The GEOS ozone data assimilation system: specification of error statistics

Ivanka Štajner,[*] Lars Peter Riishøjgaard[†] and Richard B. Rood

Data Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

February 3, 2000

SUMMARY

A global three-dimensional ozone data assimilation system has been developed at the Data Assimilation Office of the NASA/Goddard Space Flight Center. The Total Ozone Mapping Spectrometer (TOMS) total ozone and the Solar Backscatter Ultraviolet (SBUV) or (SBUV/2) partial ozone profile observations are assimilated. The assimilation, into an off-line ozone transport model, is done using the global Physical-space Statistical Analysis Scheme (PSAS). This system became operational in December 1999.

A detailed description of the statistical analysis scheme, and in particular, the forecast and observation error covariance models is given. A new global anisotropic horizontal forecast error correlation model accounts for a varying distribution of observations with latitude. Correlations are largest in the zonal direction in the tropics where data is sparse. Forecast error variance model is proportional to the ozone field. The forecast error covariance parameters were determined by maximum likelihood estimation. The error covariance models are validated using χ^2 statistics.

The analyzed ozone fields in the winter 1992 are validated against independent observations from ozone sondes and HALOE. There is better than 10% agreement between mean Halogen Occultation Experiment (HALOE) and analysis fields between 70 and 0.2 hPa. The global root-mean-square (RMS) difference between TOMS observed and forecast values is less than 4%. The global RMS difference between SBUV observed and analyzed ozone between 50 and 3 hPa is less than 15%.

1   Introduction

Ozone observations are obtained from a number of instruments with different resolutions and error characteristics. Profile information is available from a sparse network of ozone sondes and satellite instruments with a limited spatial coverage: nadir viewing only instruments (*e.g.* Solar Backscatter Ultraviolet), occultation limb sounders (*e.g.* Halogen Occultation Experiment), or limb emission sounders (*e.g.* Microwave Limb Sounder). More complete horizontal coverage is available from satellite instruments measuring total column ozone

---

[*]additional affiliation: *General Sciences Corporation, Beltsville, Maryland*

[†]additional affiliation: *Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, Maryland*

ozone (*e.g.* Total Ozone Mapping Spectrometer). which provides nearly global coverage every day. With these satellite measurements mean ozone fields are routinely produced for geophysical applications of both atmospheric chemistry and transport. In addition, perhaps with more precision than any geophysical parameter, these data have been used to identify significant trends in the mean amount of atmospheric ozone.

There are increasing demands on the ozone observations in a wide variety of applications, which stress the capabilities of the current observing system. Global, synoptic estimates of three-dimensional ozone fields have potential applications ranging from instrument calibration. to estimates of tropospheric ozone, to improvements of weather forecasts. One strategy for filling in the spatial and temporal gaps in the observations is to assimilate the observations into a predictive model. Data assimilation (Daley 1991, Cohn 1997) provides a framework for combining the available observational ozone data and their error characteristics with the ozone background field and its error characteristics to obtain the best estimate of the true ozone field.

Data assimilation has been used successfully in the analysis of meteorological data and has been central in improved weather forecasting through increasingly more optimal data usage (Kalnay *et al.* 1998). Recently, the successes in weather forecasting have motivated increased interest is using assimilated data products for climate and chemistry transport applications (*e.g. Proceedings of the First WCRP International Conference on Reanalyses, Silver Spring, Maryland. USA, 27-31 October 1997*, Douglass *et al.* 1996). Now, there is a lot of interest in applying similar techniques to analysis of constituent fields. Under the assumption of adiabatic transport on isentropic surfaces Austin (1992) completed a study to determine feasibility of constituent data assimilation into a chemistry and transport model in the stratosphere using a simple analysis scheme. Lyster *et al.* (1997) implemented a Kalman filter for the assimilation of long-lived chemical constituents on isentropic surfaces. Ménard (2000) compared a Kalman filter with suboptimal assimilation techniques including statistical interpolation for the assimilation of constituent data on isentropic surfaces. Kondratyev *et al.* (1993) assimilated simulated satellite observations into a two-dimensional transport and photochemistry model and compared the errors in ozone analyses and forecasts for different spatial resolution, frequency and type of observations. An assimilation system for total column ozone using a two-dimensional advection model on a single pressure surface and the single correction method was developed by Levelt *et al.* (1996). Khattatov *et al.* (1999) assimilated measurements of photochemically active species into trajectory and photochemical box models using the variational technique and the extended Kalman filter and provided estimates of unobserved constituents.

The first three-dimensional ozone assimilation system was developed by Grainger and Atkinson (1996). In this system observations are analyzed using statistical interpolation and the forecast model is simply persistence, *i.e.* the forecast fields are identical to the previous analyzed fields. Levelt *et al.* (1998) assimilated ozone observations into a three-dimensional chemistry and transport model using statistical interpolation on each pressure level independently.

The Goddard Earth Observing System (GEOS) ozone Data Assimilation System (DAS), presented in this paper, is the first global three-dimensional system with an ozone transport forecast model and a global physical space based statistical analysis scheme. Observations from two instruments are assimilated: total ozone measured by the Total Ozone

Mapping Spectrometer (TOMS) and partial ozone profiles measured by the Solar Backscatter Ultra Violet (SBUV) or SBUV/2 instrument. This system was developed in the Data Assimilation Office (DAO) of the NASA/Goddard Space Flight Center primarily to provide near real time three-dimensional global ozone data to satellite instrument teams, *e.g.* for Moderate-Resolution Imaging Spectrometer (MODIS) and Multi-angle Imaging Spectroradiometer (MISR) instruments on board NASA's Terra satellite who need ozone profile and/or total ozone data as an input for their retrieval algorithms.

The GEOS ozone DAS is shown schematically in Fig. 1. In one assimilation cycle an ozone analysis from the previous assimilation cycle is advected using an off-line transport (forecast) model driven by assimilated winds from the meteorological GEOS-DAS. The time step of the transport model is 15 minutes, and the observations are introduced to the model after every transport time step. The model forecast and observations are combined in the statistical analysis according to specification of their respective error statistics. Another assimilation cycle is started from the resulting analyzed ozone field. Additional products of the statistical analysis are the observed-minus-forecast residuals and analysis increments. These quantities can be used for evaluate the performance of the system and the quality of its components.

This paper will concentrate on the description of the techniques used to represent the error covariances of the both the forecast model and the observations. In the design of a data assimilation system the specification of error covariances is of great importance, as they determine the relative weights given to the forecast and observations when they are combined to form the analysis field. Temporal evolution of the error covariances from first principles remains beyond our computational capabilities. Therefore, much of the effort in building a data assimilation system is spent in modeling of the error covariance functions. For example, in the ozone assimilation all the observations are measured by instruments on board polar orbiting satellites following a regular observing pattern, but with nonuniform spatial distribution. The distribution of observations determines the correlations; they fall of more quickly in the directions and regions of dense observational coverage. A forecast error correlation function having this property is, therefore, constructed and implemented in the ozone system. Another problem is choosing a simple model for the forecast error variance that works well at all model levels. This is challenging for the ozone field because of the high variability (by more than 2 orders of magnitude) of the field in vertical.

The major components, presented in Fig. 1, of the GEOS ozone DAS are described in detail in the following sections: the forecast model in section 2, the observing system and output data in section 3. and the statistical analysis scheme in section 4. The forecast and observation error covariance models are described in sections 4.1 and 4.2, respectively. Results from the validation of the ozone system are given in section 5. They include validation against observations from independent sources and various statistics of the differences between observed and forecast values of the ozone field. Finally, a summary and conclusions are given in section 6.

2 Forecast model

The forecast model in the ozone assimilation system is the transport model developed by Lin and Rood (1996) with a flux-form semi-Lagrangian advection scheme. This model solves the constituent advection equation

$$\frac{\partial \mu}{\partial t} + \mathbf{v} \cdot \nabla \mu = 0 \tag{1}$$

where μ is the ozone mixing ratio and $\mathbf{v}$ is the wind, using the time step of 15 minutes. The ozone field is discretized in 2° latitude by 2.5° longitude horizontal resolution and on 29 hybrid levels. The uppermost 20 levels are constant pressure levels with the following pressures: 150.34, 131.24, 114.85, 99.74, 84.20, 67.70, 52.82, 40.63, 31.07, 24.01, 18.65, 14.17, 10.14, 7.05, 4.92, 3.39, 2.16, 1.22, 0.6, and 0.2 hPa, denoted by $p_{10}, p_{11}, \ldots, p_{29}$, respectively. The lowest 9 levels are given by

$$p_k = p_{int} + \sigma_k(p_s - p_{int}), \quad \text{for } k = 1, 2, ..., 9 \tag{2}$$

where p_s is the surface pressure, $p_{int} = 161.4637$ hPa is the interface pressure between σ and pressure vertical coordinates and $\sigma_1 = 0.844$, $\sigma_2 = 0.645$, $\sigma_3 = 0.52$, $\sigma_4 = 0.397$, $\sigma_5 = 0.272$, $\sigma_6 = 0.17$, $\sigma_7 = 0.102$, $\sigma_8 = 0.052$, and $\sigma_9 = 0.015$. The approximate pressure on these levels (assuming surface pressure of 1000 hPa) are 868.82, 702.00, 597.72, 494.50, 389.91, 303.63, 247.04, 205.07, and 173.63 hPa. The assimilated wind and surface pressure fields from the GEOS-DAS are used to drive the transport. The wind fields were mapped from the 70 model σ-levels of the GEOS-DAS to the 29 hybrid levels of the ozone system using an algorithm that preserves the vertical integral of the horizontal mass flux (S.-J. Lin, personal communication).

