Assimilation diagnostics from an ocean 3D-Var/4D-Var system

Anthony Weaver

CERFACS, Toulouse

With special thanks to

Nicolas Daget, Serge Gratton and Jean Tshimanga (CERFACS)
Magdalena Balmaseda and Kristian Mogensen (ECMWF)

Outline

1. Assimilation diagnostics from a global ocean ensemble 3D-Var system (Daget, Weaver and Balmaseda, QJRMS, 2009)
2. Conclusions (1)
3. Minimization diagnostics from a global ocean 4D-Var system (Tshimanga, Gratton, Weaver and Sartenaer, QJRMS, 2008)
4. Conclusions (2)

Outline

1. Assimilation diagnostics from a global ocean ensemble 3D-Var system (Daget, Weaver and Balmaseda, QJRMS, 2009)

2. Conclusions (1)

3. Minimization diagnostics from a global ocean 4D-Var system (Tshimanga, Gratton, Weaver and Sartenaer, QJRMS, 2008)

4. Conclusions (2)

General context and objective

- An ensemble 3D-Var system was developed for the European project ENSEMBLES to provide multiple ocean analyses for estimating the uncertainty in ocean initial conditions for seasonal forecasts.
- An ensemble data assimilation system provides flow-dependent information on analysis and background error.
  - This information can be exploited to improve the estimate of the background-error covariance matrix ($\mathbf{B}$) on each assimilation cycle.
  - In the ENSEMBLES experiments, we made no attempt to use the ensemble to update $\mathbf{B}$.
- The objective here is explore the possibility of using the ensemble 3D-Var system to improve $\mathbf{B}$.

How to use the ensemble information?

- Construct a low-rank approximation to **B** directly from the sample covariance of the ensemble of model forecast states.
  (Houtekammer and Mitchell 2001; Keppenne and Reinecker 2002; Ott *et al.* 2004; Buehner and Charron 2007; Oke *et al.* 2007).
  - Covariance localization is necessary to minimize spurious effects due to sampling error.

- or -

- Use the ensemble indirectly to define parameters of a (localized) covariance model in a full-rank (operator) representation of **B**.
  (Fisher 2003; Žagar *et al.* 2005; Belo Pereira and Berre 2006; Berre *et al.* 2006; Küçükkaraca and Fisher 2006).
  - A flexible covariance model (inhomogeneous, anisotropic) is required to make best use of the ensemble information.

Objective of this study

- Here, we adopt the covariance model approach.
- In particular, we investigate the potential of an ensemble of ocean states to provide useful flow-dependent estimates of the background-error **variances** in the 3D-Var system.
- This approach will be compared with a simpler approach for incorporating (weak) flow dependence in the variances, based on a parameterization in terms of the background state.
- This study is a first step towards making more comprehensive use of an ensemble for specifying additional parameters of the covariance model.

Basic features of the assimilation system

- The ocean model is a global 2° configuration of OPA8.2 (Madec *et al.* 1998).
- The surface forcing fields are derived from ERA40 (Uppala *et al.* 2005).
- The assimilation method is a multivariate 3D-Var version of the OPAVAR system (Weaver *et al.* 2005).
- First-Guess at Appropriate Time (FGAT) and Incremental Analysis Updates (IAU) are employed.
- The data are quality-controlled temperature and salinity profiles from ENSEMBLES (EN3) data-base (Ingleby and Huddleston 2007).

The 3D-Var FGAT cost function

$$J = \frac{1}{2}\delta\mathbf{w}^{\mathrm{T}}\,\mathbf{B}_{(\mathbf{w})}^{-1}\,\delta\mathbf{w} + \frac{1}{2}(\mathbf{H}\delta\mathbf{w} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{w} - \mathbf{d})$$

where

$$\mathbf{d} = \begin{pmatrix} \mathbf{d}_0 \\ \vdots \\ \mathbf{d}_i \\ \vdots \\ \mathbf{d}_N \end{pmatrix} = \begin{pmatrix} \mathbf{y}_0^o - \mathbf{H}_0\mathbf{w}^b(t_0) \\ \vdots \\ \mathbf{y}_i^o - \mathbf{H}_i\mathbf{w}^b(t_i) \\ \vdots \\ \mathbf{y}_N^o - \mathbf{H}_N\mathbf{w}^b(t_N) \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_i \\ \vdots \\ \mathbf{H}_N \end{pmatrix}.$$

