

Observation Sensitivity Calculations Using the Adjoint of the Gridpoint Statistical Interpolation (GSI) Analysis System

YANQIU ZHU* AND RONALD GELARO

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 6 October 2006, in final form 23 February 2007)

ABSTRACT

The adjoint of a data assimilation system provides an efficient way of estimating sensitivities of analysis or forecast measures with respect to observations. The NASA Global Modeling and Assimilation Office (GMAO) has developed an exact adjoint of the Gridpoint Statistical Interpolation (GSI) analysis scheme developed at the National Centers for Environmental Prediction (NCEP). The development approach is unique in that the adjoint is derived from a line-by-line tangent linear version of the GSI. Availability of the tangent linear scheme provides an explicit means of assessing not only the fidelity of the adjoint, but also the effects of nonlinear processes in the GSI itself. In this paper, the development of the tangent linear and adjoint versions of the GSI are discussed and observation sensitivity results for a near-operational version of the system are shown. Results indicate that the GSI adjoint provides accurate assessments of the sensitivities with respect to observations of wind, temperature, satellite radiances, and, to a lesser extent, moisture. Sensitivities with respect to ozone observations are quite linear for the ozone fields themselves, but highly nonlinear for other variables. The sensitivity information provided by the adjoint is used to estimate the contribution, or impact, of various observing systems on locally defined response functions based on the analyzed increments of temperature and zonal wind. It is shown, for example, that satellite radiances have the largest impact of all observing systems on the temperature increments over the eastern North Pacific, while conventional observations from rawinsondes and aircraft dominate the impact on the zonal wind increments over the continental United States. The observation impact calculations also provide an additional means of validating the observation sensitivities produced by the GSI adjoint.

1. Introduction

Modern atmospheric data assimilation systems ingest millions of observations each day to produce initial conditions for weather and climate forecasts. The vast majority of the observations are from satellites, the number and variety of which will continue to increase significantly during the next decade. Because it is unlikely that even next-generation data assimilation systems will be able to accommodate all available observations, there is an increasing need to develop intelligent strategies for data selection and utilization. Even now, the need for such strategies is made clear by the gross un-

derutilization of the current observation set. For example, operational forecast centers that assimilate Atmospheric Infrared Sounder (AIRS) data use roughly one out of every several thousand of the available measurements provided by the instrument (Goldberg et al. 2003). At the same time, these observations have increased the total number of assimilated observations significantly, while producing small to moderate gains in forecast skill (Le Marshall et al. 2006). Even if much of the data are redundant and could be compressed by several orders of magnitude, it is unlikely that current data selection strategies adequately capture the available information. To make optimal use of the increasing volume of observations, flexible and efficient tools for quantifying the “value” of observations are required.

A common method for assessing the observation value in the context of numerical weather prediction is to perform so-called observing system experiments (OSEs), in which selected subsets of observations are removed from a data assimilation system. This is a di-

* Current affiliation: Science Applications International Corporation, Beltsville, Maryland.

Corresponding author address: Dr. Ronald Gelaro, Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD 20771.
E-mail: ron.gelaro@nasa.gov

rect way to measure the value of such subsets on forecasts and assimilation products. Meaningful comparisons require an appreciable spinup period, followed by a long period over which resulting output statistics are computed. Interpretation of results in terms of analysis quality generally requires the production of short-term forecasts. OSEs are intermittently performed at operational centers (e.g., Lord et al. 2004; Kelly et al. 2004; English et al. 2004) but, because of their expense, usually involve relatively small numbers of independent experiments, each considering variations in only large subsets of observations. So, for example, it is prohibitive to investigate the impacts of all individual channels for a given satellite observing system. In addition, each variation of the observing system changes the gain, or relative weights given to the observations, with respect to the original baseline experiment.

Recently, Baker and Daley (2000) have shown that the adjoint of a data assimilation system provides an efficient way to estimate the sensitivities of an analysis or forecast measure with respect to observations. The sensitivities may be computed with respect to any or all of the observations simultaneously based on a single execution of the adjoint system. This permits arbitrary aggregation of the sensitivities, for example, by data type, channel, or location. It also allows for the estimation of the *impact* of any subset of data on an analysis or forecast measure. This approach has been used to diagnose the effectiveness of specific targeted observations (Doerenbecher and Bergot 2001; Fourrie et al. 2002), as well as to perform comprehensive assessments of observing system impacts on short-range forecast errors (Langland and Baker 2004). Other, somewhat related, methods for estimating observation sensitivity include the second-order adjoint approach proposed by Le Dimet et al. (1995), the data resolution matrix (Menke 1984), the entropy reduction method (Rabier et al. 2002), and the influence matrix diagnostic (Cardinali et al. 2004).

The key to the adjoint approach is to compute the transpose of the gain matrix that determines the weights given to the observation-minus-background residuals, either explicitly or through a sequence of available operators. For modern data assimilations, the size and complexity of these operators render this task non-trivial. There are, however, several possible approaches to producing the adjoint, the suitability of which depend on the design of the assimilation system and the acceptability of any inherent assumptions. To date, for practical reasons, most implementations have relied on modification of the existing (forward) analysis solver to produce an approximate adjoint, as opposed to the de-

velopment of a line-by-line tangent linear model from which an exact adjoint is derived.

In this study, a tangent linear model and exact adjoint of the National Centers for Environmental Prediction (NCEP) Gridpoint Statistical Interpolation (GSI) analysis scheme (Wu et al. 2002) are developed and tested in the context of the Goddard Earth Observing System atmospheric data assimilation system (GEOS DAS; Rienecker et al. 2007). The GSI is expected to become the operational analysis scheme at both NCEP and the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) in the near future. The choice to develop an exact adjoint is motivated by design aspects of the GSI algorithm, as described in later sections. The current study focuses on the development of the tangent linear and adjoint versions of the GSI, their validation, and preliminary results.

Section 2 provides a brief description of the GSI algorithm as well as overviews of the theoretical and practical aspects of developing the tangent linear and adjoint systems. In section 3 we examine the behavior of the tangent linear model, and compare this with the behavior of the full GSI in response to a range of perturbations applied to the input innovations. In section 4 we present observation sensitivity results produced by the GSI adjoint, which in turn are used to estimate the impact of various observing systems on selected measures of the analyzed increments of temperature and zonal wind. As applied here, the observation impact calculations serve primarily as a validation tool for the adjoint results. Concluding remarks and plans for future work are presented in section 5.

2. Problem formulation

a. The GSI algorithm

The GSI is a three-dimensional variational data assimilation (3DVAR) analysis scheme based on NCEP's current operational Spectral Statistical Interpolation (SSI) system (Parrish and Derber 1992), but with the spectral definition of the background error covariance operator replaced by a gridpoint version based on recursive filters. The current implementation of GSI incorporates a set of recursive filters that produce approximately Gaussian smoothing kernels and isotropic correlation functions (Wu et al. 2002). However, by superpositioning Gaussian kernels with different length scales, it is possible to generate a large class of flow-dependent inhomogeneous background error covariance models as described by Purser et al. (2003a,b).

The analysis is obtained by minimizing the scalar cost function:

$$\begin{aligned}
 J = & \frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) \\
 & + \frac{1}{2} [h(\mathbf{x}) - \mathbf{y}]^T \mathbf{R}^{-1} [h(\mathbf{x}) - \mathbf{y}] + J_{q1} + J_{q2} \quad (1)
 \end{aligned}$$

with respect to the control vector $\mathbf{x}(\psi, \chi, T_v, q, oz, \ln p_s, T_s)$, where ψ is the streamfunction, χ is the unbalanced velocity potential, T_v is the unbalanced virtual temperature, q is the (scaled) specific humidity, oz is the ozone mixing ratio, $\ln p_s$ is the logarithm of surface pressure, and T_s is the surface skin temperature. The vector \mathbf{x}^b represents the background or prior estimate of \mathbf{x} , and \mathbf{B} is its expected error covariance. The vector \mathbf{y} contains the available observations, the operator $h(\mathbf{x})$ simulates these observations from \mathbf{x} , and \mathbf{R} is the expected covariance of the instrument plus representativeness errors associated with the observations. The superscript T denotes the transpose operation.

The terms J_{q1} and J_{q2} are penalties for negative humidity and supersaturation, respectively, defined as

$$J_{q1} = \begin{cases} 0 & \text{if } q \geq 0 \\ \lambda_1 q^2 & \text{if } q < 0 \end{cases} \quad \text{and} \quad (2)$$

$$J_{q2} = \begin{cases} 0 & \text{if } q \leq q_s \\ \lambda_2 (q - q_s)^2 & \text{if } q > q_s \end{cases}, \quad (3)$$

where q_s is the saturation value of q , and λ_1 and λ_2 are parameters. To simplify the presentation, we omit these additional penalty terms from the development that follows. These terms are however included in the actual tangent linear and adjoint versions of the GSI developed for this study, and their impact on the sensitivity calculations is examined in section 3.

