Tellus (2007), 59A, 273–276
Printed in Singapore. All rights reserved

TELLUS

LETTER TO THE EDITOR

Interpretations of an adjoint-derived observational impact measure

By RONALD M. ERRICO*, *Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County and Global Modeling and Assimilation Office, Goddard Space Flight Center, Code 610.1, Greenbelt, MD 20771, USA*

(Manuscript received 30 January 2006; in final form 31 October 2006)

ABSTRACT

The measure developed by Langland and Baker to estimate and compare the utility of arbitrary sets of observations for reducing forecast error is re-derived here. The formula uses adjoints of both a forecast model and a data assimilation system. Although the final expression here is nearly identical to that developed by Langland and Baker, the third-order character of their measure is now explicitly presented. The appropriateness of going beyond first order is explained and a warning of the measure's nonlinearity possibly clouding its interpretation is offered.

1. Introduction

In Langland and Baker (2004, hereafter LB), adjoints of a forecast model and data assimilation system (the latter as in Baker and Daley, 2000) are used to efficiently estimate the reductions of a forecast-error measure due to arbitrary sets of observations used by the data assimilation system. Although the utility of their particular measure is demonstrated in both LB and as yet unpublished works, there is an apparent misconception regarding what they denote as the gradient of their measure with respect to values of the observations. Although this does not affect their primary conclusions, it does affect the precise interpretation of their results, most notably aspects relying on linearity.

2. Derivation

We begin by considering a forecast model

$$\mathbf{x}^f = \mathbf{m}(\mathbf{x}^0), \tag{1}$$

where $\mathbf{x}$ is an atmospheric model state vector with components x_i (also using j, k, l, n as component subscripts), superscripts f and 0 denote a model forecast and initial condition, respectively, and $\mathbf{m}$ denotes a nonlinear model. The forecast error is measured with respect to a verification state $\mathbf{x}^t$ (considered as 'truth') using

$$e = \sum_{i,j} \left(x_i^f - x_i^t\right) c_{i,j} \left(x_j^f - x_j^t\right), \tag{2}$$

where $c_{i,j}$ denotes a symmetric (more typically, diagonal) weighting matrix. Although e is a quadratic expression in $\mathbf{x}^f$, it is an even higher-order expression in terms of $\mathbf{x}^0$ when $\mathbf{m}$ is nonlinear. The component, rather than vector, form of (2) has been presented because tensors will be introduced at a later stage.

Usually, $\mathbf{x}^t$ is an analysis produced at the verification time using the same sequential data assimilation system that earlier produced $\mathbf{x}^0$ from observations and background information (described later). Such an $\mathbf{x}^t$ therefore also depends on $\mathbf{x}^0$. If the forecast period is sufficiently long, however, such that many additional observations have been assimilated during intervening assimilation cycles, then any such dependence can be ignored. The dependence will be ignored here even if applied to rather short forecast periods.

The first goal here is to express changes of e due to changes of $\mathbf{x}^0$ using a Taylor series approximation

$$\delta e = \sum_i \frac{\partial e}{\partial x_i^0} \delta x_i^0 + \frac{1}{2} \sum_{i,j} \frac{\partial^2 e}{\partial x_i^0 \partial x_j^0} \delta x_j^0 \, \delta x_i^0 + \frac{1}{6} \sum_{i,j,k} \frac{\partial^3 e}{\partial x_i^0 \partial x_j^0 \partial x_k^0} \delta x_k^0 \, \delta x_j^0 \, \delta x_i^0 + \cdots,$$

truncated to some order, where $\delta \mathbf{x}^0$ represents a sufficiently small change to $\mathbf{x}^0$ so that a truncated approximation is a good one. By