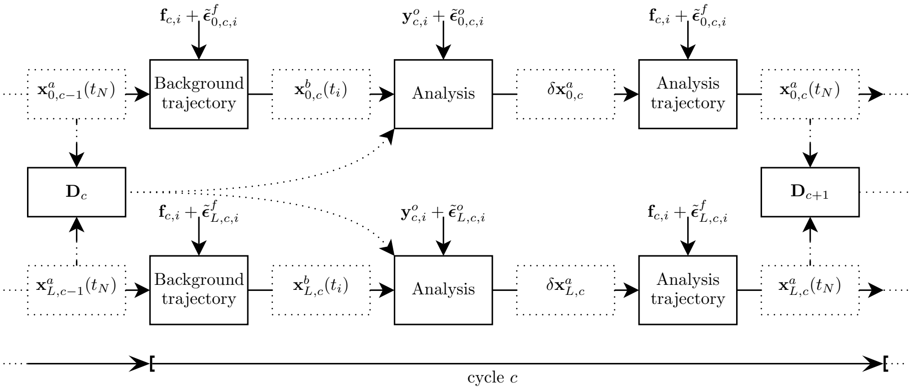
- $\delta\mathbf{w} = (\delta T, \delta S)^{\mathrm{T}}$ is the vector of temperature and salinity increments.
- $\mathbf{y}_i^o = (T_i^o, S_i^o)^{\mathrm{T}}$ is the vector of temperature and salinity observations.
- Increments for sea-surface height and velocity are obtained using balance constraints applied to the analysis increment $\delta\mathbf{w}^a$.

$$\mathbf{B}_{(\mathbf{w})} = \mathbf{K}_{(\mathbf{w})}\,\mathbf{D}^{1/2}_{(\widehat{\mathbf{w}})}\,\mathbf{F}_{(\widehat{\mathbf{w}})}\mathbf{F}^{\mathrm{T}}_{(\widehat{\mathbf{w}})}\,\mathbf{D}^{1/2}_{(\widehat{\mathbf{w}})}\mathbf{K}^{\mathrm{T}}_{(\mathbf{w})}$$

where

$$\mathbf{F}_{(\widehat{\mathbf{w}})} = \begin{pmatrix} \mathbf{F}_{TT} & 0 \\ 0 & \mathbf{F}_{S_U S_U} \end{pmatrix}, \quad \mathbf{D}^{1/2}_{(\widehat{\mathbf{w}})} = \begin{pmatrix} \mathbf{D}^{1/2}_{T} & 0 \\ 0 & \mathbf{D}^{1/2}_{S_U} \end{pmatrix}, \quad \mathbf{K}_{(\mathbf{w})} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{K}_{ST} & \mathbf{I} \end{pmatrix}$$

- $\widehat{\mathbf{w}} = (T, S_U)^{\mathrm{T}}$ where S_U corresponds to “unbalanced” salinity.
- $\mathbf{K}_{(\mathbf{w})}$ is a multivariate balance operator: $\widehat{\mathbf{w}} \mapsto \mathbf{w}$.
- $\mathbf{F}_{(\widehat{\mathbf{w}})}\mathbf{F}^{\mathrm{T}}_{(\widehat{\mathbf{w}})}$ is a quasi-Gaussian 3D univariate correlation operator, modelled using a diffusion operator.
- $\mathbf{D}_{(\widehat{\mathbf{w}})}$ is a variance matrix (for $\widehat{\mathbf{w}}$) whose estimation is the focus of this study.



- The background-error variance matrix ($\mathbf{D}_c$) used for the analysis on cycle c is estimated from the sample variance matrix computed from the ensemble of background states ($\mathbf{x}^b_{l,c}(t_0)$) at the start of cycle c.
- In our set-up, $\mathbf{x}^b_{l,c}(t_0) = \mathbf{x}^a_{l,c-1}(t_N)$.

Estimate σ^b from the difference between background states of successive ensemble members, $l = 0, \ldots, L-1$:

$$\begin{aligned}
\mathsf{D}_{(\widehat{\mathbf{w}})} &= \operatorname{diag}\left\{(\sigma^b_T)^2, (\sigma^b_{S_U})^2\right\} \\
&= \operatorname{diag}\left\{\frac{1}{2(L-1)} \sum_{l=0}^{L-1} \left[\mathsf{K}^{-1}_{(\mathbf{w})}\left(\mathbf{w}^b_l(t_0) - \mathbf{w}^b_{l+1}(t_0)\right)\right]\right. \\
&\qquad \left. \times \left[\mathsf{K}^{-1}_{(\mathbf{w})}\left(\mathbf{w}^b_l(t_0) - \mathbf{w}^b_{l+1}(t_0)\right)\right]^{\mathrm{T}}\right\}
\end{aligned}$$

where $\mathbf{w}^b_L(t_0) = \mathbf{w}^b_0(t_0)$.