Because $h(\mathbf{x})$ is generally nonlinear, the most efficient means for minimizing J is through an incremental approach (Courtier et al. 1994) in which the problem is repeatedly linearized about an updated reference solution (the *outer loop*). A gradient-based iterative algorithm (the *inner loop*) is then used to minimize the resulting cost function:

$$\begin{aligned}
 J_k = & \frac{1}{2} [\delta \mathbf{x}_k - (\mathbf{x}^b - \mathbf{x}_k)]^T \mathbf{B}^{-1} [\delta \mathbf{x}_k - (\mathbf{x}^b - \mathbf{x}_k)] \\
 & + \frac{1}{2} (\mathbf{H}_k \delta \mathbf{x}_k - \mathbf{d}_k)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \delta \mathbf{x}_k - \mathbf{d}_k), \quad (4)
 \end{aligned}$$

where $k = 0, \dots, K$ is the outer loop index. The variables

$$\mathbf{d}_k = \mathbf{y}_k - h_k(\mathbf{x}_k), \quad (5)$$

and

$$\delta \mathbf{x}_k = \mathbf{x}_{k+1} - \mathbf{x}_k, \quad (6)$$

are the residual (or innovation) vector and increment, respectively. The matrix \mathbf{H}_k is the Jacobian of \mathbf{h} linearized about \mathbf{x}_k . In practice, a preconditioned conjugate gradient descent algorithm with ~ 100 inner iterations and two outer loops is found to produce satisfactorily converged increments in most cases. The second outer loop accounts for changes in quality control (especially for radiance data) and weak nonlinearities in some observation operators (e.g., surface wind speed) but, generally speaking, produces relatively small changes to the analysis increments. In this paper, we focus on observation sensitivities based on tangent linear and adjoint versions of the GSI with a single outer loop iteration. Multiple outer loops will be addressed in a sequel.

b. Observation sensitivity

Setting $\partial J_k / \partial \delta \mathbf{x}_k = 0$ in (4) and neglecting all but the first outer loop, we obtain the analytical form of the analysis increment:

$$\delta \mathbf{x}_0 = (\mathbf{B}^{-1} + \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{H}_0)^{-1} \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{d}_0, \quad (7)$$

where $\mathbf{x}_0 = \mathbf{x}^b$. Substituting (5) and (6) into (7) and linearizing about \mathbf{x}_0 , we obtain the tangent linear analog of the analysis increment:

$$\delta \tilde{\mathbf{x}} = \mathbf{K}(\tilde{\mathbf{y}} - \mathbf{H}\tilde{\mathbf{x}}^b), \quad (8)$$

where \mathbf{K} is the gain matrix:

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}. \quad (9)$$

Here, the tildes denote tangent linear variables and the subscript for the outer loop index has been dropped for convenience. Following Baker and Daley (2000), we define the sensitivity of the analysis increment with respect to the observations as

$$\frac{\partial \delta \tilde{\mathbf{x}}}{\partial \mathbf{y}} = \mathbf{K}^T, \quad (10)$$

where \mathbf{K}^T is referred to as the analysis adjoint. By application of the chain rule, the sensitivity of any scalar aspect J , not to be confused with J in (1), of either the analysis or forecast¹ with respect to the observations is given by

$$\frac{\partial J}{\partial \mathbf{y}} = \mathbf{K}^T \frac{\partial J}{\partial \mathbf{x}}. \quad (11)$$

Note that \mathbf{K}^T maps a vector in physical space to a vector in observation space, while the mapping by \mathbf{K} is in the opposite sense. Moreover, (11) indicates that, for a

¹ If J is based on a model forecast, then the calculation of $\partial J / \partial \mathbf{x}$ in (11) will generally require the model adjoint. For the purposes of this discussion we need only assume that $\partial J / \partial \mathbf{x}$ exists.

given J , the sensitivity can be computed with respect to any or all observations simultaneously with a single execution of the adjoint system. This permits arbitrary aggregation of the results (e.g., by data type, location, channel, etc.). From (9), we see that the sensitivity depends on the characteristics of the assimilation system and on attributes of the observations such as their locations and assumed errors, but not on the observed values themselves. In contrast, the “impact” or expected change in J produced by assimilating the observations will depend on both the sensitivities and observed values. We examine measures of the observation impact in section 4.

c. Development of the adjoint system

For modern data assimilation systems that include complex observation operators and large numbers of observations, estimation of \mathbf{K}^T is nontrivial. There is no “best” approach, but rather several possible approaches, the suitability of which depend on the design of the system (e.g., is it formulated in physical or observation space?) and careful consideration of the underlying approximations and assumptions.

For example, in the observation space-based data assimilation system developed at the Naval Research Laboratory (NAVDAS; Daley and Barker 2001), the analysis increment is obtained by solving an alternative form of (7) given by

$$\delta\mathbf{x} = \mathbf{B}\mathbf{H}^T\mathbf{z}, \quad (12)$$

where

$$\mathbf{z} = (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}. \quad (13)$$

An iterative algorithm is used to obtain the vector \mathbf{z} with \mathbf{d} as input, where the gain matrix is $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$. Because $(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ is self-adjoint and the operators \mathbf{H} and \mathbf{B} are explicitly available in this formulation, the sensitivity with respect to observations can be calculated using nearly the same algorithm, but with the columns of $\mathbf{H}\mathbf{B}$ (or the vector $\mathbf{H}\mathbf{B}\partial J/\partial\mathbf{x}$) replacing \mathbf{d} as input. The obvious appeal of this approach is that it requires only minor modification of the existing forward analysis code. A caveat is that the modified algorithm solves a different minimization problem than that used to obtain the analysis increment. Thus, care should be taken when choosing the convergence or stopping criterion since, strictly speaking, the correct adjoint is obtained only when the solutions are completely converged. This caveat notwithstanding, this approach has been used effectively at NRL (e.g., Langland and Baker 2004).

In the GSI, (4) is minimized directly, so there is no analog of (13) which can be easily manipulated to ob-

tain $\mathbf{K}^T = \mathbf{R}^{-1}\mathbf{H}(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$. While it is possible, in principle, to reformulate the cost function to obtain an adjoint, this would essentially amount to building another data assimilation system in observation space having little in common with the GSI. In particular, the preconditioning and handling of quality control operations and observation error assignment performed in the inner and outer loops of the GSI would differ substantially in the reformulated system. It is unclear how inconsistencies in these operations with respect to their counterparts in the forward algorithm would affect the accuracy and interpretability of the adjoint results.

For these and other reasons it was decided to develop an exact adjoint of the GSI, that is, based on an exact transpose of a line-by-line tangent linear model (TLM) of the forward minimization algorithm. The approach is analogous to that generally used to derive the adjoint of a numerical forecast model. In this case, the iterations of the inner loop, including the intermediate analysis increments and the sequence of conjugate gradient directions, serve as the trajectory for the TLM and adjoint systems. The correctness of the adjoint is evaluated using the equality $\langle\mathbf{x}, \mathbf{G}\mathbf{y}\rangle = \langle\mathbf{G}^T\mathbf{x}, \mathbf{y}\rangle$, where \mathbf{G} is the TLM of the GSI; \mathbf{G}^T is the adjoint; \mathbf{x} and \mathbf{y} are perturbation vectors of the analysis increments and observations, respectively; and \langle, \rangle denotes the Euclidean inner product. Availability of the TLM provides a direct means of assessing not only the mathematical correctness of the adjoint, but also its usefulness for describing the behavior of the GSI with respect to a wide range of perturbations. This is examined in detail in sections 3 and 4. Another benefit of this approach is that it can be extended in the future (e.g., to examine sensitivities with respect to observation error variances and other parameters in the GSI).

Development of an exact adjoint generally requires a significant initial development effort. In particular, the development of the tangent linear and adjoint versions of the conjugate gradient descent algorithm requires careful consideration of the nonlinear procedure for generating the sequence of conjugate directions and step sizes used in the minimization. Other sources of nonlinearity in the inner loop include the observation operator for wind speed and precipitation rate (although we do not assimilate precipitation observations in the current study). Also, as we show in section 3, the moisture penalty terms J_{q1} and J_{q2} act like switches that may introduce strong nonlinearity in the GSI.

3. Tangent linear experiments

The first step in evaluating the usefulness of the GSI adjoint is to determine whether the TLM accurately

describes the behavior of the GSI in response to meaningful perturbations. The accuracy of the TLM may be assessed by comparing the TLM responses, $\delta\tilde{\mathbf{x}}$, corresponding to a given set of perturbed innovations with the differences between the increments, $\Delta\mathbf{x}$, produced by the GSI with and without these perturbations. The primary measures of agreement used here are the ratio of the root-mean-squared (RMS) values of $\delta\tilde{\mathbf{x}}$ and $\Delta\mathbf{x}$, and the correlation between $\delta\tilde{\mathbf{x}}$ and $\Delta\mathbf{x}$. If the GSI behaves approximately linearly, then both the amplitude and structure of the TLM response and perturbed GSI differences should agree well (i.e., both the ratio and correlation measures should be close to 1). Exact agreement is not expected owing to the various sources of nonlinearity in the GSI described in section 2.

The perturbed innovations have the form $\mathbf{d}' = (1 + \alpha)\mathbf{d}$, where the parameter α determines the perturbation amplitude. Because different analysis variables may exhibit valid tangent linear behavior for different perturbation amplitudes and perturbed innovation types, the choice of α requires careful consideration. In this study, we examined results for values of α ranging from 10^{-6} to 1, applied to innovations of individual observation types as well as all observation types collectively. We focus most of the discussion in this section on results for $\alpha = 0.1$ because these are sufficiently representative of the results obtained for the range of values tested. Statistics produced by the GEOS DAS (not shown) indicate that while global RMS values of the innovations for most observation types vary by roughly 10%–50% from day to day, individual innovations can exhibit much larger variations. The response to actual innovations is examined in the context of the observation impact results presented in section 4.

