Recent developments in the use and understanding of adjoint-derived estimates of observation impact in NWP

Ron Gelaro
NASA/GMAO

Thanks to R. Todling (GMAO) and R. Langland (NRL)

8th Workshop on Adjoint Model Applications in Dynamic Meteorology
May 2009 – Tannersville, PA

Background / Outline for this Talk

• Adjoint-based estimates of obs impact are now an accepted alternative / complement to traditional data denial experiments (OSEs) for assessing the impact of observations in NWP

✓ Used currently by several centers for experimentation or routine monitoring of the global observing system

✓ Intercomparison project between centers in progress

• For linear analysis problems, observation impact is closely related to (is an extension of) observation sensitivity…discussed at previous Adjoint Workshops

• This talk touches on:

✓ Initial intercomparison of results for two centers

✓ Need for, implications of >1st order estimates of impact

✓ Extension to nonlinear analysis problems

✓ Comparison, complementarity with OSEs

The Data Assimilation System

- Consider a forecast model: $\mathbf{x}^f = \mathbf{m}(\mathbf{x}_0)$

  and atmospheric analysis: $\mathbf{x}_\mathrm{a} = \mathbf{x}_\mathrm{b} + \mathbf{K}[\mathbf{y} - \mathbf{h}(\mathbf{x})]$

  where $\mathbf{x}_\mathrm{b}$ is a short forecast, $\mathbf{y}$ are observations, $\mathbf{h}$ is a (possibly nonlinear) observation operator and $\mathbf{K}$ determines the weight, or gain, given to each observation

  …the difference $\delta\mathbf{y} = \mathbf{y} - \mathbf{h}(\mathbf{x})$ is the innovation vector

- Assume, for now, that $\mathbf{h}$ is either <u>linear</u> or only a function of $\mathbf{x}_\mathrm{b}$, and define the analysis increment:

$$\delta\mathbf{x}_0 = \mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b} = \mathbf{K}\delta\mathbf{y} \qquad (1)$$

  Note that (1) may be viewed as a transformation between a perturbation $\delta\mathbf{x}_0$ in *state space* a perturbation $\delta\mathbf{y}$ in *observation space*

Observation Sensitivity: Data Assimilation System Adjoint

Baker and Daley (2000) showed that the sensitivity of the analysis to observations could be computed using the adjoint of the DAS

$$\partial \mathbf{x}_{\mathrm{a}} / \partial \mathbf{y} = \mathbf{K}^{\mathrm{T}}$$

- The sensitivity of a measure J with respect to the initial conditions (analysis) is then extended into observation space as

$$\frac{\partial J}{\partial \mathbf{y}} = \frac{\partial \mathbf{x}_{\mathrm{a}}}{\partial \mathbf{y}} \frac{\partial J}{\partial \mathbf{x}_{\mathrm{a}}} = \mathbf{K}^{\mathrm{T}} \frac{\partial J}{\partial \mathbf{x}_{\mathrm{a}}}$$

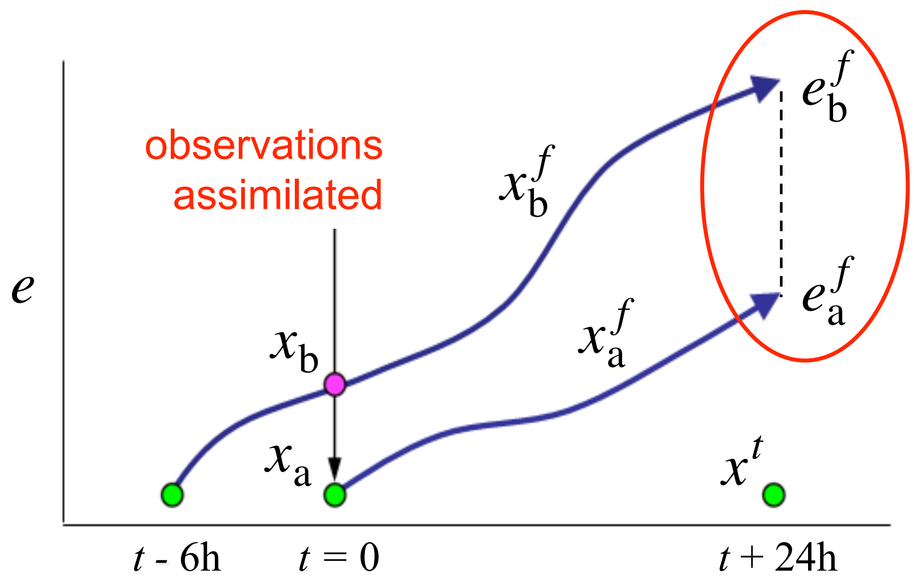
- If J is based on a model forecast, then the sensitivity of J with respect to the observations is

$$\frac{\partial J}{\partial \mathbf{y}} = \mathbf{K}^{\mathrm{T}} \mathbf{M}^{\mathrm{T}} \frac{\partial J}{\partial \mathbf{x}^{f}}$$

where $\mathbf{M}^{\mathrm{T}}$ is the adjoint of $\mathbf{m}$

Estimating the Impact of Observations on Forecasts

Langland and Baker (2004) showed that the adjoint of a data assimilation system could be used effectively to measure the impact of observations on forecast skill



• Consider forecasts from an analysis x_a and background state x_b , and energy-based measure of forecast error

$$e = (\mathbf{x}^f - \mathbf{x}^t)^T \mathbf{C} (\mathbf{x}^f - \mathbf{x}^t)$$

where x^t is a verification analysis state

• The difference $\delta e = e_a^f - e_b^f$ measures the combined impact of all obs assimilated at $t = 0$…

…it can be estimated as a sum of contributions from *__individual__* *obs using information from the model and analysis adjoints*

LB04 Observation Impact Estimate

$$\delta e \approx (\delta \mathbf{y})^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} [\mathbf{M}_{\mathrm{b}}^{\mathrm{T}} \mathbf{C}(\mathbf{x}_{\mathrm{b}}^{f} - \mathbf{x}^{t}) + \mathbf{M}_{\mathrm{a}}^{\mathrm{T}} \mathbf{C}(\mathbf{x}_{\mathrm{a}}^{f} - \mathbf{x}^{t})]$$

analysis adjoint   model adjoint

- The impact of arbitrary subsets of observations can be quantified by summing only terms involving the desired elements of $\delta \mathbf{y}$

- The vector $\mathbf{K}^{\mathrm{T}}[...]$ is computed only once and involves the **entire set** of observations

  …removing or changing the properties of one observation changes the scalar measure of all other observations

- Application is subject to assumptions and simplifications in $\mathbf{M}^{\mathrm{T}}$

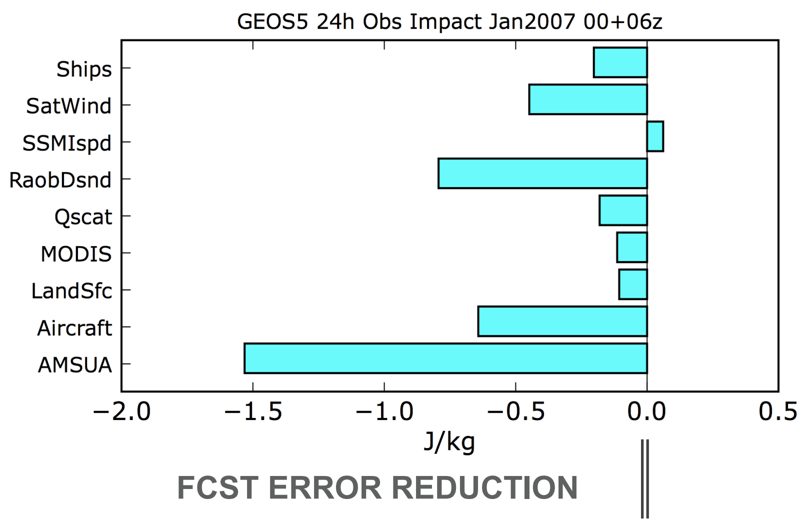
$\delta e < 0$ …the observation **improves** the forecast

$\delta e > 0$ …the observation **degrades** the forecast

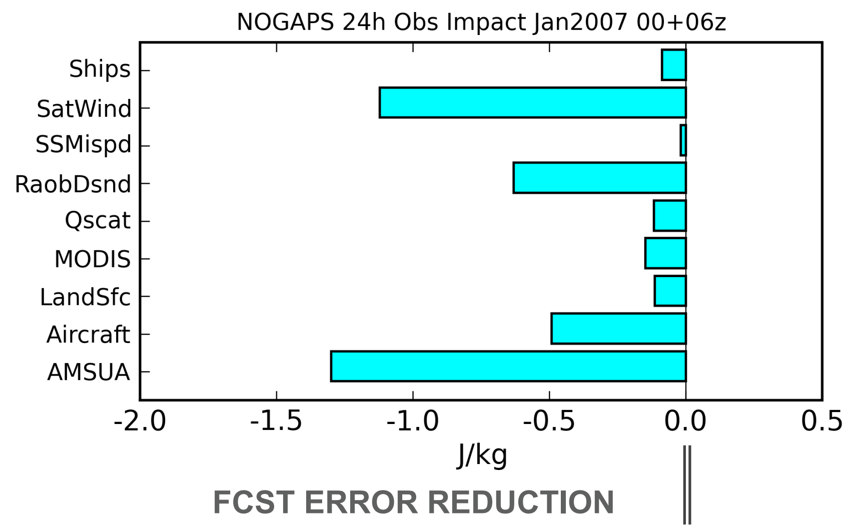
Daily Average Impacts of Major Observing Systems

