Ensemble **B** in NEMOVAR

A. Weaver[1]
M. Chrust[2], B. Ménétrier[3], A. Piacentini[1], J. Tshimanga[1], Y. Yang[4], S. Gürol[1] and H. Zuo[2]

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

December 12, 2017

PROJECT DELIVERABLE

Work Package 2: Future coupling methods

Deliverable 2.3: Using ensemble-estimated background error variances and correlation scales in the NEMOVAR system

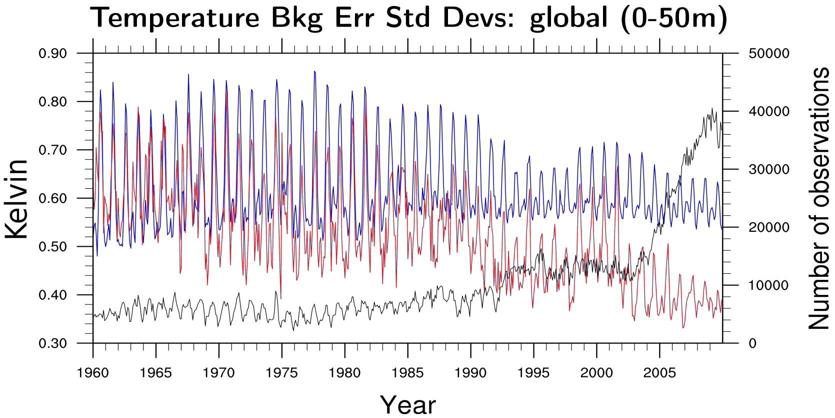
Type: Documented code and test results

Author(s): Anthony Weaver (CERFACS), Marcin Chrust (ECMWF), Benjamin Ménétrier (Météo-France), Andrea Piacentini (CERFACS), Jean Tshimanga (CERFACS), Yin Yang (IRIT/CERFACS), Selime Gürol (CERFACS) and Hao Zuo (ECMWF)

Reviewer: Matthew Martin (WP2 leader)

Delivered: 30/11/2017

Outline



- There are discrepencies between the **specified** and **expected** background-error standard deviations (**blue** and **red** curves, resp.).
- The specified **B** (same one used in CERA) does not “see” the observing system (**black curve**).
- Improved estimation and modelling methods for **B** are needed.

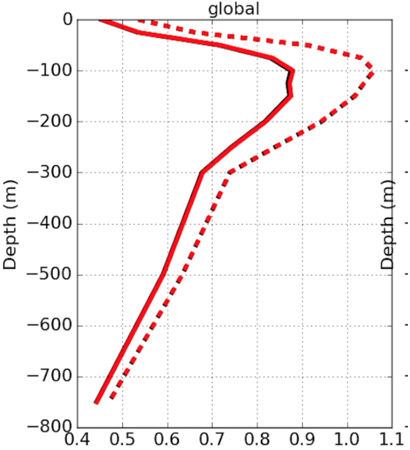
1. Motivation
2. Summary of developments
3. Using ensembles to estimate **B**
4. Improving **B** for future reanalyses

- Two methods have been developed to use ensembles to define **B**.
  1. Estimate parameters of the modelled covariance matrix.
  2. Compute a low-rank sample estimate of the covariance matrix and localize it (EnVar).
- Hybrid formulations of both 1 and 2 have also been developed.
- 1 and 2 include optimally-based algorithms for filtering and estimating parameters.
- All methods have been included in a new version of NEMOVAR (v5) in the central source code repository at ECMWF.
- The new version has been integrated into the ECMWF scripting environment for running reanalysis experiments.
- Validation experiments have been performed using both ORCA1 and ORCA025 global configurations.

