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

Anthony Weaver

CERFACS, 42 avenue Gaspard Coriolis
31057 Toulouse, France
weaver@cerfacs.fr

ABSTRACT

Examples of diagnostics used to assess the performance of a variational data assimilation system for a global ocean model are presented. Two classes of diagnostics are considered. The first class consists of assimilation statistics based on information from innovations and analysis increments obtained from multi-cycle 3D-Var experiments. In particular, these diagnostics are used to evaluate the performance of an ensemble version of the 3D-Var system. The second class consists of diagnostics based on information from the minimization algorithm. These diagnostics are used to evaluate the convergence characteristics of a single-cycle incremental 4D-Var experiment. While the focus is on ocean data assimilation, most of the diagnostics presented here are applicable in atmospheric data assimilation as well.

1 Introduction

Several diagnostics have been developed to assess the performance of a variational data assimilation system (Weaver *et al.* 2003; Weaver *et al.* 2005; Daget *et al.* 2009) for a global version of the OPA model (Madec *et al.* 1998). Two versions of the assimilation system have been developed. One version produces an ensemble of multivariate 3D-Var analyses. This version of the system was developed and applied for multi-year reanalysis as part of the ENSEMBLES project. The second version of the system is based on an incremental 4D-Var method. This version of the system was developed mainly for short-duration research applications.

Two classes of diagnostics are illustrated in this paper. The first class is designed to evaluate the statistical properties of the ensemble 3D-Var system. These diagnostics exploit information from innovations and analysis increments obtained from multiple 3D-Var cycles. The second class of diagnostics exploits information from the minimization algorithm and are used for assessing the convergence properties of incremental 4D-Var system. A more comprehensive discussion of the diagnostics presented here can be found in the recent articles by Daget *et al.* (2009) and Tshimanga *et al.* (2008).

2 Assimilation statistics from an ensemble 3D-Var FGAT system

The ensemble data assimilation system is based on a multivariate 3D-Var FGAT (Weaver *et al.* 2003, 2005; Daget *et al.* 2009). The 3D-Var analysis increment $\delta\mathbf{w}^{\mathrm{a}}$ is the minimizing solution of the quadratic cost function

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

where

$$\mathbf{d} = \begin{pmatrix} \mathbf{d}_0 \\ \vdots \\ \mathbf{d}_i \\ \vdots \\ \mathbf{d}_N \end{pmatrix} = \begin{pmatrix} \mathbf{y}_0^{\mathrm{o}} - \mathbf{H}_0 \mathbf{w}^{\mathrm{b}}(t_0) \\ \vdots \\ \mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i \mathbf{w}^{\mathrm{b}}(t_i) \\ \vdots \\ \mathbf{y}_N^{\mathrm{o}} - \mathbf{H}_N \mathbf{w}^{\mathrm{b}}(t_N) \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_i \\ \vdots \\ \mathbf{H}_N \end{pmatrix}. \tag{2}$$

A separate 3D-Var analysis is performed for each ensemble member. Incremental Analysis Updates (Bloom *et al.* 1996) is used to initialize the model with the analysis increment $\delta\mathbf{w}^{\mathrm{a}}$. The observation vector $\mathbf{y}_i^{\mathrm{o}}$ consists of temperature and salinity profiles from the ENSEMBLES quality-controlled data-set EN3 (Ingleby and Huddleston 2007). A 10-day assimilation window is used. Perturbations are added to these observations to construct the ensemble. The perturbations are drawn from a Gaussian distribution with covariance matrix equal to the diagonal observation-error covariance matrix $\mathbf{R}$ used in the assimilation system. For each ensemble member, the background state $\mathbf{w}^{\mathrm{b}}(t_i)$ is obtained by integrating the model from t_0 to t_i from the appropriate (perturbed) background initial state $\mathbf{x}^{\mathrm{b}}(t_0)$. A different set of surface forcing fields was used for each ensemble member. The observation operators $\mathbf{H}_i$ are 3D interpolation operators. The background-error covariance matrix is non-diagonal and modelled implicitly via a preconditioning transformation (Weaver *et al.* 2005). In the ensemble system, the background-error variances are updated on each 10-day assimilation cycle based on the spread of the background states at that time. Nine ensemble members were computed on each assimilation cycle. In order to increase the sample size for estimating the $\mathbf{B}$ variances, a sliding window was used to include the ensemble of background states from the previous 9 cycles, thus giving an effective ensemble size of 81.

Diagnostics are presented from four experiments that were designed to test the sensitivity of the analyses to different background- and observation-error standard deviation (σ^{b} and σ^{o}) formulations. The test period covers the years 1993–2000. The control experiment (CTL) is one in which no data are assimilated. Experiment B1R1 uses simple parameterized formulations of σ^{b} and σ^{o}. Experiment B1R2 uses the parameterized σ^{b}, and σ^{o} values estimated from the CTL experiment using the method of Fu *et al.* (1993). Finally, experiment B2R2 uses the σ^{b} estimated from the ensemble, and the σ^{o} from the Fu *et al.* method. The different ensemble members of B2R2 produced statistically similar results and thus only the results from the unperturbed member B2R2 are presented. The diagnostics focus on temperature.

2.1 The innovation "weights"

Figure 1 shows the ratio $(\sigma^{\mathrm{b}})^2[(\sigma^{\mathrm{b}})^2 + (\sigma^{\mathrm{o}})^2]^{-1}$ where $(\sigma^{\mathrm{o}})^2$ and $(\sigma^{\mathrm{b}})^2$ are the average observation- and background-error variances for temperature. The spatial averaging has been performed over the global region and within the vertical model grid cells, and the time averaging has been performed over the 1994-2000 period. The background-error variances have been evaluated at observation points before averaging. Ignoring correlations, this ratio indicates the average weight given to an innovation at a particular depth in determining the analysis increment. For the ensemble-σ^{b} experiment (B2R2), the weights are noticeably smaller and more uniform with depth compared to those from the parameterized-σ^{b} experiments (B1R1 and B1R2). As a result, the analysis on each cycle of B2R2 will tend to remain closer to the background state than it will in either B1R1 or B1R2 which will tend to pull it more to the observations, especially in the upper 200 m.