*Correspondence.
e-mail: rerrico@gmao.gsfc.nasa.gov
DOI: 10.1111/j.1600-0870.2006.00217.x

gathering the common factor δx_i^0, this series can be rewritten as

$$\delta e = \sum_i \delta x_i^0 \left(\frac{\partial e}{\partial x_i^0} + \frac{1}{2} \sum_j \frac{\partial^2 e}{\partial x_i^0 \partial x_j^0} \delta x_j^0 + \frac{1}{6} \sum_{j,k} \frac{\partial^3 e}{\partial x_i^0 \partial x_j^0 \partial x_k^0} \delta x_k^0 \, \delta x_j^0 + \cdots \right). \quad (3)$$

The term in parenthesis may be considered as a vector **g**. For computation of the derivatives,

$$m_{i,j} = \frac{\partial x_i^f}{\partial x_j^0} \quad (4)$$

will denote what is sometimes called either the Jacobian of the forecast model or, more properly, the resolvent matrix of the tangent linear version of the forecast model. Also,

$$\delta m_{i,j} = \sum_n \frac{\partial m_{i,j}}{\partial x_n^0} \delta x_n^0 \quad (5)$$

will denote a linear estimate of the change in the resolvent due to a change in the forecast trajectory about which the nonlinear model is linearized. This is zero when the model is linear. Also,

$$\delta x_k^f = \sum_j m_{k,j} \delta x_j^0 \quad (6)$$

is a linearized estimate of the change in the model forecast given a change to its initial condition. The derivatives appearing in (3) are

$$\frac{\partial e}{\partial x_i^0} = 2 \sum_{l,n} m_{l,i} c_{l,n} \left(x_n^f - x_n^t \right), \quad (7)$$

$$\frac{\partial^2 e}{\partial x_j^0 \partial x_i^0} = 2 \sum_{l,n} \left[\frac{\partial m_{l,i}}{\partial x_j^0} c_{l,n} \left(x_n^f - x_n^t \right) + m_{l,i} c_{l,n} m_{n,j} \right], \quad (8)$$

$$\frac{\partial^3 e}{\partial x_k^0 \partial x_j^0 \partial x_i^0} = 2 \sum_{l,n} \left[\frac{\partial^2 m_{l,i}}{\partial x_k^0 \partial x_j^0} c_{l,n} \left(x_n^f - x_n^t \right) + \frac{\partial m_{l,i}}{\partial x_j^0} c_{l,n} m_{n,k} + \frac{\partial m_{l,i}}{\partial x_k^0} c_{l,n} m_{n,j} + m_{l,i} c_{l,n} \frac{\partial m_{n,j}}{\partial x_k^0} \right]. \quad (9)$$

The next goal is to reconsider the various orders of approximations to (3) in terms of a vector of observation values $\mathbf{y}^o$. These observations are used to create an atmospheric analysis

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} \left[\mathbf{y}^o - \mathbf{h}(\mathbf{x}_b) \right] \quad (10)$$

from a prior (background) estimate $\mathbf{x}_b$, where **K** is the Kalman gain matrix and **h** is a possibly nonlinear operator used to relate the model state to the observations (called the forward observation operator). The difference

$$\delta \mathbf{y} = \mathbf{y}^o - \mathbf{h}(\mathbf{x}_b) \quad (11)$$

is the innovation vector, with elements denoted by δy_i. If **h** is either linear or defined only as a function of $\mathbf{x}_b$ (and not, for example, based on a linearization about the updated state $\mathbf{x}_a$) then **K** is independent of the values of $\mathbf{y}^o$, which will be assumed throughout this derivation.

For examining the impacts of observations on e, the appropriate choice for $\mathbf{x}^0$ is $\mathbf{x}_b$ so that

$$\delta \mathbf{x}^0 = \mathbf{x}_a - \mathbf{x}_b = \mathbf{K} \delta \mathbf{y}, \quad (12)$$

in which case when $\delta \mathbf{y} = \mathbf{0}$, $\delta e = 0$, and there is no impact. For any vector **g** in state space there is a corresponding vector $\tilde{\mathbf{g}} = \mathbf{K}^{\mathrm{T}} \mathbf{g}$ in observation space such that

$$(\delta \mathbf{x}^0)^{\mathrm{T}} \mathbf{g} = (\delta \mathbf{y})^{\mathrm{T}} \tilde{\mathbf{g}} \quad (13)$$

according to the definition of an adjoint. Various-order approximations to δe can therefore be expressed in terms of $\delta \mathbf{y}$.

