Diagnostics for Evaluating the Impact of Satellite Observations

Nancy L. Baker and Rolf H. Langland

Marine Meteorology Division
Naval Research Laboratory
Monterey, CA USA

1. Introduction

This research was originally motivated more than 10 years ago by preparations for the then upcoming FASTEX (Joly et al., 1999) experiment. The adaptive observation targeting methods for FASTEX included the singular vector approach (Palmer *et al.* 1998; Gelaro *et al*. 1999; Bergot *et al.* 1999; Buizza and Montani 1999) and the gradient sensitivity technique (Langland and Rohaly 1996; Bergot *et al.* 1999). These methods sought to identify the sensitive regions where small errors could amplify and dominate the short- to medium-range forecast error (Rabier et al, 1996). However, neither one of these targeting methods was able to take into account how the observations would be used by the data assimilation system, nor the presence of other observations in the region. Partly as a result of this, assimilation of the special targeted observations from FASTEX led to both degraded and improved forecasts (Doerenbecher and Bergot, 2001).

In 1996, Roger Daley[1] derived the equations for this missing piece – that of the adjoint of the data assimilation system. Baker (2000) and Baker and Daley (2000) explored observation adjoint sensitivity in an idealized context. By 2000, Roger Daley had constructed the adjoint of the NRL[2] Atmospheric Variational Data Assimilation System (NAVDAS). As it turned out, the observation-space formulation of NAVDAS was ideal for developing the adjoint, as it simply required applying the analysis operators in a reverse sense. Since that time, this area of research has expanded rapidly, and several operational NWP centers routinely use the adjoint of their data assimilation systems to monitor the quality of the observing network, diagnose assimilation system deficiencies, and make decisions on which observations to assimilate.

An overview of the data assimilation adjoint theory will be presented in Section 2, with idealized examples presented in section 3. A discussion of the choice of an appropriate forecast metric is in Section 4, along with the definition of the observation impact, while section 5 presents several applications of the observation adjoint technique for data monitoring, channel selection for satellite radiometers, and identifying systematic errors in the observations or the way the observations are used by the data assimilation system. Section 6 presents conclusions and discussion of ideas for future use of observation impact.

2. Derivation of observation and background sensitivity

Following Baker and Daley (2000), the derivation of the observation and background sensitivity begins with the three-dimensional analysis problem and the analysis equation (Daley 1991),

$$\mathbf{x_a} = \mathbf{x_b} + \mathbf{K}(\mathbf{y} - H\{\mathbf{x_b}\}) \tag{1}$$

[1] Roger Daley (1943 – 2001) was a senior visiting scientist at the Naval Research Laboratory from 1995 to 2001.
[2] Naval Research Laboratory

Following the notation given by Ide et al. (1997), the vector of observations is given by **y**, the background vector is given by $\mathbf{x_b}$, and the analysis vector is given by $\mathbf{x_a}$. In general, the application of the observation or forward operator H represents any necessary spatial and temporal interpolations from the forecast model background to the observation location and time. If the observed quantity is not directly related to the model state variables, then H also represents the transformation from the forecast values to the observed quantity. For satellite radiances (or brightness temperatures), $H\{x_b\}$ represents the forward radiative transfer model applied to $\mathbf{x_b}$ and computes forecast or background radiances. The differences between the observation and the background $(\mathbf{y} - H\{\mathbf{x_b}\})$ are the components of the *innovation vector*, while $\mathbf{x_a} - \mathbf{x_b} = \mathbf{K}(\mathbf{y} - H\{\mathbf{x_b}\})$ is the *correction vector*. If one considers linear analysis problems only, then

$$\mathbf{x_a} = \mathbf{x_b} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x_b}). \tag{2}$$

The matrix **H** is the Jacobian matrix corresponding to the forward operator $H\{x_b\}$ linearized about the background state vector. This approximation is valid for the radiative transfer relationship between temperatures and radiances, but is not valid for moisture retrievals. The Kalman gain (or weight) matrix, **K**, is given by

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{3}$$

where **B** is the background error covariance matrix and **R** is the observation error covariance matrix.

Equation (2) may be rewritten as

$$\mathbf{x_a} = \mathbf{x_b} - \mathbf{K}\mathbf{H}\mathbf{x_b} + \mathbf{K}\mathbf{y} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x_b} + \mathbf{K}\mathbf{y}, \tag{4}$$

where **I** is the identity matrix. The sensitivity of the analysis to the observations $\partial\mathbf{x_a}/\partial\mathbf{y}$ and the sensitivity of the analysis to the background $\partial\mathbf{x_a}/\partial\mathbf{x_b}$ is derived first. Following Gelb (1974), the vector gradient of a vector is a matrix and is given by $\partial\mathbf{x}^T/\partial\mathbf{y} = \mathbf{A}$, or $\partial x_i/\partial y_k = a_{ki}$. Using this relationship, the following equations may be derived:

$$\partial\mathbf{x_a}/\partial\mathbf{y} = \mathbf{K}^T, \quad \textbf{(5a)}$$

$$\partial\mathbf{x_a}/\partial\mathbf{x_b} = (\mathbf{I} - \mathbf{K}\mathbf{H})^T = \mathbf{I} - \mathbf{H}^T\mathbf{K}^T, \tag{5b}$$

where

$$\mathbf{K}^T = (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}. \tag{6}$$

The cost function J is a scalar measure of some quantity of interest over the forecast verification domain. The gradient or sensitivity of J with respect to the analysis or initial conditions for the forecast is given by $\partial J/\partial\mathbf{x}_a$, and is determined by the adjoint of the tangent forward propagator based on the nonlinear trajectory. In this paper, this quantity will be referred to as the analysis sensitivity vector.

The sensitivity of the forecast aspect to the observations ($\partial J/\partial\mathbf{y}$), is referred to as the *observation sensitivity vector*, while the sensitivity of the forecast aspect to the background field ($\partial J/\partial\mathbf{x}_b$), is referred to as the *background sensitivity vector*. Using the chain rule and (5), the observation and background sensitivity vectors may be written as

$$\partial J/\partial\mathbf{y} = \frac{\partial\mathbf{x_a}}{\partial\mathbf{y}}\frac{\partial J}{\partial\mathbf{x_a}} = \mathbf{K}^T\,\partial J/\partial\mathbf{x}_a, \tag{7a}$$

$$\partial J/\partial \mathbf{x_b} = \frac{\partial \mathbf{x_a}}{\partial \mathbf{x_b}} \frac{\partial J}{\partial \mathbf{x_a}} = (\mathbf{I} - \mathbf{H}^T \mathbf{K}^T)\, \partial J/\partial \mathbf{x}_a \tag{7b}$$

Expanding the terms in the transposed Kalman gain matrix ($\mathbf{K}^T$) in (6) using (3) gives the following expressions for the observation and background sensitivity vectors,

$$\partial J/\partial \mathbf{y} = (\mathbf{HBH}^T + \mathbf{R})^{-1} \mathbf{HB}\, \partial J/\partial \mathbf{x}_a\,, \tag{8a}$$

$$\partial J/\partial \mathbf{x}_b = \left[\mathbf{I} - \mathbf{H}^T (\mathbf{HBH}^T + \mathbf{R})^{-1} \mathbf{HB}\right] \partial J/\partial \mathbf{x}_a\,. \tag{8b}$$

In the analysis problem (1), $\mathbf{K}$ is the matrix of weights given to the *innovation* or difference between the observation and the background value (in observation space). In an analogous sense, the matrix $\mathbf{K}^T$ is the matrix of weights given to the sensitivity of *J* with respect to the analysis for the observation sensitivity problem (7a).

It is also convenient for display purposes to define the *analysis space projection of the observation sensitivity vector,*

$$\mathbf{H}^T\, \partial J/\partial \mathbf{y} = \partial J/\partial \mathbf{x}_a - \partial J/\partial \mathbf{x}_b\,. \tag{9}$$

Thus, (8a) and (8b) give a method of calculating the gradient (or sensitivity) of the forecast aspect with respect to the observations $\partial J/\partial \mathbf{y}$ and the gradient of the forecast aspect with respect to the background field $\partial J/\partial \mathbf{x}_b$.

