Chemical data assimilation of Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft measurements

Tianfeng Chai,[1] Gregory R. Carmichael,[1] Adrian Sandu,[2] Youhua Tang,[1] and Dacian N. Daescu[3]

Received 15 February 2005; revised 7 September 2005; accepted 28 September 2005; published 17 January 2006.

[1] In this paper, the four-dimensional variational (4D-Var) technique is applied to assimilate aircraft measurements during the Transport and Chemical Evolution over the Pacific (TRACE-P) field experiment into a chemical transport model, Sulfur Transport Eulerian Model, version 2K1 (STEM-2K1). Whether data assimilation would produce better analyzed fields is examined. It is found that assimilating ozone observations from one of two independent flights improves model prediction of the other flight ozone measurements, which are withheld as validation data. The adjusted initial fields after only assimilating the total reactive nitrogen (NO_y) observations lead to better predictions of NO, NO_2, and PAN, based on their agreement with the withheld measurements. One experiment simultaneously assimilating the observations of O_3, NO, NO_2, HNO_3, PAN, and RNO_3 demonstrates that the model is able to match those measurements well by changing the initial fields. In addition, the model predictions of NO_y improve significantly after assimilating the aforementioned multiple observation species, which are independent of the withheld NO_y measurements. In the paper, we also show that the key species whose initial mixing ratios would significantly affect the agreement between model and measurements can be identified using adjoint sensitivity analysis. Such information can be used to reduce the number of control variables in the 4D-Var data assimilation. To speed up the optimization process in the 4D-Var, we enforce the concentration upper bounds through the limited memory–Broyden-Fletcher-Goldfarb-Shanno-B (L-BFGS-B) algorithm, and this proves to be effective.

Citation: Chai, T., G. R. Carmichael, A. Sandu, Y. Tang, and D. N. Daescu (2006), Chemical data assimilation of Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft measurements, *J. Geophys. Res.*, *111*, D02301, doi:10.1029/2005JD005883.

1. Introduction

[2] Chemical transport models (CTMs) have developed into a major tool in air quality and atmospheric chemistry analysis. Regional CTMs play an indispensable role in assessing, predicting, and interpreting the air-quality-related hazardous events. On the other hand, large amounts of atmospheric chemistry observation data have been obtained through many field campaigns and a wide range of monitoring networks provide measurements on a regular basis. While CTMs and observations complement each other, the optimal analysis by integrating both components requires advanced data assimilation techniques.

[3] Two methods have been extensively studied in meteorology and oceanography fields, the Kalman filter method (including many variations) using a sequential approach, and the four-dimensional variational assimilation (4D-Var) with a variational approach. Both methods have been put into operational forecasting systems to assimilate the routinely available meteorological measurements. The 4D-Var method has been implemented at the European Centre for Medium-Range Weather Forecasts (ECMWF) [*Rabier et al.*, 2000; *Mahfouf and Rabier*, 2000; *Klinker et al.*, 2000]. The National Centers for Environmental Prediction (NCEP) have built an operational regional analysis-forecast system, the Rapid Update Cycle (RUC), using 4D-Var with a hourly assimilation cycle [*Benjamin et al.*, 2004]. An operational implementation of Ensemble Kalman filter (EnKF) has been under development at the Canadian Meteorological Center (CMC) [*Houtekamer et al.*, 2005].

[1] Center for Global and Regional Environmental Research, University of Iowa, Iowa City, Iowa, USA.
[2] Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.
[3] Department of Mathematics and Statistics, Portland State University, Portland, Oregon, USA.

0148-0227/06/2005JD005883

[4] Advanced data assimilation in atmospheric chemistry was first carried out by *Fisher and Lary* [1995]. They applied 4D-Var to O_3 and NO_2 observations from the Upper Atmosphere Research Satellite (UARS), assimilating them into a simple chemistry model, which included 6 species and 19 reactions. A Lagrangian scheme was adopted for the transport occurring in the stratosphere. They showed that data assimilation is capable of getting synoptic analysis from asynoptic observations. *Errera and Fonteyn* [2001] presented a stratospheric chemical assimilation system with a three-dimensional (3-D) CTM using 4D-Var, and applied it to the Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere (CRISTA) measurements. Using the variational approach, *Elbern et al.* [1997] developed an adjoint model of the European Air Pollution model (EURAD), a regional Eulerian model that focuses on tropospheric chemistry, and they have used it in various studies [*Elbern and Schmidt*, 1999; *Elbern et al.*, 2000; *Elbern and Schmidt*, 2001; *Hoelzemann et al.*, 2001]. For instance, *Elbern and Schmidt* [2001] applied the 4D-Var data assimilation to study an enhanced summer ozone episode over central Europe during August 1997. They demonstrated a significant improvement in short-term forecast of ozone after a 6-hour data assimilation interval. Other Eulerian 3-D CTMs that have developed adjoint models include Sulfur Transport Eulerian Model, version 2K1 (STEM-2K1) [*Daescu and Carmichael*, 2003; *Carmichael et al.*, 2003a; *Sandu et al.*, 2005], and the CHIMERE model [*Menut et al.*, 2000; *Schmidt and Martin*, 2003; *Menut*, 2003]. In addition to the variational data assimilation work, there are a number of atmospheric chemistry data assimilation applications using sequential approaches. These include various Kalman filter methods [*Zhang et al.*, 1999; *Segers et al.*, 2000; *van Loon et al.*, 2000; *Clerbaux et al.*, 2001; *Lamarque et al.*, 2002; *Lamarque and Gille*, 2003; *Lamarque et al.*, 2004; *Hanea et al.*, 2004].

[5] Many aircraft field campaigns conducted in the past several years have utilized CTMs for flight planning, as well as for the analyses of the observations [*Lawrence et al.*, 2002; *Carmichael et al.*, 2003b]. It is our premise that the value derived from large field experiments can be enhanced by a closer integration of modeled and measured quantities, with the two merged together to provide a consistent and best estimate of the chemical state of the atmosphere. In this paper we apply the 4D-Var technique to assimilate different chemical measurements from two TRACE-P flights into STEM-2K1 model. We use these flights to address issues related to the following: which species can be assimilated; what is the impact of the choice of control variables; and what is the impact of simultaneous assimilation of multiple species. Whether data assimilation would produce better analyzed fields is examined using various combinations of assimilated and withheld measurements.

[6] The paper is organized as follows. The three components of the 4D-Var assimilation framework, including the current CTM model, its adjoint, and the optimization aspects, are introduced in section 2. In section 3, we demonstrate and explore various issues related to the assimilation of ozone, NO_y and CO observations. These results are discussed in terms of how analyzed fields can be produced using 4D-Var. It is followed by a summary in section 4. Results for missionwide application will be presented in a future paper.

2. Four-Dimensional Variational Data Assimilation Method

2.1. Chemical Transport Model

[7] STEM-2K1 is a regional CTM, which in this study uses the SAPRC-99 chemical mechanism [*Carter*, 2000]. In total 213 gas phase reactions and 93 chemical constituents are included in this mechanism. In the model, the evolution of the chemical constituent concentration vector c in time (t) is described as

$$\frac{\partial c}{\partial t} = -u \cdot \nabla c + \frac{1}{\rho}\nabla \cdot (\rho K \cdot \nabla c) + \frac{1}{\rho} f + E \tag{1}$$

Here we denote by u the wind field vector, ρ the air density, K the turbulent diffusivity tensor, and f the chemical transformation rate. E is the elevated emission rate which describes the emission sources above the ground surface. The boundary conditions are

$$c(t,x) = c^{\mathrm{IN}}(t,x) \quad \text{for} \quad x \in \Gamma^{\mathrm{IN}}, \tag{2}$$

$$k_{nn}\frac{\partial c}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^{\mathrm{OUT}}, \tag{3}$$

$$k_{nn}\frac{\partial c}{\partial n} = V^{\mathrm{dep}} c - Q \quad \text{for} \quad x \in \Gamma^{\mathrm{GR}}. \tag{4}$$

where n is the unit outward normal vector on the domain surface, and $k_{nn} = n \cdot K \cdot n$. V^{dep} is the deposition velocity. Q is the rate of surface emissions. Γ^{GR} stands for the ground surface boundary. Lateral boundaries are partitioned into inflow Γ^{IN} and outflow Γ^{OUT} regions. On the domain top, concentrations are fixed.

[8] In STEM-2K1, the evolution is solved using an operator splitting approach, in which the transport along each direction and the chemistry steps are taken successively. A Rosenbrock method is used as the chemistry solver. The Kinetic Preprocessor (KPP) tool [*Sandu et al.*, 2003; *Daescu et al.*, 2003] was used to generate both the forward and the adjoint chemistry code. Table 1 provides a summary of STEM-2K1. More details on the CTM model are given by *Carmichael et al.* [2003b] and *Tang et al.* [2003b].

2.2. Cost Functional and Adjoint Model

[9] In 4D-Var, a cost functional is generally defined as

$$J = \frac{1}{2}[c_0 - c_b]^T B^{-1}[c_0 - c_b] + \frac{1}{2}[y - h(c)]^T O^{-1}[y - h(c)] \tag{5}$$

where B and O are error covariance matrices for background and observations in discrete spaces, respectively. The presence of the background term is meant to guarantee the uniqueness of the optimal solution. B is defined at the end of this section, and O will be defined in section 3.1. h is a projection operator, calculating the observation vector y from the model space c. In the current study the initial concentrations c_0 are chosen as the control parameters to be

Table 1. Description of STEM-2K1

	Description
Gas-phase chemistry	SAPRC99 chemical mechanism [*Carter*, 2000], 93 chemical constituents, 213 reactions (including 30 photolysis reactions)
Meteorological data	calculated using RAMS [*Pielke et al.*, 1992]
Radiation model	NCAR Tropospheric Ultraviolet-Visible (TUV) [*Madronich and Flocke*, 1999]
Chemistry solver	implicit second-order Rosenbrock [*Daescu et al.*, 2000]
Transport scheme	finite differences, Crank-Nicholson scheme [*Sandu et al.*, 2005]

adjusted. Hereafter, the subscript "0" is used to denote variables at instant $t = 0$. Assuming that the operator h is linear, $h(c)$ can be written as $h(c) = H \cdot c$. In our application, H reflects trilinear interpolation in space and linear interpolation in time when constructing model counterparts of observations.

[10] The optimal solution of 4D-Var minimizes J in equation (5), while imposing the model as a strong constraint, i.e., assuming a perfect model. To effectively solve the optimization problem, the gradients of J with respect to the control parameters are required and they can be efficiently calculated using adjoint method. The adjoint can be derived using Lagrange multipliers [*Daley*, 1991; *Wang et al.*, 2001]. The derivation of STEM-2K1 adjoint is described in detail by *Carmichael et al.* [2003a], and *Daescu et al.* [2000] show the derivation of the adjoint of the Rosenbrock scheme. The adjoint of the tangent linear model of equation (1) defines the evolution of the adjoint variable vector λ, which reads as

$$\frac{\partial \lambda}{\partial t} + \nabla \cdot (u\lambda) = -\nabla \cdot \left(\rho K \cdot \nabla \frac{\lambda}{\rho}\right) - (F \cdot \lambda) - \phi \qquad (6)$$

where the forcing term ϕ in equation (6) appears as

$$\phi = H^T \cdot O^{-1} \cdot [y - H \cdot c] \qquad (7)$$

In equation (6), F is a tensor function, obtained by linearizing the incremental f as $\delta f \approx F \cdot \delta(\rho c)$. As f is a nonlinear function of (ρc), F also varies with (ρc), i.e., $F = F(\rho c)$. The corresponding boundary conditions are

$$\lambda(t, x) = 0 \quad \text{for} \quad x \in \Gamma^{\text{IN}}, \qquad (8)$$

$$\lambda u_n + \rho k_{nn} \frac{\partial(\lambda/\rho)}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^{\text{OUT}}, \qquad (9)$$

$$\rho k_{nn} \frac{\partial(\lambda/\rho)}{\partial n} = V^{\text{dep}} \lambda \quad \text{for} \quad x \in \Gamma^{\text{GR}}. \qquad (10)$$

To obtain the surface boundary condition we use the fact that $u_n = u \cdot n = 0$ at ground level. $\lambda(t, x) = 0$ holds on the top boundary. Note that the adjoint initial conditions hold at the final time T.