2.2 Residual and innovation vector statistics

The innovation vector $\mathbf{d}$ and analysis increment $\delta\mathbf{w}^{\mathrm{a}}$ provide valuable information for assessing the statistical performance and internal consistency of the assimilation system (Desroziers *et al.* 2005).

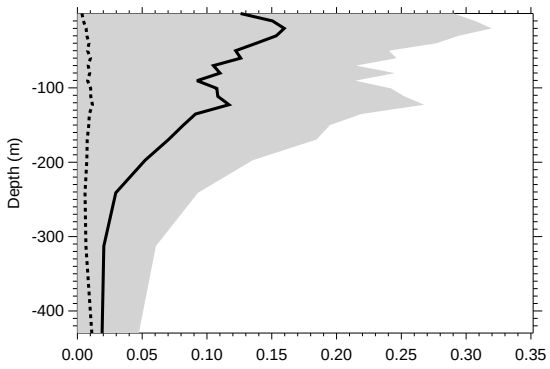
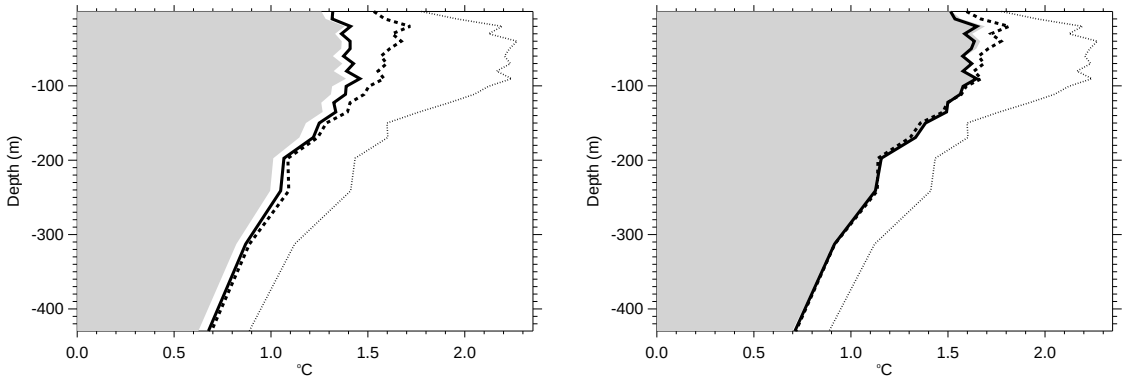


Figure 1: Vertical profiles of the ratio $(\sigma^{\mathrm{b}})^2[(\sigma^{\mathrm{b}})^2+(\sigma^{\mathrm{o}})^2]^{-1}$ *for temperature in B1R1 (grey shaded areas), B1R2 (solid curves) and B2R2 (dashed curves).*

Figure 2 shows vertical profiles of the standard deviation (sd) of the innovation vector ($\mathbf{d}$) and residual vector ($\mathbf{r}=\mathbf{d}-\mathbf{H}\delta\mathbf{w}^{\mathrm{a}}$):

$$\mathrm{sd}(\mathbf{z})=\sqrt{\overline{(\mathbf{z}-\overline{\mathbf{z}})^2}} \tag{3}$$

where $\mathbf{z}=\mathbf{d}$ or $\mathbf{r}$, and the overbar indicates spatial average over the globe and within vertical model grid cells, and temporal average over the 1994–2000 period. The standard deviation indicates how well the model fits the observed temporal and spatial variability. Experiment CTL exhibits large temperature errors, particularly in the upper 150 m where signals associated with seasonal and interannual variability are largest. Relative to CTL, all assimilation experiments improve the fit to the observed temperature variability at all depths. Differences between B1R1 and B1R2 are small (shaded and solid curves). The differences arising from using the ensemble σ^{b} (B2R2; dashed curves) are larger, with sd($\mathbf{d}$) being increased relative to those in B1R1 and B2R2, especially in the upper 100 m.



(a) residuals (b) innovations

Figure 2: Vertical profiles of the standard deviation of the analysis residuals ($\mathbf{r}=\mathbf{d}-\mathbf{H}\delta\mathbf{w}^{\mathrm{a}}$*; left panel) and innovations (*$\mathbf{d}$*; right panel) for temperature for CTL (thin dotted curves), B1R1 (grey shaded areas), B1R2 (solid curves) and B2R2 (dashed curves). Values have been averaged onto model levels. For CTL the innovation and residual are identical (*$\delta\mathbf{w}^{\mathrm{a}}=0$*).*

2.3 An efficiency index

At first sight it appears that the use of the ensemble σ^{b} has slightly degraded the performance of the assimilation system. Closer inspection of Fig. 2, however, reveals that while the innovations are larger in B2R2, the difference between the residuals and innovations is smaller than in B1R1 and B1R2, particularly in the upper 100 m where the difference is 0.1°C smaller. This result indicates that the error growth in a 10-day forecast cycle is smaller in B2R2 than in B1R1 and B1R2, which in turn suggests that the analyses in B2R2 are better balanced. Since the error growth is compensated by the assimilation increment, a smaller error growth in B2R2 should be indicative of smaller increments in B2R2. This is confirmed by Fig. 3a which displays the vertical profiles of the root-mean-square (rms) of the temperature analysis increments at observation points ($\mathrm{rms}(\mathbf{H}\delta\mathbf{w}^{\mathrm{a}})$) from the different experiments. The increments are smallest in B2R2 and largest in B1R1.

