JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 108, NO. D15, 4456, doi:10.1029/2003JD003500, 2003

Chemical data assimilation: A case study of solar occultation data from the ATLAS 1 mission of the Atmospheric Trace Molecule Spectroscopy Experiment (ATMOS)

D. J. Lary[1,2]
Data Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

B. Khattatov
National Center for Atmospheric Research, Boulder, Colorado, USA

H. Y. Mussa
Unilever Cambridge Centre/Atmospheric Research Centre, University of Cambridge, Cambridge, UK

Received 12 February 2003; revised 3 April 2003; accepted 18 April 2003; published 7 August 2003.

[1] A key advantage of using data assimilation is the propagation of information from data-rich regions to data-poor regions, which is particularly relevant to the use of solar occultation data such as from the Atmospheric Trace Molecule Spectroscopy Experiment (ATMOS). For the first time, an in-depth uncertainty analysis is included in a photochemical model-data intercomparison including observation, representativeness, and theoretical uncertainty. Chemical data assimilation of solar occultation measurements can be used to reconstruct full diurnal cycles and to evaluate their chemical self-consistency. This paper considers as an example the measurements made by the ATMOS instrument ATLAS 1 during March 1992 for a vertical profile in flow-tracking coordinates at an equivalent potential vorticity (PV) latitude of 38°S. ATMOS was chosen because it simultaneously observes several species. This equivalent PV latitude was chosen as it was where ATMOS observed the atmosphere's composition over the largest range of altitudes. A single vertical profile was used so that the detailed diurnal information that the assimilation utilizes could be highlighted. There is generally good self-consistency between the ATMOS ATLAS 1 observations of O_3, NO, NO_2, N_2O_5, HNO_3, HO_2NO_2, HCN, $ClONO_2$, HCl, H_2O, CO, CO_2, CH_4, and N_2O and between the observations and photochemical theory. *INDEX TERMS:* 0340 Atmospheric Composition and Structure: Middle atmosphere—composition and chemistry; 0394 Atmospheric Composition and Structure: Instruments and techniques; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; *KEYWORDS:* data assimilation, chemical Kalman filter, skill scores, ATMOS, ozone assimilation

Citation: Lary, D. J., B. Khattatov, and H. Y. Mussa, Chemical data assimilation: A case study of solar occultation data from the ATLAS 1 mission of the Atmospheric Trace Molecule Spectroscopy Experiment (ATMOS), *J. Geophys. Res.*, *108*(D15), 4456, doi:10.1029/2003JD003500, 2003.

1. Introduction

[2] For more than two decades, vertical profile measurements of trace gases have been made by solar occultation (e.g., from the Stratospheric Aerosol and Gas Experiment (SAGE), the Halogen Occultation Experiment (HALOE), Polar Ozone and Aerosol Measurement (POAM), the Atmospheric Trace Molecule Spectroscopy Experiment (ATMOS), and the Improved Limb Atmospheric Sounder (ILAS)). The observations are self-calibrating and are therefore useful for the analysis of temporal trends. However, by definition, the use of solar occultation limits the measurement opportunities to satellite sunrise and sunset. It is therefore useful to use chemical data assimilation cast in flow-tracking coordinates to reconstruct full diurnal cycles and to check the chemical self-consistency of the solar occultation measurements. The use of flow-tracking coordinates allows a more complete global view to be obtained from the sparse latitudinal coverage of the solar occultation measurements.

2. Why Use Assimilation?

[3] The question could be asked, How does using chemical data assimilation differ from using reverse domain filling (RDF) isentropic trajectories, where ozone depletion

[1] Also at Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, Maryland, USA.
[2] Also at Department of Chemistry, University of Cambridge, Cambridge, UK.

0148-0227/03/2003JD003500$09.00

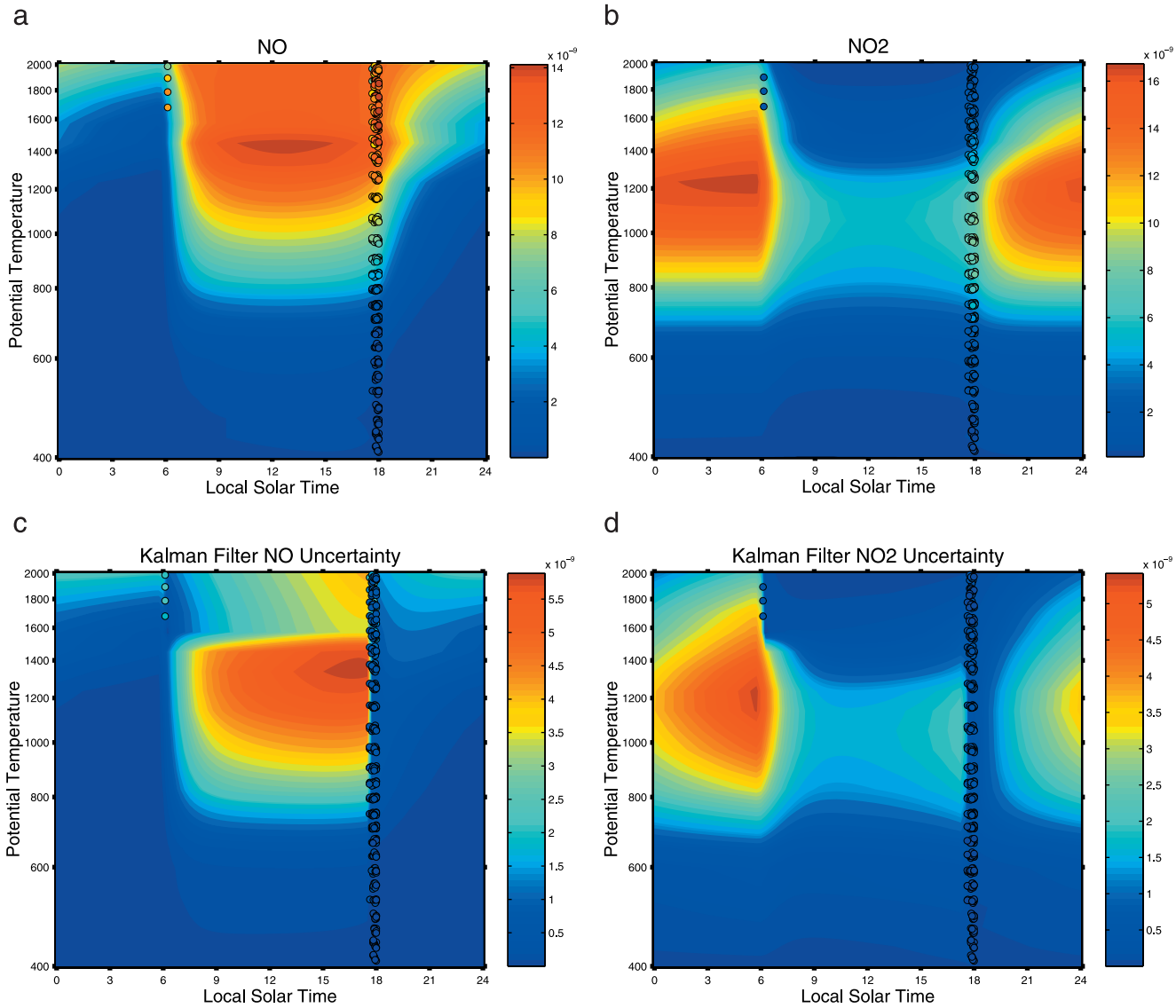

Figure 1. Diurnal cycles of (a and c) NO and (b and d) NO_2 vertical profiles for the 30 March 1992 assimilation of ATMOS ATLAS 1 data at 38°S. (top) The chemical analyses produced by assimilation overlaid with the observations (colored circles). (bottom) The analyses' uncertainties overlaid with the observation uncertainties (colored circles). The reduction in analysis uncertainty can be seen in the bottom plots when new observational information comes in. The dots indicate the locations of the assimilation analysis grid. A single vertical profile was used so that the detailed diurnal information that assimilation utilizes could be highlighted.