The first-order approximation to δe expressed in terms of $\delta \mathbf{y}$ is

$$\delta e_1 = 2 (\delta \mathbf{y})^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{M}^{\mathrm{T}} \mathbf{C} (\mathbf{x}^f - \mathbf{x}^t), \quad (14)$$

where **M** and **C** are the matrices with elements m and c, respectively. Note that both $\mathbf{x}^f$ and **M** here are formally those computed for the model forecast trajectory initialized using $\mathbf{x}_b$, although using the trajectory initialized with $\mathbf{x}_a$ would only create a second-order change to this first-order approximation. In (14), $2\, \mathbf{K}^{\mathrm{T}} \mathbf{M}^{\mathrm{T}} \mathbf{C}\, (\mathbf{x}^f - \mathbf{x}^t)$ is formally $\partial e / \partial \mathbf{y}$. For an $\mathbf{x}^f$ determined by $\mathbf{x}_b$, $\partial e / \partial \mathbf{y}$ is independent of $\delta \mathbf{y}$.

For the higher-order approximations it is useful to formally label $\mathbf{x}^f$ for the forecast initialized with $\mathbf{x}_b$ as $\mathbf{x}_b^f$. Correspondingly, the resolvent for that model trajectory is denoted as $\mathbf{M}_b$. Then

$$\mathbf{x}_a^f = \mathbf{x}_b^f + \delta \mathbf{x}^f, \quad (15)$$

$$\mathbf{M}_a = \mathbf{M}_b + \delta \mathbf{M}, \quad (16)$$

where $\delta \mathbf{M}$ has the elements δm. These may be interpreted as respective approximations to the forecast produced from the analysis and the resolvent produced by the trajectory of that forecast. With this notation, the second-order approximation may be written as

$$\delta e_2 = (\delta \mathbf{y})^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \left[\mathbf{M}_b^{\mathrm{T}} \mathbf{C} \left(\mathbf{x}_a^f - \mathbf{x}^t \right) + \mathbf{M}_a^{\mathrm{T}} \mathbf{C} \left(\mathbf{x}_b^f - \mathbf{x}^t \right) \right]. \quad (17)$$

It is derived by substituting (7) and (8) into (3), simplyfing by using (5) and (6), and finally applying (12). This approximation is like eqs (7) and (A10) in LB, except that the adjoint of the forecast model evaluated for the forecast begun from $\mathbf{x}_b$ is applied to the forecast error evaluated for the distinct forecast begun from $\mathbf{x}_a$, and vice versa. This is the opposite of what LB describe in their text. The expression that is multiplied with $\delta \mathbf{y}$ in (17) may be interpreted as a general weighting vector $\tilde{\mathbf{g}}$ as in (13) but not as a formal gradient, unlike as denoted in LB.

The third-order expression is

$$\delta e_3 = (\delta \mathbf{y})^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \left[\mathbf{M}_b^{\mathrm{T}} \mathbf{C} \left(\mathbf{x}_b^f - \mathbf{x}^t \right) + \mathbf{M}_a^{\mathrm{T}} \mathbf{C} \left(\mathbf{x}_a^f - \mathbf{x}^t \right) \right] + \text{term involving } \frac{\partial^3 \mathbf{m}}{\partial \mathbf{x}^3}, \quad (18)$$

derived analogously to (17). This expression is exactly like eqs (7) and (A10) in LB except for the term involving $\partial^3\, \mathbf{m} / \partial \mathbf{x}^3$