[11] The backward integration of equation (6) gives adjoint variables at any time, which are the sensitivities of J_O (the observation part of the cost functional in equation (5)) with respect to state variables (concentrations), i.e.,

$$\delta J_O = \lambda^T \cdot \delta c \qquad (11)$$

Note that the background part of the cost functional (J_B) in equation (5) adds one more term to the gradient of the cost functional with respect to initial concentrations. From equation (5), it is easy to obtain

$$\delta J = \delta(J_O + J_B) = \left[\lambda_0^T + (c_0 - c_b)^T \cdot B^{-1}\right] \cdot \delta c_0 \qquad (12)$$

where $\lambda_0^T + (c_0 - c_b)^T \cdot B^{-1}$ is the gradient information needed for the minimization. The optimal initial condition c_0 can be found efficiently by applying minimization routines. Quasi-Newton limited memory–Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [*Byrd et al.*, 1995] is used by most 4D-Var applications. In this article, an updated version, L-BFGS-B [*Zhu et al.*, 1997], is used. We found that adding bound constraints improve the computation efficiency. This is discussed further in section 3.3.3.

[12] In minimization, the control variable is chosen as $\ln(c_0/c_r)$, where c_r is a reference concentration. The logarithm change of control variables acts as a crude preconditioning in the optimization, and it was implemented by *Elbern et al.* [1997]. In the current study the predictions using the same CTM from the previous analyses [*Carmichael et al.*, 2003b] were chosen as the background fields. Lacking knowledge of the background error statistics, we assume B to be a diagonal matrix and the background field is considerably more uncertain than the measurements. It is noted that the introduction of error correlation is beneficial, especially for the estimation of the emission inventories using adjoint method. This can be achieved by "NMC method" [*Parrish and Derber*, 1992; *Heald et al.*, 2004] or covariance modeling [*Lamarque et al.*, 1999; *Khattatov et al.*, 2000; *Clerbaux et al.*, 2001; *Lamarque and Gille*, 2003; *Hoelzemann et al.*, 2001]. We are currently studying error correlation and this is the subject of a future paper. In the data assimilation analyses presented here, the optimization proceeds until the cost functional is reduced to 0.001 of its initial value, or the number of forward-backward model integrations exceeds 20.

3. Data Assimilation Analysis

3.1. Observations and Model Setup

[13] As part of the NASA Global Tropospheric Experiment (GTE), the TRACE-P mission was conducted over the western Pacific in February–April 2001. It used two NASA

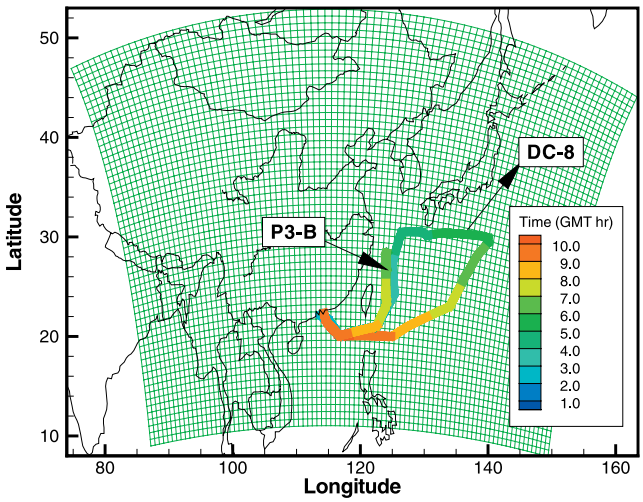

Figure 1. Flight routes of the DC-8 and the P3-B on 7 March 2001. The P3-B return trajectory overlaps with and hides the first part of the trajectory. The computational domain and grid are also shown.

aircraft, the DC-8 (ceiling 12 km) and the P-3B (ceiling 7 km) operating out of Yokota Air Force Base (near Tokyo, Japan) and Hong Kong. With the chemical composition and evolution of the Asia outflow as the mission focus, the aircraft carried various instruments to measure long-lived greenhouse gases, photochemical oxidants, aerosols, and their precursors (details of measurements and instrument descriptions are available at http://www-gte.larc.nasa.gov/trace/TP_Investigator_Measurements.htm). Measurements from two flights, flight 7 of the DC-8 and flight 9 of the P3-B, were chosen to demonstrate on utilization of the 4D-Var technique. An overlap between two flight tracks is considered helpful for the data assimilation validation, but otherwise the choice is arbitrary. The coverage of the two flight routes is given in Figure 1, while Figure 2 displays the flight heights and the distances between the two aircraft as three time series. The observations are 5-min averaged, and run through 0112–1045 UT for the DC-8 and 0220–1000 UT for the P3-B. Note that the local time is 8 hours (for Hong Kong) or 9 hours (for Tokyo) earlier than universal time (UT).

[14] The aircraft measurement errors are assumed to be uncorrelated, thus the observation error covariance O is a diagonal matrix. Note that the P3-B measurements of NO_y and its underlying components were independent of each other (separate measurements). The variances are obtained by approximating the measurement uncertainty for O_3, NO_y, NO, NO_2, and CO to be 8%, 18%, 20%, 20%, and 1%, respectively. The approximations were based on the measurement accuracy and/or precision provided at the instrument description Web site (http://www-gte.larc.nasa.gov/trace/TP_Investigator_Measurements.htm). Adjustment factors (between 1 and 2, according to comparable measurements) were applied to obtain the measurement uncertainties. Measurement uncertainties for other species were assigned 100% when no measurement accuracy or precision information is available.

[15] STEM-2K1 was run with a 90 × 60 × 18 grid, and a horizontal resolution of 80 km × 80 km. Sigma-z coordinate system is used in the vertical direction, following the RAMS dynamic model [*Pielke et al.*, 1992], which provides the meteorological fields to STEM-2K1. ECMWF 1° × 1° reanalysis was used to drive the RAMS model. Figures 3 and 4 display the RAMS wind fields at 0000 UT and 0900 UT, at altitudes of 3.5 km and 1.0 km, respectively. The flow situation on this day was dominated by a low-pressure system, ranging from central China to northwest Pacific Ocean. Strong continental outflows associated with these flows transported pollutants from Southeast Asia (including pollutants associated with biomass burning in Thailand and Myanmar) into the western Pacific. This outflow of pollutants from southeast Asia was maximum at altitudes between 2 and 4 km. Flows below 2 km were typical of the winter monsoon, and these southerly flows transported pollutants from coastal China to the lower latitudes. The flights on this day were designed to sample the continental outflow in this frontal system. Further details of the transport of pollutants in this frontal system are given by *Carmichael et al.* [2003b]. The emission inventories implemented here are presented by *Streets et al.* [2003] and *Woo et al.* [2003].

[16] An assimilation window of 12 hours (0000–1200 UT, 0800–2000 Hong Kong local time) is used for all the data assimilation analyses presented here, while the assimilated observational data vary from case to case. Both the forward and adjoint models are parallel and implemented using the Parallelization library for Air Quality Models on Structured Grids (PAQMSG) [*Miehe et al.*, 2002].

3.2. O_3 Assimilation

[17] We start the discussion with results for the assimilation of a single species (ozone). Ozone is a secondary pollutant in troposphere, but plays a central role in the photochemical oxidant cycle. As such it depends on and influences many atmospheric chemical species. So the ability to better constrain/predict ozone in the atmosphere is an important goal. In this section, we present the data

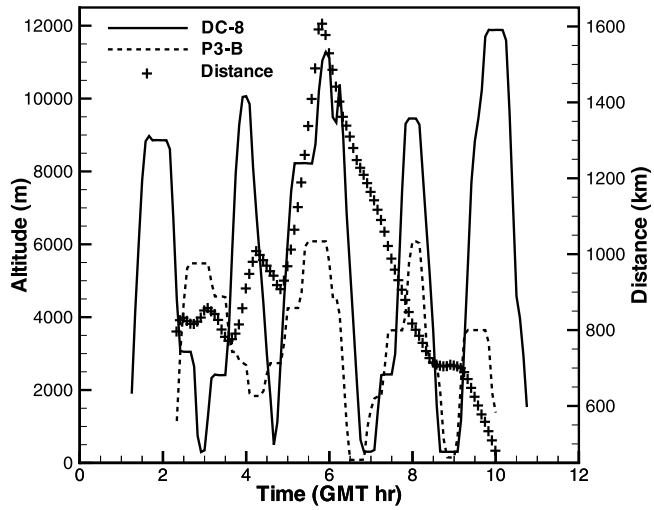

Figure 2. Time series of flight heights of the DC-8 and the P3-B on 7 March 2001. The distances between the two aircraft are also shown.

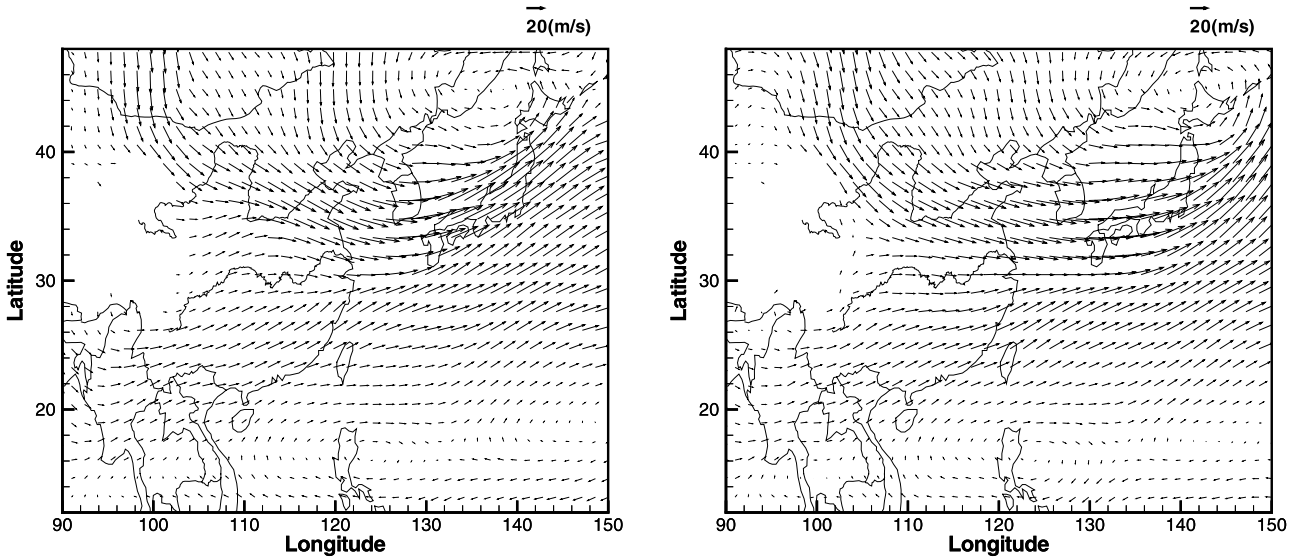

Figure 3. Wind fields at z = 3.5 km: (left) 0000 UT; (right) 0900 UT.

assimilation results of two tests, in which O_3 measurements by the DC-8 and the P3-B are assimilated separately. The ozone measurements were obtained by M. Avery from NASA Langley. In these two tests, initial O_3 concentrations are chosen to be the only control variables to be adjusted.

3.2.1. Data Assimilation

[18] Figure 5 shows the time series of model predictions of O_3 concentrations from the two tests, as well as the original observations. Without data assimilation the STEM-2K1 predictions reproduce the overall trend of the DC-8 O_3 measurements, but consistently miss the low values at 0700 UT, 0800 UT, 0900 UT, and 1000 UT. After the measurements are assimilated, the CTM results agree with the flight measurements very well. This result demonstrates that the assimilation procedure works for this case and the adjustment of ozone initial conditions only can result in a predicted field that is consistent with the specified concentrations of the precursors.

