Ensemble-variational assimilation with NEMOVAR: Part 1: design aspects and illustrations

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

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

June 19, 2017

A few words about NEMOVAR

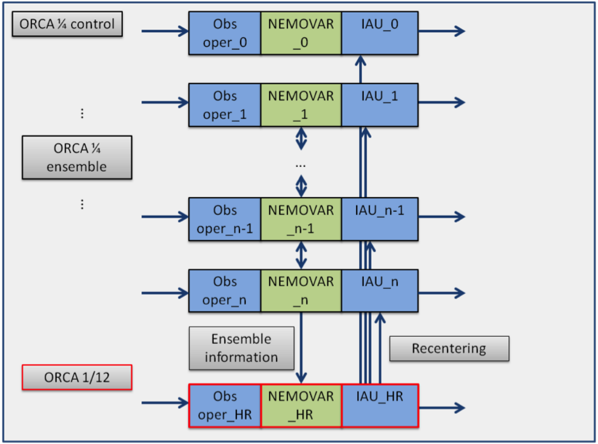
CERFACS

- A joint project by CERFACS, ECMWF, Met Office and INRIA to develop an ocean data assimilation system for the NEMO[1] model, with a variational kernel.
- Two operational centres (ECMWF, Met Office) and two research institutes (CERFACS, INRIA).
- Separate from the (much larger) NEMO consortium which oversees the development of the ocean model.
- Versions of NEMOVAR are currently operational at ECMWF and Met Office.
- Applications cover ocean forecasting, medium-range weather forecasting, monthly-to-seasonal climate forecasting, coupled and uncoupled reanalysis.
- The purpose of this talk is to outline work done in AVENUE and related projects on using ensembles to specify **B** in NEMOVAR.

[1] Nucleus for European Modelling of the Ocean.

CERFACS

- **Ensemble of Data Assimilations** framework for generating forecast and analysis ensembles.



(from the DANGOS proposal - Met Office, CERFACS, CMCC, ECMWF, INRIA)

- **Determinstic EVIL** (Aulligné *et al.* 2016) has also been developed for NEMOVAR as a cheaper alternative to the EDA.
- Transform an ensemble of background perturbations ($\mathbf{X}^{\mathrm{b}}$) into an ensemble of analysis perturbations ($\mathbf{X}^{\mathrm{a}}$) using approximate eigenvector (Ritz) information ($\mathbf{\Theta}_q, \overline{\mathbf{Z}}_q$) generated by $\mathbf{B}$-preconditioned conjugate gradient algorithms:
  - $\mathbf{B}$-PCG (Derber and Rosati 1989):

$$\mathbf{X}^{\mathrm{a}} = \left\{ \mathbf{I}_n - \mathbf{B}\overline{\mathbf{Z}}_q \big(\mathbf{I}_q - \mathbf{\Theta}_q^{-\frac{1}{2}} \big) \overline{\mathbf{Z}}_q^{\mathrm{T}} \right\} \mathbf{X}^{\mathrm{b}}$$

  - Restricted $\mathbf{B}$-PCG (Gratton and Tshimanga 2009; Gürol *et al.* 2014):

$$\mathbf{X}^{\mathrm{a}} = \left\{ \mathbf{I}_n - \mathbf{B}\mathbf{H}^{\mathrm{T}}\overline{\mathbf{Z}}_q^{\mathrm{o}} \big(\mathbf{I}_q - \mathbf{\Theta}_q^{-\frac{1}{2}} \big) \overline{\mathbf{Z}}_q^{\mathrm{o}\,\mathrm{T}} \mathbf{H} \right\} \mathbf{X}^{\mathrm{b}}$$

  where $\mathbf{H}^{\mathrm{T}}\overline{\mathbf{Z}}_q^{\mathrm{o}} = \overline{\mathbf{Z}}_q$.
- How many iterations q required to get a good estimate of $\mathbf{P}^{\mathrm{a}} = \mathbf{X}^{\mathrm{a}}\mathbf{X}^{\mathrm{a}\,\mathrm{T}}$?