is computed along trajectories using aircraft and satellite data or other two-dimensional (2D) and 3D photochemical models? Chemical data assimilation combines the observational information available from measurements with the theoretical information encapsulated into a deterministic model of atmospheric chemistry, together with the associated uncertainties of each. Specifically, assimilation uses (1) constituent observations (in this case, observations of 14 constituents simultaneously); (2) observational uncertainty; (3) theoretical knowledge to propagate information from the observations available to other species and locations; (4) evaluations of chemical self-consistency; and (5) the provision of an uncertainty on the analysis produced that accounts for observational uncertainty, temporal and spatial representativeness uncertainty, and theoretical (model) uncertainty; that is, assimilation requires a specification of the error statistics of the observations and the underlying model.

[4] For example, Figure 1 shows the time evolution of the NO and NO_2 analyses and the analyses' uncertainties. Conventional modeling studies do not present the time evolution of uncertainty associated with their predictions. Very useful as they are, techniques such as using RDF isentropic trajectories or 2D and 3D chemical models are not able to do this.

[5] Assimilation should not be seen as competing with other modeling tools; rather, its purpose is objective improvement by combining observational, theoretical, and uncertainty information formally and statistically in a mathematical framework. The key aim in developing such a powerful method is to look at scientific issues. However,

the credentials of the method must first be rigorously established. The purpose of this paper is to evaluate critically the analyses produced by chemical data assimilation and evaluate this method's skill with objective measures. Subsequent papers will use the method to look at a variety of scientific issues, for example, the role of halogens in the initiation and catalysis of methane oxidation in the upper troposphere/lower stratosphere region and the time evolution and partitioning of nitrogen, chlorine, bromine, and hydrogen species as well as the relative importance of ozone loss catalytic cycles from the start of the UARS period to the present.

3. Flow-Tracking Coordinates

[6] Under adiabatic conditions, air parcels move along isentropic surfaces (surfaces of constant potential temperature, θ). So, when considering tracer fields, θ is a suitable vertical coordinate. *McIntyre and Palmer* [1983, 1984], *Hoskins et al.* [1985], and *Hoskins* [1991] have shown the value of isentropic maps of Ertel's potential vorticity (PV) for visualizing large-scale dynamical processes. PV plays a central role in large-scale dynamics, where it behaves as an approximate material tracer [*Hoskins et al.*, 1985].

[7] As a result, PV can be used as the horizontal spatial coordinate instead of latitude and longitude [*Norton*, 1994; *Lary et al.*, 1995a]. PV is sufficiently monotonic in latitude on an isentropic surface to act as a useful replacement coordinate for both latitude and longitude, reducing the tracer field from three dimensions to two. These ideas have already led to interesting studies correlating PV and chemical tracers such as N_2O and O_3 [*Schoeberl et al.*, 1989; *Proffitt et al.*, 1989; *Lait et al.*, 1990; *Douglass et al.*, 1990; *Proffitt et al.*, 1989, 1993; *Atkinson*, 1993]. A key result of these studies is that PV and ozone mixing ratios are correlated on isentropic surfaces in the lower stratosphere, as was first pointed out by *Danielsen* [1968].

[8] Since the absolute values of PV depend strongly upon height and the meteorological condition, it is useful to normalize PV and use PV-equivalent latitude, ϕ_e, as the horizontal coordinate instead of PV itself. ϕ_e is calculated by considering the area enclosed within a given PV contour on a given θ surface. The ϕ_e assigned to every point on this PV contour is the latitude of a latitude circle which encloses the same area as that PV contour. Therefore for every level in the atmosphere, ϕ_e has the same range of values, $-90°$–$90°$. This provides a vortex-tracking and, indeed, a flow-tracking stratospheric coordinate system.

[9] We have taken these now well-established ideas and used them as a framework for our chemical data assimilation. This approach is valid for our analysis interval of 1 day and often for up to 10 days or longer in the stratosphere. Because a major component of the variability of trace gases is due to atmospheric transport, it makes sense to use a coordinate system that ''follows'' the large-scale flow pattern to perform our data assimilation.

4. ATMOS Observations

[10] As an example of how chemical data assimilation cast in flow-tracking coordinates can be used to extend the value of solar occultation data sets, we consider the ATLAS 1 data from the ATMOS instrument. The ATMOS instrument is an infrared spectrometer (a Fourier transform interferometer) that was designed to study the chemical composition of the atmosphere. The ATMOS instrument has flown four times on the space shuttle, on Spacelab-3 in April 1985, and on ATLAS 1, ATLAS 2, and ATLAS 3 in March 1992, April 1993, and November 1994, respectively. In this study we use all the data from ATLAS 1 for a vertical profile in our flow-tracking coordinates at an equivalent PV latitude, ϕ_e, of 38°S. This equivalent PV latitude was chosen as it was where ATMOS observed the atmosphere's composition over the largest range of altitudes. We wanted to study a single vertical profile so that the detailed diurnal information that assimilation provides could be highlighted. We assimilated ATMOS ATLAS 1 observations of 14 species, namely O_3, NO, NO_2, N_2O_5, HNO_3, HO_2NO_2, HCN, $ClONO_2$, HCl, H_2O, CO, CO_2, CH_4, and N_2O.

[11] The relative role played by the observations of different species in the assimilation is an interesting issue. The top 10 ranking of chemical information content for this vertical situation over an entire diurnal cycle is O_3, CO, HO_2, NO_2, ClO, CH_3OO, HCHO, NO, BrO, and H_2O. It is not at all surprising (in fact, it is rather reassuring) that O_3 comes out as number one. We also see key representatives of carbon, nitrogen, hydrogen, and bromine species, reminding us how coupled atmospheric chemistry is. It is interesting that two of the top 10 are peroxy radicals: number three is HO_2 and number six is CH_3OO. It is also noteworthy that the methyl species are rarely considered in stratospheric studies yet have significant information content.

5. Chemical Assimilation Scheme

[12] Data assimilation enables better use of observations of atmospheric chemistry. An example of the way assimilation can reconstruct a full diurnal cycle from occultation observations is shown in Figure 1. This example will be discussed more later.