[19] As the P3-B flew slower and covered a smaller region, its measurements represent some spatial variations that are difficult for the CTM to reproduce at the current resolution. The corresponding model results appear to be too smooth compared to the P3-B measurements. The time series of P3-B measurements shows O_3 concentrations over 100 ppbv at about 0400 UT and 0830 UT. The locations of these two instances are close to each other, at (121°E, 21°N), and the flight altitudes are ~2.7 km for both instances. This region of high O_3 appears to be associated with the biomass-burning-influenced outflow from Southeast Asia (see Figure 3). The DC-8 failed to pick up the high concentrations since it only flew past that region once, at 0400 UT, at a much higher altitude (>6 km). The placement

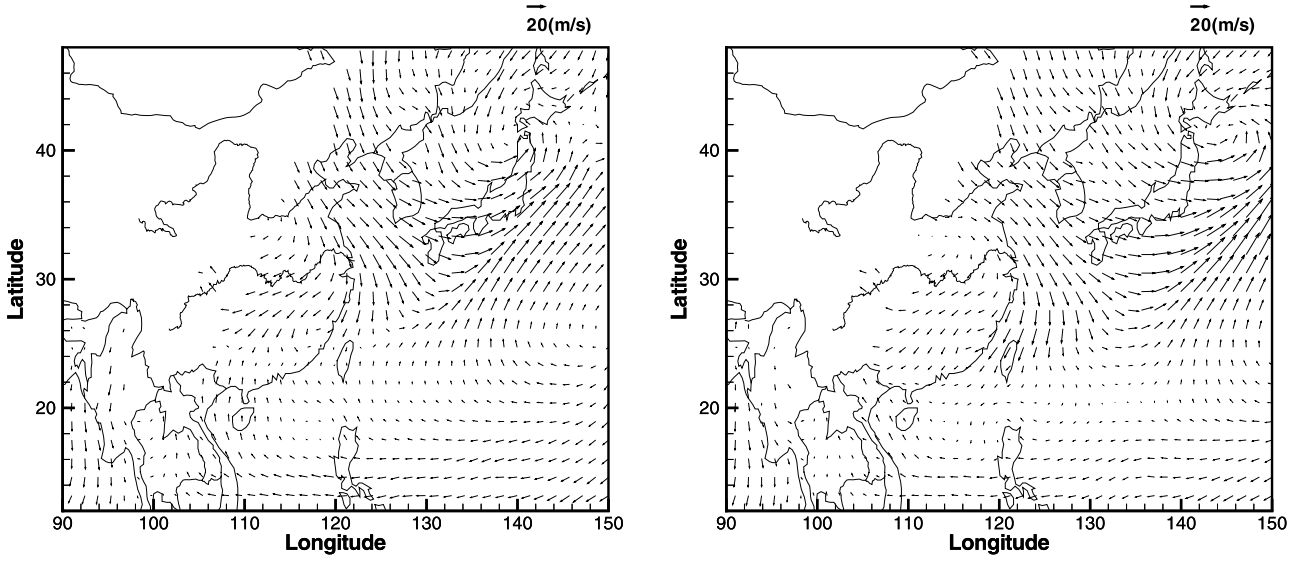

Figure 4. Wind fields at z = 1.0 km: (left) 0000 UT; (right) 0900 UT.

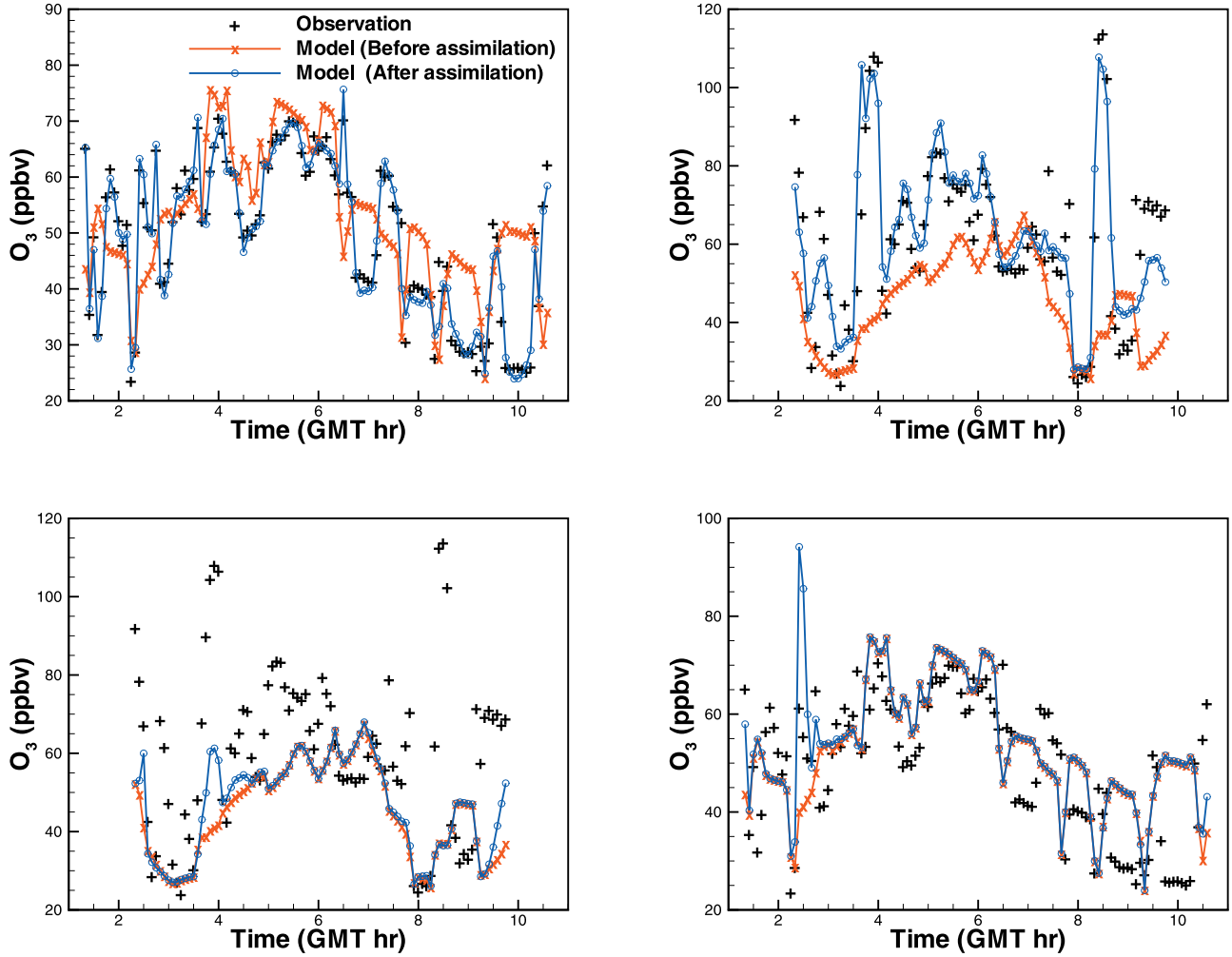

Figure 5. Results of the O_3 assimilation tests. The flight measurements and their corresponding model predictions are shown. (left) Results of the test that assimilated the DC-8 O_3 observations. (right) Results of the test that assimilated the P3-B O_3 observation. The top panels show assimilated flight observations. The bottom panels show withheld observations from the other flight. O_3 observation uncertainty is assigned 8%.

and timing of this elevated plume-like structure in the warm sector of this frontal system is difficult to precisely predict. As discussed by *Tang et al.* [2003a], the forward model meteorology had the position of the front displaced in time. This along with uncertainties in the timing and spatial distribution of the fires in Southeast Asia at this time, resulted in the displacement (<∼100 km, and <∼10 hours) of the plume from that observed (wrong place at wrong time). So the model is blind to the structure unless observations are actually injected into the model through data assimilation. It is seen in Figure 5 that a better match between the model and the P3-B observations is achieved after the measurements are assimilated, while discrepancies still remain. The meteorological fields used here were reanalyzed using observed meteorology observations, but this alone was not able to put the biomass plumes in the correct positions. In this case the assimilation of the chemical data is essential.

3.2.2. Validation

[20] After assimilating observations from the two flights separately, we applied the updated initial conditions to predict measurements from the other flight which were not used in the previous data assimilation tests. Results are also shown in Figure 5. Only slight improvement of the model prediction for the P3-B data is achieved. For both validation tests (for the P3-B data the DC-8 data are assimilated, and for DC-8 data the P3-B data are assimilated), the agreement between measurement and model results becomes better at the beginning and the end of the flights. This is due to the fact that both flights took off from and landed at the same base, shown in Figure 1. Note that the two aircraft were not close to each other at any instance during the flight, as there was about one hour difference in both takeoff and landing time between the two aircraft. Figure 2 shows that the closest distance between the DC-8 and the P3-B was about 500 km, i.e., about six grid cells, when the P3-B was about to land. This result points out that the effect of the 4D-Var assimilated measurements is felt beyond their neighboring grid points even without spatial background error covariance introduced. It is speculated that introducing the spatial background error covariance matrix would extend the assimilation effect further in

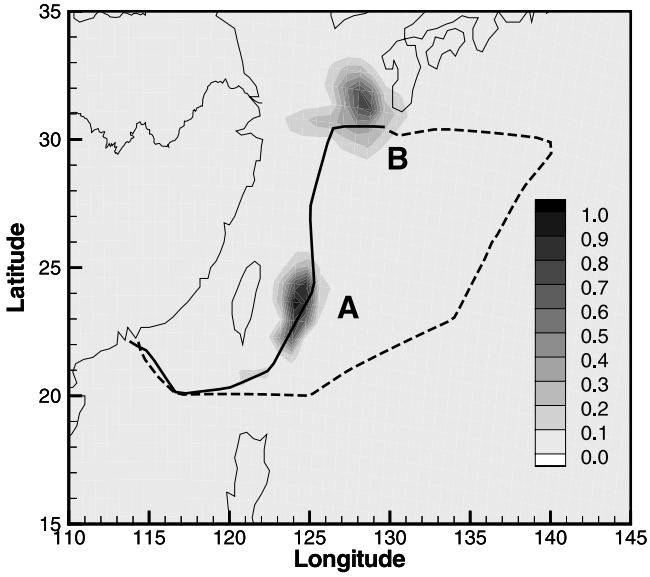

Figure 6. A top view of the "cones of influence" of ozone. Regions A and B are also marked in Figure 7. The maximum $\mathcal{C}(x)$ over 18 vertical levels is chosen at each geographical location. The values shown are normalized. The flight track of the DC-8 is also shown, where the dashed line is after 0440 UT.

distance and hence produce larger improvements on the analyzed field. This has yet to be tested using a background error covariance matrix with nonzero off-diagonal elements.

[21] Although assimilating the flight measurements does not change the model predictions of the other flight significantly, the modifications that the assimilation incurred were found to improve the agreement between the model and the withheld observations from the other flight. For example, model predictions of O_3 along the P3-B flight route became better during the period of 0340–0440 UT after assimilating the DC-8 O_3 measurements. Between 0340 UT and 0440 UT, the STEM-2K1 model predicted higher O_3 concentrations than the DC-8 measurements before data assimilation. After assimilating the lower DC-8 O_3 measurements to bring down the corresponding model results, the predicted values for the P3-B increased during this period, resulting in a better match between the model and measurements.

[22] To shed more light on the improvement of the model predictions on the P3-B measurements after the assimilation of the DC-8 O_3 measurements, we built a new cost functional only using the P3-B O_3 measurements between 0340 UT and 0440 UT. An influence function $\mathcal{C}(x)$ is calculated to show the regions that significantly affect model predictions corresponding to the selected observations. $\mathcal{C}(x)$ is obtained by averaging the magnitudes of adjoint sensitivities over time, i.e.,

$$\mathcal{C}(x) = \frac{1}{N}\sum_{n=0}^{N-1} |\lambda(x,t)|, \tag{13}$$

where N is the total number of time steps. This can be calculated for any species, showing their effects on the specific predictions corresponding to the selected observations. The spatial distribution of $\mathcal{C}(x)$ shows the "cones of influence" (see *Sandu et al.* [2005] for details). Figures 6 and 7 display the projections of the "cones of influence" of ozone. While the influence region ($\mathcal{C}(x) > 0$) actually extends to the whole domain, the current "cones of influence" plot is used to identify those more important regions. It is seen that the knowledge of ozone concentration in two regions (marked as A and B in Figures 6 and 7) is most helpful for the predictions of the selected observations and they were both covered by the DC-8 flight. As expected, one region (A in Figures 6 and 7) that

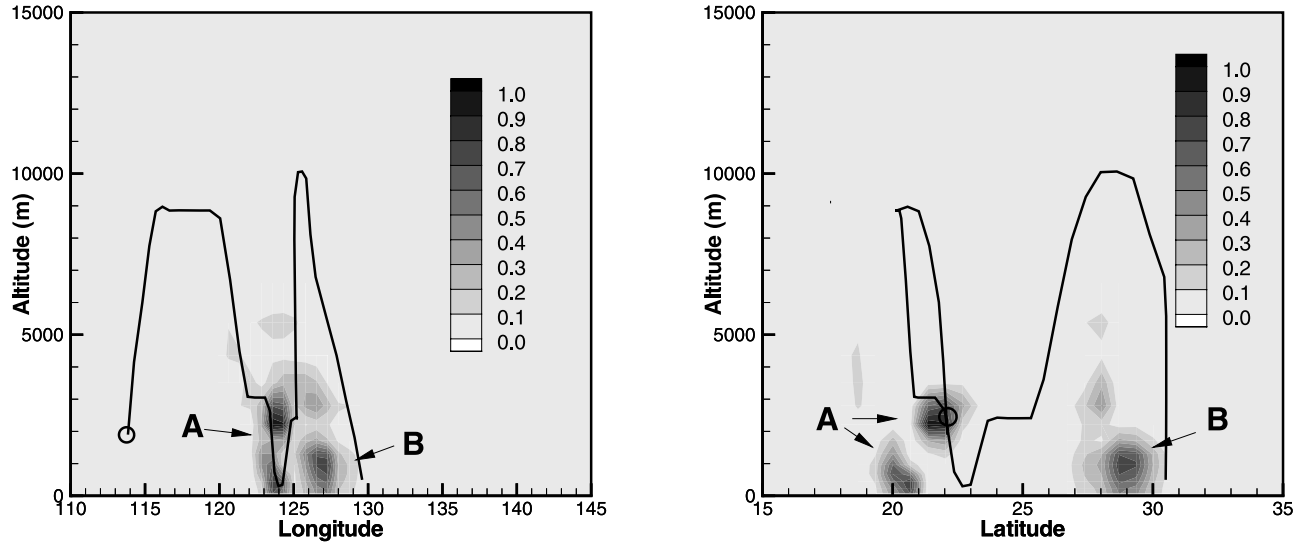

Figure 7. (left) A south view and (right) an east view of the "cones of influence" of ozone. Regions A and B are also marked in Figure 6. The projection is made by choosing the maximum value in the projection direction. The values shown are normalized. The flight track of the DC-8 before 0440 UT is also shown. The start of the track is marked by a circle.