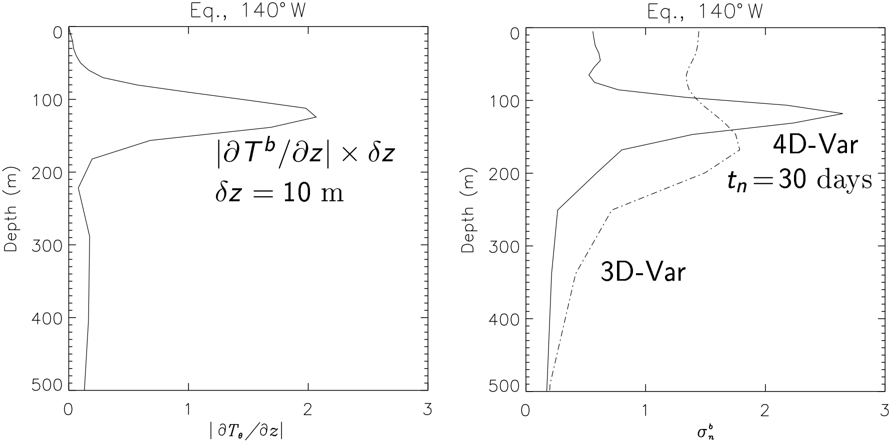
Key features of the ensemble-generation strategy

- A 9-member ensemble.
- The perturbed input parameters:
  - the surface forcing fields (heat flux, fresh-water flux, wind-stress);
  - the temperature and salinity observations;
  - the background state;
  - model error is neglected.
- Construction of the perturbations:
  - the forcing perturbations are derived from differences between different forcing analysis products (Balmaseda *et al.* 2008);
  - the observation perturbations are drawn from a Gaussian pdf with covariance matrix $\mathbf{R}$;
  - the background state is perturbed implicity via the cycling procedure;
- Reduction of sampling error:
  - A 90-day (9-cycle) sliding window is used, giving an effective ensemble size of 81 on each cycle for estimating σ^b.
  - Intraseasonal variability in σ^b is thus filtered out.

- The experimental design follows the common reanalysis procedures used in the ENSEMBLES and ENACT projects (Davey *et al.* 2006).
- The experiments are performed for the 9-year period from 1 January 1993 to 31 December 2001.
- A 10-day assimilation cycle is used.
- The experiments:
  - **CTL** : no data assimilation.
  - **B1R1** : parameterized σ^b, and σ^o defined using globally-averaged estimates from Ingleby and Huddleston (2007).
  - **B1R2** : parameterized σ^b, and σ^o estimated from Fu *et al.* method.
  - **B2R2** : ensemble σ^b, and σ^o estimated from Fu *et al.* method.
- Results will be displayed for temperature only and for the global ocean (results for salinity and in different regions are qualitatively similar).

Parameterizing σ^b in terms of $\partial T^b/\partial z$ makes some sense

(from Weaver *et al.* 2003)

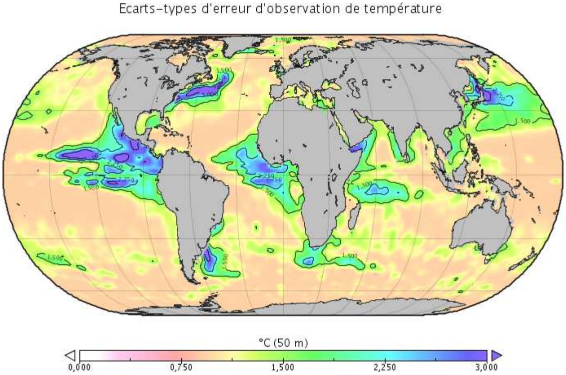


$\mathbf{P}^b(t_n) = \mathbf{B}$ in 3D-Var FGAT

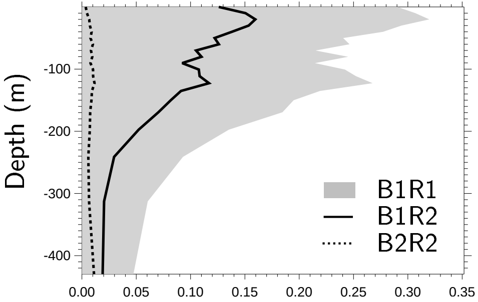
$\mathbf{P}^b(t_n) = \mathbf{M}(t_0, t_n)\,\mathbf{B}\,\mathbf{M}(t_0, t_n)^{\mathrm{T}}$ in 4D-Var (cf. EKF)