Global Baseline Jan 2007 00+06 UTC

NASA GEOS-5



Navy NOGAPS



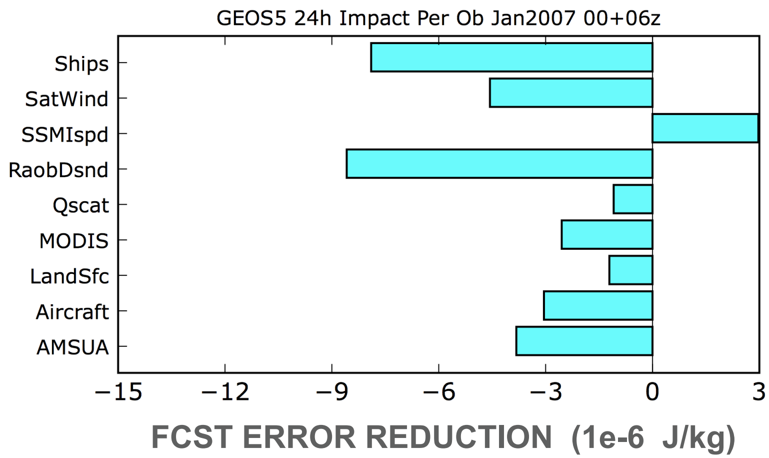
Overall impacts similar in NASA and Navy systems despite differences in algorithms, RT models, observation counts…

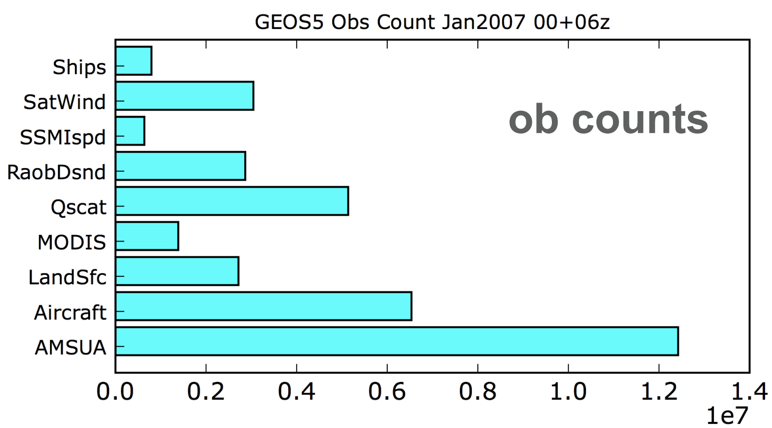
…notable differences in Satwinds, SSMI speeds

Impacts per Observation

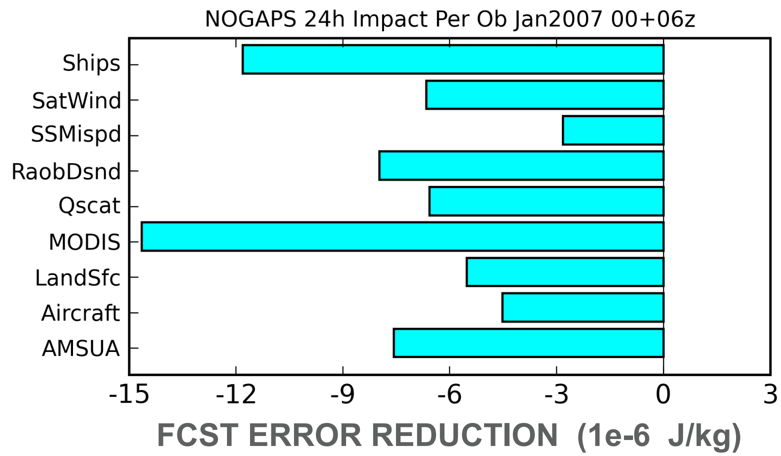
Global Baseline Jan 2007 00+06 UTC

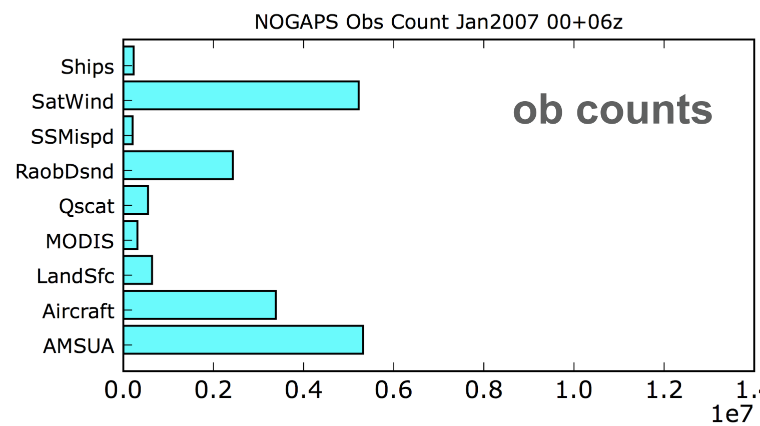
NASA GEOS-5





Navy NOGAPS

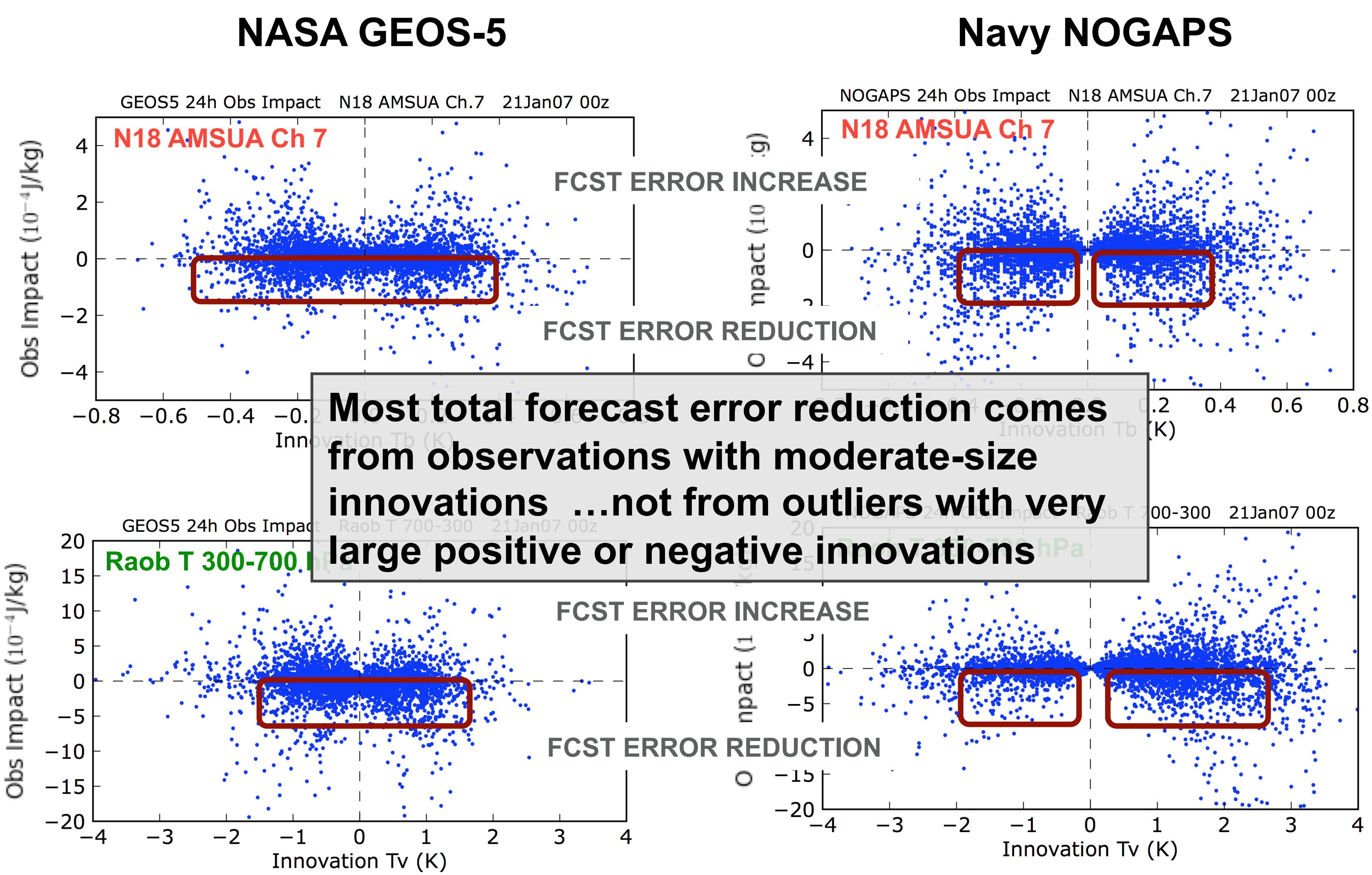




GEOS-5 has smaller impacts per-ob, because more observations are assimilated – TOTAL impacts are similar (previous slide)