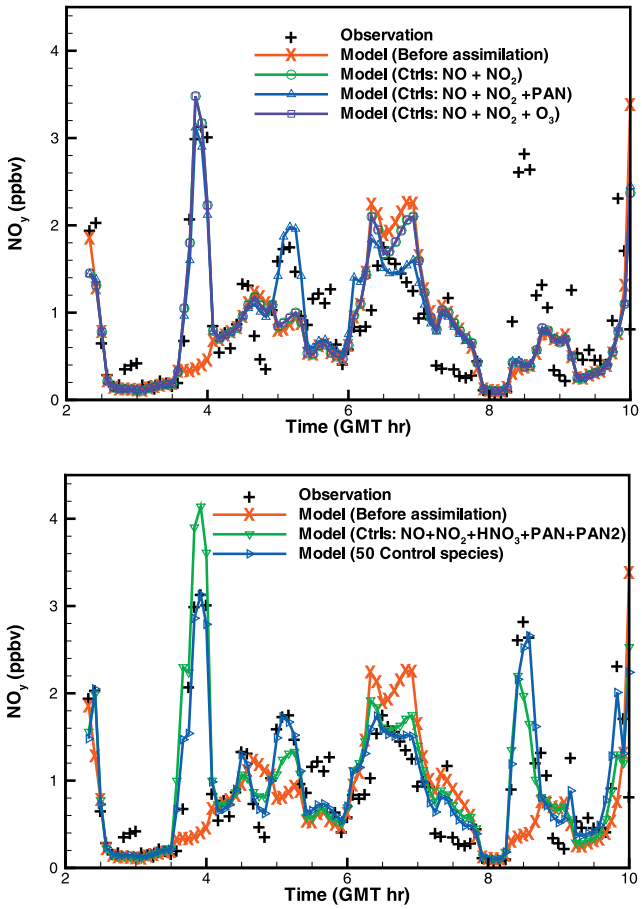

Figure 8. Data assimilation results of using different control variables in assimilating NO_y. NO_y observation uncertainty is assigned as 18%.

affects the specific predictions is located around the target region, i.e., where the P3-B O_3 measurements were made. The other part (B in Figures 6 and 7) is located to the north of the target, with an altitude around 1 km where the wind direction was mainly due south. In region B, $|\lambda(x, t)|$ at 0000 UT contributes most to the influence function $\mathcal{C}(x)$ in equation (13). However, the DC-8 flew past the region around 0400 UT. It is speculated that if the DC-8 had flown the region at an earlier time, the observations would be more beneficial to the model predictions.

[23] For the previous tests that assimilated O_3 measurements, only the initial O_3 concentrations were adjusted. As a complete chemistry mechanism is included in STEM-2K1, changes of other species at initial time would also affect O_3 predictions via the model equations. An almost identical result was found when we tried to assimilate O_3 measurements by adjusting initial concentrations of 55 species. This is not a general finding as assimilating measurements of other species often requires adjusting initial concentrations of more than one species, which will be shown later. The good O_3 assimilation results by only adjusting the initial O_3 concentrations is due to the fact that the air masses sampled by the aircraft for these flights were typically several days old [*Tang et al.*, 2004]. In regions where photochemical O_3 production dominates, the initial concentrations of its precursors, i.e., NO_x (simple combinations of oxygen and nitrogen atoms) and VOCs (volatile organic compounds), will have to be adjusted as well. In the following section, we will present the results from assimilating the P3-B NO_y observations.

3.3. NO_y Assimilation

[24] Total reactive nitrogen (NO_y) was measured on board the P3-B and provides another important example of data assimilation. The measurement of the integrated quantity NO_y represents some very important species in the chemistry of the troposphere. It is also an interesting species from a data assimilation perspective because the species comprising NO_y represent a spectrum of lifetimes from short to long lived. For instance, the oxides of nitrogen (NO, NO_2) play a key role in tropospheric ozone formation and have lifetimes on the order of hours, while some organic nitrates can have lifetimes on the order of months. Furthermore, some species have distinct diurnal cycles, while others do not. Assimilation of NO_y provides the possibility of improving the predictions of its underlying components via the model equations. Below we present results from the assimilation of the NO_y observations. The NO_y measurements were obtained by Y. Kondo's group at the University of Tokyo.

3.3.1. Assimilation

[25] For NO_y measurements by the NASA P3-B aircraft, the model counterpart is calculated from the modeled species as

$$\begin{aligned}[NO_y] = {} & [NO] + [NO_2] + [NO_3] + 2 \times [N_2O_5] + [HONO] \\ & + [HNO_3] + [HNO_4] + [RNO_3] + [PAN] + [PAN2] \\ & + [PBZN] + [MA\_PAN] \end{aligned} \quad (14)$$

where "[ ]" denotes the concentration, and the species refer to the SAPRC-99 chemical mechanism (PAN, peroxy acetyl nitrate; RNO_3, lumped organic nitrates; PAN2, peroxypropionyl nitrate and other higher-alkyl PAN analogues; PBZN, PAN analogues formed from aromatic aldehydes; MA_PAN, PAN analogue formed from methacrolein), and full species represented are given by *Carter* [2000].

[26] An attractive feature of 4D-Var is illustrated by the assimilation of NO_y. In 4D-Var, the chemical and physical processes (deposition and diurnal variation of sunlight, etc.) that partition NO_y among the various species are directly taken into account. Below we illustrate how the choice of control variables influences the assimilation results. As our first test involving the assimilation of NO_y measurements, we chose initial concentrations of the short-lived species NO and NO_2 to adjust. Model predictions of NO_y before and after the data assimilation are shown in Figure 8, as well as the observations. The significant discrepancy between model predictions and measured NO_y between 0330 UT and 0400 UT (1130 and 1200 Hong Kong local time) is corrected after assimilation. The other noticeable improvement of the NO_y predictions happen between 0600 UT and 0700 UT, as well as at the takeoff and landing times. There are still differences between the model and observations, especially around 0830 UT, and during 0430–0600 UT.

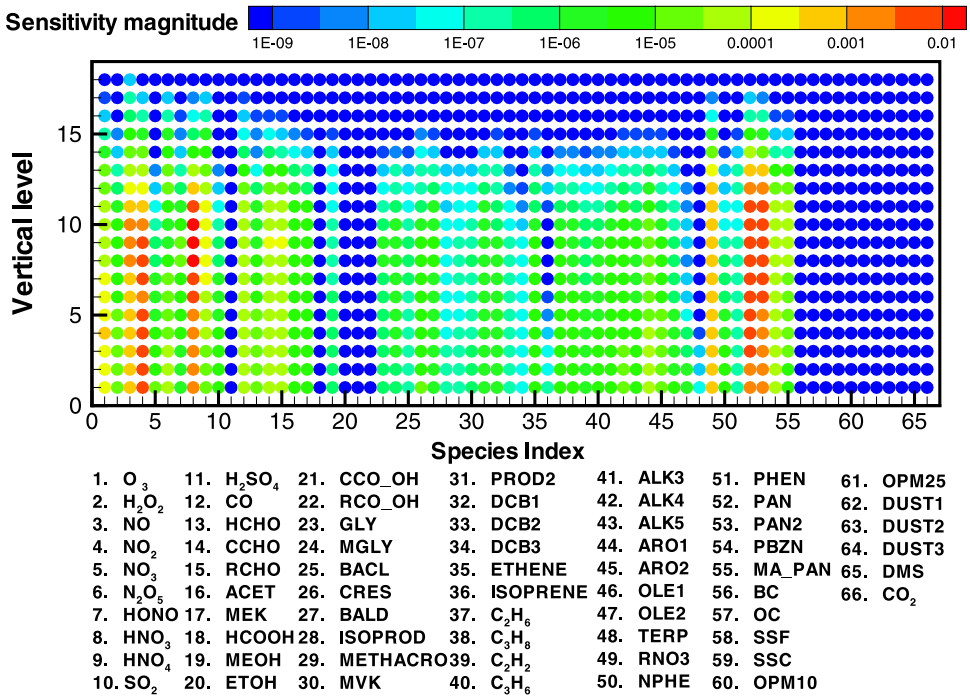

Figure 9. Sensitivity of NO_y predictions with respect to initial conditions at different levels for 66 prognostic species. Averaged magnitudes of the gradients $\partial J/\partial \ln c_0$ are shown, where J measures the difference between the NO_y observations and the corresponding model outputs as defined in equation (5) and c_0 is the initial concentration. The key species that affect the NO_y predictions are NO, NO_2, HNO_3, PAN, and PAN2 for this case. Species 56–66 are tracers in the current model. H_2SO_4, HCOOH, ETOH, CCO-OH, and RCO-OH are reaction products in the SAPRC-99 chemical mechanism. Full species represented and reaction details are given by *Carter* [2000].

[27] Two other cases, each by adjusting one additional species, are also shown in Figure 8. The addition of PAN to the control variables helps the model to produce NO_y that better match observations around 0515 UT and during 0600–0700 UT. This is due to the fact that HNO_3 and PAN dominated NO_y during the TRACE-P period [*Talbot et al.*, 2003; *Miyazaki et al.*, 2003]. On the other hand, adding O_3 does not help the data assimilation much. Cases also

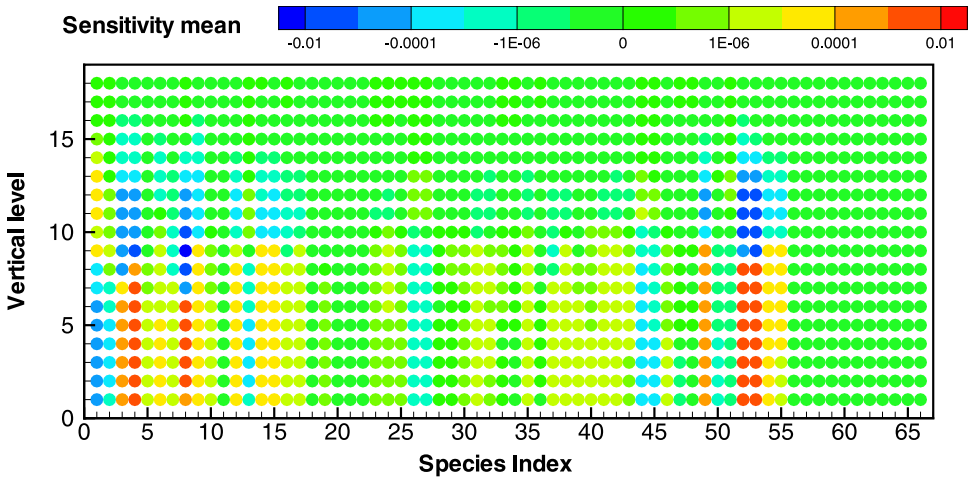

Figure 10. Sensitivity of NO_y predictions with respect to initial conditions at different levels for 66 prognostic species. Mean gradients $\partial J/\partial \ln c_0$ are shown (blue depicts negative values, and red depicts positive values), where J measures the difference between the NO_y observations and the corresponding model outputs as defined in equation (5) and c_0 is the initial concentration. The initial concentrations of NO, NO_2, HNO_3, PAN, and PAN2 at the lower levels need to be reduced to better match the measurements. See Figure 9 for names of the species.

