Ocean assimilation algorithm developments, with a focus on ensemble DA in NEMOVAR

A. Weaver[1]
M. Chrust[2], B. Ménétrier[3], Y. Yang[4], H. Zuo[2]

[1]CERFACS
[2]ECMWF
[3]Météo-France
[4]ENSEEIHT/CERFACS

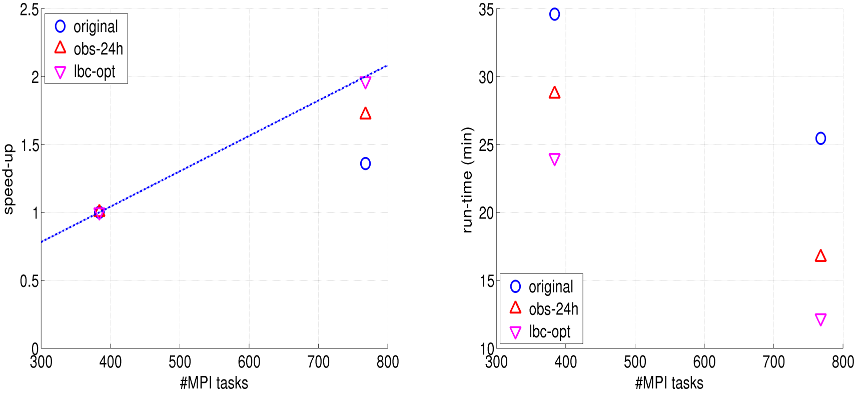
January 23, 2017

Outline

T2.3: *To improve the ocean analysis component using ensembles*
D2.3: *Ensemble-based covariance estimates*.
(Code and documented results; Month 46)

- NEMOVAR has been largely rewritten to facilitate the use of ensembles in defining the background-error covariance matrix (**B**).
- The diffusion-based correlation operator has been completely revised to make it more general and to improve its scalability (Weaver *et al.* 2016, QJRMS).
- The new code has been merged into the trunk of the central NEMOVAR repository at ECMWF.
  - Source code management under Git.
  - Atlassian tools for collaborative software development (JIRA, Bitbucket).
- The new code has been integrated into the prepIFS/SMS running environment at ECMWF (M. Chrust, H. Zuo).
  - Multi-annual reanalysis experiments (with ensembles) and comparison with "Old" NEMOVAR now possible.

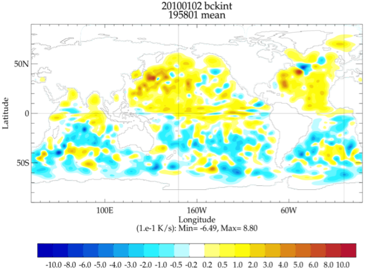
1/4° global ocean model

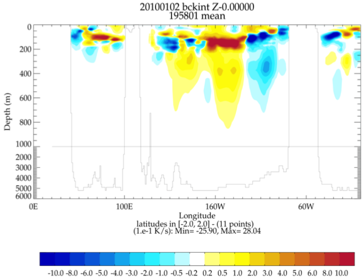


(Courtesy M. Chrust, ECMWF)

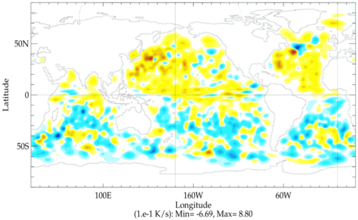
Mean temperature from 6-month experiment from 01/02/2010

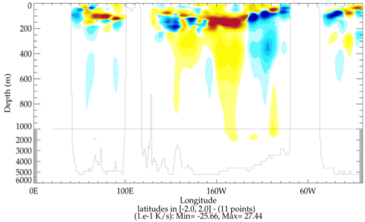
New





Old



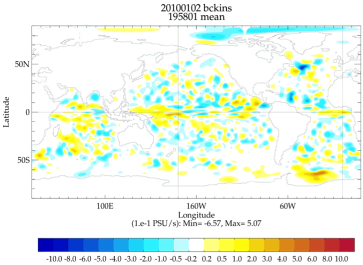


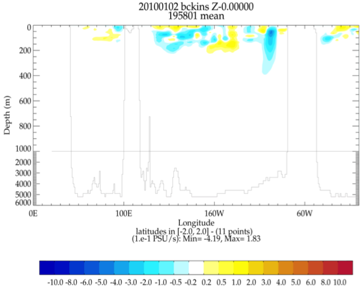
Old vs New NEMOVAR with "same" parameter settings