The smaller analysis increments in B2R2 could also be an indication that the assimilation system is under-affected by the observations. It is instructive therefore to compare the rms of the analysis increments with the actual 10-day forecast improvement as measured by the innovations. To do so, we define an "efficiency" (E) index,

$$\mathrm{E} = \frac{\mathrm{rms}(\mathbf{d}_c) - \mathrm{rms}(\mathbf{d})}{\mathrm{rms}(\mathbf{H}\delta\mathbf{w}^{\mathrm{a}})}, \tag{4}$$

which measures the ratio of the difference between the rms of the 10-day forecast error from the control and from the assimilation experiment, to the "work done" by the assimilation method (at observation points) to reduce the forecast error. Small (large) innovations and increments will act to increase (decrease) the E index. For example, one system will be more efficent than another (have a larger E value) if it can achieve, on average, a similar reduction in the innovations but with smaller increments. Positive (negative) values of the E index imply that the assimilation is beneficial (detrimental) to the model. Note that the index depends on the forecast lead-time (which influences the numerator in Eq. (4)) as well as the width of the assimilation window (which influences the denominator in Eq. (4)). Therefore, the E index cannot be used to compare experiments with different assimilation windows. Here, the forecast lead-time and assimilation window width are both equal to ten days in all experiments. Note also that the E index is defined for any assimilation experiment that is affected by observations at least once, so the denominator is always non-zero. Vertical profiles of the E index for the three assimilation experiments are shown in Fig. 3b. The E index is positive at all depths for all experiments, with highest values obtained in B2R2 and lowest values in B1R1. The E index is largest at the mean level of the thermocline (100 m). In all experiments, there is a decrease in the temperature E index near the surface. This is related to the strong SST relaxation term used in both CTL and the assimilation experiments, which acts to reduce the value of the numerator in Eq. (4).

2.4 Desroziers diagnostics

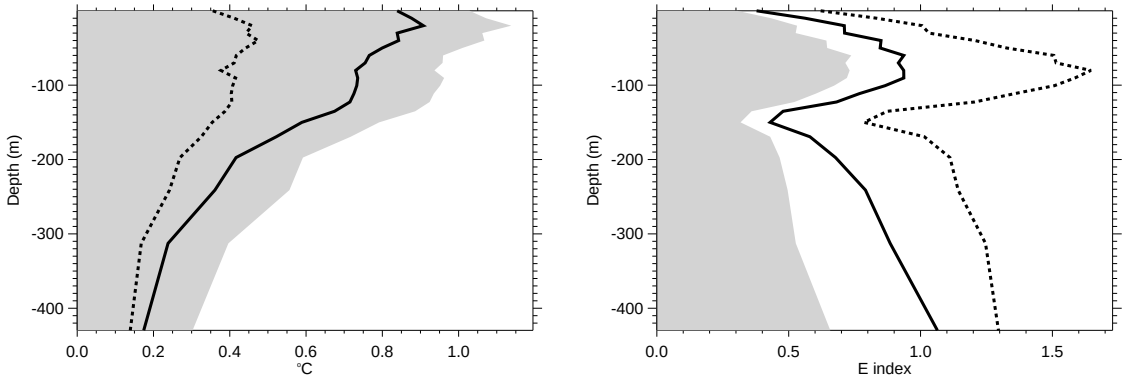
The difficulty in defining background- and observation-error statistics means that they are likely to be incorrectly specified in a practical data assimilation system. Desroziers *et al.* (2005) discuss how the innovations and analysis increments generated by a data assimilation system can be used to diagnose *a posteriori* the covariances of observation error and background error in observation space. Assuming that the background and observation errors are mutually uncorrelated, and that their covariance matrices are good approximations to the true error covariance matrices, then

$$E[\mathbf{d}\mathbf{d}^{\mathrm{T}}] \approx \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R}, \tag{5}$$

$$E[\mathbf{d}\,(\mathbf{H}\delta\mathbf{w}^{\mathrm{a}})^{\mathrm{T}}] \approx \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}, \tag{6}$$

and

$$E[\mathbf{d}\,(\mathbf{d} - \mathbf{H}\delta\mathbf{w}^{\mathrm{a}})^{\mathrm{T}}] \approx \mathbf{R}. \tag{7}$$



(a) Rms of the analysis increments

(b) The efficiency index

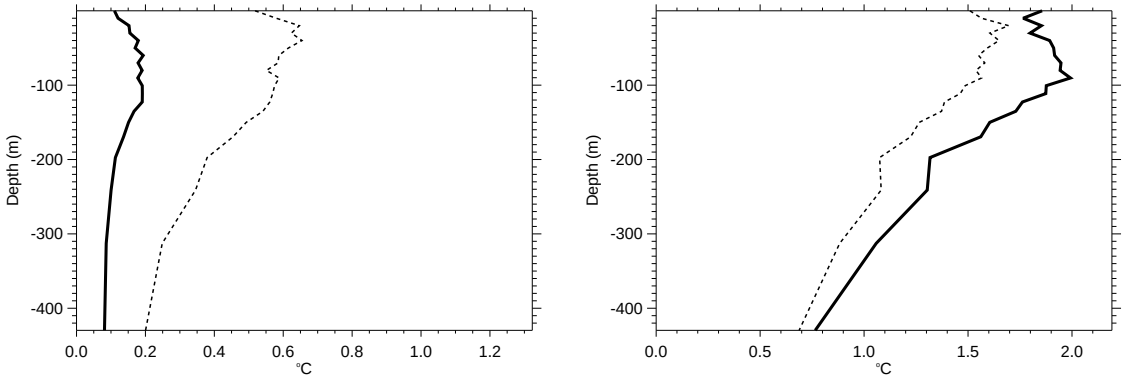
*Figure 3: Vertical profiles of the rms of the temperature assimilation increments at observation points (*rms($\mathbf{H}\delta\mathbf{w}^{\mathrm{a}}$)*; left panel) and of the efficiency index (Equation (4); right panel) for temperature in B1R1 (grey shaded areas), B1R2 (solid curves) and B2R2 (dashed curves). Values have been averaged onto model levels.*

The left-hand sides of Eqs. (6) and (7) can be estimated using statistics from the assimilation system, while the right-hand sides of these equations are the specified covariance matrices presented earlier. These expressions are used to check the consistency of the specified *standard deviations* (σ^{b} and σ^{o}) with those diagnosed using assimilation statistics.