[13] The first use of chemical data assimilation was by *Fisher and Lary* [1995]. *Khattatov et al.* [1999] have described the chemical Kalman filter used in this study. Other studies using assimilation include those by *Elbern* [1997], *Elbern et al.* [1997], *Eskes et al.* [1998, 1999], *Khattatov et al.* [1999], *Lary* [1999], *Elbern et al.* [2000], *Khattatov et al.* [2000, 2001], *Wang et al.* [2001], *Smyshlyaev and Geller* [2001], *Clerbaux et al.* [2001], *Errera and Fonteyn* [2001], *Chipperfield et al.* [2002], and *Daescu and Carmichael* [2003]. The photochemical mode used is the extensively validated AutoChem model [*Fisher and Lary*, 1995; *Lary et al.*, 1995b; *Lary*, 1996]. The model is explicit and uses an adaptive-time step, error-monitoring time integration scheme for stiff systems of equations [*Press et al.*, 1992; *Stoer and Bulirsch*, 1980]. AutoChem was the first model to perform 4D variational data assimilation (4D-Var) [*Fisher and Lary*, 1995] and now also includes a Kalman filter [*Khattatov et al.*, 1999]. AutoChem uses kinetic data largely based on the work of *Atkinson et al.* [1997] and *DeMore et al.* [1997], with updates for nitrogen chemistry based on the work of *Brown et al.* [1999a], *Portmann et al.* [1999], and *Brown et al.* [1999b].

[14] A key part of the chemical model is the calculation of photolysis rates. In this study they are calculated using full

spherical geometry and multiple scattering [*Anderson*, 1983; *Lary and Pyle*, 1991a, 1991b; *Meier et al.*, 1982; *Nicolet et al.*, 1982], corrected after *Becker et al.* [2000]. The photolysis rate used for each time step is obtained by 10-point Gaussian-Legendre integration [*Press et al.*, 1992]. These calculations are updated on every assimilation iteration to ensure that the improved ozone profile at a given equivalent latitude is used to calculate the photolysis rates.

[15] The chemical system under study here contains a total of 60 species. Fifty-five species are time integrated, namely $O(^1D)$, $O(^3P)$, O_3, N, NO, NO_2, NO_3, N_2O_5, HONO, HNO_3, HO_2NO_2, CN, NCO, HCN, Cl, Cl_2, ClO, ClOO, OClO, Cl_2O_2, $ClNO_2$, ClONO, $ClONO_2$, HCl, HOCl, CH_3OCl, Br, Br_2, BrO, $BrONO_2$, BrONO, HBr, HOBr, BrCl, H_2, H, OH, HO_2, H_2O_2, CH_3, CH_3O, CH_3O_2, CH_3OH, CH_3OOH, CH_3ONO_2, $CH_3O_2NO_2$, HCO, HCHO, CH_4, CH_3Br, CF_2Cl_2, CO, N_2O, CO_2, and H_2O. The remaining five species are not integrated and are not in photochemical equilibrium, namely O_2, N_2, $HCl_{(S)}$, $H_2O_{(S)}$, and $HNO_{3(S)}$. The model contains a total of 420 reactions, 278 bimolecular reactions, 32 trimolecular reactions, 60 photolysis reactions, 4 cosmic ray processes, and 46 heterogeneous reactions.

[16] The first guess used came from a January simulation of the Cambridge 2D chemical model [*Law and Pyle*, 1991, 1993a, 1993b]. As these are not particularly well suited to the conditions observed by ATMOS during March 1992, our chemical assimilation was cycled three times. The results presented are from the third cycle. This allowed for quite rapid convergence to the conditions observed by ATMOS. Each cycle involves a first-guess simulation, followed by two iterations of the chemical Kalman filter, and then an analysis simulation. The analysis simulation is a free-running chemical simulation starting from the chemical state vector at the end of the chemical Kalman filter. This is done to mimic the way 4D-Var works, with the added advantage of having the time evolution of the full species error covariance matrix. This is desirable as it means there are no steps in the species concentrations with time due to Kalman filter increments (i.e., no time discontinuities when new information is used).

5.1. Observations

[17] The key difference between conventional modeling and data assimilation is the use of observations and information on observational and other uncertainties. The uncertainties we consider are the observational uncertainty quoted by the observation team, the representativeness uncertainty (i.e., the spatial variability over an analysis grid cell), and the photochemical theoretical uncertainty.

[18] We obtain information on the representativeness uncertainty and improve the signal to noise by using all observations available within a grid cell and by studying the probability distribution function (PDF) of these observations. The width of the PDF is used as a measure of the variability, or representativeness uncertainty, for the grid cell. When more than one observation is available in a grid cell, we take the median of the PDF as the observation for that cell and the median of the observation uncertainty PDF as the observation uncertainty for that grid cell. So, both the observation and observation uncertainty are obtained directly from the data in a way that makes them truly representative of that grid cell.

[19] The Kalman filter assimilation is a multivariate technique which can simultaneously process the observation/uncertainty information pairs of the 55 time-integrated species whenever and wherever they are available, together with the theoretical information from our photochemical model.

5.1.1. Temporal Search Criteria

[20] The criteria used to determine at what time we use an observation are local solar time and solar zenith angle. In the case of short-lived species or those with a significant diurnal cycle, such as NO and NO_2, the observation solar zenith angle needs to be within 2° of the analysis time step solar zenith angle, and the observation local solar time needs to be within 1 hour of the analysis local solar time for the observation to be used. This ensures that sunrise measurements are used at sunrise and that sunset measurements are used at sunset. This is also important as there can be a significant difference between the geographic latitude and the equivalent PV latitude. As our analysis grid uses equivalent PV latitude, it is important that we use the local solar zenith angle as the criteria for using an observation and not just the local solar time. In the case of longer-lived species or those without a significant diurnal cycle, we can use less stringent criteria. For these longer-lived species we specify that the observation solar zenith angle needs to be within 6° of the analysis time step solar zenith angle and that the observation local solar time needs to be within 1 hour of the analysis local solar time. These, in fact, are rather strict criteria for the long-lived species. Sensitivity experiments have shown that much more lax values work just as well for the long-lived species.

5.1.2. Spatial Search Criteria

[21] The criteria used to determine at what location we use an observation are equivalent PV latitude, ϕ_e, and potential temperature, θ. An observation is used in the ϕ_e-θ grid box where it lies. The grid has 21 potential temperature levels, spaced equally in $\log(\theta)$ between 400 and 2000 K, and 32 equivalent PV latitudes, spaced evenly between $-90°$ and 90°. That means that no interaction between the boxes (meridional transport, mixing, diabatic ascent/descent) is considered.

[22] We have chosen to deal with an entire vertical profile of observations at a time as vertical profiles can contain gaps. These are easy to fill in by eye, but for an algorithm to deal with the data voids we need to consider an entire profile at a time in our flow-tracking coordinate space. In a full 3D assimilation the 3D spatial covariance matrix would be performing this task. However, in this study we are using an array of 0D boxes with a full Kalman filter and detailed chemistry. To make this computationally achievable, we use multiple 0D box models which are stacked into a series of profiles, giving us a 2D global assimilation with 21 potential temperature levels, spaced equally in $\log(\theta)$ between 400 and 2000 K, and 32 equivalent PV latitudes, spaced evenly between $-90°$ and 90°.

[23] So, to clarify, the actual assimilation is done in a 2D array of completely independent boxes. However, the observation reconstructions used for this independent array of boxes are generated by considering an entire profile at a time, where the profile is a function of potential temperature for a given equivalent PV latitude range. The equivalent PV

latitude range is the width of the grid box. This is not ideal but is a necessary and good approximation because of the computer time required to do a full 3D assimilation with a Kalman filter.

