Block iterative methods for Ensemble-Variational Assimilation in weather prediction

François Mercier

AVENUE Project workshop
20/06/2017
Cerfacs

Context: AROME

- AROME : limited area NWP model of Météo France.

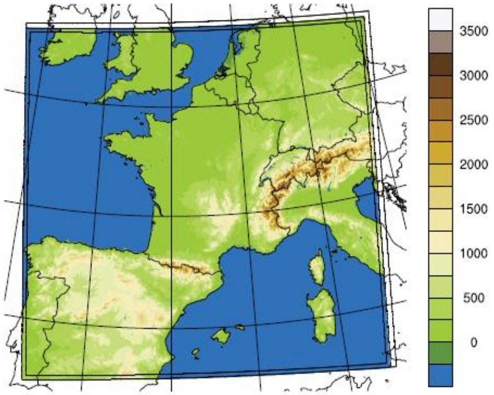


Fig. : B. Ménétrier thesis, 2014

- Forced at the boundaries by ARPEGE (global model)
- Forecasts up to $42h$

- Version used in this work : 3,8km horizontal resolution

- Size of the control vector : $540 * 512 * 90 * 4 \approx 10^8$

latitude (540), *longitude* (512), *vertical levels* (90), *physical variables* (4)

Context: data assimilation in AROME

- 3DVAR, cycled every 1 hour.
- EDA (Ensemble of 3DVAR)
- *Under development* : 3DEnVar

- Observations assimilated in AROME
  - ✓ Land synoptic stations
  - ✓ Radiosoundings
  - ✓ Plane measurements
  - ✓ Satellites
  - ✓ Radars

➢ $10^4 - 10^5$ observations

Context: data assimilation in AROME

- 3D-VAR, cycled every 3 hours.
- EDA (Ensemble of 3D-VAR)
- *Under development* : 3D-EnVar

Need to minimize an ensemble of 3DVAR cost functions.

➢ block Krylov methods !

1/ Theoretical considerations

2/ Results on a simplified geophysical model

3/ First results with AROME

Deterministic problem: 3DVAR

- 3DVAR cost function to minimize :
  - y : observations ; x^b : background state
  - $d = y - \boldsymbol{H}x^b$: innovation
  - δx an increment ($x^a = x^b + \delta x$)

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

- Solving the gradient problem ($\nabla J = 0$) with right-***B*** conditioning :

$$\underbrace{(\boldsymbol{Id} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{B})}_{\boldsymbol{A}}v = \underbrace{\boldsymbol{H}^T\boldsymbol{R}^{-1}d}_{r}$$

$$\delta x = \boldsymbol{B}v$$

➢ Linear system with hessian ***A***

$$\boldsymbol{A}v = r$$

Ensemble problem: En-3DVAR

- Ensemble with m members:
  - Perturbations of :
    - $y \rightarrow (y_1, \dots, y_m)$
    - $x^b \rightarrow (x_1^b, \dots, x_m^b)$

➢ $d \rightarrow (d_1, \dots, d_m)$

➢ m linear systems with same hessian and different right hand sides.

- Defining the block-vectors $[v_1 \dots v_m]$ and $[r_1 \dots r_m]$:

$$\boldsymbol{A}[v_1 \dots v_m] = [r_1 \dots r_m]$$

Minimization of J: which solutions ?

Which algorithm to solve this block-problem ?

- For 1 system:
  - iterative methods ($\boldsymbol{A} \in \mathbb{R}^{10^8 * 10^8}$)

- For the m systems ?
  - m independant (parallel) minimizations ?
  - Block-methods ?

Back to the 1 member problem: Krylov techniques

$\boldsymbol{A}v = r$

Incremental assimilation: we suppose $\delta x^{(0)} = 0$

- Iterative process. At iteration i :

    - We construct an orthonormal base for the Krylov space K_i:

$$K_i = span(r, \boldsymbol{A}r, \boldsymbol{A}^2 r, \dots, \boldsymbol{A}^{i-1} r)$$

    - We look for an approximate $\tilde{v}$ for the solution of $\boldsymbol{A}v = r$ assuming:

        - $\tilde{v} \in K_i$
        - $\tilde{r} = \boldsymbol{A}\tilde{v} - r \perp K_i$

- $\tilde{v} = v$ for $i \geq g$ (exact solution in g iterations)

  g degree of the minimal polynomial of $\boldsymbol{A}$ ($g \leq N$ the dimension)

Krylov techniques: naming / convergence issues

- Krylov techniques: generic name
- Arnoldi algorithm: for non symmetric $\boldsymbol{A}$
- Lanczos, CG: for symmetric $\boldsymbol{A}$ (here relatively to the $\boldsymbol{B}$-dot product).

- convergence rates highly dependent on the spectrum of the hessian
- interest of the $\boldsymbol{B}$-conditioning.

Block-Krylov methods: back to ensembles

$$\boldsymbol{A}[v_1, \dots, v_m] = [r_1, \dots, r_m]$$

- At iteration i, we define the block-Krylov space, B_i :

$$B_i = span\left(r_1, \dots, r_m, \boldsymbol{A}r_1, \dots, \boldsymbol{A}r_m, \dots, \boldsymbol{A}^{i-1}r_1, \dots, \boldsymbol{A}^{i-1}r_m\right)$$

$$\dim(B_i) \leq i * m$$

- For any member (p), we look for an approximate solution, $\tilde{v}_p$, such as:
  - $\tilde{v}_p \in B_i$
  - $\tilde{r}_p = \boldsymbol{A}\tilde{v}_p - r_p \perp B_i$

  ➡ For each member, we use information coming from all the others.

- If all vectors independant: exact solution in g/m iterations (compared to g iterations for the non block version).

Block methods: memory issues

- Considering the size of the control variables:

$$v \in \mathbb{R}^{10^8}$$

- With $m \approx 100$:

$$\boldsymbol{V} \in \mathbb{R}^{10^{10}}$$

- With $j \approx 50$ (iterations) and 2 bases stored (physical and conditioned spaces):

$$\text{total bases size} \in \mathbb{R}^{10^{12}}$$

➢ Dual space algorithms to make the calculations in the observations space

Gratton and Tshimanga 2009
Gürol et al. 2013

Block-Krylov methods in dual space: (Block) RB-Lanczos

1 member hessian problem, 2 equivalent formulations:

primal space : $$(\boldsymbol{B}^{-1} + \boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H})\delta x = \boldsymbol{H}^T \boldsymbol{R}^{-1} d$$

dual space : $$(Id + \boldsymbol{R}^{-1} \boldsymbol{H} \boldsymbol{B} \boldsymbol{H}^T)\lambda = \boldsymbol{R}^{-1} d$$

$$\delta x = \boldsymbol{B} \boldsymbol{H}^T \lambda$$

Solving this dual linear system using the $\boldsymbol{HBH}^T$ dot product:

- produces the same iterates as the primal space Krylov algorithm !
- Adaptable for block methods as well (cf. results).