Scatter of Observation Impact vs Innovation

Baseline Intercomparison 21 Jan 2007 00UTC



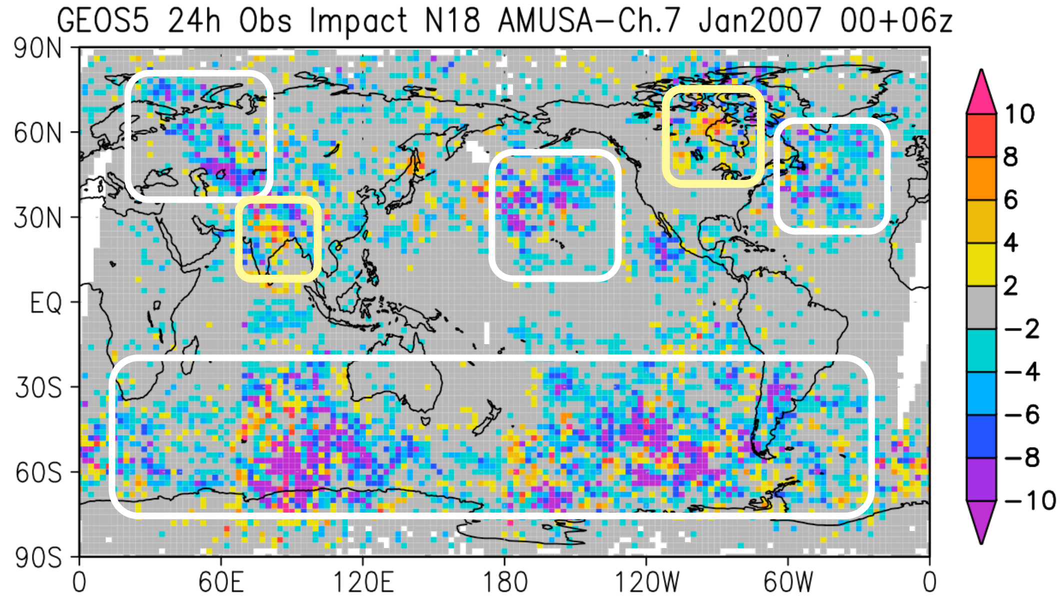
Most total forecast error reduction comes from observations with moderate-size innovations …not from outliers with very large positive or negative innovations

Observation Impacts for NOAA-18 AMSU-A Ch. 7

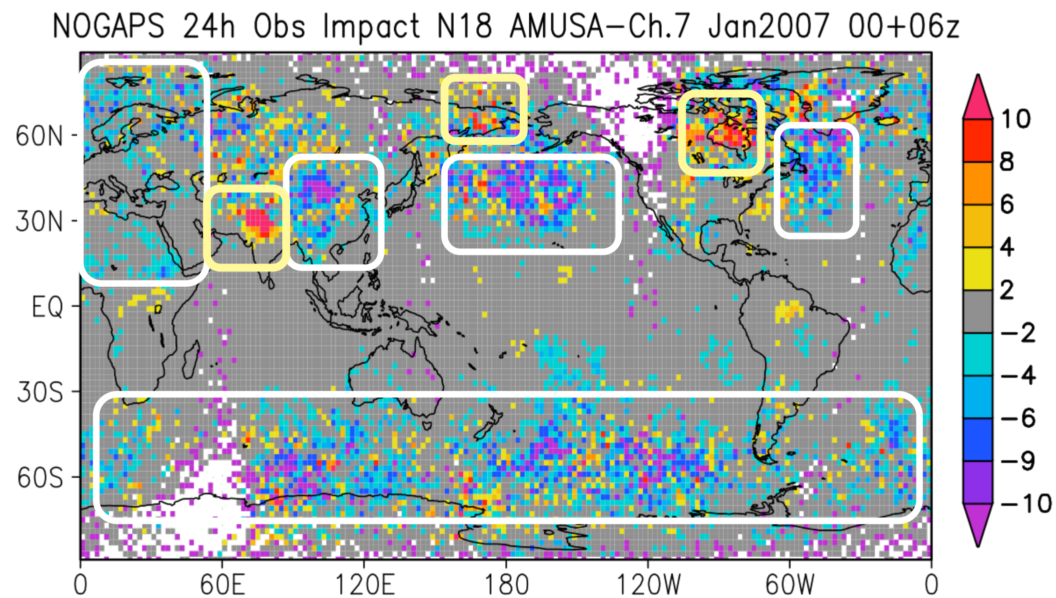
Observations that produce large forecast error reductions

Observations that produce forecast error increases in **both models**

Land or ice surface contamination of radiance data?



NASA GEOS-5



Navy NOGAPS

Baseline Intercomparison
Jan 2007 00+06 UTC

Orders of Approximation of δe

Errico (2007) placed the LB04 measure in the context of various-order Taylor series approximations of δe in terms of $\delta \mathbf{y}$:

1st order:

$$\delta e_1 = \delta \mathbf{y}^{\mathrm{T}} \underbrace{2\mathbf{K}^{\mathrm{T}}\mathbf{M}_{\mathrm{b}}^{\mathrm{T}}\mathbf{C}(\mathbf{x}_{\mathrm{b}}^{f} - \mathbf{x}^{t})}_{\tilde{\mathbf{g}}_1}$$

2nd order:

$$\delta e_2 = \delta \mathbf{y}^{\mathrm{T}} \underbrace{\mathbf{K}^{\mathrm{T}}[\mathbf{M}_{\mathrm{b}}^{\mathrm{T}}\mathbf{C}(\mathbf{x}_{\mathrm{a}}^{f} - \mathbf{x}^{t}) + \mathbf{M}_{\mathrm{a}}^{\mathrm{T}}\mathbf{C}(\mathbf{x}_{\mathrm{b}}^{f} - \mathbf{x}^{t})]}_{\tilde{\mathbf{g}}_2}$$

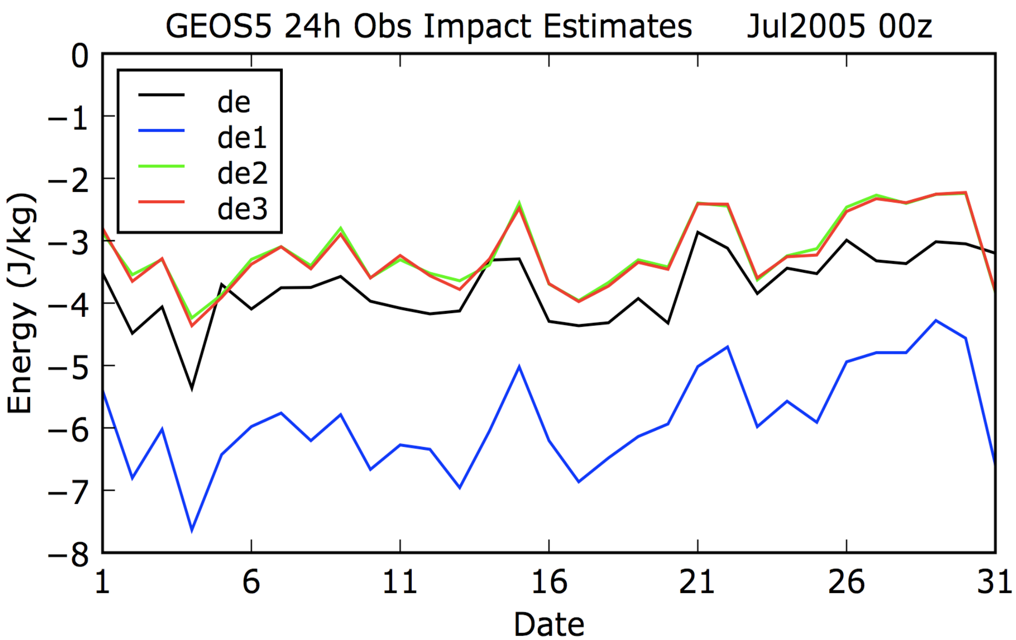
3rd order: **(LB04)**

$$\delta e_3 = \delta \mathbf{y}^{\mathrm{T}} \underbrace{\mathbf{K}^{\mathrm{T}}[\mathbf{M}_{\mathrm{b}}^{\mathrm{T}}\mathbf{C}(\mathbf{x}_{\mathrm{b}}^{f} - \mathbf{x}^{t}) + \mathbf{M}_{\mathrm{a}}^{\mathrm{T}}\mathbf{C}(\mathbf{x}_{\mathrm{a}}^{f} - \mathbf{x}^{t})]}_{\tilde{\mathbf{g}}_3}$$

