

A Practical Method to Estimate Information Content in the Context of 4D-Var Data Assimilation. II: Application to Global Ozone Assimilation

K. Singh¹, M. Jardak¹, A. Sandu¹, K. W. Bowman², and M. Lee²

¹Department of Computer Science, Virginia Polytechnic Institute and State University, 2202 Kraft Drive, Blacksburg, VA 24060, USA

²Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Correspondence to: Adrian Sandu (sandu@cs.vt.edu)

Abstract.

Data assimilation obtains improved estimates of the state of a physical system by combining imperfect model results with sparse and noisy observations of reality. Not all observations used in data assimilation are equally valuable. The ability to characterize the usefulness of different data points is important
5 for analyzing the effectiveness of the assimilation system, for data pruning, and for the design of future sensor systems.

In the companion paper (Sandu et al., 2012) we derive an ensemble-based computational procedure to estimate the information content of various observations in the context of 4D-Var. Here we apply this methodology to quantify the signal and degrees of freedom for signal information metrics of satel-
10 lite observations used in a global chemical data assimilation problem with the GEOS-Chem chemical transport model. The assimilation of a subset of data points characterized by the highest information content yields an analysis comparable in quality with the one obtained using the entire data set.

1 Introduction

The information content of observations in the context of data assimilation is defined by their con-
15 tribution to decreasing the uncertainty in the state estimate (Fisher, 1922). In the companion paper (Sandu et al., 2012) we show that the posterior statistics of the variational cost function and its gradient

can be used to quantify the information content of observations in the context of four dimensional variational (4D-Var) data assimilation. Several information metrics are considered in (Sandu et al., 2012): the Fisher information matrix, the Shannon information, and the degrees of freedom for signal, which measure the decrease in error variance, and the signal information, which measures the effects of data assimilation in terms of adjusting the mean. An efficient computational approach is developed to estimate the information metrics using ensemble averages.

While the information theoretic approach discussed in (Sandu et al., 2012) is general, our application of interest is chemical data assimilation (Carmichael et al., 2008) involving gas phase (Daescu et al., 2000; Carmichael et al., 2003; Constantinescu et al., 2007; Liao et al., 2006) and particulate phase (Sandu et al., 2005; Hakami et al., 2005; Henze et al., 2004) atmospheric tracers. Examples of large scale applications are discussed in (Chai et al., 2006, 2007). Ensemble Kalman filters have also been used in chemical data assimilation (Constantinescu et al., 2007a,b,c; Sandu et al., 2005).

In this paper we consider the problem of global ozone data assimilation using the GEOS-Chem model, and satellite observations from the Tropospheric Emission Spectrometer. We use the new methodology to estimate the signal and the degrees of freedom for signal contents of different ozone column retrievals. The assimilation of a subset of data points characterized by the highest information content gives analyses that are comparable in quality with the one obtained using the entire data set.

The paper is organized as follows. Section 2 reviews a computationally feasible estimation technique for information content developed in the companion paper (Sandu et al., 2012). All estimates require the ability to compute expected values with respect to the analysis probability distribution; obtaining such expected values is discussed in Section 3. Section 4 presents in detail the setup of the global data assimilation experiment, and Section 5 shows the information content of TES ozone column retrievals results obtained with the proposed technique. Section 6 summarizes the findings of this work and points to future research directions.

2 Information Metrics and Their Estimation

The 4D-Var analysis \mathbf{x}_0^A is the initial condition which minimizes the cost function

$$\mathcal{J}(\mathbf{x}_0) = \underbrace{\frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}_0^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B)}_{\mathcal{J}^B(\mathbf{x}_0)} + \underbrace{\frac{1}{2} \sum_{i=1}^N (\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1} (\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i)}_{\sum_{i=1}^N \mathcal{J}_i^{\text{obs}}(\mathbf{x}_0)} \quad (1)$$

subject to the model equation constraints. Here $\mathbf{x}_0^B \in \mathbb{R}^n$ is the background value of the initial state, \mathbf{B}_0 is the covariance of the initial background errors \mathbf{B}_0 , $\mathbf{y}_i \in \mathbb{R}^m$ are the observations at time t_i , $i = 1, \dots, N$, and \mathbf{R}_i are the corresponding observation error covariances.

45 The *information content* of the observations \mathbf{y} quantifies the decrease in uncertainty from before data assimilation (\mathcal{P}^B) to after data assimilation (\mathcal{P}^A). The information content depends not only on the data (\mathbf{y}_i), but also on the data accuracy (\mathbf{R}_i^{-1}), on all other observations $\mathbf{y}_j, j \neq i$, on the background uncertainty (\mathbf{B}_0^{-1}), and on the model dynamics \mathcal{M} .

In the companion paper (Sandu et al., 2012) we have developed a computationally feasible methodology
 50 way to estimate the information content of various observations in the context of 4D-Var. This approach is based on an approximate sampling from the posterior error distribution in 4D-Var. We assume that we have the ability to compute expected values with respect to the posterior density $\mathbb{E}^A[f(\mathbf{x}_0)]$.

2.1 Degrees of freedom for signal

The Degrees of freedom for signal (DFS) metric for the information content has been previously em-
 55 ployed in meteorological data assimilation (Rodgers, 1996; Fisher, 2003; Cardinali et al., 2004; Zupanski et al., 2007; Stewart et al., 2008).

The degrees of freedom for signal (DFS) measures the total reduction in variance after assimilation. For Gaussian background and analysis distributions

$$\mathcal{I}^{\text{DFS}} = n - \text{trace}(\mathbf{B}_0^{-1} \mathbf{A}_0) \quad (2)$$

where \mathbf{A}_0 denotes the analysis covariance matrix at the initial time, and n is the dimension of the system (the number of variables in the model).

It is shown in the companion paper (Sandu et al., 2012) that the contribution of each observation \mathbf{y}_ℓ to the DFS information metric can be estimated via

$$\mathcal{I}_{\mathbf{y}_\ell}^{\text{DFS}} = 2\mathbb{E}^A[\mathcal{J}_\ell^{\text{obs}}(\mathbf{x}_0)] - 2\mathcal{J}_\ell^{\text{obs}}(\mathbf{x}_0^A) \quad (3)$$

where \mathbb{E}^A represents the expected value with respect to the posterior distribution. After the data
 60 assimilation has been performed, one runs the forward model N_{ens} times. The ensemble average of the cost function, minus the cost function at the analysis, estimates the DFS information.

2.2 Signal information

The *signal part* of the relative entropy (Xu, 2006)

$$\mathcal{I}^{\text{Signal}} = \frac{1}{2} (\mathbf{x}_0^A - \mathbf{x}_0^B)^T \mathbf{B}_0^{-1} (\mathbf{x}_0^A - \mathbf{x}_0^B) \quad (4)$$

measures the reduction of uncertainty due to change in the best estimate from the background state to the analysis state. As described in the companion paper (Sandu et al., 2012), the contribution of the

data point \mathbf{y}_ℓ to the signal information can be (coarsely) approximated as:

$$\mathcal{I}_{\mathbf{y}_\ell}^{\text{Signal}} \approx \left(\mathbf{y}_\ell - \mathcal{H}_\ell(\mathbf{x}_\ell^B) \right)^T \mathbf{R}_\ell^{-1} \left(\mathcal{H}_\ell(\mathbf{x}_\ell^A) - \mathcal{H}_\ell(\mathbf{x}_\ell^B) \right) \quad (5)$$

The model is run from the analysis and the ‘‘synthetic observations’’ $\mathcal{H}_\ell(\mathbf{x}_\ell^A)$ are recorded. The model is run again starting from the background state, and (5) is evaluated for each data point \mathbf{y}_ℓ .

65 3 Expected Values With Respect to the Analysis Probability Density

The information metric estimates discussed in (Sandu et al., 2012) require expected values with respect to the analysis probability distribution. Since 4D-Var does not provide immediately an approximation of the posterior density, a discussion of how to obtain these expected values is necessary.