[24] Our observation reconstruction algorithm consists of the following steps: (1) For each level in the analysis profile we determine which observations lie between the bottom of the level below and the top of the level above; (2) if there are more than n_{thresh} (usually taken to be 25) observations, we sort them in order of their percentage uncertainty and just use the n_{thresh} most reliable data points (i.e., those with the lowest concentration uncertainty to concentration ratio); (3) for these data points we calculate the median concentration; (4) if there are any levels which have no observations bracketed above and below by levels which did have observations, we linearly interpolate using the available values above and below.

5.2. Observation Uncertainties

[25] The uncertainty of the reconstructed observation has two components. First, for each (ϕ_e,θ) grid box we have a distribution of observed concentrations and a distribution of observed concentration uncertainties. We take the observational uncertainty to be the median observed concentration uncertainty for the current distribution in the given (ϕ_e,θ) grid box. Second, we have the representativeness uncertainty. Information on the representativeness uncertainty is gained by using all observations available within a grid cell and studying the PDF of these observations. The width of the PDF is used as a measure of the variability or representativeness uncertainty for the (ϕ_e,θ) grid cell. So, both the observation and observation uncertainty used are derived directly from the data but are selected in a way that we are sure whether they are truly representative of that grid cell. The average deviation, or mean absolute deviation, is a robust estimator of the width of the distribution [*Press et al.*, 1992]:

$$\sigma_{\text{rep}} = \text{ADev}(\chi_1 \ldots \chi_N) = \frac{1}{N}\sum_{j=1}^{N} \left|\chi_j - \bar{\chi}\right|, \tag{1}$$

where ADev is the average deviation, χ is the volume mixing ratio, χ_j refers to the individual concentration observations available in the grid cell, and $\bar{\chi}$ is the mean (i.e., average) mixing ratio. The reconstructed observation for the (ϕ_e, θ) grid box is the median of the 25 most accurate observations available in the grid box. The median is used so that the reconstructed observations, and hence the analyses, are not affected by outliers.

[26] The representativeness uncertainty is an interesting quantity as it clearly picks out mixing barriers such as the polar vortices, the tropical pipe, and the tropopause. In addition, for species such as ozone in polar day we see a large representativeness uncertainty at the summer pole. It is important to note that this representativeness uncertainty is fully accounted for, goes into the analysis, and is reflected in the analysis uncertainty. Whatever coordinate system is used, such situations will be encountered; it is therefore important to account for the grid representativeness uncertainty as accurately as possible and to include them in the analysis.

[27] It is also fair to observe that this level of treatment of uncertainties is seldom, if ever, performed for atmospheric chemical modeling. It is suggested that this is a major step taken in this study.

5.3. Photochemical Theory Uncertainties

[28] Currently, we assume a growth rate in the photochemical theory uncertainty of 5% per time step (that is, the diagonal of the covariance matrix is increased by 5% at each time step to crudely account for photochemical theory uncertainties). The time step used here is 15 min, but we often use a 1 hour time step, also. This is deliberately conservative to avoid filter divergence. In the future we want to explicitly calculate the photochemical theory uncertainty by using the uncertainty in each rate coefficient and in the constituent concentrations.

5.4. Analysis Quality Measures

[29] Once the analysis has been performed, we quantify its quality by generating a set of statistics that compare the observations used in making the analysis with the analysis itself.

5.4.1. Observation Increment, ($O - F$)

[30] This is typically the best measure of forecast skill: the difference between the first guess and the observations, also known as the observed-minus-background difference or as the innovation vector [*Daley*, 1991]:

$$(O - F) = \frac{1}{n}\sum_{k=1}^{n} (o_k - f_k), \tag{2}$$

where o_k denotes the kth observation and f_k denotes the corresponding value from the first-guess forecast. The lower the absolute value of ($O - F$), the better.

5.4.2. Analysis Increment, ($A - F$)

[31] This is a good measure of model bias: the difference between the first guess and the final analysis, also known as the analysis-minus-background difference or as the correction vector [*Daley*, 1991]:

$$(A - F) = \frac{1}{n}\sum_{k=1}^{n} (a_k - f_k), \tag{3}$$

where a_k denotes the analysis and f_k denotes the corresponding value from the first-guess forecast. The lower the absolute value of ($A - F$), the better.

5.4.3. Analysis-Observations, ($A - O$)

[32] This is typically the best measure of the analysis bias: the difference between the analysis and the observations, also known as the analysis-minus-observed difference or as the innovation vector [*Daley*, 1991]:

$$(A - O) = \frac{1}{n}\sum_{k=1}^{n} (a_k - o_k), \tag{4}$$

where o_k denotes the kth observation and a_k denotes the corresponding value from the analysis. The lower the absolute value of ($A - O$), the better.

6. Assessment of the Assimilation

[33] ATMOS ATLAS 1 observed 14 of the 55 species which our assimilation scheme can simultaneously analyze. These 14 observed species include key species in reactive

nitrogen and chlorine chemistry as well as several long-lived source gases. The ATLAS 1 observations are only available at satellite sunrise and sunset. In this case study for 30 March 1992 the measurements are mainly available at sunset.

6.1. O_3

[34] Figure 2a displays vertical profiles for the 30 March 1992 assimilation (ATMOS ATLAS 1) of the O_3 analysis concentrations, uncertainties, and skill measures. The leftmost panel shows that there is excellent agreement between the ATMOS observations (diamonds with error bars) and the analysis (line with error bars).

[35] The total observational uncertainty has two components, the measurement uncertainty (solid line in the rightmost panel) and the representativeness uncertainty (dashed line). The representativeness uncertainty attempts to quantify the spatial variability of ozone over the grid box. This has been quantified by using the average deviation observed in each grid box. The analysis uncertainty is the diagonal element of the time-integrated Kalman filter error covariance matrix (solid line with overlaid circles). It can be seen that these uncertainties are of the order of 0.1 ppmv, or a few percent.

[36] The analysis bias, $(O-A)$, is small and well within the observational uncertainties. The forecast skill, $(O-F)$, and model bias, $(A-F)$, are a little larger. This is because the a priori first guess was not entirely suitable for the situation that ATMOS observed. As was mentioned in the introduction, it came from a 2D model of the atmosphere.

6.2. NO and NO_2

[37] Full diurnal cycles produced by the assimilation for NO and NO_2 from the occultation measurements are shown in Figure 1. The left plots show the analyses overlaid with the observations (colored triangles). The right plots show the analyses' uncertainties overlaid with the observation uncertainties (colored triangles). The reduction in analysis uncertainty can be seen in the right plots when new observational information comes in. The dots indicate the locations of the assimilation analysis grid.

[38] Figures 2b and 2c display vertical slices through the NO and NO_2 diurnal cycles at sunset. The leftmost panels of Figures 2b and 2c show that there is good agreement between the ATMOS observations (diamonds with error bars) and the analyses (lines with error bars).

6.3. N_2O_5, $HONO_2$, and HO_2NO_2

[39] Figures 3a–3c display vertical slices through the N_2O_5, HNO_3, and HO_2NO_2 diurnal cycles at sunset, respectively. For each of these species the analyses and observations agree to within the analyses' uncertainties.

6.4. Other Species

[40] Figure 4 displays vertical slices through the $ClONO_2$, HCl, HCN, H_2O, CH_4, N_2O, CO, and CO_2 diurnal cycles at sunset. For each of these species the analyses and observations agree to within the analyses' uncertainties. It is noteworthy that in some cases (for example, $ClONO_2$) the assimilation gives a larger uncertainty than was specified by the measurement team. This occurs for at least two reasons:

[41] (1) The assimilation also accounts for the representativeness uncertainty. In some cases where there are large spatial gradients, this uncertainty can be larger, even considerably larger, than the observational uncertainty. The analysis uncertainty as well as the photochemical model uncertainty reflects this.