+ a higher order term

Note that $\tilde{\mathbf{g}}_1$ is a gradient and independent of δy, but $\tilde{\mathbf{g}}_2$ and $\tilde{\mathbf{g}}_3$ are weights that depend on all δy through $\mathbf{x}_{\mathrm{a}}$

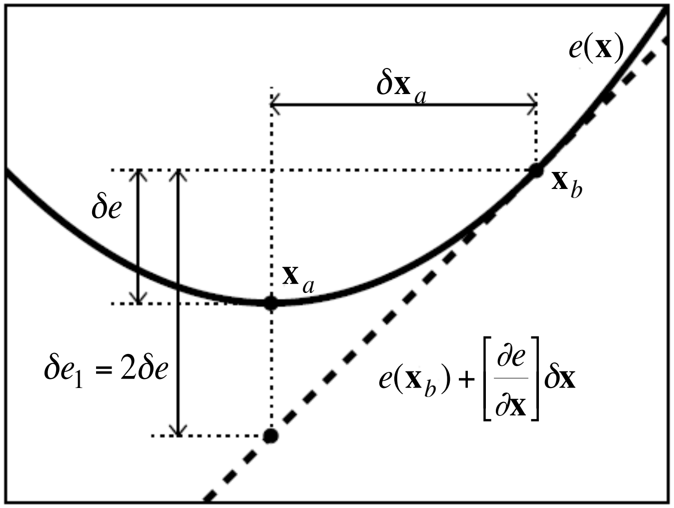
First- vs. Higher-Order Approximations of δe



Gelaro et al. (2007)

- Higher-than-first-order approximation of impact required due to quadratic nature of e

Trémolet (2007)



- If $\mathbf{x}_a$ is near the minimum of e, then the first order approximation will be twice the correct value.*

*$\delta e \approx \frac{1}{2}\delta e_1$ *is a tempting approximation, but dangerous if the forecast is poor*

The ‘Price’ of Higher-Order Accuracy

Terms beyond first-order in the approximation δe_3 have the form:

$$\delta e_3 - \delta e_1 \approx (\delta \mathbf{y})^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{M}^{\mathrm{T}} \mathbf{C} \mathbf{M} \mathbf{K} (\delta \mathbf{y})$$

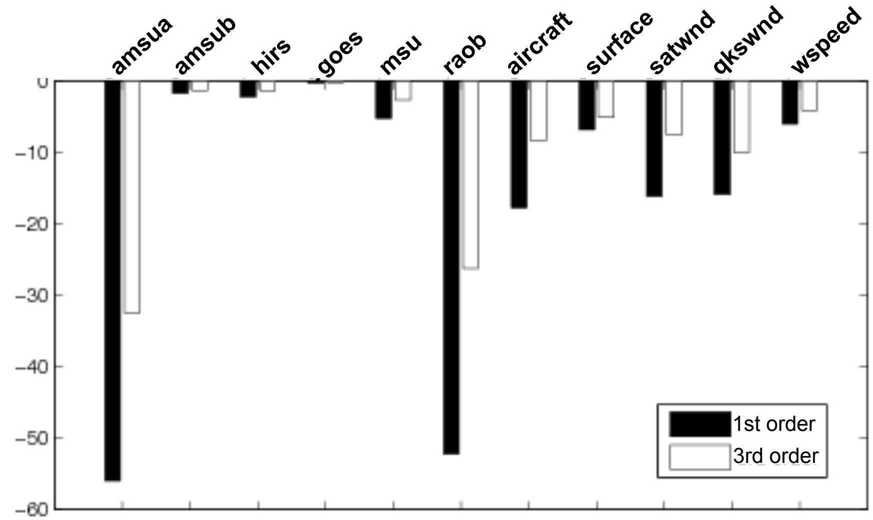
Errico (2007) pointed out that the nonlinear dependence of these terms on δy means **partial sums** of δe_3 involve **cross terms** with other observations and therefore possible ambiguities

Gelaro et al. (2007) found this effect to be small for partial sums measuring average impacts of the major observing systems…

…smaller subsets?

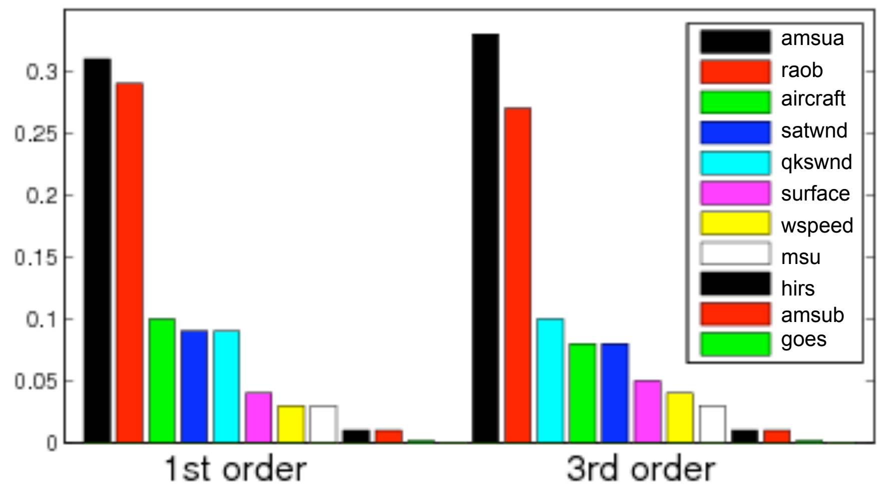
Accumulated values of forecast error reduction
(J/kg)

Totals for July 2005 00UTC



Order of approx-imation affects the magnitudes of the impact estimates (~2x)...

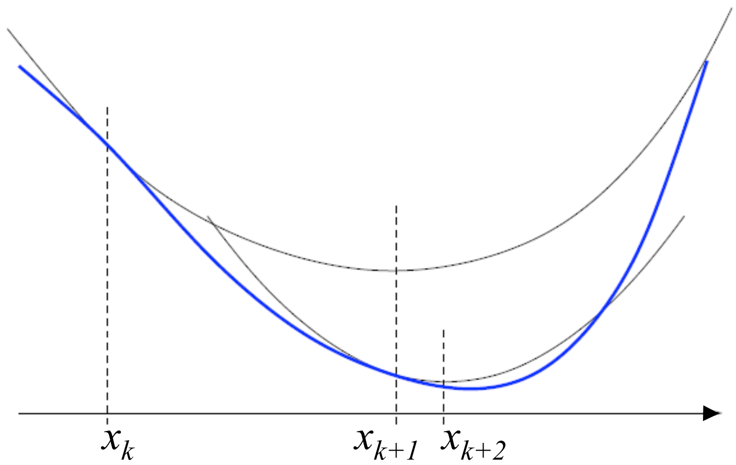
Ranked fractional contributions to forecast error reduction



..but not the relative contribution of each obs system to the overall error reduc-tion

Gelaro et al. (2007)