Figure 4 shows vertical profiles from B2R2 of the specified σ^{b} and σ^{o} (solid curves) and the diagnosed σ^{b} and σ^{o} (dashed curves) estimated from Eqs. (6) and (7) using the innovation and analysis increments from all cycles between 1994–2000. In B2R2, the specified σ^{b} are everywhere *underestimated* compared to the diagnosed values (Fig. 4a) whereas the specified σ^{o} are everywhere *overestimated* compared to the diagnosed values. Compared to B2R2, there is better consistency between the diagnosed and specified σ^{b} in B1R2 (figure not shown), although this seems to be achieved at the expense of degrading the consistency between the diagnosed and specified σ^{o}. The results in Fig. 4 suggest that the ensemble 3D-Var system produces background (and analysis) perturbations with inadequate spread on a global average. The apparent overestimation of σ^{o}, on the other hand, points to limitations in our simple model of the observation-error covariances, which ignores spatial and temporal correlations and employs flow-independent variance estimates derived from a method that is itself subject to assumptions of questionable validity. Although Eq. (7) is used purely for diagnostic purposes in this study, it provides the basis of an iterative algorithm for calibrating σ^{o} using the innovations and analysis increments generated by the assimilation system (Desroziers *et al.* 2005). In a similar way, the Desroziers diagnostics can be used to calibrate observation-space values of σ^{b}. Unlike σ^{o}, however, these are not direct inputs to the ensemble 3D-Var system. How best to use the Desroziers diagnostics to improve the estimates of σ^{b} in the space of the analysis variables and how to combine this information effectively with ensemble estimates of σ^{b} are open questions.

2.5 Temporal variability of ensemble and assimilation statistics

The results presented so far have highlighted time-averaged aspects of the assimilation performance. Here some time-varying aspects will now be evaluated, focusing on results from the ensemble experiment B2R2. Figures 5a and b show time-series of the 1993–2000 ensemble *spread* (the square root of the ensemble variance) of the observation-space analysis $\mathbf{H}_i\mathbf{w}_l^{\mathrm{a}}(t_i)$ (light grey shade) and background



(a) Specified and diagnosed σ^{b}

(b) Specified and diagnosed σ^{o}

Figure 4: Vertical profiles of the temperature σ^{b} (left panel) and σ^{o} (right panel) in B2R2. Solid curves correspond to the σ^{b} and σ^{o} that were specified in the assimilation experiment; dashed curves correspond to the σ^{b} and σ^{o} that were diagnosed a posteriori using Eqs. (6) and (7).

$\mathbf{H}_i \mathbf{w}_l^{\mathrm{b}}(t_i)$ (black shade), computed with respect to all ($L = 9$) ensemble members:

$$\mathrm{spread}\{\mathbf{H}_i \mathbf{w}^{\mathrm{a,b}}\} = \sqrt{\overline{\frac{1}{L-1}\sum_{l=0}^{L-1}\left(\mathbf{H}_i \mathbf{w}_l^{\mathrm{a,b}}(t_i) - \frac{1}{L}\sum_{l=0}^{L-1}\mathbf{H}_i \mathbf{w}_l^{\mathrm{a,b}}(t_i)\right)^2}} \tag{8}$$

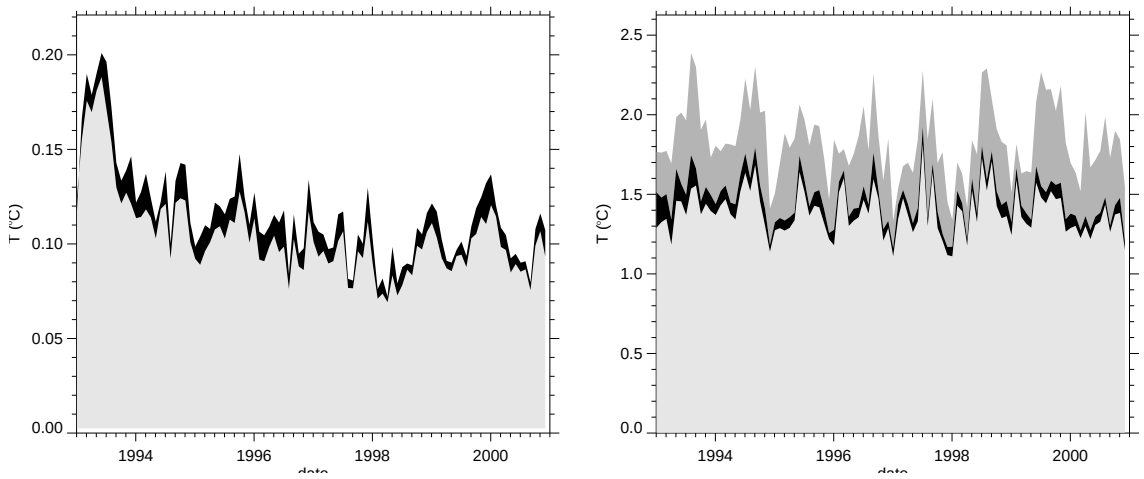
where the overbar indicates spatial average over the globe and within vertical model grid cells, and temporal average over 30-day intervals. A well-defined ensemble should have a spread characteristic of the actual uncertainty in the model state. Figure 5a shows that the temperature spread is systematically smaller in the analysis than in the background, as one would expect. The spread appears to stabilize around a mean value of 0.1°C, after an initial increase during the first 6 months of the experiment. In other words, there is no evidence of ensemble collapse. The decrease in the spread from mid-1993 onwards corresponds to the time when the parameterized σ^{b} are replaced with the ensemble σ^{b}.