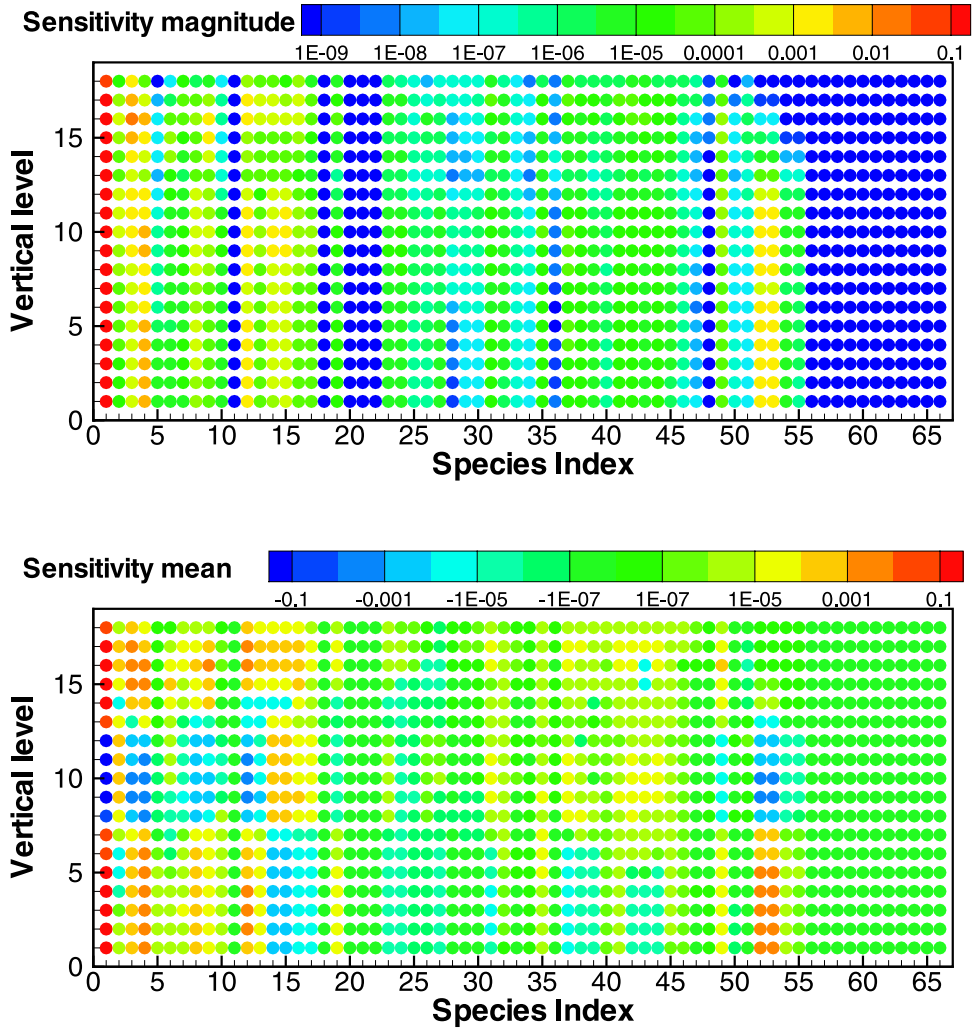

Figure 11. Sensitivity of O_3 predictions with respect to initial conditions at different levels for 66 prognostic species. (top) Averaged magnitudes and (bottom) means (blue depicts negative values, and red depicts positive values) of gradients $\partial J/\partial \ln c_0$ are shown, where J measures the difference between the O_3 observations and the corresponding model outputs as defined in equation (5) and c_0 is the initial concentration. The model prediction of O_3 is more sensitive to its own initial concentrations than those of the other species. See Figure 9 for names of the species.

shown in Figure 8 adjusted more species. One case that included the initial NO, NO_2, HNO_3, PAN, and PAN2 as control variables is able to regenerate the NO_y peak value at 0830 UT. Note that the adjustment of different species is independent of each other; that is, no a priori assumption is made on the relative contribution of the separate species to NO_y. In another test, the initial concentrations of 50 prognostic species were adjusted (16 other prognostic species in STEM-2K1 are not involved in the NO_y predictions in the SAPRC-99 mechanism; see caption of Figure 9 for the list). The results of this case show the best agreement between the predictions and assimilated observations.

3.3.2. Adjoint Sensitivity Analysis

[28] In theory, we can always choose the concentrations of all species at the initial time as control variables if an adequate preconditioning is performed before or during the minimization. However, the large number of control variables will make the full background error covariance estimate B in equation (5) impractical to implement even when statistics are available. Eliminating the species that have minimal effect on the model predictions of certain observations from the control variables improves the conditioning in the minimization, and makes it feasible for the future applications of 4D-Var data assimilation that implements the background error correlation. In the case that the initial concentrations of 50 species are chosen to be adjusted, the number of control variables is $90 \times 60 \times 18 \times 50 = 4{,}860{,}000$. The implementation is made possible by assuming B in equation (5) to be diagonal.

[29] The key species that affect the model predictions of certain measurements can be effectively determined using the 4D-Var data assimilation system. To seek those key

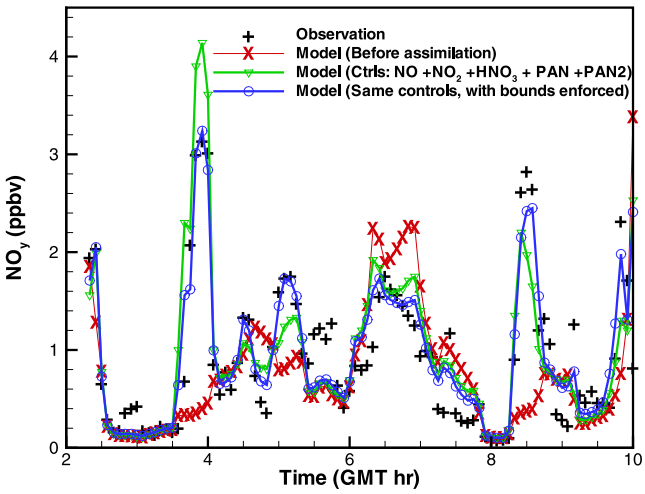

Figure 12. Data assimilation results show the effect of enforcing upper bounds through L-BFGS-B in 4D-Var. NO_y observation uncertainty is assigned as 18%.

species, we look at the gradients of the cost functional, i.e., $\partial J/\partial \ln c_0$, which is equal to $\lambda \cdot c_0$ and can be easily obtained through the adjoint model in the current 4D-Var data assimilation system. Here the cost functional is defined as equation (5), measuring the difference between the measurements and model outputs in a least square form. For instance, the 5 species chosen in the previous case to assimilate NO_y measurements are based upon such sensitivity analysis. Figure 9 shows the horizontally averaged magnitudes of the gradients for each prognostic species at 18 model vertical levels (the vertical grid spacing increases with height, e.g., the altitude of level 5 is 1.0 km and of level 10 is 3.5 km). The averaged magnitudes, calculated by averaging the absolute values of the gradients, indicate which species and at what levels would changes in initial values most drastically affect the model predictions of the measurements that are used to construct the cost functional. On the basis of Figure 9, NO, NO_2, HNO_3, PAN, and PAN2 are the key species that affect the NO_y predictions for the conditions simulated.

[30] The mean value of $\partial J/\partial \ln\ c_0$ on each level is displayed in Figure 10. This metric identifies whether the overall increases (when $\partial J/\partial \ln c_0 < 0$) or decreases (when $\partial J/\partial \ln c_0 > 0$) in the initial concentrations of individual species on a certain level are needed to reduce the cost functional, i.e., to improve the model predictions. Figure 10 shows that $\partial\ J/\partial\ \ln\ c_0$ is mostly positive for NO, NO_2, HNO_3, PAN, and PAN2 at the lower levels. This implies that we need to reduce the initial concentrations of these species to better match the measurements. Around level 10, which is 3.5 km, the aforementioned components need to increase their concentrations to best match to the observations. These results are consistent with the model predictions of NO_y before assimilation shown in Figure 8. The model tends to underestimate NO_y at higher altitudes and overestimate (slightly) at lower altitudes. The underestimation of NO_y at higher altitudes is likely due to a misplacement of the biomass burning air masses from Southeast Asia as discussed earlier.

[31] Figure 11 shows the averaged magnitude and mean of the sensitivity, with the cost functional representing the differences between the P3-B O_3 measurements and their model counterparts. We see that the model predictions of O_3 along the P3-B flight track are much more sensitive to the initial O_3 field than any other species. This explains why we were able to assimilate the O_3 measurements well by only adjusting initial O_3 concentrations. In general, photochemically active species such as NO_2 are tightly connected to many species. For instance, similar analysis applied to NO_2 measurements shows that initial O_3, CO, and NO affect NO_2 as much as its own initial condition. It is also important to note that the model prediction of O_3 is very sensitive to the initial O_3 mixing ratios at the top of the model domain due to the stratospheric O_3 source. A positive mean gradient shown in Figure 11 implies that a reduction of the O_3 mixing ratios at the top would bring the model prediction of O_3 closer to the P3-B O_3 measurements. With a negative gradient in the middle vertical layers, where the P3-B flew most of the time, the initial O_3 values are underestimated (Figure 5 shows the model predictions are generally lower than the measurements). At lower altitudes when the P3-B flew around 0700 UT and 0900 UT, the measurements are lower than predicted. Figure 11 displays positive mean gradients at lower levels, suggesting the initial O_3 values need to be reduced in order to produce a better match between the model and measurements.

3.3.3. Minimization Algorithm

[32] In 4D-Var, the L-BFGS algorithm is often chosen to update the control variables en route to the minimization of the cost functional. Without the need to save a full Hessian matrix, it proves to be very effective in optimization problems that have a large number of control variables. However, the minimization algorithm is not immune to the generation of physically or chemically unrealistic points in its line search procedure. For instance, if the initial mixing ratio c_0 is chosen as the control variable, the line search cannot guarantee to generate all positive components at the next point unless such a requirement is explicitly stated. Negative mixing ratios can be avoided by choosing $\ln (c_0/c_r)$ as control variable, which ensures positive c_0 by the relation

$$c_0 = c_r e^{\ln \frac{c_0}{c_r}}. \tag{15}$$

However, the line search (solely a mathematical procedure) can still end up with some unrealistic values in its pursuit of the optimal solution, such as choosing 1000 ppbv O_3 mixing ratio in the troposphere. Such points are destined to be thrown away since they differ too much from the background, and the forward model integration using them would produce results far away from the real measurements. Thus a large cost functional would be generated and the line search has to restart from the previous spot. Normally the iteration number reported in the optimization does not count such miscues although it is costly to waste time-consuming CTM forward integrations and backward adjoint runs. In the current study, as described before, we set the maximum forward-backward integrations to be 20.