3. Idealized Examples of Observation Sensitivity

In this section, two idealized examples from Baker (2000) are presented to illustrate some of the basic principles of observation and background sensitivity. For the first example, consider a two-dimensional horizontal univariate (e.g., geopotential height) case. The gradient of J with respect to the initial conditions $\partial J/\partial \mathbf{x}_a$ is specified on the two-dimensional domain using simple trigonometric and exponential functions and is shown in Fig. 1, where Fig. 1a has small spatial scale features and Fig. 1e has large spatial scale features. A hypothetical situation in which $\partial J/\partial \mathbf{x}_a$ straddles a coastline is simulated in Fig. 1, with a data-dense continent on the left-hand side of the figure and a data-void ocean is on the right-hand side of the figure.

The observation locations are shown by the "+" signs in Figs. 1b-c and 1f-g. An observation is assumed to be available at every gridpoint, and the observation errors are assumed to be uncorrelated. Over the continent, the observation error variance $\varepsilon_r^2 = 1.0$, while over the ocean, the observation error variance is set to 1.0 x $10^{6.}$ As the background error variance equals 1.0, this allows the observation sensitivity (which is in observation space) to be readily contoured. The background error covariance **B** is specified using the special Second Order Autoregressive Function (SOAR) where ρ_b is the background error correlation between two analysis grid points given by

$$\rho_b(x_i, x_j, y_i, y_j) = \left(1 + r/L_b\right) \exp\left(-r/L_b\right), \tag{10}$$

L_b is the background error correlation length scale, and r is the distance between grid points. The background error correlation length is specified as $L_b = 2.42\Delta x$.

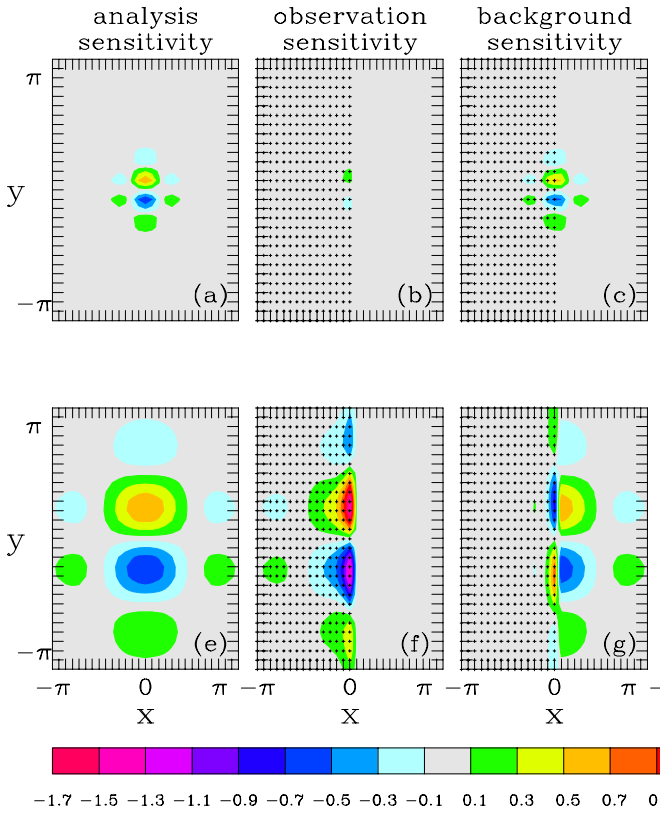

Figure 1: Simulated sensitivity vectors for small- (a-c) and large-scale (e-g) analysis sensitivity vectors. The imposed analysis sensitivity vectors are shown in (a) and (e), the observation sensitivity vectors in (b) and (f), and the background sensitivity vectors in (c) and (g). The observation locations are given by the "+". The color scale is indicated along the bottom. (Panels d and h are not shown). From Baker and Daley (2000).

The observation and background sensitivity vectors corresponding to the small-scale analysis sensitivity vector are shown in Figs. 1b and 1c, respectively, while the observation and background sensitivity vectors corresponding to the large-scale analysis sensitivity vector are shown in Figs. 1f and 1g.

We first consider the well-observed continental interior portion far from the coastline. For the small-scale case (Figs. 1a-c), even though the analysis sensitivity gradient is well sampled in the continental interior, there is no sensitivity to the observations, only to the background. For the large-scale case of Fig. 1e-f, the observation sensitivity is the same as the analysis sensitivity in the continental interior and there is no sensitivity to the background. This apparent discrepancy occurs because the background error correlation length scale is relatively large, implying that the background errors are primarily large-scale, and the analysis will use the observations primarily to reduce the large-scale errors (Daley, 1991). Because the small-scale background errors are implicitly assumed to be relatively small, the observations have very little effect on these small spatial scales. Thus, for this example, the small-scale features of the analysis are derived primarily from the background and the large-scale features are derived primarily from the observations. In an adjoint context, this means that the background sensitivity will be derived primarily from the small scales of the analysis sensitivity and the observation sensitivity will be derived primarily from the large scales.

The region along the boundary between the well sampled and the unsampled areas is considered next. For the small-scale analysis sensitivity gradient, the sensitivity to coastal observations is only slightly larger than in the well-sampled continental interior (Fig. 1b). For the large-scale case (Fig. 1e-g), the situation is completely different. In a narrow region along the coastal boundary, both the sensitivity to the observations (Fig. 1f) and the sensitivity to the background (Fig. 1g) are greater in magnitude than the analysis sensitivity

(Fig. 1e) at the same gridpoint. The background sensitivity is of opposite sign to the observation and analysis sensitivities in this coastal region, which is consistent with (9). This phenomenon was defined by Baker and Daley (2000) as observation and background super-sensitivity, respectively, and can be explained with the following example.

For this experiment, the analysis sensitivity gradient (Fig. 2a) is composed of a combination of the large- and small-scale sensitivity patterns from Fig. 1. The background error correlation is modeled using the SOAR function for heights from (1) with $L_b = 3.6\Delta x$; this value was chosen as it maximizes the observation sensitivity for multiple observations. The background error standard deviation is set to 1.0 and the observation error standard deviation is set to 0.1. Twenty observations are placed at gridpoints in a "Z" shape across the centers of the analysis sensitivity gradient pattern (Fig. 2a).

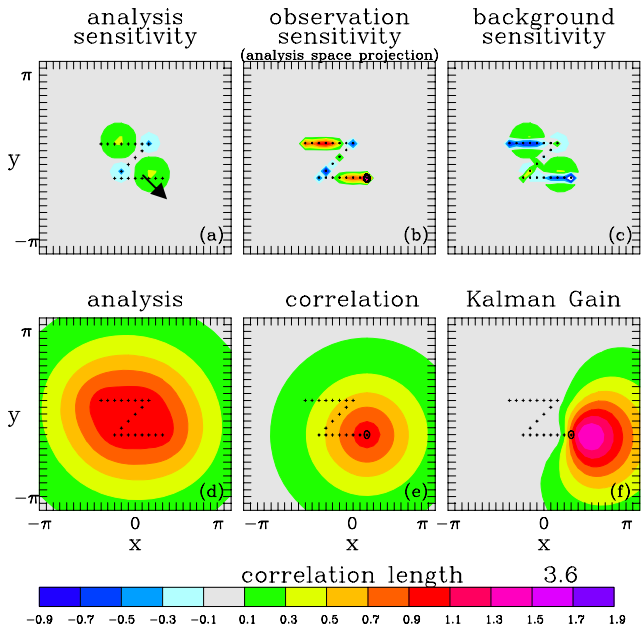

Figure 2: Two-dimensional univariate height observation sensitivity example for multiple observations. The largest sensitivity occurs for the height observation identified by the circle; the other height observations are given by the "+". In this plot, Lb = 3.6Δx, and εr/ε b = 0.1. (a) The imposed analysis sensitivity gradient, (b) the analysis space projection of the observation sensitivity vector, (c) the background sensitivity, (d) the corresponding analysis, assuming an innovation of 1.0 m at each observation location, (e) the background error correlation function corresponding to the circled observation, and (f) the row of KT for the circled observation. The color scale is at the bottom. From Baker (2000).