Block vs. non-block approaches / Parallelization issues

- Block-Lanczos vs. m independent single-member Lanczos:
  - Less iterations expected.
  - Same number of $\boldsymbol{B}$, $\boldsymbol{H}$ applications / it.
  - More dot prods / axpys (m^2 vs. m)

- Parallelization:

3 codes :

- No MPI parallelization in block code
  - Finer geographical parallelization in Arome.
- MPI parallelization, local base storage
  - No memory issues.
  - Many communication.
- MPI parallelization, full base storage
  - Not many communications, independent of the iteration.
  - Require large memory capacities.
    - ➢ ok for dual space algorithm.

1/ Theoretical considerations

2/ Results on a simplified geophysical model

3/ First results with AROME

Quasi-Geostrophic (QG) model: brief introduction

- Simple but realistic two layers model of mid-latitudes large scale atmospheric circulation.

- State vector size:

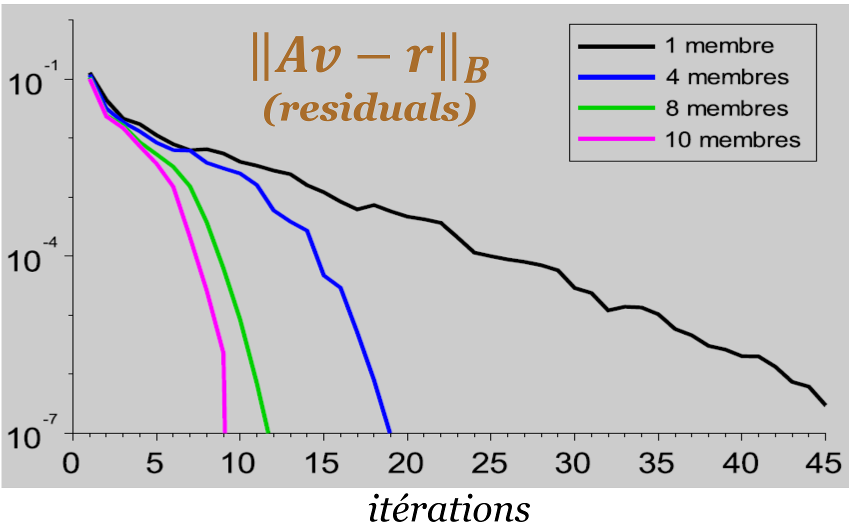
$$(40 * 20 * 2) = 1600 \qquad \text{or} \qquad (640 * 320 * 2) = 4.10^5$$

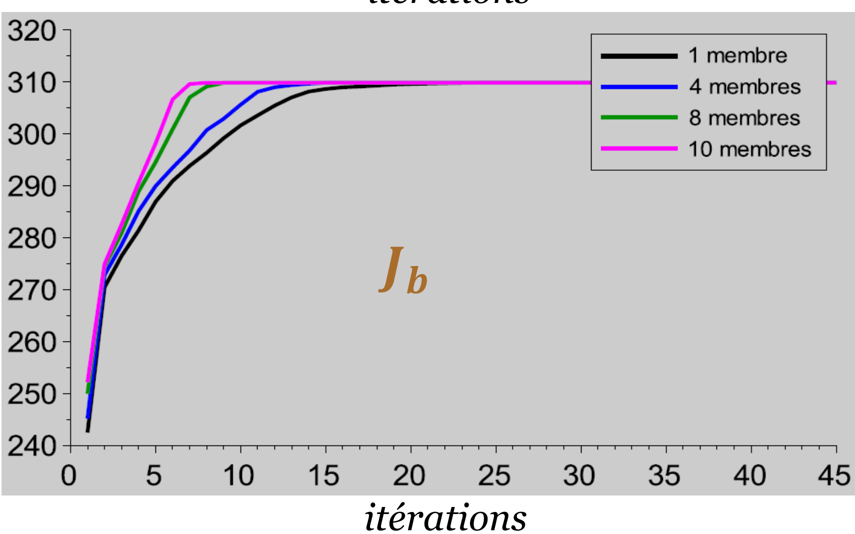
latitude *longitude* *vertical levels*

- Assimilation system:
  - 3DVAR and ensemble of 3DVAR
  - Background and observations explicitly perturbed to create an ensemble
  - 400 or 10^4 observations points

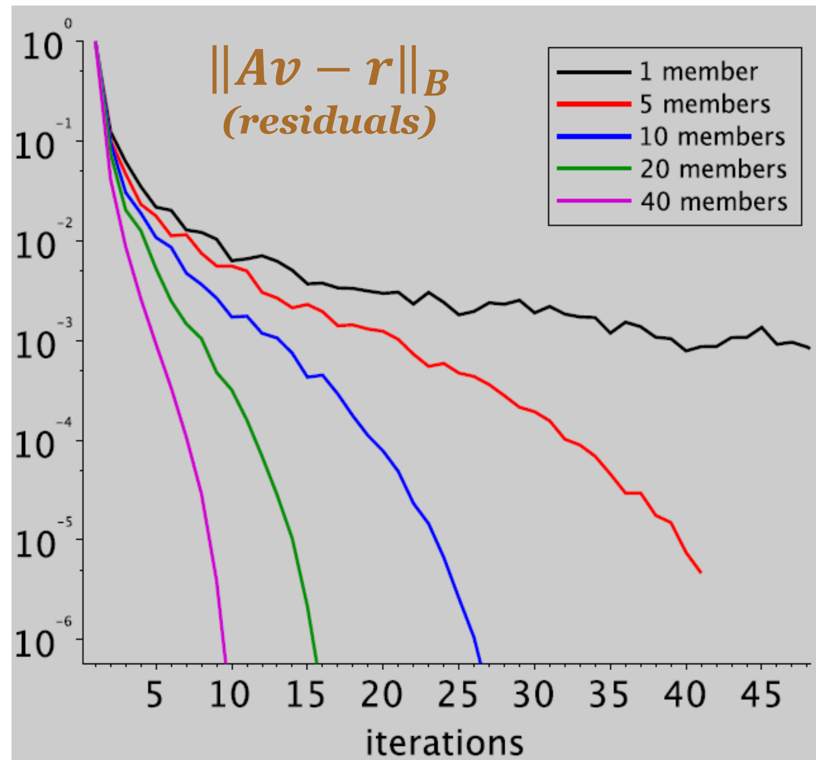
QG model: convergence of block-methods

- Small system ($N = 1600$)