Figure 5b shows corresponding time-series of the sd($\tilde{\mathbf{r}}_i$) (light grey shade) and innovation sd($\mathbf{d}_i$) (black shade) of the unperturbed ensemble member $l = 0$, as given by Eq. (3) but with the temporal averaging operator defined as in Eq. (8). Both sd($\tilde{\mathbf{r}}_i$) and sd($\mathbf{d}_i$) are about one order of magnitude larger than the analysis spread of the (observation space) analysis and background (Fig. 5a). The spread of the background state at observation points roughly corresponds to the prescribed values of σ^{b} at observation points. The magnitude of sd($\mathbf{d}_i$) is at all times comparable to that of the mean σ^{o} (not shown), which is consistent with Eq. (5) in view of the relatively small ensemble spread that defines σ^{b}. Despite the small spread, sd($\mathbf{d}_i$) of B2R2 is consistently much smaller than sd($\mathbf{d}_i$) of CTL (dark grey shade).

3 Minimization diagnostics from an incremental 4D-Var system

This section presents diagnostics to assess the convergence properties of an incremental 4D-Var version of the ocean assimilation system. The incremental algorithm approximately solves a (non-quadratic) minimization problem with nonlinear constraints by minimizing a sequence ($k = 1, \ldots, K$) of quadratic cost functions

$$J[\delta\mathbf{w}^{(k)}] = \frac{1}{2}\left(\delta\mathbf{w}^{(k)}\right)^{\mathrm{T}} \mathbf{B}^{-1}\,\delta\mathbf{w}^{(k)} + \frac{1}{2}\left(\mathbf{H}^{(k-1)}\delta\mathbf{w}^{(k)} - \mathbf{d}^{(k-1)}\right)^{\mathrm{T}} \mathbf{R}^{-1}\left(\mathbf{H}^{(k-1)}\delta\mathbf{w}^{(k)} - \mathbf{d}^{(k-1)}\right) \tag{9}$$



(a) Ensemble spread (b) Standard deviation of the residuals and innovations

Figure 5: Left panel: 1993–2000 time-series of the ensemble spread in temperature at observation points (upper panels) for the background, $\mathbf{H}_i\mathbf{w}^{\mathrm{b}}(t_i)$ (black shaded area), and analysis, $\mathbf{H}_i\mathbf{w}^{\mathrm{a}}(t_i)$ (light grey shaded area), in B2R2. Right panel: 1993–2000 time-series of the standard deviation of the temperature innovation vector, sd($\mathbf{d}_i$) (black shaded area), and of the residuals, sd($\tilde{\mathbf{r}}_i$) (light grey shaded area), with $\tilde{\mathbf{r}} = \mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{w}^{\mathrm{a}}$, in B2R2. The standard deviation of the innovation in CTL (dark grey shaded area) is also shown. Values have been been computed for the global region and averaged into 30-day intervals.

where

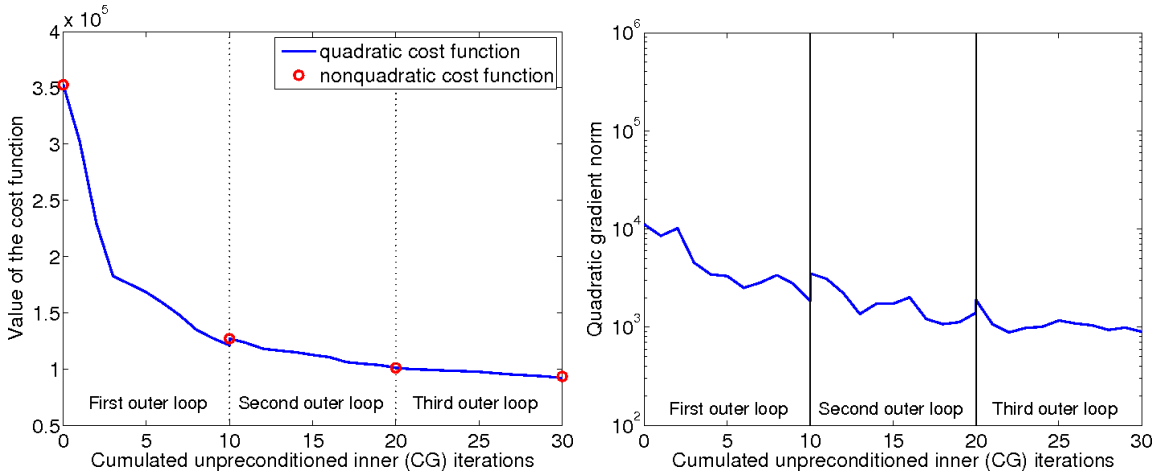
$$\mathbf{d}^{(k-1)} = \begin{pmatrix} \mathbf{d}_0^{(k-1)} \\ \vdots \\ \mathbf{d}_i^{(k-1)} \\ \vdots \\ \mathbf{d}_N^{(k-1)} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_0^{\mathrm{o}} - \mathbf{H}_0\mathbf{w}^{(k-1)}(t_0) \\ \vdots \\ \mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{w}^{(k-1)}(t_i) \\ \vdots \\ \mathbf{y}_N^{\mathrm{o}} - \mathbf{H}_N\mathbf{w}^{(k-1)}(t_N) \end{pmatrix} \quad \text{and} \quad \mathbf{H}^{(k-1)} = \begin{pmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_i\mathbf{M}^{(k-1)}(t_0,t_i) \\ \vdots \\ \mathbf{H}_N\mathbf{M}^{(k-1)}(t_0,t_N) \end{pmatrix} \quad (10)$$