σ^o estimated using the Fu *et al.* method

Example of temperature σ^o at 50 m



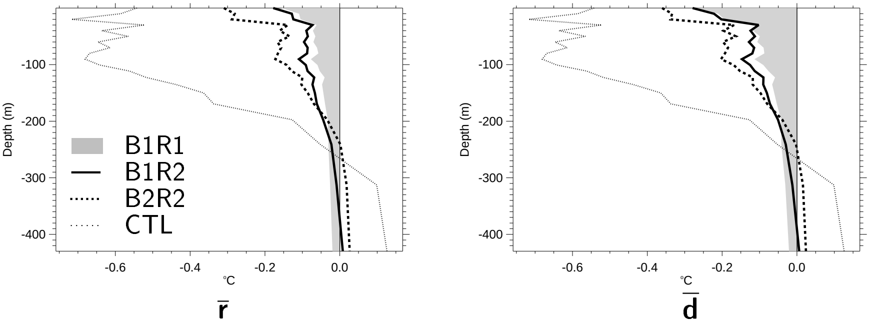
The average innovation "weights" for temperature



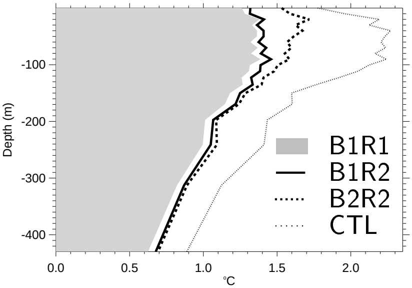
$$w_T = \frac{(\sigma_T^b)^2}{(\sigma_T^b)^2 + (\sigma_T^o)^2}$$

- Neglecting correlations, w_T is the average weight for an innovation.
- Both σ_T^b and σ_T^o have been computed at observation points, and averaged over the 1994-2000 period and the global domain.

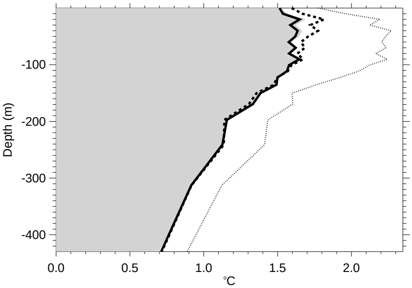
Analysis residual and innovation statistics for temperature



- $\mathbf{r} = \mathbf{d} - \mathbf{H}\delta\mathbf{w}^a$ (residual) and $\mathbf{d} = \mathbf{y}^o - \mathbf{H}\mathbf{w}^b$ (innovation).
- $\overline{\mathbf{z}}$ indicates spatial (global) and temporal (1994-2000) average.
- Mean bias in CTL is reduced substantially in all assimilation expts.



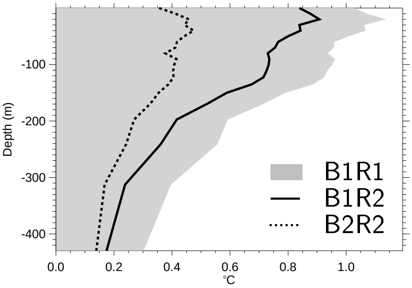
sd(**r**)



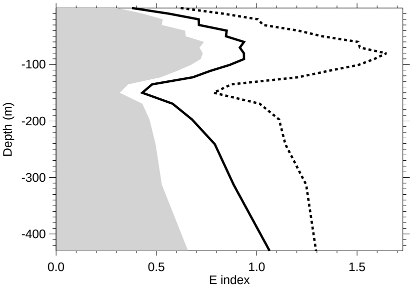
sd(**d**)

- $\mathbf{r} = \mathbf{d} - \mathbf{H}\delta\mathbf{w}^a$ (residual) and $\mathbf{d} = \mathbf{y}^o - \mathbf{H}\mathbf{w}^b$ (innovation).
- $\mathrm{sd}(\mathbf{z}) = \sqrt{\overline{(\mathbf{z} - \bar{\mathbf{z}})^2}}$
- All assimilation expts. improve the fit to the observed variability.
- The "error growth" in the 10-day forecast is smallest for B2R2.