The NEMOVAR B formulation

- The NEMOVAR **B** formulation is quite general:

$$\mathbf{B} \;=\; \beta_{\rm m}^2 \, \underbrace{(\mathbf{B}_{{\rm m}_1} + \mathbf{B}_{{\rm m}_2} + \ldots)}_{\mathbf{B}_{\rm m}} + \beta_{\rm e}^2 \, \mathbf{B}_{\rm e} + \beta_{\rm E}^2 \, \mathbf{B}_{\rm EOF}$$

  wnere $\beta_{\rm m}^2$, $\beta_{\rm e}^2$ and $\beta_{\rm E}^2$ are constant weights or switches.

- Multiple covariance models for representing different scales (Mirouze *et al.* 2016):

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

- A localized ensemble-based covariance matrix:

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

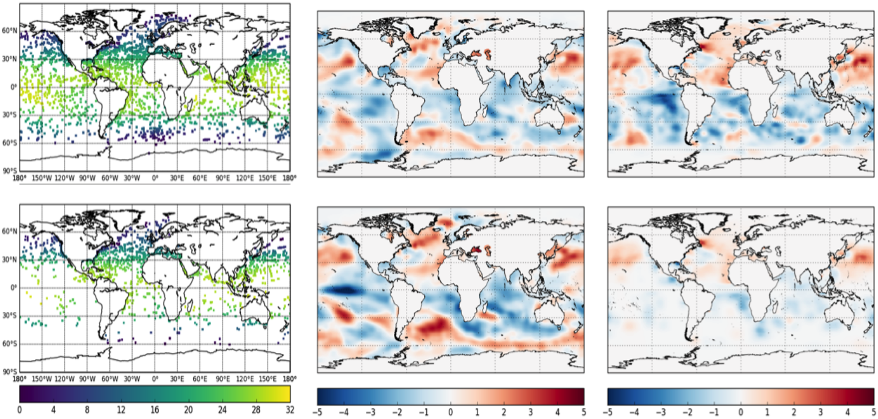
  where the columns of $\widetilde{\mathbf{X}} = \mathbf{D}_{\rm e}^{-1/2} \, \mathbf{K}_{\rm b}^{-1} \, \mathbf{X}^{\rm b}$ are unbalanced, normalized background ensemble perturbations.

- A large-scale EOF-based covariance matrix for assimilating sparse observations (Met Office):

$$\mathbf{B}_{\rm EOF} \;=\; \mathbf{P} \, \mathbf{\Lambda} \, \mathbf{P}^{\rm T}$$

- Observations have a non-local impact when using EOF-based covariances.

Data coverage | EOF assimilation | Standard assimilation



(Courtesy D. Lea, Met Office)

- The covariance model has the form

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

- $\mathbf{C}\psi$ is the discretized representation of the linear (implicit diffusion) operator $C : \psi_{0} \rightarrow \psi_{M}$, for $\psi_{0}, \psi_{M} \in \mathbb{R}^{d}$, defined by the solution of

$$(1 - \nabla \cdot \boldsymbol{\kappa} \nabla)^{M}\, \gamma^{-1/2}\, \psi_{M} = \gamma^{1/2}\, \psi_{0}, \qquad (1)$$

where M is a positive integer, $\gamma^{1/2}$ is a normalization function, and $\boldsymbol{\kappa}$ is a $d \times d$ matrix (Weaver *et al.* 2016).

- For constant $\boldsymbol{\kappa}$, the kernel of the integral solution of (1) admits covariance functions from the **Matérn class** (Guttorp and Gneiting 2006):

$$c_{d}(r) \propto r^{M-d/2}\, K_{M-d/2}(r)$$

where $r = \sqrt{(\mathbf{z}-\mathbf{z}')^{\mathrm{T}}\, \boldsymbol{\kappa}^{-1}\, (\mathbf{z}-\mathbf{z}')}$ and $K_{M-d/2}$ is the Bessel function of the 2nd kind of order $M - d/2$.

- The covariance model has the form

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

- Ensembles are used to estimate the **variances** ($\mathbf{D} \rightarrow \mathbf{D}_{\mathrm{e}}$) and the **diffusion tensor** ($\boldsymbol{\kappa} \rightarrow \boldsymbol{\kappa}_{\mathrm{e}}$) associated with the diffusion operator in $\mathbf{C}$.