In the current ozone assimilation system the chemical source and sink terms are not explicitly modeled. This decision was based on experiments using parameterized chemistry (Riishøjgaard *et al.* 2000). The known bias between the source and sink terms and the observations, especially near 40 km altitude, was found to be substantial on the time scale of the data insertion. Exclusion of the chemical terms performed better because the constant conflict between the observations and the chemical model was eliminated. Effectively, since the time interval between the introduction of the observations to the system is small, the observations themselves are acting as the source and sink terms. A notable shortcoming of neglecting the chemical terms is in the ozone hole region and the troposphere. However, since the initial applications of the system focus on use of synoptic maps of the total column, and the variability of the total column is dominated by advective processes, the exclusion of the chemical terms is not a critical issue. More rigorous and accurate inclusion of chemical processes will be required for future applications.

3 Observations and output data

Two types of ozone observations are assimilated in the GEOS ozone DAS: total column ozone and stratospheric ozone profiles. The total column measurements provide extensive horizontal coverage, and are dominated by processes close to the tropopause. Observations of

middle and upper stratospheric ozone profiles are necessary to produce realistic assimilated stratospheric profiles, especially in the regions of active photochemistry in the upper stratosphere and in the middle to lower stratosphere under ozone hole conditions. Practically, the mean amount of ozone is provided by the column measurements, with the vertical profile information, in concert with model information, fine tuning the three-dimensional structure.

The Total Ozone Mapping Spectrometer (TOMS) and Solar Backscatter Ultra Violet (SBUV) instruments were chosen to provide total ozone and stratospheric profile observations, respectively, for the following reasons. Both instruments have been operating for 20 years and there are plans to fly such instruments in the future (SBUV/2 on NOAA-L and QuikTOMS). For our near-real time system it is important that the ozone observations are available within 24 hours of the measurements. Both instruments measure ozone based on the portion of the ultraviolet (UV) sunlight scattered on air particles, aerosols, cloud particles and reflected from the Earth's surface and provide regular coverage of the sunlit portion of the Earth's atmosphere. Compared with ozone measurements based on infrared emission, these measurements are relatively insensitive to atmospheric temperature and humidity, surface temperature, ice and snow cover. However, ozone measured by TOMS and SBUV may be inaccurate when large amounts of SO_2 or aerosols are present in the atmosphere, for example after major volcanic eruptions. Ozone observations from both instruments have been validated in a number of studies. For further information see Riishøjgaard et al. (2000)and references therein.

The validation period used here is winter 1992, one of the standard time periods used in the Data Assimilation Office meteorological validation plan. This period overlaps with the EASOE (European Arctic Stratospheric Ozone Experiment; see Geophys. Res. Lett., special issue, vol. 21, no. 13, 1994) measurement campaign during which additional ozone sondes were launched. The ozone sondes provide independent measurements for validation of lower stratospheric and tropospheric ozone profiles. The presence of an aerosol layer in the stratosphere during winter 1992, following the eruption of Mt. Pinatubo in June 1991 allows investigations of the effect of the aerosols on the ozone field and satellite ozone observations. The assimilated observations for this period are the TOMS data from Nimbus 7 and the SBUV/2 data from NOAA-11.

The TOMS total ozone level 2 data, version 7 (McPeters et al. 1996) with the quality flag 0 are used. The TOMS is a scanning instrument with the resolution from about 50 × 50 km at nadir to about 50 × 190 km at the largest scan angle. As this resolution is higher than that of the analysis grid, the TOMS data are averaged onto this grid prior to the assimilation. Only averages based on a minimum number of observations are used. This minimum number varies with the cosine of latitude from a value of 3 at the pole to 11 at the equator.

The SBUV/2 partial column ozone data in Umkehr layers 3-12 are assimilated. The level 2, version 6 reprocessed (Bhartia et al. 1996) data, with quality flags 0 (good profile) and 1 (high optical path profile), are used. In the earlier version of the system (Riishøjgaard et al. 2000) only the data with quality flag 0 were assimilated. However, data in an almost 20 degrees wide latitude band near the polar night region carries the quality flag 1 because of the high solar zenith angle. Without the high latitude SBUV data the quality of analysis suffers significantly.

In Table 1 Umkehr layer indices, layer boundaries in hPa, and observation error standard deviation parameters used for the specifications of observation errors of SBUV data

j	pressure at the bottom of the layer p_b [hPa]	pressure at the top of the layer p_t [hPa]	$\alpha_{SBUV}(j)$
1	1013.25	253.31	0.247
2	253.31	126.66	0.325
3	126.66	63.33	0.165
4	63.33	31.66	0.095
5	31.66	15.83	0.082
6	15.83	7.92	0.054
7	7.92	3.96	0.039
8	3.96	1.98	0.058
9	1.98	0.99	0.101
10	0.99	0.49	0.137
11	0.49	0.25	0.229
12	0.25	0	0.394

Table 1: SBUV ozone profile information is reported in 12 Umkehr layers. The pressure at the boundaries of each layer and the coefficients used to model observation error variance for SBUV observations in each layer are given in this table.

(described in the section 4.2 below) are given.

The analyzed ozone filed is output every six hours with horizontal resolution of 2° in latitude and 2.5° in longitude. Prior to saving, the analyzed ozone is interpolated from the hybrid levels onto 29 output pressure levels (850, 700, 600, 500, 400, 300, 250, 200, 170, 150, 130, 115. 100, 85, 70, 50, 40, 30, 25. 20, 15, 10, 7, 5, 3, 2, 1, 0.5, and 0.2 hPa).

4   Statistical analysis scheme

The analysis scheme implemented in the GEOS ozone DAS is the Physical-Space Statistical Analysis System (PSAS) (Cohn *et al.* 1998). The statistical analysis is done after every model time step of 15 minutes. Each analysis uses the TOMS and SBUV observations measured within 7.5 minutes before or after the analysis time. Denote the number of observations by p ($p \sim 500$) and the number of grid points by n ($n = (144 \times 89 + 2) \times 29 = 371722$). Denote the p-vector of observations by $\mathbf{w}^o$, the n-vectors containing the values of the background (forecast) field by $\mathbf{w}^f$, the true (unknown) field by $\mathbf{w}^t$ and the analyzed field by $\mathbf{w}^a$. The forecast error is

$$\epsilon^f = \mathbf{w}^f - \mathbf{w}^t, \tag{3}$$

and the observation error is

$$\epsilon^o = \mathbf{w}^o - H\mathbf{w}^t, \tag{4}$$

where H is the $p \times n$ matrix of the linear observation operator mapping the model variables on model grid to the observation variables at the observation locations. The forecast error

6

covariance matrix is defined by

$$P^f = \langle \epsilon^f \, (\epsilon^f)^T \rangle, \tag{5}$$

and the observation error covariance matrix is defined by

$$R = \langle \epsilon^o \, (\epsilon^o)^T \rangle, \tag{6}$$

where $< \cdot >$ denotes the expectation operator. The analysis consists of solving[1] the innovation equation for vector $\mathbf{x}$

$$(H P^f H^T + R)\,\mathbf{x} = \mathbf{w}^o - H\mathbf{w}^f, \tag{7}$$

and obtaining the analyzed state $\mathbf{w}^a$ using the following equation

$$\mathbf{w}^a = \mathbf{w}^f + P^f H^T \mathbf{x}. \tag{8}$$

The vector $\mathbf{w}^o - H\mathbf{w}^f$ on the right hand side of the eq. (7) is called observed-minus-forecast residual. Under the assumptions that ϵ^f and ϵ^o are Gaussian distributed with mean 0, uncorrelated with $\mathbf{w}^t$ and with each other, $\mathbf{w}^a$ is the optimal estimate of $\mathbf{w}^t$ by three common optimality criteria: minimum variance, maximum likelihood and best linear unbiased estimate (see Cohn 1997).

In the sections 4.1 - 4.3 the models used to construct each of the matrices in equations (7) and (8) are given.

4.1 Forecast error covariance model

The forecast error covariance matrix P^f is defined in the eq. (5), and could in principle be obtained from Kalman filter equations. It is, however, modeled using a fixed forecast error correlation function and an approximation of the forecast error variance evolution. The reasons for the use of a simplified model are twofold: one is the problem of identifiability of all the parameters needed to fully specify the evolution of forecast error covariances (Dee 1991); the other is the prohibitive computational requirement for the evolution of forecast error covariances using Kalman filter equations.

The forecast model space in the ozone system is a spherical annulus parameterized by latitude, longitude and pressure with maximum pressure p_s on the surface of the inner sphere, and minimum pressure of 0.1 hPa on the surface of the outer sphere. The model grid points are indexed by the elements of the set $\mathcal{A}$ given by

$$\mathcal{A} = \Big\{ (a_1, a_2, a_3) : a_1 \in \{-45, -44, \ldots, 45\},$$

$$a_2 \in \{0, 1, \ldots, 143\}, a_3 \in \{1, 2, \ldots, 29\} \Big\}.$$

---

[1]For positive definite sparse matrices (like $H P^f H^T + R$ in the ozone system) the preconditioned conjugate gradient method (Strang 1986) is applicable and efficient. The iteration of this semi-direct method is terminated after obtaining an approximate solution $\tilde{\mathbf{x}}$ for which the Euclidean 2-norm of the residual $(H P^f H^T + R)\tilde{\mathbf{x}} - \mathbf{w}^o + H\mathbf{w}^f$ is less than 10^{-7} or after p iterations. The preconditioning matrix is a diagonal matrix whose diagonal is equal to the diagonal of $H P^f H^T + R$.