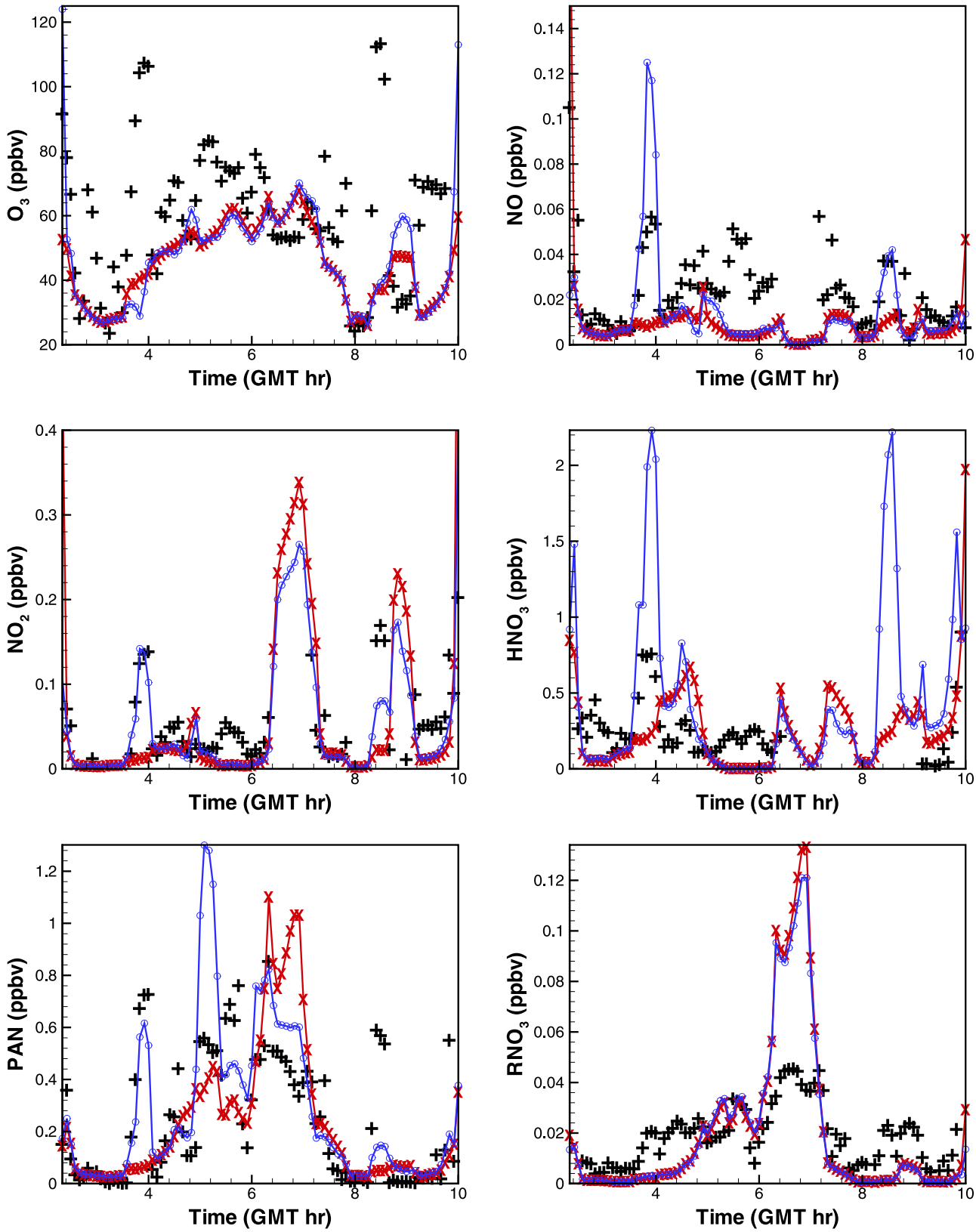

Figure 13. Model predictions before (shown in red) and after (shown in blue) assimilation of the P3-B NO_y measurements. The measurements of O_3, NO, NO_2, HNO_3, PAN, and RNO_3 (shown as pluses) were not assimilated, and their assigned uncertainties are 8%, 20%, 20%, 100%, 100%, and 100%, respectively.

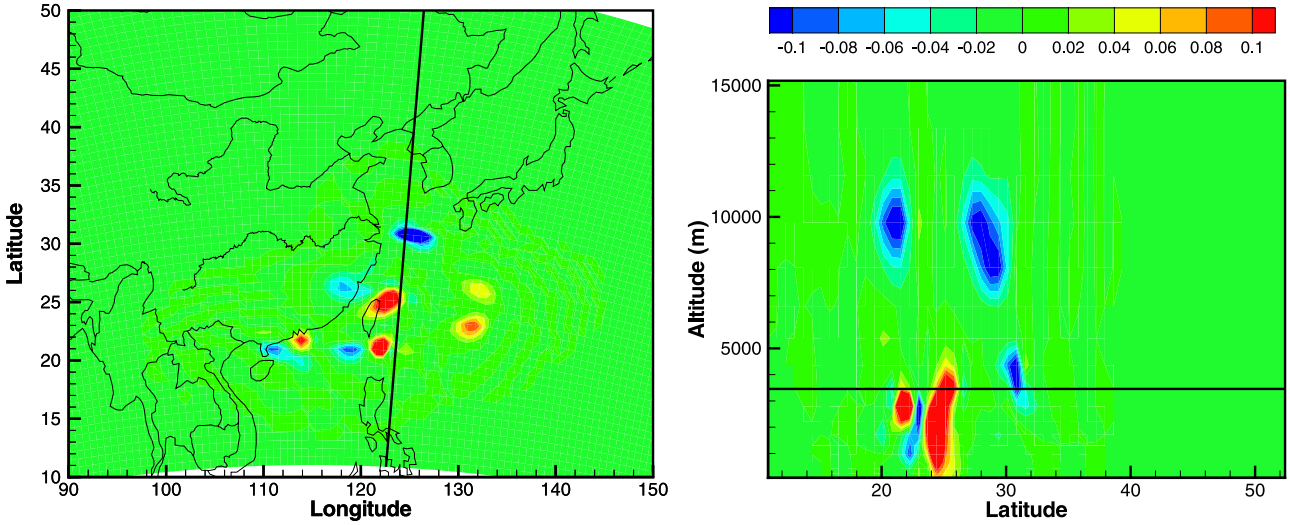

Figure 14. Relative changes in initial O_3 after assimilating the P3-B NO_y measurements. (left) At z = 3.5 km; (right) vertical cross section. The lines in the plots indicate the location of the other cross section.

[33] Additional information about the expected range of the chemical species mixing ratios can be effectively included into the minimization using the L-BFGS-B algorithm (the last "B" of L-BFGS-B stands for "bounds." The algorithm is able to solve large-scale bound-constrained or unconstrained optimization). The upper bounds for each species at every level can be specified. For instance, upper limits to the NO_2 mixing ratio at the bottom and top levels (z = 0.075 km, 15.179 km) were set to be 30 ppbv and 0.5 ppbv, respectively, while the O_3 upper bounds were set to 200 ppbv throughout the domain (these values were determined on the basis of the TRACE-P observations). Figure 12 shows the NO_y assimilation results using the 5 aforementioned species as control variables, when the upper limit of these mixing ratios were enforced through the upper bounds in the L-BFGS-B algorithm. The previous test using the same controls, but without the upper bounds, is also shown for comparison. The benefit of adding the upper bounds through L-BFGS-B algorithm is clearly shown in the figure. In fact, the new result is very comparable to the one using 50 control variables shown in Figure 8. Applying the upper bound constraint for the case using 50 control variables shows slight improvement as well. For these L-BFGS-B implementations, about 0.3% of the total upper bounds were enforced during the Cauchy search in the optimization [*Zhu et al.*, 1997]. It should be noted that to minimize with L-BFGS and then apply the upper limit constraint by replacing those large

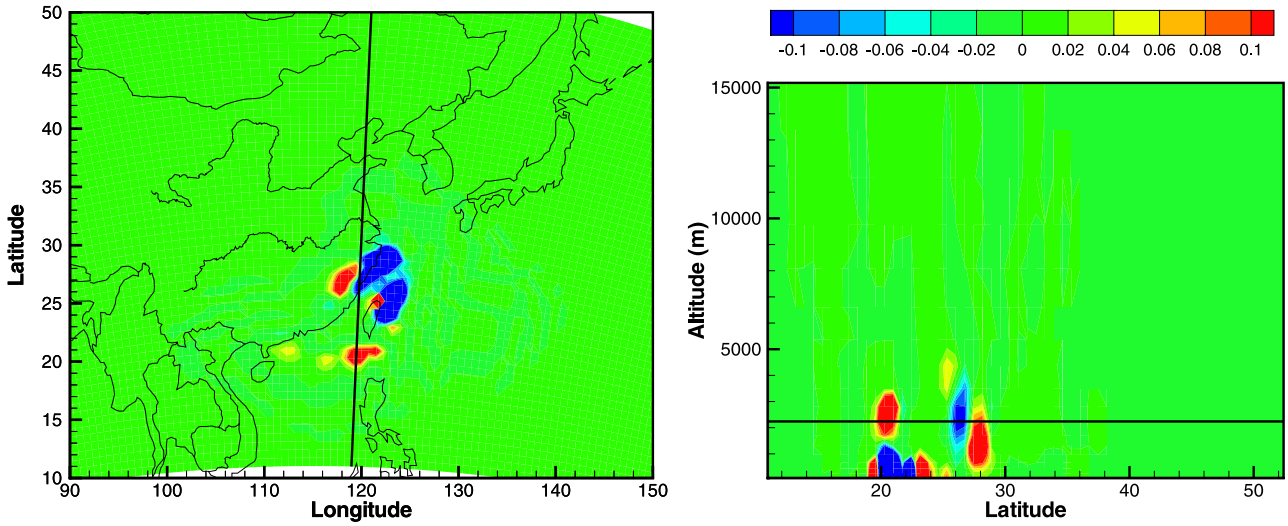

Figure 15. Relative changes in initial NO_2 after assimilating the P3-B NO_y measurements. (left) At z = 2.2 km; (right) vertical cross section. The lines in the plots indicate the location of the other cross section.

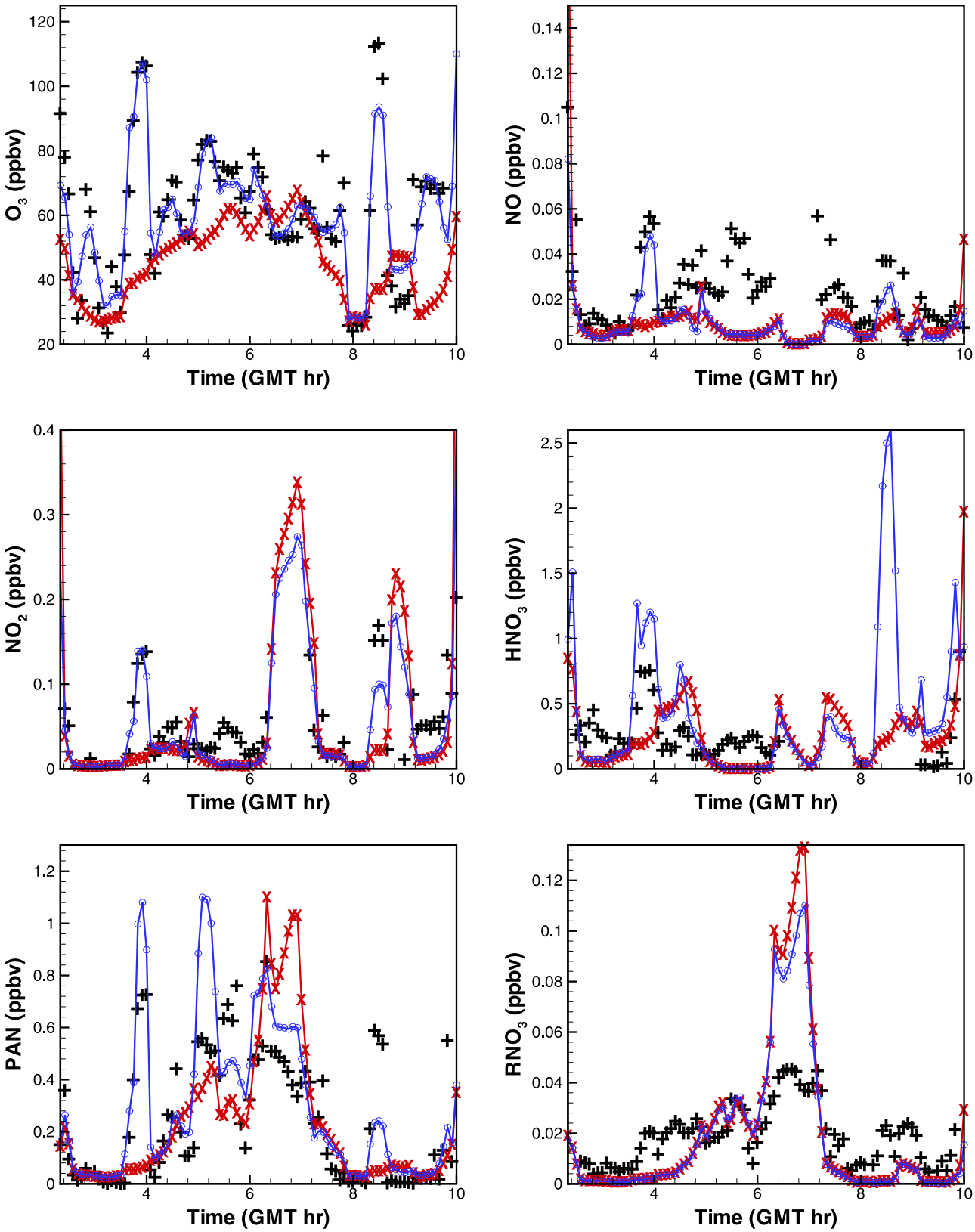

Figure 16. Model predictions before (red) and after (blue) assimilating the P3-B measurements (pluses) of O_3, NO, NO_2, HNO_3, PAN, and RNO_3. Observation uncertainties assigned for O_3, NO, NO_2, HNO_3, PAN, and RNO_3 are 8%, 20%, 20%, 100%, 100%, and 100%, respectively.