[42] (2) Information on this species is propagating through the assimilation system from observations of other chemically coupled species. This other information, with its associated weight(uncertainty), is not entirely consistent with the observations, and so the analysis acquires a higher uncertainty, for example, due to the presence of observation or model biases.

7. Data-Theory Intercomparison

[43] A key advantage of using data assimilation is the propagation of information from data-rich regions to data-poor regions, which is very relevant to the assimilation of solar occultation data such as from ATMOS.

7.1. Mathematical Framework for Intercomparison

[44] Previous studies, such as that by *Chang et al.* [1996], have shown good agreement of volume mixing ratio profiles measured by ATMOS with in situ measurements acquired from platforms on the NASA ER-2. *Sen et al.* [1998] presented volume mixing ratio profiles of NO, NO_2, HONO, HO_2NO_2, N_2O_5, and $ClONO_2$ and their composite budget (NO_y) from 20 to 39 km, measured remotely in solar occultation by the Jet Propulsion Laboratory MkIV Interferometer during a balloon flight from Fort Sumner, New Mexico (35°N), on 25 September 1993. In general, the observed profiles agree well with values calculated using a photochemical steady state model constrained by simultaneous MkIV observations of long-lived precursors and aerosol surface area from SAGE II.

[45] What is new in this study is that the comparison between observations and photochemical theory and the constraint of photochemical theory by observations is done within an objective mathematical framework, where full account is taken of information uncertainty both in the observations and in photochemical theory. Previous ATMOS intercomparisons have not done this.

[46] After examining Figures 1–4 in this paper, it may be asked, How much information is ATMOS providing, and how is this any different from what we would obtain from using a free-running model? ATMOS has provided a great deal of information for Figures 1–4. If it was not for the constraint of ATMOS ATLAS 1 observations, the sunset observations would not have agreed with the photochemical model and would not have simultaneously agreed with the O_3, NO, NO_2, N_2O_5, HNO_3, HO_2NO_2, HCN, $ClONO_2$, HCl, H_2O, CO, CO_2, CH_4, and N_2O photochemical models.

7.2. Quantifying Uncertainty

[47] If the primary interest is the likely magnitude of diurnal variation, then one does not need assimilation. The reason for using the assimilation is to simultaneously use as much information as we have and to provide an estimate of the uncertainty associated with this information, which includes a treatment of observational uncertainty, spatial and temporal representativeness uncertainty, and photochemical theory uncertainty. Assimilation also allows ob-

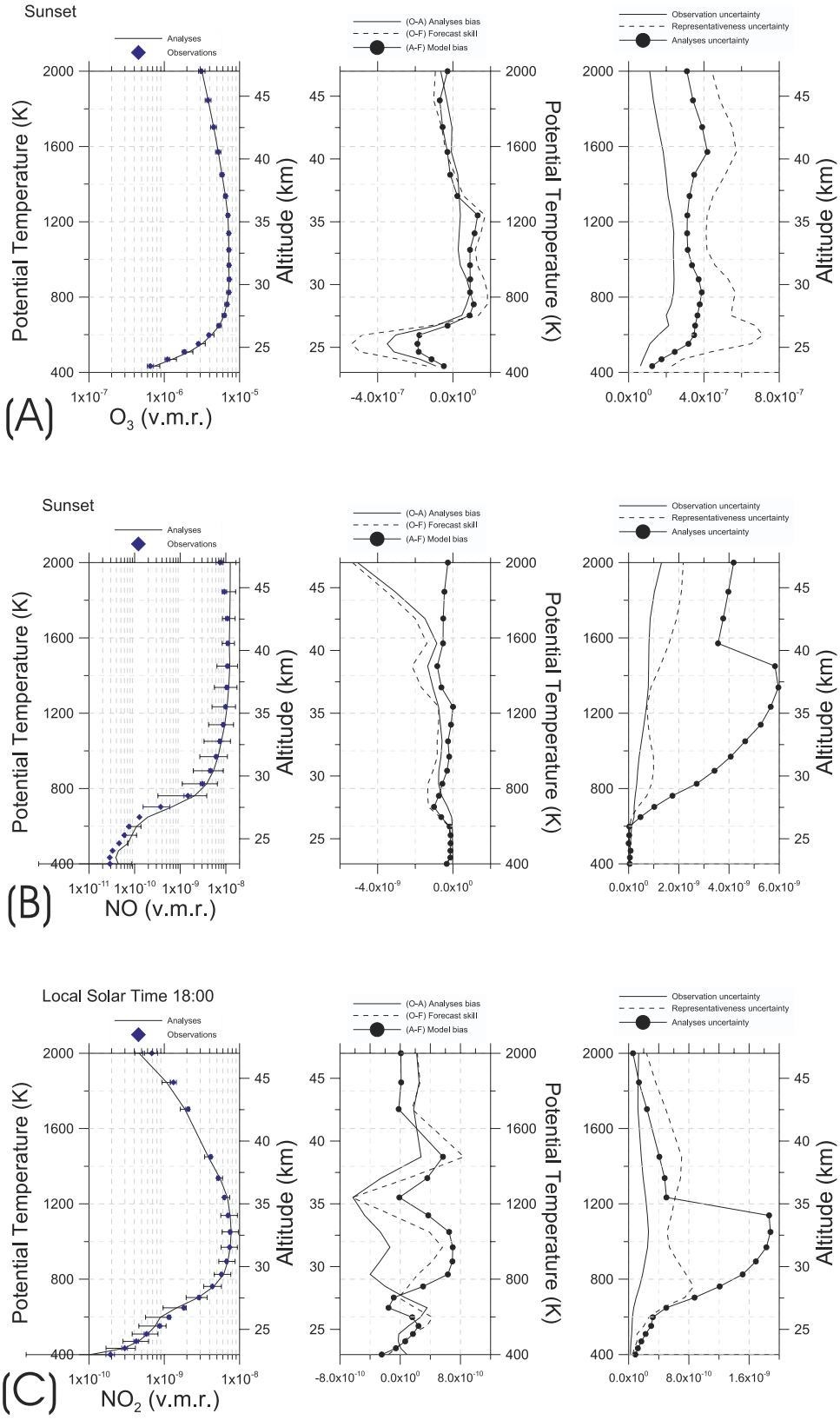

Figure 2. (left) Vertical profiles of analyses and observations: (a) O_3, (b) NO, and (c) NO_2. (middle) Skill measures described in section 4.4, namely the observation increment, $(O-F)$, which is a measure of forecast skill, the analysis increment, $(A-F)$, which is a measure of model bias, and the analysis-observations, $(A-O)$, which is a measure of the analysis bias. (right) The uncertainties: observational, representativeness, and analysis. All data are from the 30 March 1992 assimilation of ATMOS ATLAS 1 data at 38°S. Observations and observation error bars are in blue.