and the fact that in LB both $\mathbf{M}_a$ and $\mathbf{x}_a^f$ are computed using the full nonlinear forecast trajectory rather than the linear approximations (15–16). If the model nonlinear terms are restricted to quadratic ones (a restriction that does not omit the dominant advective terms) then the missing term in (18) vanishes. Thus, eq. (7) in LB may be considered as a third-order approximation to δe due to the consideration of observations at the analysis time. As for δe_2, the implied vector appearing as an inner product with $\delta\mathbf{y}$ in (18) is not a formal gradient, contrary to what is indicated by the notation in eqs (7) and (A11) in LB.

3. Discussion

If the distinction between $\mathbf{M}_a$ and $\mathbf{M}_b$ is ignored,

$$\delta e_3 - \delta e_1 = (\delta\mathbf{y})^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{C}\left(\mathbf{x}_a^f - \mathbf{x}_b^f\right). \tag{19}$$

This expression looks like that for e_1 except for the missing factor of 2 and the fact that the last factor on the right hand side is now the difference between the two forecasts rather than the background forecast and truth. The approximation appearing as (19) is second-order in $\delta\mathbf{y}$ since the difference between the forecasts is proportional to $\delta\mathbf{y}$ when the model is linear:

$$\delta e_3 - \delta e_1 = (\delta\mathbf{y})^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{CMK}(\delta\mathbf{y}). \tag{20}$$

As smaller magnitudes of $\delta\mathbf{y}$ are considered, (20) generally becomes negligibly smaller than (14). When (20) is compared to (14) in the context of current operational data assimilation systems, however, the sizes of the first- and second-order terms are in practice not greatly different: For a 2-day forecast, for example, the magnitude of $\mathbf{x}_a^f - \mathbf{x}_b^f$ is approximately half that of $\mathbf{x}_b^f - \mathbf{x}^t$, as shown in Appendix B. It should therefore not be expected that (14) is necessarily an adequate approximation to the change of e due to analysed observations. In contrast, if the model nonlinearity is insignificant, (17–18) are both essentially exact expressions of the δe.

The quadratic (or higher) order nature of (18) has another critical consequence. Although (18) can be written in component form as

$$\delta e_3 = \sum_i \delta y_i \tilde{g}_i^3, \tag{21}$$

where superscript 3 indicates the weights for the third-order approximation, it hides the fact that these $\tilde{g}_i^3$ are also functions of δy (as revealed in (20)). Consequently, while (21) appears to permit the formal consideration of partial sums for arbitrary subsets of observations $i \in S$, those partial sums generally depend on components of $\delta\mathbf{y}$ that are not included in the subset S. In other words, it is problematic to attribute

$$\delta e_{3S} = \sum_{i \in S} \delta y_i \tilde{g}_i^3 \tag{22}$$

entirely to the effect of the observations included in S, as suggested by the presentation in LB. The loss of independent contributions to the approximation δe_3 by independent subsets of observations contrasts with such independence for the approximation δe_1.

4. Conclusion

The expression used in LB to describe the impact of sets of observations on decreasing forecast error has been shown to be a partial sum of quadratic or higher order expressions in terms of the observation innovations. If the model were linear, it would be exactly quadratic. The difference between this expression and a first order expression can be substantial as applied in LB because the magnitude of an error of a forecast generated from a background is not much greater than the magnitude of a difference between forecasts generated from an analysis and its corresponding background. This assertion is based partly on the assumptions and experience in using time-lagged or perturbed-analysis techniques to estimate forecast skill. Although formally of different orders in Taylor series expansion, the first-order terms are not, in application, strongly dominant. Thus, in the application described in LB, considerations beyond first-order may be necessary in order to get a reliable estimate of the effect of observations, as implied by their eq. (7).

