Block methods for solving an ensemble of data assimilations

F. Mercier, Y. Michel, T. Montmerle, P. Jolivet and S. Gürol.
Météo-France & CNRS
Workshop on Sensitivity Analysis and Data Assimilation in Meteorology and Oceanography, 1-6 July 2018, Aveiro, Portugal

The Ensemble of Data Assimilations (EDA)

The EDA is :

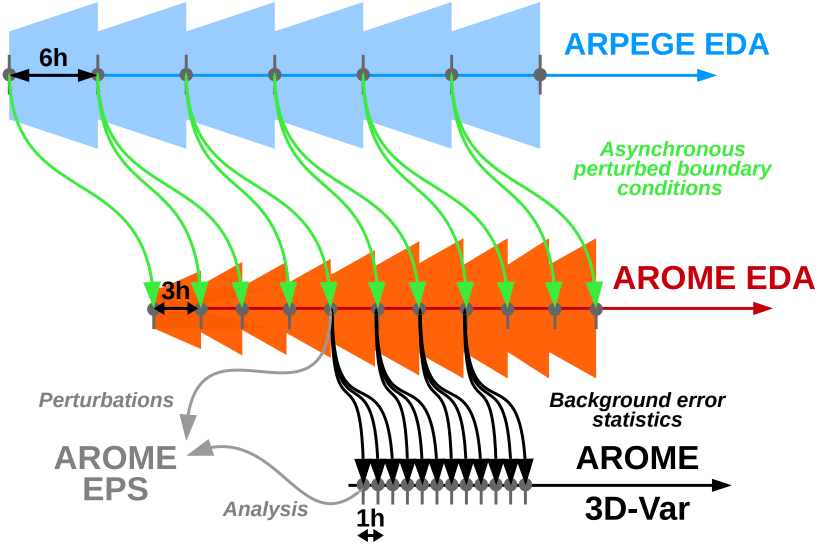
- an ensemble of cycled 3D or 4DVars with perturbed model, observations and surface/boundary conditions.
- the variational counterpart of the stochastic EnKF

Why running an EDA ?

The EDA provides an ensemble of analyses and short-range forecasts (backgrounds) that can be used to :

- build flow-dependent background error statistics for deterministic variational schemes
  (*e.g.*, the **B**-matrix of a 3D-Var or EnVar scheme) ;
- initialize an Ensemble Prediction System.

The AROME EDA in our NWP suite



The EDA based on 3DVars : formulation

Ensemble of 3DVars, for every member $k \in [\![1:m]\!]$

$$\mathcal{J}(\mathbf{x}^k) = \frac{1}{2}(\mathbf{x}_b^k - \mathbf{x}^k)\mathbf{B^k}^{-1}(\mathbf{x}_b^k - \mathbf{x}^k) + \frac{1}{2}\left(\mathbf{y}_o^k - \mathcal{H}^k(\mathbf{x}^k)\right)\mathbf{R^k}^{-1}\left(\mathbf{y}_o^k - \mathcal{H}^k(\mathbf{x}^k)\right)$$

- $\mathbf{y}_o^k$: perturbed observations for member k
- $\mathbf{x}^k$: (perturbed) background for member k.

Incremental formulation, assuming common **B** and **R**

- $\mathbf{x}^k = \mathbf{x}_b^k + \delta\mathbf{x}^k$
- $\mathcal{H}^k(\mathbf{x}^k) \approx \mathcal{H}^k(\mathbf{x}_b^k) + \mathbf{H}^k\delta\mathbf{x}^k$

Solution of a sequence of quadratic problems :

$$2\mathbf{J}(\delta\mathbf{x}^k) = \|\delta\mathbf{x}^k\|^2_{\mathbf{B}^{-1}} + \|\mathbf{d}^k - \mathbf{H}^k\delta\mathbf{x}^k\|^2_{\mathbf{R}^{-1}}$$

where $\mathbf{d}^k = \mathbf{y}_o^k - \mathcal{H}^k(\mathbf{x}_b^k)$ is the innovation.

The EDA based on 3DVars : formulation

Equating the gradient to zero :

$$\left(\mathbf{B}^{-1} + \mathbf{H}^{k\,\mathsf{T}}\mathbf{R}^{-1}\mathbf{H}^{k}\right)\delta\mathbf{x}^{k} = \mathbf{H}^{k\,\mathsf{T}}\mathbf{R}^{-1}\mathbf{d}^{k}$$