Results are presented for August 2004 based on analyses produced at 0000 UTC. For practical reasons, we use a relatively low horizontal resolution version of GSI corresponding to 1.875° in latitude and longitude, with 64 vertical levels defined on σ surfaces. The background forecast is provided by the *GEOS-5* model (Rienecker et al. 2007). Analyses are produced using a 6-h assimilation cycle that includes all conventional observations and satellite radiances assimilated operationally during the study period. Conventional observations are assimilated from rawinsondes, aircraft, surface ships, and land stations. Additional conventional data types include winds from geostationary satellites, profilers, and scatterometers, wind speeds from the Special Sensor Microwave Imager (SSM/I), and ozone observations from the Solar Backscatter Ultraviolet (SBUV/2) instrument. Satellite radiances are assimilated from the Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) Ad-

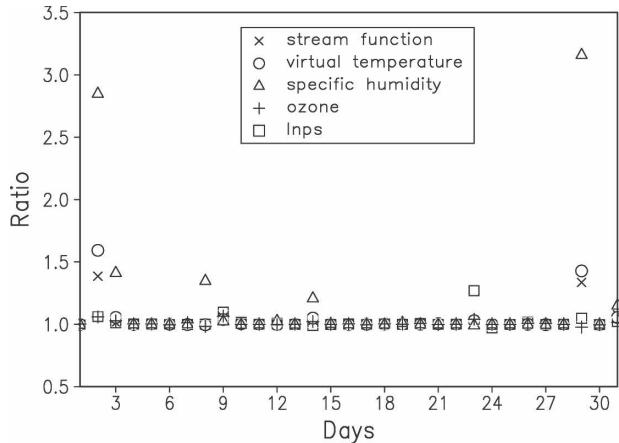


FIG. 1. The ratio between the RMS TLM response and GSI difference at analysis level 25 during August 2004 when all of the observation innovations are perturbed by 10%.

vanced Microwave Sounding Unit (AMSU-A and AMSU-B), the High-Resolution Infrared Radiation Sounder (HIRS-2 and HIRS-3), the Geostationary Operational Environmental Satellites (GOES), and the Microwave Sounding Unit (MSU). In accordance with the development in section 2, both the GSI and TLM are run with a single outer loop.

Figure 1 shows daily time series of the RMS ratios of the TLM response and perturbed GSI differences for several variables at analysis level 25 (approximately 500 hPa) for experiments in which the innovations for all observation types have been perturbed by 10%. The results are representative of those at other vertical levels. The ratios are close to 1 for all variables on most days except 2 and 29 August. On these days, the specific humidity response of the TLM is roughly 3 times larger than the perturbed GSI differences, although the ratios for the remaining variables are much closer to 1. The disagreement in the humidity responses for these cases is examined in more detail below. Ratios slightly larger than 1 also occur on a few other days for some variables, but the responses are generally in good agreement.

An example of the response to perturbed innovations of an individual observation type is shown in Fig. 2 for the case of 6 August. This example is representative of the good agreement between the TLM and perturbed GSI differences observed in the vast majority of cases. The figure shows the TLM response (Fig. 2a) and GSI differences (Fig. 2b) in terms of zonal wind at level 30 (approximately 300 hPa) when all satellite radiances are perturbed by 10%. In this example it can be seen that the perturbed radiances have a much larger impact in the Southern Hemisphere, reflecting the greater in-

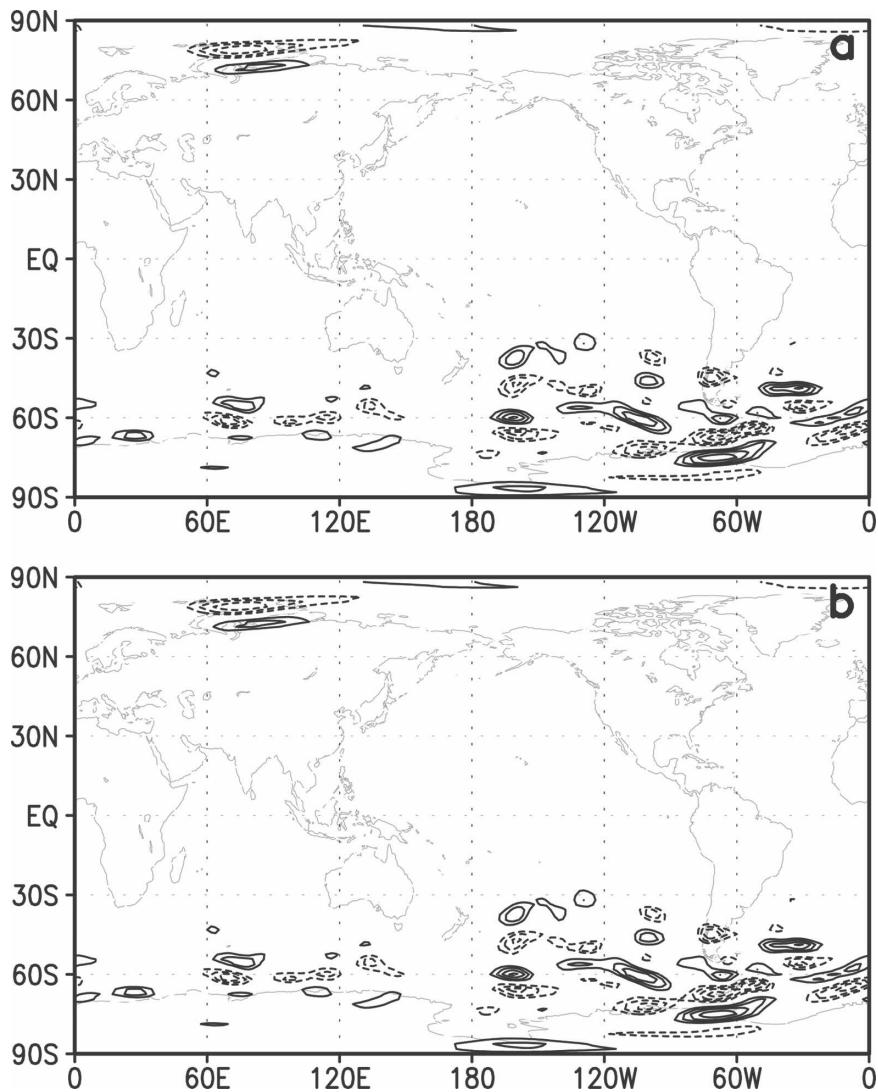


FIG. 2. The zonal wind at analysis level 30 at 0000 UTC 6 Aug for the (a) TLM response and (b) GSI difference when all radiance innovations are perturbed by 10%. The contour interval is 0.08 m s^{-1} . Negative contours are dashed and the zero contour is omitted.

fluence of satellite data where conventional observations are more sparse.

The moisture penalty terms in (2) and (3), although strongly nonlinear, are continuous and differentiable in nature, with continuous first derivatives. In most cases, these terms make adjustments that are therefore well modeled by the TLM. However, in 2 of the 31 cases examined, these terms exhibited strong nonlinearity to a degree that degraded the TLM solution noticeably. Figure 3a shows the TLM response (contours) and perturbed GSI differences (shaded) in terms of specific humidity at approximately 500 hPa for the poorly modeled case of 29 August noted earlier. The GSI differences are distributed across all longitudes with maxi-

imum variance in the tropics and subtropics. The TLM response bears no likeness to the GSI differences, with two isolated extrema over the Gulf of Alaska and the equatorial Atlantic Ocean.

Figure 3b is similar to Fig. 3a, except that the moisture penalty terms have been excluded from both the TLM and GSI for this single analysis cycle. In this case, the TLM response and perturbed GSI differences agree well. Moreover, it can be seen that the GSI differences with and without the penalty terms are very similar, indicating that the penalty terms have not changed the GSI increments themselves significantly during one analysis cycle. Direct comparison of the GSI increments with and without the moisture penalty terms for

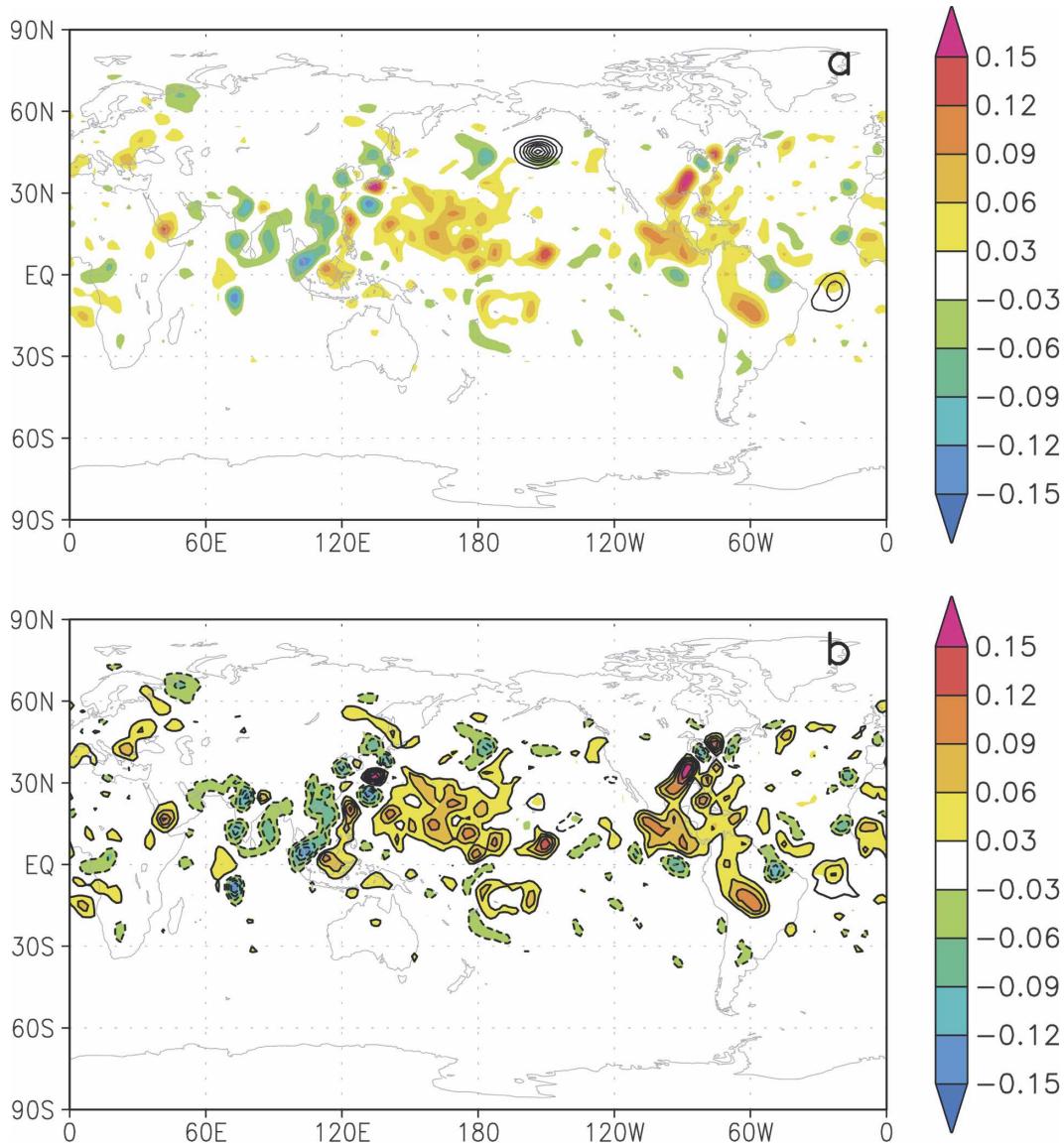


FIG. 3. The specific humidity at analysis level 25 at 0000 UTC 29 Aug for the TLM response (contours) and GSI difference (shaded) when all innovations are perturbed by 10%. The moisture penalty terms in the TLM and GSI are (a) included and (b) excluded. The values have been multiplied by 10^3 , with negative contours dashed and the zero contour omitted.

