An approach to assess observation impact based on observation-minus-forecast residuals

Ricardo Todling

Global Modeling and Assimilation Office, NASA/GSFC,
Greenbelt, Maryland

Abstract

An adjoint-based, state-space, technique is now becoming common practice to assess the impact of observations on the forecast. The present note introduces an alternative, observation-space, approach that appears less restrictive and is simpler to implement in practice than its counterpart. A brief summary of a comparison between these two approaches is given here using the NASA GEOS-5 data assimilation system.

1 Introduction

Langland and Baker (2004) introduce a technique to examine the impact of observations on the short-range forecast. In that, a *state-space* aspect of the forecast is defined and changes to the aspect are then associated to changes in the observing system, that are ultimately recast into individual observation impacts. This is an adjoint-based approach involving the sensitivity operator of the entire data assimilation system and various approximations as elaborated by Errico (2007), Trémolet (2008), and Daescu and Todling (2009). As seen by this author, this technique has some limitations, mainly: the subjectivity of the definition of forecast aspect, the need to rely on a verification state, and the usual constraints associated with adjoint-based methods.

Motivated by estimation theory arguments, we introduce an *observation-space* metric that uses observation-minus-forecast (OMF) residuals to calculate observation impacts on the forecast and ameliorates the limitations just mentioned. Under the assumption that the observing system is relatively homogeneous in time, this approach has the additional advantage of being nearly cost-free and simple to implement. As such, the proposed metric is also easily applicable to related ensemble techniques (e.g., Tan et al. 2007; and Liu and Kalnay 2008).

2 Background

The state-space approach introduces a forecast aspect

$$e^s_{k|\ell} \quad \equiv \quad (\mathbf{x}^f_{k|\ell} - \mathbf{x}^t_k)^T \mathbf{T}_k (\mathbf{x}^f_{k|\ell} - \mathbf{x}^t_k)\,, \tag{1}$$

that measures the error in the forecast $\mathbf{x}^f_{k|\ell}$ at time t_k, initiated at time $t_\ell < t_k$, and inquires how changes to the observing system affect this measure. First, an answer can only be obtained by replacing the unknown true state $\mathbf{x}^t_k$ with an available verification state $\mathbf{x}^v_k$. Second, to be able to relate changes δe^s_k to (1) to changes to the forecast initial conditions and ultimately to the observing system, one must assume these changes to be infinitesimal, and only approximations to δe^s_k can be thus be derived. And finally, a suitable choice of weighting matrix $\mathbf{T}_k$ must be made that is not always easy to justify.

This article proposes to evaluate observation impacts directly in observation-space. This can be done by introducing the following, OMF-based, metric

$$e^o_{k|\ell} \quad \equiv \quad \left[\mathbf{h}_k(\mathbf{x}^f_{k|\ell}) - \mathbf{y}^o_k\right]^T \mathbf{C}_k \left[\mathbf{h}_k(\mathbf{x}^f_{k|\ell}) - \mathbf{y}^o_k\right] , \tag{2}$$

where the forecast $\mathbf{x}^f_{k|\ell}$ is converted to observation space by the observation operator $\mathbf{h}_k$, at time t_k. Observation impacts can be derived by calculating the difference $\delta e^o_k \equiv e^o_{k|\ell+1} - e^o_{k|\ell}$, which requires simply evaluation of $e^o_{k|\ell}$ for forecasts issued from two consecutive analyses. Metric (2) has the following advantages over (1): (i) by choosing the weighting matrix to be the inverse of the observation error covariance matrix, $\mathbf{C}_k = \mathbf{R}_k^{-1}$, it defines a rather natural forecast aspect that directly relates to the way observations are weighted in the analysis system, and provides full assessment of the observations — though this is still an arbitrary choice of weights; (ii) it naturally avoids introducing undesirable correlations by circumventing the need for a verifying state; and (iii) it does not involve linearization and use of adjoints, therefore being applicable to any length of forecast.