The observation matrix $\mathbf{H}^{(k-1)}$ extends that of 3D-Var FGAT (Eq. (2)) by including a dynamical model $\mathbf{M}^{(k-1)}$ for propagating the increment from initial time to the observation times. Here $\mathbf{M}^{(k-1)}$ is the tangent-linear model with simplifications in the vertical and isopycnal mixing schemes. The superscript $(k-1)$ indicates linearization with respect to the state estimate $\mathbf{w}^{(k-1)}(t_i)$ on outer iteration k, where $\mathbf{w}^{(0)}(t_i) = \mathbf{w}^{\mathrm{b}}(t_i)$. The resolution used for the increment in the linearized model is identical to that used for the state vector in the nonlinear model. The initial state is updated directly with the increment after each quadratic minimization. The quadratic cost function is minimized using a conjugate gradient (CG) algorithm based on a close variant of the CONGRAD software (Fisher 1998). The experimental framework is similar to that of the 3D-Var experiment B1R1 described earlier. Here we focus on the results from a single cycle minimization with a 10-day assimilation window. Three outer iterations were used ($K = 3$), and ten CG (inner) iterations were used per outer iteration.

3.1 The cost function and gradient norm

Figure 6a shows the value of the quadratic (incremental) cost function on the inner loop as a function of the inner iteration number for the three outer iterations of the 4D-Var experiment. No second-level

preconditioning was used between outer iterations in this experiment (only the first-level **B** preconditioning was used). The value of the nonquadratic (nonincremental) cost function at the end-points of each outer iteration is also shown in Fig. 6a (open circles). These correspond to the linearization points ($k = 0, 1, 2$ and 3) in the incremental algorithm. The discrepancy between the values of the quadratic cost function and the nonquadratic cost function at the outer loop end-points gives an indication of the accuracy of the linear approximation. As can be seen from Fig. 6a, the differences are small and only distinguishable between the first and second outer iterations ($k = 1$) where the relative error is 4.5% (for $k = 2$ the relative error is less than 0.1%). This suggests that the linear approximation is quite accurate in this experiment.



(a) Cost function

(b) Euclidean gradient norm

Figure 6: a) The values of the a) quadratic cost function (solid curve) and nonquadratic cost function (open circles), and b) Euclidean norm of the gradient of the cost function, as a function of the inner (CG) iteration number in each of the three outer iterations of the (unpreconditioned) 4D-Var experiment. The curves are placed one after the other in sequence and the inner iterations are cumulated.

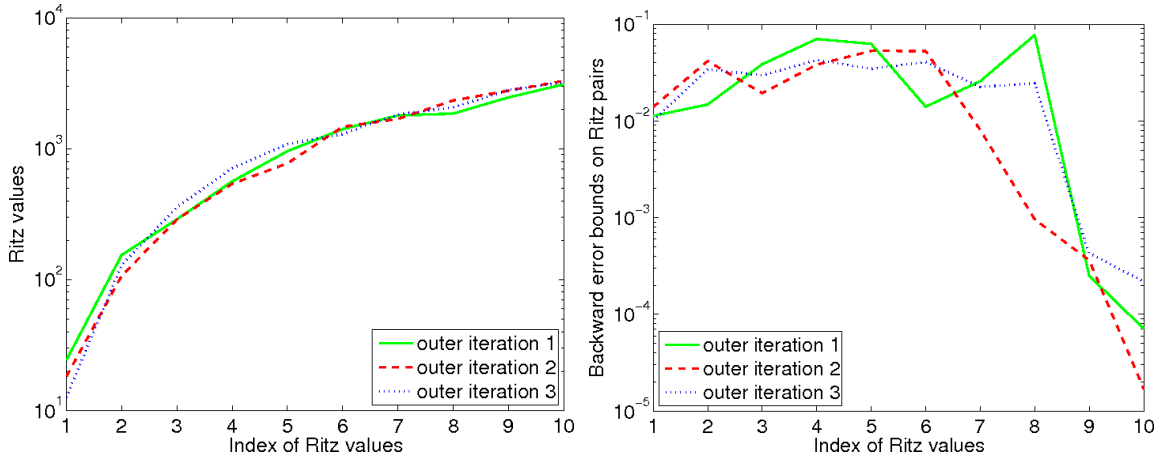
The Euclidean norm of the gradient of the quadratic cost function on each inner iteration has been proposed as a practical measure of convergence of CG minimization (Lawless and Nichols 2006). Figure 6b shows the Euclidean gradient norm as a function of inner iteration for the three outer iterations. Contrary to the quadratic cost function, the Euclidean gradient norm does not decrease smoothly and monotonically with increasing inner iteration. The Euclidean gradient norm is therefore not a good measure of convergence of the CG minimization for this experiment. Alternative convergence measures that decrease monotonically with CG iteration are available, such as the relative reduction in the quadratic cost (Meurant 2006). Operational systems, however, tend to use, as in this experiment, a stopping criterion based on a maximum number of inner iterations since it gives full control over the computational time for the assimilation step.

3.2 Ritz values/vectors

Limited-memory preconditioners (LMPs) such as spectral (Fisher 1998), quasi-Newton (QN; Morales and Nocedal 2000) and Ritz (Tshimanga *et al.* 2008) are constructed from scalar and vector information gathered, at no extra computational cost, during the CG minimization on previous outer iterations. They are based under the assumption that the Hessian matrix does not change significantly from one outer iteration to the next. One way of analyzing the validity of this assumption is to examine, between each

outer iteration, the variation of the Ritz values (approximate Hessian eigenvalues) which are available from the CG minimization. Figure 7a shows that the Ritz values for the unpreconditioned experiment are indeed similar in each of the three outer iterations. The maximum relative difference for the largest Ritz value is 3% and occurs between the first and third outer iterations. This result then suggests that the Hessian matrix associated with previous outer iterations provides relevant information for preconditioning the minimization on subsequent outer iterations.