- 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:

$$\begin{aligned}\mathbf{D} \;&=\; \alpha_{\mathrm{m}}^{2}\,\mathbf{D}_{\mathrm{m}} \;+\; \alpha_{\mathrm{e}}^{2}\,\mathbf{D}_{\mathrm{e}}\\ \boldsymbol{\kappa} \;&=\; \gamma_{\mathrm{m}}^{2}\,\boldsymbol{\kappa}_{\mathrm{m}} \;+\; \gamma_{\mathrm{e}}^{2}\,\boldsymbol{\kappa}_{\mathrm{e}}\end{aligned}$$

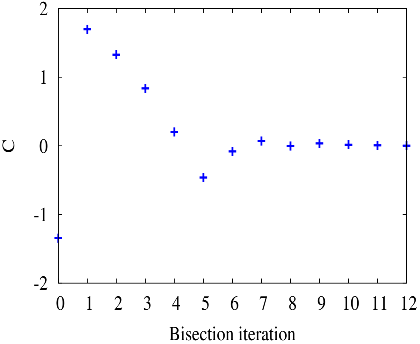
  where $\mathbf{D}_{\mathrm{m}}$ and $\boldsymbol{\kappa}_{\mathrm{m}}$ are modelled ("climatological") estimates, and $\alpha_{\mathrm{m,e}}$ and $\gamma_{\mathrm{m,e}}$ are constant weights.

- Find the scale (L_{opt}) that corresponds to the unique zero-crossing of an optimality function $C(L)$ (Ménétrier *et al.* 2015).
- For Gaussian statistics

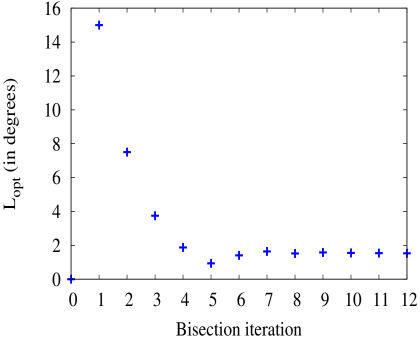
$$C(L) = E[\widetilde{\mathbf{v}}^2] - \frac{N_\mathrm{e}+1}{N_\mathrm{e}-1} E[\widetilde{\mathbf{v}}\,\widehat{\mathbf{v}}]$$

where $\widetilde{\mathbf{v}}$ is the raw variance, $\widehat{\mathbf{v}}$ the filtered variance, and N_e the ensemble size.

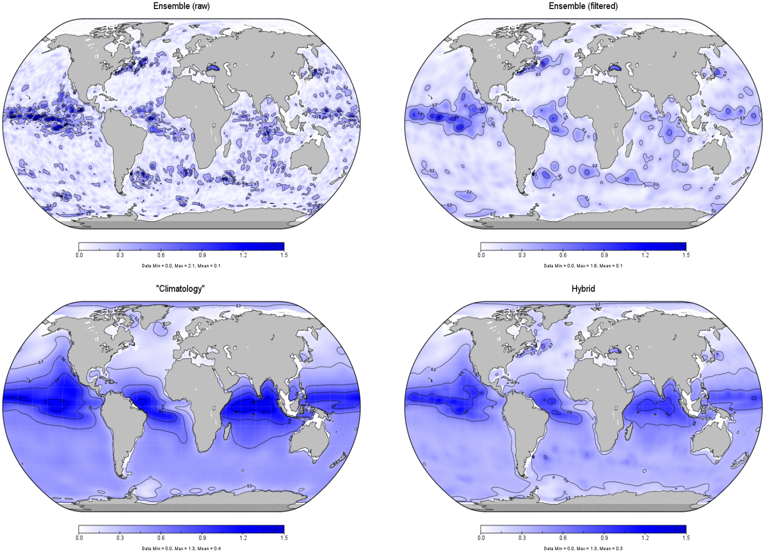
Optimality function



Filter scale



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



- $\boldsymbol{\kappa}^{-1}$ can be related to the local correlation Hessian tensor $\boldsymbol{H}$ (Hristopulos 2002; Weaver and Mirouze 2013):