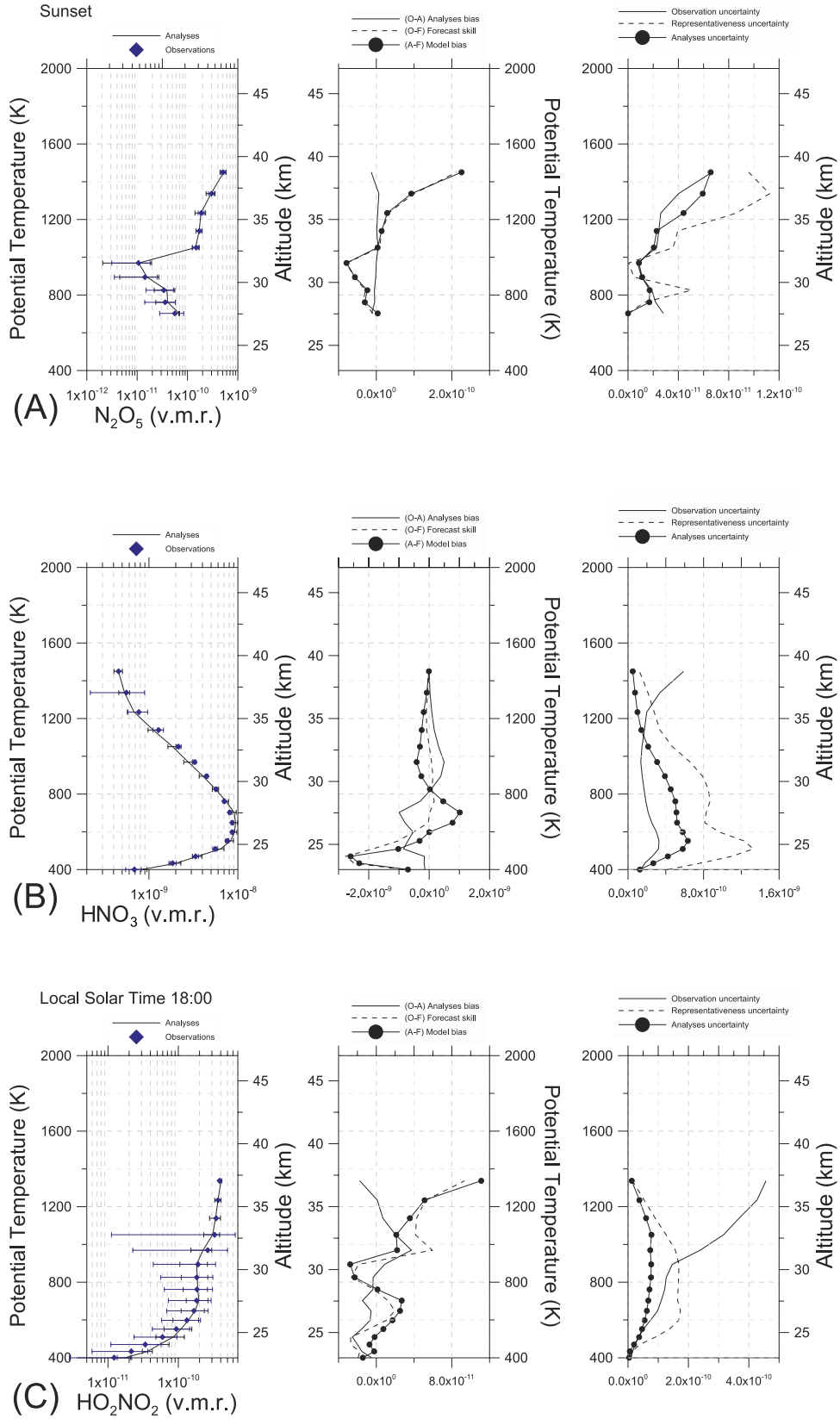

Figure 3. Same as Figure 2, except for (a) N_2O_5, (b) HNO_3, and (c) HO_2NO_2.

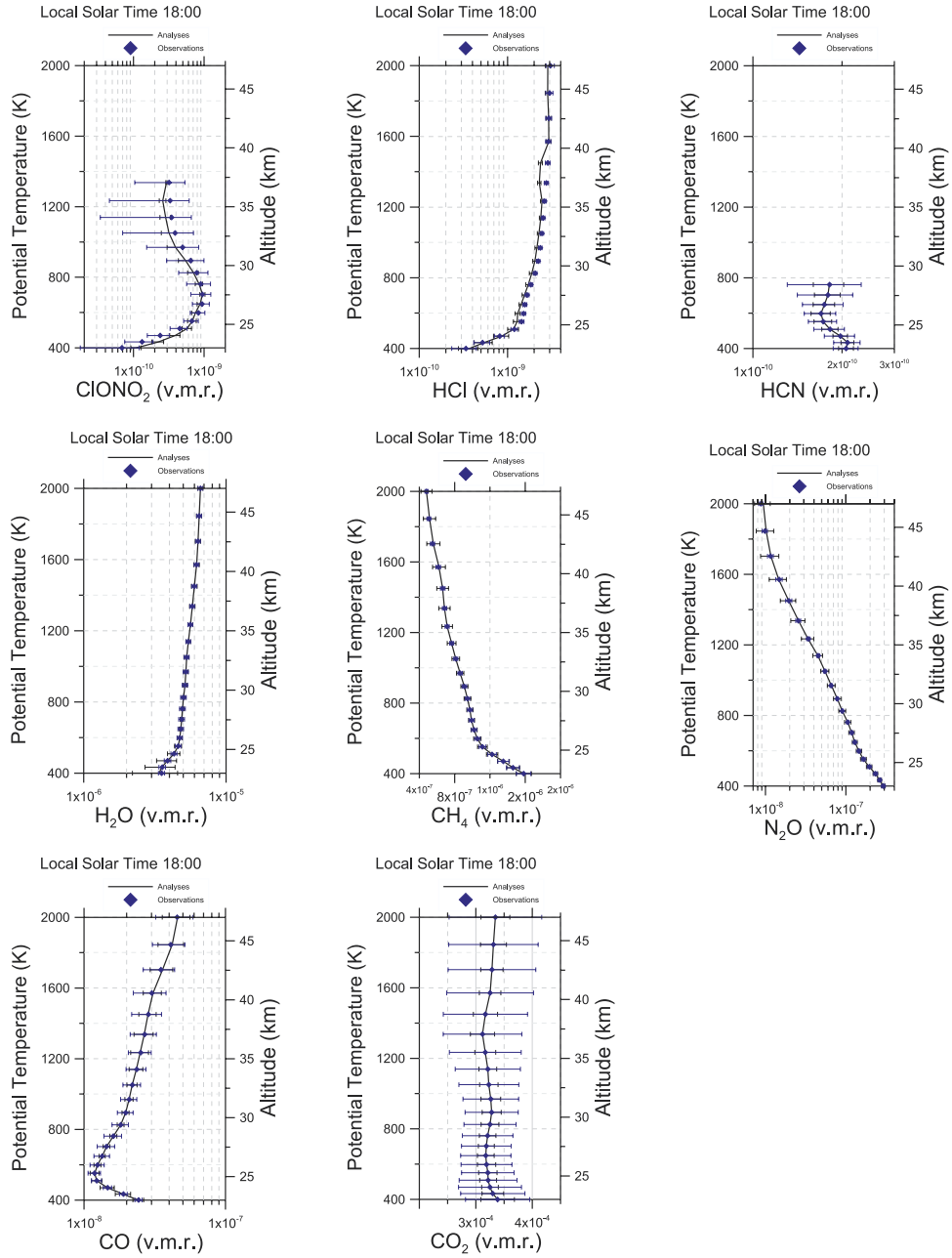

Figure 4. Vertical profiles for $ClONO_2$, HCl, HCN, H_2O, CH_4, N_2O, CO, and CO_2 analyses from the 30 March 1992 assimilation of ATMOS ATLAS 1 data at 38°S. Observations and observation error bars are in blue.

jective self-consistency tests such as those described in section 4.4 and by the χ^2 test of $(O-F)$ proposed by *Dee* [1995] and applied by *Menard et al.* [2000], *Menard and Chang* [2000], and *Khattatov et al.* [2000]. The currently available techniques, other than assimilation, do not provide this capability.