An “efficiency” index



$\text{rms}(\mathbf{H}\delta\mathbf{w}^a)$

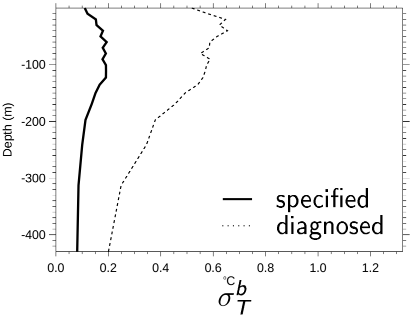


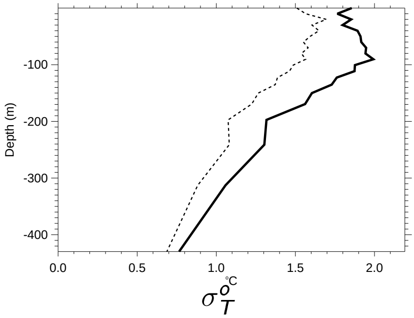
$$\mathrm{E} = \frac{\mathrm{rms}(\mathbf{d}_{\mathrm{CTL}}) - \mathrm{rms}(\mathbf{d})}{\mathrm{rms}(\mathbf{H}\delta\mathbf{w}^a)}$$

- $\mathrm{E} = \dfrac{\text{10-day forecast error from CTL – 10-day forecast error from assim.}}{\text{“work done” by assimilation method to reduce forecast error}}$
- $\mathrm{E} > 0$ ($\mathrm{E} < 0$) implies assimilation is beneficial (detrimental).
- E increases (decreases) if $\mathbf{d}$ or $\delta\mathbf{w}^a$ decreases (increases).

Specified versus diagnosed σ^b and σ^o for temperature in B2R2

(method of Desroziers *et al.* 2005)





- If **B** and **R** are good estimates of the true background- and observation-error covariance matrices then

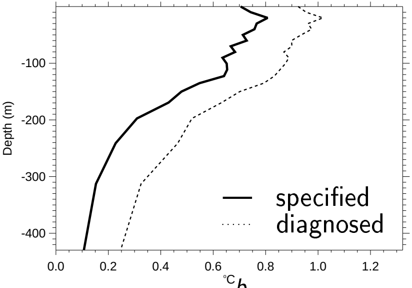
$$E[\mathbf{d}\,(\mathbf{H}\delta\mathbf{w}^a)^{\mathrm{T}}] \approx \mathbf{H}\mathbf{B}_{(\mathbf{w})}\mathbf{H}^{\mathrm{T}}$$

$$E[\mathbf{d}\,(\mathbf{d} - \mathbf{H}\delta\mathbf{w}^a)^{\mathrm{T}}] \approx \mathbf{R}$$

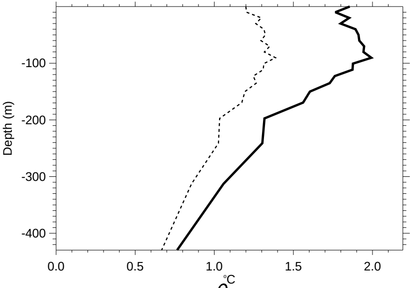
- Here, σ_T^b is underestimated, and σ_T^o is overestimated.

Specified versus diagnosed σ^b and σ^o for temperature in B1R2

(method of Desroziers *et al.* 2005)



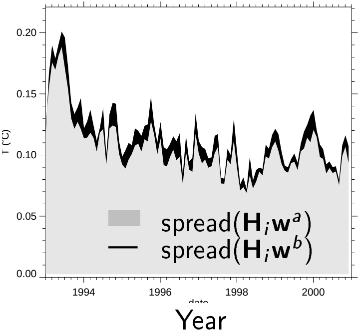
σ_T^b

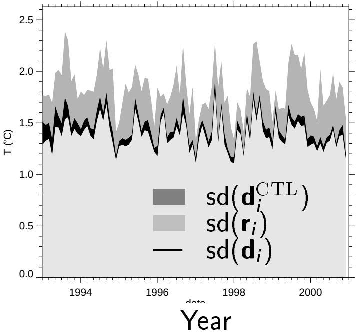


σ_T^o

- σ_T^b is also underestimated (to a lesser extent than in B2R2).
- σ_T^o is also overestimated (to a greater extent than in B2R2).