3.1 Expected values as weighted background ensemble averages

Consider the following sample from the background distribution:

$$\mathbf{x}_0^q \in \mathcal{P}^B(\mathbf{x}_0), \quad q = 1, \dots, N_{\text{ens}}$$

The drawing is such that each sample has an equal weight $1/N_{\text{ens}}$. We can approximate expected values with respect to the posterior density by weighted ensemble averages as follows (Wikle and Berliner, 2007):

$$\mathbb{E}^A[f(\mathbf{x}_0)] \approx \sum_{q=1}^{N_{\text{ens}}} \mathcal{P}^A(\mathbf{x}_0^q) f(\mathbf{x}_0^q) = \sum_{q=1}^{N_{\text{ens}}} \frac{\mathcal{P}^A(\mathbf{x}_0^q)}{\mathcal{P}^B(\mathbf{x}_0^q)} \mathcal{P}^B(\mathbf{x}_0^q) f(\mathbf{x}_0^q) = \sum_{q=1}^{N_{\text{ens}}} w^q f(\mathbf{x}_0^q)$$

The posterior average can be calculated as a weighted average of samples taken from the background distribution. Using the Bayes’ theorem, the new weights are:

$$w^q = \frac{\mathcal{P}^A(\mathbf{x}_0^q)}{\mathcal{P}^B(\mathbf{x}_0^q)} \frac{1}{N_{\text{ens}}} = \frac{\mathcal{P}(\mathbf{y}|\mathbf{x}_0^q)}{\mathcal{P}(\mathbf{y})} \frac{1}{N_{\text{ens}}}$$

With the relationship that the observation part of the cost function is the logarithm of the observation likelihood (Sandu et al., 2012), we can compute the weights as:

$$v^i = \exp\left(\mathcal{J}^{\text{obs}}(\mathbf{x}_0^i)\right), \quad w^q = \frac{v^q}{\sum_{i=1}^{N_{\text{ens}}} v^i}$$

The computational procedure is as follows. Start with an equally weighted sample \mathbf{x}_0^q of the background probability density. For each sample run the model, and compute the observations part of the

4D-Var cost function $\mathcal{J}(x_0^q)$, as well as the metric of interest $\mathcal{J}(x_0^q)$. The analysis mean is a weighted average of the obtained values:

$$\mathbb{E}^A [f(\mathbf{x}_0)] = \sum_{q=1}^{N_{\text{ens}}} \left(\frac{\exp(\mathcal{J}^{\text{obs}}(\mathbf{x}_0^q))}{\sum_{i=1}^{N_{\text{ens}}} \exp(\mathcal{J}^{\text{obs}}(\mathbf{x}_0^i))} \right) f(\mathbf{x}_0^q)$$

70 3.2 Approximate posterior sampling by estimated posterior covariance

In this simple approach one assumes that the correlation structures of \mathbf{B}_0 and \mathbf{A}_0 are similar, and that the difference comes from changes in variances. The background and analysis variances can be estimated roughly by comparing the background and analysis against data, and by measuring the model-data discrepancies. It is then assumed that the decrease in the model-data discrepancy for each
75 variable, vertical level, area, etc. is representative of the corresponding decrease in variance. The analysis variances are obtained by rescaling the background variances (for each variable, vertical level, area, etc.). These variances, together with the specified correlation structure of \mathbf{B}_0 , define the analysis covariance \mathbf{A}_0 .

Random draws are taken from the normal distribution $\mathbf{x}_0^r \in \mathcal{N}(\mathbf{x}_0^A, \mathbf{A}_0)$, $r = 1, \dots, N_{\text{ens}}$. The expected values with respect to the posterior density are estimated by the ensemble averages:

$$\mathbb{E}^A [f(\mathbf{x}_0)] \approx \frac{1}{N_{\text{ens}}} \sum_{r=1}^{N_{\text{ens}}} f(\mathbf{x}_0^r)$$

3.3 Approximate posterior sampling by principal components

This approach uses the fact that the analysis covariance matrix is approximated by the inverse Hessian of the cost function, evaluated at the optimum (Thacker, 1989; Gejadze et al., 2008)

$$\mathbf{A}_0 \approx \left(\nabla_{\mathbf{x}_0, \mathbf{x}_0}^2 \mathcal{J} \right)^{-1}$$

80 Several eigenvectors corresponding to the smallest eigenvalues of the Hessian are computed. The inverses of these eigenvalues, together with their eigenvectors, approximate the principal components of the posterior error and can be used for approximate sampling from the posterior distribution. The computation of the smallest eigenpairs of the Hessian can be performed by a Lanczos procedure using only Hessian vector products, obtained via a second order adjoint model (Sandu and Zhang, 2008).

85 3.4 Approximate posterior sampling by subspace analysis

The hybrid of 4D-Var and ensemble approach discussed in (Cheng et al., 2010) is able to sample the posterior distribution. Suppose we are given the background state $\mathbf{x}_0^B \in \mathbb{R}^n$ and an ensemble of

perturbation vectors drawn from the background error distribution

$$\left(\Delta\mathbf{x}_0^{\text{B}}\right)^r \in \mathcal{N}(\mathbf{0}, \mathbf{B}_0), \quad r = 1, 2, \dots, N_{\text{ens}} \quad (6)$$

Starting from \mathbf{x}_0^{B} , we save the first k iterates $\mathbf{x}_0^{(j)}, j = 1, \dots, k$, generated by the numerical optimization routine used in the 4D-Var assimilation. We also save the corresponding gradients $\nabla \mathcal{J}(\mathbf{x}_0^{(j)}), j = 1, \dots, k$.

Let \mathbf{U} be the subspace spanned by successive increments

$$\mathbf{U} = \text{span} \left\{ \mathbf{x}_0^{(1)} - \mathbf{x}_0^{(0)}, \dots, \mathbf{x}_0^{(k)} - \mathbf{x}_0^{(k-1)} \right\}$$

In essence, 4D-Var reduces the error components in \mathbf{U} , and leaves unmodified the error components in the orthogonal complement of \mathbf{U} .

- 90 Compute the orthogonal projector \mathbf{P} onto the orthogonal complement of \mathbf{U} (the multiplication $\mathbf{P} \cdot \Delta\mathbf{x}_0$ removes the components of $\Delta\mathbf{x}_0$ in the subspace \mathbf{U}). From the saved sequence of gradients a low rank quasi-Newton approximation of the inverse Hessian H_{QN}^{-1} is constructed; this approximates the posterior covariance in the subspace \mathbf{U} .

The analysis perturbations are obtained by removing the background error components in \mathbf{U} , and replacing them with random draws from the approximate analysis error distribution in \mathbf{U}

$$\left(\Delta\mathbf{x}_0^{\text{A}}\right)^r = \mathbf{P} \cdot \left(\Delta\mathbf{x}_0^{\text{B}}\right)^r + H_{\text{QN}}^{-1/2} \zeta^r, \quad \zeta^r \in \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad r = 1, 2, \dots, N_{\text{ens}} \quad (7)$$

4 Application to Data Assimilation of Global Ozone

- 95 We apply the estimation methodology to a 4D-Var data assimilation study with a global chemical transport model. The data assimilation experiment focuses on ozone. Ozone is an important constituent of stratosphere which absorbs the high energy UV-B and UV-C rays, thus preventing the disintegration of DNA molecules and supporting the existence of life. However, ozone present in mid to low troposphere is a pollutant, a powerful oxidizing agent leading to destruction of tissues, damaging fibers and
100 creating breathing problems.

The data are satellite ozone column retrievals. We estimate the information content of satellite observations taken at different times using different information theoretic metrics.

4.1 The model: GEOS-Chem

- The model used for the numerical experiments in this paper is GEOS-Chem (<http://acmg.seas.harvard.edu/geos>), a global three-dimensional chemical transport model (CTM) driven by assimilated meteorological observations from Goddard Earth Observing System. A detailed description of the model
105

is presented in (Bey et al., 2001). GEOS-Chem accounts in detail for emissions from both natural and anthropogenic sources, for gas phase chemistry, aerosol processes, long range transport of pollutants, troposphere-stratosphere exchanges, etc. GEOS-Chem is being widely used world-wide for global atmospheric chemistry studies. We use GEOS-Chem v7-04-10. Subsequent model releases and references can be found at <http://geos-chem.org>.

The GEOS-Chem-Adjoint system (http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint) has been developed through a joint effort of groups at Caltech, University of Colorado, Virginia Tech, Harvard, and Jet Propulsion Laboratory (Henze et al., 2007; Singh, 2009a,b; Eller et al., 2009). The system can perform adjoint sensitivity analyses and 4D-Var chemical data assimilation. Inverse modelling studies with GEOS-Chem-Adjoint are exemplified in (Henze et al., 2009; Kopacz et al., 2007; Zhang et al., 2009).

4.2 The data: Tropospheric Emission Spectrometer (TES) ozone column retrievals

We assimilate ozone profile retrievals from the Tropospheric Emission Spectrometer (TES), in order to obtain improved estimates of the ozone initial conditions. TES (Beer et al., 2001), one of four science instruments aboard NASA's Aura satellite, measures the infrared-light energy (radiance) emitted by Earth's surface, and by the chemical tracers in the atmosphere (<http://tes.jpl.nasa.gov>). Vertical profiles of chemical concentrations are retrieved from the radiance measurements using an off-line inversion process.

A-priori information about the vertical concentration profile of the species of interest is needed to solve the retrieval inverse problem (the prior information does not come from the measurement). Let $\mathbf{x}^{\text{prior}}$ be the prior vertical ozone concentration profile (in volume mixing ratio units), and let $\mathbf{z}^{\text{prior}} = \ln \mathbf{x}^{\text{prior}}$. Let $\mathbf{x}^{\text{radiance}}$ be the atmospheric profile as resulting directly from the radiances and $\mathbf{z}^{\text{radiance}} = \ln \mathbf{x}^{\text{radiance}}$.