CERFACS

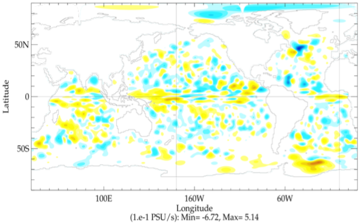
Mean salinity from 6-month experiment from 01/02/2010

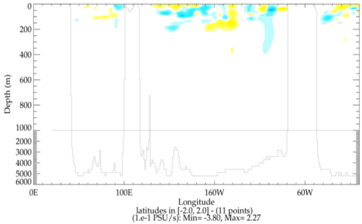
New





Old



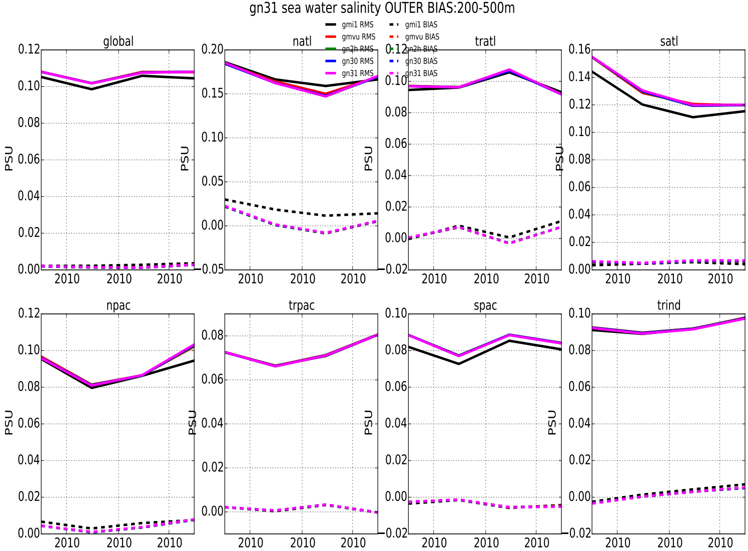


Old vs New NEMOVAR with "same" parameter settings

CERFACS

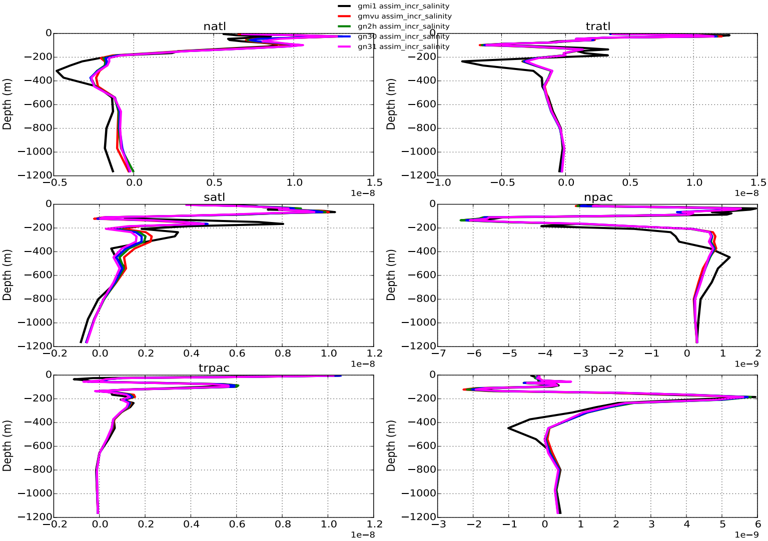
Mean and RMS fit to salinity observations for 6-month expt. from 01/02/2010