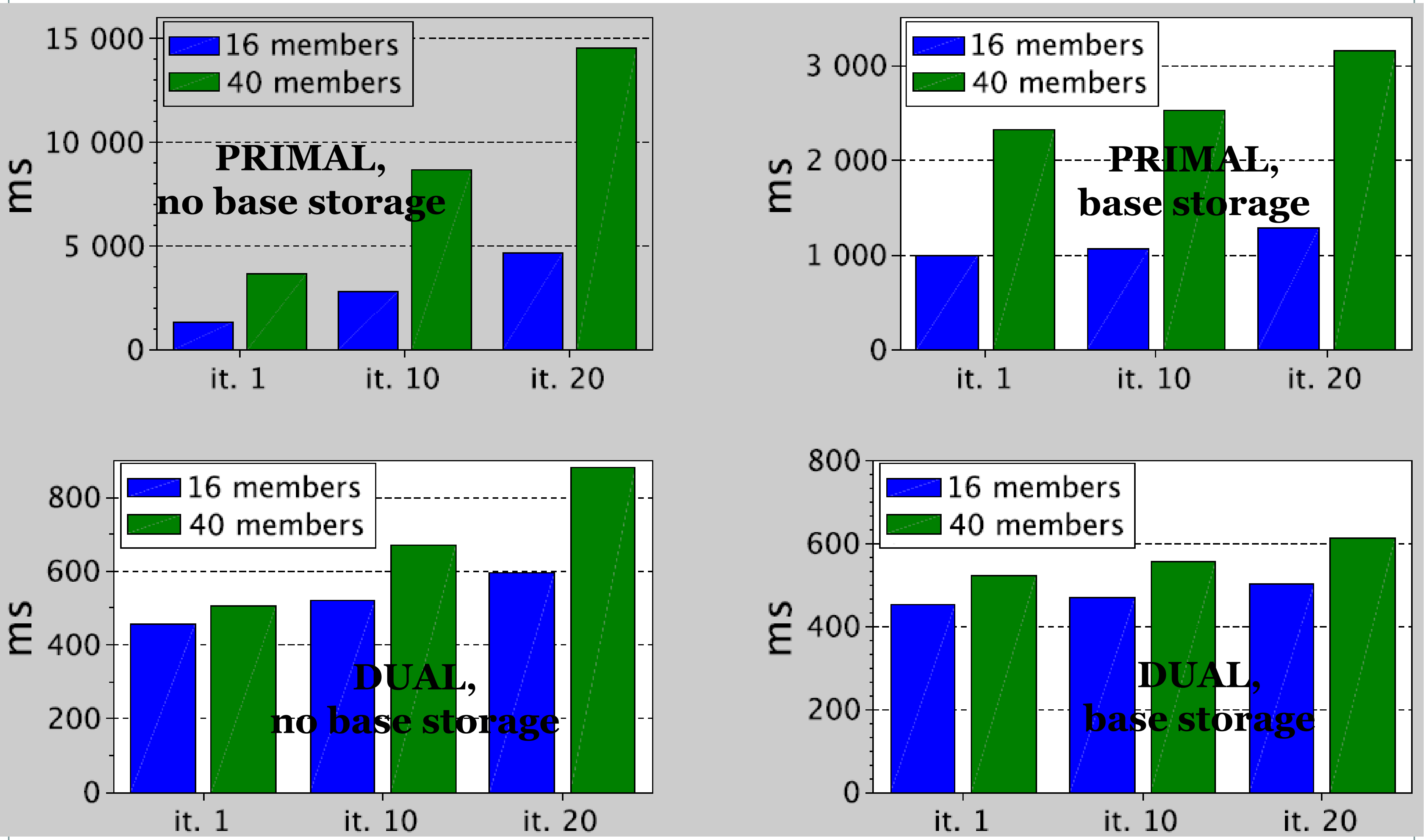


- Large system ($N = 4.10^5$)

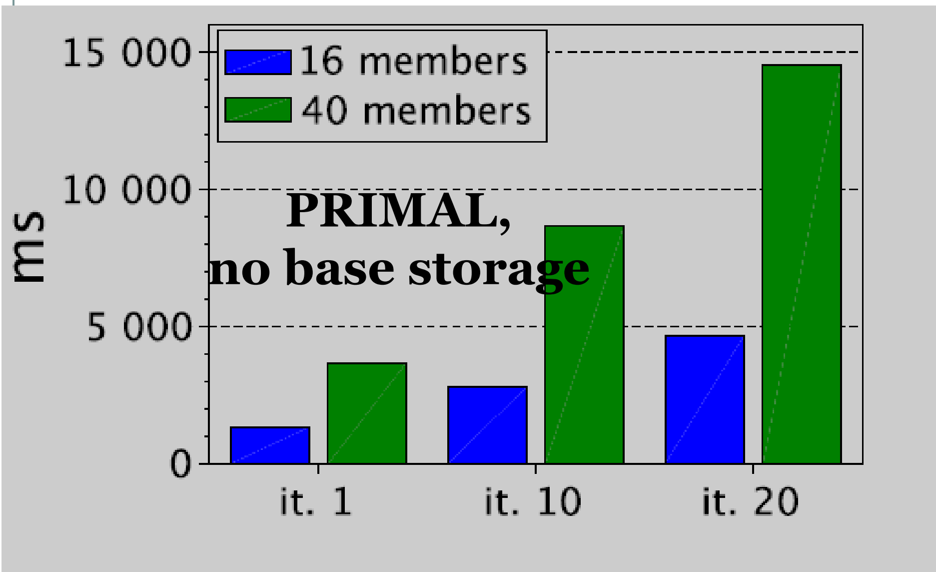


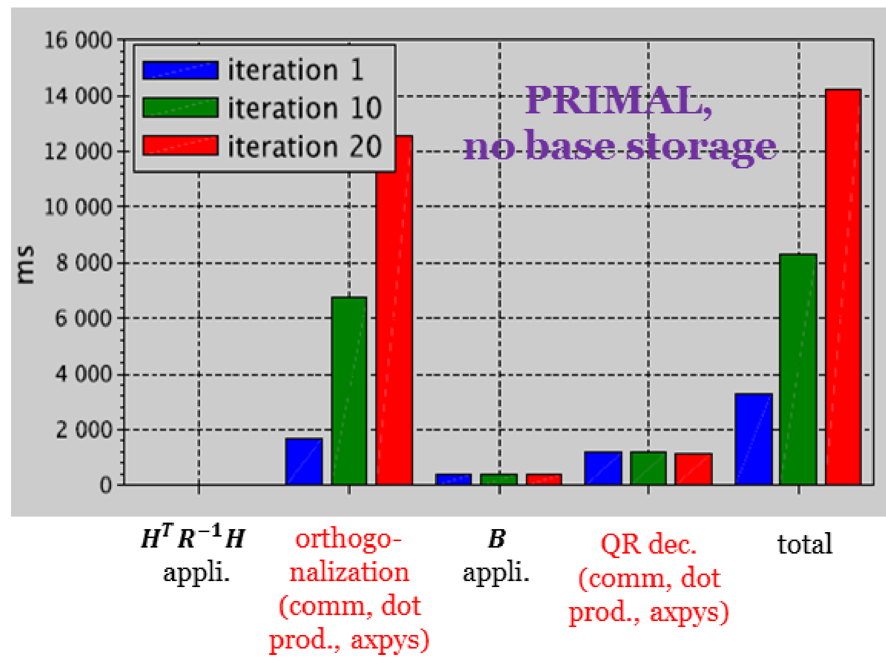
Results for the first member in any case