this analysis cycle (not shown) reveals that their differences are approximately an order of magnitude smaller than the increments themselves.² Because the TLM without the penalty terms is capable of representing the behavior of the GSI without the penalty terms, and because the results from the GSI with and without the penalty terms for a single analysis are quite similar, the

² If the GSI is run with two outer loops, the differences can become somewhat larger, but are still significantly smaller than the values of the increments in most locations.

TLM (or adjoint) can be run without the penalty terms when necessary and still produce a reasonable estimate of the sensitivity. In such cases, it is not necessary to remove the penalty terms from the GSI analysis itself. In more recent versions of the GSI than the one used in this study, the control variable for humidity has been redefined in a way that reduces the need for these penalty terms substantially (J. C. Derber 2006, personal communication).

The average correlations between the TLM responses and GSI differences at approximately 500 hPa

TABLE 1. Average correlations between the TLM responses and GSI differences at analysis level 25 for the month of August 2004 when the innovations have been perturbed by 10%. See text for details.

Increment variable	Perturbed innovation type								
	rad	T_v	u	v	q	oz	p_s	spd	all
T_v	0.967	0.991	0.931	0.897	0.973	0.787	0.914	0.881	0.971
u	0.966	0.973	0.998	0.970	0.947	0.781	0.984	0.965	0.989
v	0.969	0.964	0.995	0.984	0.943	0.799	0.983	0.955	0.987
ψ	0.958	0.972	0.993	0.972	0.944	0.784	0.985	0.975	0.982
χ	0.877	0.946	0.999	0.992	0.944	0.764	0.959	0.896	0.994
q	0.944	0.877	0.680	0.720	0.999	0.630	0.749	0.726	0.925
oz	0.897	0.816	0.458	0.478	0.935	0.999	0.762	0.505	0.998
$\ln p_s$	0.890	0.890	0.989	0.969	0.910	0.728	0.999	0.986	0.986
T_s	0.996	0.996	0.978	0.947	0.998	0.909	0.981	0.972	0.994

for the month of August 2004 are summarized in Table 1. The top row lists the perturbed innovation types, which correspond to satellite radiances (rad), followed by conventional observations of virtual temperature (T_v), zonal wind (u), meridional wind (v), specific humidity (q), ozone mixing ratio (oz), surface pressure (p_s), and near-surface wind speed (spd), plus all observation types collectively (all). The leftmost column lists the analysis increment variables. The numerical values show the average correlations between the TLM response and GSI differences for each analysis increment variable on the far left in response to the perturbed innovation type at the top. Except for results in the rightmost column (all), only the specified innovation type has been perturbed for each set of experiments, while the remaining innovation types are left unperturbed.

Overall, the results show that the GSI behavior is well represented by the TLM, consistent with the RMS ratios in Fig. 1. The results shown here are representative of those at other vertical levels. The vast majority of the correlations are close to or greater than 0.9, with some notable exceptions involving the responses of some variables to perturbed innovations of ozone and, to a lesser extent, specific humidity. For example, when the ozone innovations are perturbed, the average response correlation is greater than 0.9 for ozone itself, but substantially lower for most other variables. This occurs because, while the inner-loop problem in the GSI is linear (except for certain observation operators and the moisture penalty terms), the minimization algorithm used to solve this problem is nonlinear. Ozone and moisture are not strongly coupled to other variables through the background error covariance operator, \mathbf{B} , and so cross responses involving these variables may be significantly affected by the nonlinear nature of the minimization algorithm. In contrast, cross responses between wind, temperature, and pressure are strongly coupled through \mathbf{B} and therefore not significantly af-

ected by nonlinearity. The direct responses of all variables to perturbations of the same observation variable tend to be strong and linear as a result of the diagonal elements of \mathbf{B} . Because cross responses tend to be weaker than direct responses in general, the correlations for all variables are high when all the innovations are perturbed simultaneously (rightmost column).

4. Adjoint experiments

Having demonstrated the ability of the TLM to represent the general behavior of the GSI, we now examine results produced by the adjoint. In contrast with the TLM, which takes a perturbation $\alpha \mathbf{d}$ in observation space as input and produces a perturbation $\delta \mathbf{x}$ in analysis space as output, the adjoint takes a gradient $\partial J / \partial \mathbf{x}$ of a response function J in analysis space as input and produces a gradient (or sensitivity) $\partial J / \partial \mathbf{y}$ in observation space as output. The response function J can be any differentiable scalar measure of interest defined globally or for a particular region of interest.

In this study, we examine results for four response functions of the following form:

$$J = \frac{1}{2} \langle \delta \mathbf{x}, \mathbf{S} \delta \mathbf{x} \rangle, \quad (14)$$

where \mathbf{S} is a projection operator that selects only a subset of the total analysis increment $\delta \mathbf{x}$. The first pair of measures, denoted $J_{T_{NP}}$ and $J_{U_{NP}}$, include only increments of either temperature or zonal wind, respectively, at all vertical levels in a $23^\circ \times 60^\circ$ box centered over the eastern North Pacific. The second pair, denoted $J_{T_{US}}$ and $J_{U_{US}}$, are similar to the first, but for a box centered over the continental United States (see Figs. 4 and 6). The measures, while not necessarily of meteorological significance, allow us to examine sensitivities of the wind and temperature increments in regions where the mixture of satellite and conventional observations differs markedly. The quadratic forms of these

measures prevent cancelation due to increments of opposite sign and have the added convenience that their gradients with respect to the analysis, required as input to the adjoint, are equivalent to the increments themselves where the response functions are defined.

a. Observation sensitivity results

Observation sensitivities were computed once each day for each response function based on the 0000 UTC analyses for the month of August 2004. The observing system and resolution of the GSI adjoint are the same as those used for the TLM experiments in section 3. Figures 4–6 show examples of $\partial J_{T_{NP}}/\partial \mathbf{y}$ and $\partial J_{U_{US}}/\partial \mathbf{y}$ for selected observing systems on 5 August. The results for this case are representative of those throughout the study period. They are presented here to highlight basic characteristics of the sensitivities including their dependence on the type and location of the observations, and on the density of surrounding observations. In fact, the sensitivities depend in a complex way on all aspects of the gain matrix \mathbf{K} in (9). The reader is referred to Baker (2000) and Baker and Daley (2000) for more detailed descriptions of these dependencies.

Figure 4a shows the sensitivity of $J_{T_{NP}}$ to rawinsonde temperature observations at 500 hPa. Note that the sensitivities have units of the gradient $\partial J_{T_{NP}}/\partial \mathbf{y}$ that, for temperature observations, are $\text{K}^2/\text{K} \sim \text{K}$. To highlight observations with the greatest sensitivity, values within one contour interval of 0 are shaded gray. The box over the eastern North Pacific outlines the area where $J_{T_{NP}}$ is defined; for convenience we refer to this as the target area in the discussion that follows. There are no rawinsonde observations located within the target area in Fig. 4a. However, $J_{T_{NP}}$ is sensitive to several rawinsonde temperature observations to the northeast of the target area, along the west coast of North America. In particular, the sensitivity with respect to the observation over Vancouver Island is close to 6.5 K, implying that, to first-order accuracy, a 1-K increase (decrease) in the temperature of this observation would increase (decrease) $J_{T_{NP}}$ by approximately 6.5 K^2 . Observations along the U.S. west coast, while equally close to the target area, have much smaller sensitivity values. This is most likely due to the greater density of observations over the western United States compared with western Canada, which tends to reduce the sensitivity to individual observations. Because of the global nature of the 3DVAR solution, observations far removed from the target area exhibit small, but nonzero, sensitivity values.

Figure 4b shows the sensitivity of $J_{T_{NP}}$ to channel-5 brightness temperatures on *NOAA-I6* AMSU-A.

These observations provide temperature information in a deep vertical layer that peaks in the midtroposphere. In this case, observations from three orbits of the satellite lie within the target area, producing the largest sensitivity values of all observation types examined for this case. The sensitivities exhibit a complex pattern of positive and negative values, reflecting small-scale features in the increment field at various vertical levels. A more detailed view of this dense pattern in the vicinity of the target area is shown in Fig. 5, along with the squared analysis increments of temperature at approximately 550 hPa used to compute the response function $J_{T_{NP}}$. This level is close to where channel 5 on AMSU-A has its peak response. There is a clear correspondence between the extrema in the increments and large values of observation sensitivity. An exact correspondence is not expected because the sensitivities also depend on the increments at other vertical levels, as well as on various aspects of \mathbf{K} as noted earlier.