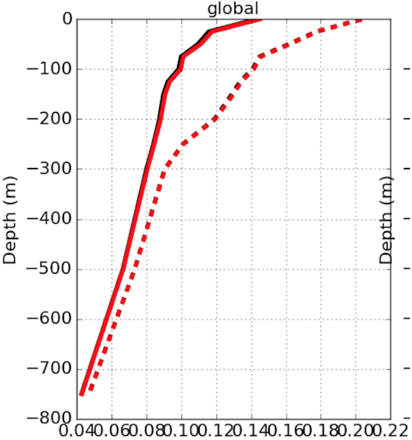
RMS of OmA (solid) and OmB (dashed) from 6-month experiment with ORCA1

Old (v3.4) and New (v5) NEMOVAR with "same" parameters

Temperature



Salinity



CERFACS

$$\mathbf{B} \;=\; \underbrace{\left(\mathbf{B}_{\mathrm{m}_1} + \mathbf{B}_{\mathrm{m}_2} + \ldots\right)}_{\mathbf{B}_{\mathrm{m}}} + \mathbf{B}_{\mathrm{e}} + \mathbf{B}_{\mathrm{EOF}}$$

(where the terms are weighted by diagonal matrices $\boldsymbol{\Upsilon}_{\mathrm{m}}$, $\boldsymbol{\Upsilon}_{\mathrm{e}}$ and $\boldsymbol{\Upsilon}_{\mathrm{EOF}}$)

- Multiple covariance models for representing different scales:

$$\mathbf{B}_{\mathrm{m}_i} \;=\; \mathbf{K}_{\mathrm{b}}\, \boldsymbol{\Upsilon}_{\mathrm{m}}^{1/2} \left(\mathbf{D}_{\mathrm{m}_i}^{1/2}\, \mathbf{C}_{\mathrm{m}_i}\, \mathbf{D}_{\mathrm{m}_i}^{1/2} \right) \boldsymbol{\Upsilon}_{\mathrm{m}}^{1/2}\, \mathbf{K}_{\mathrm{b}}^{\mathrm{T}}$$

- A localized ensemble-based covariance matrix:

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

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

- A large-scale EOF-based covariance matrix for assimilating sparse observations (D. Lea; part of Deliverable 2.1):

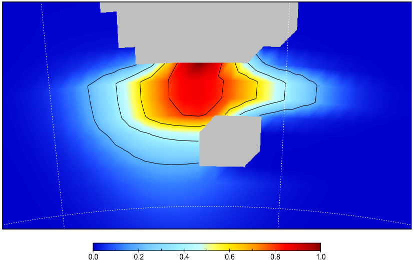
$$\mathbf{B}_{\mathrm{EOF}} \;=\; \boldsymbol{\Upsilon}_{\mathrm{EOF}}^{1/2}\, \mathbf{P}\, \boldsymbol{\Lambda}\, \mathbf{P}^{\mathrm{T}}\, \boldsymbol{\Upsilon}_{\mathrm{EOF}}^{1/2}$$

A revised correlation operator

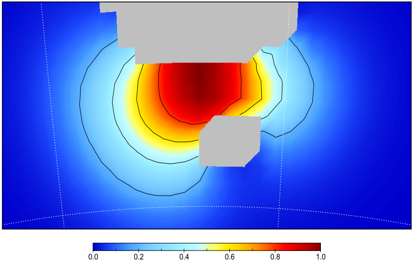
CERFACS

- The correlation operator, localization operator and parameter filter are based on an algorithm that involves solving an implicitly formulated diffusion equation.
- The diffusion model has been completely revised to make it more general, to eliminate numerical artefacts near complex boundaries, and to improve computational efficiency and scalability on high-performance computers (Weaver *et al.*, 2016; 2017).

Old (v3.4)



New (v5)



- Ensemble perturbations are used to estimate the **variance** matrix (**D**), as well as the **diffusion tensor** ($\boldsymbol{\kappa}$) associated with the diffusion operator in **C**.
- To remove sampling error with small ensemble sizes, the raw estimates are filtered using a diffusion operator with an optimally-based algorithm to determine the filtering scale (Ménétrier *et al.* 2015; Michel *et al.* 2016).
- A hybrid parameter formulation has also been developed:

$$\mathbf{D}_{\rm h} = \alpha_{\rm m}^2 \, \mathbf{D}_{\rm m} + \alpha_{\rm e}^2 \, \mathbf{D}_{\rm e}$$
$$\boldsymbol{\kappa}_{\rm h}^{-1} = \gamma_{\rm m}^2 \, \boldsymbol{\kappa}_{\rm m}^{-1} + \gamma_{\rm e}^2 \, \boldsymbol{\kappa}_{\rm e}^{-1}$$

  where $\mathbf{D}_{\rm m}$ (resp., $\mathbf{D}_{\rm e}$) and $\boldsymbol{\kappa}_{\rm m}$ (resp., $\boldsymbol{\kappa}_{\rm e}$) are modelled (resp., ensemble) estimates, and $\alpha_{\rm m,e}^2$ and $\gamma_{\rm m,e}^2$ are constant weights.
- The spatially averaged hybrid parameters are fixed to a reference value (either the spatially averaged modelled or spatially averaged ensemble values).

- The optimality criteria of Ménétrier *et al.* (2015) translate into the following requirements on the design of the spatial filter:

$$\left.\begin{array}{rcl} \mu^{\mathrm{S}}[\widehat{\mathbf{v}}] & = & \mu^{\mathrm{S}}[\widetilde{\mathbf{v}}] \\ \mathcal{C} & = & 0 \end{array}\right\}$$

where $\widetilde{\mathbf{v}}$ and $\widehat{\mathbf{v}}$ are the raw and filtered variances, and μ^{S} is spatial average.

- For Gaussian (G) statistics:

$$\mathcal{C}^{\mathrm{G}} = \mu^{\mathrm{S}}[\widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}}] - \left(\frac{N_{\mathrm{e}}+1}{N_{\mathrm{e}}-1}\right)\mu^{\mathrm{S}}[\widetilde{\mathbf{v}} \circ \widehat{\mathbf{v}}].$$

- For Non-Gaussian (NG) statistics:

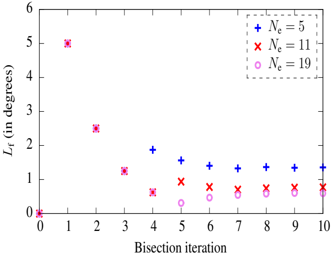
$$\begin{aligned} \mathcal{C}^{\mathrm{NG}} &= \mu^{\mathrm{S}}[\widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}}] - \frac{N_{\mathrm{e}}(N_{\mathrm{e}}-2)(N_{\mathrm{e}}-3)}{(N_{\mathrm{e}}-1)(N_{\mathrm{e}}^2-3N_{\mathrm{e}}+3)}\mu^{\mathrm{S}}[\widetilde{\mathbf{v}} \circ \widehat{\mathbf{v}}] \\ &\quad - \frac{N_{\mathrm{e}}^2}{(N_{\mathrm{e}}-1)(N_{\mathrm{e}}^2-3N_{\mathrm{e}}+3)}\mu^{\mathrm{S}}[\widetilde{\boldsymbol{\xi}}] \end{aligned}$$

where $\widetilde{\boldsymbol{\xi}}$ is the raw fourth-order moment.

