

Department of Mathematics

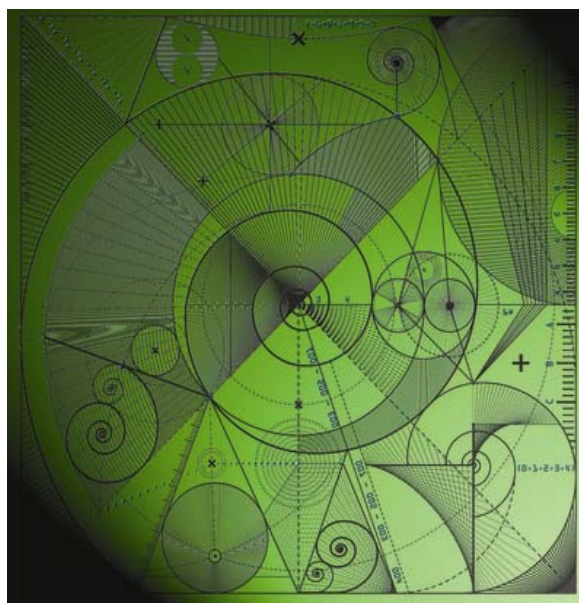
Preprint MPS_2010-07

19 March 2010

A note on the analysis error associated with
3D-FGAT

by

A.S. Lawless





A note on the analysis error associated with 3D-FGAT

A.S. Lawless

Department of Mathematics, University of Reading, U.K.

Abstract: The analysis error variance of a 3D-FGAT assimilation is examined analytically using a simple scalar equation. It is shown that the analysis error variance may be greater than the error variances of the inputs. The results are illustrated numerically with a scalar example and a shallow-water model. Copyright © 0000 Royal Meteorological Society

KEY WORDS First guess at appropriate time; variational data assimilation

Received

1 Introduction

Data assimilation is widely used in weather and ocean forecasting to provide initial conditions for numerical forecast models. By combining observational data with an a priori, or background, estimate of the model state, it is possible to obtain an improved estimate of the current state of the system, known as the analysis. Many data assimilation techniques are based on Bayes' rule, which in the case of Gaussian errors is equivalent to a least squares fitting. For such techniques, provided that the errors in the observations and background state are correctly represented, the analysis obtained will be at least as accurate as the most accurate piece of input information, in a statistical sense. Thus the addition of more information into the assimilation procedure cannot degrade the analysis.

In practice many approximations must be made in designing data assimilation schemes for practical use. One such approximation is known as 3D-FGAT (first-guess at appropriate time), which can be considered as a half-way step between incremental three-dimensional variational data assimilation (3D-Var) and incremental four-dimensional variational data assimilation (4D-Var). The aim of 3D-FGAT is to provide some of the time information of a sequence of observations, without the need to code a full tangent linear and adjoint model, as needed in incremental 4D-Var. Such a scheme was used to generate the ERA-40 reanalysis (Uppala *et al.*, 2005) and has been used in several atmospheric, oceanic and chemical assimilation systems (for example, Lee *et al.*, 2004; Vialard *et al.*, 2003; Barret *et al.*, 2008). In this note we examine the effect of the 3D-FGAT approximation on the analysis errors for a simple problem. By considering the scheme as an approximation to incremental 4D-Var we show that the effect of the approximation is to increase the variance of the analysis error. In particular, we show that whereas for a perfect incremental 4D-Var system the

analysis error variance can be no larger than the smallest of the variances on the inputs, the analysis error variance in 3D-FGAT can exceed the error variance of the inputs.

2 Assimilation schemes

Incremental 4D-Var requires the minimization of a sequence of linearized cost functions of the form

$$\mathcal{J}[\delta\mathbf{x}_0] = \frac{1}{2}\delta\mathbf{x}_0^T \mathbf{B}^{-1} \delta\mathbf{x}_0 + \frac{1}{2} \sum_{i=0}^N (\mathbf{H}_i \delta\mathbf{x}_i - \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i \delta\mathbf{x}_i - \mathbf{d}_i), \quad (1)$$

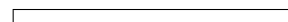
where \mathbf{B} is the background error covariance matrix, \mathbf{R}_i , \mathbf{H}_i and \mathbf{d}_i are the observation error covariance matrices, linearized observation operators and innovation vectors at times t_i and $\delta\mathbf{x}_i$ satisfies the tangent linear model equation. The method is equivalent to solving the full nonlinear 4D-Var problem using a Gauss-Newton method (Lawless *et al.*, 2005).