QG model: clock time for 16 vs. 40 members

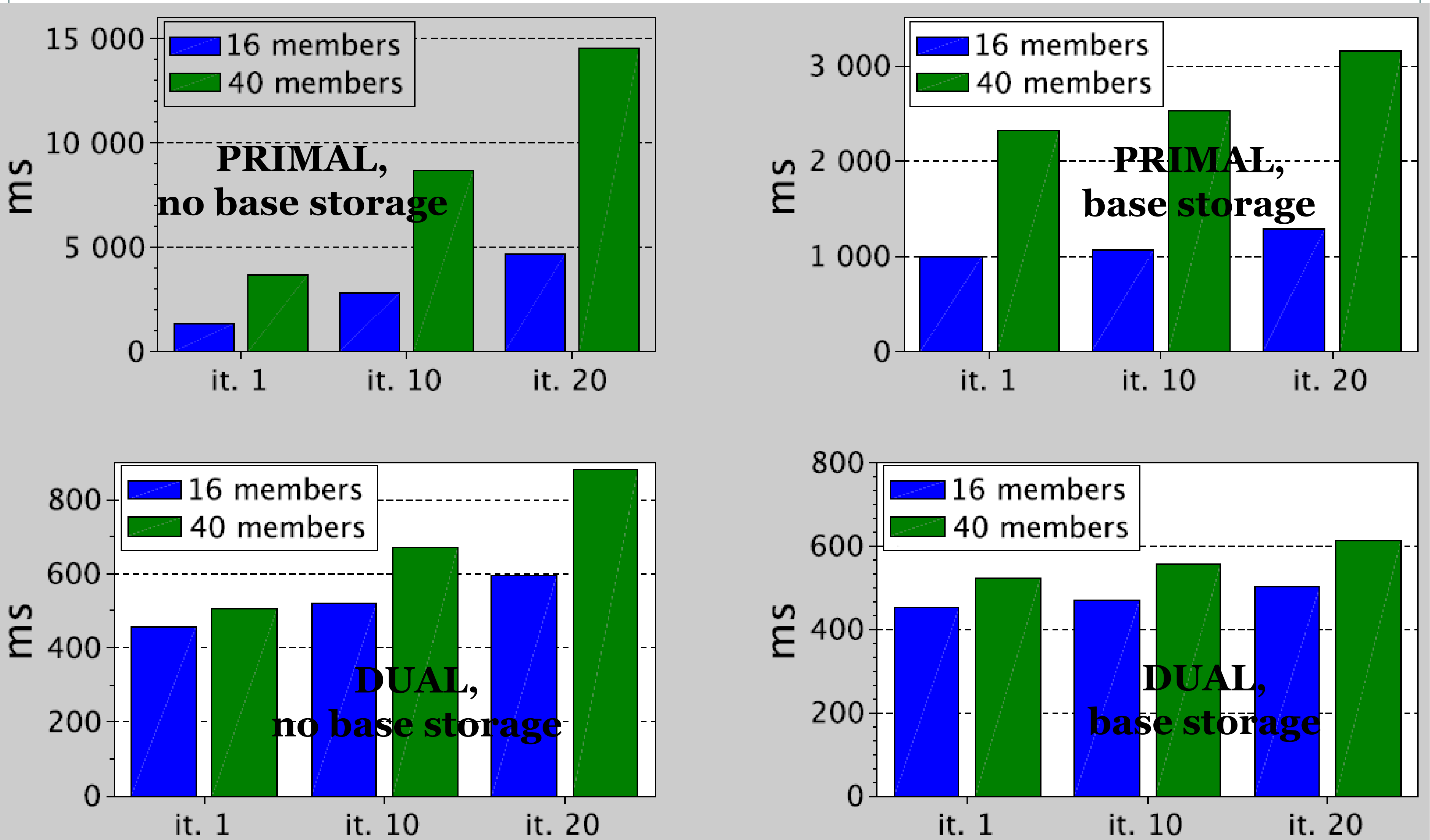


QG model: clock time for 16 vs. 40 members

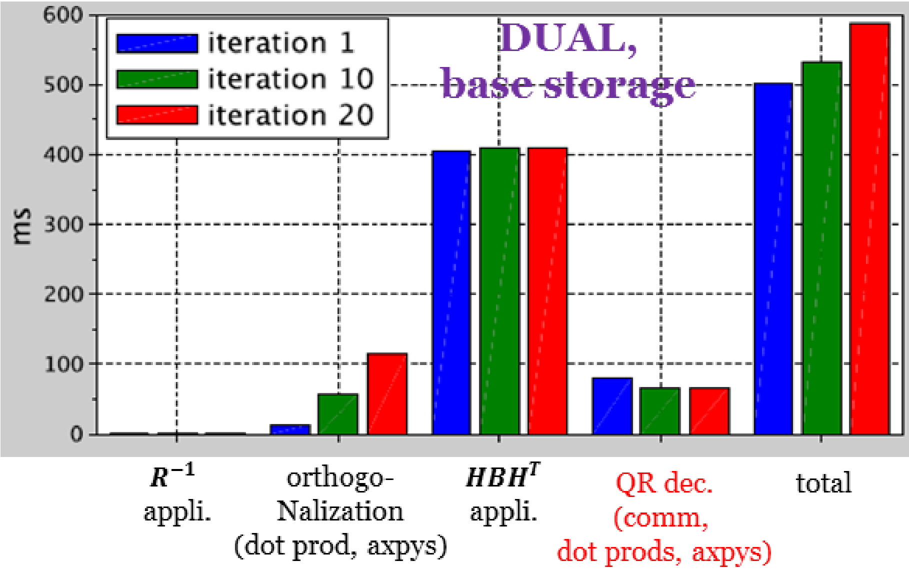


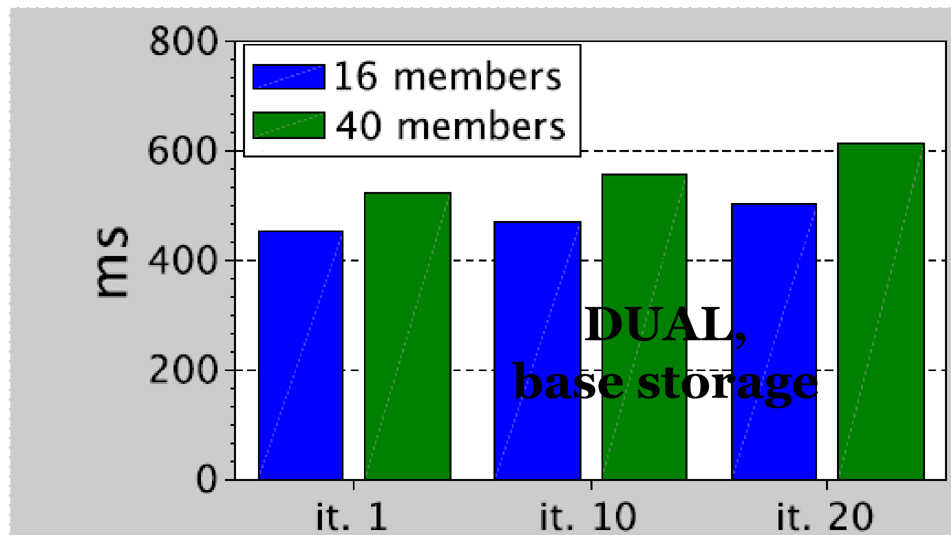


QG model: clock time for 16 vs. 40 members



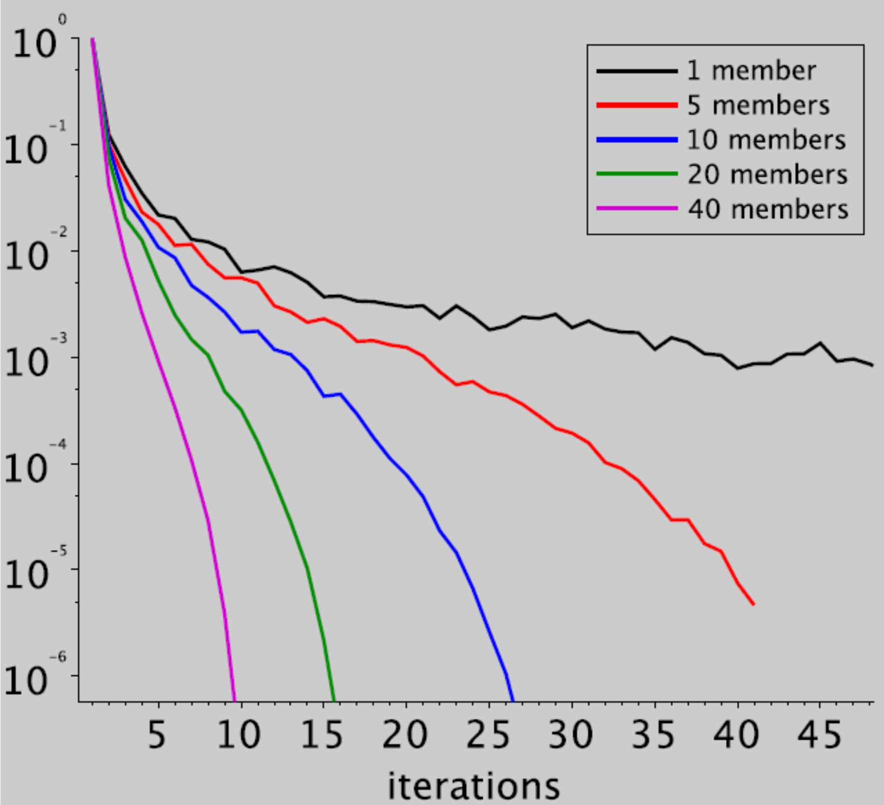
QG model: clock time for 16 vs. 40 members