The necessity of considering non-linear approximations to the skill improvement unfortunately means that a measure of the impact due to a particular set of observations using a simple partial sum as in (22) is rendered somewhat ambiguous because it includes cross products of innovations within and outside the particular subset. This is a trade-off of obtaining greater accuracy with the higher-order approximation. Experience with the measure in LB and their subsequent, as yet unpublished work, suggests that attribution of their measure to distinct subsets yields reasonable interpretations. Potential ambiguity in the partial sums is not reasonably explained by postulating a lack of correlation between innovations due to distinct observation types. Innovations by distinctly different observations are generally correlated, even if the observation errors are uncorrelated because, at least for nearby locations, the background errors are correlated.

Due to nonlinearity, without evidence to the contrary, it is conceivable that bad observations could alter the estimated impact of nearby good observations computed as a partial sum, or vice versa. It would be useful if experiments could be designed to explore the consequence of the nonlinear nature of (17–18) and (22) so that interpretations of their partial sums could be more unequivocal. If no confusion due to innovation cross products occurs, an explanation is desirable since it may reveal something as yet unclear about how a data assimilation system behaves.

5. Acknowledgments

This work was supported by the United States National Aeronautics and Space Administration grant MAP/04–0000–0080, the Office of Naval Research grant N00014–0310554, and the National Science Foundation grant ATM–0351938. The author thanks Nancy Baker, Rolf Langland, and Ronald Gelaro for suggestions regarding earlier versions of the manuscript.

6. Appendix A

As an example to show that the magnitude of $\mathbf{x}_a^f - \mathbf{x}_b^f$ is not an order smaller than that of $\mathbf{x}_b^f - \mathbf{x}^t$, results from four pairs of forecasts are presented in Table 1. The forecasts and analyses were produced using the state of the art forecast model and 3-D variational data assimilation system employed at the Global Modelling and Assimilation Office at the United States National Aeronautics and Space Administration (GMAO/NASA). The grid resolution was 1° latitude by 1.25° longitude by 55 levels. The four $\mathbf{x}_a^f$ were produced from 24 and 36 h forecasts begun from two analyses begun at 00 UTC on the two randomly selected days 21 January 2005 and 17 February 2005. The corresponding $\mathbf{x}_b^f$ were produced from forecasts begun from analyses 6 hr earlier but valid at the same times as the corresponding $\mathbf{x}_a^f$.

The measures used to evaluate the differences were square roots of the northern hemisphere, area- and mass- weighted mean squared differences in either the temperature T or northward wind component v. The mass-weighting was fractional, using the ratio of the pressure thickness of each model layer to the surface pressure at each grid point. The results indicate that for both cases, forecast times, and sets of fields, the magnitudes of $\mathbf{x}_a^f - \mathbf{x}_b^f$ are approximately one-half those of the corresponding $\mathbf{x}_b^f - \mathbf{x}^t$. Although this will tend to make (14) smaller than (20), it will not render it negligible.

Table 1. The root mean squared values of the indicated differences. The superscripts indicate the forecast durations in hours as begun from each corresponding analysis. Each $\mathbf{x}_a^f - \mathbf{x}_b^f$ should be compared with the $\mathbf{x}_b^f - \mathbf{x}^t$ value immediately below it

Case rms field	January v(m s^{-1})	January T(K)	February v(m s^{-1})	February T(K)
$\mathbf{x}_a^{24} - \mathbf{x}_b^{30}$	1.65	0.68	1.50	0.64
$\mathbf{x}_b^{30} - \mathbf{x}^t$	3.18	1.33	3.05	1.30
$\mathbf{x}_a^{36} - \mathbf{x}_b^{42}$	1.98	0.77	1.75	0.73
$\mathbf{x}_b^{42} - \mathbf{x}^t$	4.16	1.68	3.66	1.55

References

Baker, N. L. and Daley, R. 2000. Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Quart. J. Roy. Meteor. Soc.* **126**, 1431–1454.

Langland, R. H. and Baker, N. L. 2004. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus* **56A**, 189–201.