# 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





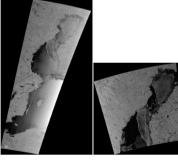
## NEMO data assimilation: An introduction

- 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:

$$J(w) = \frac{1}{2} w^{T} w + \frac{1}{2} \sum_{i=1}^{I} (H M_{i} U_{i} w - d_{i})^{T} R^{-1} (H M_{i} U_{i} w - d_{i})$$
  

$$\nabla_{w} J = w + \sum_{i=1}^{I} U_{i}^{T} M_{i}^{T} H^{T} R^{-1} (H M_{i} U_{i} w - d_{i})$$
Tangent Linear model

• NEMOVAR:

Adjoint model

- 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!

and the second s
and the constants of the second secon
approximation of No is n > 10, the calculated of the concentration of No is approximately approximately in a much is exclusive and the calculated on the concentration of No. (In the calculated of the calculated
dimension of % is much amaz dimension of % is much as a calculated for the to N which is much a is vector and performed on the to N which is much and the vector and the contract of the to N which is the vector a in vector and in the to N which is the vector a in vector and the to N which is the vector a second the vector a second the to N which is the vector a second th
dimension which is in vector a is vector f according $r \times N$ which is in vector in vector f according to $r \times N$ ations), ation space), according to $r \times N$ ations to $r \times N$ at $r \to N$
an × N ations), ations in space), connect states the show
m × reations vation space Han we encounts obtained on the source of the
dimension of X is in a sub- matrix state in the second state in the second state is a sub- of the second state is a sub- matrix state is a sub- matrix state is a sub- matrix state is a sub- state is a sub- state is a sub- matrix state is
$\begin{array}{c} \label{eq:constraints} \begin{tabular}{lllllllllllllllllllllllllllllllllll$
$ \begin{array}{l} \label{eq:constraint} \begin{tabular}{lllllllllllllllllllllllllllllllllll$
ents begin ion i with all
elemento function -1 (Here intali
The ozer cost in aro mathematic
$\begin{array}{c} Interview of the second seco$
The elements of the control restore entrol elements of the control elements of the con
mining vector of a start ion
In atroi 1 - 1 + 2 are ma walk
contraction was another and another and
1 = 2 tran varia to ou ff.a.
I(W) the cover the 11 or
otes rol that an in
The elements of the control where value dependence of the set of the control where value dependence of the control where value dependence of the control vector with the vector of a vector of the ve
The elements of the control of the control of the set of the control of the cont
and it sin collo game star of a
here weer sesure inthe alle the
whe abs also he in the age
the and wich the alone
in the senter where at his is done it
15 and i the des A
The elements beginner of the transformed of the observation of the transformed of the transformed of the transformed of the observation of the transformed of the transformed of the observation of the transformed of the tr
ally the one offer offer.
the relie of the method of
all and the second
100 Mar 100 Co 13

# 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://pdaf.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 reccommended by AWI
  - (L<u>)</u>NETF

"L" means "localized"

Localized often best for our applications

# PDAF:

## **NEMO-Nordic implementation**

#### NEMO-Nordic-PDAF:

- Localized Sequential Evoluative 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 (**EOF**s) 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 histroical model run