The vertical ozone profile is retrieved according to the formula (Parrington et al., 2009)

$$\hat{\mathbf{z}} = \mathbf{z}^{\text{prior}} + A_v \left(\mathbf{z}^{\text{radiance}} - \mathbf{z}^{\text{prior}} \right) + G \eta, \quad \hat{\mathbf{x}} = \exp(\hat{\mathbf{z}}) \quad (8)$$

Here A_v is the averaging kernel matrix, G is the gain matrix, and η is the spectral measurement error (assumed to have mean zero and covariance S_η). More details can be found in (Worden et al., 2004; Jones et al., 2003; Bowman et al., 2002).

The corresponding TES observation operator is linear with respect to the logarithm of the concentrations, but nonlinear with respect to the concentration profile:

$$\mathcal{H}(\mathbf{x}) = \mathbf{z}^{\text{prior}} + A_v \left(\ln(L(\mathbf{x})) - \mathbf{z}^{\text{prior}} \right)$$

The ozone column \mathbf{x} represented on the N_{lev} GEOS-Chem grid vertical levels is interpolated by the

operator L to an ozone column $L(\mathbf{x})$ represented on the p TES profile retrieval levels.

For this reason several chemical data assimilation studies based on TES retrieved profiles (Jones et al., 2003; Bowman et al., 2006; Parrington et al., 2009) have opted to perform the suboptimal Kalman filtering step in the logarithm of the concentrations:

$$\ln \mathbf{x}^A = \ln \mathbf{x}^f + K \left(\hat{\mathbf{z}} - \mathcal{H}(\mathbf{x}^f) \right)$$

In case of 4D-Var data assimilation, the forcing calculation is carried out in the model state space. For this reason an adjoint of the observation operator needs to be derived explicitly to update the gradients.

$$\begin{aligned} (\mathcal{H}'(\mathbf{x}))^T \cdot v &= \left(\frac{\partial}{\partial \mathbf{x}} (A_v \ln(L(\mathbf{x}))) \right)^T \cdot v \\ &= \left(\frac{\partial L}{\partial \mathbf{x}} \right)^T \cdot \begin{pmatrix} \frac{1}{[L(\mathbf{x})]_0} & 0 & \cdots & 0 \\ 0 & \frac{1}{[L(\mathbf{x})]_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{[L(\mathbf{x})]_p} \end{pmatrix} \cdot A_v^T \cdot v \end{aligned}$$

Here $(\mathcal{H}'(\mathbf{x}))^T$ is a matrix and $v = \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y})$. The product $A_v \cdot v$ is scaled by the diagonal matrix with the i -th diagonal entry $1/[L(\mathbf{x})]_i$. The result is fed to $(\partial L/\partial \mathbf{x})^T$, the adjoint of the interpolation operator, which entities from TES profile retrieval domain back to the GEOS-Chem model domain.

4.3 The validation data: INTEX ozonesonde profiles

In order to assess the quality of the data assimilation results, we compare the respective analyses against an independent data set. The independent data are the ozonesonde profiles measured during the INTEX Ozonesonde Network Study 2006 (IONS-6) (<http://croc.gsfc.nasa.gov/intexb/ions06.html>, (Thompson et al., 2007a, 2007b)) for the month of August. There were 418 ozonesondes launched from 22 stations across North America. A detailed description of the number of ozonesondes launched per station with longitude and latitude information can be found in (Parrington et al., 2008).

We use ozonesonde parameters such as launch time, longitude, latitude and pressure level to interpolate the concentration fields generated by the model. Differences between the ozone concentrations from ozonesonde observations, model forecasts, and model analyses are averaged individually over longitude, latitude and time to create vertical profiles of model errors. We report the vertical distribution of the mean and the standard deviation of model errors.

4.4 Experimental Setting

The GEOS-Chem simulations are carried out at a resolution of $4^\circ \times 5^\circ$. At this resolution, each latitude-longitude grid box on the ground level covers an area of about $400\text{Km} \times 500\text{Km}$. The chemical system accounts for 43 different chemical species. The dimension of the state space in our simulations is $n \approx 8$ million (72 longitude grid points, times 46 latitude grid points, times 55 vertical levels, times 43 chemical tracers).

The control variables are the initial concentrations of ozone throughout the simulation domain. While GEOS-Chem is capable of performing simulations up to 75 Km (55 vertical levels), the model error increases with height and the model bias is non-negligible in the upper troposphere and into the stratosphere. For this reason we perform data assimilation only up to 21 Km (the bottom 23 vertical levels). The dimension of the control vector for data assimilation is $n_c \approx 80,000$ (72 longitude grid points, times 46 latitude grid points, times 23 vertical levels, times 1 chemical tracer – ozone).

The assimilation time window has a length of 5 days, starting on August 1st, 2006 (00 GMT) and ending on August 6th, 2006 (00 GMT). The observation time window is 4 hours, i.e., the observation operator treats all retrievals available in a 4 hour window as a single data point. Specifically, the observation y_i at time t_i consists of all the data available for the time interval $[t_i - 2\text{hours}, t_i + 2\text{hours}]$. There were a total of 85011 unique individual observation locations during this 5-day interval.

We estimate the information content of ozone profile retrievals from TES when used to improve the ozone initial conditions in GEOS-Chem through 4D-Var data assimilation. The main computational costs come from: (1) the 4D-Var run, which requires 11 iterations of the optimization routine, with each iteration performing a forward and adjoint model run; and (2) an ensemble of 20 additional model runs, including adjoints, to gather the data needed for the estimation of different information content metrics. Concentrations and other time dependent variables are checkpointed during the forward runs, and are read during the adjoint runs. The adjoint forcing calculations are performed every observation window (4 hours). The numerical optimization method is the limited memory bound-constrained BFGS method (Zhu et al., 1997), which has become the "gold standard" in solving large scale 4D-Var chemical data assimilation problems (Sandu et al., 2005). The total computational time is 14 minutes and 46 seconds per forward plus adjoint model runs. All the simulations are parallel and use eight cores; they were performed on a Dell Precision T5400 workstation with 2 quadcore Intel(R) Xeon(R) processors with clock speed 2.33GHz and a RAM of 16GB shared between the eight cores.

We consider a diagonal background error covariance matrix (\mathbf{B}_0) in all our experiments for simplicity. The setting can be easily extended to use a non-diagonal \mathbf{B}_0 that captures spatial error correlations (Singh et al., 2010). The initial variances (the diagonal entries of the \mathbf{B}_0 matrix) are constructed

from the average background concentrations \mathbf{x}_0^B on each of the N_{lev} vertical layers

$$\mathbf{B}_0 = \begin{bmatrix} \mathbf{B}_0^{(0)} & 0 \dots & 0 \\ 0 & \mathbf{B}_0^{(1)} \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \dots & \mathbf{B}_0^{(N_{lev})} \end{bmatrix} \quad (9)$$

where

$$\mathbf{B}_0^{(\ell)} = \begin{bmatrix} \sigma_\ell^2 & 0 \dots & 0 \\ 0 & \sigma_\ell^2 \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \dots & \sigma_\ell^2 \end{bmatrix}_{dim \times dim}, \quad dim = N_{lon} \cdot N_{lat}, \quad (10)$$

with

$$\sigma_\ell = \frac{rel}{dim} \sum_{i=1}^{N_{lon}} \sum_{j=1}^{N_{lat}} \mathbf{x}_0^B(i, j, \ell, s_{O3}), \quad \ell = 1, \dots, N_{lev}, \quad s_{O3} = \text{index of ozone} \quad (11)$$

The relative uncertainty level in the background initial conditions is taken to be 50%, i.e., $rel = 0.5$.

The following simple technique is employed to approximately sample the analysis distribution. We
 185 perform data assimilation and compare the background and the analysis fields against the INTEX
 ozonesonde validation data set. This provides a vertical distribution of mean errors and of their vari-
 ance. We make the following assumptions: the analysis covariance matrix is diagonal (the correlation
 length is smaller than one grid size); the relative error reduction realized through data assimilation is
 similar in all gridpoints at the same vertical level; and the relative error reduction is similar throughout
 190 the assimilation window. Under these assumptions the error reduction measured against the INTEX
 ozonesonde data is representative of the reduction in error at the initial time throughout the entire
 computational grid. Consequently, the analysis error standard deviation at a given grid point is ob-
 tained by scaling the background standard deviation. The scaling factor is the ratio of the standard
 deviation of the analysis against INTEX data over the standard deviation of the background against
 195 INTEX data; the same scaling factor is applied to all grids at the same vertical level. In summary,
 the analysis mean is provided by the result of the data assimilation. The analysis covariance matrix is
 diagonal, with the diagonal entries obtained by scaling the corresponding background variances. The
 scaling factors are obtained by comparing the background and the analysis against the validation data
 set.