Old New



Salinity increment profile on 01/02/2010

Old New



- The covariance model has the form

$$\mathbf{B}_{\rm m} \;=\; \mathbf{K}_{\rm b}\,\mathbf{D}_{\rm m}^{1/2}\,\mathbf{C}_{\rm m}\,\mathbf{D}_{\rm m}^{1/2}\,\mathbf{K}_{\rm b}^{\rm T}$$

- Ensembles are used to estimate the variances ($\mathbf{D}_{\rm m} \rightarrow \mathbf{D}_{\rm e}$) and the local correlation tensor ($\boldsymbol{\kappa}_{\rm m} \rightarrow \boldsymbol{\kappa}_{\rm e}$) associated with the diffusion operator in $\mathbf{C}_{\rm m}$.
- The estimates are filtered using a diffusion operator with an optimally-based algorithm to determine the filtering scale (Ménétrier *et al.* 2015).
- A hybrid parameter formulation has also been developed:

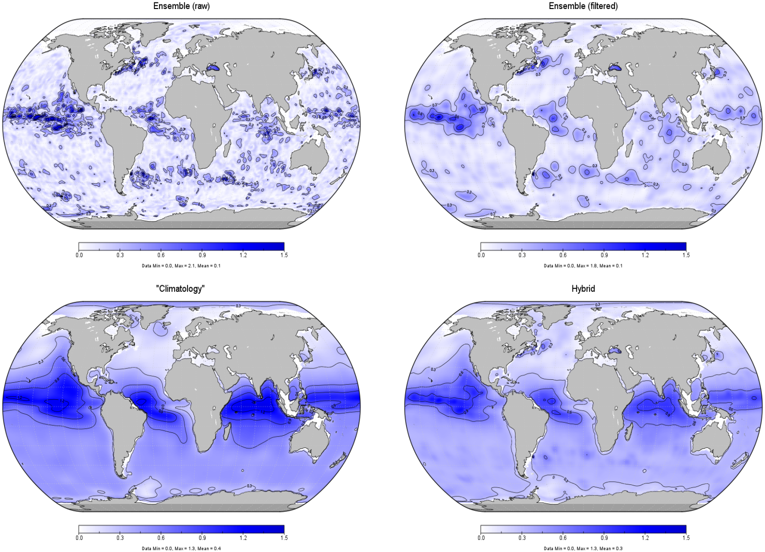
$$\mathbf{D} \;=\; \alpha_{\rm m}^2\,\mathbf{D}_{\rm m} \;+\; \alpha_{\rm e}^2\,\mathbf{D}_{\rm e}$$
$$\boldsymbol{\kappa} \;=\; \gamma_{\rm m}^2\,\boldsymbol{\kappa}_{\rm m} \;+\; \gamma_{\rm e}^2\,\boldsymbol{\kappa}_{\rm e}$$

where $\mathbf{D}_{\rm m}$ and $\boldsymbol{\kappa}_{\rm m}$ are modelled ("climatological") estimates, and $\alpha_{\rm m,e}$ and $\gamma_{\rm m,e}$ are constant weights.

Ensemble variances using the ECMWF ocean reanalysis

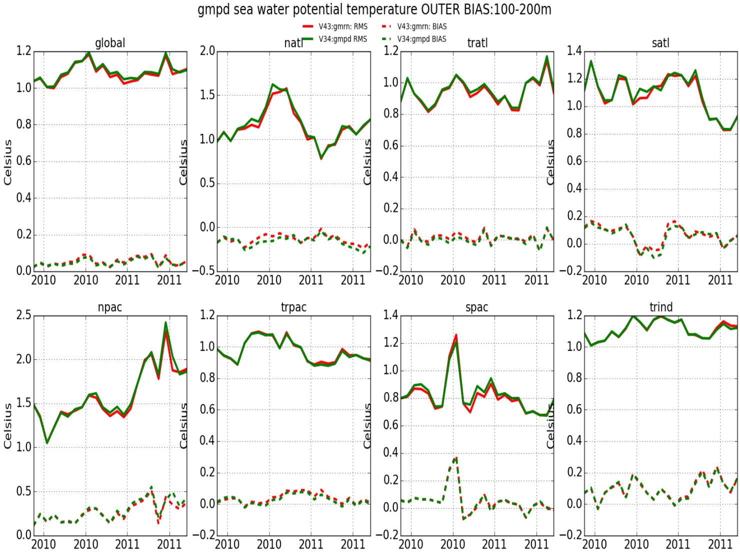
CERFACS

5-member ensemble (4 perturbed + 1 unperturbed) from 31/05/2015.
Background temperature error standard deviations at 100 m.