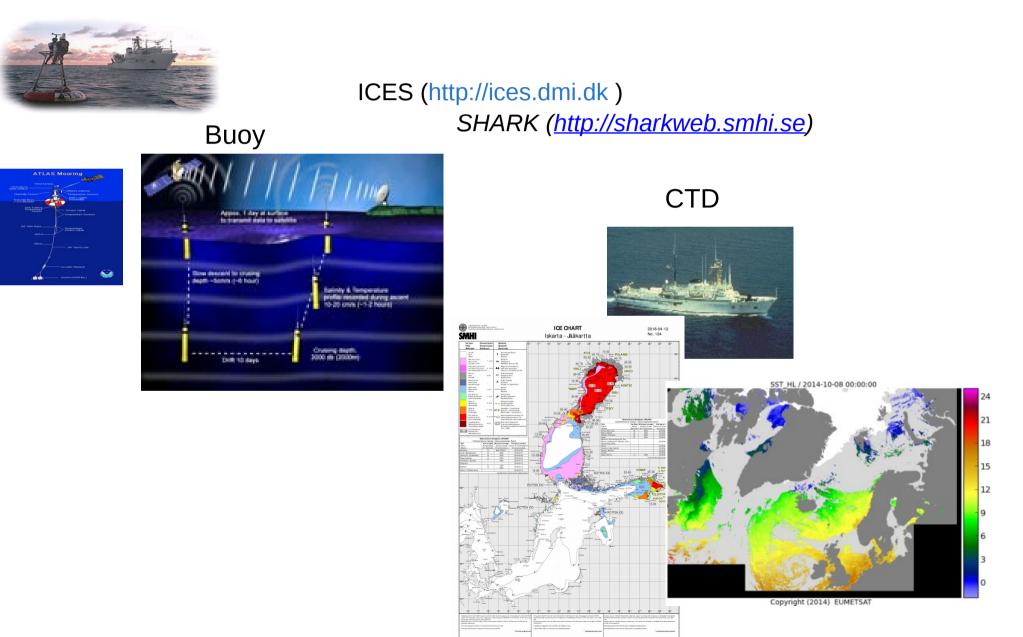
#### PROS & CONS

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..