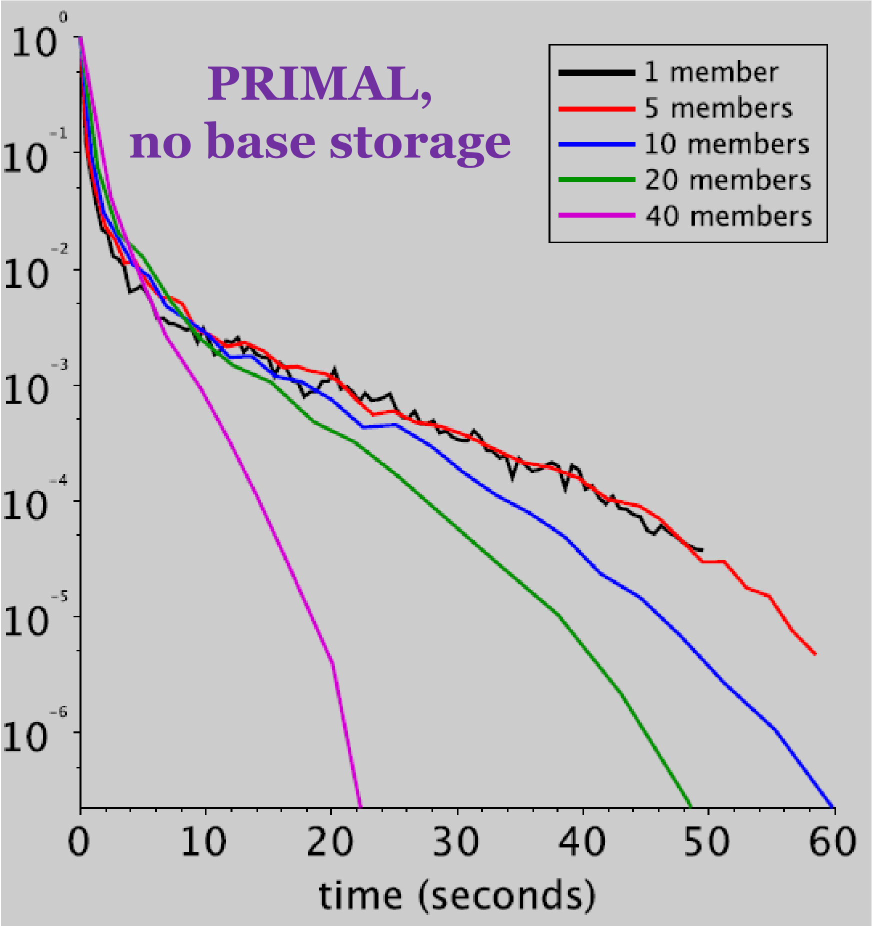
QG model: convergence time

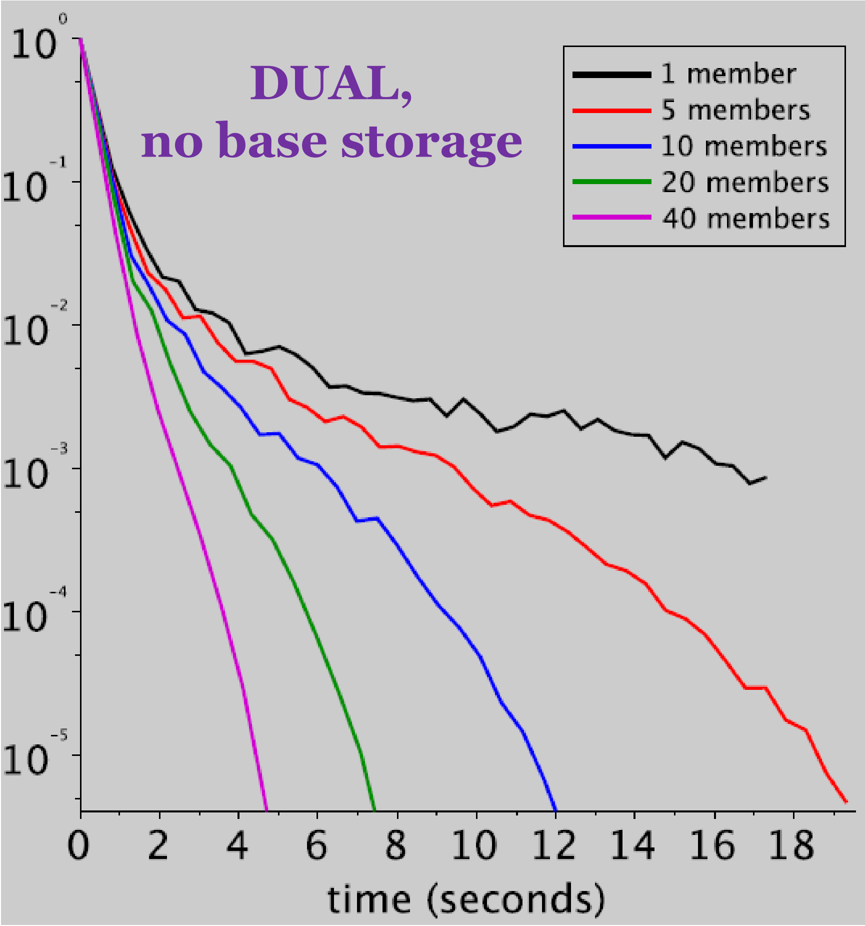
convergence vs. iterations



QG model: convergence time

convergence vs. time





1/ Theoretical considerations

2/ Results on a simplified geophysical model

3/ First results with AROME

AROME: first experiment parameters

- Ensemble of 3DVAR, $04/02/2017$, $00h$.