Figure 4c shows the sensitivity of $J_{T_{NP}}$ to GOES infrared cloud drift observations of zonal wind at 500 hPa. There are several observations within or near the target area that exhibit the largest sensitivity values for this observation type. These values are one to two orders of magnitude smaller than the maximum sensitivity values for the AMSU-A radiances or rawinsonde temperatures in Figs. 4a,b (although, strictly speaking, it is difficult to make a quantitative comparison between sensitivities with respect to wind and temperature). Wind observations at this level affect the analyzed temperatures above and below this level indirectly through balance conditions imposed by the background error covariance. The results show that the sensitivity of $J_{T_{NP}}$ to these wind observations is relatively small.

Moving to results for $\partial J_{U_{US}}/\partial \mathbf{y}$, we show in Fig. 6a the sensitivity with respect to rawinsonde zonal wind observations at 500 hPa. The largest sensitivity values occur along the edges of the target area where the observation density is lowest. While the results are not strictly comparable to those in Fig. 4a, the magnitudes of the responses in these figures suggest that temperature and zonal wind increments are each similarly sensitive to rawinsonde observations of the same variables. Conversely, comparing the sensitivities with respect to (the same) cloud drift wind observations in Figs. 6c and 4c, and comparing their magnitudes with those of the other observing systems in each figure, it is clear that cross sensitivities may be considerably weaker. However, the relatively large sensitivity of $J_{U_{US}}$ with respect to channel-5 AMSU-A brightness temperatures shown in Fig. 6b indicates that cross sensitivities are not always small.

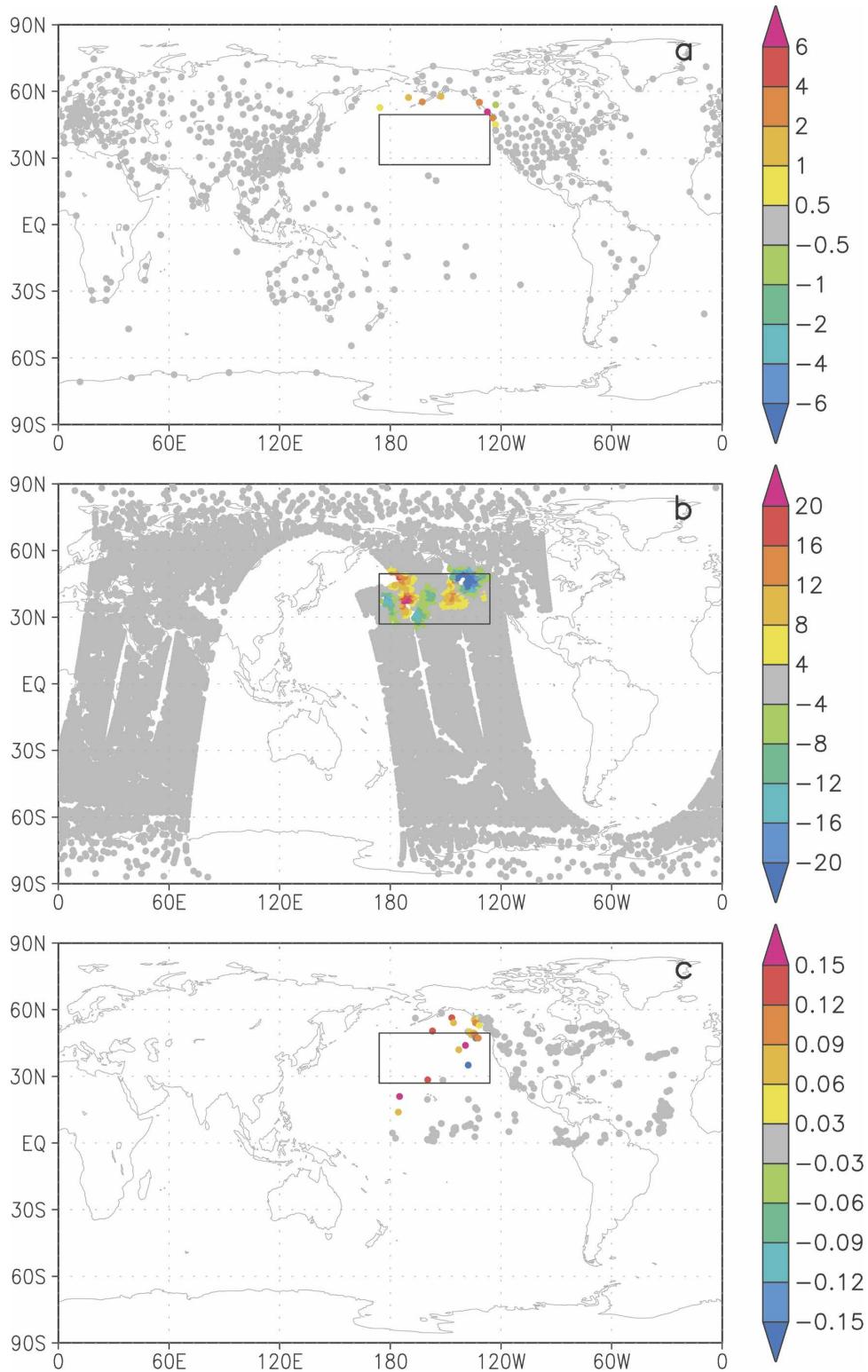


FIG. 4. The sensitivity of J_{TNP} at 0000 UTC 5 Aug with respect to (a) rawinsonde temperature (K) observations at 500 hPa, (b) channel-5 brightness temperatures (K) on NOAA-16 AMSU-A, and (c) GOES IR cloud drift zonal wind ($K^2 m^{-1} s$) observations at 500 hPa. The box outlines the area in which J_{TNP} is defined.

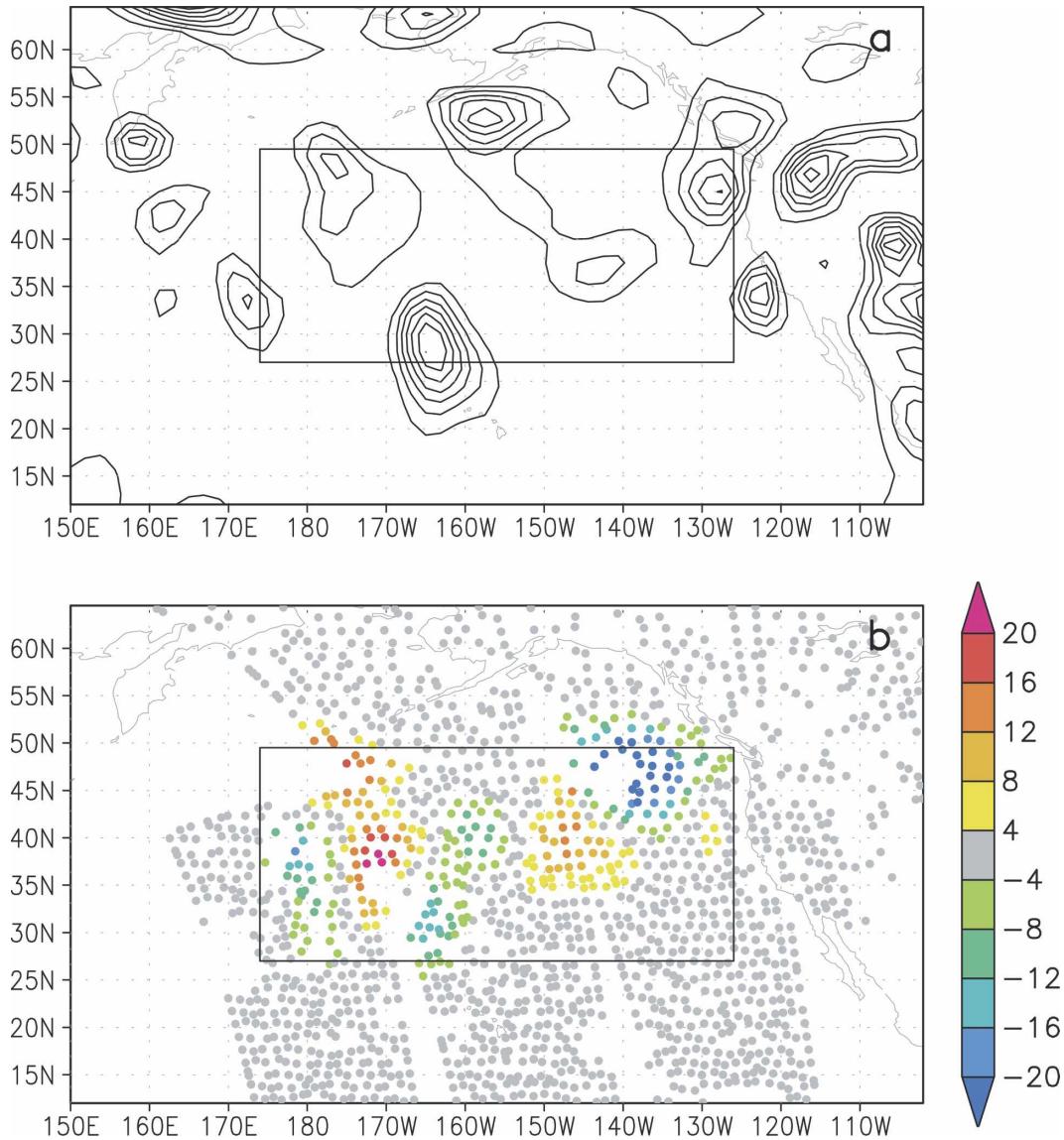


FIG. 5. Limited-area view of (a) the squared analysis increments of temperature at approximately 550 hPa (contour interval is 0.2 K^2) and (b) the sensitivity of $J_{T_{NP}}$ with respect to channel-5 brightness temperatures (K) on NOAA-16 AMSU-A at 0000 UTC 5 Aug. The box outlines the area in which $J_{T_{NP}}$ is defined.

b. Observation impact

The sensitivity information produced by the GSI adjoint can be used effectively to estimate the impact of observations on the response function J . As applied here, this provides not only a powerful diagnostic tool for data assimilation, but also a means of verifying the accuracy of the observation sensitivities themselves.