200 5 Information content of TES ozone column retrievals

We exhibit three different sets of results that provide estimates of the information content of individual
 observation locations in the context of 4D-Var data assimilation. Essentially, we demonstrate assimi-

tion results using subsets of observation locations, with high and low DFS information content, high and low signal information content, and high information content common to both signal and DFS
205 metrics.

5.0.1 The DFS information content

Loosely speaking, the DFS metric (discussed in Section 2.1) indicates the number of states that benefit from the assimilation of observations. The closer the \mathcal{I}^{DFS} is to the total number of model states n , the more information the observations have brought into the system through data assimilation. Tehnically,
210 the DFS measures the relative decrease in the error covariance through assimilation.

The DFS information content for individual data points \mathbf{y}_i is estimated using equation (3). Here one data point \mathbf{y}_i consists of all the ozone retrievals available in the 4 hours interval $[t_i - 2 \text{ hours}, t_i - 2 \text{ hours}]$. As the Aura satellite orbits the Earth, observations are taken over different locations and at different times of the day. It is therefore expected that some data points will contain more information
215 than other, i.e., are more useful in reducing uncertainty when assimilated. We utilize an ensemble of $N_{\text{ens}} + 1$ model runs initialized with states drawn from an approximate analysis distribution, as discussed in Section 3.2. During each run, the cost function contribution of each observation location is saved. These results are used to estimate \mathcal{I}^{DFS} via (3).

Figure 1 presents the DFS acquired from the data in each observation window. The data in the window
220 16:00 GMT - 20:00 GMT, August 3rd, 2006 has the highest information content. The DFS information content decreases with time, and the impact of the observations taken later in the assimilation window is smaller.

We next study the DFS contributions of individual retrievals to the assimilation result. Each column and each vertical level is considered separately: one data point corresponds to one retrieved ozone
225 value in one column, and at one vertical level.

All the data points are ranked according to their estimated DFS information values. We consider the most informative set of observations (those that belong to the top 20%), as well as the least informative set of observations (that forms the bottom 20%). Figure 2 presents the locations of observations within the top and within the bottom twentieth percentiles. Perspective, top, latitudinal, and longitudinal
230 views of these locations are presented. The DFS information content of each observation location has been normalized to a value between 0 and 1. Red points represent locations of observations with high DFS information content, while blue points correspond to locations with a low DFS information content.

Panels (a) and (b) show that the locations with the highest information content are located within

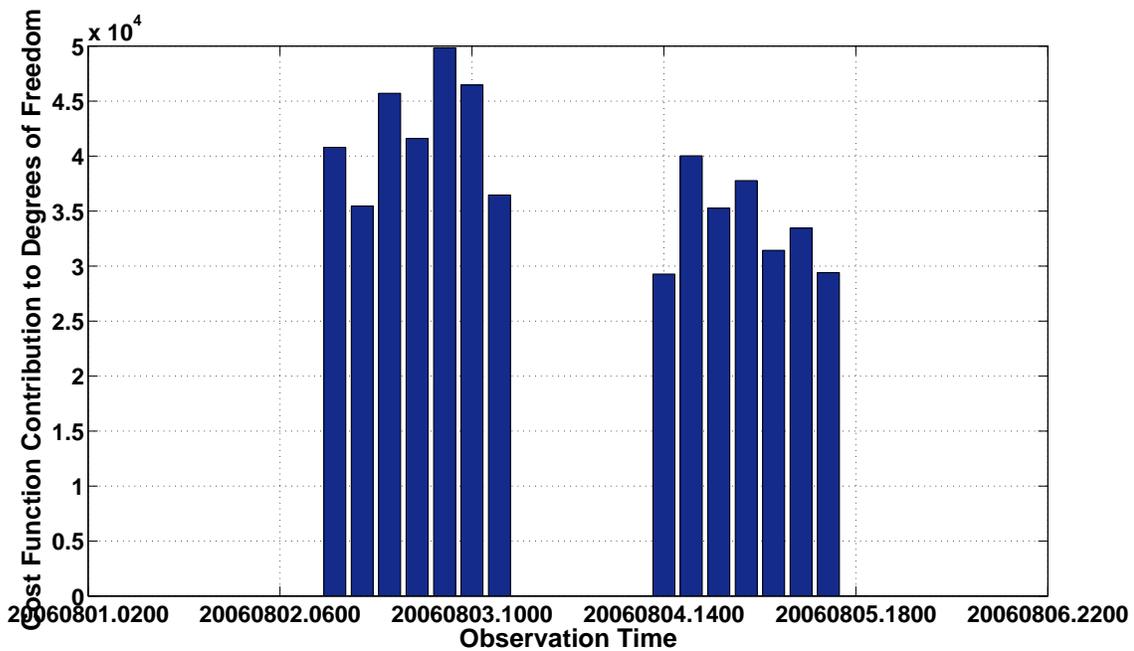


Fig. 1. The DFS information content of observations taken at different times within the 5 days assimilation window starting at 00:00 GMT August 1, 2006, and ending at 00:00 GMT August 6, 2006.

235 model vertical levels 0 to 21 (roughly, below 100 hPa). Vice-versa, the least informative observations are taken above 100 hPa (model vertical levels 22 to 55). As seen in panel (a), among the most informative observations, those located near tropopause (100 hPa) bring the highest DFS. The non-informative observations are evenly distributed along latitudes and longitudes, see panel (b). The horizontal distribution of the high information content locations does not display any localized pattern.

240 Next, we perform additional data assimilation experiments using only subsets of observations. Specifically, in one experiment we utilize only the data points in the top 20%, and in another only the data points in the bottom 20%, filtered by their DFS information content. The results are presented in Figure 3. The assimilation results using only the top 20% most informative observations are almost as accurate as the result using all observations. The quality of the analysis obtained using only the bottom 20% observations is similar to that of free model run. The fact that almost all information is captured by the top 20%, and almost no information is captured by the bottom 20%, suggests that the DFS provides a sharp diagnostic to distinguish between the most and the least important data points.

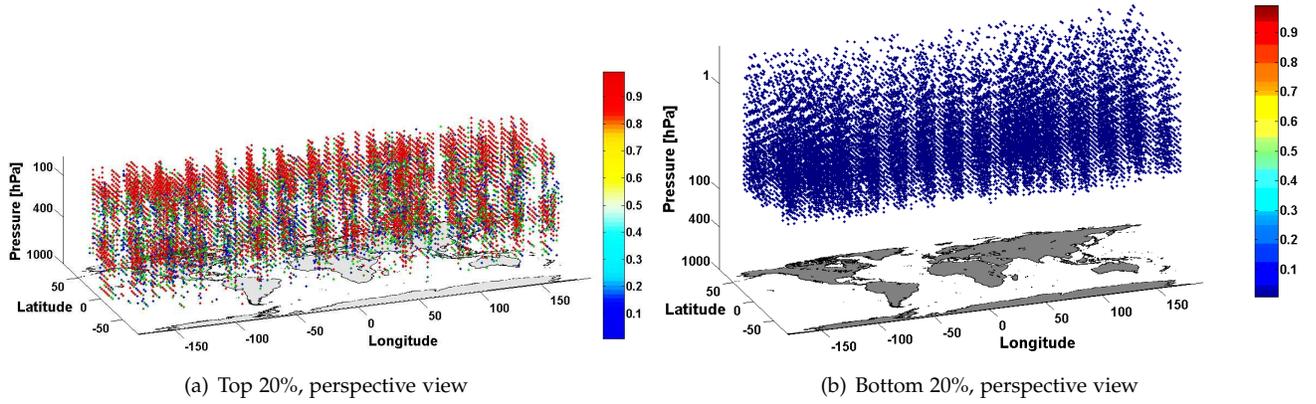


Fig. 2. The location of observations within the top and within the bottom twentieth percentiles, when ranked according to their DFS information content. The DFS contribution of each observation is normalized to a value between 0 to 1.