Nonlinear Analysis Problems

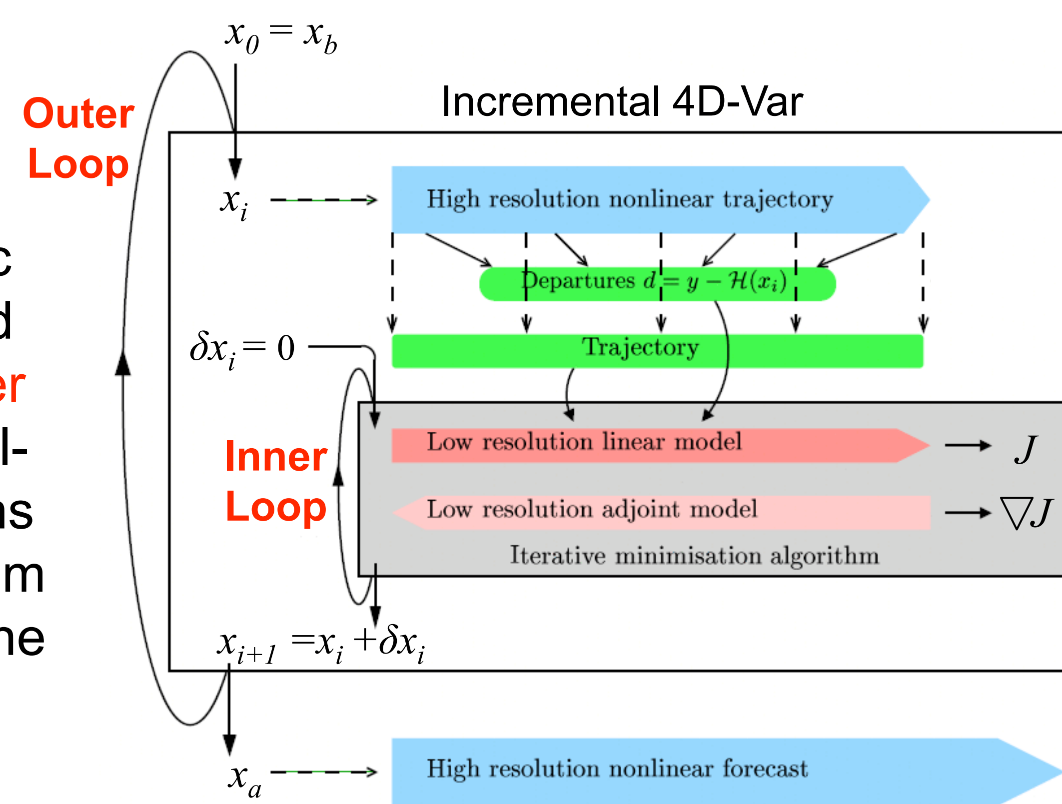


• In general, the analysis cost function is nonlinear and difficult to minimize

• One complex problem is replaced by a series of slightly easier ones

…***incremental formulation***

• An approximate quadratic cost function is defined and minimized repeatedly (outer loop) until a satisfactory solution is found; the iterations of the minimization algorithm within each outer loop define the inner loop



Observation Impact in Incremental Variational Data Assim.

Trémolet (2008) examined observation impact in a variational data assimilation system, accounting for $j = 1,...,m$ outer loops

Increment is not: $\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b} = \mathbf{K}\,\delta\mathbf{y}$

It is, after loop j : $\mathbf{x}_j - \mathbf{x}_\mathrm{b} = \mathbf{K}_j\delta\mathbf{y}_j + \mathbf{K}_j\mathbf{H}_j(\mathbf{x}_{j-1} - \mathbf{x}_\mathrm{b})$

or $\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{b} = \sum_{j=1}^{m}\mathbf{L}_j\mathbf{K}_j\,\delta\mathbf{y}_j$

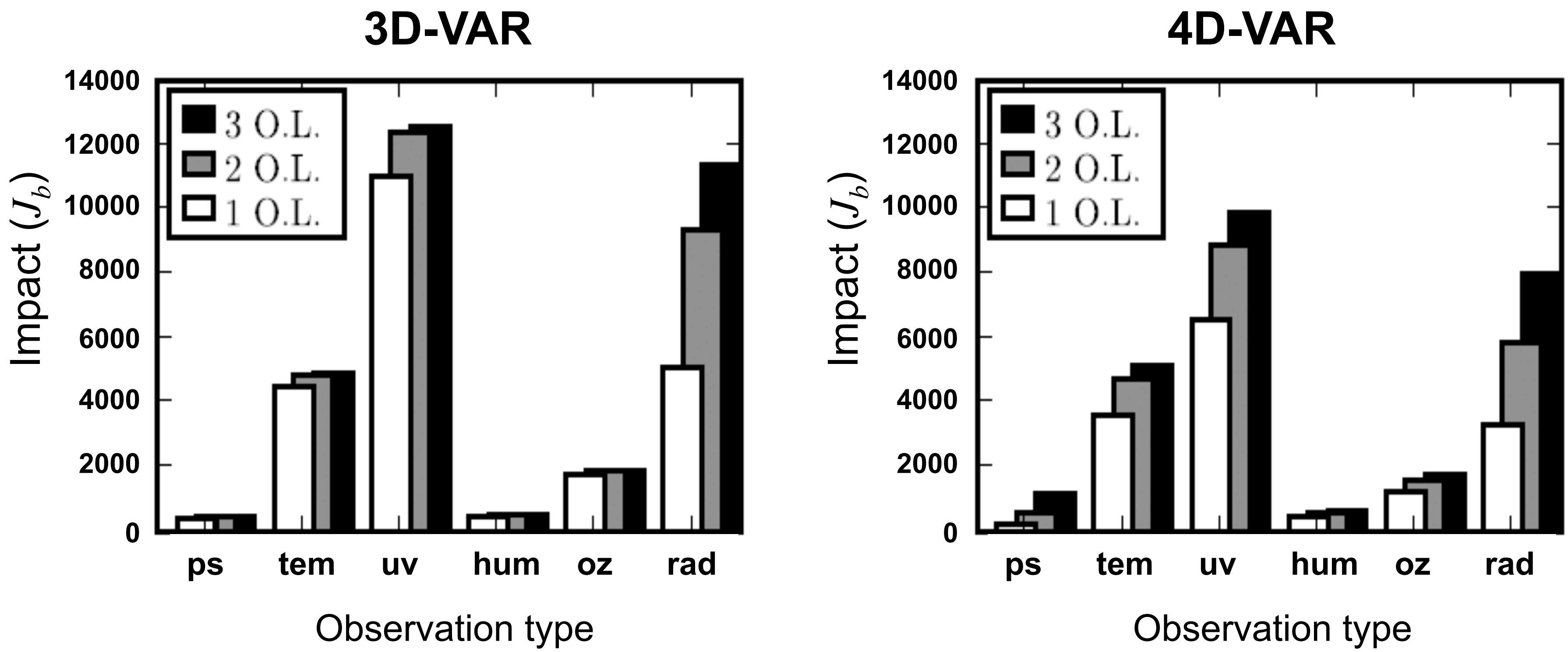
where $\mathbf{L}_j = \mathbf{K}_m\mathbf{H}_m ...\mathbf{K}_{j+1}\mathbf{H}_{j+1}$ and $\mathbf{L}_m = \mathbf{I}$

Then observation impact is: $I \approx \sum_{j=1}^{m}\left\langle \mathbf{K}_j^\mathrm{T}\mathbf{L}_j^\mathrm{T}\mathbf{g}, \delta\mathbf{y}_j \right\rangle$

where $\mathbf{g}$ is a gradient or weight in model space

For example, with m=2 outer loops: $I \approx \left\langle \mathbf{K}_1^\mathrm{T}\mathbf{H}_2^\mathrm{T}\mathbf{K}_2^\mathrm{T}\mathbf{g}, \delta\mathbf{y}_1 \right\rangle + \left\langle \mathbf{K}_2^\mathrm{T}\mathbf{g}, \delta\mathbf{y}_2 \right\rangle$

Observation Impact with Outer Loops



Impact per observation type on the analysis increment with 1, 2, and 3 outer loop iterations

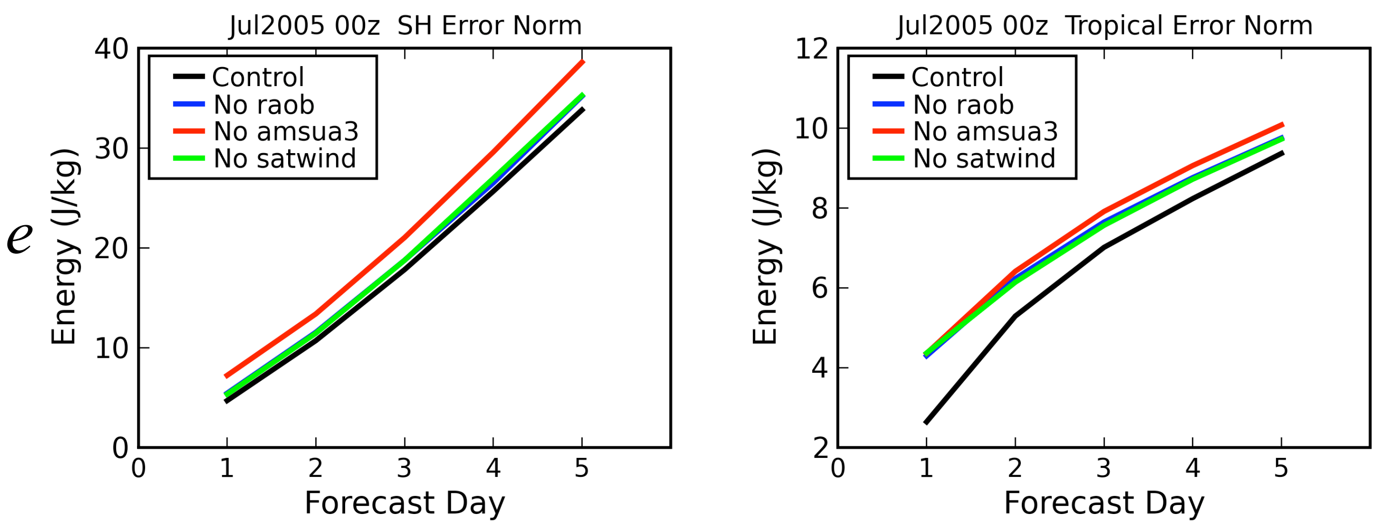
- Outer loop (nonlinear) effects are larger in 4D-Var
- Overall observation impact is smaller in 4D-Var