Mean temperature bias (100-200m) from 2-year experiment, 20 members (first expt, no tuning, equal weights to each hybrid component).

Modelled Hybrid

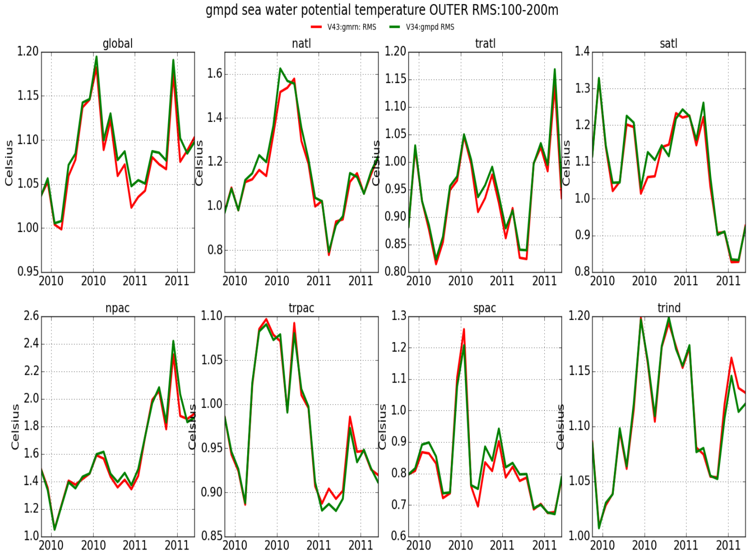


Modelled ("climatological") versus hybrid variances

CERFACS

Mean temperature RMS fit (100-200m) from 2-year experiment, 20 members

Modelled Hybrid



Mean salinity bias and RMS fit (100-200m) from 2-year experiment, 20 members

Modelled Hybrid

Outline

CERFACS

1. "New" versus "Old" NEMOVAR
2. Preliminary experiments with ensemble variances
3. Developments towards a fully hybrid **B**
4. Plans for the final year

- Hybrid **B** has the form

$$\mathbf{B} \;=\; \beta_{\mathrm{m}}^2\,\mathbf{B}_{\mathrm{m}} \;+\; \beta_{\mathrm{e}}^2\,\mathbf{B}_{\mathrm{e}} \quad \left(\; +\beta_{\mathrm{E}}^2\,\mathbf{B}_{\mathrm{EOF}}\right)$$

where β_{m}^2 and β_{e}^2 are constant weights.

- $\mathbf{B}_{\mathrm{e}}$ is a localized sample estimate of the covariance matrix:

$$\mathbf{B}_{\mathrm{e}} \;=\; \mathbf{K}_{\mathrm{b}}\,\mathbf{D}_{\mathrm{e}}^{1/2}\left(\mathbf{L}\circ\widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^{\mathrm{T}}\right)\mathbf{D}_{\mathrm{e}}^{1/2}\,\mathbf{K}_{\mathrm{b}}^{\mathrm{T}}$$

where the columns of $\widetilde{\mathbf{X}} = \mathbf{D}_{\mathrm{e}}^{-1/2}\,\mathbf{K}_{\mathrm{b}}^{-1}\,\mathbf{X}^{\mathrm{b}}$ are transformed background ensemble perturbations.