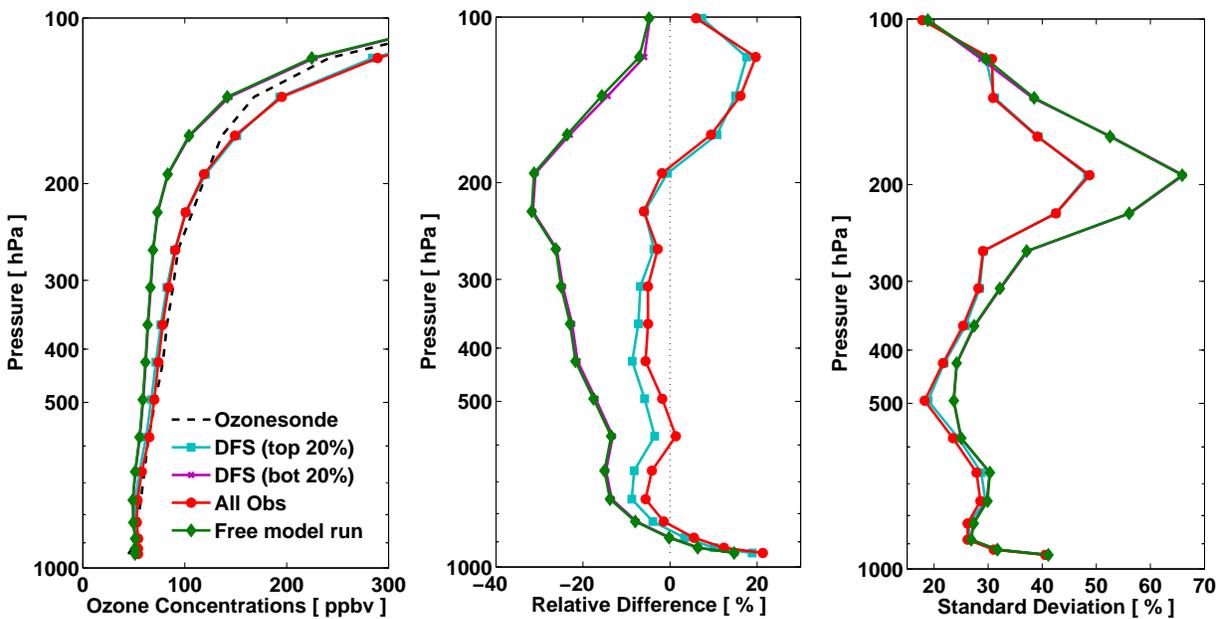


Fig. 3. Plot of ozonesonde data, free model run, and 4D-Var analysis trajectories obtained using subsets of observation locations. The subsets are selected according to their DFS information content.

5.0.2 The Signal information content

The signal information content of individual data points \mathbf{y}_ℓ is estimated using the formula (5). No gradient calculations are necessary. The estimate depends only on the innovation vectors associated with the background trajectory $\mathbf{d}_\ell^B = \mathbf{y}_\ell - \mathcal{H}(\mathbf{x}_\ell^B)$, and with the analysis trajectory $\mathbf{d}_\ell^A = \mathbf{y}_\ell - \mathcal{H}(\mathbf{x}_\ell^A)$. Equation (5) can be written as

$$\mathcal{I}_{\mathbf{y}_\ell}^{\text{Signal}} \approx \left(\mathbf{d}_\ell^B\right)^T \mathbf{R}_\ell^{-1} \left(\mathbf{d}_\ell^B - \mathbf{d}_\ell^A\right) \quad (12)$$

We first perform a forward model run starting with the optimal initial condition \mathbf{x}_0^A and save the innovation vectors \mathbf{d}_ℓ^A for each observation location and for all observation windows. We then perform a second run starting with the background initial condition \mathbf{x}_0^B . During this run we compute the innovation vectors \mathbf{d}_ℓ^B , and, using the saved \mathbf{d}_ℓ^A values, we also compute the Signal information content (12).

The time series of the Signal information content per each observation window is shown in Figure 4. The difference between the contribution of observations taken earlier and taken later during the assimilation window is small. This difference is relatively large for the DFS information metric, as will be seen in Figure 1.

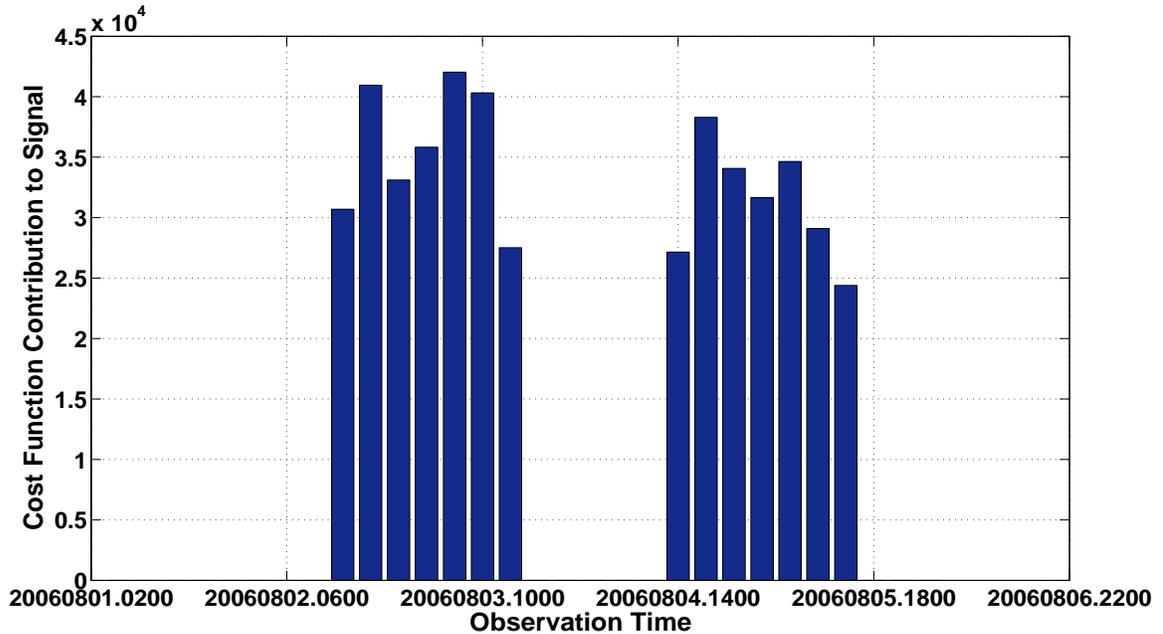


Fig. 4. The Signal information content of observations taken at different times within the assimilation window.

We next relate the signal information content of each observation with its location. This approach reveals the spatial distribution of observations that contribute more information to the data assimilation

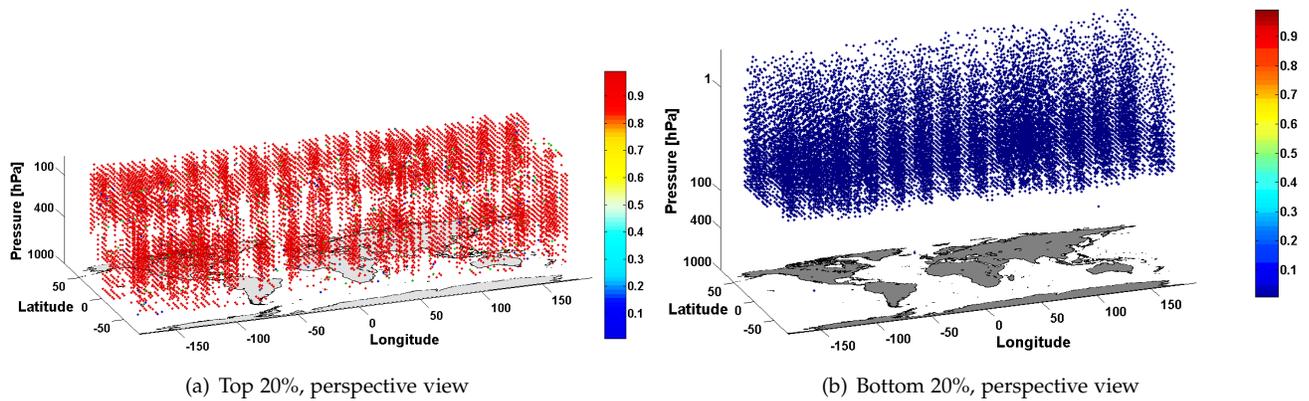


Fig. 5. The location of observations with the highest, and with the lowest signal information content. The colors represent vertical layer numbers of the model.

260 process.

Panels (a) and (b) in Figure 5 present the locations with the highest and lowest signal information content respectively. Similar to DFS, observations with highest information content are located within model vertical levels 0 to 21 (roughly, below 100 hPa), while, the least informative observations are located above 100 hPa (model vertical levels 22 to 55). However, observations within top 20% $T_{y_i}^{\text{Signal}}$ have similar signal information values as opposed to DFS case, where, observations within top 20% had varied range of DFS values.

We now investigate the relationship between the estimated signal information content, and the benefit that the respective observations bring to the 4D-Var data assimilation process. Specifically, we perform a 4D-Var data assimilation using only the subset of observations within the top twentieth percentile, and another 4D-Var data assimilation using only the subset of observations within the bottom twentieth percentile when ranked by their signal information content. All data assimilation experiments use the same covariance matrices and the same background field x_0^B .

Figure 6 presents the results of the different data assimilation experiments. The errors are measured against the independent data set of INTEX Ozone-sonde Network Study 2006 (IONS-6). The leftmost panel presents the mean ozone concentration vertical profiles. The central panel shows the mean errors, i.e., the relative difference between the mean model profiles and ozonesondes. The rightmost panel presents the corresponding error standard deviations. A detailed discussion of the 4D-Var data assimilation results using all the observations is provided in Singh et al. (2010).