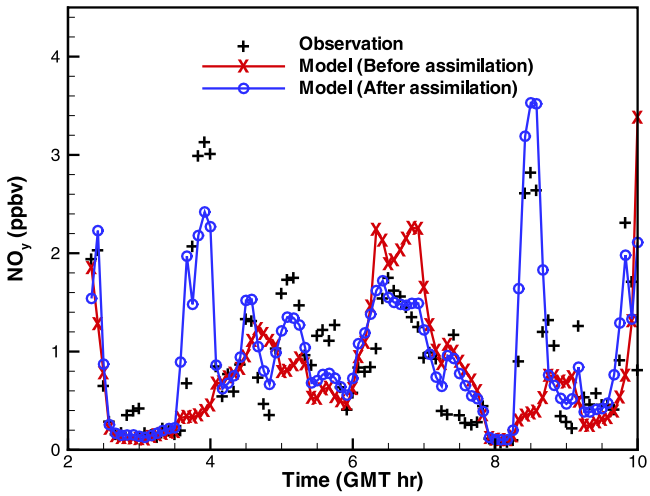

Figure 17. Model predictions of NO_y before and after assimilating O_3, NO, NO_2, HNO_3, PAN, and RNO_3 measurements. The NO_y measurements were not used in the assimilation test. NO_y observation uncertainty is assigned as 18%.

values is not equivalent to enforcing the upper bound inside the L-BFGS-B algorithm.

3.3.4. Model Response to NO_y Assimilation

[34] In section 3.2 we showed that improvement of the model predictions by assimilating O_3 measurements extends hundreds of kilometers away from the measurements. In fact, assimilating measurements of one kind can have positive effects on many other closely related species. Figure 13 shows the model predictions of NO, NO_2, HNO_3, PAN, PAN2, and RNO_3, before and after assimilating the P3-B NO_y measurements. The measurements of these species by the P3-B are also plotted to evaluate the assimilation effect, but they were kept from the data assimilation test. The data assimilation presented here uses 50 control species and applies the L-BFGS-B algorithm.

[35] The model predictions of NO, NO_2, and PAN significantly improve around 0400 UT after assimilating the NO_y observations. This is consistent with the fact that the large discrepancy between model predicted and measured NO_y between 0330 UT and 0400 UT is reduced in the data assimilation, as displayed in Figure 8 and Figure 12. The model results of NO_y also change greatly at around 0830 UT after assimilation, as shown in Figure 12. The 4D-Var determines what form of NO_y need to be adjusted by enforcing all the physical and chemical processes through the model. For instance, the diurnal cycles of NO and NO_2 play an important role in the adjustment. In this case the assimilation of NO_y resulted in the close agreement between model predictions of NO after assimilation and the actual NO measurements by the P3-B which were not assimilated. Another significant improvement made by assimilating NO_y is on the PAN prediction during 0500–0700 UT. Under these conditions the model determined the NO_y to be in the form of PAN. The model results do not change much at the other times. For species that do not contribute to NO_y directly under these conditions, such as O_3, locally (along the P3-B flight track) the model has little response. There is almost no change of RNO_3 predictions after assimilating NO_y measurements for this case.

[36] After assimilating the P3-B NO_y measurements, the adjustments on the initial conditions of the chosen 50 control species are quite different. The most significant relative changes over the original initial values are on O_3, NO_2, HNO_3, and PAN. For each species,

$$\mathcal{R}(x) = \frac{c_0(x) - c_b(x)}{\overline{c_b(x)}} \tag{16}$$

is calculated, where $\overline{c_b(x)}$ is the average of the background mixing ratios. Figures 14 and 15 show $\mathcal{R}(x)$ of O_3 and NO_2, respectively. While the largest relative changes are located around the flight region, the adjustment of the initial conditions reaches much further. The changes are not simple increases or decreases, and show various patterns for different species. While the patterns share more similarity between NO_2, HNO_3, and PAN, that of O_3 is distinctive in that it reaches more eastward and further in height, due in part to the long lifetime of ozone in the upper troposphere and the contributions of stratospheric/tropospheric exchange processes. As Figure 3 shows westerly winds at z = 3.5 km in the P3-B flight region, the adjustments on the initial conditions of NO_2, HNO_3, and PAN shift to the west of the P3-B flight route, reflecting the transport effect.

3.4 Other Assimilation Tests

[37] In the previous tests we focused on the effect of assimilating measurements of a single parameter in the CTM analysis. However, it is important to note that different combinations of the components can lead to the same NO_y value. Thus no unique solution is guaranteed only on the basis of the assimilation of NO_y measurements. The chemistry and physical processes of the model provide the guidance for the partitioning of NO_y among the various species. The simultaneous assimilation of multiple species should provide additional constraints. In this section, we present results from simultaneously assimilating O_3, NO, NO_2, HNO_3, PAN, and RNO_3 measurements on board the P-3B (see *Eisele et al.* [2003] for more measurement information). The initial concentrations of the same 50 prognostic species, as in the NO_y assimilation, were used as control variables. Upper bounds were enforced in the L-BFGS-B algorithm.

[38] Figure 16 shows measurements and the predictions of the 6 assimilated species. The overall prediction improved when adding additional assimilated species. In the first half of the assimilation time window the improvements over model predictions are visible for all species except for RNO_3. In the second half, the improvement is not as strong, but still shows the positive effect of assimilation. Since the O_3 measurements are the most accurate, with an estimated uncertainty of 8%, it is the major component in the cost functional defined in equation (5). The minimization of the cost functional then drives the model predictions to match O_3 measurements as the primary task. The uncertainties for NO and NO_2 are 20%, much smaller than the uncertainties assumed for HNO_3, PAN, and RNO_3, which are 100%. This leads to slightly better agreement between

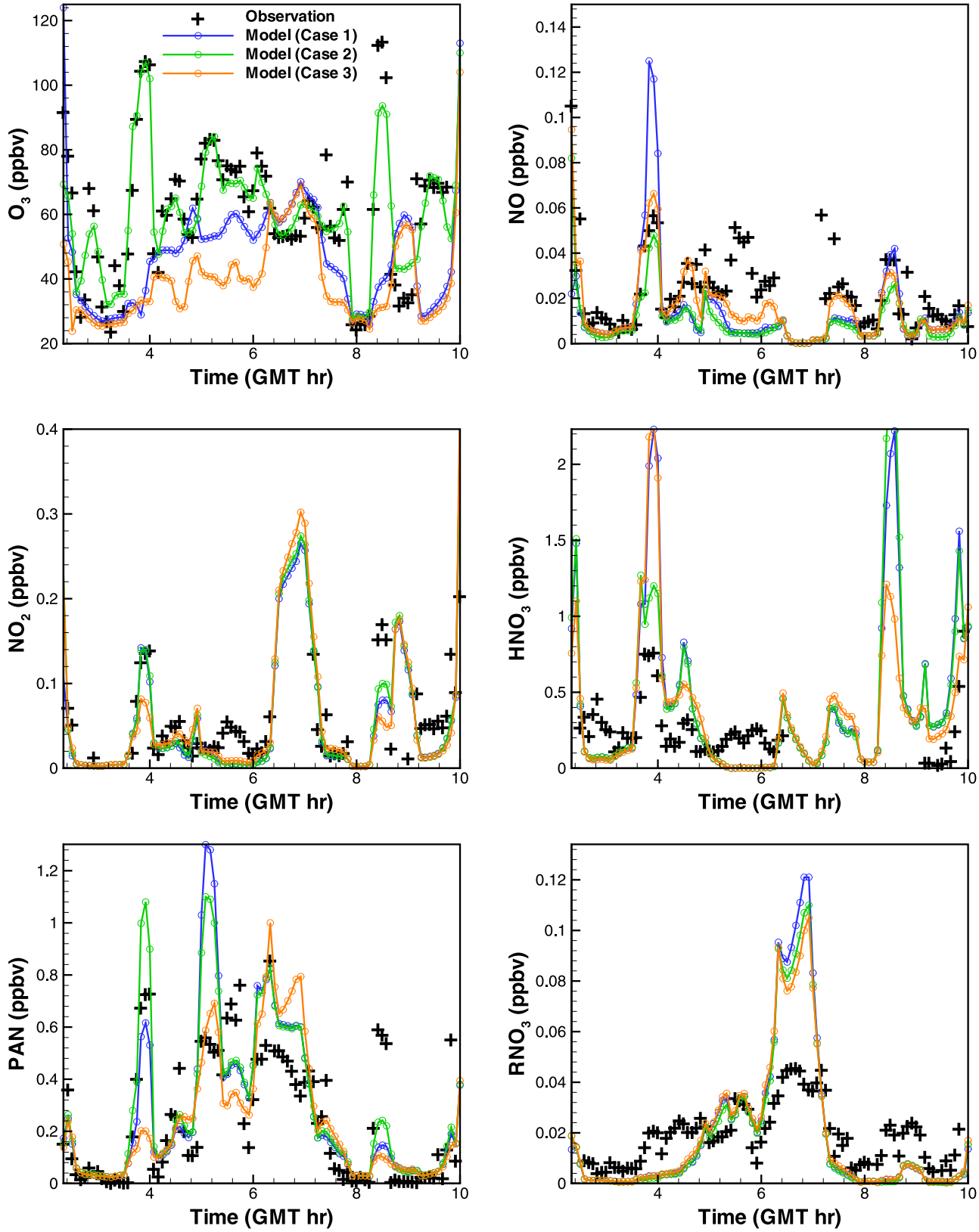

Figure 18. Comparison of model predictions between three assimilation tests. The measurements that were assimilated: case 1, NO_y; case 2, O_3, NO, NO_2, HNO_3, PAN, and RNO_3 (also shown in Figure 16); case 3, NO_y, NO, NO_2, HNO_3, PAN, and RNO_3. All measurements are from the P3-B. Observation uncertainties assigned for O_3, NO, NO_2, NO_y, HNO_3, PAN, and RNO_3 are 8%, 20%, 20%, 18%, 100%, 100%, and 100%, respectively.

model and observations for NO and NO_2 than for HNO_3, PAN, and RNO_3 after assimilation.

[39] Figure 17 shows the NO_y calculated by the model after the simultaneous assimilation of observed O_3, NO, NO_2, HNO_3, PAN, and RNO_3, compared to the observed values of NO_y. Since the NO_y measurements were not used in the assimilation and are independent of the observations that were assimilated, it also serves as a validation test. The results demonstrate that the assimilation of various observations on species comprising NO_y significantly improves the model predictions of NO_y (as one would hope). Unlike the early validation test using measurements from two different flights, here the withheld NO_y observations and the assimilated measurements from the same flight share the same spatial and temporal sampling patterns.

[40] An additional simulation was performed where we simultaneously assimilated measured NO_y and NO, NO_2, HNO_3, PAN, and RNO_3. This assimilation also produced slightly different results. A comparison of the effect of assimilation of NO_y only, assimilation of O_3, NO, NO_2, HNO_3, PAN, and RNO_3 together (also shown in Figure 16), and the assimilation of NO_y along with NO, NO_2, HNO_3, PAN, and RNO_3 is shown in Figure 18. As discussed above assimilation of NO_y by itself does not provide a unique solution to the adjustment of the various species contributing to NO_y. Adding additional species in theory adds determinism. Adding some of the individual components and the integral quantity NO_y can supply conflicted pieces of information. For example, contributions of PAN to NO_y just before 0400 UT and at ~0500 UT are significantly different for the three cases. The fact that the results are not consistently improved with more information may reflect a systematic error in the forward model (e.g., in regard to a deficiency in the chemical mechanism), correlations between measurement errors (not accounted for in these results), and/or errors in other parameters not adjusted in these initial-condition-only adjustments (e.g., emissions). Further studies are ongoing into these aspects and will be the subject of future papers.

4. Summary

[41] We presented variational data assimilation analyses using TRACE-P flight measurements and STEM-2K1 chemical transport model. By adjusting certain initial chemical fields, the CTM can generate results that closely match the observations. Using the adjusted initial field after assimilating O_3 measurements, the agreement between model predictions and measurements by the other independent flight improves. It is also found that the model predictions of NO, NO_2, and PAN better match the measurements when only NO_y observations are assimilated. The case that assimilates measurements of different species shows the effect of different observations used in the assimilation highly depends on their measurement uncertainties.

[42] In this study we also showed the benefit of adjoint sensitivity analysis in identifying the key control parameters to improve the model-observation agreement. To apply the 4D-Var technique using CTMs, it is helpful to include additional information into the minimization procedure. The upper bounds of mixing ratios provided to the L-BFGS-B algorithm lead to a speedup of the 4D-Var optimization.

[43] The error statistics of the initial field used in our STEM-2K1 is absent. The knowledge of such information is crucial to provide the optimal analysis in the future studies. Other than the initial field, the emission inventories and boundary conditions are far from accurate as well. In addition, the meteorological fields, physical parameters such as diffusion, and chemical reaction rates are also uncertain. Complete calibration of these uncertainties is a daunting task that can only be pursued step by step. Nevertheless, a great promise of data assimilation is shown in its ability to blend the real measurements with our accumulated knowledge of atmospheric chemistry through CTMs.