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

- $\boldsymbol{H}$ can be approximated locally from sample statistics using the formulae:

1. Belo-Pereira and Berre (2006); Weaver and Mirouze (2013):

$$\widetilde{\boldsymbol{H}}(\mathbf{z}) = \frac{\left(\overline{\nabla\epsilon(\mathbf{z})\left(\nabla\epsilon(\mathbf{z})\right)^{\mathrm{T}}} - \nabla\sigma(\mathbf{z})\left(\nabla\sigma(\mathbf{z})\right)^{\mathrm{T}}\right)}{\left(\sigma(\mathbf{z})\right)^2} \quad \text{where } \left(\sigma(\mathbf{z})\right)^2 = \overline{\left(\epsilon(\mathbf{z})\right)^2}$$

2. Michel (2013); Michel *et al.* (2016):

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

3. Sato *et al.* (2009), Varella *et al.* (2011):

$$\widetilde{\boldsymbol{H}}(\mathbf{z}) \approx \frac{\overline{\nabla\widetilde{\epsilon}(\mathbf{z})\left(\nabla\widetilde{\epsilon}(\mathbf{z})\right)^{\mathrm{T}}}}{\left(\sigma(\mathbf{z})\right)^2} \quad \text{assuming } \sigma(\mathbf{z}) \text{ is slowy varying}$$

- $\boldsymbol{\kappa}^{-1}$ can be related to the local correlation Hessian tensor $\boldsymbol{H}$ (Hristopulos 2002; Weaver and Mirouze 2013):

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

- $\boldsymbol{H}$ can be approximated locally from sample statistics using the formulae:

1. Belo Pereira and Berre (2006); Weaver and Mirouze (2013):

$$\widetilde{\boldsymbol{H}}(\mathbf{z}) = \frac{\left(\overline{\nabla\epsilon(\mathbf{z})\left(\nabla\epsilon(\mathbf{z})\right)^{\mathrm{T}}} - \nabla\sigma(\mathbf{z})\left(\nabla\sigma(\mathbf{z})\right)^{\mathrm{T}}\right)}{\left(\sigma(\mathbf{z})\right)^2} \quad \text{where } \left(\sigma(\mathbf{z})\right)^2 = \overline{\left(\epsilon(\mathbf{z})\right)^2}$$

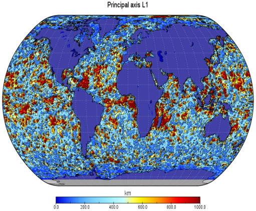
2. Michel (2013); Michel *et al.* (2016):

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

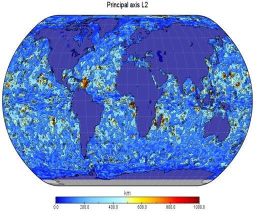
3. Sato *et al.* (2009), Varella *et al.* (2011):

$$\widetilde{\boldsymbol{H}}(\mathbf{z}) \approx \frac{\overline{\nabla\widetilde{\epsilon}(\mathbf{z})\left(\nabla\widetilde{\epsilon}(\mathbf{z})\right)^{\mathrm{T}}}}{\left(\sigma(\mathbf{z})\right)^2} \quad \text{assuming } \sigma(\mathbf{z}) \text{ is slowy varying}$$

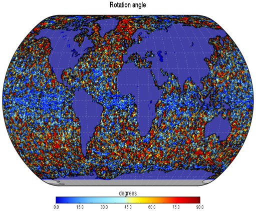
- Level 1 temperature; **5 members**; **raw estimates**



(a) Principal axis L1

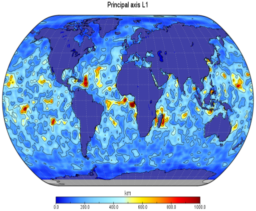


(b) Principal axis L2

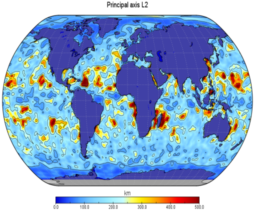


(c) Rotation angle

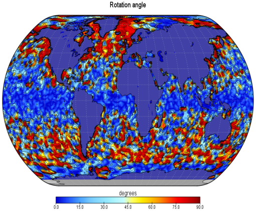
- Level 1 temperature; **5 members**; **filtered estimates** (Michel *et al.* 2016)