Optimal spatial filtering

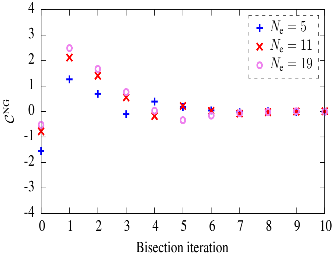
CERFACS

Filter scale

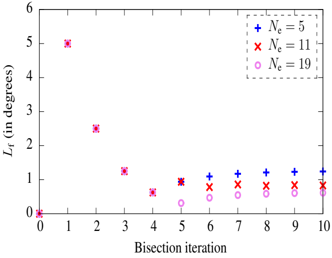


(a) Non-Gaussian

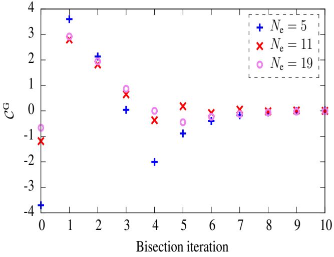
Optimality function



(b) Non-Gaussian

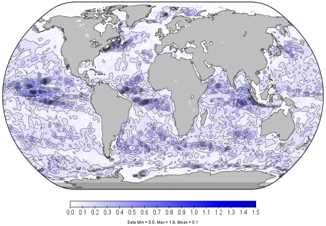


(c) Gaussian

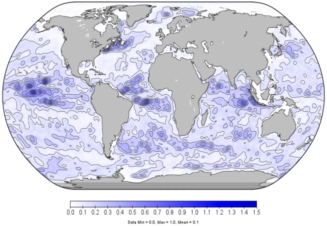


(d) Gaussian

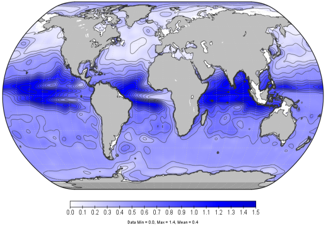
11-member ensemble (10 perturbed + 1 unperturbed) from 04/06/2011.
Background T error standard deviations at 100 m.



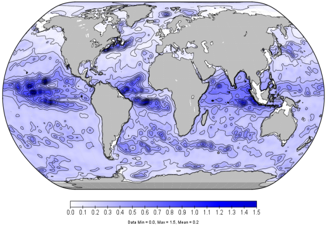
(a) Raw



(b) Filtered



(c) Parameterized



(d) Hybrid

- $\boldsymbol{\kappa}^{-1}$ can be related to the local correlation tensor (LCT) $\boldsymbol{H}$ (Chorti and Hristopulos 2008; Weaver and Mirouze 2013):

$$\left(\frac{1}{2M-d-2}\right)\boldsymbol{\kappa}^{-1} = -\nabla\nabla^{\mathrm{T}} c_d\big|_{r=0} = \boldsymbol{H}$$

where c_d is the analytical form of the (Matérn) correlation function in $\mathbb{R}^d$.

- $\boldsymbol{H}$ can be approximated locally using ensemble perturbations $\epsilon(\mathbf{z})$ (Michel 2013; Michel *et al.* 2016):

$$\widetilde{\boldsymbol{H}}(\mathbf{z}) = \overline{\nabla\widetilde{\epsilon}(\mathbf{z})\left(\nabla\widetilde{\epsilon}(\mathbf{z})\right)^{\mathrm{T}}} \text{ where } \widetilde{\epsilon}(\mathbf{z}) = \epsilon(\mathbf{z})/\sigma(\mathbf{z})$$

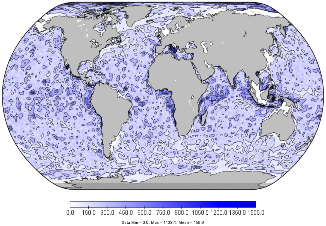
- $\widetilde{\boldsymbol{H}}^{-1}$ is a tensor of "squared length-scales":

$$\widetilde{\boldsymbol{H}}^{-1} = \boldsymbol{R}\,\boldsymbol{L}\,\boldsymbol{R}^{\mathrm{T}}$$

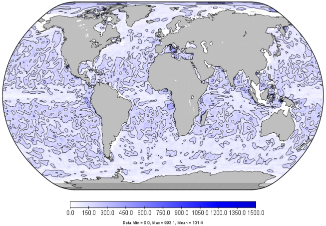
where $\boldsymbol{R}$ a rotation matrix, and $\boldsymbol{L} = \mathrm{diag}\left(L_1^2, \ldots, L_d^2\right)$ where L_1 etc are the length-scales along the principal axes.