The model grid point $\mathbf{q}_a$ with the index $a = (a_1, a_2, a_3)$ has latitude $2a_1$ degrees, longitude $2.5a_2$ degrees and belongs to the model level a_3 with pressure p_{a_3} (as given in and above eq. (2)). Denote by $h(a) = (a_1, a_2, 1)$ and by $v(a) = (0, 0, a_3)$. The point $\mathbf{q}_{h(a)}$ has the same latitude and longitude as $\mathbf{q}_a$, but it belongs to the inner sphere with the Earth's radius with pressure p_1. The point $\mathbf{q}_{v(a)}$ has the same pressure p_{a_3} as $\mathbf{q}_a$, but its latitude and longitude are equal to $0°$.

The forecast error covariance model used here is based on that for multilevel univariate covariances presented in (DAO 1996). The matrix P^f is a grid evaluation of a forecast error covariance function of the form

$$P^f_{ab} = \sigma^f_a \sigma^f_b f(\mathbf{q}_{h(a)} \cdot \mathbf{q}_{h(b)}) g(\mathbf{q}_{v(a)}, \mathbf{q}_{v(b)}). \tag{9}$$

where the forecast error standard deviations σ^f, the horizontal f and the vertical g correlation functions are specified in the sections 4.1.1, 4.1.2 and 4.1.3 respectively. Given any σ^f that is nonnegative for every $a \in \mathcal{A}$ and functions f and g that are correlation functions on $\mathbf{R}^3$ the matrix P^f is positive semidefinite and thus a valid covariance matrix[2].

4.1.1 Forecast error variance

The dynamics of the forecast error variance $(\sigma^f)^2$ in the time interval between two consecutive analysis times t_k and t_{k+1} is given by the modified advection equation (Cohn 1993)

$$\frac{\partial (\sigma^f)^2}{\partial t} + \mathbf{v} \cdot \nabla (\sigma^f)^2 = (\sigma^m)^2 \tag{10}$$

where $(\sigma^m)^2$ is the model error variance and $\mathbf{v}$ is the wind. The initial condition is

$$(\sigma^f)^2|_{t=t_k^+} = (\sigma^a)^2|_{t=t_k^+}$$

where $(\sigma^a)^2$ is the analysis error variance and the evaluation of functions at t_k^+ represents their values immediately after the analysis at time t_k. If the growth of the forecast error variance due to model error and the reduction of forecast error variance due to the analysis are both neglected, i.e. $(\sigma^m)^2 = 0$ and $(\sigma^f)^2|_{t=t_k^+} = (\sigma^f)^2|_{t=t_k^-}$, the forecast error variance satisfies the constituent advection equation (1). Two readily available solutions of this equation are the following

$$(\sigma^f)^2 = \alpha_f \, \mathbf{w}^f, \tag{11}$$

$$\sigma^f = \beta_f \, \mathbf{w}^f, \tag{12}$$

with parameters α_f and β_f. Better ozone profiles in comparison with SBUV, Halogen Occultation Experiment (HALOE) (Bruhl et al. 1996) and ozone sonde measurements were obtained using the model in eq. (11) where forecast error variance is proportional to the forecast ozone field. In particular, the mixing ratio peak in the profile obtained with the model in eq. (12) was at lower altitude, higher in magnitude and the gradient in vertical

---

[2] A product of two correlation functions is a correlation function for the following reason. For every finite grid the matrix formed by grid evaluation of the product of functions is equal to the Hadamard product of positive semidefinite matrices (grid evaluations of the factors) and thus, it is positive semidefinite.

direction near the peak was higher in magnitude than that of profile measurements. Thus, we use model in eq. (11) with parameter $\alpha_f = 8.1 \cdot 10^{-4}$ ppmv. This value of α_f was obtained by experimentation. Maximum likelihood estimation (Dee and da Silva 1999) indicates that α_f varies with latitude: larger values are at higher latitudes, especially in the dynamically active winter hemisphere.

Further evidence in support of the forecast error variance model proportional to the ozone field as in eq. (11) comes from the following comparison of independent ozone measurements from HALOE and ozone sondes with the analyzed ozone. For given time period and a model level, let μ_i^h, for $i = 1, \ldots, N$ denote all the HALOE measurements. Each μ_i^h is interpolated from nearest two HALOE measurements above and below this level. Denote by μ_i^a the value of ozone analysis at the nearest analysis grid point in time and space to μ_i^h, for $i = 1, \ldots, N$. The ratio α of the mean-square difference between HALOE measurements and analysis with the mean of HALOE measurements,

$$\alpha = \frac{\sum_{i=1}^{N}(\mu_i^h - \mu_i^a)^2}{\sum_{i=1}^{N}\mu_i^h},$$

is shown in Fig. 2. This ratio α increases significantly only around the stratopause where a larger error is likely in both HALOE (Bruhl et al. 1996) and analysis, at the top model level of 0.1 hPa where it is likely due to the analysis error, and below 70 hPa where number and quality of HALOE observations decreases. The difference between HALOE measurements and analysis includes HALOE observation error, analysis error and the error from collocating HALOE observations with the nearest analysis profile in space and time. The ratio formed like α, but for ozone sonde observations, also shown in Fig. 2, increases significantly around tropopause and above 20 hPa where the number and the quality of sonde measurements decreases. These are the indications that the future forecast error variance models should describe larger errors around tropopause and stratopause.

4.1.2 Horizontal forecast error correlation function

Anisotropic models for forecast error correlations have been used in oceanic data assimilation: variable length scale with latitude (Derber and Rosaty 1989) and variable length scale with direction (Carton and Hackert 1990). In our system the function f in the eq. (9) is chosen to be an anisotropic correlation function obtained by composing the component-wise application of a one-to-one mapping and an isotropic correlation function. For our choice of the one-to-one mapping, the effect of the composition is that the length scales for f in zonal and meridional directions on the sphere differ, and they also change with the latitudes of points. The length scale of f is the longest in the zonal direction at the equator.

Lemma 1 *Let* $\tau : \boldsymbol{R}^3 \to \boldsymbol{R}^3$ *be any mapping. Let* $\rho : \boldsymbol{R}^3 \times \boldsymbol{R}^3 \to \boldsymbol{R}$ *be any correlation function on* $\boldsymbol{R}^3$. *The function* $f : \boldsymbol{R}^3 \times \boldsymbol{R}^3 \to \boldsymbol{R}$ *given by*

$$f(\mathbf{x}, \mathbf{y}) = \rho(\tau(\mathbf{x}), \tau(\mathbf{y}))$$

is a correlation function.[3] *If* τ *is one-to-one and the grid evaluation of* ρ *is a positive definite*

---

[3]Given any finite grid of n points $\{q_1, q_2, \ldots, q_n\}$ the matrix obtained by evaluating f on this grid is equal to the matrix obtained by evaluating ρ on the grid $\{\tau(q_1), \tau(q_2), \ldots, \tau(q_n)\}$, and thus it is positive semidefinite.

matrix for every finite grid of distinct points, then every grid evaluation of f on a finite grid of distinct points is a positive definite matrix.[4]

The horizontal forecast error correlation function f is given by Lemma 1 with ρ and τ specified as follows. The function ρ is the second order autoregressive (SOAR) correlation function

$$\rho(\mathbf{q}_a \cdot \mathbf{q}_b) = (1 + ||\mathbf{q}_a - \mathbf{q}_b||/L)\, exp(-||\mathbf{q}_a - \mathbf{q}_b||/L), \tag{13}$$

with $L = 385$ km and $|| \cdot ||$ denoting the Euclidean distance in $\mathbf{R}^3$.

The mapping τ on $\mathbf{R}^3$ is given in Cartesian coordinates by

$$\tau(x, y, z) = (l(z)x, l(z)y, z) \tag{14}$$

where

$$l(z) = \begin{cases} (z/(z+s))^{1/2} & \text{for} & \frac{s}{2} \le z \le s, \\ s/(2(s^2 - z^2)^{1/2}) & \text{for} & -\frac{s}{2} \le z \le \frac{s}{2}, \\ (z/(z-s))^{1/2} & \text{for} & -s \le z \le -\frac{s}{2}, \end{cases} \tag{15}$$

and $s = ||(x, y, z)||$. This mapping contracts a sphere centered at the origin towards the axis through its poles given by the equation $x = y = 0$ as shown in Fig. 2. The contraction is the strongest at the equator ($z = 0$) where the contraction factor l(z) is the smallest. Consequently, the length scale for f in the zonal direction at the equator is the largest. The image of the Earth's surface (sphere with radius $s = 6371$ km) is a surface consisting of the union of a finite cylinder and two hemispheres given by

$$x^2 + y^2 + \left(z - \frac{s}{2}\right)^2 = \frac{s^2}{4} \quad \text{for} \quad \frac{s}{2} \le z \le s, \tag{16}$$

$$x^2 + y^2 = \frac{s^2}{4} \quad \text{for} \quad -\frac{s}{2} \le z \le \frac{s}{2}, \tag{17}$$

$$x^2 + y^2 + \left(z + \frac{s}{2}\right)^2 = \frac{s^2}{4} \quad \text{for} \quad -s \le z \le -\frac{s}{2}. \tag{18}$$

Anisotropic forecast error correlations with longer length scale in the longitudinal than meridional directions, and longer in the tropics than at high latitudes were expected for the following reasons. Due to the observing geometry of the TOMS and the SBUV instruments there are more observations per unit area of the Earth's surface at higher latitudes than in the tropics. For example, the SBUV instrument measures at nadir points only. The distance of the neighboring observations along the orbit (roughly meridional direction) is about 200 km. The distance between neighboring observations in longitudinal direction (on two subsequent orbits) varies from about 400 km at high latitudes to about 2800 km in the tropics. The TOMS observations provide almost global coverage in one day. However, frequency of observations per unit area varies with latitude. A region in the tropics is observed once per day. At high latitudes subsequent orbits overlap. Thus, a region at high latitudes is observed several times a day. This distribution of SBUV and TOMS observations is expected to result in longer analysis and forecast error length scales in the zonal direction

---

[4]The fact that τ is one-to-one guarantees that a every grid of n distinct points is mapped onto a grid of n distinct points.

in the tropics (data sparse direction and region) than in the meridional direction in the tropics and any direction at high latitudes.