From (7) and (9), we can express the analysis increment as

$$\delta \mathbf{x} = \mathbf{Kd}. \tag{15}$$

Combining (15) and (11), and using the definition of an adjoint, we obtain

$$\langle \partial J / \partial \mathbf{x}, \delta \mathbf{x} \rangle = \langle \partial J / \partial \mathbf{y}, \mathbf{d} \rangle. \tag{16}$$

For quadratic measures of the form (14), we have $\partial J / \partial \mathbf{x} = \mathbf{S} \delta \mathbf{x}$ that, when substituted into (16), yields

$$J = \frac{1}{2} \langle \partial J / \partial \mathbf{y}, \mathbf{d} \rangle. \tag{17}$$

Equation (17) is an estimate of J computed in observation space, based on the inner product between the observation sensitivities and the innovations. The impact of any or all observations on J is therefore easily computed by summing only the elements cor-

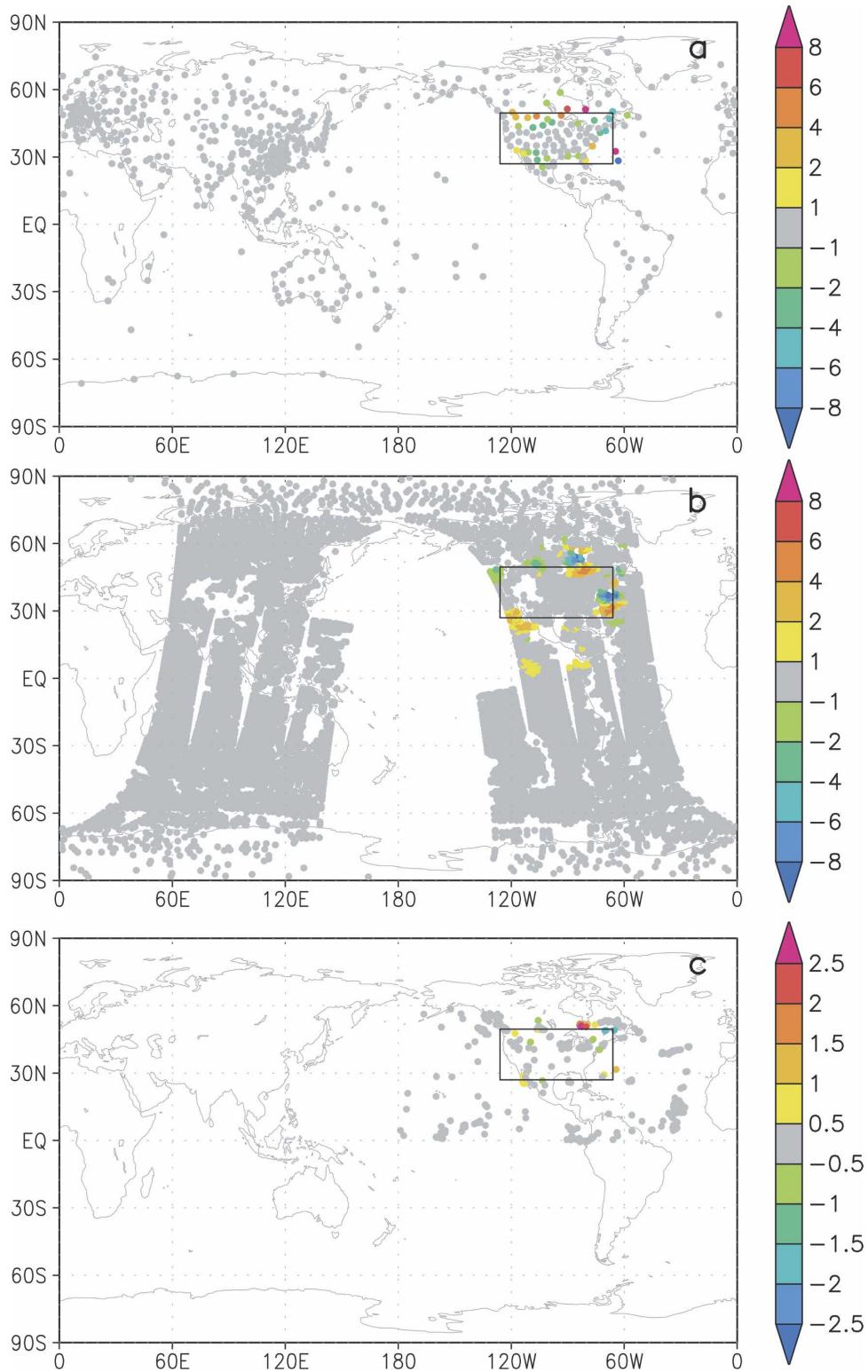


FIG. 6. The sensitivity of $J_{U_{US}}$ at 0000 UTC 5 Aug with respect to (a) rawinsonde zonal wind (m s^{-1}) observations at 500 hPa, (b) channel-5 brightness temperatures ($\text{K}^{-1} \text{m}^2 \text{s}^{-2}$) on NOAA-15 AMSU-A, and (c) GOES IR cloud drift zonal wind (m s^{-1}) observations at 500 hPa. The box outlines the area in which $J_{U_{US}}$ is defined.

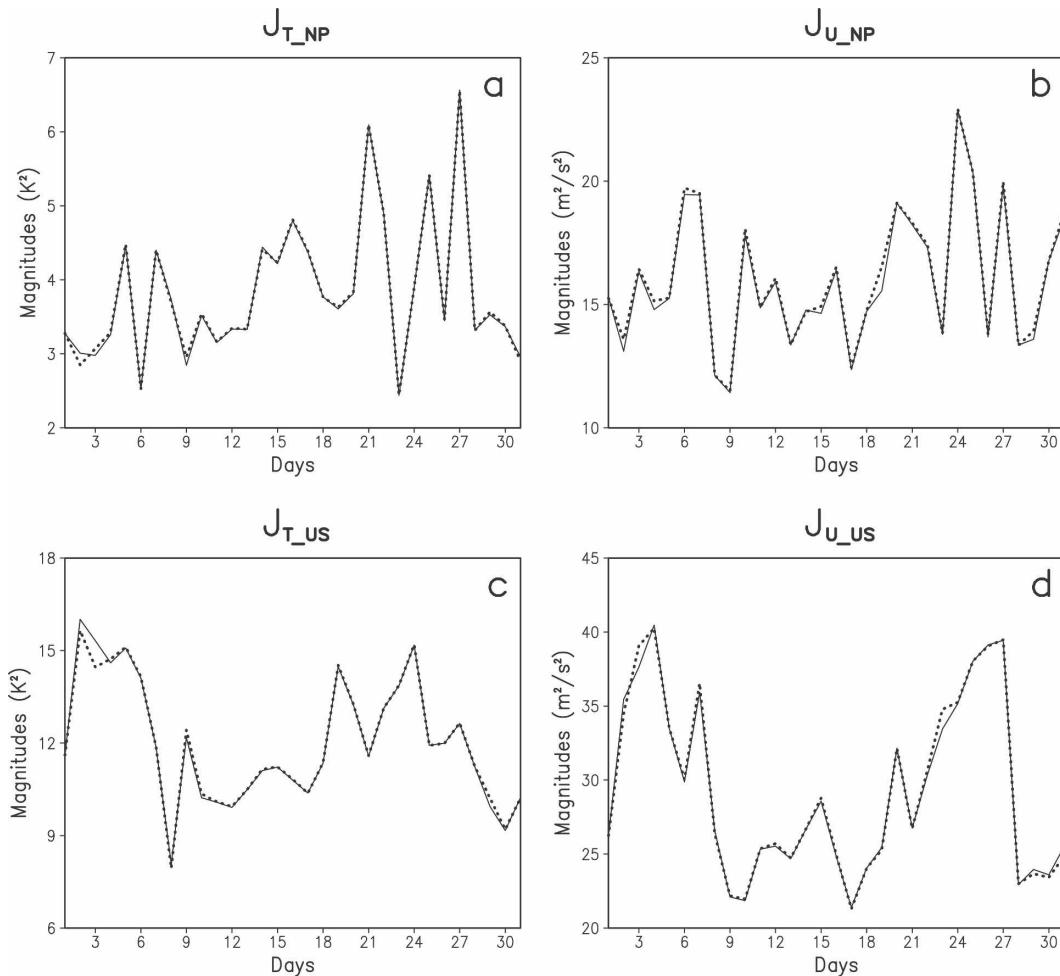


FIG. 7. Time series of (a) J_{T_NP} (K²), (b) J_{U_NP} (m² s⁻²), (c) J_{T_US} (K²), and (d) J_{U_US} (m² s⁻²) during August 2004 calculated from the analysis increments directly (solid) and from the total observation impact estimate (dashed). All values have been multiplied by 10⁻³.

responding to a selected set of observations. If all the observations are included in (17), then the result may be compared with (14) as a measure of the accuracy of the observation space estimate of J (and thus the observation sensitivities). Figure 7 shows daily time series of this comparison for each of the response functions examined in this study. The values agree extremely well in all cases and, in many cases, the curves are nearly indiscernible. The results demonstrate the accuracy of the adjoint response to realistic perturbations.