The results in Figure 6 reveal that the observations with higher signal information content contribute more to the 4D-Var analysis. The quality of the analysis using only the top 20% of observations is similar to the quality of the analysis using all observations, and the analysis based on the bottom 20%

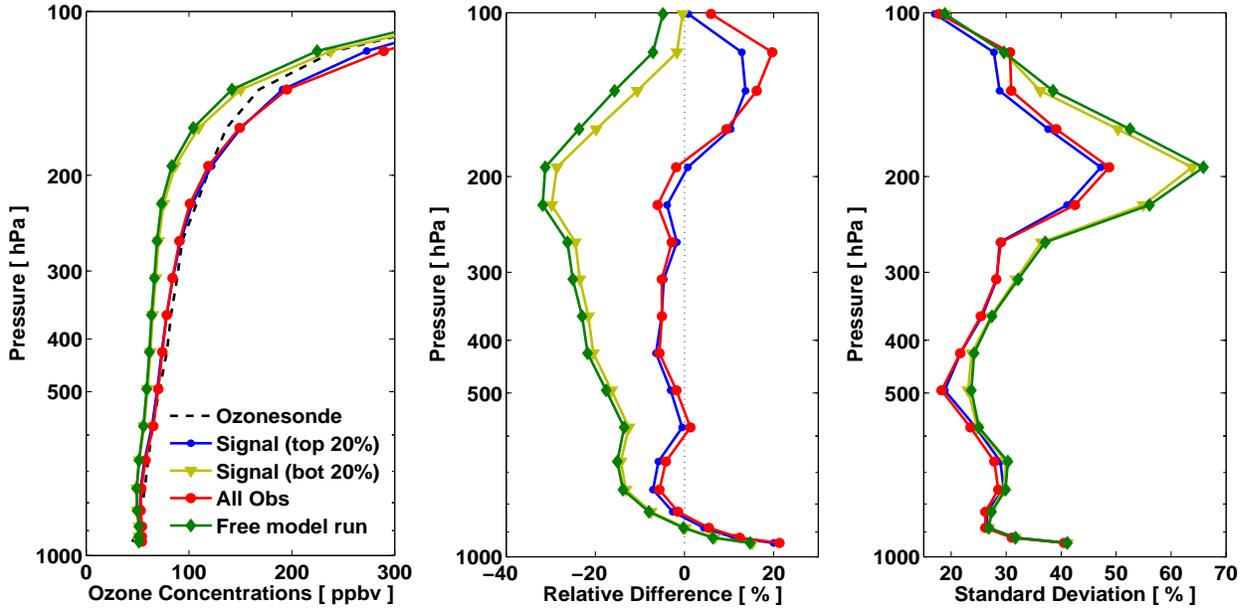


Fig. 6. Plot of ozonesonde data, free model run, and 4D-Var analysis trajectories obtained using subsets of observation points. The subsets are selected according to their signal information content.

of the observations is comparable to the free model run.

5.0.3 Common to signal and DFS information content

As described in the previous section, DFS and signal provide complementary measures of the information content. Therefore, It would be of interest to consider observation points that have high DFS as well as high signal information content. We choose the top 20% of all observation points that rank high on both DFS and signal metrics. In order to come up with such a selection, we arranged the complete set of observation data points in two different arrays, first array with descending DFS information content and second with descending signal information content. Starting from index 17002 (20% of 85011), we move to higher indexes and take the intersection of the grid coordinates of the two arrays until we find 20% observation locations with high DFS and signal contents.

We perform data assimilation using the common top 20% of all observations. Figure 7 compares the quality of the vertical ozone profiles generated by the free model run, the 4D-Var analysis using all observations, the top 20% signal, the top 20% DFS, and the top 20% data points common to both DFS and signal. Results indicate that the data points satisfying combined signal and DFS criterion provide the most accurate analysis overall. The analysis generated using these points follows closely

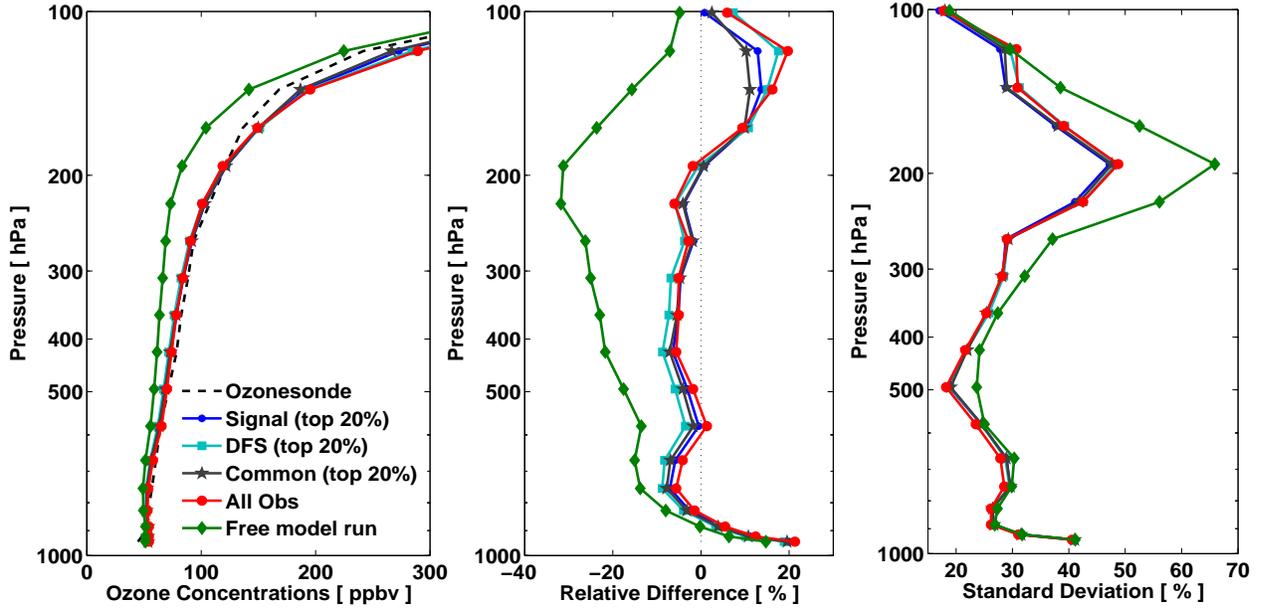


Fig. 7. Plot of ozonesonde data, free model run, and 4D-Var analysis trajectories obtained using subsets of observation points. The subsets are selected according to their information content common to DFS and Signal.

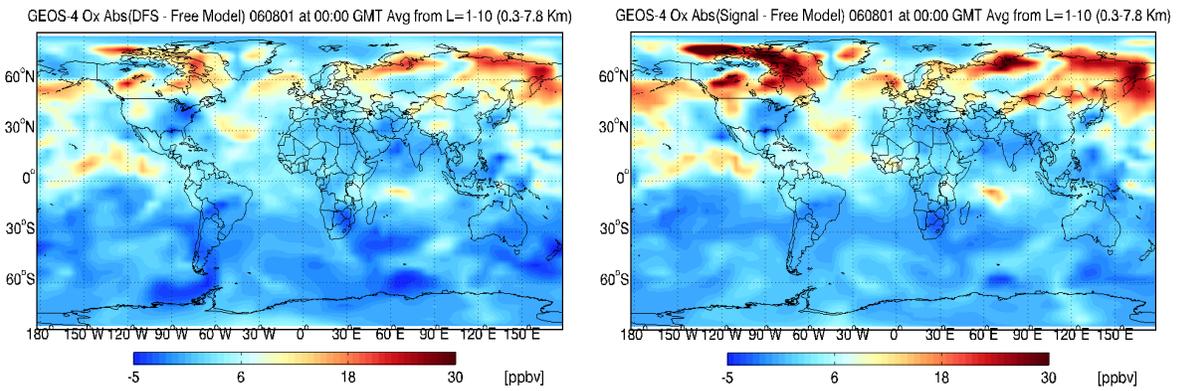
the analysis generated by using full observation data set from ground level up to 300 hPa and is better than other observation data sets in the 100–300 hPa vertical region. This indicates that pruning the least informative data points may actually improve the quality of the overall analysis.

300 A direct comparison of different assimilation results is provided in Figure 8. Specifically, we plot the differences in global ozone concentrations at the beginning of the assimilation window (00:00 GMT on August 6, 2006) averaged over the first 10 GEOS-Chem vertical levels. Panels (a)-(c) show differences between the 4D-Var analysis fields and the model forecast (solution without data assimilation); the analyses use observation data points with top 20% $\mathcal{I}_{y_i}^{\text{DFS}}$, top 20% $\mathcal{I}_{y_i}^{\text{Signal}}$, and top 20% signal and DFS, respectively. Panels (d)-(f) show absolute differences between 4D-Var analyses using all observation data and using only the data within the top 20% $\mathcal{I}_{y_i}^{\text{DFS}}$, top 20% $\mathcal{I}_{y_i}^{\text{Signal}}$ and top 20% $\mathcal{I}_{y_i}^{\text{Common}}$.