$\boldsymbol{B}$: climatological
$\boldsymbol{R}$: diagonal
$\boldsymbol{H}$: linearized around the mean background state.

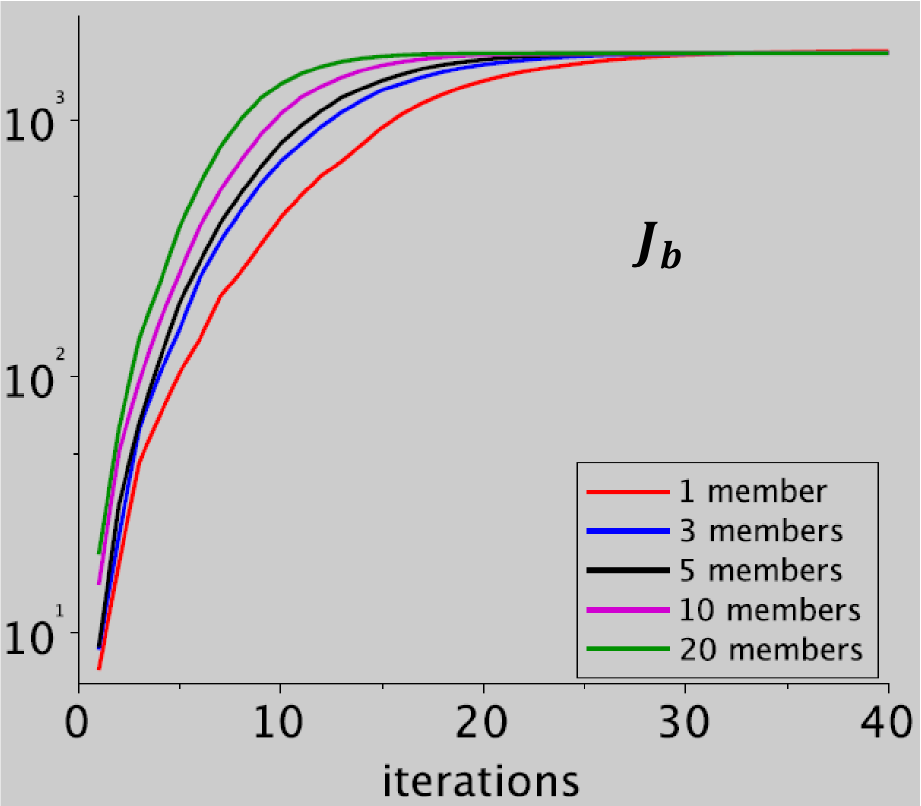
$N = 10^8$
$Nobs = 10^5$

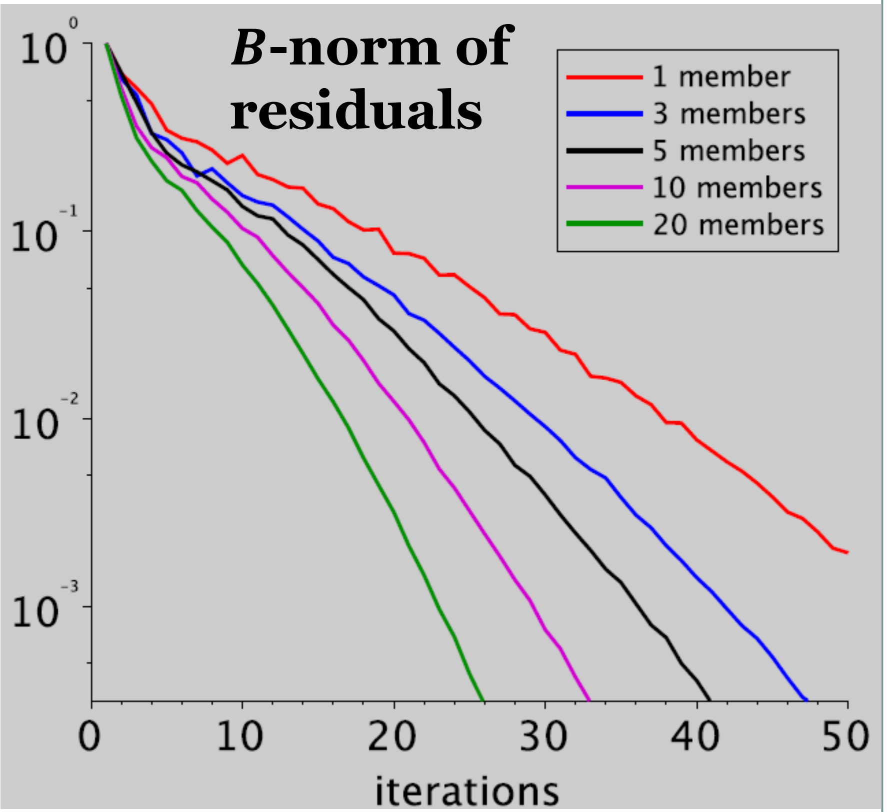
- 1 to 20 members

- Observations explicitly perturbed ($\sim\mathcal{N}(0, \boldsymbol{R}_{ij}^2)$)
  Backgrounds implicitly perturbed