Theoretical considerations (not presented here) allow us to derive expressions relating the two measures in (1) and (2), and their corresponding observation impacts. Particularly, when the verification is taken to be the analysis, one can explicitly calculate the dependence of the state-space measure on the various intermediate observation-minus-background residuals between the initial and final times of the forecast. Interestingly, in the linear, optimal case, the dependence on these terms disappears as a consequence of the whiteness of the innovations process. Furthermore, it is simple to show that in the linear, optimal case, the choice $\mathbf{T}_k = \mathbf{H}_k\mathbf{C}_k\mathbf{H}_k^T$, where now $\mathbf{H}_k$ is the observation operator, produces a state-space *expected* observation impact that is identical to that calculated in observation space, that is, $< e^s_k > \overset{opt}{=} < e^o_k >$, for $< \bullet >$ representing the expectation operator.

3 Analysis

Evaluation of the various conjectures above has been done using an upgraded version of the GEOS-5 data assimilation system (DAS; Rienecker et al. 2009) that includes the model adjoint used in Errico et al. (2007), and the additional analysis adjoint of Trémolet (2008). We start by evaluating the subjectivity of the state-space metric (1) with respect to the choice of weighting matrix $\mathbf{T}_k$. For that, we employ two slightly different versions of the total (dry) energy norm forming the elements of this matrix. The first, uses the common expression of the linearized total (dry) energy whose fractional weights largely emphasize the troposphere (Fig. 1, left panel, blue curve); the second, uses the mild modification of the energy expression due to Errico et al. (2007), and whose fractional weights largely emphasize the stratosphere (Fig. 1, left panel, red curve). The right panel of Fig. 1 displays observation impacts on the 24-hr forecasts, for all 00 UTC analysis times of August 2007. Separate observation impacts are shown for the main observing systems used in GEOS-5 DAS. The radiosonde and dropsonde network (RaobDsnd) appearing as the dominant observing system under the tropospheric measure, appears secondary when evaluated with respect to the metric emphasizing the stratosphere under which, not surprisingly, AMSU-A dominates. Clearly, for a metric emphasizing the stratosphere, land, ships and surface wind measurements appear not to be of relevance to reduce the 24-hr forecast error; the conclusion is simply that results are subject to the metric chosen.

Next, the left panel of Fig. 2 shows how observation impacts derived with the observation-space approach compare with those from the state-space approach. Fractional impacts are displayed for the cases discussed in Fig. 1, and joined with similar results obtained using the observation-space approach when $\mathbf{C}_k = \mathbf{R}_k^{-1}$ (labeled R-omf). Overall, the observation-space approach indicates that AMSU-A and AIRS observations contribute most to the 24-hr forecast error reduction. A striking result relates to the contribution of the radiosonde and dropsonde network that seems rather small when obtained directly in observation space. Further investigation is taking place to carefully corroborate this result.

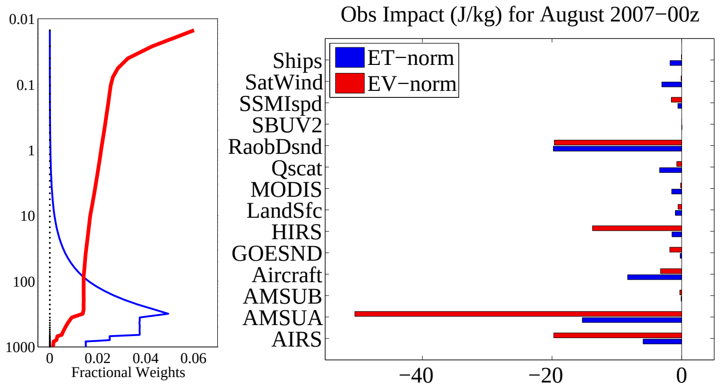

Figure 1: The left panel shows the vertical fractional weights used to calculate the total (dry) energy norm emphasizing: the troposphere (ET; blue curve); and the stratosphere (EV; red curve) — both are shown for a point where $p_s = 1000$ hPa; the model top pressure is 0.01 hPa. (Similar to Fig. 1 of Errico et al. 2007). The right panel shows observation impacts on the 24-hr forecast for various observing systems used in GEOS-5 DAS for the two norms in question.