The accuracy of the observation space estimate of J , in turn, allows meaningful aggregation of the results according to observation type, location, channel, etc. Figure 8 shows a basic application of this capability. In this case, the total impact of the observations has been separated into the contributions from the satellite radiances and conventional observations for each of the

response functions examined in this study. For the temperature increments over the North Pacific (Fig. 8a), the abundant satellite radiances in this region (cf. Fig. 4b) affect the analysis of temperature significantly. The radiances account for most of the impact on all days except 13 August, when much of the polar-orbiting satellite data were missing for technical reasons. As might be expected, the situation is reversed for the zonal wind increments over the North Pacific (Fig. 8b). For these increments, conventional wind observations (primarily from *GOES-10* and commercial aircraft, not shown) dominate over the indirect impact of the radiances. Over the U.S. region (Figs. 8c,d), conventional observations have the dominant impact on both the temperature and zonal wind increments. This is not surprising given the abundance of observations from both rawinsondes and aircraft over the continent, which are given significant weight in the

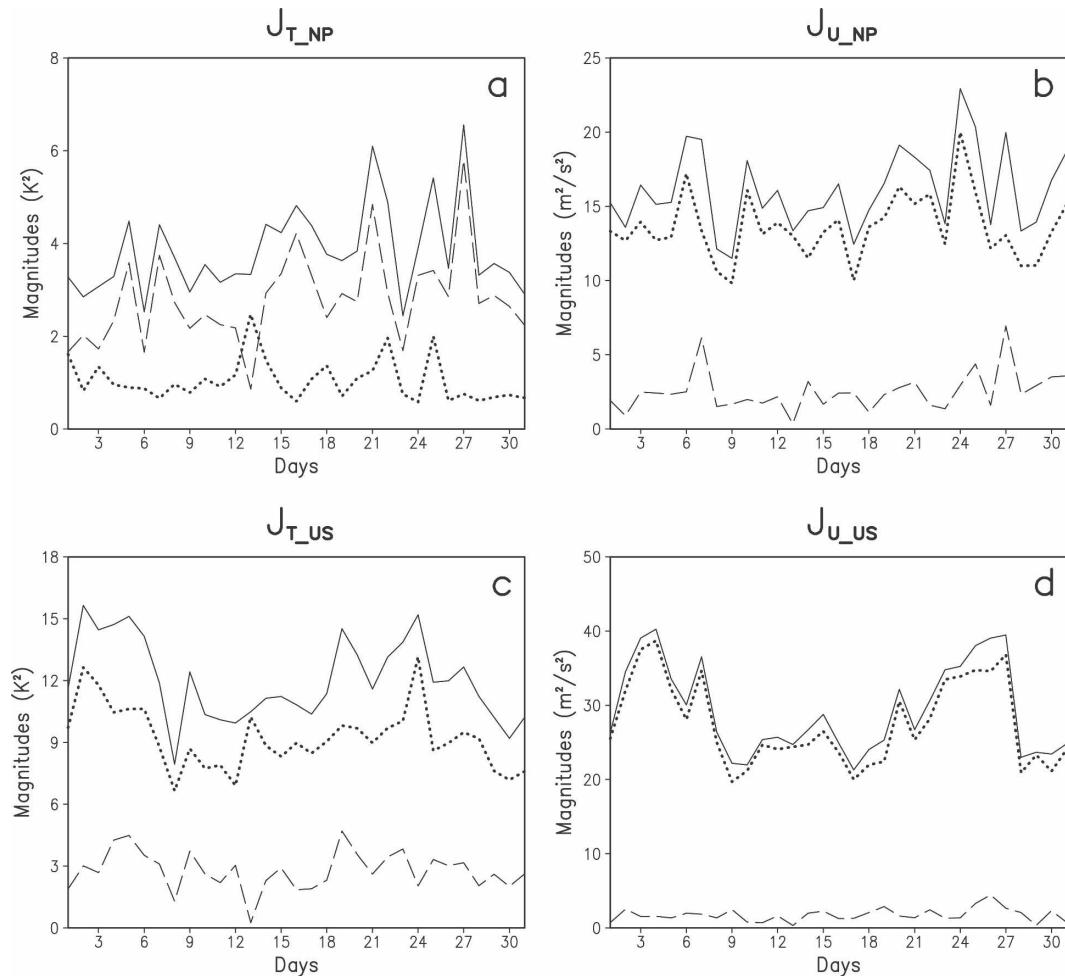


FIG. 8. The contributions of the satellite radiances (dashed) and conventional observations (dotted) to the total observation impact (solid) for (a) $J_{T_{NP}}$ (K^2), (b) $J_{U_{NP}}$ ($m^2 s^{-2}$), (c) $J_{T_{US}}$ (K^2), and (d) $J_{U_{US}}$ ($m^2 s^{-2}$) during August 2004. All values have been multiplied by 10^{-3} .

analysis. For the zonal wind increments in particular, there is almost no impact from satellite radiances despite the coverage by, for example, AMSU-A radiances shown in Fig. 6b.

The results in Fig. 8 can be further separated into the contributions from individual observing systems. Figure 9 shows daily time series of the impact of several different observing systems on $J_{T_{NP}}$. It can be seen that the NOAA-16 AMSU-A radiances have the largest impacts on average (note the different ordinate scalings), and account for most of the large impact from satellite radiances on $J_{T_{NP}}$ observed in Fig. 8a. Interestingly, on 13 August, when much of the radiance data are missing, the impacts of other observing systems (e.g., rawinsondes and surface marine observations) increase significantly compared with their average values for the period. This illustrates the relative nature of the impacts of different observing systems, and possible re-

dundancies between them. Other observing systems, including rawinsondes, HIRS-3 radiances, and aircraft observations from the Meteorological Data Collection and Reporting System (MDCRS; Moninger et al. 2003) have significant but smaller impacts on $J_{T_{NP}}$ than AMSU-A throughout the period. Also note that while the total impact of all observations on $J_{T_{NP}}$ must be positive, the contributions from individual components of the observing system may be negative.

Finally, the impacts of these same observing systems on $J_{U_{US}}$ are presented in Fig. 10. In this case, the impacts of rawinsondes and MDCRS aircraft observations are larger by nearly an order of magnitude, on average, than those of the other observing systems shown. The results are consistent with those in Fig. 8d, which show that conventional observations account for virtually all of the impact on $J_{U_{US}}$ throughout the study period.

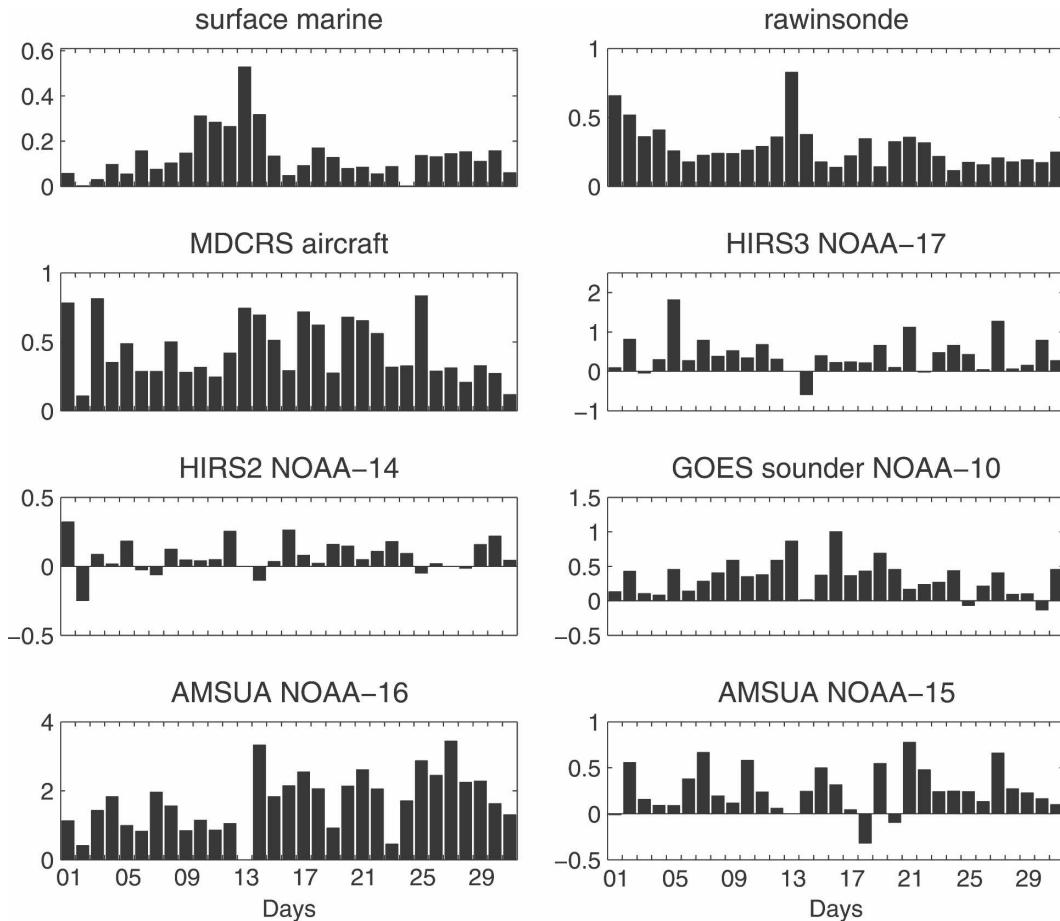


FIG. 9. The impact of various observing systems on J_{TNP} (K^2) during August 2004. All values have been multiplied by 10^{-3} .

5. Conclusions

An exact adjoint of the GSI analysis scheme was developed and tested in the context of the GMAO *GEOS-5* atmospheric data assimilation system. Development of an exact adjoint was deemed an appropriate strategy given the formulation of the existing nonlinear minimization problem, including the quality control procedures and observation error assignment performed in the inner and outer loops of the GSI. As a prerequisite, a line-by-line TLM of the GSI, including the conjugate gradient descent algorithm used in the minimization, was developed and thoroughly tested with realistic-sized perturbations of the input innovations. Comparison of the TLM responses to perturbed innovations with differences between the GSI increments with and without these perturbations reveal that the overall behavior of the GSI is well represented by the TLM. Perturbed innovations of wind, temperature, moisture, and satellite radiances produce highly linear

responses for most of the analyzed variables except ozone and, to a lesser extent, specific humidity. The response to perturbed ozone innovations is linear for the ozone itself, but nonlinear for other variables. Because direct responses—that is, the response of one variable to perturbed innovations of the same observation variable—tend to be large and highly linear, the tangent linear assumption holds strongly for all variables when all the innovations are perturbed simultaneously. Penalty terms for supersaturation and negative humidity in the GSI cost function may cause intermittent, but severe, nonlinearities, which cannot be modeled by the TLM. However, these terms can be omitted or reduced in amplitude when computing the observation sensitivity in such cases without altering the results significantly.

