visual

Start

Init. parallelization

Init_parallel_pdaf

Initialize Model (Mesh, fields)

Init_pdaf

Do i=1,nsteps

Time Step
Consider BC
Consider forcing

assimilate_pdaf

Post-processing

Stop
2-level parallelization of the assimilation system

MPI and OpenMP are used for parallelization:
- OpenMP is used for parallelization of NEMO
- MPI is used to distribute the ensemble members