The resulting analysis space projection of the observation sensitivity vector ($\mathbf{K}^{\mathrm{T}}\,\partial J/\partial \mathbf{y}$) and the background sensitivity vector are displayed in Figs. 2b,c respectively. The largest observation sensitivity occurs for the circled observation at the lower right end of the "Z" in Fig. 2b. The effect of the observation density gradient is more pronounced when the density change occurs in a region of significant analysis sensitivity gradient amplitude (e.g., for the circled observation). When the observations extend all of the way across the analysis sensitivity gradient, the amplitude of the observation sensitivity is less (for example, for the observation indicated by the arrow). Thus, the largest observation sensitivity does not necessarily occur where the analysis sensitivity gradient is a maximum, but where there is a large change in observation density and the analysis sensitivity gradient is both large scale and sufficiently large in magnitude. This result is analogous to the coastal example in Fig. 1.

Assuming that the background field is zero and the observations equal one, the resulting two-dimensional analysis from (2) is given in Figure 2d. The homogenous, isotropic nature of the correlation function, shown

for the circled location in Fig. 2e, is evident and only a hint of the "Z" configuration of the observations can be seen.

The dependence of the univariate observation sensitivity on the observation density can be understood by graphically examining the various terms in the observation sensitivity equation, e.g.,

$$\partial J/\partial \mathbf{y} = \mathbf{K}^T \, \partial J/\partial \mathbf{x}_a = (\mathbf{HBH}^T + \mathbf{R})^{-1} \mathbf{HB} \, \partial J/\partial \mathbf{x}_a \, .$$

The term $\mathbf{HB}$ is the background error correlation between the observation locations and every gridpoint. The row of this matrix corresponding to the circled observation is shown in Fig. 2e. It is symmetric in appearance and is essentially the same for every observation location (given the constraints of a finite domain). The inverse error covariance matrix $(\mathbf{HBH}^T + \mathbf{R})^{-1}$ is in observation space and is not plotted. The term $\mathbf{HBH}^T$ is the background error correlation between observation locations. Since the observation errors are assumed to be spatially uncorrelated, the matrix **R** is simply the diagonal matrix of the observation error variances ε_r^2. The row of the transpose of the Kalman gain matrix ($\mathbf{K}^T$) corresponding to the circled observation is plotted in Fig. 2f. The resulting pattern is not symmetric, but has large values adjacent to the circled observation.

Plots of the appropriate row of $\mathbf{K}^T$ for all of the 20 observations are shown in Fig. 3. The most striking features are the very large lobes (in both size and magnitude) that occur for observations that are relatively isolated from their neighbors. Observations that are located near the center of the pattern have much smaller maxima and minima of the Kalman gain (in both size and magnitude). These variations in $\mathbf{K}^T$ are due to the matrix $(\mathbf{HBH}^T + \mathbf{R})^{-1}$. Observations that are farther from the other observations are less correlated with them and this leads to the large asymmetry in the Kalman gain. For the forward analysis problem, this implies that isolated observations contain more independent information than observations with close neighbors and are thus given greater weight in the analysis. In the adjoint sense, this indicates that isolated observations have larger "adjoint weights" ($\mathbf{K}^T$) and potentially greater observation sensitivity.

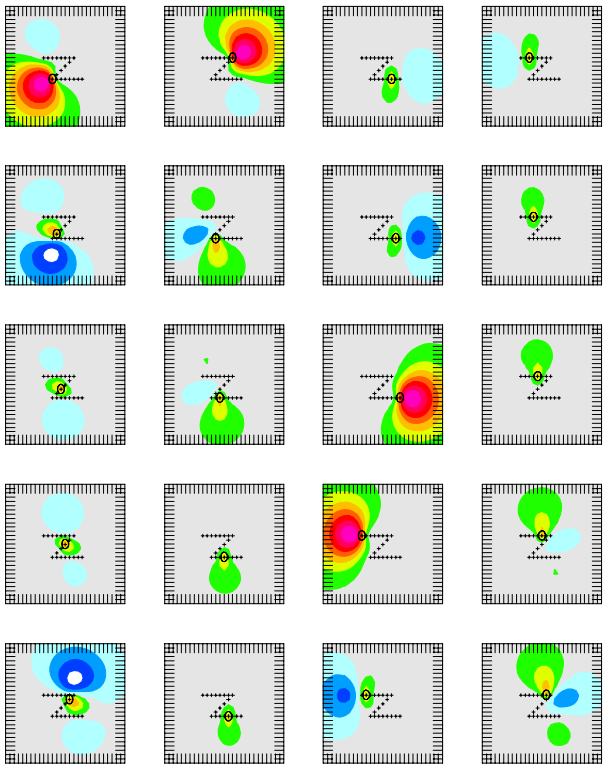

Figure 3: Plots of the row of K^T for the 20 observations shown in Fig. 2. The row of K^T corresponds to the circled observation in each panel. The grid domain and color scale corresponds to Fig. 2. From Baker (2000).

The resulting observation sensitivity vector (8a) is the result of the matrix-vector multiplication between $\mathbf{K}^{\mathrm{T}}$ and $\partial J/\partial \mathbf{x}_a$. This can be visualized for any given observation by mentally summing, gridpoint by gridpoint over the domain, the product of $\partial J/\partial \mathbf{x}_a$ (Fig. 2a) and the row of the $\mathbf{K}^{\mathrm{T}}$ corresponding to that observation (Fig. 2e or Fig. 3). This exercise can be used to explain the larger observation sensitivity value for the observation at the lower-right end of the "Z" (indicated by the circle) when compared to the observation at the upper-left end of the "Z" (indicated by an arrow). The magnitude and shape of the row of $\mathbf{K}^{\mathrm{T}}$ is the same for both observations. However, the circled observation is located in a region with larger values of the analysis sensitivity gradient, so that the non-zero portions of $\mathbf{K}^{\mathrm{T}}$ overlap larger values of the analysis sensitivity gradient. Hence, the circled observation has the larger observation sensitivity.

It is evident that the sensitivity for a given observation depends upon the overlap between the amplitude and spatial extent (length scale) of the appropriate row of $\mathbf{K}^{\mathrm{T}}$, and the amplitude and spatial extent (length scale) of the analysis sensitivity gradient $\partial J/\partial \mathbf{x}_a$. The observation sensitivity is maximized when the maxima or minima of the analysis sensitivity gradient coincide with the maxima or minima of the row of $\mathbf{K}^{\mathrm{T}}$ such that the overall observation sensitivity contributions from the projection of $\mathbf{K}^{\mathrm{T}}$ onto $\partial J/\partial \mathbf{x}_a$ are of the same sign.

The Kalman gain or weight matrix is arguably the most important term for both the linear analysis and observation adjoint sensitivity problems. The properties (amplitude, sign, and length scale) of this matrix are a function of several factors, with the background error correlation length scale being the dominant factor. The scale of $\mathbf{K}^{\mathrm{T}}$ decreases as L_b decreases, and the influence of the analysis sensitivity gradient in the immediate vicinity of the observation becomes more marked (with correspondingly less influence from the adjacent sub-structures of the analysis sensitivity gradient). In addition, neighboring observations will not contribute to the observation sensitivity at a particular location if the correlation length scale is too small. If L_b is too long, then the maximum amplitude of $\mathbf{K}^{\mathrm{T}}$ may not coincide with either the observations or analysis sensitivity gradient extrema (particularly for the more isolated observations).

4. Application to Real Problems

Observation sensitivity, while an interesting quantity in its own right in that it lends insight into how observations are used by the assimilation system, is not capable of quantifying whether a given observation/observing system improves or degrades the forecast. An observation may have large observation sensitivity, but if the innovation is zero, its impact on the resulting analysis and forecast will be minimal, other than through its effects on $\mathbf{K}^{\mathrm{T}}$. Similarly, an observation may have small observation sensitivity, but that does not imply that the resulting changes to the analysis will be small, only that they will not be in the direction needed to effectively change for forecast aspect *J*. To estimate the impact of the observations on the forecast, some other measure of observation impact is required.

4.1. Forecast Metrics and the Cost Function J

The first step is to define the cost function *J* or aspect of the forecast of interest. In principle, it could be defined as any first-order differentiable function of the model forecast. The tangent linear approximation generally limits the forecast length to three days or less. One frequently-used forecast aspect is the quadratic measure of forecast error given by

$$e_f = \left\langle (\mathbf{x}_f - \mathbf{x}_t), \mathbf{C}(\mathbf{x}_f - \mathbf{x}_t) \right\rangle \tag{11}$$

where $\mathbf{x}_f$ is the forecast and $\mathbf{x}_t$ is the verification state, given by the analysis valid at the forecast time. The matrix $\mathbf{C}$ is symmetric matrix of weights, typically for either a dry or moist static total energy, for all levels between the surface and around 250 hPa.

