Developments in ensemble-variational data assimilation for the global ocean

Anthony Weaver
CERFACS, Toulouse

Acknowledgements:
S. Gürol\textsuperscript{1}, B. Ménétrier\textsuperscript{1}, A. Piacentini\textsuperscript{1}, J. Tshimanga\textsuperscript{1}, Y. Yang\textsuperscript{1}, M. Balmaseda\textsuperscript{2}, M. Chrust\textsuperscript{2}, A. Vidard\textsuperscript{3}, D. Lea\textsuperscript{4}, M. Martin\textsuperscript{4}
\textsuperscript{1}CERFACS, Toulouse; \textsuperscript{2}ECMWF, Reading; \textsuperscript{3}INRIA, Grenoble; \textsuperscript{4}Met Office, Exeter
A few words about NEMOVAR

- A joint project by CERFACS, ECMWF, Met Office and INRIA to develop an ocean data assimilation system for the NEMO\textsuperscript{1} model, with a variational kernel.

- Two operational centres (ECMWF, Met Office) and two research institutes (CERFACS, INRIA).

- The project is formalized by an MoU.

- 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:

1. Outline recent developments aimed at using ensembles for specifying $\mathbf{B}$ in NEMOVAR.
2. Discuss computational issues related to representing $\mathbf{B}$ in NEMOVAR.

\textsuperscript{1}Nucleus for European Modelling of the Ocean.
The baseline is an **Ensemble of Data Assimilations** framework for generating forecast and analysis ensembles (in development).

**Deterministic EVIL** (Auligné *et al.* 2016) has also been developed as a cheaper alternative to the EDA for generating analysis ensembles, using Ritz information from B-preconditioned CG minimization algorithms:

\[
X^a = \left\{ I_n - \hat{Z}_q \left( I_q - \Theta_q^{-1/2} \right) \bar{Z}_q^T \right\} X^b 
\]

\[
X^a = \left\{ I_n - BH^T \bar{Z}^o_q \left( I_q - \Theta_q^{-1/2} \right) \bar{Z}_q^o H \right\} X^b 
\]

(B-PCG; Derber & Rosati 1989)  
(Restricted B-PCG; Gürol *et al.* 2014)
The NEMOVAR B formulation

- The NEMOVAR B formulation is quite general:

\[ B = \beta_m^2 \left( B_{m1} + B_{m2} + \ldots \right) + \beta_e^2 B_e + \beta_{EOF}^2 B_{EOF} \]

where \( \beta_m^2, \beta_e^2 \) and \( \beta_{EOF}^2 \) are constant weights or switches.

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

\[ B_{m_i} = K_b D_i^{1/2} C_{m_i} D_i^{1/2} K_b^T \]

- A localized ensemble-based correlation matrix:

\[ B_e = K_b D_e^{1/2} \left( L \circ \tilde{X} \tilde{X}^T \right) D_e^{1/2} K_b^T \]

where the columns of \( \tilde{X} = D_e^{-1/2} K_b^{-1} X^b \) are transformed background ensemble perturbations.

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

\[ B_{EOF} = P \Lambda P^T \]
We have developed two ways of defining $B$ from ensembles:

1. Estimate the variances and local correlation tensor of a diffusion-based covariance model $B_m$.

2. Form the sample covariance matrix and localize it (in model space) using a Schur product with a specified correlation matrix $L$:

$$B_e = L \odot \tilde{X}\tilde{X}^T = L \odot \tilde{B} \iff (B_e)_{ij} = L_{ij} \tilde{B}_{ij}$$

The equivalent operator form used in the variational minimization is

$$B_e v = \sum_{p=1}^{N_e} \left( \tilde{x}_p \circ L \left( \tilde{x}_p \circ v \right) \right) \quad \text{where} \quad \tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_{N_e})$$

We also consider the hybrid variant:

$$B = \beta_e^2 B_e + \beta_m^2 B_m$$

where $B_m$ employs climatological or modelled parameters.