Further support for the anisotropy in correlations given by f comes from the following experiment. According to Kalman filter equations the forecast error covariance matrix at time t_k is

$$P_k^f = A_{k-1} P_{k-1}^a A_{k-1}^T + Q_{k-1},$$

where P_{k-1}^a is the analysis error covariance matrix at the preceding time t_{k-1}, A_{k-1} is the matrix specifying the action of the forecast model between times t_{k-1} and t_k, and Q_{k-1} is the model error covariance matrix for the time interval between t_{k-1} and t_k. The relative contribution of analysis and model error terms to the forecast error covariances was discussed by Cohn (1993). Initially, the analysis term dominates. The contribution of the model error term increases with increasing length of the interval between t_{k-1} and t_k. The model error is particularly important in the limiting case when ozone is simulated by integrating forecast model for a long time without assimilating any observations. In contrast, the analysis error term largely determines the forecast error covariances when observations are assimilated frequently. Recall that in the ozone system observations are assimilated every 15 minutes. The anisotropy in forecast error covariances in the ozone system is expected to be largely determined by the anisotropy in analysis error covariances. Thus, we performed a simplified experiment in which horizontal distribution of observations simulates that of the SBUV observations. The analysis error correlations were computed using the Kalman filter equation

$$P^a = (I - P^f H^T (H P^f H^T + R)^{-1} H) P^f,$$

where both forecast and observation error variances were spatially constant, observation errors were uncorrelated. and forecast error correlations were modeled using an isotropic function. The analysis error correlation length was found to be the largest in the longitudinal direction in the tropics. The degree of anisotropy in the analysis error correlations was found to increase with decreasing observation error variance.

Finally, maximum likelihood estimation (Dee and da Silva 1999) of the forecast error length scale was done. The realizations of the observed-minus forecast residuals (right hand side of the eq. (7)) in the ozone system were used. The forecast error length scale was estimated for different latitude regions and for longitudinal and latitudinal directions. These estimates of forecast error length scale were found to vary with latitude and direction similarly to the way that length scale of f varies with latitude and direction.

4.1.3 Vertical forecast error correlation function

The vertical correlation function g is constructed using Lemma 1,

$$g(\mathbf{q}_{v(a)}, \mathbf{q}_{v(b)}) = \nu((\ln(p_{a_3} + p_0) - \ln(p_{b_3} + p_0))/c). \tag{19}$$

The function ν depending on an additional parameter e is a member of the family of compactly supported correlation functions constructed by Gaspari and Cohn (1999) and it is given in the Appendix. The parameter values $p_0 = 1.3$, $c = 0.5 \ln 2.4$, and $e = -1.4$ are static, determined in offline calculations. For the purpose of determining their values it is assumed that the forecast error is proportional to the change in the ozone field due to the

forecast step. *i.e. the ozone transport*. Thus, the vertical correlations of the forecast error are identical to the vertical correlations of the change to the ozone field due to the transport. The change to the ozone field by the transport, accumulated over 24 hours. in each of the model profiles is treated as one element of the sample. The sample correlation matrix C is constructed. and the parameters e, p_0 and $c \leq 0.5 \ln 2.4$ were chosen so that the grid evaluation of g approximates C. The value of $c = 0.5 \ln 2.4$ guarantees that the correlation coefficient for any pair of points on our model grid whose level indices differ by six or more is zero.

Computational savings in the PSAS algorithm can be obtained by enforcing the sparsity of the forecast error covariance matrix P^f. One way to achieve this is through the choice of a vertical forecast error correlation function. For a function that is equal to zero for pairs of model grid points whose vertical separation exceeds 5 model levels the matrix P^f is less than 33% full. Five such correlation functions were tested: the function g, uncorrelated (*i.e.* 1 for zero separation, 0 otherwise), function falling off linearly to 0 at separation of 6 model levels and equal to 0 for separation of 6 or more levels, compactly supported correlation functions (Gaspari and Cohn 1999) of two different vertical coordinates: the logarithm of the pressure and the average potential temperature. Among these candidate functions, g provided the closest fit to the sample correlation matrix C and its use resulted in the analyzed ozone profiles closest to the SBUV and Halogen Occultation Experiment (HALOE) measurements.

4.2   Observation error covariance model

The TOMS observation errors are assumed to be uncorrelated. both among themselves and with respect to SBUV observation errors. The SBUV observation errors are assumed to be correlated in the vertical.

The block of the observation error covariance matrix R corresponding to TOMS observations is therefore diagonal:

$$R_{ij} = \begin{cases} (\sigma^o(i))^2 & \text{for } i = j \\ 0 & \text{otherwise,} \end{cases} \tag{20}$$

Recall that TOMS observations are averaged onto the forecast model grid to form super-observations. Observation error variance $(\sigma^o(i))^2$ for i^{th} super-observation created from TOMS data is modeled by

$$(\sigma^o(i))^2 = (\alpha_{TOMS} \, \mathbf{w}^o(i))^2 + (\sigma^{rep}(i))^2 \tag{21}$$

where $\alpha_{TOMS} = 0.015$ and $\sigma^{rep}(i)$ is the sample error standard deviation of the TOMS observations averaged to form the ith super-observation. The quantity $\sigma^{rep}(i)$ is used to model representativeness error (Cohn 1997) standard deviation for i^{th} TOMS super-observation.

Observation error variance $(\sigma^o(j))^2$ for j^{th} SBUV observation is modeled by

$$(\sigma^o(j))^2 = (\alpha_{SBUV}(u_j) \, \mathbf{w}^o(j))^2, \tag{22}$$

where u_j is the index of the Umkehr layer corresponding to the observation $\mathbf{w}^o(j)$ and parameters $\alpha_{SBUV}(u_j)$ are given in Table 1.

12

The matrix of correlations of SBUV errors in Umkehr layers 3 to 12 is modeled by a symmetric matrix whose lower left triangle is:

$$
\begin{bmatrix}
1 & & & & & & & & & \\
.125 & 1 & & & & & & & & \\
.101 & .363 & 1 & & & & & & & \\
.042 & .040 & .501 & 1 & & & & & & \\
.059 & .011 & .196 & .414 & 1 & & & & & \\
.214 & .068 & .120 & .088 & .305 & 1 & & & & \\
.003 & -.049 & -.187 & -.122 & -.061 & .438 & 1 & & & \\
-.027 & -.155 & -.097 & -.032 & .031 & .055 & .363 & 1 & & \\
.218 & .040 & .197 & .172 & .035 & .150 & .009 & .502 & 1 & \\
-.047 & -.177 & -.218 & -.084 & .032 & -.015 & .225 & .348 & .125 & 1
\end{bmatrix} . \tag{23}
$$

The parameters in the SBUV error covariance model were determined using SBUV observed-minus-forecast (O-F) values from an earlier assimilation experiment. Under the assumption of uncorrelated observation and forecast errors, the covariance matrix of observed-minus-forecast residuals is the sum of the observation error covariance matrix R and $HP^{f}H^{T}$, where H is the matrix of the linear operator mapping model ozone mixing ratio to partial column ozone in Umkehr layers. For the purpose of determining SBUV observation error statistics the forecast error correlation matrix is assumed to have multivariate Gaussian distribution with mean 0, variance given in the eq. (11) and the correlation matrix C described at the end of section 4.1.3. This specifies P^{f}. For each pair of Umkehr layers, an element of $HP^{f}H^{T}$ corresponding to this pair is a weighted sum of forecast error covariances of pairs of model layers, one intersecting the first Umkehr layer and the other intersecting the second Umkehr layer. Each weight is the pressure difference in the intersection of an Umkehr layer with a model layer multiplied by the conversion factor into Dobson units. The vector of SBUV observed-minus-forecast residuals is assumed to have mean 0. The difference between SBUV observed-minus-forecast sample covariance matrix and $HP^{f}H^{T}$ is interpreted as the SBUV observation error covariance matrix. This matrix is not guaranteed to be positive definite by construction. Thus, its positive definiteness was verified subsequently by showing that all its eigenvalues are positive.