Trémolet (2008)

Observing System Experiments (OSEs)

- Subsets of observations are **removed** from the assimilation system and forecasts are compared against a control system that includes all observations

- Because of expense, usually involve a relatively small number of independent experiments, each considering a relatively large subset of observations



Gelaro and Zhu (2009)

Comparison and Interpretation of ADJ and OSE Results

…a few things to keep in mind…

ADJ: measures the impacts of observations in the context of all other observations present in the assimilation system

OSE: removal of observations changes or degrades the system…$\mathbf{K}$ differs for each member

❖

ADJ: measures the impact of observations in each analysis cycle separately and against the control background

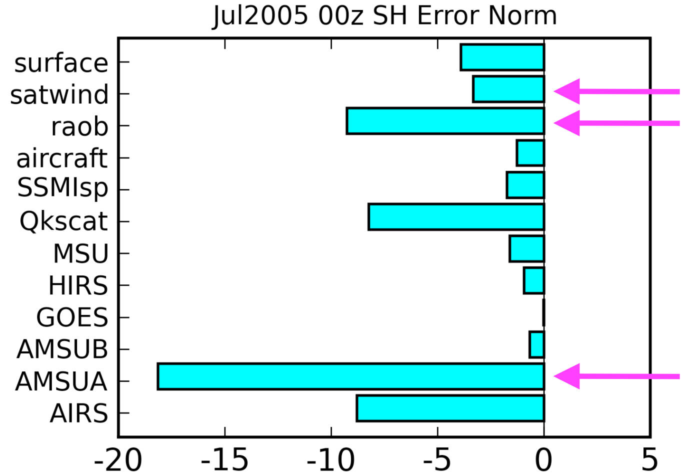
OSE: measures the impact of removing information from both the background and analysis in a cumulative manner

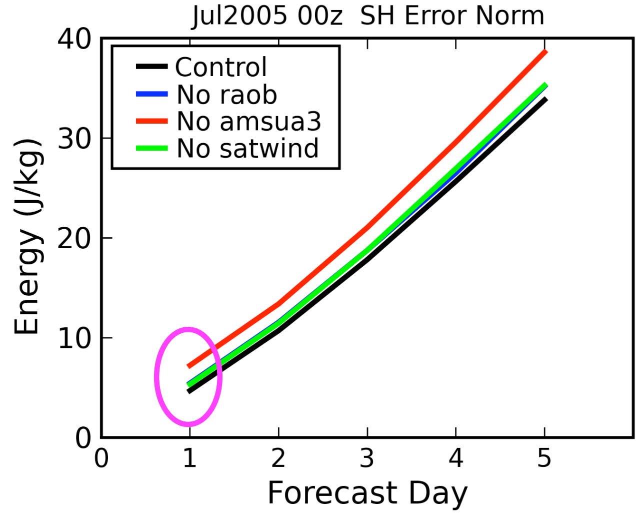
❖

ADJ: measures the response of a single forecast metric to all perturbations of the observing system

OSE: measures the effect of a single perturbation on all forecast metrics

Quantitative Comparison of ADJ and OSE Results





- Strictly speaking, quantitative comparison is limited to the forecast range and metric for which the ADJ results are valid on the one hand (e.g. 24h SH e-norm) and to the selected observing systems removed in the OSEs on the other hand

- Even then, comparisons between the ADJ and OSE results are complicated by the fact that values/changes in e measured in the OSE context are not directly comparable to values of δe measured in the ADJ context

Quantitative Comparison of ADJ and OSE Results

OSE: $$e = (\mathbf{x}_0^f - \mathbf{x}^t)^{\mathrm{T}} \mathbf{C} (\mathbf{x}_0^f - \mathbf{x}^t)$$

ADJ: $$\delta e = (\delta \mathbf{y})^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} [\mathbf{M}_{\mathrm{b}}^{\mathrm{T}} \mathbf{C} (\mathbf{x}_{\mathrm{b}}^f - \mathbf{x}^t) + \mathbf{M}_{\mathrm{a}}^{\mathrm{T}} \mathbf{C} (\mathbf{x}_{\mathrm{a}}^f - \mathbf{x}^t)]$$

Gelaro and Zhu (2009) defined a fractional impact F_j of observing system j for each approach:

$$F_j(\mathrm{ADJ}) = \delta e_j / \delta e$$

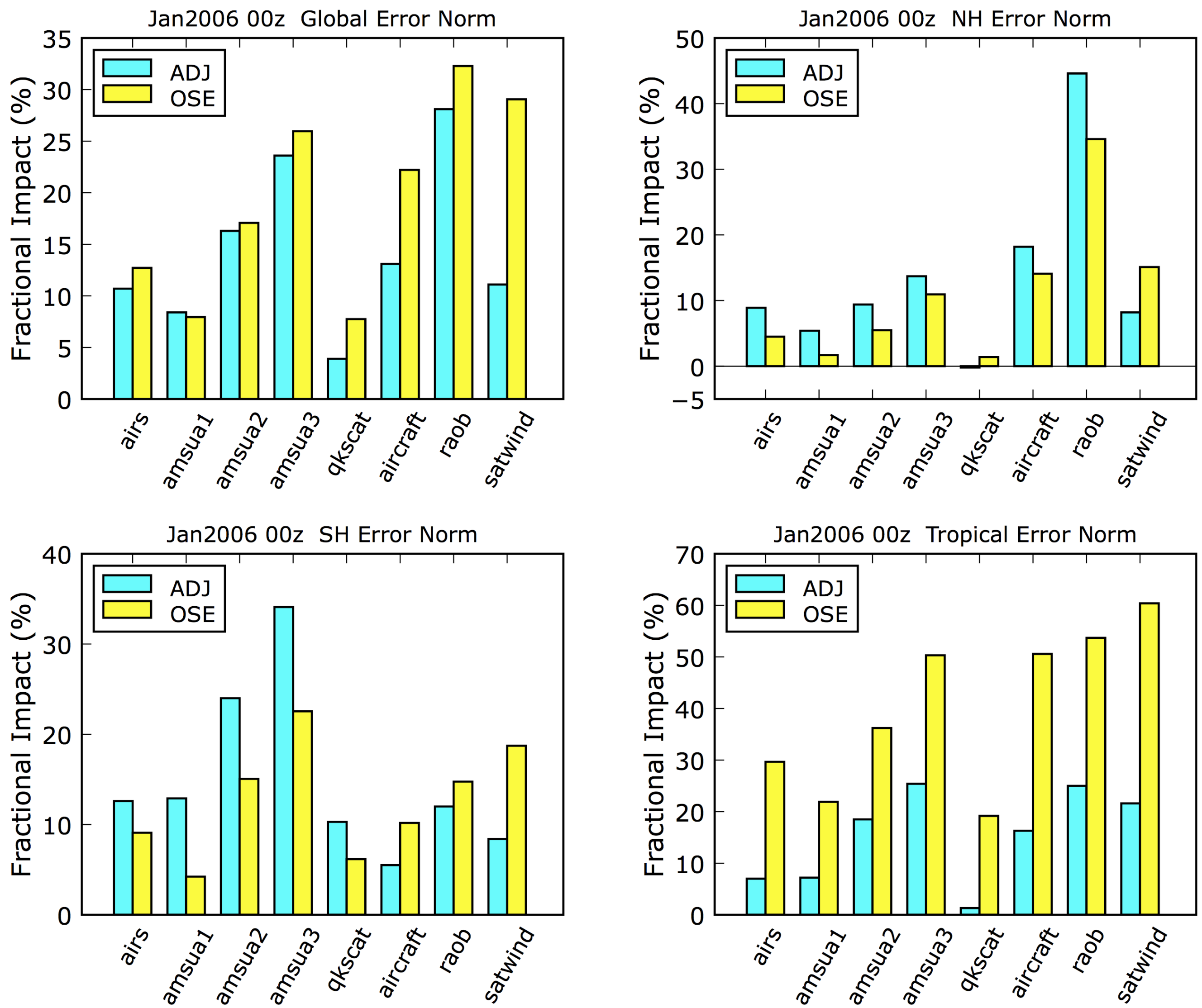
- Measures the % **decrease** in error due to the **presence** of obs system j with respect to the background forecast
- $\sum_j F_j(\mathrm{ADJ}) = 1$

$$F_j(\mathrm{OSE}) = (e_{\mathrm{no}\,j} - e_{\mathrm{ctl}}) / e_{\mathrm{ctl}}$$

- Measures the % **increase** in error due to the **removal** of obs system j with respect to the control forecast
- $\sum_j F_j(\mathrm{OSE}) \neq 1$