7.3. Inferring Information on Unobserved Quantities

[48] In the case of a quality data set such as ATMOS, the assimilation can reliably produce full diurnal cycles from just occultation observations and relatively reliably infer analyses for unobserved species. In this case study, full diurnal analyses and uncertainties of 55 species were produced from observations of just 14 of these species. In addition, the assimilation has provided a useful check for photochemical self-consistency between the ATMOS observations and between the observations and photochemical theory. Self-consistency is not always present between the multiconstituent observations made from a given platform; for example, UARS observed NO_2 from three instruments which were not self-consistent with each other. This will be the subject of further papers.

[49] The reason for choosing 38°S for this case study is that at this latitude, there was the greatest vertical extent of ATMOS ATLAS 1 observations. The good agreement obtained between ATMOS observations and photochemical

theory reported here was not the result of a fortuitous choice of data. Many other studies have been performed with data from numerous platforms, including other ATMOS missions, MkIV balloon flights, ER-2 data, and UARS data. Space does not us allow to present these here; they will be the subject of other papers.

[50] Observations from other platforms and measurement campaigns have been considered such as the Stratospheric Photochemistry Aerosol and Dynamics Experiment (SPADE) in May 1993 and the Cryogenic Limb Array Etalon Spectrometer (CLAES), which have much more diurnal coverage than just sunrise and sunset and which show the same result. Chemical data assimilation can reconstruct well an entire diurnal cycle from reliable observations at one time, whether the observations are at sunrise, sunset, or spaced throughout the day.

[51] This means that multispecies chemical data assimilation is likely to be useful for data product monitoring and quality assessment. This is relevant to the use of data assimilation within a validation campaign of new instruments such as those to be placed on board EOS Aura. Solar occultation data (such as from the Canadian Atmospheric Chemistry Experiment, the Aerosol Characterization Experiment (ACE), and HALOE), aircraft data (such as from the DC-8 and ER-2), and balloon data (such as the MkIV) could be used to reconstruct full diurnal cycles of species such as NO and NO_2, thus relaxing the local solar time/zenith angle coincidence requirements needed to validate Aura observations.

8. Summary

[52] Multispecies chemical data assimilation is useful for constituent data product monitoring/validation and quality assessment. Two key advantages of using data assimilation are the ability to check the chemical self-consistency of multispecies observations and information prorogating from data-rich regions to data-poor regions. This is particularly relevant to the assimilation of solar occultation data such as from ATMOS.

[53] Chemical data assimilation has been able to take solar occultation measurements of 14 species at sunrise and sunset and produce full diurnal cycles for 55 observed and unobserved species, together with their associated uncertainty. The associated uncertainty includes a treatment of the observation, representativeness, and photochemical theory uncertainty.

[54] We confirm that there is self-consistency between the ATMOS ATLAS 1 observations of O_3, NO, NO_2, N_2O_5, HNO_3, HO_2NO_2, HCN, $ClONO_2$, HCl, H_2O, CO, CO_2, CH_4, and N_2O and between ATMOS and photochemical theory. If there had been inconsistency between the observations and photochemical theory, the assimilation skill scores such as $(O-F)$ and $(O-A)$ would have highlighted these and quantified their magnitude.

[55] Future studies in preparation using chemical data assimilation include the halogen-catalyzed oxidation of hydrocarbons in the upper troposphere and lower stratosphere; the partitioning of hydrogen, nitrogen, chlorine, bromine, and carbon species as a function of temperature, solar illumination, and aerosol loading; the relative roles of catalytic ozone loss cycles and how they have changed with time and location over the UARS period; and satellite instrument validation including Envisat, ACE, and Aura.

[56] **Acknowledgments.** It is a pleasure to acknowledge NASA for a distinguished Goddard Fellowship in Earth Science; the Royal Society for a Royal Society University Research Fellowship; the government of Israel for an Alon Fellowship; the NERC, EU, and ESA for research support; and Simon Hall of Cambridge University, who has provided such excellent computational support.

References

Anderson, D., The troposphere-stratosphere radiation-field at twilight—A spherical model, *Planet. Space Sci.*, *31*, 1517–1523, 1983.

Atkinson, R., An observational study of the austral spring stratosphere: Dynamics, ozone transport and the "ozone dilution effect," Ph.D. thesis, Mass. Inst. of Technol., Cambridge, 1993.

Atkinson, R., D. Baulch, R. Cox, R. Hampson, J. Kerr, M. Rossi, and J. Troe, Evaluated kinetic and photochemical data for atmospheric chemistry: Supplement VI, IUPAC subcommittee on gas kinetic data evaluation for atmospheric chemistry, *J. Phys. Chem. Ref. Data*, *26*, 1329–1499, 1997.

Becker, G., J. Grooss, D. McKenna, and R. Muller, Stratospheric photolysis frequencies: Impact of an improved numerical solution of the radiative transfer equation, *J. Atmos. Chem.*, *37*, 217–229, 2000.

Brown, S., R. Talukdar, and A. Ravishankara, Rate constants for the reaction OH + NO_2 → HNO_3 + M under atmospheric conditions, *Chem. Phys. Lett.*, *299*, 277–284, 1999a.

Brown, S., R. Talukdar, and A. Ravishankara, Reconsideration of the rate constant of the reaction of hydroxyl radicals with nitric acid, *J. Phys. Chem.*, *103*, 3031–3037, 1999b.

Chang, A., et al., A comparison of measurements from atmos and instruments aboard the ER-2 aircraft: Tracers of atmospheric transport, *Geophys. Res. Lett.*, *23*, 2389–2392, 1996.

Chipperfield, M. P., B. V. Khattatov, and D. J. Lary, Sequential assimilation of stratospheric chemical observations in a three-dimensional model, *J. Geophys. Res.*, *107*(D21), 4585, doi:10.1029/2002JD002110, 2002.

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

Daescu, D., and G. Carmichael, An adjoint sensitivity method for the adaptive location of the observations in air quality modeling, *J. Atmos. Sci.*, *60*, 434–450, 2003.

Daley, R., *Atmospheric Data Analysis*, *Cambridge Atmos. Space Sci. Ser.*, vol. 2, Cambridge Univ. Press, New York, 1991.

Danielsen, E. F., Stratospheric-tropospheric exchange based on radioactivity, ozone and potential vorticity, *J. Atmos. Sci.*, *25*, 502–518, 1968.

Dee, D., Online estimation of error covariance parameters for atmospheric data assimilation, *Mon. Weather Rev.*, *123*, 1128–1145, 1995.

DeMore, W. B., C. J. Howard, S. P. Sander, A. R. Ravishankara, D. M. Golden, C. E. Kolb, M. J. Hampson, R. F. Molina, and M. J. Kurylo, Chemical kinetics and photochemical data for use in stratospheric modeling, *JPL Publ., 97-4*, 269 pp., 1997.