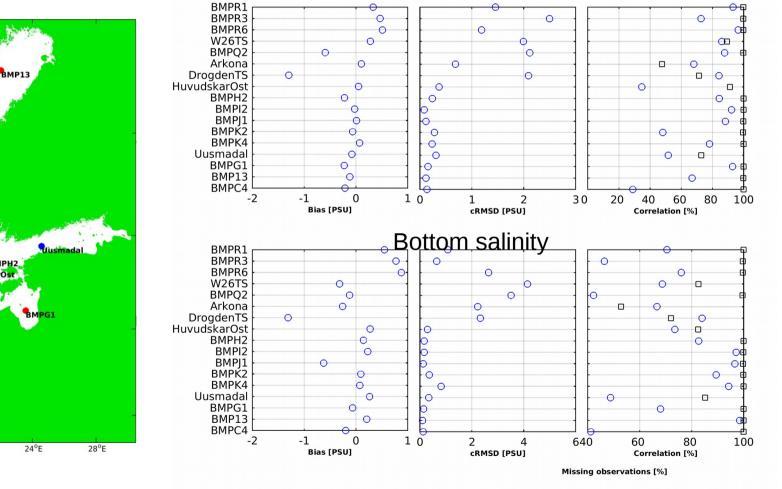
#### Observations



#### Moorings



# Surface and bottom salinity (V4) VS. independent observation



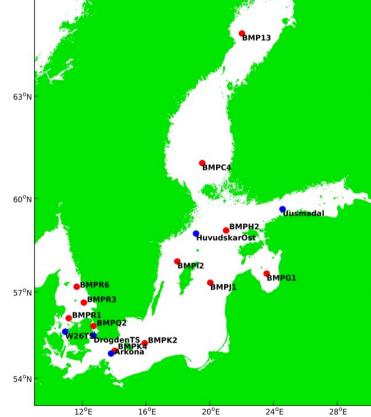
#### Surface salinity

CMEMS BALTIC SEA MFC

Tract 70-CMEMS MFC BAL

Until March 31 2021

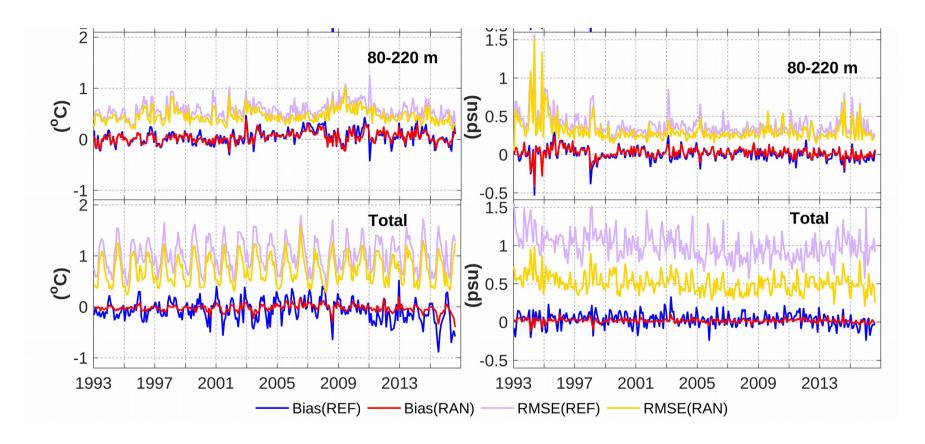
missing OBS O V4NEMO



66°N

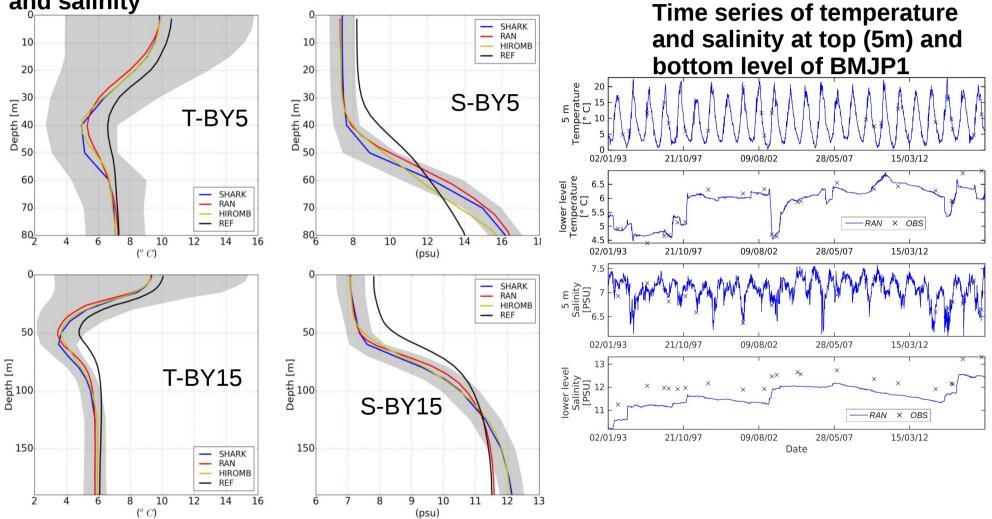


# Monthly-averaged bias and RMSE of temperature (left) and salinity (right).

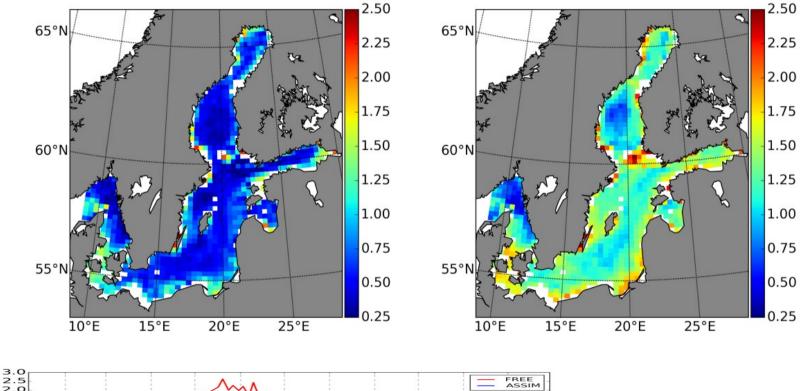




### Average (1993-2016) vertical profiles for temperature and salinity



#### Map of the RMSE of SST from ASSIM run (left ) and FREE run (right ) calculated against IceMap SST in 2010, respectively.