[44] In this paper we have demonstrated how the 4D-Var method can be used to assimilate aircraft data. The technique can be easily extended to assimilate observations from a variety of data sources (e.g., we are now assimilating ozone data from aircraft, surface networks, lidar, and ozonesondes). The integration of modeled and measured quantities is critical in the interpretation of observational data such as those obtained during intensive field campaigns, and to more completely utilize the increasing amount of operational atmospheric chemical and aerosol data provided by satellite and monitoring networks. In addition this information can be used to help identify model deficiencies, and thus lead to better models, and ultimately an improvement in our capability to predict air quality. Further efforts focused on improving the forward model are planned and needed.

[45] Finally, results from such analysis studies should also provide important insights into experimental design. The 4D-Var framework allows us to assess the impact on the assimilation results of adding additional observations of different species, and different combinations, as well as exploring issues related to time resolution (e.g., using 10-s versus 3-min average values in the assimilation), and spatial sampling.

[46] **Acknowledgments.** The authors gratefully thank the National Science Foundation for support through award NSF ITR AP&IM 0205198. This work was also supported by NASA GTE TRACE-P, ACMAP, and the NSF Atmospheric Chemistry Program. The work of A. Sandu was partially supported by award NSF CAREER ACI 0093139.

References

Benjamin, S. G., et al. (2004), An hourly assimilation-forecast cycle: The RUC, *Mon. Weather Rev.*, *132*(2), 495–518.

Byrd, R., P. Lu, and J. Nocedal (1995), A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Stat. Comput.*, *16*(5), 1190–1208.

Carmichael, G. R., D. N. Daescu, A. Sandu, and T. Chai (2003a), Computational aspects of chemical data assimilation into atmospheric models, in *Computational Science: ICCS 2003*, *Lect. Notes Comput. Sci.*, vol. 4, pp. 269–278, Springer, New York.

Carmichael, G. R., et al. (2003b), Regional-scale chemical transport modeling in support of the analysis of observations obtained during the TRACE-P experiment, *J. Geophys. Res.*, *108*(D21), 8823, doi:10.1029/2002JD003117.

Carter, W. (2000), Documentation of the SAPRC-99 chemical mechanism for VOC reactivity assessment, *Tech. Rep. 92-329*, Calif. Air Resour. Board, Sacramento.

Clerbaux, C., J. Hadji-Lazaro, D. Hauglustaine, G. Megie, B. Khattatov, and J. F. Lamarque (2001), Assimilation of carbon monoxide measured from satellite in a three-dimensional chemistry-transport model, *J. Geophys. Res.*, *106*(D14), 15,385–15,394.

Daescu, D. N., and G. R. Carmichael (2003), An adjoint sensitivity method for the adaptive location of the observations in air quality modeling, *J. Atmos. Sci.*, *60*, 434–449.

Daescu, D. N., G. R. Carmichael, and A. Sandu (2000), Adjoint implementation of rosenbrock methods applied to variational data assimilation problems, *J. Comput. Phys.*, *165*(2), 496–510.

Daescu, D. N., A. Sandu, and G. R. Carmichael (2003), Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: II, Numerical validation and applications, *Atmos. Environ.*, *37*, 5097–5114.

Daley, R. (1991), *Atmospheric Data Analysis*, 457 pp., Cambridge Univ. Press, New York.

Eisele, F., et al. (2003), Summary of measurement intercomparisons during TRACE-P, *J. Geophys. Res.*, *108*(D20), 8791, doi:10.1029/2002JD003167.

Elbern, H., and H. Schmidt (1999), A 4D-Var chemistry data assimilation scheme for Eulerian chemistry transport modeling, *J. Geophys. Res.*, *104*(D15), 18,583–18,598.

Elbern, H., and H. Schmidt (2001), Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, *106*(D4), 3569–3590.

Elbern, H., H. Schmidt, and A. Ebel (1997), Variational data assimilation for tropospheric chemistry modeling, *J. Geophys. Res.*, *102*(D13), 15,967–15,985.

Elbern, H., H. Schmidt, O. Talagrand, and A. Ebel (2000), 4D-variational data assimilation with an adjoint air quality model for emission analysis, *Environ. Modell. Software*, *15*, 539–548.

Errera, Q., and D. Fonteyn (2001), Four-dimensional variational chemical assimilation of CRISTA stratospheric measurements, *J. Geophys. Res.*, *106*(D11), 12,253–12,265.

Fisher, M., and D. J. Lary (1995), Lagrangian four-dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc.*, *121*, 1681–1704.

Hanea, R. G., G. J. M. Velders, and A. Heemink (2004), Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model, *J. Geophys. Res.*, *109*, D10302, doi:10.1029/2003JD004283.

Heald, C., et al. (2004), Comparative inverse analysis of satellite (MOPITT) and aircraft (TRACE-P) observations to estimate Asian sources of carbon monoxide, *J. Geophys. Res.*, *109*, D23306, doi:10.1029/2004JD005185.

Hoelzemann, J. J., H. Elbern, and A. Ebel (2001), SAS and 4D-var data assimilation for chemical state analysis by urban and rural observation sites, *Phys. Chem. Earth, Part B: Hydrol. Oceans Atmos.*, *26*(10), 807–812.

Houtekamer, P. L., L. M. Herschel, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen (2005), Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, *Mon. Weather Rev.*, *133*(3), 604–620.

Khattatov, B. V., J. F. Lamarque, L. V. Lyjak, R. M. P. Levelt, X. X. Tie, G. P. Brasseur, and J. C. Gille (2000), Assimilation of satellite observations of long-lived chemical species in global chemistry transport models, *J. Geophys. Res.*, *105*(D23), 29,135–29,144.

Klinker, E., F. Rabier, G. Kelly, and J. F. Mahfouf (2000), The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration, *Q. J. R. Meteorol. Soc.*, *126*(564), 1191–1215.

Lamarque, J. F., and J. C. Gille (2003), Improving the modeling of error variance evolution in the assimilation of chemical species: Application to MOPITT data, *Geophys. Res. Lett.*, *30*(9), 1470, doi:10.1029/2003GL016994.

Lamarque, J. F., B. V. Khattatov, and J. C. Gille (1999), Assimilation of Measurement of Air Pollution from Space (MAPS) CO in a global three-dimensional model, *J. Geophys. Res.*, *104*(D21), 26,209–26,218.

Lamarque, J. F., B. V. Khattatov, and J. C. Gille (2002), Constraining tropospheric ozone column through data assimilation, *J. Geophys. Res.*, *107*(D22), 4651, doi:10.1029/2001JD001249.

Lamarque, J. F., et al. (2004), Application of a bias estimator for the improved assimilation of Measurements of Pollution in the Troposphere (MOPITT) carbon monoxide retrievals, *J. Geophys. Res.*, *109*, D16304, doi:10.1029/2003JD004466.

Lawrence, M. G., et al. (2002), Global chemical weather forecasts for field campaign planning: Predictions and observations of large-scale features during MINOS, CONTRACE, and INDOEX, *Atmos. Chem. Phys.*, *3*, 267–289.

Madronich, S., and S. Flocke (1999), *The Role of Solar Radiation in Atmospheric Chemistry*, pp. 1–26, Springer, New York.

Mahfouf, J. F., and F. Rabier (2000), The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics, *Q. J. R. Meteorol. Soc.*, *126*(564), 1171–1190.

Menut, L. (2003), Adjoint modeling for atmospheric pollution process sensitivity at regional scale, *J. Geophys. Res.*, *108*(D17), 8562, doi:10.1029/2002JD002549.

Menut, L., R. Vautard, M. Beekmann, and C. Honor (2000), Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model, *J. Geophys. Res.*, *105*(D12), 15,379–15,402.

Miehe, P., A. Sandu, G. R. Carmichael, Y. Tang, and D. Daescu (2002), A commmunication library for the parallelization of air quality models on structured grids, *Atmos. Environ.*, *36*, 3917–3930.

Miyazaki, Y., Y. Kondo, M. Koike, and TRACE-P Science Team, (2003), Synoptic-scale transport of reactive nitrogen over the western Pacific in spring, paper presented at EGS-AGU-EUG Joint Assembly, Eur. Geophys. Soc., Nice, France, 6–11 April.

Parrish, D., and J. Derber (1992), The National Meteorological Center's spectral statistical-interpolation analysis system, *Mon. Weather Rev.*, *120*(8), 1747–1763.

Pielke, R. A., et al. (1992), A comprehensive meteorological modeling system: RAMS, *Meteorol. Atmos. Phys.*, *49*, 69–91.

Rabier, F., H. Jarvinen, E. Klinker, J. F. Mahfouf, and A. Simmons (2000), The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Q. J. R. Meteorol. Soc.*, *126*(564), 1143–1170.

Sandu, A., D. N. Daescu, and G. R. Carmichael (2003), Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: I. Theory and software tools, *Atmos. Environ.*, *37*, 5083–5096.

Sandu, A., D. Daescu, G. R. Carmichael, and T. Chai (2005), Adjoint sensitivity analysis of regional air quality models, *J. Comput. Phys.*, *204*(1), 222–252.

Schmidt, H., and D. Martin (2003), Adjoint sensitivity of episodic ozone in the Paris area to emissions on the continental scale, *J. Geophys. Res.*, *108*(D17), 8561, doi:10.1029/2001JD001583.

Segers, A. J., A. W. Heemink, M. Verlaan, and M. van Loon (2000), A modified RRSQRT-filter for assimilating data in atmospheric chemistry models, *Environ. Modell. Software*, *15*, 663–671.

Streets, D. G., K. F. Yarber, J.-H. Woo, and G. R. Carmichael (2003), Biomass burning in Asia: Annual and seasonal estimates and atmospheric emissions, *Global Biogeochem. Cycles*, *17*(4), 1099, doi:10.1029/2003GB002040.

Talbot, R., et al. (2003), Reactive nitrogen in Asian continental outflow over the western Pacific: Results from the NASA Transport and Chemical Evolution over the Pacific (TRACE-P) airborne mission, *J. Geophys. Res.*, *108*(D20), 8803, doi:10.1029/2002JD003129.

Tang, Y., et al. (2003a), Influences of biomass burning during the Transport and Chemical Evolution over the Pacific (TRACE-P) experiment identified by the regional chemical transport model, *J. Geophys. Res.*, *108*(D21), 8824, doi:10.1029/2002JD003110.

Tang, Y., et al. (2003b), Impacts of aerosols and clouds on photolysis frequencies and photochemistry during TRACE-P: 2. Three-dimensional study using a regional chemical transport model, *J. Geophys. Res.*, *108*(D21), 8822, doi:10.1029/2002JD003100.

Tang, Y., et al. (2004), Multiscale simulations of tropospheric chemistry in the eastern Pacific and on the U.S. West Coast during spring 2002, *J. Geophys. Res.*, *109*, D23S11, doi:10.1029/2004JD004513.

van Loon, M., P. J. H. Builtjes, and A. J. Segers (2000), Data assimilation of ozone in the atmospheric transport and chemistry model LOTOS, *Environ. Modell. Software*, *15*, 603–609.

Wang, K. Y., D. J. Lary, D. E. Shallcross, S. M. Hall, and J. A. Pyle (2001), A review on the use of the adjoint method in four-dimensional atmospheric-chemistry data assimilation, *Q. J. R. Meteorol. Soc.*, *127*(576), 2181–2204.

Woo, J.-H., et al. (2003), Contribution of biomass and biofuel emissions to trace gas distributions in Asia during the TRACE-P experiment, *J. Geophys. Res.*, *108*(D21), 8812, doi:10.1029/2002JD003200.

Zhang, X. F., A. W. Heemink, H. J. M. L. Janssen, P. H. M. Janssen, and F. J. Sauter (1999), A computationally efficient Kalman smoother for the evaluation of the CH_4 budget in Europe, *Appl. Math. Modell.*, *23*(2), 109–129.

Zhu, C., R. H. Byrd, and J. Nocedal (1997), L-BFGS-B–Fortran routines for large scale bound constrained optimization, *Trans. Math. Software*, *23*(4), 550–560.

G. R. Carmichael, T. Chai, and Y. Tang, Center for Global and Regional Environmental Research, University of Iowa, 429 IATL, Iowa City, IA 52246, USA. (tchai@cgrer.uiowa.edu)

D. N. Daescu, Department of Mathematics and Statistics, Portland State University, P. O. Box 751, Portland, OR 97207-0751, USA.

A. Sandu, Department of Computer Science, Virginia Polytechnic Institute and State University, 660 McBryde Hall (0106), Blacksburg, VA 24061, USA.