In accordance with the TLM results, the GSI adjoint produces accurate estimates of the sensitivities with respect to observations. In a series of experiments using locally defined response functions based on the analysis

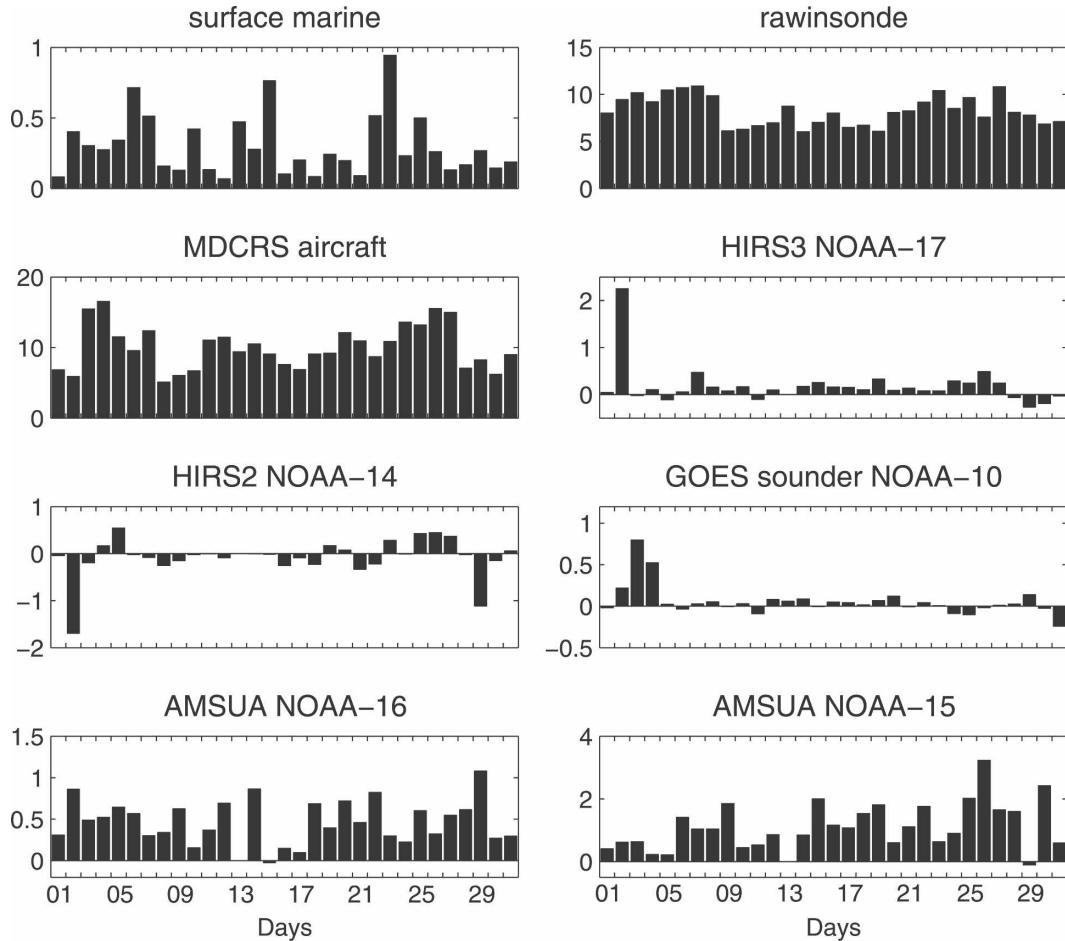


FIG. 10. Same as in Fig. 9, but for $J_{U_{US}}$ ($\text{m}^2 \text{s}^{-2}$). All values have been multiplied by 10^{-3} .

increments of temperature and zonal wind as input to the adjoint, the sensitivities are found to be in good agreement with Baker (2000) and Baker and Daley (2000) in terms of their magnitudes and dependence on the type, distribution, and density of surrounding observations. Larger sensitivities are observed with respect to observations close to where the response function is defined, while much smaller sensitivities are observed with respect to observations elsewhere. For observations of the same type and in the same general location, the sensitivity is largest when the density of surrounding observations is low, and vice versa.

A powerful application of the observation sensitivity information is in estimating the impact of a set of observations on a given response function. For the response functions studied here, the impact is easily computed from the inner product between the observation sensitivities and the corresponding innovations. It was found, for example, that AMSU-A radiances have the largest impact of all observing systems on the temperature increments over the eastern North Pacific, while

conventional observations from rawinsondes and aircraft dominate the impact on the zonal wind increments over the continental United States. The combined impact of all observations provides an observation space-based estimate of the total response function, which may be compared with the response function computed from the analysis increments directly. The observation and analysis space values were found to be in extremely close agreement in all cases examined, confirming the accuracy of the observation sensitivities. As shown by Langland and Baker (2004), the combined use of sensitivity information from the analysis and forecast model adjoints can be used effectively to estimate the impact of observations on short-range forecasts. Experiments combining the GSI and *GEOS-5* model adjoints are in progress at the GMAO and will be reported on in a future study.

The adjoint results presented in this study were produced using a single outer loop of the minimization algorithm, while the GSI itself is usually run with multiple (usually two) outer loops to accommodate small

nonlinear effects from observation operators such as for wind speed and precipitation. This capability is currently being developed for the adjoint. However, as successive outer loops tend to make only small changes to the increments, we do not anticipate significant qualitative differences with the adjoint results presented here. In addition, more recent versions of the GSI include variational quality control procedures and a redefined control variable for humidity that reduces the need for the highly nonlinear moisture penalty terms in the present version. These features are currently being incorporated into the adjoint and their impacts will also be reported in future studies.

Acknowledgments. The authors thank Ricardo Todling and Ron Errico of the Global Modeling and Assimilation Office (GMAO) for their helpful discussion of this work and review of the manuscript. We also thank John Derber, Russ Treadon, and Daryl Kleist of the National Centers for Environmental Prediction (NCEP) for their discussion and guidance on various development aspects of this work. The comments of two anonymous reviewers were very helpful in improving the original manuscript. This work was supported by the Atmospheric Data Assimilation Development component of the NASA Modeling, Analysis and Prediction program (MAP/04-0000-0080).

REFERENCES

- Baker, N., 2000: Observation adjoint sensitivity and the adaptive observation-targeting problem. Ph.D. thesis, Naval Postgraduate School, Monterey, CA, 332 pp.
- , and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Quart. J. Roy. Meteor. Soc.*, **126**, 1431–1454.
- Cardinali, C., S. Pezzulli, and E. Andersson, 2004: Influence-matrix diagnostic of a data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **130**, 2767–2786.
- Courtier, P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy operational implementation of 4-D VAR using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387.
- Daley, R., and E. Barker, 2001: NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.*, **129**, 869–883.
- Doerenbecher, A., and T. Bergot, 2001: Sensitivity to observations applied to FASTEX cases. *Nonlinear Processes Geophys.*, **8**, 467–481.
- English, S., R. Saunders, B. Candy, M. Forsythe, and A. Collard, 2004: Met Office satellite data OSEs. *Proc. Third WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction*, WMO/Tech. Doc. 1228, Alpbach, Austria, WMO, 146–156.
- Fourrie, N., A. Doerenbecher, T. Bergot, and A. Joly, 2002: Adjoint sensitivity of the forecast to TOVS observations. *Quart. J. Roy. Meteor. Soc.*, **128**, 2759–2777.
- Goldberg, M. D., Y. Qu, L. M. McMillin, W. Wolf, L. Zhou, and M. Divakarla, 2003: AIRS near-real-time products and algorithms in support of numerical weather prediction. *IEEE Trans. Geosci. Remote Sens.*, **41** (2), 379–389.
- Kelly, G., T. McNally, J.-N. Thepaut, and M. Szyndel, 2004: OSEs of all main data types in the ECMWF operation system. *Proc. Third WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction*, WMO/Tech. Doc. 1228, Alpbach, Austria, WMO, 63–94.
- Langland, R. H., and N. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56**, 189–201.
- Le Dimet, F.-X., H. E. Ngodock, and J. Vernon, 1995: Sensitivity analysis in variational data assimilation. *Proc. Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography*, WMO/Tech. Doc. 651, Vol. 1, Tokyo, Japan, WMO, 99–104.
- Le Marshall, J., and Coauthors, 2006: Improving global analysis and forecasting with AIRS. *Bull. Amer. Meteor. Soc.*, **87**, 891–894.
- Lord, S., T. Zapotocny, and J. Jung, 2004: Observing system experiments with NCEP's global forecast system. *Third WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction*, WMO/Tech. Doc. 1228, Alpbach, Austria, WMO, 56–62.
- Menke, W., 1984: *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, 289 pp.
- Moninger, W. R., R. D. Mamrosh, and P. M. Pauley, 2003: Automated meteorological reports from commercial aircraft. *Bull. Amer. Meteor. Soc.*, **84**, 203–216.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Purser, R. J., W. Wu, D. F. Parrish, and N. M. Roberts, 2003a: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1524–1535.
- , —, —, and —, 2003b: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1536–1548.
- Rabier, F., N. Fourrie, D. Chafai, and P. Prunet, 2002: Channel selection methods for infrared atmospheric sounding interferometer radiances. *Quart. J. Roy. Meteor. Soc.*, **128**, 1011–1027.
- Rienecker, M. M., and Coauthors, 2007: The GEOS-5 data assimilation system—Documentation of version 5.0.1 and 5.1.0. Technical Report Series on Global Modeling and Data Assimilation, Vol. 27, NASA Tech. Memo. 104606.
- Wu, W., R. J. Purser, and D. F. Parrish, 2002: Three dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.