4.2. Observation Impact Calculations

Doerenbecher and Bergot (2001) used a linear combination of the innovation and the observation sensitivity observation impact to assess the impact of the FASTEX special observations. Subsequently, Langland and Baker (2004) developed the following definition for observation impact.

We begin by considering two 24-hr forecast trajectories, one starting from the analysis background (nominally a 6-hr forecast from the previous update cycle), and one starting from the current analysis (t = 0 hrs) as shown in Fig. 4. The forecast error for the background trajectory is given by e_g, while the forecast error in the analysis trajectory is given by e_f. The cost function *J* is defined to be the difference between the forecast errors ($\Delta e_f^g = e_f - e_g$), and is due solely to the assimilation of the observations at the analysis time (t = 0 hrs). Following Langland and Baker (2004), Δe_f^g can be estimated using the observation sensitivity gradients as

$$\delta e_f^g = \left\langle \left(\mathbf{y} - \mathbf{H}\mathbf{x}_b\right), \frac{\partial J_f^g}{\partial \mathbf{y}} \right\rangle = \left\langle \left(\mathbf{y} - \mathbf{H}\mathbf{x}_b\right), \mathbf{K}^T \left\{ \frac{\partial e_f}{\partial \mathbf{x}_a} + \frac{\partial e_g}{\partial \mathbf{x}_b} \right\} \right\rangle, \tag{12}$$

where $\partial J_f^g / \partial \mathbf{y}$ is the observation sensitivity obtained using the two trajectories.

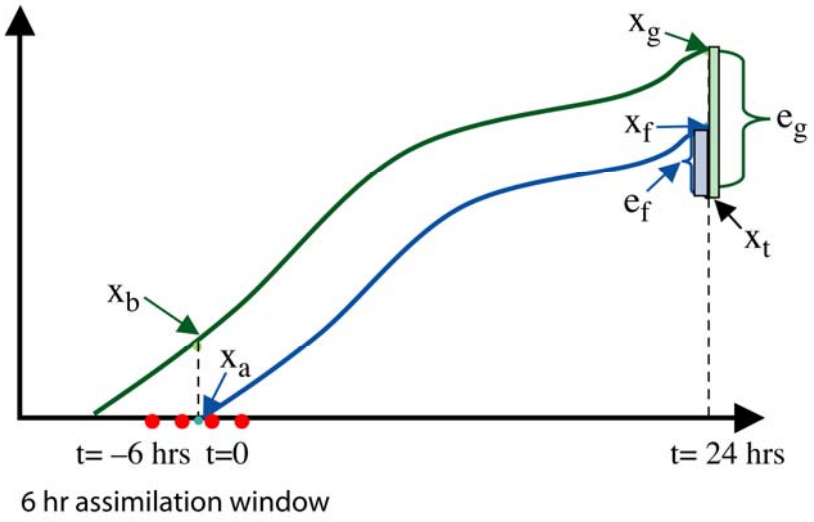

Figure 4: Illustrations of the two trajectories used for the calculation of the observation impact. The trajectory ending at x_g, with error e_g, begins from the background x_b, while the trajectory ending at x_f, with error e_f, begins from the analysis x_a. Both trajectories are verified against the analysis x_t at t=24 hrs. The assimilation of the observations at t=0 hrs moves the forecast trajectory from the dark green line to the blue line. From Langland and Baker (2004).

Equation (12) provides a single measure of the global observation impact for all observations. Negative values indicate a reduction in forecast error, while positive values indicate an increase in forecast error. The global observation impact can be easily partitioned into observation subsets, for example, for a given observing platform (e.g. radiosondes), or by satellite channel. It is important to remember that the computation of the observation impact in (12) always involves the entire set of observations, and changing the properties of even one observation changes the scalar measure for all other observations.

The accuracy of the observation impact results depends upon the accuracy of the adjoint-based calculations in (12). One way to assess this accuracy is to compare the nonlinear forecast error differences $e_f - e_g$ to

those computed using adjoint techniques. In Fig. 5, the red line indicates the full nonlinear NOGAPS[3]model moist static energy error in J kg^{-1} (ordinate) for the forecasts starting from $\mathbf{x}_b$ for the period from 14 Nov. 2006 to 3 Jan. 2007 (abscissa). The yellow line indicates the same error metric, but for the forecast trajectory starting from $\mathbf{x}_a$. The difference in these two errors ($e_f - e_g$) is given by the blue line, while the green line is the adjoint estimate of the error difference, as computed by (12) using the adjoint of NAVDAS[4]. The adjoint-based estimate recovers, on average, approximately 85% of the nonlinear forecast error differences. This percentage is very similar to values obtained using the same observation impact measure with the NASA GEOS-5[5] system (Gelaro, personal communication) and Meteorological Research Division of Environment Canada (EC) systems[6] (Pellerin, personal communication). The systematic underestimation of the actual nonlinear forecast error differences Δe_f^g by observation adjoint impact δe_f^g is likely due to neglected moist physics in the TLM (i.e. convection), and nonlinearities in the dry dynamics. It is interesting to note how closely the adjoint-based estimates track the day-to-day variations in the nonlinear forecast error differences.

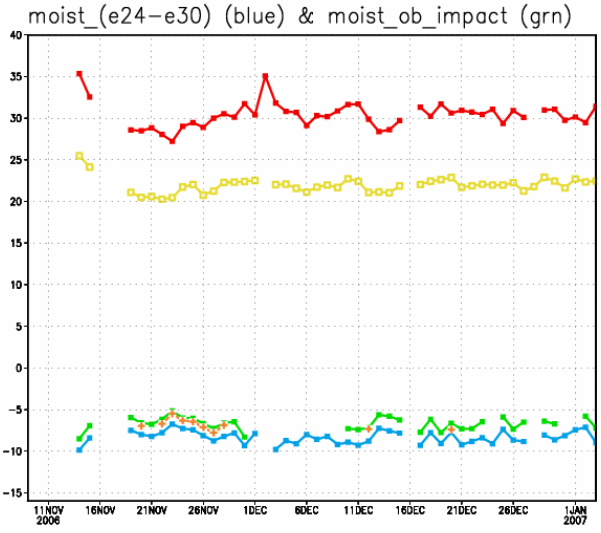

Figure 5: Comparison of the nonlinear forecast error for the 30-hr forecast starting from x_b (red line) and the 24-hr forecast starting from x_a (yellow line). The difference in these errors is given by the blue line, while the green line denotes the adjoint-based impact computed from (12).

4.3. Nonlinearity Considerations

Errico (2007) and Gelaro et al. (2007) have examined the effects on nonlinearity on the interpretation of the partial sums used to bin the observation impact by type (platform, station, channel, etc.) Their results indicate that the observation impact measure proposed in Langland and Baker (2004) is essentially equivalent to the 3rd order approximation of the change in a quadratic measure of forecast error (see above) due to the observations. Gelaro et al. (2007) demonstrated that a higher than first order approximations is required to accurately assess the observation impact. The dominant nonlinearity in (12) is apparently due to

[3] Navy Operational Global Atmospheric Prediction System. NOGAPS is the operational spectral forecast model and is described in Hogan et al. (1999). The adjoint is described in Rosmond (1997).

[4] NAVDAS is an observation-space 3D-Var system, and is described in Daley and Barker (2001). The NAVDAS adjoint development is discussed in Langland and Baker (2004b).

[5] The GEOS-5 system consists of the Goddard Modeling and Assimilation Office atmospheric forecast model and Gridpoint Statistical Interpolation (GSI) data assimilation system and respective adjoints (see Zhu and Gelaro, 2007)

[6] The EC group used observation-space formulations of their 3D-Var and 4D-Var systems (Morneau et al., 2006).

the quadratic nature of the cost function (11), although other nonlinearities arise because the second and third order terms have a dependence upon the innovations and the forecast starting from the analysis $\mathbf{x}_a$. Despite these nonlinearities, the authors concluded that observation impact measure proposed by Langland and Baker (2004) yields reasonable estimates. However, they noted that their assessment was made using large numbers of observations, and cautioned that these conclusions may not hold for small subsets of observations, or even large subsets with comparable impact magnitudes.