% Contributions to 24hr Forecast Error Reduction

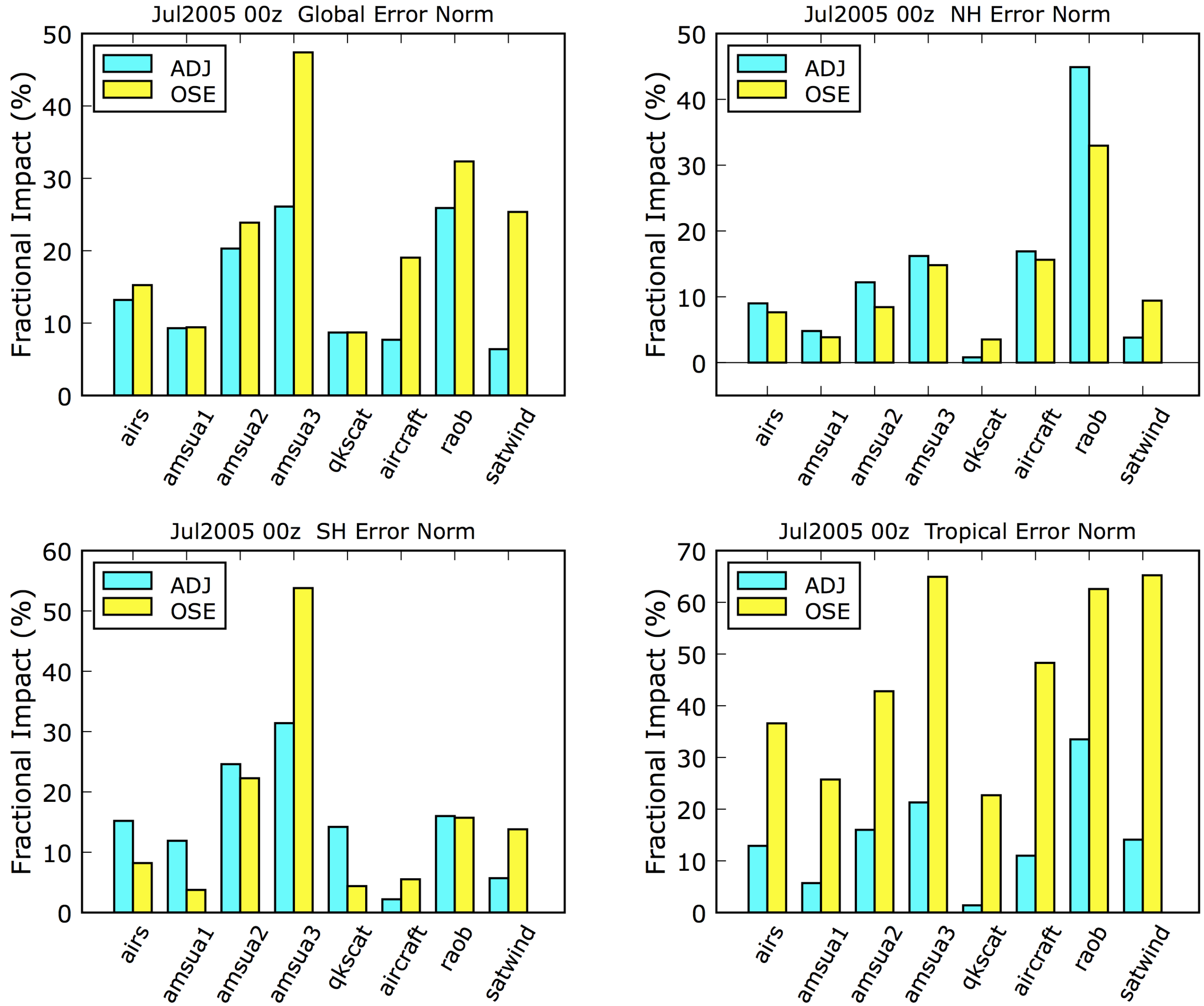
January 2006



Gelaro and Zhu (2009)

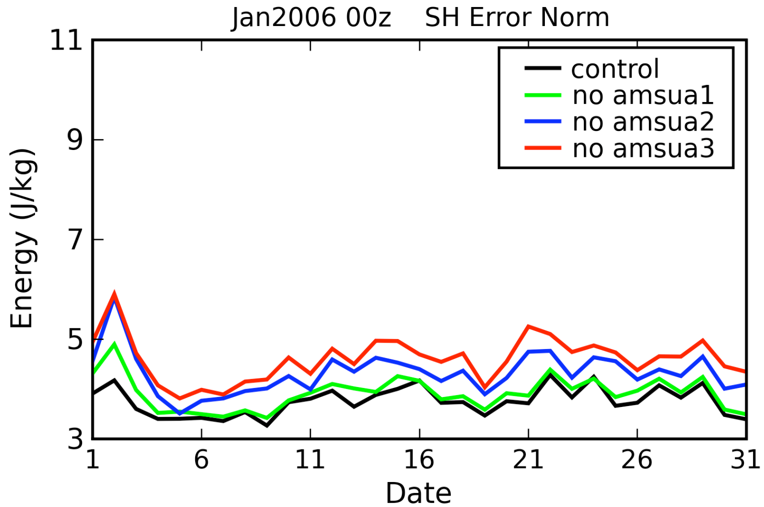
% Contributions to 24hr Forecast Error Reduction

July 2005

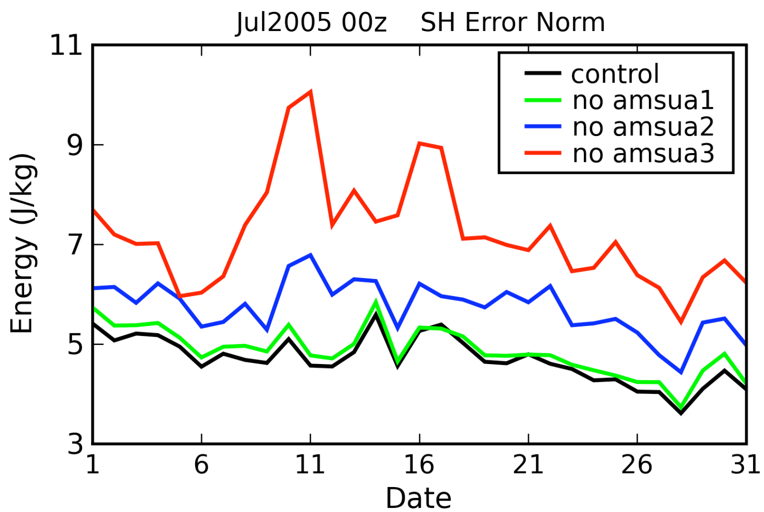


Gelaro and Zhu (2009)

OSE Time Series of SH 24-hr Forecast Error Norm



January 2006

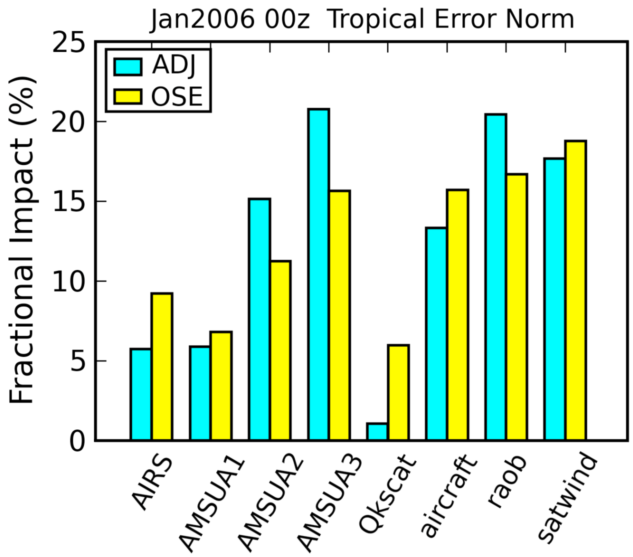


July 2005

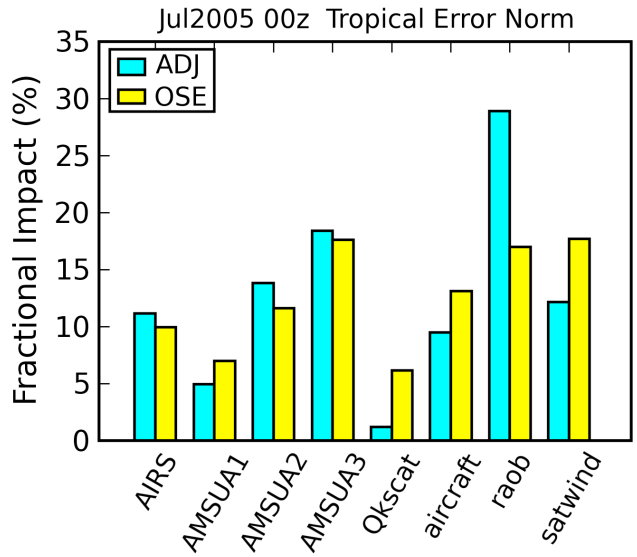
Skill collapses when all AMSUA removed during SH winter…OSE and ADJ results become difficult to compare