Estimated horizontal correlation scales

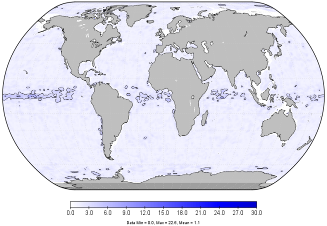
Background T error scales at 100 m



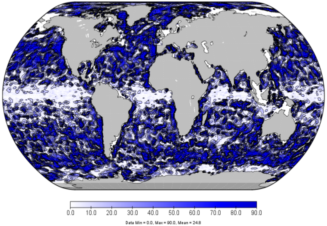
(a) L_1



(b) L_2

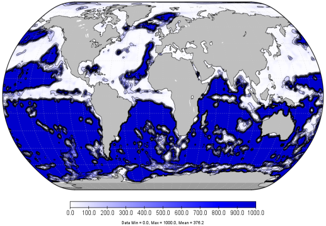


(c) L_1/L_2

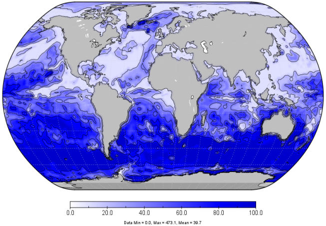


(d) R_{12}

Background T error vertical scales at 5 m



(a) L_3



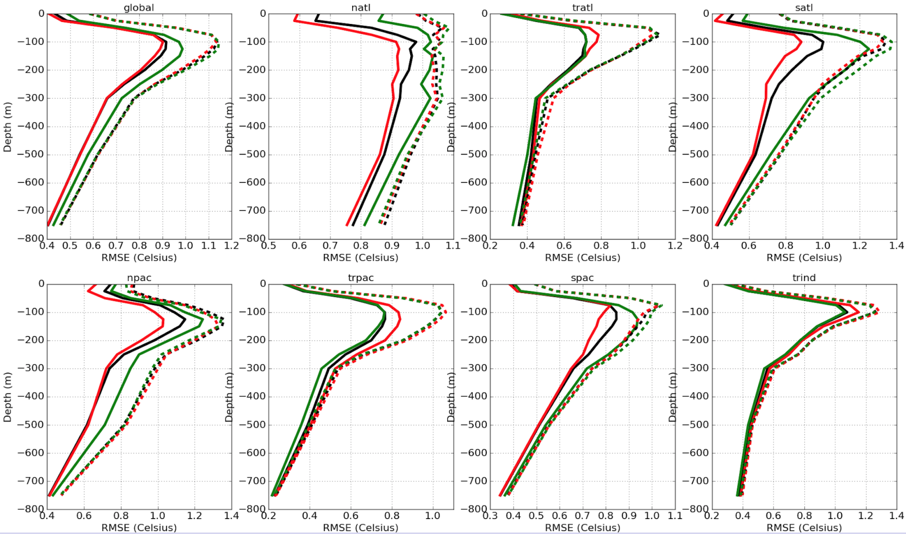
(b) Background MLD depth

- Parameterizing the vertical length-scales in terms of the mixed-layer depth makes sense.
- Such a parameterization is available in NEMOVAR but has not been used at ECMWF.

Impact of ensemble and hybrid variances

RMS of OmA (solid) and OmB (dashed) for **temperature** from 6-month experiment in ORCA025

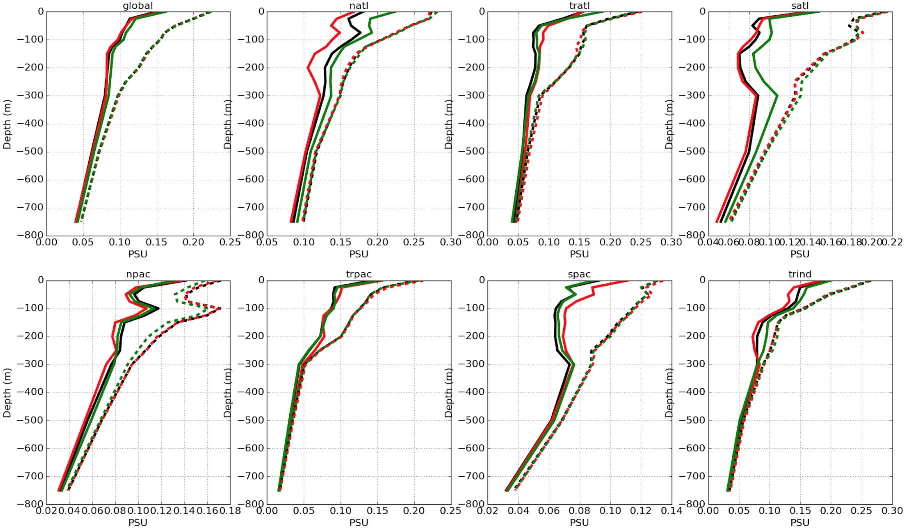
ensemble-only (15 members); **hybrid** (0.5, 0.5); **parameterized-only**