Primal formulation, right-**B** preconditioning :

$$\left(\mathbf{I} + \mathbf{H}^{k\,\mathsf{T}}\mathbf{R}^{-1}\mathbf{H}^{k}\mathbf{B}\right)\mathbf{v}^{k} = \mathbf{H}^{k\,\mathsf{T}}\mathbf{R}^{-1}\mathbf{d}^{k}$$

with $\delta\mathbf{x}^{k} = \mathbf{B}\mathbf{v}^{k}$

Linear system solved with the **B**-inner product

Dual formulation, left-$\mathbf{R}^{-1}$ preconditioning :

$$\left(\mathbf{I} + \mathbf{R}^{-1}\mathbf{H}^{k\,\mathsf{T}}\mathbf{R}^{-1}\mathbf{H}^{k}\right)\lambda^{k} = \mathbf{R}^{-1}\mathbf{d}^{k}$$

with $\delta\mathbf{x}^{k} = \mathbf{B}\mathbf{H}^{k\,\mathsf{T}}\lambda^{k}$

Linear system solved with the $\mathbf{H}^{k}\mathbf{B}\mathbf{H}^{k\,\mathsf{T}}$-inner product

The EDA based on 3DVars : formulation

Best method to solve for these m linear systems ?

- m independent minimizations with a Krylov subspace method ?
- A Block-Krylov method ?

Krylov subspace methods

- Linear systems

$$\mathbf{A}_k \mathbf{v}^k = \mathbf{b}_k$$

- Krylov subspace methods searchs for an approximate solution $\widetilde{\mathbf{v}}_i^k$ from a subspace $\widetilde{\mathbf{v}}_0^k + \mathcal{K}_i(\mathbf{A}_k, \mathbf{r}_0^k)$ where

$$\mathcal{K}_i(\mathbf{A}_k, \mathbf{r}_0^k) = \mathrm{Span}\big(\mathbf{b}_k, \mathbf{A}_k\mathbf{b}_k, \mathbf{A}_k^2\mathbf{b}_k, \cdots, \mathbf{A}_k^{i-1}\mathbf{b}_k\big)$$

- Look for the projection $\widetilde{\mathbf{v}}^k \in \mathcal{K}_i$:

$$\mathbf{A}_k\widetilde{\mathbf{v}}^k - \mathbf{b}_k \perp \mathcal{K}_i$$

- For symmetric and positive definite $\mathbf{A}_k$, this leads to Conjugate Gradient method.
- For unsymmetric systems this algorithm leads to the Full Orthogonal Method (FOM).

Block-FOM

- Linear systems

$$\mathbf{A}[\mathbf{v}^1, \cdots, \mathbf{v}^m] = [\mathbf{b}_1, \cdots, \mathbf{b}_m]$$

Note : this requires the same Hessian for every member !

- Krylov subspace methods searchs for an approximate solution $\widetilde{\mathbf{v}}_i^k$ from a subspace space B_i :

$$B_i = \text{Span}\big(\mathbf{b}_1, \cdots, \mathbf{b}_m, \mathbf{A}\mathbf{b}_1, \cdots, \mathbf{A}\mathbf{b}_m, \cdots, \mathbf{A}^{i-1}\mathbf{b}_1, \cdots, \mathbf{A}^{i-1}\mathbf{b}_m\big)$$

Search subspace is enlarged ($\dim(B_i) \leq i \times m$), and every member uses the information from all other ones.

- Look for the projection $\widetilde{\mathbf{v}}^k \in B_i$:

$$\mathbf{A}\widetilde{\mathbf{v}}^k - \mathbf{b}_k \perp B_i$$

- For symmetric and positive definite $\mathbf{A}$, this leads to Block-CG ; for unsymmetric systems Block-FOM.

Block versus non-block-Krylov : operation count

m independent Krylov	block-Krylov with m members
♣ Application of operators $\mathbf{B}, \mathbf{H}^\mathrm{T}, \mathbf{H}, \mathbf{R}^{-1}$:	
$m/$iteration	$m/$iteration
♣ Orthonormalisation of basis (scalar products, axpys)	
$\sim i \times m/$iteration	$\sim i^2 \times m/$iteration
	Use dual formulation !
♣ Handling of matrices of size i or $i \times m$	
low cost ($i \leq 100$)	moderate cost ($i \times m \leq 5000$)

Block-Krylov : the only way to win is to have less iterations / much faster convergence.

AROME EDA

Main characteristics

	AROME-France	AROME EDA
Spatial resolution	1.3 km	3.25 km
Timestep	50 s	100 s
Dynamical core	Non-Hydrostatic	Hydrostatic
Domain size	1440×1536 L90	600×640 L90
Assimilation	1 deterministic 3D-Var	25 perturbed 3D-Var
Frequency	1H	3H
Radar thinning	8 km	20 km
Nodes (64 Go)	48+2	6+1
MPI tasks (fc)	384	120

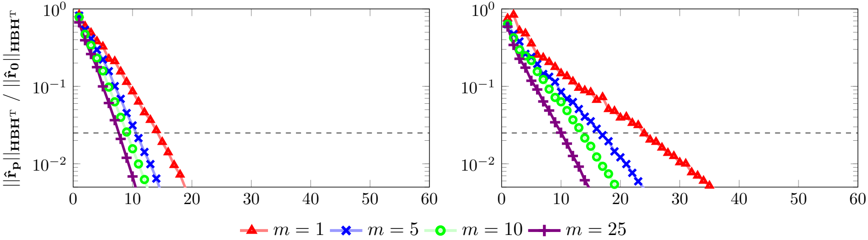
Adjustements w/r operational settings

- We use a common set of perturbed observations $\mathbf{y}_o^k = \mathbf{y}_o + \epsilon^k$
- The obs. operator is linearized around the ensemble mean $\mathbf{H}^k = \overline{\mathbf{H}}$.