The practical implementation of the spectral LMP involves approximating exact eigenpairs by Ritz pairs. The so-called backward error (Rigal and Gaches 1967) associated with this approximation is shown in Fig. 7b. This figure shows that the information related to the two largest Ritz values (indices 9 and 10) is the most accurate on all three outer iterations (cf. Figs. 7a and b), with the backward error being between 10^{-3} and 10^{-5}. For the other Ritz values, the error is between 10^{-1} and 10^{-2}, except for index 8 on the second outer iteration where the error is nearer 10^{-3}. Since the largest Ritz value is an accurate estimate of the largest eigenvalue of the Hessian and since, with the **B** preconditioning, the smallest eigenvalue of the Hessian is equal to one, the condition number κ can be estimated to a good approximation. For this unpreconditioned experiment, the values of κ are 3099, 3300 and 3202 on the first, second and third outer iterations, respectively.

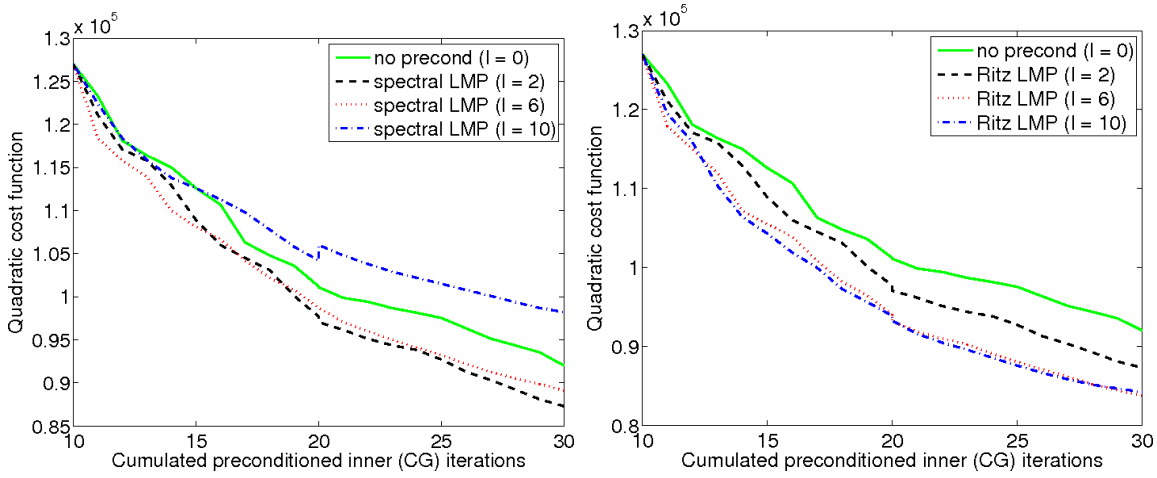


(a) Ritz values (b) Backward error

Figure 7: a) The ten Ritz values and b) the associated backward error, evaluated at the end of each of the three outer iterations of the (unpreconditioned) 4D-Var experiment.

The impact of different LMPs can be assessed by examining the inner-loop minimization after the first outer iteration. Figure 8 illustrates the effect of increasing the number of preconditioning vectors (l) used to construct the spectral and Ritz LMPs. For constructing these LMPs, the Ritz pairs with smallest backward error were used (Fig. 7). Figure 8a shows the performance of the minimization with the spectral LMP (approximated with Ritz information) tends to deteriorate when the number of vectors increases. With $l = 10$, the preconditioned experiment actually degrades the minimization compared to the unpreconditioned experiment. This can be explained by the fact that the Ritz pairs are not all good approximations to eigenpairs of the Hessian matrix, and therefore inappropriate for use within the spectral LMP. However, with accurate eigenpair information, which is the case when $l = 2$ and to a lesser extent when $l = 6$, the spectral LMP is clearly beneficial to the minimization. Figure 8b shows that, contrary to the spectral LMP, the Ritz LMP tends to improve the minimization when the number of preconditioning vectors is increased. The amount gained by increasing l from 0 to 2 vectors and from 2 to 6 vectors is similar. There is only marginal impact, however, from using $l = 10$ compared to $l = 6$. This behaviour is consistent with the changes in κ which is equal to 2378 when $l = 2$, 1290 when $l = 6$

and 1247 when $l = 10$. The small impact from using 10 compared to 6 vectors suggests that it may not be necessary to use the maximum number of preconditioning vectors to obtain near-optimal benefit from the Ritz LMP, an important point if computer memory space is limited.



(a) Spectral LMP (b) Ritz LMP

Figure 8: The values of the quadratic cost function on the second and third outer iterations of a 4D-Var experiment without preconditioning (green solid curve) and with preconditioning using different numbers of preconditioning vectors: $l = 2$ (black dashed curve); $l = 6$ (red dotted curve); and $l = 10$ (blue dashed-dotted curve). Displayed are the results for the a) spectral LMP and b) Ritz LMP. Note that for the spectral LMP, the Ritz values/vectors are used as an approximation to exact eigenvalues/eigenvectors.

3.3 Behaviour of the nonquadratic cost function

Figure 9 shows the behaviour of the quadratic and nonquadratic cost functions as a function of inner iteration on the third outer iteration of the 4D-Var experiment employing the Ritz LMP with $l = 10$ preconditioning vectors. In the incremental 4D-Var algorithm, the value of the nonquadratic cost function is only computed at the outer loop end-points and is not readily available on intermediate inner iterations as displayed in Fig. 6. Here, the intermediate values were computed by performing a sequence of 4D-Var experiments with a different number of inner iterations on the third outer iteration. This figure illustrates a very different behaviour in the quadratic and nonquadratic cost functions on the third outer iteration. While the quadratic cost function decreases monotonically during the entire inner-loop minimization, the nonquadratic cost function, after an initial reduction during the first 5 or 6 inner iterations (25 or 26 cumulated iterations), starts to diverge to the extent that by iteration 10 it has returned to a value similar to that obtained after 4 inner iterations. In this experiment, it is clearly detrimental to the global minimization to iterate the CG algorithm beyond 6 inner iterations on the third outer iteration. In general, the appropriate number of inner iterations for each outer iteration will depend on the particular LMP employed.