4.4. Practical Considerations for DA Adjoint Design

Since NAVDAS is in observation space, and solves (1) directly, it is a relatively straightforward task to construct the adjoint using the same subroutines (analysis operators), but with different input and output, and applied in a different order (see Langland and Baker, 2004b for details). The adjoint of NASA GEOS-5 was constructed as a line-by-line adjoint, in part because the GSI is in model space, and to specifically account for and quantify the contributions of GSI nonlinearities on the observation impact. The Meteorological Research Division of EC opted to build the dual (PSAS) formulations of their 4D-Var and 3D-Var systems, and to construct the adjoint from those systems using the analysis operators (Morneau et al., 2006). ECMWF solves for the observation sensitivity by rewriting (7a) as

$$\partial J/\partial \mathbf{y} = \mathbf{K}^T \, \partial J/\partial \mathbf{x}_a = \mathbf{R}^{-1}\mathbf{H}\mathbf{A} \, \partial J/\partial \mathbf{x}_\mathbf{a} \,, \quad (13) \tag{13}$$

where **A** is the analysis error covariance, and the observation sensitivity is computed using the Krylov subspace method (Cardinali, personal communication). The observation impact is computed by

$$\delta J = \frac{\partial J}{\partial \mathbf{y}} (\mathbf{y} - \mathbf{H}\mathbf{x}_b) \,. \tag{14}$$

5. Application of Observation Adjoint Techniques

The observation impact has been successfully used by scientists at the NRL and Fleet Numerical Meteorology and Oceanography Center (FNMOC) to diagnose a number of data quality issues in operational observation suite used by NAVDAS. The observation impact is routinely computed at 00 UTC for the operational real-time NAVDAS/NOGAPS forecast run. The archived observation impact statistics have been used to identify errors in the observation meta data, such as incorrect station elevations in the Master Station List. These statistics also enabled us to correctly identify a temporary wind direction bias of around 30^o with the Lihue (91165) radiosonde observations during December 2004. We have also used it to justify the continuation of certain U.S. military radiosonde stations.

One obvious question is whether removing observations flagged as non-beneficial translates to an increase in forecast skill. The following example illustrates that this is indeed the case. During January and February 2006, we noticed that the MTSAT atmospheric motion vectors (AMVs) produced by the Cooperative Institute for Meteorological Satellite Studies at University of Wisconsin were systematically providing non-beneficial impact (as indicated by the red shading) in the southern hemisphere (see Fig. 6a). The corresponding mean zonal wind innovations were also unusually large and negative (the blue shading in Fig. 6b). After reviewing the observation impact statistics, we opted to remove all MTSAT AMVs more that 39^o from the satellite sub-point. An Observation System Experiment (OSE) was run, removing this data from the assimilation system for the period of 16 February to 27 March 2006. The 500 hPa height anomaly correlations indicated around 3-hrs forecast improvement in both hemispheres at 5 days. We simultaneously contacted Chris Velden and colleagues at CIMSS concerning the problem. The problem was quickly traced to a processing issue related to a lack of a unique time stamp for each scan-line, resulting in a wind speed bias that increased to the south (as a function of true scan time). CIMSS rapidly implemented a fix to their

wind processing system, and the observation impact and mean innovation values decreased to within their normal ranges (not shown).

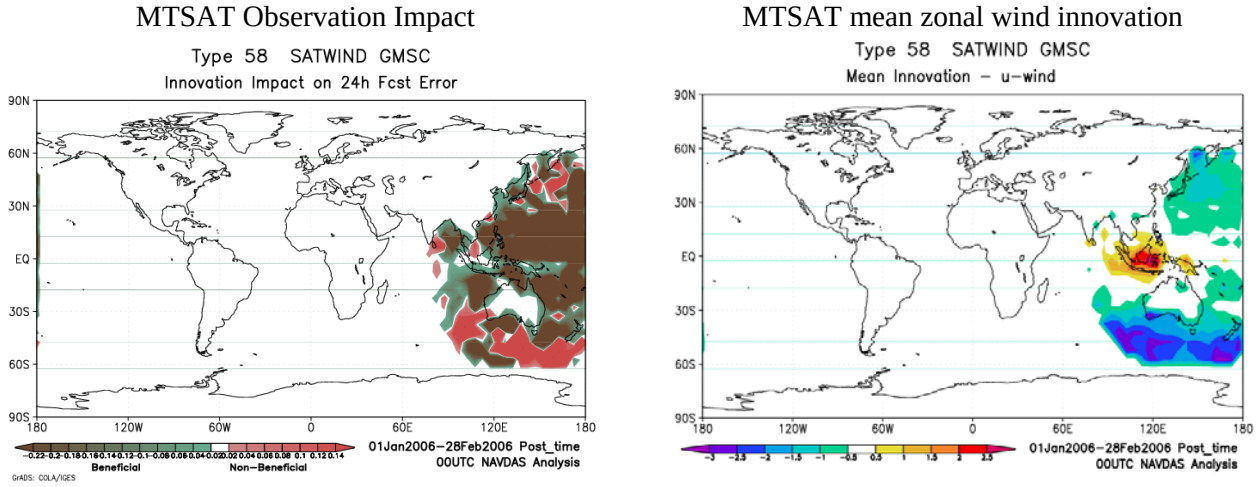

Figure 6: Total observation impact (left panel; J kg^{-1}) for MTSAT winds for the period from 01 Jan. 2006 to 28 Feb. 2006. The green shading indicates beneficial impact, while the red shading indicates non-beneficial impact. The mean zonal wind innovation (m s^{-1}) for the same period is shown in the right panel. The slow wind speed biases due to the observation processing problem are shown by the blue shading in the southern hemisphere.

5.1. Channel Selection for AIRS

The observation impact has also been used with the NASA AQUA AIRS (Advanced Infrared Sounder) radiance observations as a method to identify beneficial and non-beneficial channels for assimilation (Ruston, 2006). With the observation impact approach, the observation impact is generated from a several-week assimilation run using the subset of channels under consideration. The observation impact statistics are generated, and the non-beneficial observations are eliminated from the next forecast trial. This approach differs from other channel selection methods, such as entropy-reduction, adjoint sensitivity and Kalman filter sensitivity. The entropy- reduction method iteratively selects the satellite channel that most effectively reduces the background error (Rodgers, 1996; Fourrié and Rabier, 2004). The adjoint sensitivity method (Doerenbecher and Bergot, 2001) is also an iterative approach, and preferentially selects the channel with the maximum observation sensitivity. The Kalman filter sensitivity (Bergot and Doerenbecher, 2002) iteratively chooses the channel that gives the maximum decrease in the forecast error variance. Fourrié and Rabier (2004) showed that the last 3 methods produced similar results (in terms of the computed degrees of freedom for signal), even though the selected channels are not the same. In particular, the adjoint-based methods tend to favor channels with information in the sensitive regions of the lower troposphere, while the entropy-reduction methods tend to include more information from the upper troposphere.

An example of the AIRS and AMSU observation impact for the NAVDAS/NOGAPS system for an assimilation run from 19-25 August 2006 is shown in Fig. 7. For this example, the initial channel set was selected based on channel usage at other operational NWP centers. The channels are grouped by frequency and labeled by channel number. Positive numbers along the abscissa indicate an increase in forecast error (in J kg^{-1}) due to that channel, while negative numbers indicating a reduction in forecast error. The forecast metric is the moist static energy error norm, as described earlier.

Some of the non-beneficial impact is easily explained – in the operational version of NAVDAS, only linear observation operators are allowed, and only those with sensitivity to the temperature state vector. Thus, it is

not surprising that the water vapor channels increase the forecast error. The responses for AMSU channels 8 and 10 are not as readily explained, nor are the varied responses for the short wave CO_2 channels.

The channels that contributed to the forecast degradation were removed, along with the ozone and water vapor channels, and the assimilation run was repeated. The results are presented in Fig. 8.