(d) Principal axis L1



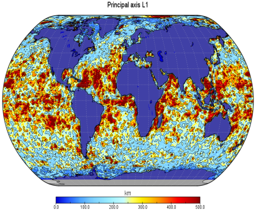
(e) Principal axis L2



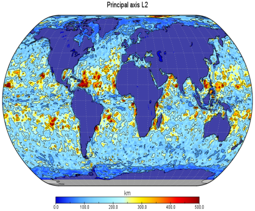
(f) Rotation angle

Estimating correlation scales from an ocean ensemble

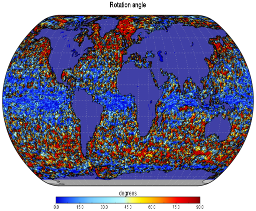
- Level 1 temperature; **19 members**; **raw estimates**



(g) Principal axis L1

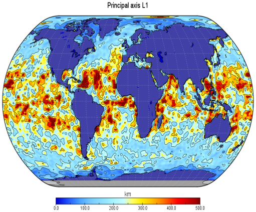


(h) Principal axis L2

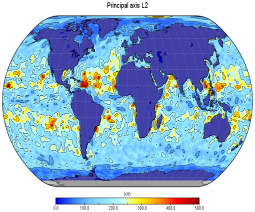


(i) Rotation angle

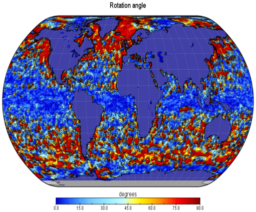
- Level 1 temperature; **19 members**; **filtered estimates** (Michel *et al.* 2016)



(j) Principal axis L1



(k) Principal axis L2



(l) Rotation angle

What are some of the key issues?

- Computing derivatives of ensemble perturbations is inherently noisy.
- Computing the elements of $\widetilde{\boldsymbol{H}}$ using finite differences on a staggered grid (C-grid in NEMO) requires some averaging to colocate elements. This introduces inaccuracies.
- Filtering $\widetilde{\boldsymbol{H}}$ while ensuring symmetry and positive-definiteness is tricky (Michel *et al.* 2016).
- Filtering with a constant scale does not seem appropriate for the global ocean.
- The numerical representation of the non-diagonal elements of $\boldsymbol{\kappa}$ in the diffusion operator is complicated (self-adjointness, positive definiteness).
- Solving the linear system of an implicitly-formulated diffusion equation with non-diagonal $\boldsymbol{\kappa}$ is expensive.
- Ignoring the non-diagonal elements may result in artificially short length scales.
- Recomputing $\boldsymbol{\kappa}$ on each assimilation cycle would require recomputing normalization factors on each cycle, which is expensive.

What are some of the key issues?

- Computing derivatives of ensemble perturbations is inherently noisy.
- Computing the elements of $\widetilde{\boldsymbol{H}}$ using finite differences on a staggered grid (C-grid in NEMO) requires some averaging to colocate elements. This introduces inaccuracies.
- Filtering $\widetilde{\boldsymbol{H}}$ while ensuring symmetry and positive-definiteness is tricky (Michel *et al.* 2016).
- Filtering with a constant scale does not seem appropriate for the global ocean.
- The numerical representation of the non-diagonal elements of $\boldsymbol{\kappa}$ in the diffusion operator is complicated (self-adjointness, positive definiteness).
- Solving the linear system of an implicitly-formulated diffusion equation with non-diagonal $\boldsymbol{\kappa}$ is expensive.
- Ignoring the non-diagonal elements may result in artificially short length scales.
- Recomputing $\boldsymbol{\kappa}$ on each assimilation cycle would require recomputing normalization factors on each cycle, which is expensive.
- **Is this the right approach?**

Using ensembles to estimate $\mathbf{B}_\mathrm{e}$

- The localized ensemble-based covariance matrix has the form:

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

- The operator form of Schur product localization used for minimization in variational assimilation is