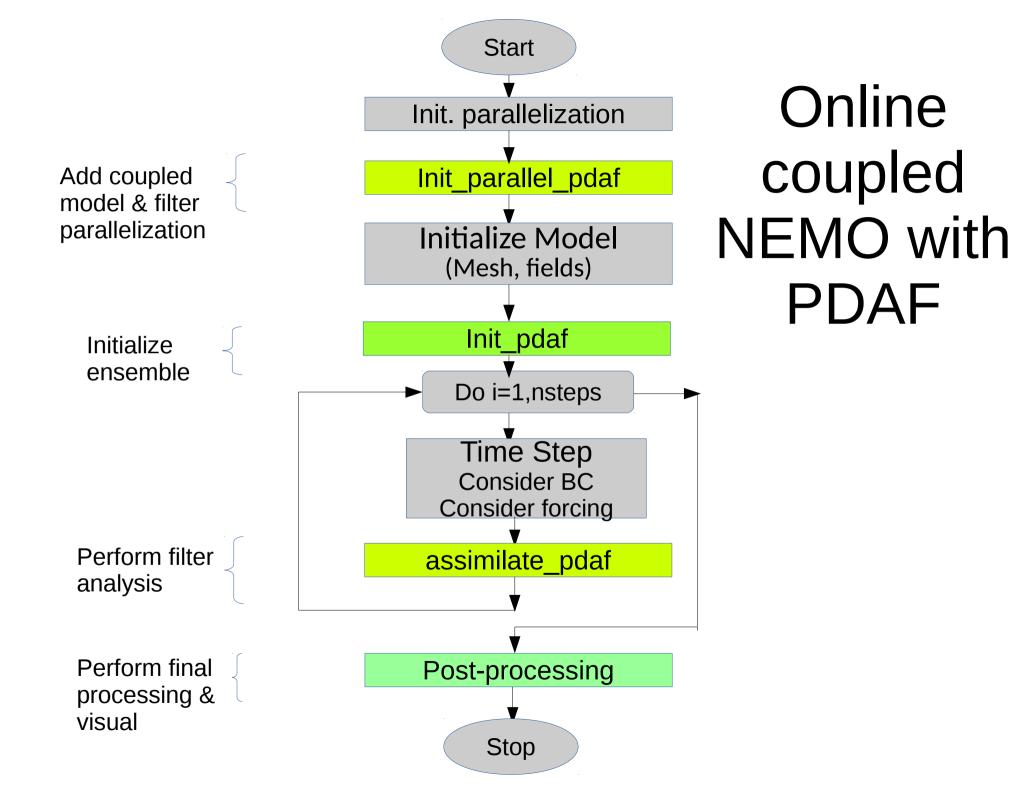
The evolution of the simulation SST bias, error and number of IceMap SST in 2010.

## PDAF: Cooperation

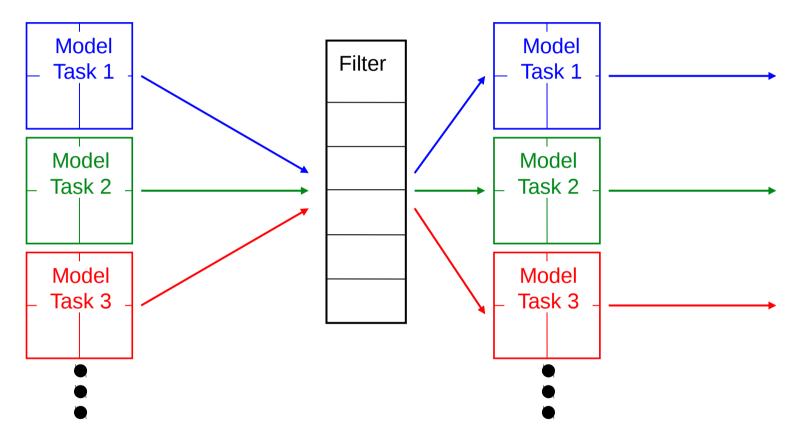
- Comparison on-line vs. off-line versions; different?
- Comparison of different ensemble filters:
  - LSEIK Filter
  - LESTK Filter
  - ...
- Development of ensemble generation techniques
- Technical implementation for NEMO-Nordic?
- Test implementations for sea ice, SLA, chlorophyll-a, ...
- Regular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

### The end...

### About on-line version of PDAF



# 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