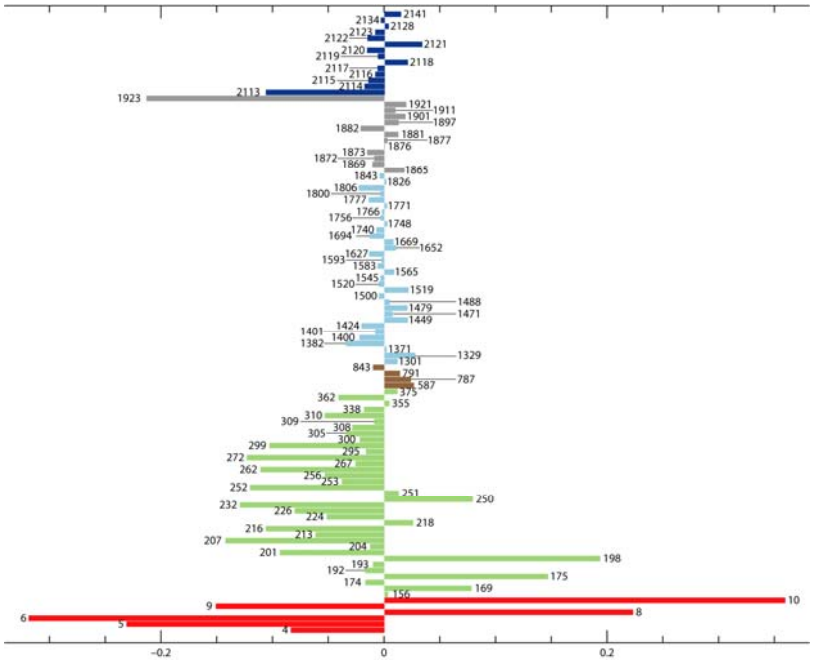

Figure 7: Observation impact (absicca, in J kg^{-1}) computed using moist static energy-weighted error norm for the NAVDAS/NOGAPS system for Aug. 19-25, 2006. The impact for the AIRS and AMSU channels is plotted along the ordinate, where the numbers beside the bars indicate the channel number, with color-coding as follows: short-wave CO_2 (blue and gray); water vapor (light blue); long wave window (brown), long wave CO_2 (green), and the AMSU microwave (red). Figure courtesy of Ben Ruston.

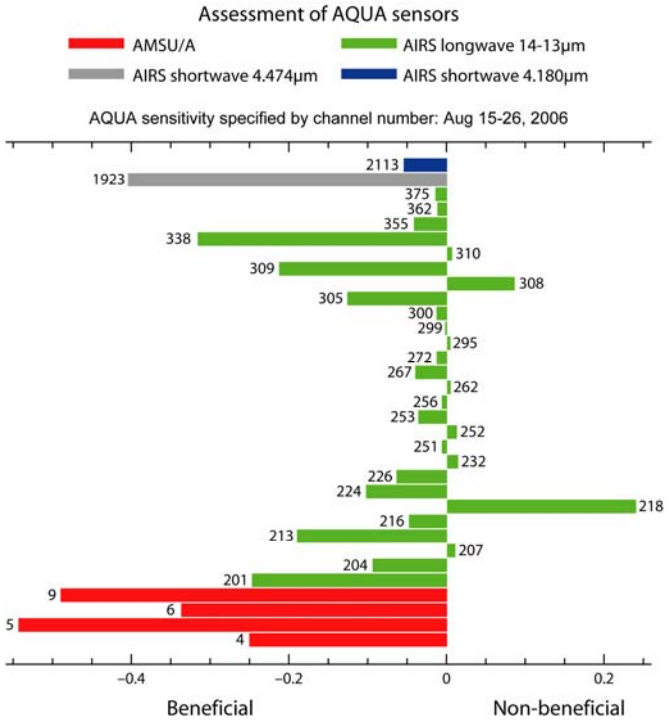

Figure 8: Same as Fig. 7, but for the observation impact after the non-beneficial channels were removed. Figure courtesy of Ben Ruston.

Overall, the majority of the channels now act to reduce the forecast error. Interestingly, some of the channels that were previously beneficial are now non-beneficial. We don't know why this occurs, however it's possible that examining the problem in a 1D-Var context would lend some insight into its nature. It's also possible that this is an indication of the nonlinearity aspects of the problem discussed in section 4.3.

5.2. NWP Center Inter-comparison Studies

The Joint Center for Satellite Data Assimilation (JCSDA) has recently funded a collaborative effort between NRL and GMAO to evaluate and compare the observation impact for all in-situ and satellite observations used in the NAVDAS-NOGAPS and GEOS-5 forecast systems. The results will be used as guidance for data assimilation quality control, data thinning and observation selection, bias correction, and tuning of observation and background error. An example of this approach is shown in Figs. 9a-d, for the AMSU-A observation impact for the NRL, GMAO, EC and ECMWF systems, respectively.

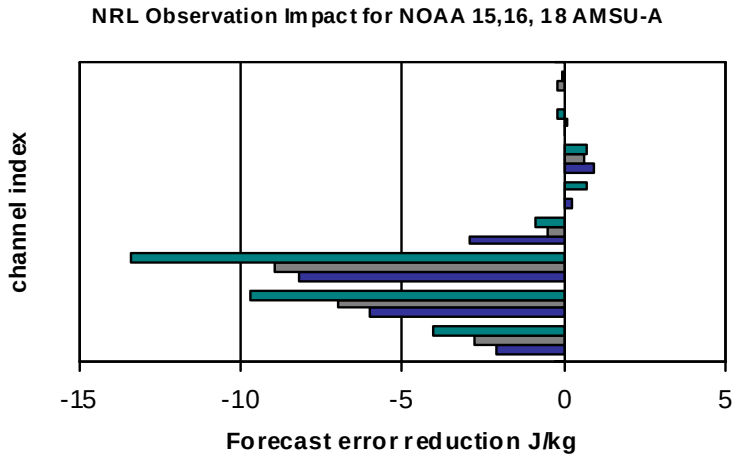

Figure 9a: Total observation impact (abscissa in J kg^{-1}) versus channel index (ordinate) for NOAA-15 (blue), 16 (gray) and 18 (green) AMSU-A observations assimilated by NAVDAS/NOGAPS for 1 Jan. to 28 Feb. 2006 at 00 UTC.

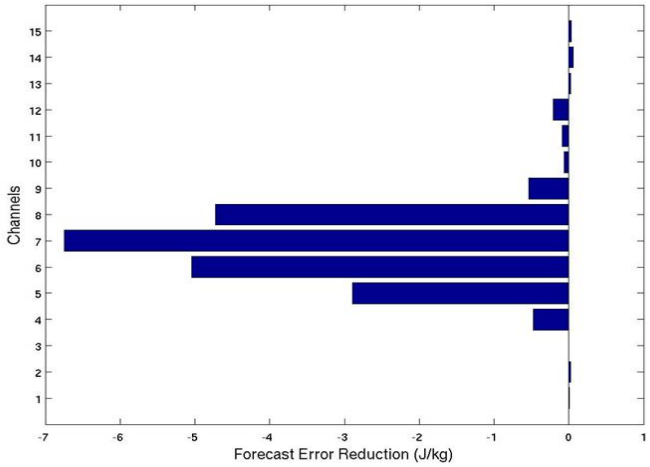

Figure 9b: Same as for Figure 9a, but for the GMAO GEOS-5 system for July 2005 at 00 UTC. Note that this is for a one-month period, so the total forecast impact is roughly half that in Figure 9a. The impact for NOAA-15 and NOAA-16 are summed together. Figure provided courtesy of Ron Gelaro, GMAO.