The algorithm for 3D-FGAT is very similar to incremental 4D-Var, except that the tangent linear model is approximated by the identity, so that in the linearized cost function (1) we have $\delta\mathbf{x}_i = \delta\mathbf{x}_0$ for all times t_i . This introduces a discrepancy between the nonlinear model used to calculate the innovations in the outer loop and the linear model used to evolve the perturbations in the inner loop. We note that in operational implementations of 3D-FGAT the increment is usually defined to be at the centre of a time window of observations rather than at the beginning. In this case the linearized cost function takes the form

$$\mathcal{J}[\delta\mathbf{x}_0] = \frac{1}{2}\delta\mathbf{x}_0^T \mathbf{B}^{-1} \delta\mathbf{x}_0 + \frac{1}{2} \sum_{i=-N/2}^{N/2} (\mathbf{H}_i \delta\mathbf{x}_0 - \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i \delta\mathbf{x}_0 - \mathbf{d}_i). \quad (2)$$

In the next section we examine the analysis error of 3D-FGAT for a simple scalar example.

*Correspondence to: Department of Mathematics, University of Reading, Whiteknights, PO Box 220, Reading, Berkshire, RG6 6AX, U.K. E-mail: a.s.lawless@reading.ac.uk



3 Analysis error for a simple example

We consider the analysis error after one outer loop of incremental 4D-Var and of 3D-FGAT. In order to illustrate the effect of the FGAT approximation we consider the simple example of a linear model for a scalar variable x . We define the full discrete system model to be

$$x_{i+1} = \alpha x_i, \quad (3)$$

where α is a scalar constant and $x_i \equiv x(t_i)$. We consider an example in which we have a time window $[t_0, t_2]$ with observations y_0, y_2 at times t_0 and t_2 respectively. We assume that the errors on each observation have error variance σ_o^2 and that the errors are uncorrelated. For this system, since the full model (3) is already linear, the tangent linear model has the same form as the full model and is given by

$$\delta x_{i+1} = \alpha \delta x_i, \quad (4)$$

where δx_i is a small perturbation to the state x_i . We note that by using a linear system we expect the incremental 4D-Var algorithm to give the same solution as the minimization of the full 4D-Var cost function, since no approximation is being made in the linearization step. Although this is a big simplification in order to make the mathematical analysis tractable, it does not take away from the validity of the approach, but allows us to analyse the effects of 3D-FGAT in the simplest system possible.

We must be very precise about the properties of the background field at the different points in the time window. We will assume that the background field for the 3D-FGAT scheme is $x_b(t_1)$, defined at the centre of the window, time t_1 , with error variance σ_b^2 . In order to derive a 4D-Var scheme for this system we need to have a background field at the start of the time window, time t_0 . Since the model is linear then the background field at time t_0 is simply given by

$$x_b(t_0) = \frac{1}{\alpha} x_b(t_1) \quad (5)$$

and this will have an error variance of σ_b^2/α^2 . The innovation vectors for this problem are

$$d_0 = y_0 - x_b(t_0), \quad d_2 = y_2 - \alpha^2 x_b(t_0). \quad (6)$$

The inner loop cost function for incremental 4D-Var is then

$$\begin{aligned} \mathcal{J}[\delta x_0] = & \frac{1}{2} \alpha^2 \frac{\delta x_0^2}{\sigma_b^2} + \frac{1}{2} \frac{(d_0 - \delta x_0)^2}{\sigma_o^2} \\ & + \frac{1}{2} \frac{(d_2 - \alpha^2 \delta x_0)^2}{\sigma_o^2}. \end{aligned} \quad (7)$$

We minimize this to obtain δx_0 and add this to the background field $x_b(t_0)$ at time t_0 to obtain the incremental 4D-Var analysis

$$x_a(t_0) = \frac{1}{\alpha^2 \sigma_o^2 + \sigma_b^2 (1 + \alpha^4)} [\alpha^2 \sigma_o^2 x_b + \sigma_b^2 y_0 + \alpha^2 \sigma_b^2 y_2], \quad (8)$$

which has analysis error

$$\sigma_a^2(t_0) = \frac{\sigma_o^2 \sigma_b^2}{\alpha^2 \sigma_o^2 + \sigma_b^2 (1 + \alpha^4)}. \quad (9)$$

As expected we find that the analysis error variance is both less than the observation error variance and less than the error variance of the background field used, which in this case is given by σ_b^2/α^2 at time t_0 .