How to estimate the localization matrix $L$ and hybridization weights $\beta_m^2$ and $\beta_e^2$?
Localization and hybridization

- Localization by $\mathbf{L} +$ hybridization with $\mathbf{B}_m$ to combat sampling error:

\[
\mathbf{B} = \underbrace{\beta_e^2 \mathbf{L}}_{\text{Gain } \mathbf{L}^h} \circ \tilde{\mathbf{B}} + \underbrace{\beta_m^2 \mathbf{B}_m}_{\text{Offset}}
\]

Localization + hybridization = linear filtering of $\tilde{\mathbf{B}}$

$\mathbf{L}^h$ and $\beta_m^2$ have to be optimized together

- Optimal localization/hybridization minimizes (Ménétrier and Auligné 2015)

\[
e^h = \mathbb{E} \left[ \| \mathbf{L}^h \circ \tilde{\mathbf{B}} + \beta_m^2 \mathbf{B}_m - \tilde{\mathbf{B}}^* \|_F^2 \right]
\]

where $\tilde{\mathbf{B}}^* = \lim_{N_e \to \infty} \tilde{\mathbf{B}}$ is the target.

It can be shown that, with optimal parameters, whatever the static $\mathbf{B}_m$:

Localization + hybridization is better than localization alone
Optimal hybridization weights

Practical expressions can be derived for the optimal weights (Ménétrier and Auligné 2015).

Example from NEMOVAR

As expected:

- $\beta_e^2$ increases with the ensemble size.
- $\beta_m^2$ decreases with the ensemble size.
Optimal localization

Localization and hybridization are optimized simultaneously.

Example from NEMOVAR

Correlation (black) and localization (colors)

Ensemble size:
- 10
- 20
- 30
- 40
- 50
Aspects of the practical computation

A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E} \left[ \cdot \right]$ in the optimal formulae.

Estimation of correlation and localization (10 members, 5 m temperature)
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\cdot]$ in the optimal formulae.

Estimation of correlation and localization (10 members, 5 m temperature)
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[ \cdot ]$ in the optimal formulae.

Estimation of correlation and localization (10 members, 5 m temperature)
A spatial ergodicity assumption is required to estimate the statistical expectations $E[\cdot]$ in the optimal formulae.

Estimation of correlation and localization (10 members, 5 m temperature)
Aspects of the practical computation

A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E} \left[ \cdot \right]$ in the optimal formulae.

Estimation of correlation and localization (10 members, 5 m temperature)
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\cdot]$ in the optimal formulae.

Estimation of correlation and localization (10 members, 5 m temperature)
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\cdot]$ in the optimal formulae.

Estimation of correlation and localization (10 members, 5 m temperature)
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\cdot]$ in the optimal formulae.

Estimation of correlation and localization (10 members, 5 m temperature)
Hybrid correlations from NEMOVAR

Example of 5 m T-T correlations at a point in the North Atlantic

Fit the localization function to an $M$th-order AR-function and model it with a diffusion operator ($L$).

Optimized parameters: $\beta_e^2 = 0.56$, $\beta_m^2 = 0.54$, $L_{\text{loc}} = 260$ km, $M = 6$. 
Diffusion-based filtering operators

- In NEMOVAR, implicitly formulated diffusion operators are used for defining the univariate correlation operators in $B_m$, the localization operator in $B_e$, and for parameter filtering.
- The associated smoothing kernels are closely related to those from the well-known Matérn family (Weaver and Mirouze 2013).
- In the original NEMOVAR formulation, 2D and 3D diffusion operators were represented as products of simpler 1D operators (cf. recursive filter).
- The 1D approach has some nice features but is difficult to make anisotropic, produces numerical artefacts near complex boundaries, and scales poorly.