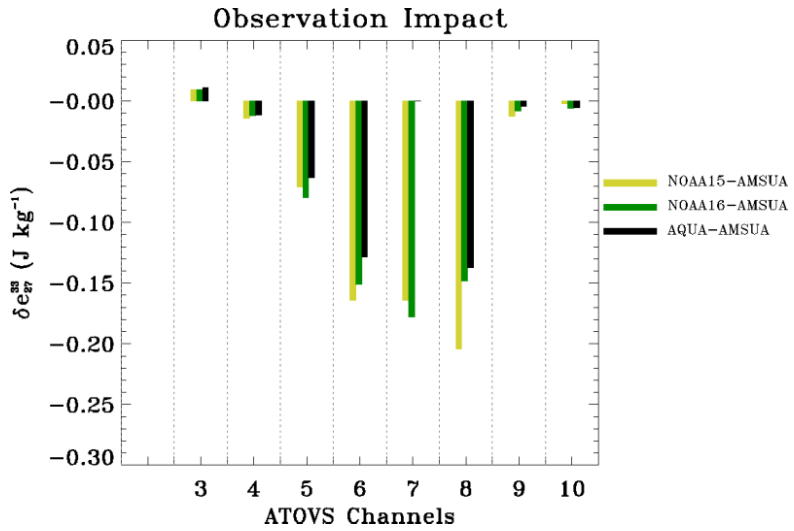

Figure 9c: AMSU-A observation impact per observation, for the EC 4D-PSAS system for August 2004. Note that the forecast impact (J kg^{-1}) is on the ordinate, and the channel index is on the abscissa. The pale green bars are for NOAA-15, while the dark green bars are for NOAA-16 and the black bars are for AQUA. Figure provided courtesy of Simon Pellerin, Stéphane Laroche, Josée Morneau, and Monique Tanguay, EC.

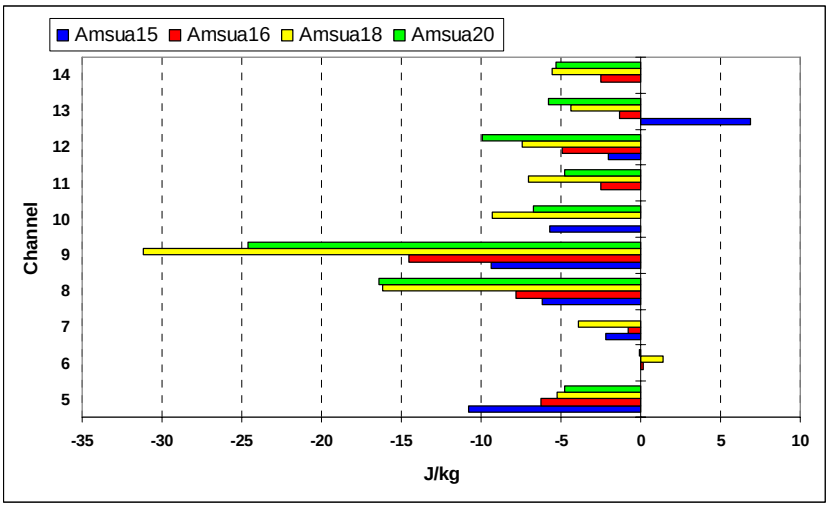

Figure 9d: As in Figure 9a, but for the ECMWF 4D-Var system, with observation impact computed using (14) for 15 June to 15 July 2006, for 00 and 12 UTC. The blue bars are for NOAA-15, the red bars are for NOAA-16, the yellow bars for NOAA-18, and the green bars are for METOP-A. Figure provided courtesy of Carla Cardinali (ECMWF).

The results are quite similar overall, with NRL, GMAO and EC showing the strongest forecast reductions for the mid-tropospheric channels (AMSU ch 5-8). The ECWMF results differ, in that they indicate stronger observation impact for the upper tropospheric channels. It is unclear whether this is due to the higher model top for ECMWF forecast system, or the observation impact calculation itself.

The most striking difference is the non-beneficial impact for AMSU channels 8 and 9 for the NRL system. These results suggest a problem with the assimilation of these channels, possibly due to the relatively low model top (effectively around 4 hPa), or insufficient model and analysis resolution near the tropopause. This example also highlights the fact that although the adjoint can tell us when there is a problem, it can not tell us how to fix that problem!

5.3. Observation Impact for the Global Observing Network

The observation impact for all observations assimilated by the NAVDAS/NOGAPS system daily at 00 UTC is shown in Fig. 10 for the 30 days ending 25 Sept. 2007. In the left panel, the observation impact is binned according to observing platform and expressed as the total impact, or reduction in forecast error, over the 30 days. In the right panel, the total impact is first scaled (by 1000) and then divided by the number of observations in each bin.

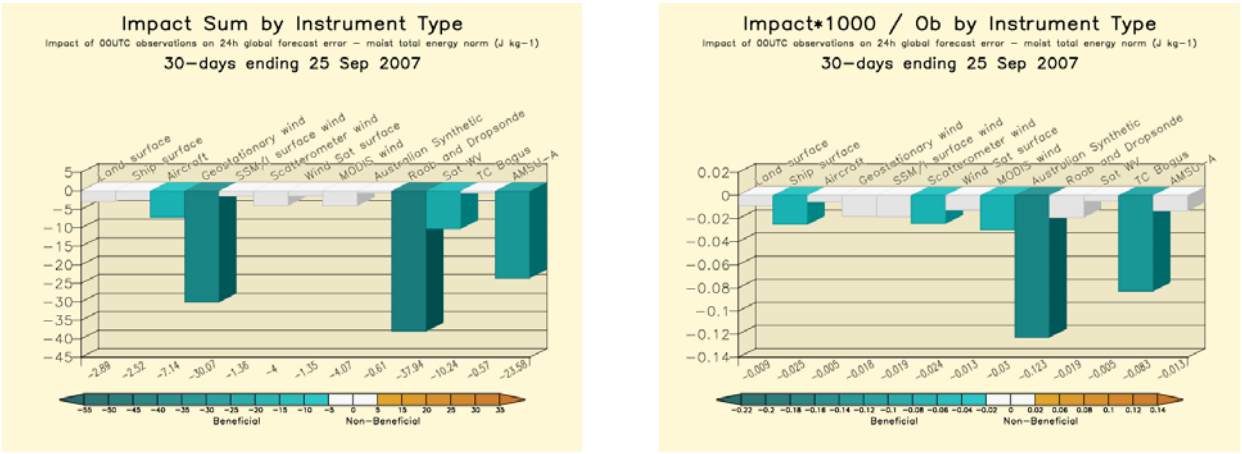

Figure 10: Total (left panel) and per ob (right panel) 30-day 00 UTC observation impact (J kg^{-1}) for the major observing platforms. Negative values (green) indicate a reduction in forecast error. White bars indicate essentially neutral impact. Orange bars (not present) would indicate an increase in forecast error.

In terms of total impact, for the NRL system, radiosondes are still the dominant source of information in the 00 UTC analysis, followed by the AMVs and AMSU-A radiances. The satellite total precipitable water observations rank as fourth. These results are analogous to those found with OSEs, although it should be noted that the two are not equivalent. The adjoint-based methods quantify the reduction in forecast error due to an observing system, in the context of all observations assimilated. The OSE approach gives the impact relative to a control run when an observing system is removed from the assimilation.

On a per observation basis, the two observing platforms with the most beneficial impact are the Australian surface pressure pseudo-observations (PAOBs) and the tropical cyclones pseudo-observations generated from the tropical cyclone warning messages. The total observation counts for each type are small, so it is possible that some of the nonlinear aspects discussed in 4.3 are relevant; however, these relative rankings have remained consistent over several years of monitoring.

5.4. Observation Impact as a Function of Observation Time

The final example (Fig. 11) is from EC, and shows the mean Southern Hemisphere observation impact per observation for the 4D-PSAS system as a function of the time of the observation within the assimilation window. A clear trend is apparent, indicating that the observations at the end of the assimilation window have the largest impact. This result is particularly relevant for satellite-data providers, as it shows that reduced data latency is of critical importance.