4.3 Observation operator

Let $\mu(p)$ denote the ozone mixing ratio in an atmospheric column as a function of pressure p. The partial column ozone between the levels with pressures p_1 and p_2 is

$$
\mathcal{I}(\mu, p_1, p_2) = \gamma \int_{p_1}^{p_2} \mu(p) dp. \tag{24}
$$

where γ is the constant for conversion from units of pressure to Dobson units. For example, $\mathcal{I}(\mu, 0, p_s)$, where p_s is the surface pressure, is the total column ozone. If p_t is the pressure at the top and p_b is the pressure at the bottom of an Umkehr layer, then $\mathcal{I}(\mu, p_t, p_b)$ is the partial column ozone in this Umkehr layer.

There are 29 hybrid levels in the ozone assimilation system and therefore the model profile is given by a vector μ of size 29. A discrete approximation of $\mathcal{I}(\mu, a, b)$ is given by

the following sum

$$I(\mu, p_1, p_2) = \gamma \sum_{k=1}^{29} \mu(k) \triangle p_k \qquad (25)$$

where $\triangle p_k$ is the pressure increment of the intersection of the pressure interval corresponding to the model layer k and the interval $[p_1, p_2]$.

After the averaging procedure described in section 3, every TOMS observation is given at a horizontal location of a model profile μ. The observation operator for TOMS total column ozone consists of evaluating $I(\mu, 0, p_s)$.

The observation operator for SBUV partial ozone columns consists of two parts. First is the linear interpolation in longitude and latitude of the neighboring four model ozone mixing ratio profiles. A profile of ozone mixing ratio μ at model levels and with the horizontal location of the SBUV observation is obtained. Denote by p_t the pressure at the top and by p_b the pressure at the bottom of the Umkehr layer for which the SBUV measured the partial ozone column. The second part of the observation operator is the evaluation of $I(\mu, p_t, p_b)$.

5   Validation

5.1   Validation against TOMS and SBUV observations

One measure of the quality of the assimilation is how well the data assimilation system predicts the observations being blended into the system. This allows us to evaluate how consistent the observations are with our knowledge of the atmosphere as described by the forecast model. In this section global synoptic analysis and forecast fields are compared with TOMS and SBUV observations. An example showing a partitioning of analyzed ozone across the tropopause is included to illustrate the influence of the forecast model and the assumptions used in the statistical analysis on the analyzed ozone.

An example of the analyzed total ozone field is given in the upper plate of Fig. 4 together with a daily map of TOMS level 3 data on the same day. The analyzed ozone field is synoptic and global, while TOMS needs 24 hours to obtain the coverage shown, and it does not observe in the polar night region. There is a close qualitative agreement between these two fields. Total ozone values are the highest in the northern high latitudes, lower in the southern hemisphere, and the lowest in the tropics. Note that a transient feature of very low ozone values over Europe is captured in both fields. This feature is typical for the large variability, of about 200 DU, around the circle of latitude 60° N.

In Figs. 5 and 6 a measure of differences between TOMS observations and model forecast of total ozone is shown. The observed-minus-forecast residuals $\mathbf{w}^o - H\mathbf{w}^f$ (right hand side of the eq. (7)) indicate how well are the observations predicted by the system. The evolution of daily global root mean square (RMS) of TOMS observed-minus-forecast residuals in the ozone system during the validation period is given in Fig. 5. There is a sharp drop in the RMS over the first two days of the assimilation from about 30 (due to the initial condition) to about 10 Dobson units. The RMS stays below 4% of the average total ozone, but there is an increase of $\sim$ 2 Dobson units during the validation period. In Fig. 6 the daily RMS of TOMS observed-minus-forecast residuals in 2° degree latitude bands is shown. Most of the increase in the RMS occurs at the northern middle and southern high latitudes. In the

14

northern middle latitudes. due to the decreasing polar night region, the coverage of TOMS is increasing to include more of the dynamically active region with relatively high variability in total ozone. In the southern high latitudes the variability of the total ozone field increased during the month of January and consequently the RMS of observed-minus-forecast residuals increased.

The quality of analyzed ozone profiles is evaluated in a comparison with SBUV data. The SBUV level 2 ozone mixing ratio product, given on pressure levels, is used in this comparison. For each synoptic time (0, 6, 12, 18 z) approximately half of the corresponding Umkehr layer data (the observations in the three hours preceding the synoptic time) has been assimilated by the system. The RMS differences between analysis and SBUV values over one day are shown in Fig. 7 (solid curve). Also shown are the RMS differences between analysis and SBUV for the system described by Riishøjgaard *et al.* (2000), which included a biased chemistry model. For the current system the values are generally smaller. There are three main reasons for the improvements. In the upper stratosphere the main cause is removal of the biased chemistry parameterization from the forecast model. Below 60 hPa the improvements are due to the assimilation of the additional SBUV observations from Umkehr layer 3. In addition, the global RMS difference decreased at most levels due to assimilation of additional SBUV observations measured under high solar zenith angles (increasing the SBUV data coverage by about 20° latitude in the northern high latitudes).

Partitioning of the ozone across the tropopause is shown in Figure 8. Total ozone in the tropical region measured by TOMS exhibits a distinct wave one feature (Shiotani 1992), with the maximum in total ozone often over the Atlantic ocean and the minimum over the Pacific ocean. This feature was found to be limited to the tropospheric part of the column by Ziemke *et al.* (1996). The monthly means of the analyzed ozone mixing ratio at 150 hPa and 50 hPa are shown in Fig. 8. In the analyzed ozone fields the wave one feature is indeed limited to below 150 hPa (upper plate). The analyzed ozone field in the stratosphere has less zonal variability as it can be seen in the field at 50 hPa (lower plate). However, note that the shape of the ozone analysis at 50 hPa depends strongly on the assumptions used in the assimilation system. It can be seen from the statistical analysis equations (7) and (8) that the analysis field is formed using the forecast model, assimilated observations and specification of their error characteristics. Changes in either of these components results in changes of the mean analyzed field at 50 hPa as it is illustrated by the following three experiments. In the first experiment higher forecast error variances were used and thus more weight was given to observations. In this experiment the maximum of the mean analyses is over the eastern Atlantic. This maximum agrees with that of the SBUV observations in the Umkehr layer 4, *i.e.* of the ozone column between 31.66 and 63.33 hPa. In another experiment parameterized ozone production and loss rates were used as a part of the forecast model, as in Riishøjgaard *et al.* (2000). The resulting monthly mean analysis field is almost identical to the field shown in Fig. 8. The variability in the ozone field at these altitudes is largely determined by advective rather than chemical processes. In the last experiment the forecast model using parameterized chemistry and transport was integrated without assimilating any observations. The resulting monthly mean ozone field has a maximum over western and a minimum over eastern Pacific. Thus, the analyzed ozone field in the tropics at 50 hPa results as a combination of the influence of dynamics (with a maximum over western Pacific) and SBUV observations (with a maximum over eastern Atlantic) according to the specification

of error covariances in the statistical analysis scheme.

5.2 Validation against independent observations

Comparison of the analysis fields with standard independent observations that are not assimilated is another common validation procedure. For example, statistics of their differences can reveal biases due to input data that may not be apparent from observed-minus-forecast residuals. Furthermore, in the ozone system the vertical resolution of the analysis field exceeds the vertical resolution of SBUV and TOMS observations. Thus, comparisons with independent profile observations with high vertical resolution are important in evaluating the quality of analyzed ozone profiles.

Two sources of independent ozone profile observations are used in the validation: WMO balloon-borne ozone sondes and HALOE (Halogen Occultation Experiment) on Upper Atmosphere Research Satellite. Ozone sondes provide ozone profiles in the troposphere and lower to middle stratosphere. Sonde measurements have high vertical resolution, but stations are nonuniformly distributed (with the majority in northern middle to high latitudes) and sonde launches are infrequent. Fewer than 400 profiles are available from the WMO database during our validation period and less than half of the sondes reached levels higher than 7 hPa. All the available sonde profiles were used in the following comparison, without any quality control. Each sonde profile was interpolated to the analysis pressure levels (ranging from 850 to 10 hPa) from the nearest measurements above and below each analysis level using linear interpolation in the logarithm of pressure. The mean profiles of the sondes and the analysis are shown in the upper plate of Fig. 9 and they are generally in good agreement. The analysis mean around 150 hPa is larger than sonde mean. A larger error at this altitude is a likely consequence of two characteristic features: proximity of the tropopause with sharp changes in the ozone profile and this being the highest altitude not covered by assimilated SBUV profile observations. The RMS difference between sondes and analysis (lower plate of Fig. 9) generally increases with altitude. The RMS is relatively large around 130 hPa, but decreases around 85 hPa due to assimilation of the SBUV data from Umkehr layer 3, which contains the pressure levels at 70, 85. 100 and 115 hPa.

We now digress from the statistical validation to show the ability of the system to reproduce variability in the ozone field. This is illustrated by reproducing an event in which ozone field differed significantly from the climatological values. At the end of January 1992 subtropical tropospheric air was transported into lower midlatitude stratosphere (Vaughan and Timmis 1998). Low values of total ozone are visible in Fig. 4 over Europe in analyzed fields and TOMS observations. In Fig. 10, lower stratospheric ozone profiles over Hohenpeissenberg, Germany are shown. On January 24, before the event occurred, the analyzed profile and ozone sonde measurements have values typical for this latitude. After the event occurred, on January 29, there is a very good agreement between sonde measurements and GEOS ozone DAS analysis, both exhibiting unusually low ozone values.