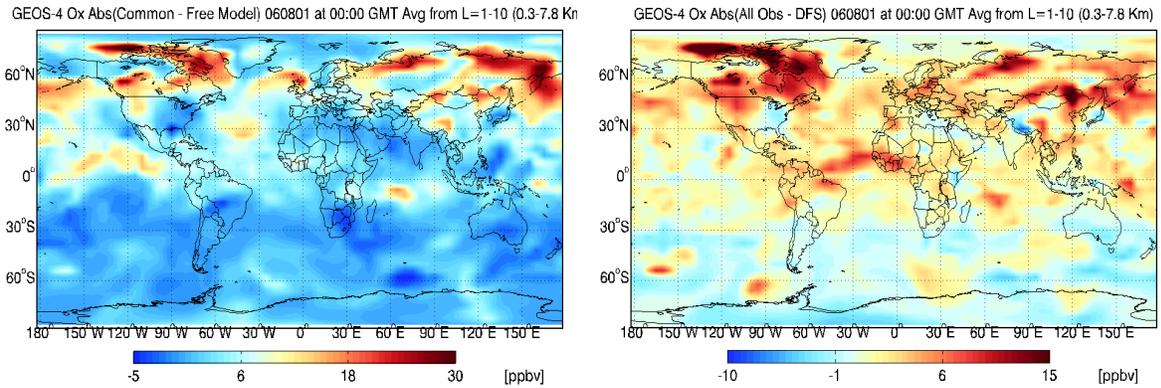
305

Since Figure 8 does not provide any comparison of 4D-Var analysis with real ozone observations, we use Figure 7 as a baseline to assess the results. There is a limitation to this assessment however which is the fact that the IONS-06 ozonesondes data used in Figure 7 are available only for North America (mainly United States). Comparing Figures 7 and 8 reflect that in lower to mid troposphere (up to 10 GEOS-Chem levels, 400 hPa), 4D-Var analysis using all observations is slightly different from analysis using observations with highest DFS information while it is closer to analysis using observations with highest signal information and observations which rank high on both DFS and signal metrics.

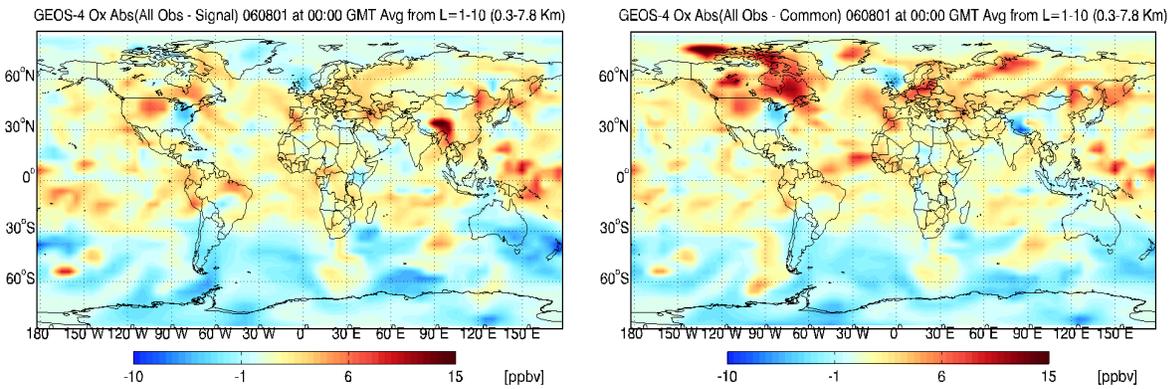
310



(a) Absolute difference between the 4D-Var analysis using data points with top 20% DFS and the free model run (b) Absolute difference between the 4D-Var analysis using data points with top 20% signal and the free model run



(c) Absolute difference between the 4D-Var analysis using data points with top 20% common to signal and DFS and the free observations and data points with top 20% DFS model run (d) Absolute difference between the 4D-Var analysis using all observations and data points with top 20% DFS



(e) Absolute difference between the 4D-Var analysis using all observations and data points with top 20% signal (f) Absolute difference between the 4D-Var analysis using all observations and data points with top 20% common signal and DFS

Fig. 8. Direct comparison of different assimilation results using various subsets of the data. Differences in global ozone concentrations are shown at 00:00 GMT on August 6, 2006 and averaged over the first 10 GEOS-Chem vertical levels.

6 Conclusions

315 This paper discusses a characterization of the information content of observations in the context of
four dimensional variational (4D-Var) data assimilation framework. The ability to characterize the use-
fulness of different data points is important for analyzing the effectiveness of the assimilation system,
for data pruning, and for the design of future sensor systems. The companion paper (Sandu et al.,
2012) considers several metrics from information theory are used to quantify the information content
320 of data, including the trace of the Fisher information matrix, the signal information, and the degrees
of freedom for signal. That paper shows how several of these metrics can be evaluated from expected
values of the 4D-Var cost function and its gradient.

The estimates require the computation of expected values with respect to the posterior distribution,
which is not readily available in 4D-Var data assimilation. In Section 3 we discuss several approaches
325 to obtain posterior expected values. The experiments reported here use random draws from a normal
distribution with the mean given by the assimilation result, a diagonal covariance matrix, and the anal-
ysis variances obtain by properly scaling the background variances. The scaling factors are obtained
by comparing both the background and analysis results against an independent data set.

We consider a global ozone data assimilation problem using TES satellite observations and the GEOS-
330 Chem chemical transport model. The quality of the analysis is assessed by comparing the results
against an independent data set (INTEX ozonesonde measurements). Assimilations of subsets (20%) of
the observation locations characterized by the highest signal, degrees of freedom for signal, and com-
bined information contents provide analyses that are comparable in quality with the analysis obtained
using all available observations. These results are highly encouraging as they indicate the effectiveness
335 of the proposed approach as a diagnostic tool for the value of observations used during data assimi-
lation. The issue of pruning the least informative observations without degrading the analysis deserves
future investigation.

Acknowledgements

This work has been supported in part by NASA through the ROSES-2005 AIST project, by NSF through
340 the awards NSF CCF-0635194, NSF OCI-0904397, NSF CCF-0916493, and NSF DMS-0915047, and by
the Computational Science Laboratory at Virginia Tech.

References

- Beer, R., Glavich, T. A., and Rider, D. M., Tropospheric emission spectrometer for the Earth Observing System's Aura satellite. *Applied Optics*, 2001; **40(15)**, 2356-2367.
- 345 Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B., Fiore, A. M., Li, Q., Liu, H., Mickley, L. J., and Schultz, M., Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research*, 2001; **106**, 23, 073-23,095.
- Bowman, K. W., Worden, J., Steck, T., Worden, H. M., Clough, S., and Rodgers, C., Capturing time and vertical variability of tropospheric ozone: A study using TES nadir retrievals. *Journal of Geophysical Research*, 2007; **107**,
350 D23.
- Bowman, K. W., Rodgers, C. D., Kulawik, S. S., Worden, J., Sarkissian, E., Osterman, G., Steck, T., Luo, M., Eldering, A., Shepherd, M., Worden, H., Lampel, M., Clough, S., Brown, P., Rinsland, C., Gunson, M., and Beer, R., Tropospheric Emission Spectrometer: Retrieval method and error analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 2006; **44(5)**, 1297-1307.
- 355 Cardinali, C., Pezzulli, S., Andersson, E., Influence-matrix diagnostic of data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 2004; **130**, 2767-2786.
- Carmichael, G. R., Daescu, D. N., Sandu, A., Chai, T., Computational aspects of chemical data assimilation into atmospheric models. *Computational Science - ICCS*, 2003, PT IV, Book series title: Lecture Notes in Computer Science , 2660: 269-278 , 2003.
- 360 Carmichael, G. R., Sandu, A., Chai, T., Daescu, D. N., Constantinescu, E. M., Tang, Y., Predicting air quality: Improvements through advanced methods to integrate models and measurements. *Journal of Computational Physics*, **227** (7): 3540-3571 , 2008
- Chai, TF;Carmichael, GR; Sandu, A; Tang, Y., Daescu, D. N., Chemical data assimilation of Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft measurements. *Journal of Geophysical Research – Atmospheres*, **111**
365 (D2): Art. No. D02301, 2006.
- Chai, TF; Carmichael, GR; Tang, YH; Sandu, A., Hardesty, M., Pilewskie, P., Whitlow, S., Browell, E. V., Avery, M. A., Nedelec, P., Merrill, J. T., Thompson, A. M., Williams, E., Four-dimensional data assimilation experiments with International Consortium for Atmospheric Research on Transport and Transformation ozone measurements. *Journal of Geophysical Research – Atmospheres*, **112** (D12): Art. No. D12S15, 2007.
- 370 Cheng, H., Jardak, M., Alexe, M. and Sandu, A., A hybrid approach to estimating error covariances in variational data assimilation. *Tellus A*. **Vol. 62**, Number 3, May 2010 , pp. 288-297(10).
- Constantinescu, E. M., Sandu, A., Chai, T. F., Ensemble-based chemical data assimilation. I: General approach. *Quarterly Journal of the Royal Meteorological Society*, **133** (626): 1229-1243 Part A , 2007
- Constantinescu, E. M., Sandu, A., Chai, T. F., Carmichael, G. R., Ensemble-based chemical data assimilation. II:
375 Covariance localization. *Quarterly Journal of the Royal Meteorological Society*, **133** (626): 1245-1256 Part A , 2007
- Constantinescu, E. M., Sandu, A., Chai, T. F., Carmichael, G. R., Assessment of ensemble-based chemical data assimilation in an idealized setting. *Atmospheric Environment*, **41** (1): 18-36 , 2007.
- Constantinescu, E. M., Chai, T. F., Sandu, A., Carmichael, G. R., Autoregressive models of background errors for chemical data assimilation. *Journal of Geophysical Research – Atmospheres*, **112** (D12): Art. No. D12309 , 2007.
- 380 Daescu, D., Carmichael, G. R., Sandu, A., Adjoint implementation of Rosenbrock methods applied to variational