If the analysis is evolved to the centre of the time window using the model (3), then the analysis error of the incremental 4D-Var scheme at that time is

$$\sigma_a^2(t_1) = \frac{\alpha^2 \sigma_o^2 \sigma_b^2}{\alpha^2 \sigma_o^2 + \sigma_b^2 (1 + \alpha^4)}, \quad (10)$$

which is less than σ_b^2 , the background error variance at time t_1 . This result is a simple extension in time of a standard result for minimum variance estimation (for example, Daley, 1991, section 4.1).

For the 3D-FGAT scheme applied to this problem the increment is considered to be valid at the centre of the time window t_1 and the inner loop cost function (2) is given by

$$\begin{aligned} \mathcal{J}[\delta x_1] = & \frac{1}{2} \frac{\delta x_1^2}{\sigma_b^2} + \frac{1}{2} \frac{(d_0 - \delta x_1)^2}{\sigma_o^2} \\ & + \frac{1}{2} \frac{(d_2 - \delta x_1)^2}{\sigma_o^2}, \end{aligned} \quad (11)$$

which has a minimum at

$$\delta x_1 = \frac{\sigma_b^2 (d_0 + d_2)}{\sigma_o^2 + 2\sigma_b^2}. \quad (12)$$

For this scheme the analysis is found by adding the increment to the background field in the centre of the time window. Thus we have

$$\begin{aligned} x_a(t_1) = & x_b(t_1) + \frac{\sigma_b^2}{\sigma_o^2 + 2\sigma_b^2} (y_0 - x_b(t_0)) \\ & + \frac{\sigma_b^2}{\sigma_o^2 + 2\sigma_b^2} (y_2 - \alpha^2 x_b(t_0)). \end{aligned} \quad (13)$$

A calculation of the analysis error variance gives

$$\begin{aligned} \sigma_a^2(t_1) = & \frac{\sigma_b^2 \sigma_o^2}{\sigma_o^2 + 2\sigma_b^2} \\ & + \frac{\sigma_b^4}{(\sigma_o^2 + 2\sigma_b^2)^2} (2 - \beta) ((2 - \beta) \sigma_b^2 + 2\sigma_o^2), \end{aligned} \quad (14)$$

where

$$\beta = \alpha + \frac{1}{\alpha}. \quad (15)$$

We note that where $\alpha = 1$, so that there is no approximation in 3D-FGAT, we have $\beta = 2$ and the analysis error variance is the same as that for incremental 4D-Var. In this case the analysis error variance is bounded above by the smallest of the variances of the observational and background errors. For $\alpha \neq 1$ the second term of (14) introduces an error dependent on the factor $2 - \beta$. Thus this

term measures how close the identity approximation used in 3D-FGAT is to the true tangent linear model. The further the tangent linear model is from the identity, the larger this term will become. It is particularly important to note that for values of α far from unity this term may be arbitrarily large and so the analysis error variance at the initial time may exceed the error variance of the inputs.

In this example we have assumed that the variance information of the background field is correct at the centre of the time window. However, by removing the model evolution of the perturbation, the evolution of the variance information is neglected, so that the innovations are weighted incorrectly. For the case where $\alpha > 1$ so that the variance grows throughout the assimilation window, then the innovation at time t_0 is over-weighted with respect to the background and the innovation at time t_2 is under-weighted (with the opposite occurring for $\alpha < 1$). It is this incorrect use of the statistical information which leads to a sub-optimal analysis.

If we consider 3D-Var scheme applied to this system, so that the observations are assumed to be valid at the centre of the time window, then the analysis is found to be no longer unbiased. This arises from the fact that the innovations are calculated as

$$d_0 = y_0 - \alpha x_b(t_0), \quad d_2 = y_2 - \alpha x_b(t_0). \quad (16)$$

This introduces terms in the expected analysis error dependent on the change in the true state between observation times. Terms involving the true state then also occur in the expression for the variance. Hence we see that 3D-FGAT theoretically removes a major source of error in 3D-Var, even if the analysis error variance may be large.