Statistical validation of stratospheric and mesospheric parts of the profiles is done using independent HALOE measurements. From the middle of January to the end of February 1992 HALOE measured 651 profiles at the sunrise between the latitudes 75S and 49N. Their mean is in a close agreement with the mean of analysis profiles and they are both shown in Fig. 11. Some of the disagreement is likely to be caused by biases in the SBUV observations

due to the aerosols. The aerosol layer formed by the eruption of Mt. Pinatubo reached about 25 km altitude in the tropics during our validation period (Torres *et al.* 1995b). The effect of such high altitude aerosol layer on SBUV is underestimation of ozone between 32 and 8 hPa and overestimation between 63 and 32 hPa (Torres and Bhartia 1995a). Similar biases are expected in the analyzed ozone because our statistical analysis scheme does not account for biases in observations.

5.3 Validation of the forecast and observation error covariance models

The analyzed ozone fields were validated in the preceding two sections. In this section we focus on the validation of the forecast and observation error covariance models implemented in the ozone system using statistics of the realizations of observed-minus-forecast residuals.

The true state ϵ^t and the errors ϵ^f and ϵ^o in equations (3) and (4) are unknown. However, the observed-minus-forecast residuals

$$\mathbf{w}^o - H\mathbf{w}^f = \epsilon^o - H\epsilon^f$$

are available. Under the assumptions that forecast error ϵ^f and observation error ϵ^o are uncorrelated Gaussian distributed with mean 0 and covariance matrices P^f and R, respectively, the observed-minus-forecast residuals $\mathbf{w}^o - H\mathbf{w}^f$ are Gaussian distributed with mean 0 and covariance matrix $HP^fH^T + R$. Let z be the random variable given by

$$z = (\mathbf{w}^o - H\mathbf{w}^f)^T(HP^fH^T + R)^{-1}(\mathbf{w}^o - H\mathbf{w}^f) = \mathbf{x}^T(\mathbf{w}^o - H\mathbf{w}^f)$$

where

$$\mathbf{x} = (HP^fH^T + R)^{-1}(\mathbf{w}^o - H\mathbf{w}^f)$$

is the solution of eq. (7). Under the above assumptions, z has χ^2 distribution with p degrees of freedom. Its mean is p and its variance is $2p$. Our samples are based on different numbers of observations, *i.e.* p is not constant, thus we consider two related variables whose moments are independent of p. The mean of $q = z/p$ is 1 and the variance of $y = z/\sqrt{p}$ is 2. Denote the number of samples by k. Each

$$z_i = \mathbf{x}_i{}^T(\mathbf{w}^o{}_i - H_i\mathbf{w}^f{}_i)$$

for $i = 1, \ldots, k$ should be a realization of a random variable with χ^2 distribution with p_i degrees of freedom. Let $q_i = z_i/p_i$ and $y_i = z_i/\sqrt{p_i}$ for $i = 1, \ldots, k$. Denote by $\hat{\mu}$ the sample mean of q_i

$$\hat{\mu} = \frac{1}{k}\sum_{i=1}^{k}\frac{z_i}{p_i},$$

by $\hat{v}_0$ the sample variance of y_i assuming the known mean $\sqrt{p_i}$ of y_i

$$\hat{v}_0 = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{z_i}{\sqrt{p_i}} - \sqrt{p_i}\right)^2 = \frac{1}{k}\sum_{i=1}^{k}\frac{(z_i - 1)^2}{p_i},$$

exp.	$\hat{\mu}$	$\hat{v}_0$	$\hat{v}_1$	κ_1	κ_2	k
	1	2	2	68.26	95.44	
initial	20.067	148557.016	72206.987	3.37	13.48	89
current	1.519	142.098	67.665	12.09	23.08	91

Table 2: The values of sample statistics computed using data from an initial version (Riishøjgaard *et al.* 2000) and the current version of GEOS ozone DAS are given. All the statistics were computed for the samples from noon on February 15 to noon on February 16, 1992. The values that each statistics should approximate, provided that assumptions leading to χ^2 distribution were correct, are given in the second row of this Table. Note that the assimilation of SBUV data with the quality flag 1 increases the number of analysis times for which there are data to assimilate, or equivalently the number of samples k.

and by $\hat{v}_1$ the sample variance of y_i with respect to the estimated mean $\hat{\mu}\sqrt{p_i}$ of y_i

$$\hat{v}_1 = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{z_i}{\sqrt{p_i}} - \hat{\mu}\sqrt{p_i}\right)^2 = \frac{1}{k}\sum_{i=1}^{k}\frac{(z_i - \hat{\mu})^2}{p_i}.$$

For p large and a random variable z with χ^2 distribution with p degrees of freedom, the distribution of $h(z,p) = \sqrt{2z} - \sqrt{2p}$ can be approximated by the standard normal distribution. The relative frequency of samples for which $h(z_i, p_i) \in [-1, 1]$ and $h(z_i, p_i) \in [-2, 2]$ are denoted by κ_1 and κ_2, respectively.

The computed values of variables $\hat{\mu}$, $\hat{v}_0$, $\hat{v}_1$, κ_1 and κ_2 for an initial (Riishøjgaard *et al.* 2000) and the current version of GEOS ozone DAS are given in Table 2. The values of all the above variables for the current system are closer to the values that they should be approximating, some even by several orders of magnitude.

There is a number of reasons why the variances $\hat{v}_0$ and $\hat{v}_1$ are larger than 2. For example, the aerosol layer from the eruption of Mt. Pinatubo volcano in Philippines on June 15, 1991 was in the stratosphere during our validation period. This caused a scan angle dependent bias in TOMS total ozone on the order of 5 DU (Bhartia *et al.* 1993) and an altitude and solar zenith angle dependent bias in SBUV ozone profiles (Torres and Bhartia 1995a). The errors in the forecast due to the neglected photochemical processes are not modeled. These errors are expected to be larger at higher model levels where photochemical processes have short time scales. The transport model without chemistry is biased because it cannot maintain the absolute ozone peak in the tropics around 10 hPa. The relative error of SBUV observations depends on the solar zenith angle at which the measurements were done. However, the relative error standard deviation is modeled in the ozone system by a constant for each SBUV Umkehr layer (see Table 1).

The χ^2 related diagnostic statistics given in Table 2 are important for evaluating error covariance models and the validity of probabilistic assumptions used in the statistical analysis. However, while it is convenient to evaluate a complex system using several χ^2 related statistics there is also a danger in oversimplifying as different changes in covariance models can result in similar changes in these statistics. For example, a change in covariance models that attempts to account for one of the problems in error modeling listed in the previous

paragraph can introduce an imbalance between observation and forecast error models. This change can result in a significant temporary improvement of the χ^2 related statistics. But, after more than a day, an overall degradation in performance can be seen using χ^2 related statistics and other validation metrics for analyzed fields presented earlier in the paper.

6   Summary and conclusions

A description of a new three-dimensional global ozone data assimilation system and examples from its validation were given. In this system the TOMS total ozone and the SBUV partial ozone profile observations are assimilated into an ozone transport model. A global physical space based statistical analysis scheme is used to assimilate available data at every model time step (15 minutes).

In the development of the ozone assimilation system system special attention was paid to the specification of forecast and observation error characteristics. These error characteristics determine the relative contributions of forecast field and observations when they are combined in the statistical analysis to form the analyzed ozone field.

Models for error characteristics and reasons for their choices were described. For example, forecast error variance is modeled proportional to the ozone field. Unique models are used for the forecast error correlations. A general procedure for generating anisotropic from isotropic correlation models through a stretching mapping was given and successfully applied. The choice of the stretching mapping is determined by the application. In the ozone system the anisotropy and the mapping for horizontal forecast error correlations are determined by the distribution of observations. The correlations are falling off faster in the data dense than in data sparse regions and directions. The resulting horizontal forecast error correlations are anisotropic with longer correlation length in the zonal than in the meridional direction. The degree of anisotropy is the largest in the tropics with the longest length scale in the zonal (data sparse) direction. The maximum likelihood estimation of forecast error length scales and a simplified Kalman filter experiment both support this choice of an anisotropic correlation model. Vertical forecast error correlations are modeled using another anisotropic function. In addition, this function is zero for grid points separated by more than 5 model levels in order to limit computational burden in the statistical analysis.

The following examples from validation in winter 1992 period were presented. Statistical validation of analyzed fields was done against SBUV and independent (HALOE and ozonesondes) observations. The root-mean-square of TOMS observed-minus-forecast residuals is less than 4%. The ozone system was found to be able to reproduce some variability in ozone fields. For example, the analyzed fields were in an excellent agreement with independent ozonesonde observations within a "minihole". The wave one feature in tropical analyzed ozone is confined to the tropospheric part of the column, consistent with the vertical extent of this feature found by Ziemke et al. (1996). The accuracy of error covariance models was evaluated using χ^2 related statistics showing a large improvement in covariance modeling compared to an initial version (Riishøjgaard et al. 2000) of the system.

Thus, an ozone system was developed and it was found to perform well during the validation period in the winter of 1992. The ozone fields that are assimilated using this system are available to scientific community in near-real-time since the launch of NASA's

Terra satellite.