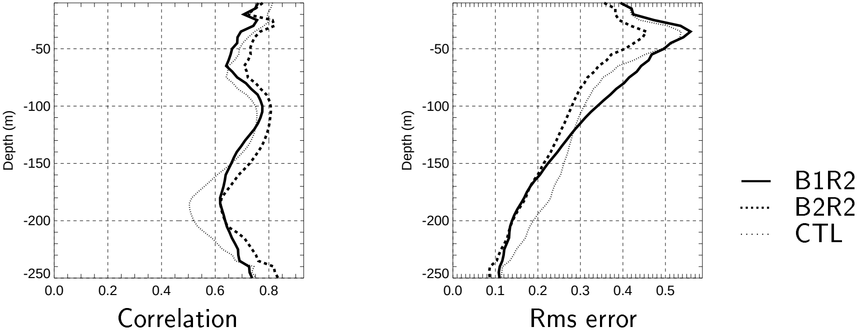
Experiment B2R2





- $\mathrm{spread}\{\mathbf{H}_i\mathbf{w}^{a,b}\} = \sqrt{\frac{1}{L-1}\sum_{l=0}^{L-1}\left(\mathbf{H}_i\mathbf{w}_l^{a,b}(t_i) - \frac{1}{L}\sum_{l=0}^{L-1}\mathbf{H}_i\mathbf{w}_l^{a,b}(t_i)\right)^2}$
- Spread of the analysis < spread of the background.
- No evidence of ensemble collapse.
- Spread($\mathbf{H}_i\mathbf{w}^{a,b}$) is approximately a factor 10 smaller than sd($\mathbf{r}_i$), sd($\mathbf{d}_i$).

Example from the eastern Pacific (110°W)



- B2R2 outperforms B1R2 (and B1R1) at all moorings.
- B2R2 outperforms CTL in the central and eastern Pacific, but slightly worse in the western Pacific.

Outline

Conclusions (1)

- Both the parameterized and ensemble σ^b formulations produce a significant reduction in the rms of the innovations (compared to the control), with the parameterized σ^b slightly better above 150 m.
- Evidence that the ensemble σ^b analyses are better "balanced".
  - Reduced error growth between cycles.
  - Smaller analysis increments.
  - Closer to independent data (sea-level anomalies from T/P and current-meter data from TAO).
- Desroziers *et al.* statistics suggest that the ensemble σ^b are underestimated.
  - The parameterized σ^b are also underestimated but to a lesser extent.
- The apparent underestimation of the ensemble spread points to the need to improve the ensemble generation strategy.
  - Simple inflation techniques did not give satisfactory results.

Outline

The numerical experiments

- An incremental 4D-Var version of the OPAVAR system.
- Global 2^o configuration.
- Same resolution in the outer and inner loops.
- Tangent-linear model with simplified vertical mixing and simplified isopycnal diffusion.
- Assimilation of temperature and salinity profiles.
- A single 10-day cycle (Jan. 1-10, 1993).
- No. of control variables $\sim 1.7 \times 10^6$; no. of observations $\sim 1.4 \times 10^5$.
- 3 outer iterations with 10 inner iterations per outer iteration.
- Inner-loop minimization done using a close variant of the CONGRAD routine (Fisher 1998).
- CONGRAD is a Lanczos implementation of a $\mathbf{B}$-preconditioned conjugate gradient algorithm.

The incremental 4D-Var cost function

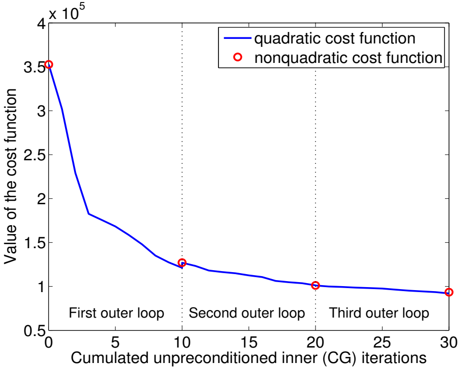
$$J = \frac{1}{2}\delta\mathbf{w}^{\mathrm{T}}\,\mathbf{B}_{(\mathbf{w})}^{-1}\,\delta\mathbf{w} + \frac{1}{2}(\mathbf{H}\delta\mathbf{w} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{w} - \mathbf{d})$$

where

$$\mathbf{d} = \begin{pmatrix} \mathbf{d}_0 \\ \vdots \\ \mathbf{d}_i \\ \vdots \\ \mathbf{d}_N \end{pmatrix} = \begin{pmatrix} \mathbf{y}_0^o - \mathbf{H}_0\mathbf{w}^b(t_0) \\ \vdots \\ \mathbf{y}_i^o - \mathbf{H}_i\mathbf{w}^b(t_i) \\ \vdots \\ \mathbf{y}_N^o - \mathbf{H}_N\mathbf{w}^b(t_N) \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_i\mathbf{M}(t_0, t_i) \\ \vdots \\ \mathbf{H}_N\mathbf{M}(t_0, t_N) \end{pmatrix}.$$

- $\delta\mathbf{w} = (\delta T, \delta S, \delta\eta, \delta u, \delta v)^{\mathrm{T}}$ is the vector of temperature, salinity, SSH and velocity increments.
- $\mathbf{y}_i^o = (T_i^o, S_i^o)^{\mathrm{T}}$ is the vector of temperatpure and salinity observations.
- Direct initialization (not IAU) and outer iterations.