4 Numerical results

In order to illustrate the problems associated with the analysis error for 3D-FGAT we present numerical results based on the example presented in the previous section and on a shallow-water model. In practice we may expect the effect of the 3D-FGAT approximation to depend on the ratio of σ_b^2/σ_o^2 , which in turn will depend on the physical system and variable being modelled. Fisher (2007) states that this ratio is approximately equal to one for a numerical weather prediction system, the figures of Weaver et al. (2003) indicate a ratio of approximately 2 for the temperature of the central Pacific ocean, whereas the figures presented in Daget et al. (2009) indicate a ratio of 0.1 – 0.5 for a global ocean model. The different combinations of the background and observation error variances σ_b^2 and σ_o^2 presented here are chosen to illustrate possible effects of the 3D-FGAT approximation rather than to imitate any particular physical system.

4.1 Scalar example

We consider the simple model given by (3). We define a true value of the state at the initial time, $x_t(t_0)$ and specify the value of the model parameter α . Values of the true

Table I. Analysis error variance for different values of α for the scalar example using incremental 4D-Var and 3D-FGAT.

α	4D-Var	3D-FGAT
0.25	0.02	1.9
0.5	0.09	0.20
1.5	0.14	0.17
2.0	0.09	0.20
2.5	0.06	0.37
2.75	0.05	0.51

state at subsequent times in the assimilation window are calculated using the model equation (3). We consider a two-time-step window $[t_0, t_2]$, with observations at times t_0 and t_2 and the background field $x_b(t_1)$ defined at time t_1 . The observations are taken to be equal to the true field at the appropriate time plus a random, unbiased, Gaussian error, with a specified variance σ_o^2 . The background field is defined to be equal to the true state at time t_1 plus a random, unbiased, Gaussian error with specified variance σ_b^2 . We calculate the analyses from the incremental 4D-Var and 3D-FGAT schemes over a total of 100,000 independent cases, using different random noise inputs with the same specified variances, and then calculate the variance of the analysis error over all of these cases. The incremental 4D-Var scheme uses the true tangent linear model, with the specified value of α , whereas 3D-FGAT approximates this by $\alpha = 1$.

We choose the true initial state to be $x_t(t_0) = 5$ and the input variances on the observational and background errors to be $\sigma_o^2 = 0.4$ and $\sigma_b^2 = 2.0$ respectively. The values of the analysis error variance from the different experiments are shown in Table I. For the incremental 4D-Var scheme the analysis error is calculated at the centre of the time window, so as to provide a direct comparison with the 3D-FGAT error. For the values of α shown the analytical values calculated from the formulae (10) and (14) match the numerical values to the accuracy given, and so these values are not duplicated in the table. For values of the true model parameter α close to one, where the 3D-FGAT scheme is a good approximation, we find that the analysis errors for the two schemes are very similar. However, as the value of α moves away from one and the 3D-FGAT approximation breaks down, the analysis error variance for this scheme also increases. We see that for $\alpha = 0.25$ and for $\alpha = 2.75$ the analysis error variance is greater than the observation error variance. Thus the use of a sub-optimal assimilation system with an inaccurate background field has resulted in a loss of the accurate information present in the observations. It is also possible to find values of the parameters such that the analysis error variance is greater than both the observational and background error variances. For example, if we choose $\alpha = 0.25$, $\sigma_o^2 = 0.5$ and $\sigma_b^2 = 5$, then the analysis error variance using 3D-FGAT is 5.46, which is greater than either of the input variances.

4.2 Shallow water model

As a second example we consider a more realistic system, the one-dimensional nonlinear shallow water system for the flow of a fluid over an obstacle in the absence of rotation. The model equations are given by

$$\frac{Du}{Dt} + \frac{\partial\phi}{\partial x} = -g\frac{\partial\bar{h}}{\partial x}, \quad \frac{1}{\phi}\frac{D\phi}{Dt} + \frac{\partial u}{\partial x} = 0, \quad (17)$$

with $D/Dt = \partial/\partial t + u\partial/\partial x$. In these equations $\bar{h} = \bar{h}(x)$ is the height of the bottom orography, u is the velocity of the fluid and $\phi = gh$ is the geopotential, where g is the gravitational acceleration and $h > 0$ the depth of the fluid above the orography. The system is discretized using a semi-implicit semi-Lagrangian integration scheme, as described in Lawless *et al.* (2003). We define the problem on a periodic domain of 1000 grid points, with a spacing $\Delta x = 0.01$ m between them, so that $x \in [0$ m, 10 m] and assume a model time step $\Delta t = 0.0092$ s. Other parameters for the problem are as defined in Lawless *et al.* (2005).