AROME: results in dual space

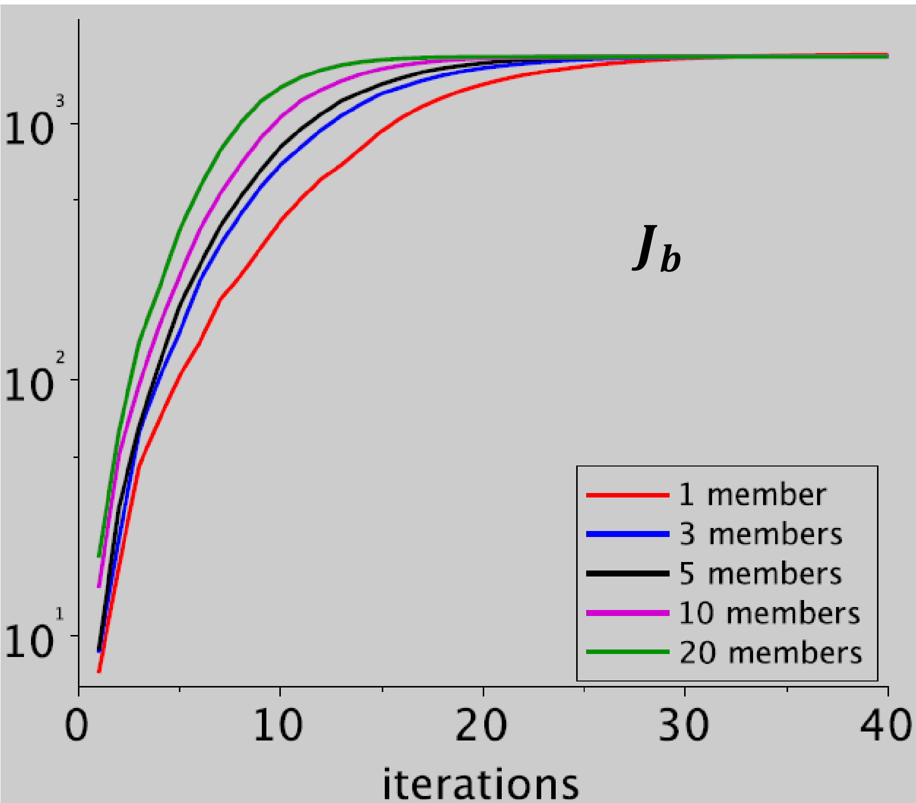
A set of conventional observations (3.10^4, no radar, no sats), dual space

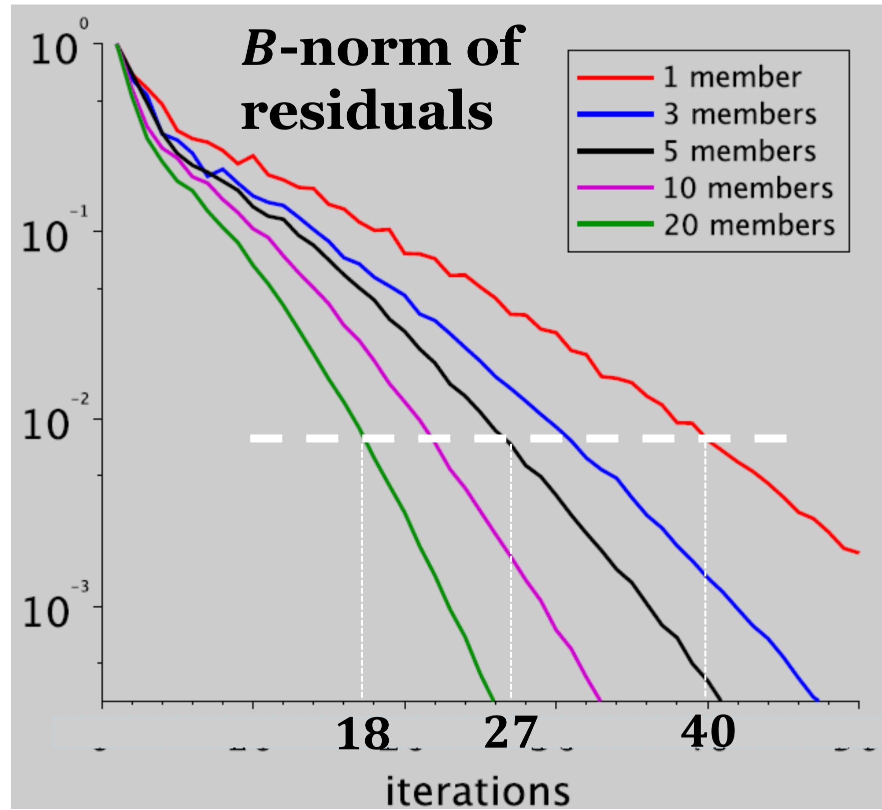




AROME: results in dual space

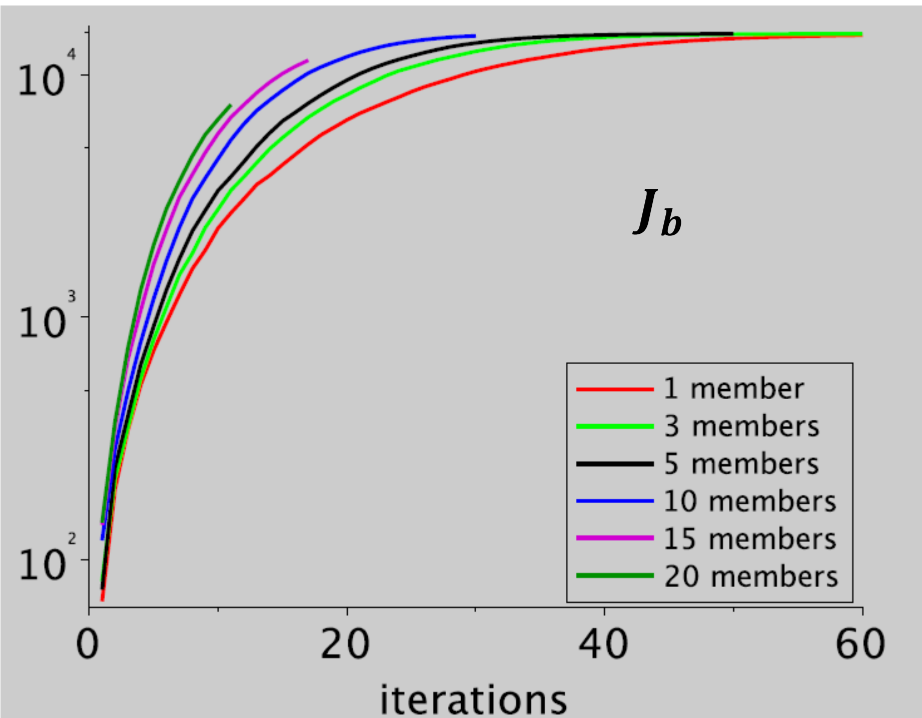
A set of conventional observations (3.10^4, no radar, no sats), dual space