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

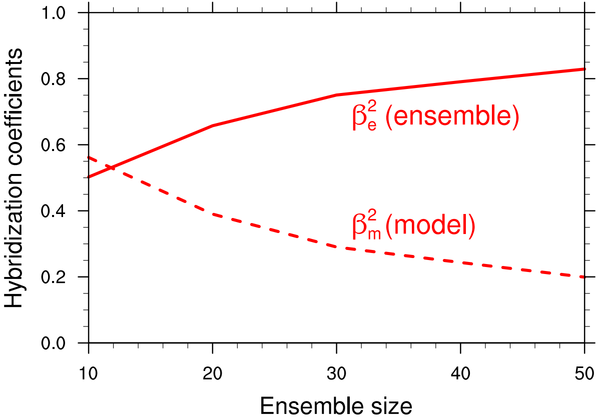
- We use a diffusion operator for the localization operator $\mathbf{L}$.
- We also consider the **hybrid** variant:

$$\mathbf{B} = \beta_\mathrm{e}^2\, \mathbf{B}_\mathrm{e} + \beta_\mathrm{m}^2\, \mathbf{B}_\mathrm{m}$$

  where $\mathbf{B}_\mathrm{m}$ employs climatological or modelled parameters.
- $\mathbf{L}$ and the hybridization weights β_m^2 and β_e^2 can be estimated using an optimally-based procedure (Ménétrier and Auligné 2015).
- The algorithm is applied offline using the *hybrid_diag* software (B. Ménétrier)
- It has been interfaced with NEMOVAR (Y. Yang).

Optimal hybridization weights

Example from NEMOVAR

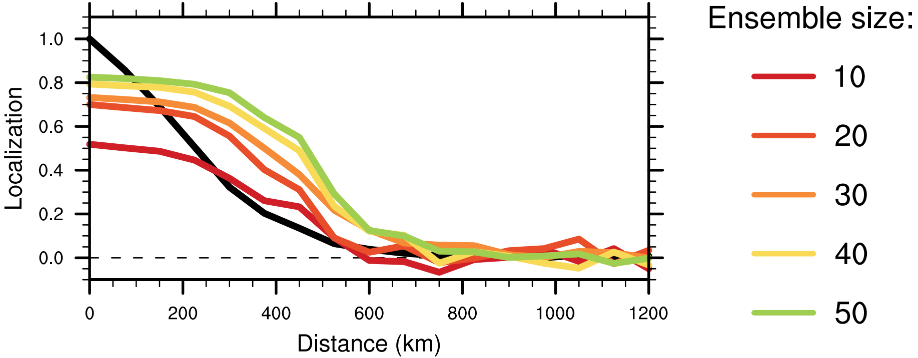


As expected:

- β_e^2 increases with the ensemble size.
- β_m^2 decreases with the ensemble size.

Localization and hybridization are optimized simultaneously.

Example from NEMOVAR



Correlation (black) and localization (colors)

- 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:

$$\mathbf{L} = \mathrm{diag}\left(\mathbf{L}_1, \dots, \mathbf{L}_M\right)$$

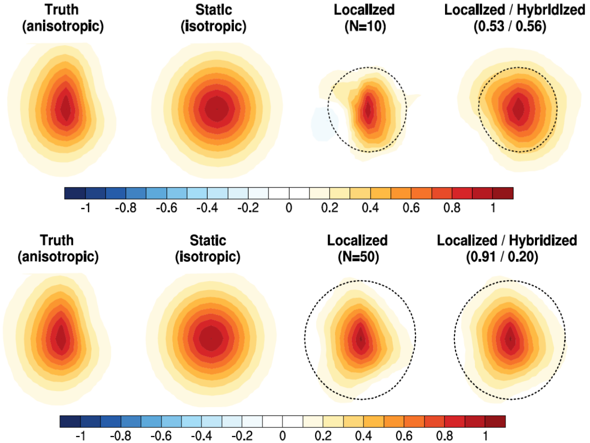
  3. Multivariate and common localization for each variable:

$$\mathbf{L} = \begin{pmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{pmatrix} \mathbf{L}_1 \begin{pmatrix} \mathbf{1} & \cdots & \mathbf{1} \end{pmatrix}$$