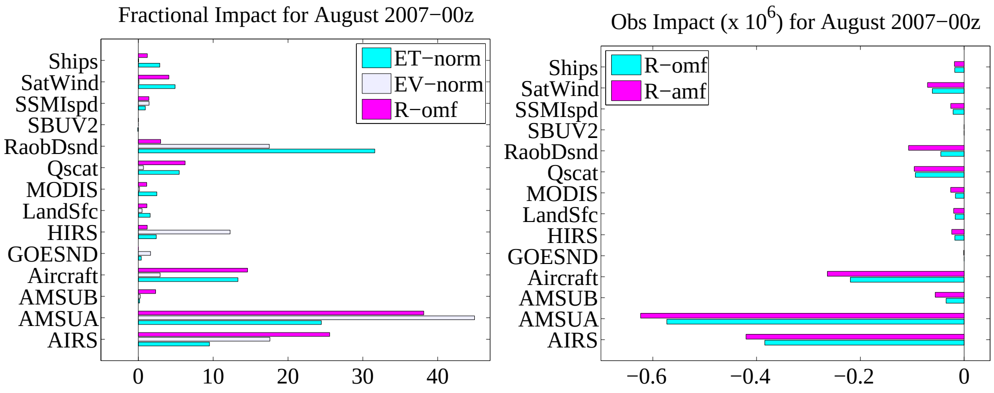

Figure 2: The panel on the left shows fractional impacts calculated with the state-space approach using norms emphasizing the tropospheric (ET-norm) and the stratosphere (EV-norm); the panel also displays fractional impacts calculated with the observation-space approach using OMF residuals (R-omf). The right panel examines the role of the verification by comparing observation impacts on the 24-hr forecasts calculated using the observation-space approach (R-omf) with the impacts calculated in observation-space but when the observations are replaced with the analysis evaluated by the observation operator (R-amf).

A clue in what might perhaps explain some of the main differences between the two approaches is presented in the right panel of Fig. 2, where the observation-space approach is used to examine the role of using the analysis for verification. Results labeled as R-amf have been obtained after applying the observation operator to the analyses and using them in place of the observations to calculate the metric (2). They indicate that use of the analysis for verification amounts to an overestimate of the contribution of the observations to reducing forecast error. Indeed, it seems this is particularly noticeable in the case of radiosondes and dropsondes.

4 Summary

The present work revisits the approach to assess impact of observations on the forecast by introducing a simple and possibly less restrictive technique than what is becoming practice in many data assimilation centers. Theoretical insights (not fully discussed here) provide the motivation for re-examining the subject. Studies of the impact of observations on the 24-hr forecasts indicate that use of observation-minus-forecast residuals may suffice to obtain a fully comprehensive assessment of the observing system, and if so, at almost no cost to operational practice.

5 References

Daescu, D. N., and R. Todling, 2008: Adjoint estimation of the variation in a model functional output due to assimilation of data. *Mon. Wea. Rev.,* **137**, 1705-1716.

Errico, R. M., 2007: Interpretation of an adjoint-derived observational impact measure. *Tellus,* **59A**, 273-276.

______, R. Gelaro, E. Novakovskaia, and R. Todling, 2007: General characteristics of stratospheric singular vectors. *Meteorologische Zeitschrift*, **16**, 621-634.

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

Liu, J., and E. Kalnay, 2008: Estimation of observation impact without adjoint model in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **134**, 1327-1335.

Rienecker, M. M., and co-authors, 2009. The GEOS-5 Data Assimilation System - Documentation of Versions 5.0.1, 5.1.0, and 5.2.0 - NASA TM 104606, Technical Report Series on Global Modeling and Data Assimilation.

Tan, D., E. Andersson, M. Fisher, and L. Isaksen, 2007: Observing system impact assessment using a data assimilation ensemble technique: application to the ADM-Aeolus wind profiling mission. Tech. Memo. 510, ECMWF.

Trémolet, Y. 2008: Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus,* **60A**, 964-978.