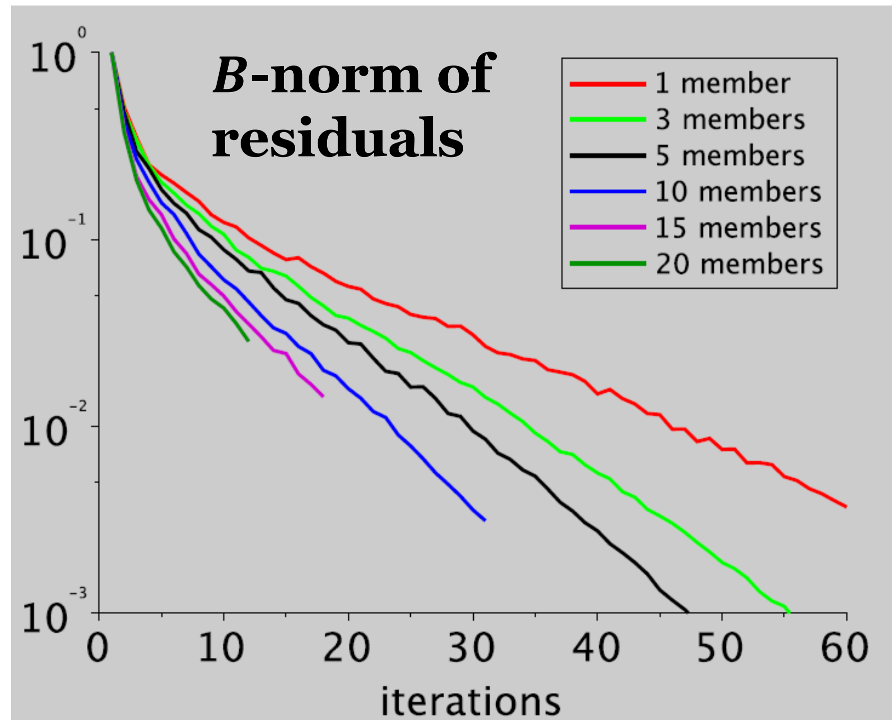




AROME: results in primal space

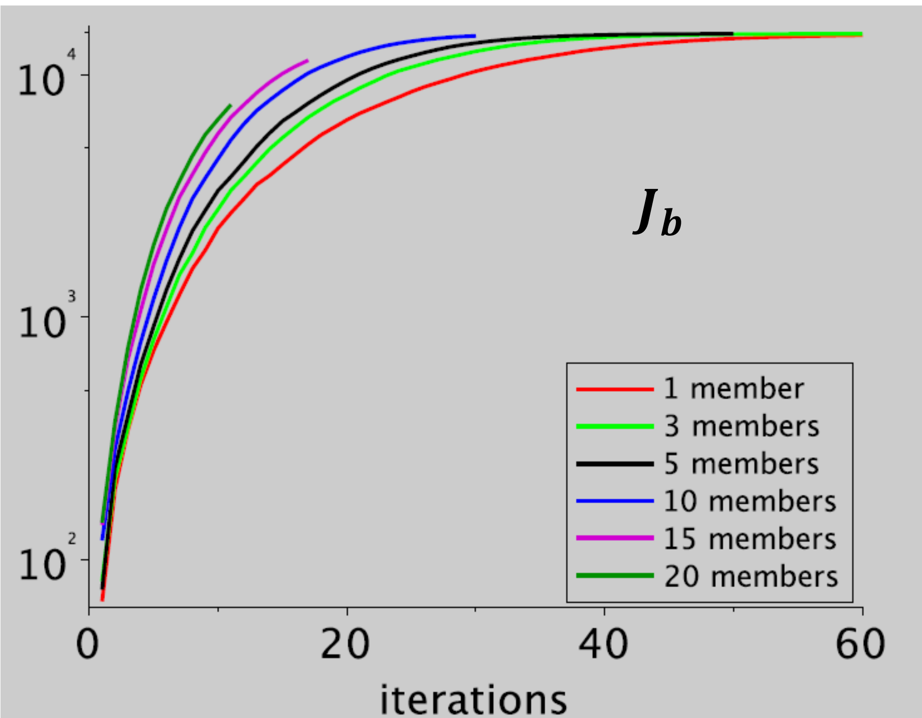
All observations (10^5), primal space

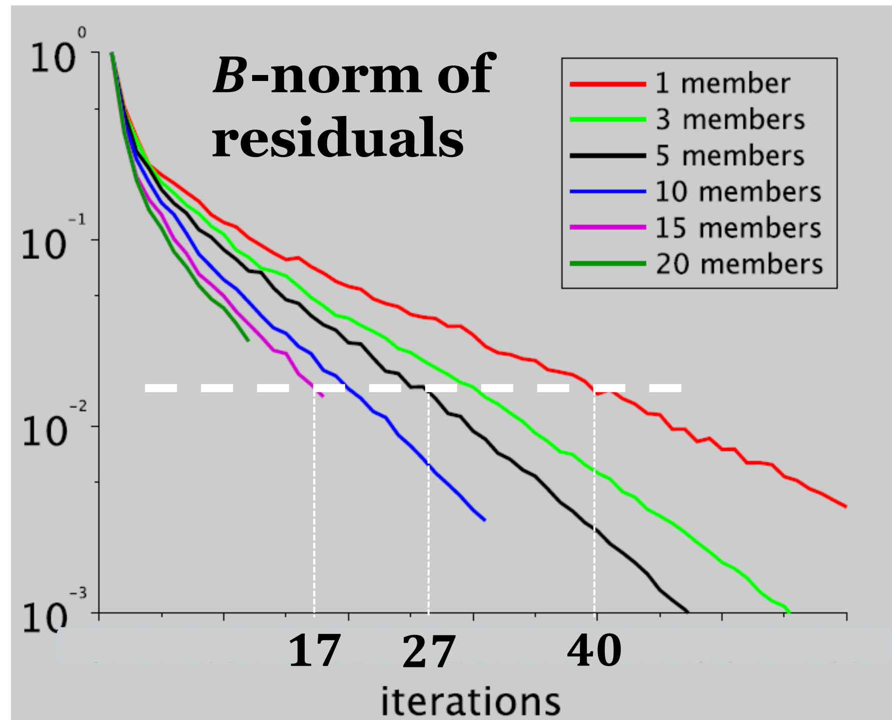




AROME: results in primal space

All observations (10^5), primal space





Conclusions

- Development of block methods for large systems with / without MPI parallelization.
- We have shown the efficiency of dual space algorithms to deal with block problems with high dimensions.

- QG
  - ➢ Block methods speed up the minimization for MPI versions with base storage in primal space.
    - But memory limitations can occur.
  - ➢ Dual block algorithms equivalent but in smaller dimension, no memory limitations.

- AROME
  - ➢ Very good results in terms of convergence.
  - ➢ Work still needed to add the MPI code and to run the dual algorithms with all the observations.