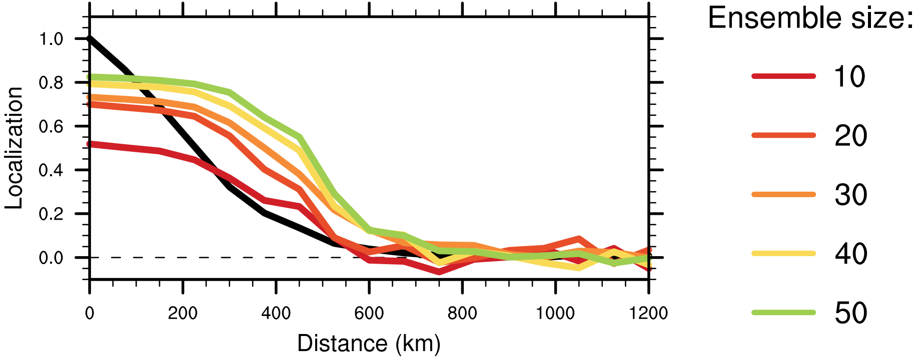
Impact of ensemble and hybrid variances

RMS of OmA (solid) and OmB (dashed) for **salinity**
from 6-month experiment in ORCA025

ensemble-only (15 members); **hybrid** (0.5, 0.5); **parameterized-only**

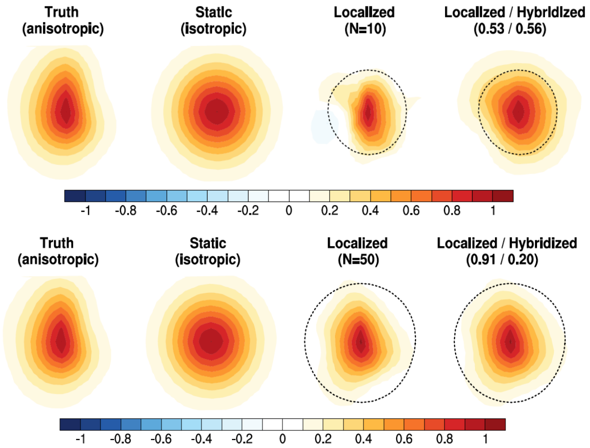


- Different formulations of the localization operator have been developed.
- Localization functions and hybrid weights are estimated simultaneously using the hdiag_nicas software (B. Ménétrier).



Correlation (black) and localization (colors)

Example of T-T correlations using two different ensembles sizes (10 and 50)



- Several options have been developed for defining **B** from ensembles.
- Computational cost is a determining factor.
- A three-step strategy for progressively improving **B**:
  1. *Flow-dependent ensemble-derived variances*
     - ⋆ Low cost-overhead; ensemble-only or hybrid; tuning inflation factor and hybrid weights; improving ensemble-generation procedure
     - ⋆ Include sea-ice concentration
  2. *Climatological ensemble-derived correlation scales*
     - ⋆ Ensemble horizontal scales combined with flow-dependent MLD-dependent vertical scales; efficient normalization; ensemble gradient-method or function-fitting approach (hybrid_nicas)
  3. *Hybrid EnVar*
     - ⋆ Climatological ensemble-derived $\mathbf{B}_{\mathrm{m}}$; efficient localization in $\mathbf{B}_{\mathrm{e}}$ (hybrid_nicas); opens the way to coupled (cross-domain) covariances.
- Continued efforts are needed to improve the computational efficiency and scalability of the diffusion operator.

Results obtained with the hdiag_nicas software (B. Ménétrier)