Do we converge faster ?

Algorithm : block-RB-FOM

- block version of the left-$\mathbf{R}^{-1}$ preconditionned Full Orthogonal Method ;
- less prone to round-off errors than the Conjugate Gradient ;
- with local storage of the basis to reduce the communications.



Reduction of the norm of the residuals for various ensemble sizes with block-RB-FOM, for (left) a good and (right) a worse conditionning.

A parallelization strategy

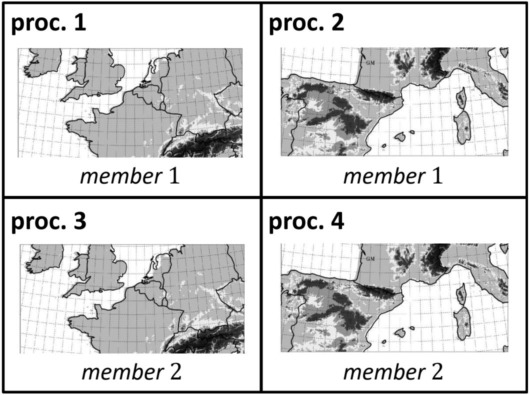
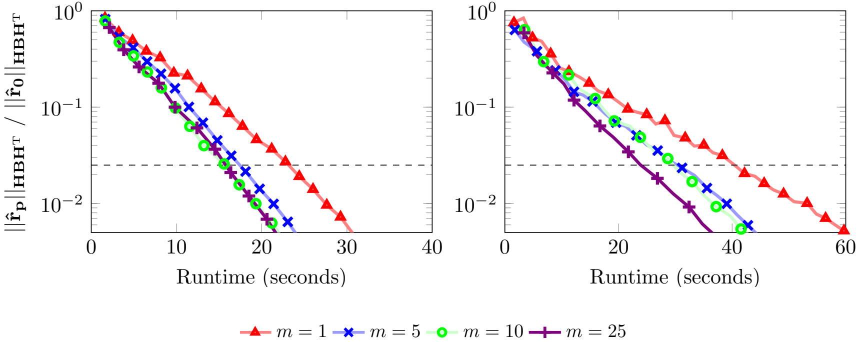


Illustration of the workload distributions implemented in OOPS on the AROME-France domain for four MPI processes and two members in the ensemble : workload distribution by member, combined with an underlying geographical distribution.

Do we run faster ?

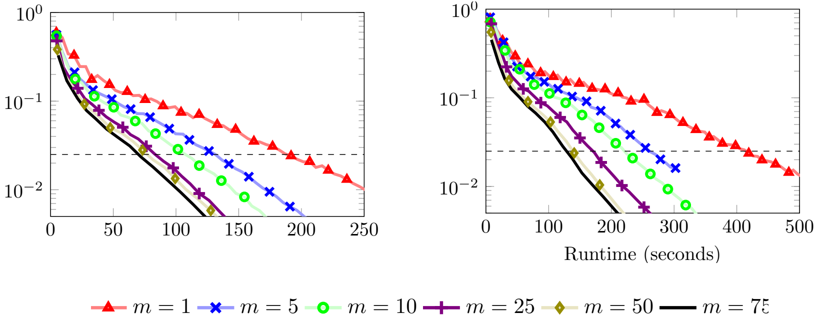


Runtime for various ensemble sizes with block-RB-FOM, for (left) a good and (right) a worse conditionning.

Extended experiments

Scalability

- Increasing ensemble size up to 75 members
- Using 10 times more observations (radar, satellite)



Runtime for various ensemble sizes with block-RB-FOM for extended experiments

Conclusions

- The block Krylov methods may be used to solve simultaneously an ensemble of perturbed minimizations, as encountered in the EDA.
- We derive a block algorithm (block-RB-FOM) that works in observation space to reduce the size of the control vectors (when $p \ll n$)
- Implemented under OOPS with advanced parallelization strategy
- With 25 members, gains in the range 45–55% (in the number of iterations) and in the range 25–50% (in terms of computational gain).
- Block methods are even more advantageous with larger ensemble sizes and more observations.
- Requires however the same Hessian, but may be combined with the "Mean-Pert" approach to deal with non-linearity in the mode.

There is more in two papers to come in QJRMS :

♣ Mercier et al.,2018b : Speeding up the ensemble data assimilation system of the limited area model of Météo-France using a block Krylov algorithm (revised).
♣ Mercier et al, 2018a : Block Krylov methods for accelerating ensembles of variational data assimilations (accepted).