- The practical form of the Schur product term used in NEMOVAR is

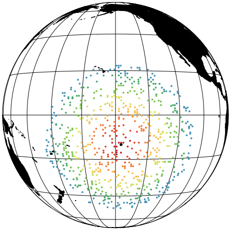
$$\left(\mathbf{L}\circ\widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^{\mathrm{T}}\right)\mathbf{v} = \sum_{p=1}^{N_{\mathrm{e}}}\left(\widetilde{\mathbf{x}}_p\circ\mathbf{L}\big(\widetilde{\mathbf{x}}_p\circ\mathbf{v}\big)\right) \qquad \text{where} \qquad \widetilde{\mathbf{X}} = \left(\widetilde{\mathbf{x}}_1,\ldots,\widetilde{\mathbf{x}}_{N_{\mathrm{e}}}\right)$$

where **L** is represented by a diffusion operator.

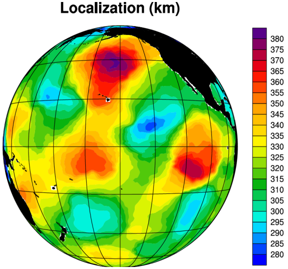
Localization modelling

- Four formulations of $\mathbf{L}$ have been implemented in NEMOVAR:
  1. No localization:
  $$\mathbf{L} = \begin{pmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{1} & \cdots & \mathbf{1} \end{pmatrix}$$
  2. Univariate and separate localization for each of the M variables:
  $$\mathbf{L} = \mathrm{diag}\left(\mathbf{L}_1, \ \ldots, \ \mathbf{L}_M\right)$$
  3. Multivariate and common localization for each variable:
  $$\mathbf{L} = \begin{pmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{pmatrix} \mathbf{L}_1 \begin{pmatrix} \mathbf{1} & \cdots & \mathbf{1} \end{pmatrix}$$
  4. Multivariate and separate localization for each of the M variables:
  $$\mathbf{L} = \begin{pmatrix} \mathbf{L}_1^{1/2} \\ \vdots \\ \mathbf{L}_M^{1/2} \end{pmatrix} \begin{pmatrix} \mathbf{L}_1^{\mathrm{T}/2} & \cdots & \mathbf{L}_M^{\mathrm{T}/2} \end{pmatrix}$$

- The hybridization weights and localization functions are determined using an optimally-based algorithm (Ménétrier and Auligné 2015; B. Ménétrier, unpublished research).
- The algorithm is applied offline; it requires the evaluation of statistical moments given as input the ensemble perturbations and randomized vectors that sample the modelled $\mathbf{B}_{\mathrm{m}}$.
- The algorithm has been interfaced with NEMOVAR (Y. Yang).

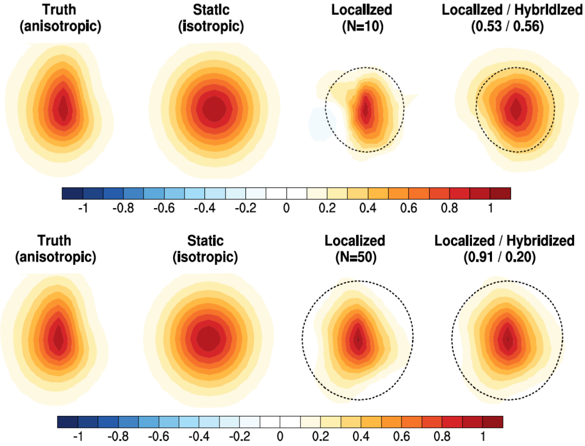


At each grid point, sample using the nearest neighbours at increasing distances.



The procedure can be used to estimate a spatial map of localization scales.

Example of surface T-T correlations at a point in the North Atlantic using two different ensembles sizes (10 and 50)



Summary and plans

- Major code developments are complete and have been integrated into the ECMWF system.
- Need to resolve the discrepancy between Old and New NEMOVAR (with similar parameter settings).
- First results (no tuning) with ensemble/hybrid variances are encouraging but more work is needed.
  - Estimation/tuning of parameters (inflation factor, hybridization weights).
  - Ensemble-estimation and filtering of the local correlation tensor.
- In parallel, continue evaluating the fully hybrid **B**.
- Further improvements to computational aspects of the diffusion operator.
  - Implicit solver.
  - Normalization factors.
  - Coarse grid for "large" scales.
- Several visits planned to ECMWF in 2017 to work on ensemble DA.