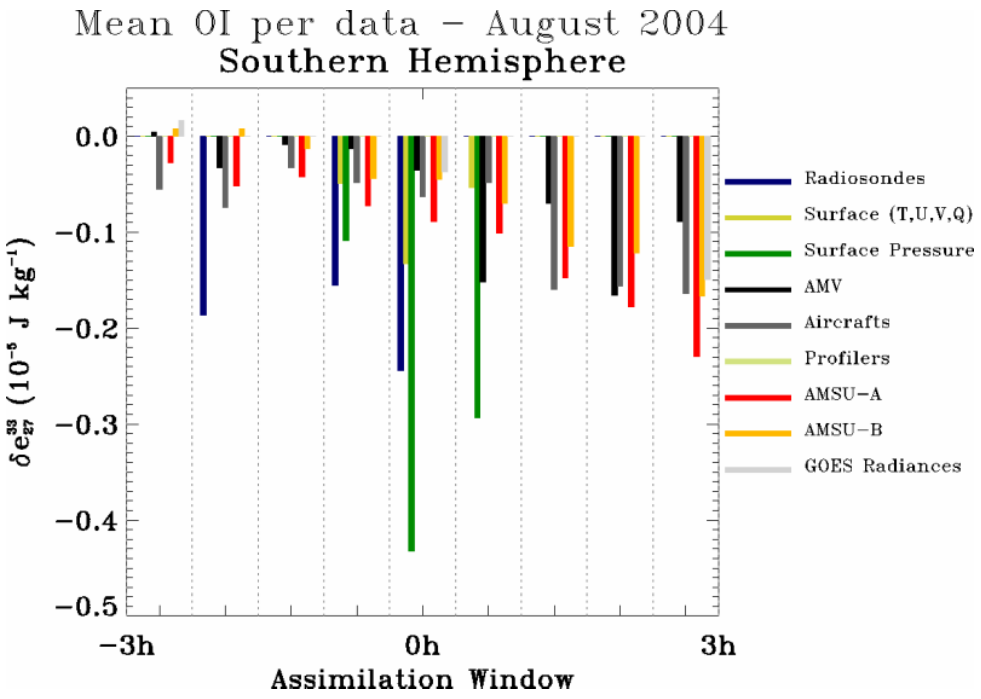

Figure 11: Mean observation impact per observation (ordinate in J kg^{-1}) for the EC 4D-PSAS system, as a function of observation time relative to the assimilation window. The observing platforms are color-coded and given in the legend. Figure provided courtesy of Simon Pellerin, Stéphane Laroche, Josée Morneau, and Monique Tanguay, EC.

6. Concluding Remarks

The adjoint of the data assimilation system has proven to be a useful tool to help identify systematic problems with the observing network, to quantify the value of different observing platforms, and as a way to compare different data assimilation systems. One must keep in mind that the observation impact, as calculated via (12) or (14) is a very specific measure of forecast impact, as it depends upon the choice of J, the suite of observations assimilated, the data assimilation system, and the NWP forecast model. Despite the similar results obtained using different NWP systems, the observation impact is, strictly speaking, valid only for the data assimilation system and NWP forecast system used to compute the statistics[7]. Ultimately, that is its greatest strength, and potentially, its greatest weakness.

The ongoing formal collaboration between NRL and GMAO, and the informal ones with EC and ECMWF will help clarify the extent that the observation impact informs us about issues with the global observing network and data assimilation systems in general.

Acknowledgements

The authors gratefully acknowledge support from the Naval Research Laboratory, under program element 062435N, project number BE-435-046, and the Joint Center for Satellite Data Assimilation. We thank our Monterey colleagues Tim Hogan (NRL), Randy Pauley (FNMOC), Ben Ruston (NRL) for their contributions. We also thank Ron Gelaro (GMAO), Carla Cardinali (ECMWF), and Simon Pellerin, Stéphane Laroche, Josée Morneau, and Monique Tanguay (EC) for their contributions to this paper.

[7] The same is true for OSEs and OSSEs (Observing System Simulation Experiments).

7. References

Baker, N.L., 2000: Observation adjoint sensitivity and the adaptive observation-targeting problem. Ph.D. dissertation, Naval Postgraduate School, 265 pp. *Available from the Naval Research Laboratory, Monterey, CA 93943.*

Baker, N.L. and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptive observation-targeting problem. Q. J. R. Meteorol. Soc., **126**, 1431-1454.

Bergot, T., 1999: Adaptive observations during FASTEX: A systematic survey of upstream flights. *Quart. J. Roy. Meteor. Soc.*, **125**, 3271-3298.

Bergot, T and A. Doerenbecher, 2002: A study of the optimization of the deployment of target observations using adjoint-based Methods. *Q.J.R. Meteorol. Soc*., **128**, 1689-1712.

Buizza, R., and A. Montani, 1999: Targeting observations using singular vectors. *J. Atmos. Sci*., **56**, 2965-2985.

Cardinali, C., 2004: Influence-matrix diagnostic of a data assimilation system. *Q. J. R. Meteor. Soc*., **130**, 2767-2786.

Daley, R., 1991: *Atmospheric Data Assimilation*, Cambridge University Press, 457 pp.

Doerenbecher, A. and T. Bergot, 2001: Sensitivity to observations applied to FASTEX cases. *Nonlinear. Proc. Geophys.*, **8**, 467-481.

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

Fourrié, N.A., A. Doerenbecher, T. Bergot, and A. Joly, 2002: Adjoint sensitivity of the forecast to TOVS observations. *Q.J.R. Meteor. Soc*., **128**, 2759-2777.

Fourrié, N.A, and F. Rabier, 2004: Cloud characteristics and channel selection for IASI radiances in meteorologically sensitive areas. *Q. J. R. Meteor. Soc.*, **130**, 1839-1856.

Gelaro, R., R.H. Langland, G.D. Rohaly, and T.E. Rosmond, 1999: An assessment of the singular vector approach to targeted observing using the FASTEX data set. *Quart. J. Roy. Meteor. Soc.*, **125**, 3299-3327.

Gelaro, R., Y. Zhu and R. Errico, 2007: Examination of various-order adjoint-based approximations of observation impact. Submitted to *Meteorologische Zeitschrift*.

Gelb, A., 1974: *Applied optimal estimation.* MIT Press, 374 pp.

Hogan, T.F., T.E. Rosmond, and R.L. Pauley, 1999: The navy operational global atmospheric prediction system: Recent changes and testing of gravity wave and cumulus parameterizations. Preprints, *13th Conf. Num. Wea. Pred.*, Denver, CO, Amer. Meteor. Soc., 60-65.

Ide, K., P. Courtier, M. Ghil, and A.C. Lorenc, 1997: Unified notation for data assimilation: operational, sequential and variational. *J. Meteor. Soc. Japan*, **75 1 B**, 181-189.

Joly, A, D. Jorgensen, M.A. Shapiro, A. Thorpe, P. Bessemoulin, K.A. Browning, J.-P. Chalon, S.A. Clough, K.A. Emanuel, L. Eymard, R. Gall, P.H. Hildebrand, R.H. Langland, Y. Lemaitre, P. Lynch, J.A. Moore, P.O.G. Persson, C. Snyder, and R. Wakimoto, 1997: The Fronts and Atlantic Storm-track Experiment (FASTEX): Scientific objectives and experimental design. *Bull. Amer. Meteor. Soc.,* **78**, 1917-1940.

Langland, R.H., and G.D. Rohaly, 1996: Adjoint-based targeting of observations for FASTEX cyclones. Preprints, *Seventh Conf. Mesoscale Processes*, Reading, United Kingdom, Amer. Meteor. Soc., 369-371.

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

Langland, R.H. and N.L. Baker, 2004b: A technical description of the NAVDAS adjoint system. NRL/MR/7530-04-8746, 62 pp. *Available from the Naval Research Laboratory, Monterey, CA, 93943-5502.*

Morneau, J., S. Pellerin, S. Laroche, and M. Tanquay, 2006: Estimation of the adjoint sensitivity gradients in observation space using the dual (PSAS) formulation of the Environment Canada operational 4D-Var. Proceeding, Second THORPEX International Science Symposium, December 4-8, 2006. pp 162-163.

Palmer, T.N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics and adaptive observations. *J. Atmos. Sci.*, **55**, 633-653.

Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121-150.

Rosmond, T.E., 1997: A technical description of the NRL Adjoint Modeling System, Tech. Rep. NRL/MR/7532/97/7230, Naval Research Laboratory, Monterey, CA 93943-5502, 55 pp.

Ruston, B., C. Blankenship, W. Campbell, R. Langland and N. Baker, 2006: Assimilation of AIRS data at NRL. *15th International TOVS Study Conference,* Maratea, IT, 4-10 Oct., 2006.

Xu, L., R. Langland, N. Baker and T. Rosmond, 2006: Development of the NRL 4D-Var data assimilation adjoint system. Operational Numerical Weather Prediction and Data Assimilation, European Geosciences Union, General Assembly 2006, Vienna, Austria, 02 – 07 April 2006.

Zhu, Y. and R. Gelaro, 2007: Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Accepted for publication in Monthly Weather Review.*