Acknowledgments

Appendix

The function ν, used in defining the vertical forecast error correlation function g, is a member of the family of compactly supported correlation functions (depending on a parameter e) constructed by Gaspari and Cohn (1999) defined by

$$\nu(r) = \begin{cases} \nu_1(r) & 0 \le |r| < 1/2 \\ \nu_2(r) & 1/2 \le |r| < 1 \\ \nu_3(r) & 1 \le |r| < 3/2 \\ \nu_4(r) & 3/2 \le |r| < 2 \\ 0 & 2 \le |r|, \end{cases} \tag{26}$$

where the even functions $\nu_1, \ldots, \nu_4$ are defined for $r \ge 0$ by

$$\begin{aligned}
\nu_1(r) &= -\frac{1}{3d}\Big(48er^4 - 48e^2r^4 + 60er^3 - 9e + 160e^2r^2 - 40er^2 - 66e^2 \\
&\quad - 120e^2r^3 + 20r^2 + 56e^2r^5 - 64er^5 - 15r^3 + 24r^5 - 24r^4 - 3\Big), \\
\nu_2(r) &= \frac{1}{12rd}\Big(-120r^2 - 96r^5 + 32r^6 - 4 + 48r + 29e - 42e^2 + 80r^3 \\
&\quad + 160e^2r^6 - 192er^6 + 60r^4 + 612e^2r - 880er^3 + 800e^2r^3 \\
&\quad - 1080e^2r^2 + 780er^2 - 210er - 384e^2r^5 + 480er^5\Big), \\
\nu_3(r) &= -\frac{e}{12rd}\Big(243 - 230e + 96er^6 - 64r^6 - 720r^3 + 1620r^2 \\
&\quad - 1134r + 732er - 1320er^2 + 240er^4 + 288r^5 - 384er^5 \\
&\quad - 240r^4 + 800er^3\Big), \\
\nu_4(r) &= \frac{4e^2}{3rd}\Big(-120r^2 - 16 + 96r + 40r^3 + 2r^6 - 12r^5 + 15r^4\Big),
\end{aligned}$$

and $d = 3e + 22e^2 + 1$.

References

Austin. J. Toward the Four Dimensional Assimilation of Stratospheric Constituents *J. Geophys. Res.*, 97, 2569-2588, 1992.

Bhartia. P. K.. J. Herman. R. D. McPeters, and O. Torres, Effect of Mount Pinatubo Aerosols on Total Ozone Measurements From Backscatter Ultraviolet (BUV) Experiments. *J. Geophys. Res.*, 98, 18,547-18,554, 1993.

Bhartia. P.K., R.D. McPeters. C.L. Mateer, L.E. Flynn, and C. Wellemeyer, Algorithm for the estimation of vertical ozone profile from the backscattered ultraviolet (BUV) technique, *J. Geophys. Res.*, 101,18793-18806, 1996.

Bruhl. C.. S. R. Drayson. J. M. Russell III, P. J. Crutzen, J. M. McInerney, P. N. Purcell, H. Claude, H. Gernandt, T. J. McGee. I. S. McDermid. and M. R. Gunson. Halogen Occultation Experiment ozone channel validation, *J. Geophys. Res.*, 101, 10217-10240, 1996.

Carton. J. A., and E. C. Hackert, Data assimilation applied to the temperature and circulation in the tropical Atlantic, 1983-84, *J. Phys. Oceanogr.*, 20, 1150-1165, 1990.

Cohn. S. E., Dynamics of Short-Term Univariate Forecast Error Covariances, *Mon. Wea. Rev.*, 121, 3123-3149. 1993.

Cohn. S. E., Introduction to Estimation Theory, *J. Met. Soc. Japan*, 75, 257-288, 1997.

Cohn S. E., A. da Silva, J. Guo, M. Sienkiewicz and D. Lamich, Assessing the Effects of Data Selection with the DAO Physical-space Statistical Analysis System, *Mon. Wea. Rev.*, 126, 2913-2926, 1998.

DAO. *Algorithm Theoretical Baseline Document Version 1.01*, Data Assimilation Office, NASA Goddard Space Flight Center. 1996.

Daley. R.. *Atmospheric Data Analysis*, Cambridge University Press. Cambridge, 457pp., 1991.

Dee. D. P.. Simplification of the Kalman filter for meteorological data assimilation, *Q. J. R. Meteorol. Soc.*, 117. 365-384, 1991.

Dee. D. P. and A. M. da Silva. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology *Mon. Wea. Rev.*, 124, 1822-1834, 1999.

Derber, J. and A. Rosati. A global oceanic data assimilation system. *J. Phys. Oceanogr.*, 19, 133-1347, 1989.

Douglass. A. R., C. J. Weaver. R. B. Rood and L. Coy, A three-dimensional simulation of the ozone annual cycle using winds from a data assimilation system, *J. Geophys. Res.*, 101, 1463-1474, 1996.

Gaspari. G. and S. E. Cohn. Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, 125, 723-757, 1999.

Grainger. S., and R. J. Atkinson, Three dimensional analysis of the global ozone distribution by Statistical Interpolation of satellite ozone data *Proceedings of the XVIII Quadrennial Ozone Symposium, L'Aquila, Italy. 1996, ed. R. D. Bojkov and G. Visconti*, pp. 41-44, International Ozone Commission.

Kalnay. E.. S. J. Lord, and R. D. McPherson, Maturity of Numerical Weather Prediction: Medium Range, *Bull. Amer. Meteor. Soc..* 79, 2753-2769, 1998.

Khattatov, B. V., J. C. Gille. L. V. 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. 18715-18737, 1999.

Kondratyev, K. Ya, A. A. Buznikov, O. M. Pokrovskii, and Yu. B. Yanushanets, Methods of Assimilating Satellite Information for Analysis and Prediction of Atmospheric Ozone, *Sov. J. Remote Sensing*, 10(3), pp. 407-420, 1993.

Levelt, P. F., M. A. F. Allaart and H. M. Kelder, On Assimilation of Total Ozone Satellite Data, *Ann. Geophysicae* 14, 1111-1118, 1996.

Levelt, P. F., B. V. Khattatov, J. C. Gille, G. P. Brasseur, X. X. Tie, and J. W. Waters, Assimilation of MLS ozone measurements in the global three-dimensional chemistry transport model ROSE, *Geophys. Res. Lett.*, 25, 4493-4496. 1998.

Lin, S.-J. and R. B. Rood, Multidimensional Flux-Form Semi-Lagrangian Transport Schemes, *Mon. Wea. Rev.*, 124, 2046-2070. 1996.

Lyster, P. M., S. E. Cohn. R. Ménard, L.-P. Chang, S.-J. Lin, and R. G. Olsen, Parallel Implementation of a Kalman filter for Constituent Data Assimilation, *Mon. Wea. Rev.*, 125, 1674-1686, 1997.

McPeters. R.D., P. K. Bhartia, A. J. Kruger, J. R. Herman, B. M. Schlesinger, C. G. Wellemeyer, C. J. Seftor, G. Jaross, S. L. Taylor, T. Swissler, O. Torres, G. Labow. W. Byerly, and R. P. Cebula, *Nimbus-7*

21

Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide, NASA Reference Publication 1384, National Aeronautical and Space Administration, Washington DC 1996.

Ménard, R., Tracer assimilation, *Inverse Methods in Global Biogeochemical Cycles*, AGU Geophysical Monograph Series vol. 114 (eds. Kasibhatla, P., M. Heimann, P. Rayner, N. Mahowald, R. Prinn, and D. Hartley), 67-79. AGU, Washington, DC, 2000.

Riishøjgaard, L. P., I. Štajner and G.-P. Lou, The GEOS Ozone Data Assimilation System *Adv. Space Res., in press.* Available from http://dao.gsfc.nasa.gov/DAO_people/ivanka/

Shiotani, M., Annual, quasi-biennial and El Nino-Southern Oscillation (ENSO) timescale variations in equatorial total ozone. *J. Geophys. Res.*, 97, 7625-7633, 1992.

Strang, G., *Introduction to Applied Mathematics* Wellesley-Cambridge Press, Wellesley, Mass., 1986.

Torres, O. and P. K. Bhartia, Effect of stratospheric aerosol on ozone profile from BUV measurements *Geophys. Res. Lett.*, 22, 235-238, 1995a.

Torres, O., J. R. Herman, P. K. Bhartia, and Z. Ahmad, Properties of Mount Pinatubo aerosols as derived from Nimbus 7 total ozone mapping spectrometer measurements *J. Geophys. Res.*, 100, 14,043-14,055, 1995b.

Vaughan, G. and C. Timmis, Transport of near-tropopause air into the lower midlatitude stratosphere *Q. J. R. Meteorol. Soc.*,124, 1559-1578, 1998.