A 3D-FGAT scheme for this system is set up over a time window $[-T, T]$, with observations of u and ϕ at every spatial point at times $-T, 0, T$. The true state at time $-T$ is defined using the initial conditions from Case I of Lawless *et al.* (2005) and we take $T = 0.23$, so that there are 50 time steps in the assimilation interval. The solution to this problem is given by a stationary field over the orography and two outgoing gravity waves moving away from the orography in opposite directions. The assimilation time window corresponds to approximately 40% of the time scale of the gravity waves. For the assimilation experiments the observations and background field are generated by adding random, uncorrelated Gaussian noise to the true state with given variances. The assimilation is then run using these variances and a background error covariance matrix defined using an exponential correlation function with length scale 0.05 m.

In order to calculate the analysis error variance for this system, instead of sampling over different assimilation experiments, we assume that the analysis errors at different spatial points are independent statistical samples. In Table II we show the analysis error variance from assimilation experiments in which the comparative accuracy of the background and observations changes. For the ‘accurate’ observations or background the variances are set to 10^{-4} $m^2 s^{-1}$ for u and 4×10^{-4} $m^2 s^{-2}$ for ϕ . For the ‘inaccurate’ data these are increased by two orders of magnitude to 10^{-2} $m^2 s^{-1}$ and 4×10^{-2} $m^2 s^{-2}$ respectively. These correspond to standard deviations of approximately 1% of the full fields for the accurate data and 10% of the full fields for the inaccurate data. We see that when the background and observations are both accurate, or when the background is more accurate than the observations, then the analysis error variances for both u and ϕ are both of order 10^{-5} , which is less than all the input variances. However, when the background is inaccurate and the observations are accurate, then the variance of the analysis errors for both u and ϕ are greater than the variances of the observations. Thus, as for the simple scalar

Table II. Analysis error variances for experiments using the shallow water model with different comparative accuracies of the observations and background.

		Accurate observations	Inaccurate observations
Accurate background	u	1.95×10^{-5}	1.94×10^{-5}
	ϕ	3.24×10^{-5}	3.97×10^{-5}
Inaccurate background	u	8.16×10^{-4}	-
	ϕ	1.06×10^{-3}	-

Table III. Analysis error variances for experiments using the shallow water model with different time windows.

Time window	u error	ϕ error
30	7.35×10^{-4}	9.95×10^{-4}
10	1.87×10^{-4}	3.69×10^{-4}
2	1.70×10^{-5}	3.44×10^{-5}

example, the inaccurate background is able to degrade the information available in the good observations. The results for inaccurate observations and background are not discussed, since this is qualitatively the same as for accurate observations and background.

From the results with the simple example we may expect this problem to arise when the linear model is far from the identity, and so to be worse for a larger assimilation window. To test this hypothesis we perform further experiments using an inaccurate background and accurate observations for time windows of 30 time steps, 10 time steps and 2 time steps. The analysis errors for these experiments are shown in Table III. We find that as the time window is shortened, so that the FGAT approximation is more accurate, then the analysis error is reduced. However, for the time windows of 30 and 10 time steps the variance of the analysis errors for u are still higher than those of the observations. It is only when we have a very small window that the analysis is more accurate than the observations.

5 Concluding remarks

Assimilation schemes based on 3D-FGAT are widely used as an extension to 3D-Var in cases where the implementation of a 4D-Var system is prohibitively expensive. In this note we have demonstrated a property of 3D-FGAT which has hitherto been unremarked on in the literature, namely that the variance of the analysis error may be greater than the variance of any or all of the inputs. This is an inherent property which arises from the approximation of the tangent linear model by the identity within the assimilation scheme, which is not accounted for statistically. It can be interpreted as a kind of representativeness error within the linear problem, where the error is in how well the approximate linearization of the nonlinear observation operator represents the exact linearization, rather than how well the observation operator represents the true mapping between the state and observation space. The neglect of

this component of the error within the assimilation system may lead to a sub-optimal analysis, which can cause an increase in the analysis error even when all other prior statistical information is correctly specified. It is important not to confuse this assumption with the tangent linear assumption used in 4D-Var. Even in a linear situation the assumption of 3D-FGAT may not hold.