4 Conclusions

Assimilation diagnostics used to evaluate the ensemble ocean 3D-Var system have suggested that the analyses produced with the ensemble-estimated background-error variances were in better balance than

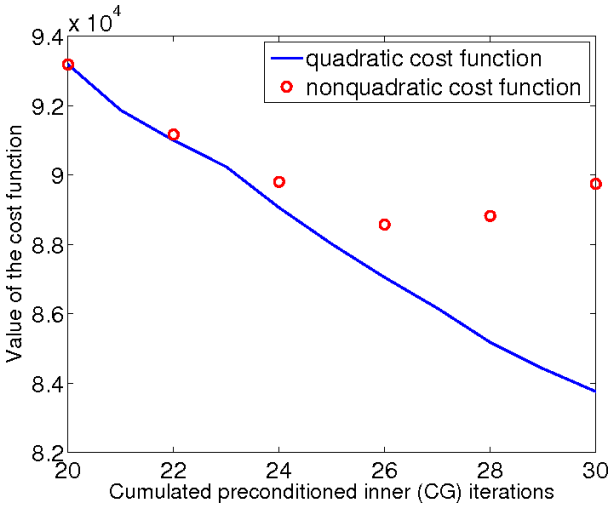


Figure 9: The values of the quadratic cost function (blue solid curve) and nonquadratic cost function (red open circles) as a function of the cumulated inner (CG) iteration number on the third outer iteration of the preconditioned 4D-Var experiment using the Ritz LMP with $l = 10$ vectors.

those produced with the parameterized background-error variances: there was reduced error growth between cycles, the analysis increments were smaller, and the analyses were closer to independent measurements of variables not directly constrained by the assimilated temperature and salinity profiles (results not shown here; see Daget *et al.* (2009)). These positive results, however, were accompanied by a slight worsening of the innovation statistics in the upper 100 m. In this region, the Desroziers diagnostics indicated that the ensemble-estimated variances (and to a lesser extent the parameterized variances) were underestimated. This apparent understimation of the ensemble spread points to the need to improve the ensemble generation strategy.

The minimization diagnostics used to evaluate the convergence properties of the incremental ocean 4D-Var emphasized the importance of developing an appropriate stopping criterion for the inner-loop minimization. The Euclidean gradient norm was shown not to be a robust measure of convergence of the inner loop. Furthermore, full convergence of the inner loop was shown to be not necessarily desirable in a nonlinear system. In particular, too many inner iterations could lead to divergence on the outer loop, thus illustrating the importance of carefully monitoring the outer-loop convergence in addition to that of the inner-loop.

Acknowledgements

The diagnostics from the ensemble 3D-Var system were produced in collaboration with Magdalena Balmaseda and Nicolas Daget. The minimization diagnostics were produced in collaboration with Serge Gratton, Annick Sartenaer and Jean Tshimanga. Support for this work was obtained from the ENSEMBLES project (contract No. GOCE-CT-2003-505539), the Groupe Mission Mercator/Coriolis and LEFE-ASSIM.

References

Bloom S. C., L. L. Takacs, A. M. da Silva and D. Ledvina (1996). Data assimilation using Incremental Analysis Updates. *Mon. Weather Rev.*, **124**, 1256–1271. Courtier, P., J.-N. Thépaut and A. Hollingsworth (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367–1388.

Daget, N., A. T. Weaver and M. A. Balmaseda (2009). Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Q. J. R. Meteorol. Soc.*, **135**, 1071–1094.

Desroziers, G., L. Berre, B. Chapnik and P. Poli (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.*, **131**. 3385–3396.

Fisher, M. (1998). Minimization algorithms for variational data assimilation. *Recent Developments in Numerical Methods for Atmospheric Modelling*, ECMWF, pp. 364–385.

Fu, L. L., I. Fukumori and R. N. Miller (1993). Fitting dynamic models to the Geosat sea level observations in the tropical Pacific Ocean. Part II: A linear, wind-driven model. *J. Phys. Oceanogr.*, **23**, 2162–2181.

Golub, G. H., C. F. Van Loan (1996). *Matrix Computations* (3rd edn). Johns Hopkins University Press, Baltimore.

Ingleby, B. and M. Huddleston (2007). Quality control of ocean temperature and salinity profiles - historical and real-time data. *J. Mar. Sys.*, **65**, 158–175.

Lawless, A. S. and N. K. Nichols (2006). Inner-loop stopping criteria for incremental four-dimensional variational assimilation. *Mon. Weather. Rev.*, **134**, 3425–3435.

Madec, G., P. Delecluse, M. Imbard and C. Levy (1998). OPA8.1 Ocan General Circulation Model reference manual. Technical note no. 11. LODYC/IPSL. Paris. France.

Meurant, G. (2006). *The Lanczos and Conjugate Gradient Algorithms: From Theory to Finite Precision Computations*. SIAM, Philadelphia, USA.

Morales, J. and J. Nocedal (2000). Automatic preconditioning by limited memory quasi-Newton updating. *SIAM J. Optimization*, **10(4)**, 1079–1096.

Rigal, R. L. and J. Gaches (1967). On the Compatibility of a given solution with the data of a linear system. *J. ACM.*, **14**, 543–548.

Tshimanga, J., S. Gratton, A. T. Weaver and A. Sartenaer (2008). Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **134**, 753–771.

Weaver, A. T., J. Vialard and D. L. T. Anderson (2003). Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks. *Mon. Weather Rev.*, **131**, 1360–1378.

Weaver, A. T., C. Deltel, E. Machu, S. Ricci and N. Daget (2005). A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.*, **131**, 3605–3625.