Normalized % Contributions to 24hr Forecast Error Reduction

January 2006



July 2005



…ADJ and OSE responses differ in magnitude in the tropics, but assign similar relative ‘value’ to the various observing systems

Combined Use of ADJ and OSEs

- Both OSEs and ADJ measure the net effect of observations on the forecast

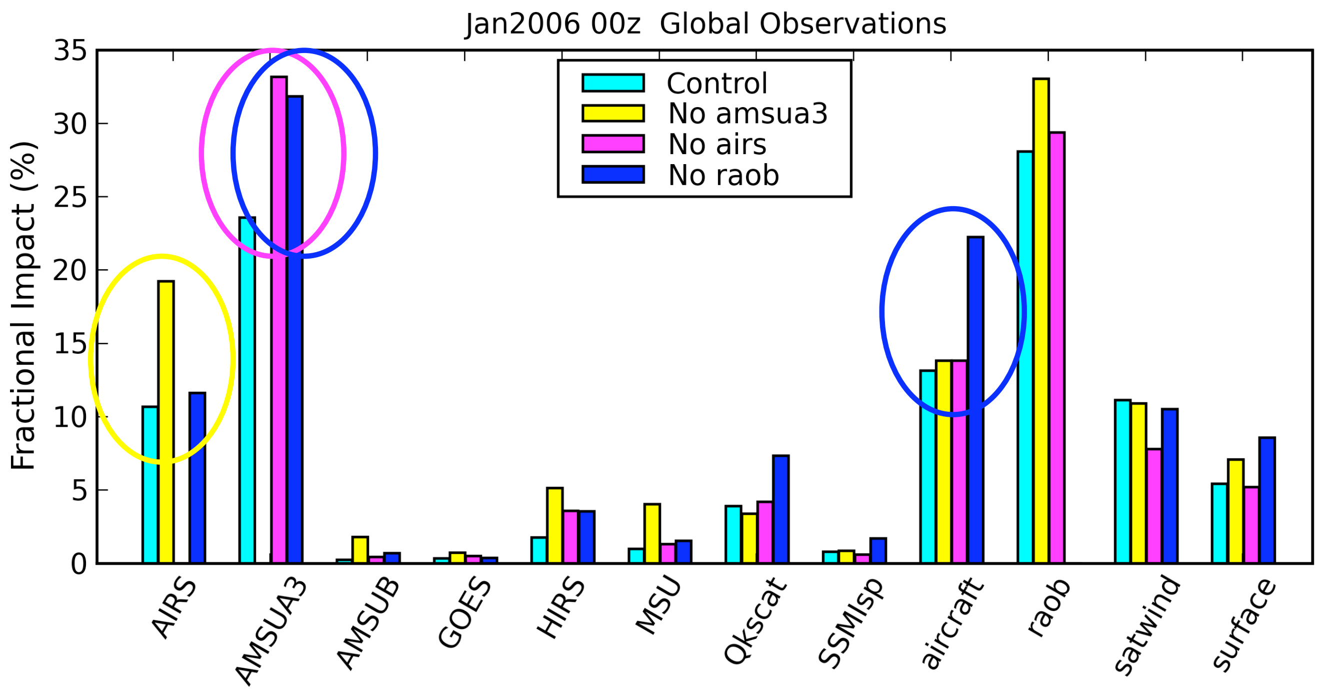
- We are also interested in dependencies and redundancies between observing systems as observations are added or removed …inform current data selection, future data needs

- Such information is implicitly available in an OSE in terms of the responses of the remaining observing systems when a given set of observations is removed

- These responses can be measured through the combined use of OSEs and ADJs, by applying the ADJ to the perturbed (vs. only the control) members of an OSE

Combined Use of ADJ and OSEs

ADJ applied to perturbed OSE members to examine how changing the mix of observations influences their impacts

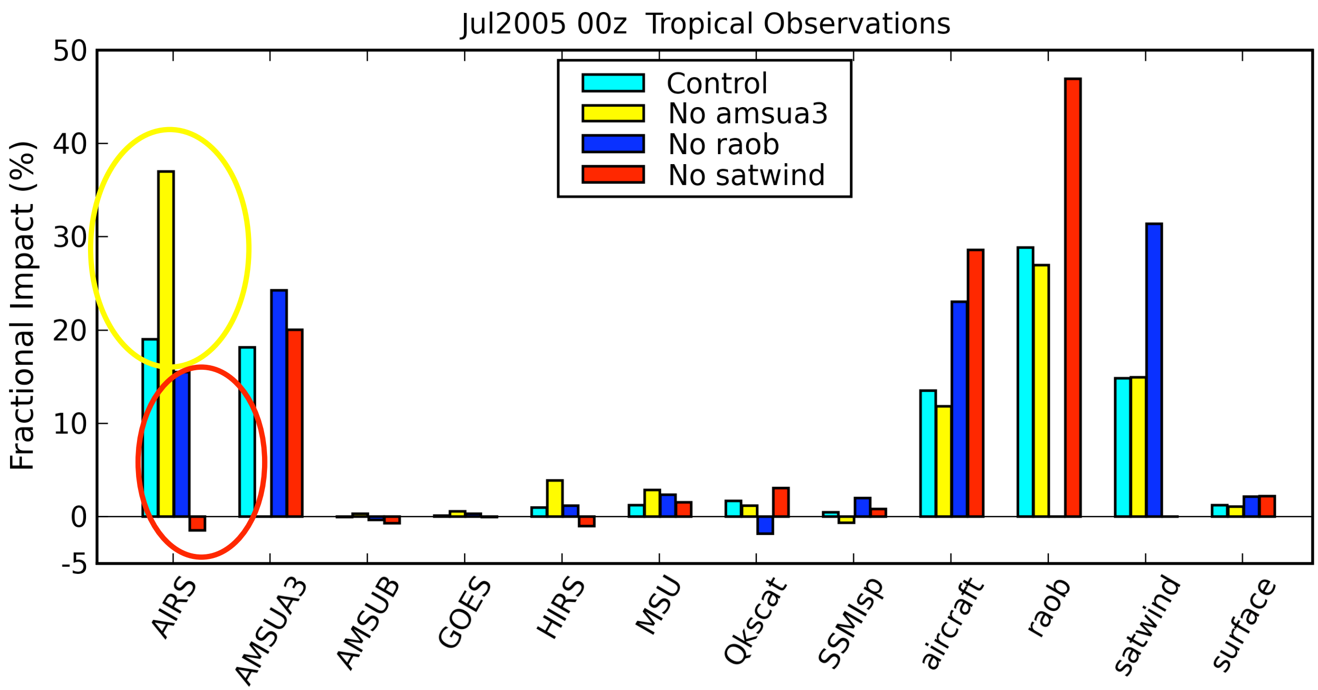


Gelaro and Zhu (2009)

- Removal of AMSUA results in large increase in AIRS (and other) impacts
- Removal of AIRS results in significant increase in AMSUA impact
- Removal of raobs results in significant increase in AMSUA, aircraft and other impacts (but not AIRS)

Combined Use of ADJ and OSEs

ADJ applied to perturbed OSE members to examine how changing the mix of observations influences their impacts



Gelaro and Zhu (2009)

- Removal of AMSUA results in large increase in AIRS impact in tropics
- **Removal of wind observations** results in significant **decrease** in AIRS impact in tropics (in fact, AIRS degrades forecast without satwinds!)

Conclusions on the Complementarity of ADJ and OSE

- Despite fundamental differences in how impact is measured, ADJ and OSE methods provide comparable estimates of the overall ‘value’ of <u>most</u> observing systems

- Differences in OSE and ADJ results should be expected and do not point to shortcomings in either:
  - ✓ different treatment of background information
  - ✓ removal of whole observing systems that contribute disproportionately to analysis quality (AMSU-A)

- Information gleaned from OSEs and ADJs should be viewed as complementary; ADJ extends, not replaces, OSEs:
  - ✓ applicable forecast range, metrics differ
  - ✓ ADJ well suited for routine monitoring

- The combined use of ADJs and OSEs illuminates the complex, complementary nature of how observations are used by the assimilation system

References Cited in this Talk

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

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

Gelaro, R. and Zhu Y. 2009. Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus* **61A,** 179–193*.*

Gelaro, R., Zhu,Y. and Errico, R.M. 2007. Examination of various-order adjoint-based approximations of observation impact. *Meteorologische Zeitschrift* **16,** 685–692*.*

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

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

Trémolet, Y. 2007. First-order and higher-order approximations of observation impact. *Meteorologische Zeitschrift* **16,** 693–694.