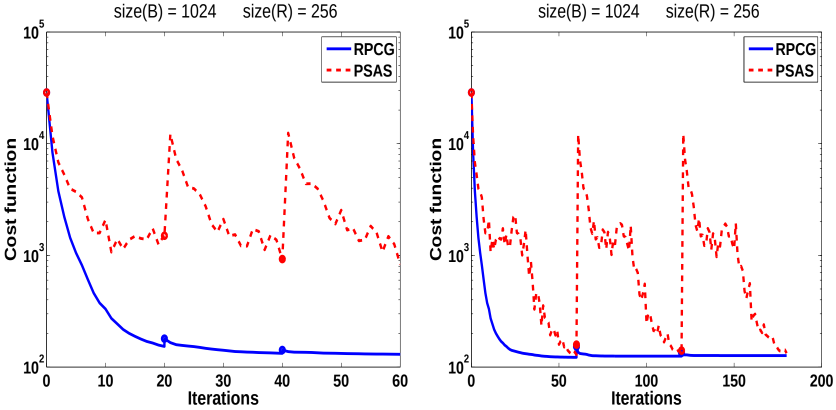
Monitoring the "jumps" on outer iterations (1)



- The jumps on outer iterations give an indication of the accuracy of the linear approximation.
- Largest jump between 1st and 2nd outer iterations (rel. error $\sim 4.5\%$).

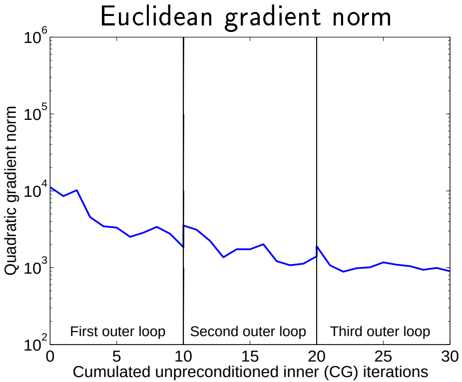
Monitoring the "jumps" on outer iterations (2)

Comparison of 4D-PSAS and 4D-Var in a "toy" problem
(from Gratton and Tshimanga (2009), submitted to QJRMS)



- The "jumps" can be particularly problematic in PSAS if the inner-loop minimization is stopped before full convergence.

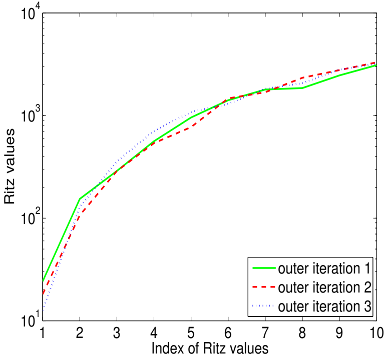
Convergence criterion for the inner loop



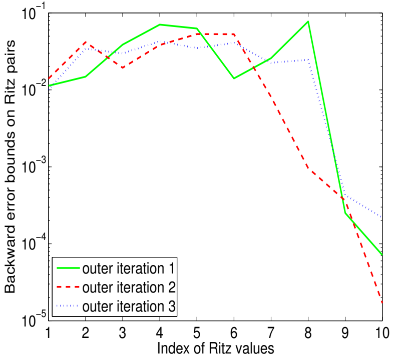
- The Euclidean gradient norm is not necessarily a good measure of convergence of the CG minimization.
- Convergence diagnostics that decrease monotonically are preferable.
  - ▷ Relative reduction in quadratic cost.
  - ▷ Gradient norm based on the inverse Hessian metric.