Douglass, A. R., R. B. Rood, R. S. Stolarski, M. R. Schoeberl, M. H. Proffitt, M. Loewenstein, J. R. Podolske, and S. Strahan, Global three-dimensional constituent fields derived from profile data, *Geophys. Res. Lett.*, *17*, 525–528, 1990.

Elbern, H., Parallelization and load balancing of a comprehensive atmospheric chemistry transport model, *Atmos. Environ.*, *31*, 3561–3574, 1997.

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

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

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

Eskes, H., A. Piters, P. Levelt, M. Allaart, and H. Kelder, On the assimilation of total-ozone satellite data, *Earth Obs. Q.*, *58*, 35–38, 1998.

Eskes, H., A. Piters, P. Levelt, M. Allaart, and H. Kelder, Variational assimilation of gome total-column ozone satellite data in a 2D latitude-longitude tracer-transport model, *J. Atmos. Sci.*, *56*, 3560–3572, 1999.

Fisher, M., and D. Lary, Lagrangian 4-dimensional variational data assimilation of chemical-species, *Q. J. R. Meteorol. Soc.*, *121*, 1681–1704, 1995.

Hoskins, B., Towards a pv-theta view of the general-circulation, *Tellus, Ser. A.*, *43*, 27–35, 1991.

Hoskins, B., M. McIntyre, and A. Robertson, On the use and significance of isentropic potential vorticity maps, *Q. J. R. Meteorol. Soc.*, *111*, 877–946, 1985.

Khattatov, B. V., J. C. Gille, L. Lyjak, G. P. Brasseur, V. L. Dvortsov, A. E. Roche, and J. W. Waters, Assimilation of photochemically active species and a case analysis of UARS data, *J. Geophys. Res.*, *104*, 18,715–18,737, 1999.

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

Khattatov, B., L. Lyjak, and J. Gille, On applications of photochemical models to the design of measurement strategies, *Geophys. Res. Lett.*, *28*, 2377–2380, 2001.

Lait, L., et al., Reconstruction of O_3 and N_2O fields from ER-2, DC-8, and balloon observations, *Geophys. Res. Lett.*, *17*, 521–524, 1990.

Lary, D. J., Gas phase atmospheric bromine photochemistry, *J. Geophys. Res.*, *101*, 1505–1516, 1996.

Lary, D., Data assimilation: A powerful tool for atmospheric chemistry, *Philos. Trans. R. Soc. London, Ser. A*, *357*, 3445–3457, 1999.

Lary, D., and J. Pyle, Diffuse-radiation, twilight, and photochemistry. 1., *J. Atmos. Chem.*, *13*, 373–392, 1991a.

Lary, D., and J. Pyle, Diffuse-radiation, twilight, and photochemistry. 2., *J. Atmos. Chem.*, *13*, 393–406, 1991b.

Lary, D., M. Chipperfield, J. Pyle, W. Norton, and L. Riishojgaard, 3-dimensional tracer initialization and general diagnostics using equivalent pv latitude-potential-temperature coordinates, *Q. J. R. Meteorol. Soc.*, *121*, 187–210, 1995a.

Lary, D., M. Chipperfield, and R. Toumi, The potential impact of the reaction OH + ClO → HCl + O_2 on polar ozone photochemistry, *J. Atmos. Chem.*, *21*, 61–79, 1995b.

Law, K., and J. Pyle, Modeling the response of tropospheric trace species to changing source gas concentrations, *Atmos. Envrion.*, *25*, 1863–1871, 1991.

Law, K. S., and J. A. Pyle, Modeling trace gas budgets in the troposphere: 1. Ozone and odd nitrogen, *J. Geophys. Res.*, *98*, 18,377–18,400, 1993a.

Law, K. S., and J. A. Pyle, Modeling trace gas budgets in the troposphere: 2. CH_4 and CO, *J. Geophys. Res.*, *98*, 18,401–18,412, 1993b.

McIntyre, M., and T. Palmer, Breaking planetary-waves in the stratosphere, *Nature*, *305*, 593–600, 1983.

McIntyre, M., and T. Palmer, The surf zone in the stratosphere, *J. Atmos. Terr. Phys.*, *46*, 825–849, 1984.

Meier, R., D. Anderson, and M. Nicolet, Radiation-field in the troposphere and stratosphere from 240–1000 nm. 1., General–analysis, *Planet. Space Sci.*, *30*, 923–933, 1982.

Menard, R., and L. Chang, Assimilation of stratospheric chemical tracer observations using a Kalman filter. part II: χ^2-validated results and analysis of variance and correlation dynamics, *Mon. Weather Rev.*, *128*, 2672–2686, 2000.

Menard, R., S. Cohn, and L. Chang, Assimilation of stratospheric chemical tracer observations using a Kalman filter. part I: Formulation, *Mon. Weather Rev.*, *128*, 2654–2671, 2000.

Nicolet, M., R. Meier, and D. Anderson, Radiation-field in the troposphere and stratosphere. 2. Numerical-analysis, *Planet. Space Sci.*, *30*, 935–983, 1982.

Norton, W., Breaking Rossby waves in a model stratosphere diagnosed by a vortex-following technique for advecting material contours, *J. Atmos. Sci.*, *51*, 654–673, 1994.

Portmann, R. W., S. S. Brown, T. Gierczak, R. K. Talukdar, J. B. Burkholder, and A. R. Ravishankara, Role of nitrogen oxides in the stratosphere: A reevaluation based on laboratory studies, *Geophys. Res. Lett.*, *26*, 2387–2390, 1999.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in Fortran-The Art of Scientific Computing*, 2nd ed., Cambridge Univ. Press, New York, 1992.

Proffitt, M., et al., Antarctic ozone hole from a high-altitude ER-2 aircraft, *J. Geophys. Res.*, *94*, 16,547–16,555, 1989.

Proffitt, M., K. Aikin, J. Margitan, M. Loewenstein, J. Podolske, A. Weaver, K. Chan, H. Fast, and J. Elkins, Ozone loss inside the northern polar vortex during the 1991–1992 winter, *Science*, *261*, 1150–1154, 1993.

Schoeberl, M., et al., Reconstruction of the constituent distribution and trends in the Antarctic polar vortex from ER-2 flight observations, *J. Geophys. Res.*, *94*, 16,815–16,845, 1989.

Sen, B., G. C. Toon, G. B. Osterman, J.-F. Blavier, J. J. Margitan, R. J. Salawitch, and G. K. Yue, Measurements of reactive nitrogen in the stratosphere, *J. Geophys. Res.*, *103*, 3571–3585, 1998.

Smyshlyaev, S. P., and M. A. Geller, Analysis of SAGE II observations using data assimilation by the SUNY-SPB two-dimensional model and comparison to TOMS data, *J. Geophys. Res.*, *106*, 32,327–32,335, 2001.

Stoer, J., and R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1980.

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

B. Khattatov, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, USA. (boris@ucar.edu)

D. J. Lary, Data Assimilation Office, NASA Goddard Space Flight Center, Code 910.3, Greenbelt, MD 20771, USA. (dlary@dao.gsfc.nasa.gov)

H. Y. Mussa, Unilever Cambridge Centre/Atmospheric Research Centre, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. (hym21@cam.ac.uk)