Although these results have been demonstrated for simple examples, there is no reason to think that this problem will disappear as the model becomes more complicated. Rather, the problem may arise whenever the true tangent linear model matrix is far from the identity. However, it must be recognized that 3D-FGAT is still likely to be an improvement over 3D-Var, which assumes that the observations in a time window are all valid at the same time. In fact, operational practice has shown great benefits from moving to a 3D-FGAT scheme. Hence, these results are not designed to discourage the use of 3D-FGAT. Rather they illustrate the importance of understanding the assumptions in this assimilation approach and their possible effects on the analysis error. One indication arising from this work is the importance of testing the validity of the 3D-FGAT approximation, in much the same way as a 4D-Var system is tested. Such a test was implemented by Weaver *et al.* (2003) in the design of a 3D-FGAT assimilation system for an ocean circulation model. More routine tests of this type in a 3D-FGAT system would indicate where the error in the approximation is high and so provide an indication of possible uncertainties in the 3D-FGAT analysis.

Finally we comment that the approximation of the tangent linear model by the identity in 3D-FGAT can be considered an extreme example of using an approximate linear model in incremental 4D-Var. In most operational 4D-Var systems the linear model is not an exact linearization of the nonlinear model used in the outer loop, but often contains simplifications, such as different physical parametrizations or a different spatial resolution. Although we wished to concentrate on the 3D-FGAT approach in this note, a simplified incremental 4D-Var scheme could be analysed in the same way. It is clear from the analysis presented here that if the error between the approximate and true linear model is not accounted for statistically, then this may lead to the increase in analysis error variance illustrated in this note.

Acknowledgement

This work was supported by the UK Natural Environment Research Council through the National Centre for Earth Observation. The author would like to thank the two anonymous reviewers whose comments greatly helped to improve a previous version of this manuscript.

References

Barret, B, Ricaud, P, Mari, C, Attié, J-L, Bousserez, N, Josse, B, Le Flochmœn, E, Livesey, NJ, Massart, S, Peuch, V-H, Piacentini, A, Sauvage, B, Thouret, V, Cammas, J-P. 2008. Transport pathways of

- CO in the African upper troposphere during the monsoon season: a study based upon the assimilation of spaceborne observations. *Atmos. Chem. Phys.* **8**: 3231–3246.
- Daget, N., Weaver, A.T., Balmaseda, M. 2005. Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Quart. J. Royal Met. Soc.* **135**: 1071–1094.
- Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press.
- Fisher, M. 2007. The sensitivity of analysis errors to the specification of background error covariances. In *Proceedings of ECMWF workshop on flow dependent aspects of data assimilation*, 11–13 June 2007, 27–36.
- Lawless, AS, Gratton, S, Nichols, NK. 2003. A comparison of two methods for developing the linearization of a shallow-water model. *Quart. J. Royal Met. Soc.* **129**: 1237–1254.
- Lawless, AS, Gratton, S, Nichols, NK. 2005. An investigation of incremental 4D-Var using non-tangent linear models. *Quart. J. Royal Met. Soc.* **131**: 459–476.
- Lee, M-S, Barker, D, Huang, W, Kuo, Y-H. 2004. First guess at appropriate time (FGAT) with WRF 3DVAR. WRF/MM5 users' workshop, Boulder, Colorado, 22–25 June 2004.
- Uppala, SM *et al.* 2005. The ERA-40 re-analysis. *Quart. J. Royal Met. Soc.* **131**: 2961–3012.
- Vialard, J, Weaver, AT, Anderson, DLT, Delecluse, P. 2003. Three- and four-dimensional variational assimilation with a general circulation model of the Tropical Pacific Ocean. Part II: Physical validation. *Mon. Wea. Rev.* **131**: 1379–1395.
- Weaver, AT, Vialard, J, Anderson, DLT. 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. Wea. Rev.* **131**: 1360–1378.