- data assimilation problems. *Journal of Computational Physics*, 165 (2): 496-510, 2000.
- Eller, P., Singh, K., Sandu, A., Bowman, K., Henze, D. K., and Lee, M., Implementation and evaluation of an array of chemical solvers in a global chemical transport model. *Geophysical Model Development*, 2009; **Vol. 2**, 1-7.
- Fisher, R. A., On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, 1922; **Series A**, **222**, 309-368, URL: <http://www.jstor.org/stable/91208>.
- 385 Fisher, M., Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *ECMWF Technical Memoranda*, 2003; **397**.
- Gejadze, I. Y., Le Dimet, F. X., and Shutyaev, V., On analysis error covariances in variational data assimilation. *SIAM Journal on Scientific Computing*, 2008; **30(4)**, 1847-1874.
- 390 Hakami, A., Henze, D. K., Seinfeld, J. H., Chai, T. F., Tang, Y., Carmichael, G. R., and Sandu, A., Adjoint inverse modeling of black carbon during the Asian Pacific Regional Aerosol Characterization Experiment. *Journal of Geophysical Research – Atmospheres*, 110, (D14): Art. No. D14301, 2005.
- Henze, D. K., Seinfeld, J. H., Liao, W., Sandu, A., Carmichael, G. R., Inverse modeling of aerosol dynamics: Condensational growth. *Journal of Geophysical Research – Atmospheres*, 109 (D14): Art. No. D14201 , 2004.
- 395 Henze, D. K., Hakami, A., and Seinfeld, J. H., Development of the adjoint of GEOS-Chem. *Atmospheric Chemistry and Physics*, 2007; **7**, 2413-2433.
- Henze, D. K., Seinfeld, J. H., and Shindell, D. T., Inverse modeling and mapping U.S. air quality influences of inorganic PM_{2.5} precursor emissions with the adjoint of GEOS-Chem. *Atmospheric Chemistry and Physics*, 2009; **9**, 5877-5903.
- 400 Jones, D. B. A., Bowman, K. W., Palmer, P. I., Worden, J. R., Jacob, D. J., Hoffman, R. N., Bey, I., and Yantosca, R. M., Potential of observations from the Tropospheric Emission Spectrometer to constrain continental sources of carbon monoxide. *Journal of Geophysical Research*, 2003; **108**, D24.
- Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., and Zhang, Q., A comparison of analytical and adjoint Bayesian inversion methods for constraining Asian sources of CO using satellite (MOPITT) measurements of CO columns. *Journal of Geophysical Research*, 2009; **114**, D04305.
- 405 Liao, W. Y., Sandu, A., Carmichael, G. R., Chai, T. F., Singular vector analysis for atmospheric chemical transport models. *Monthly Weather Review*, 134 (9): 2443-2465 SEP 2006.
- Parrington, M., Jones, D. B. A., Bowman, K. W., Horowitz, L. W., Thompson, A. M., Tarasick, D. W., Witte, J. C., Estimating the summertime tropospheric ozone distribution over North America through assimilation of observations from the Tropospheric Emission Spectrometer. *Journal of Geophysical Research*, 2008; **Vol 113**, D18307.
- 410 Parrington, M., Jones, D. B. A., Bowman, K. W., Thompson, A. M., Tarasick, D. W., Merrill, J., Oltmans, S. J., Leblanc, T., Witte, J. C., Millet, D. B., Impact of the assimilation of ozone from the Tropospheric Emission Spectrometer on surface ozone across North America. *Geophysical Research Letters*, 2009; **36(4)**, L04802.
- 415 Rodgers, C. D., Information content and optimization of high spectral resolution measurements. *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research*, **SPIE Volume 2830**, 136-147.
- Sandu, A., Liao, W., Carmichael, G. R., Henze, D. K., and Seinfeld, J. H., Inverse modeling of aerosol dynamics using adjoints: Theoretical and numerical considerations. *Aerosol Science and Technology*, 2005, 39 (8): 677-694.
- Sandu, A., Constantinescu, E. M., Liao, W. Y., Carmichael, G. R., Chai, T., Seinfeld, J. H., and Daescu, D. N.,

- 420 Ensemble-based data assimilation for atmospheric chemical transport models. *Computational Science - ICCS*, 2005, PT 2 Book series title: Lecture Notes in Computer Science , 3515: 648-655.
- Sandu, A., Daescu, D. N., Carmichael, G. R., and Chai, T., Adjoint sensitivity analysis of regional air quality models. *Journal of Computational Physics*, 2005; **Volume 204**, 222-252.
- Sandu, A. and Zhang, L., Discrete second order adjoints in atmospheric chemical transport modeling. *Journal of*
425 *Computational Physics*, 2008; **227(12)**, 5949-5983.
- Sandu, A., Singh, K., Jardak, M., Bowman, K., Lee, M., A Practical Method to Estimate Information Content in the Context of 4D-Var Data Assimilation. I: Methodology. *Atmospheric Chemistry and Physics*, submitted, 2012.
- Singh, K., Eller, P., Sandu, A., Bowman, K. W., Jones, D., Lee, M., Improving GEOS-Chem model forecasts through profile retrievals from Tropospheric Emission Spectrometer. *International Conference on Computational Science*,
430 2009, Lecture Notes on Computational Science **volume 5545**, 302-311.
- Singh, K., Eller, P., Sandu, A., Henze, D., Bowman, K., Kopacz, M. and Lee, M., Towards the construction of a standard adjoint GEOS-Chem model. *High Performance Computing Symposium at Spring Simulation Multiconference*, 2009.
- Singh, K., Jardak, M., Sandu, A., Bowman, K. W., Lee, M., Jones, D., Construction of non-diagonal background error covariance matrices in global chemical data assimilation. *Geophysical Model Development*, 4, 299-316, doi:10.5194/gmd-4-299-2011, 2011.
- Stewart, L. M., Dance, S. L., Nichols, N. K., Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids*, 2008; **56**, 1521-1527.
- Thacker, W. C., The role of the Hessian matrix in fitting models to measurements. *Journal of Geophysical Research*,
440 1989; **94(C5)**, 6177-6196.
- Thompson, A.M., et al. (2007a), "Intercontinental chemical transport experiment ozonesonde network study (IONS) 2004: 1. Summertime upper troposphere/lower stratosphere ozone over northeastern North America". *Journal of Geophysical Research*, **112**, D12S12.
- Thompson, A.M., et al. (2007b), "Intercontinental chemical transport experiment ozonesonde network study
445 (IONS) 2004: 2. Tropospheric ozone budgets and variability over northeastern North America". *Journal of Geophysical Research*, **112**, D12S13.
- Xu, Q. (2006), Measuring information content from observations for data assimilation: relative entropy versus Shannon entropy difference. *Tellus, A.*, 198-209.
- Wikle, C.K., and Berliner, L.M. A Bayesian tutorial for data assimilation. *Physica D*, 2007; **230**, 1-16.
- 450 Worden, J. R., Bowman, K. W., and Jones D. B. A., Characterization of atmospheric profile retrievals from Limb Sounding Observations of an inhomogeneous atmosphere. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 2004; **86**,(03)00274-7.
- Zhang, L., Jacob, D. J.,Kopacz, M., Henze, D. K., Singh, K., Jaffe, D. A., Intercontinental source attribution of ozone pollution at western U.S. sites using an adjoint method. *Geophysical Research Letters*, 2009; **Volume 36**, L11810.
- 455 Zhu, C., Byrd, R. H., and Nocedal., J., L-BFGS-B: Algorithm 778, FORTRAN routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software*, 1997; **Vol 23, Num. 4**,550-560.
- Zupanski, D., Hou, A.Y., Zhang, S.Q.: Applications of information theory in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*; Vol. 133, pp. 1533-1545, 2007.