  4. Multivariate and separate localization for each of the M variables:

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_1^{1/2} \\ \vdots \\ \mathbf{L}_M^{1/2} \end{pmatrix} \begin{pmatrix} \mathbf{L}_1^{\mathrm{T}/2} & \cdots & \mathbf{L}_M^{\mathrm{T}/2} \end{pmatrix}$$

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



- We can employ localization for the correlations and spatial filtering for the variances.
- In the hybrid formulation, the modelled component could be based on a time-averaged ensemble for the variances and diffusion tensor.
- Localization scales increase with ensemble size.
- The conditioning of the system matrix of implicit diffusion degrades as the length scales increase, so the cost of localization gets more expensive.
- We can (should) do localization on a coarse grid to reduce cost. Developing appropriate coarse grids for the global ocean is tricky (work in progress).
- It is difficult to represent flat-like localization functions with a diffusion operator.
- It is probably not necessary to estimate localization scales on every assimilation cycle, so the cost of re-normalization can be reduced.

Concluding remarks

- A lot of work has been done in AVENUE and related projects (ERA-CLIM2) to develop different methods for defining ensemble covariances in NEMOVAR.
- Operational experience is now needed to determine the best choices.
- This has started recently at ECMWF, as discussed next in Marcin Chrust's presentation.

References

- Auligné T, Ménétrier B, Lorenc A, Buehner M. 2016. Ensemble-Variational Integrated Localized data assimilation. *Mon. Weather Rev.* **144**, 3677–3696.
- Belo Pereira M, Berre L. 2006. The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Weather Rev.* **134**: 2466–2489.
- Derber J, Rosati A. 1989. A global oceanic data assimilation system. *J. Phys. Oceanogr.* **19**: 1333–1347.
- Gratton S, Tshimanga J. 2009. An observation-space formulation of variational assimilation using a Restricted Preconditioned Conjugate-Gradient algorithm. *Q. J. R. Meteorol. Soc.* **135**: 1573–1585.
- Gürol S, Weaver AT, Moore AM, Piacentini A, Arango HG, Gratton S. 2014. **B**-preconditioned minimization algorithms for variational data assimilation. *Q. J. R. Meteorol. Soc.* **140**: 539–556.
- Guttorp P, Gneiting T. 2006. Miscellanea studies in hte history of probability and statistics XLIX: on the Matérn correlation family. *Biometrika*. **93**: 989–995.
- Hristopulos DT. 2002. New anisotropic covariance models and estimation of anisotropic parameters based on the covariance tensor identity. *Stoch. Environ. Res. Risk Assess.* **16**: 43–62.

- Ménétrier B, Montmerle T, Michel Y, Berre L. 2015. Linear filtering of sample covariances for ensemble-based data assimilation. Part I: optimality criteria and application to variance filtering and covariance localization. *Mon. Weather Rev.* **143**: 1622–1643.
- Ménétrier B, Auligné T. 2015. Optimized localization and hybridization to filter ensemble-based covariances. *Mon. Weather Rev.* **143**: 3931–3947.
- Michel Y, Ménétrier B, Montmerle T. 2016. Objective filtering of the local correlation tensor. *Q. J. Roy. Meteorol. Soc.* **142**: 2314–2323.
- Mirouze I, Blockley EW, Lea DJ, Martin MJ, Bell MJ. 2016. A multiple length scale correlation operator for ocean data assimilation. *Tellus A*. **68**.
- Sato Y, De Pondeca MSFV, Purser RJ, Parrish DF. 2009. Ensemble-based background error covariance implementations using spatial recursive filters in NCEP's grid-point statistical interpolation system. Office Note 459, National Centers for Environmental Prediction: Camp Springs, MD.
- Varella H, Berre L, Desroziers G. 2011. Diagnostic and impact studies of a wavelet formulation of background-error correlations in a global model. *Q. J. Roy. Meteorol. Soc.* **137**: 1369–1379.

References

- Weaver AT, Mirouze I. 2013. On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Q. J. Roy. Meteorol. Soc.* **139**: 242–260.
- Weaver AT, Tshimanga J, Piacentini A. 2016. Correlation operators based on an implicitly formulated diffusion equation solved with the Chebyshev iteration. *Q. J. Roy. Meteorol. Soc.* **142**: 455–471.