These problems, especially lack of scalability, compelled us to develop a new approach.
Iterative solvers for implicit diffusion

- Solve the 2D and 3D implicit diffusion problems approximately using an iterative algorithm based on the **Chebyshev iteration (CI)** (Weaver et al. 2015).
- We formulate an $M$-step implicit diffusion operator as $\mathbf{U U}^T$ where $\mathbf{U}$ and $\mathbf{U}^T$ each involve solving a sequence of $M/2$ linear systems with SPD matrix $\mathbf{A}$. 
- Several attractive properties of CI for this problem.
  - Extreme eigenvalues of $\mathbf{A}$ required by CI can be pre-computed using a Lanczos method.
  - Similar convergence properties as CG.
  - No scalar products $\implies$ no global MPI communications.
  - Straightforward to implement.
  - Linear solver $\implies$ exact numerical symmetry of $\mathbf{U U}^T$ can be enforced for arbitrary convergence threshold.
We do not need a strict convergence criterion to get an adequate solution.

(c) $K = 2$ ($\epsilon_k = 0.3$)  
(d) $K = 4$ ($\epsilon_k = 10^{-1}$)  
(e) $K = 9$ ($\epsilon_k = 10^{-2}$)  
(f) $K = 13$ ($\epsilon_k = 10^{-3}$)  
(g) $K = 43$ ($\epsilon_k = 10^{-10}$)
A “time”-parallel variant of implicit diffusion

- Rewrite the sequence of $M/2$ symmetric linear systems as a single nonsymmetric linear system:

\[
\begin{align*}
A\psi_1 &= \psi_0 \\
A\psi_2 &= \psi_1 \\
&\vdots \\
A\psi_{M/2} &= \psi_{M/2-1}
\end{align*}
\]

\[
\begin{align*}
\Rightarrow -\psi_1 + A\psi_2 &= 0 \\
&\vdots \\
-\psi_{M/2-1} + A\psi_{M/2} &= 0
\end{align*}
\]

This has the form $A\psi = \zeta$ where

\[
A = \begin{bmatrix}
A & -I & A \\
-1 & A & \ddots & \ddots \\
& \ddots & \ddots & \ddots & \ddots \\
& & -1 & A & -I \\
& & & & -1 & A
\end{bmatrix}, \quad \psi = \begin{bmatrix}
\psi_1 \\
\psi_2 \\
\vdots \\
\psi_{M/2}
\end{bmatrix}, \quad \zeta = \begin{bmatrix}
\psi_0 \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

- $A$ and $A$ have the same eigenspectrum.
- We can solve the nonsymmetric system using the Chebyshev iteration.
- The advantage of the $A$ system is that the $A$-matrix operators can be applied in parallel on each Chebyshev iteration.
A “time”-parallel variant of implicit diffusion

- Define “run-time cost” as the number of sequential $A$-matrix products.
- The red crosses and blue circles show the total number of $A$-matrix products required to achieve a residual reduction of $10^{-4}$ for the symmetric and nonsymmetric systems, for experiments with different values of $M$.
- The violet circles represent the potential gain from “time” parallelism.
Restricted Additive Schwarz (RAS) preconditioning

- Solve the implicit diffusion problem exactly for each of the (small) local subdomains of the parallel domain decomposition.
- Include an overlap region for the direct solution method but only retain the solution within the subdomain for the RAS preconditioner.
- The RAS preconditioner is **linear but nonsymmetric**.
- Can be applied with the Chebyshev iteration.

(Courtesy of Marcin Chrust, ECMWF)

- Improved convergence \(\implies\) increased computation-to-communication ratio \(\implies\) improved code scalability.
Concluding remarks

NEMOVAR was largely rewritten to facilitate the use of ensembles in defining $B$.

Maintaining flexibility in the choices of algorithms important:

1. Primal and dual space minimization algorithms.
2. Modelled covariances, localized sample covariances, and hybrid variants.
3. (hybrid-)3D-Var and (hybrid-)4D-Var.

The computational efficiency of the diffusion-based filtering operators is crucial, especially for high-resolution configurations.

The existing algorithms were completely revised to improve the robustness and performance of the diffusion operators on massively parallel machines.

Work is ongoing to make further improvements to code modularity and flexibility; e.g., to allow for multiple grids at various resolutions within the same executable.

This work is part of a larger effort to integrate NEMOVAR under Object Oriented Prediction Systems (OOPS).