Approximate Hessian eigenvalues/vectors from the Lanczos algorithm

Approx. eigenvalues (Ritz values)

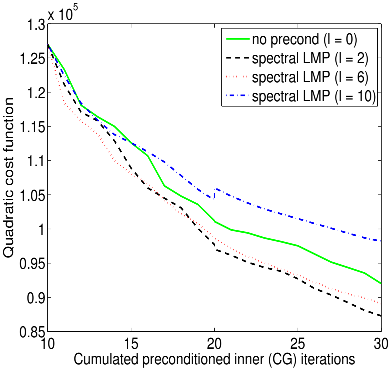


Backward error $||\Delta \mathbf{A}||/||\mathbf{A}||$



- Here the Ritz values are similar between outer iterations $k = 1, 2$ and 3.
  - ▷ Ritz pairs from $k < K$ can be used to precondition iterations $k \geq K$.
- The largest Ritz value is the most accurate (error $\sim 10^{-3}$–10^{-5}).
  - ▷ Can be used to provide a good estimate of the condition number.
- Most of the other Ritz values are much less accurate (error$\sim 10^{-1}$–10^{-2}).
  - ▷ Caution when using these Ritz pairs in spectral preconditioners.

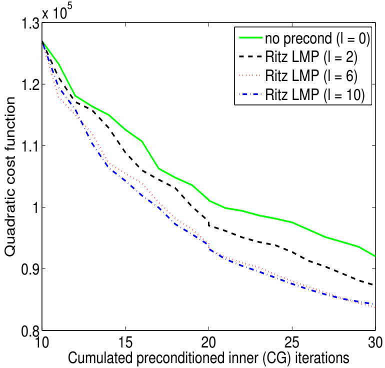
The spectral limited-memory preconditioner (Fisher 1998)



$$\mathsf{K}_l^{\text{spectral}} = \mathbf{I}_n + \sum_{i=1}^{l} \left(\frac{1}{\theta_i} - 1 \right) \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}} \quad \text{where} \quad \mathbf{A}\mathbf{z}_i \approx \theta_i \mathbf{z}_i$$

- Inaccurate eigenpairs $(\theta_i, \mathbf{z}_i)$ (Ritz pairs) used in the spectral LMP can be worse than no preconditioning at all.

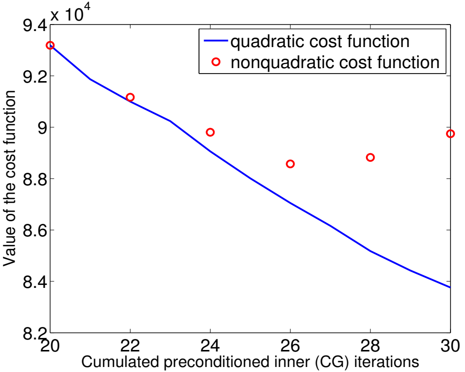
The Ritz limited-memory preconditioner (Tshimanga *et al.* 2008)



$$\mathbf{K}_l^{\mathrm{Ritz}} = \left(\mathbf{I}_n - \sum_{i=1}^{l} \frac{\mathbf{z}_i \mathbf{z}_i^{\mathrm{T}}}{\theta_i} \mathbf{A}\right) \left(\mathbf{I}_n - \sum_{i=1}^{l} \mathbf{A} \frac{\mathbf{z}_i \mathbf{z}_i^{\mathrm{T}}}{\theta_i}\right) + \sum_{i=1}^{l} \frac{\mathbf{z}_i \mathbf{z}_i^{\mathrm{T}}}{\theta_i}$$

- A “good” preconditioner for (**A**-conjugate) Ritz vectors as well as exact eigenvectors (a “stablized” spectral LMP).
- A more accurate formula (with Ritz vectors) for computing analysis self-sensitivities (Cardinali *et al.* 2004).

Monitoring the behaviour of the non-quadratic cost function



- In incremental 4D-Var the value of the non-quadratic cost function is only computed at the outer-loop end-points.
- Here they are diagnosed at *intermediate* points (expensive!).
- Divergence in the non-quadratic cost occurs after 6 inner iterations (on the 3rd outer loop).

Outline

1. Assimilation diagnostics from a global ocean ensemble 3D-Var system (Daget, Weaver and Balmaseda, QJRMS, 2009)
2. Conclusions (1)
3. Minimization diagnostics from a global ocean 4D-Var system (Tshimanga, Gratton, Weaver and Sartenaer, QJRMS, 2008)
4. Conclusions (2)

Conclusions (2)

- The inner-loop minimization requires an appropriate stopping criterion (in 3D-Var FGAT as well as 4D-Var).
  - The Euclidean gradient norm is not a robust measure of convergence.
  - Beware of "jumps" and divergence on the outer loop (see also Trémolet 2007).

- For a fixed number of outer iterations, the optimal number of inner iterations (per outer iteration) can be diagnosed *a priori*.
  - Requires multiple cost function evaluations on the outer-loop $\rightarrow$ Very expensive!
  - Periodic tuning of the number of inner iterations would be more practical.
  - Results will depend on the preconditioner as well as the characteristics of the problem.