Ziemke, J. R., S. Chandra, A. M. Thompson, and D. P. McNamara, Zonal asymmetries in southern hemisphere column ozone: Implications of biomass burning, *J. Geophys. Res.*, 101, 14,421-14,427, 1996.
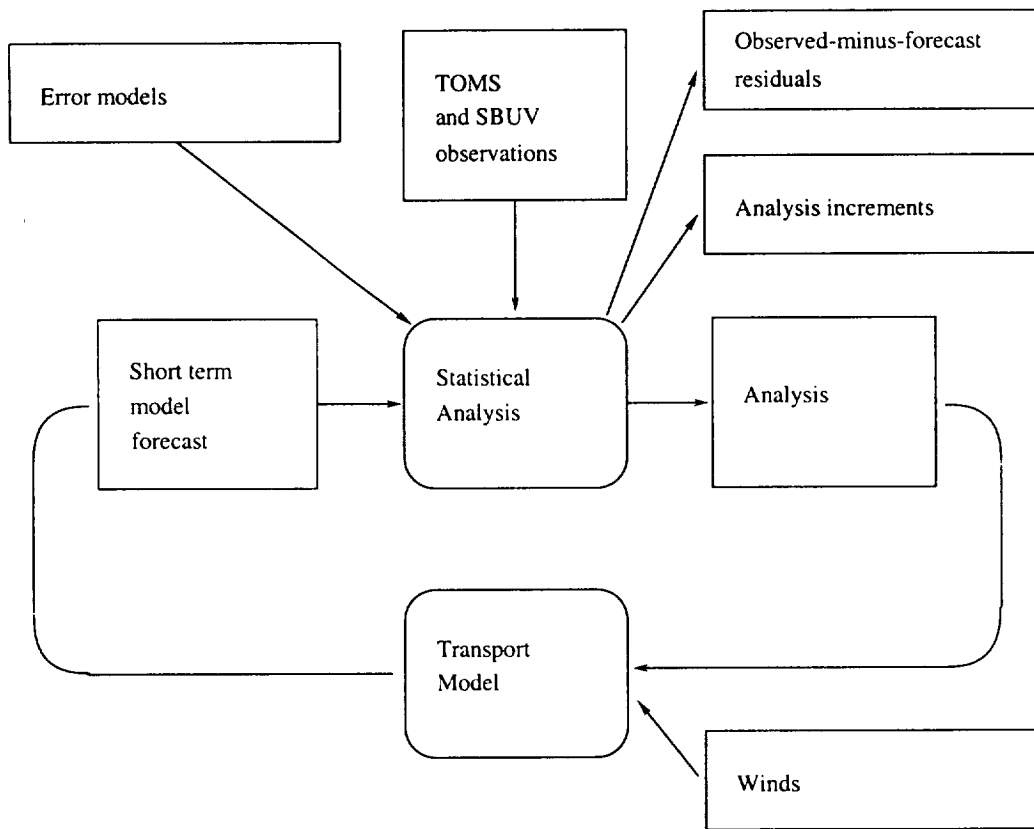
Figure 1: Schematic diagram of the ozone data assimilation system.
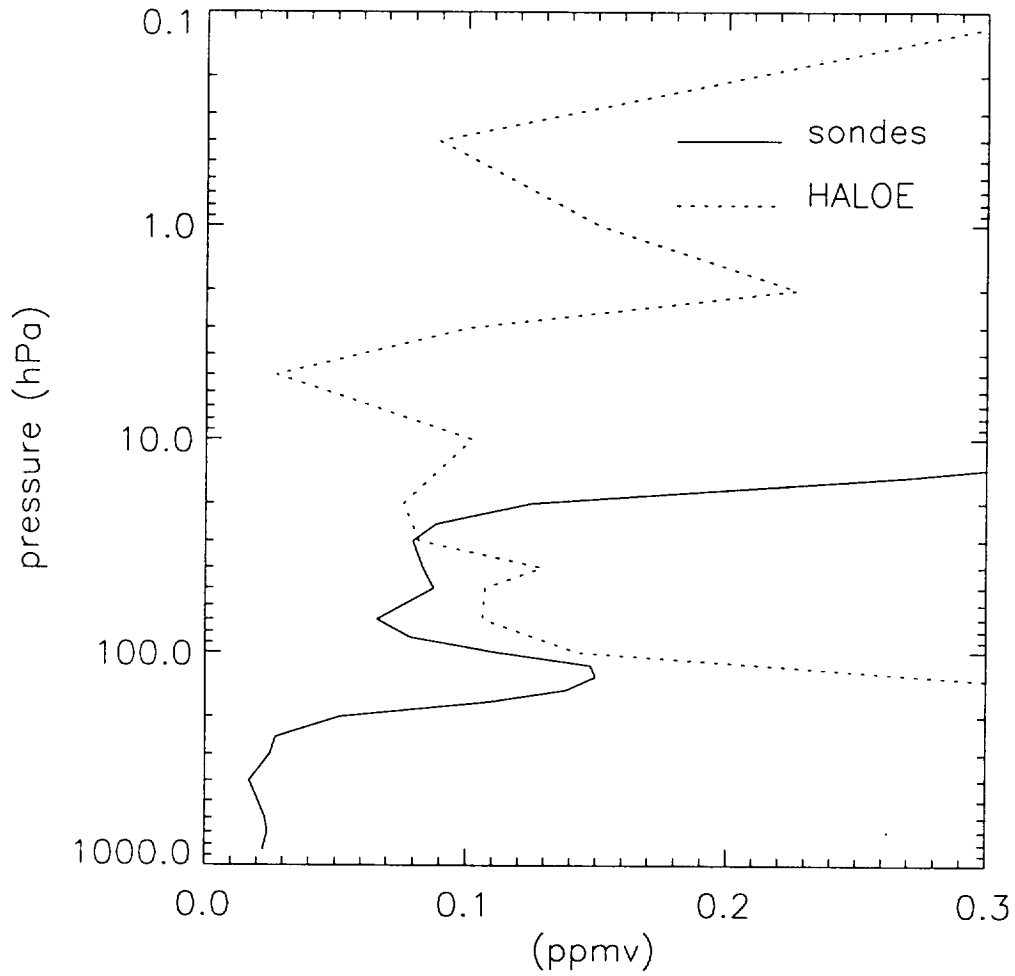
Figure 2: The ratio of the mean square difference between independent observations and analysis over the mean of the observations is shown for HALOE and ozone sondes.

Mapping between sphere and cylindrical surface
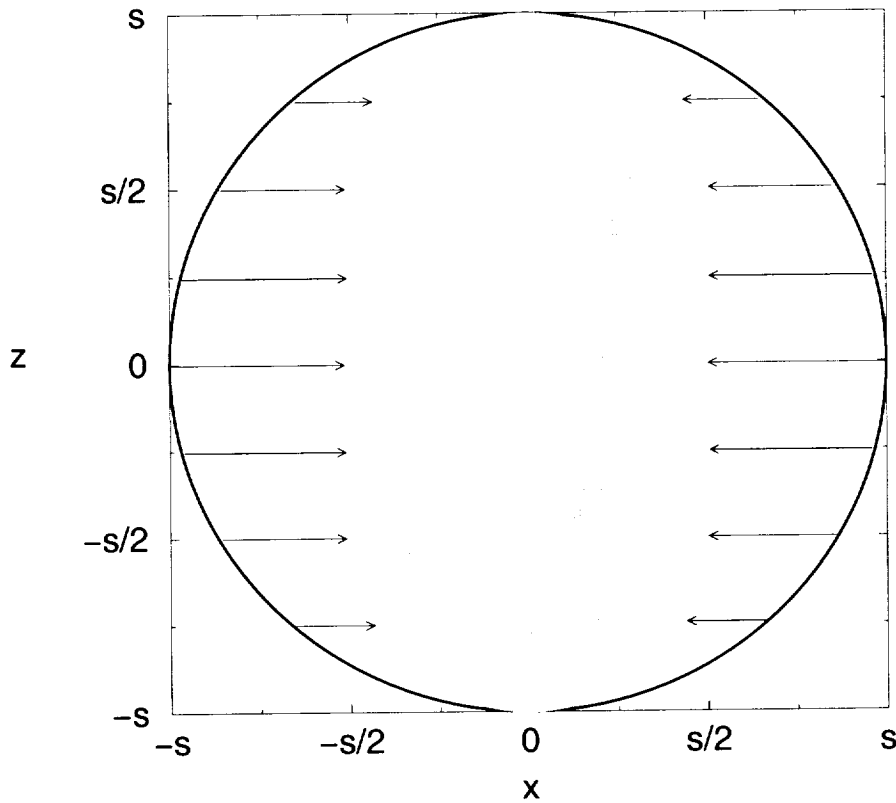
Figure 3: The action of the mapping τ in the plane $y = 0$ is illustrated. The sphere of radius s (representing Earth's surface) is mapped to a surface consisting of a finite cylinder and two hemispheres.

Total ozone analysis at 12 on 920129
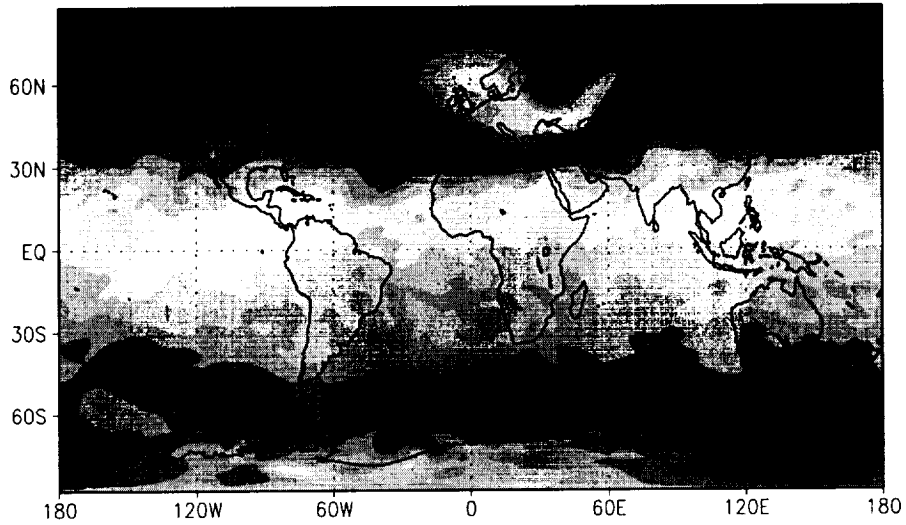
TOMS observed total ozone on 920129

250 270 290 310 330 350 370 390 410 430
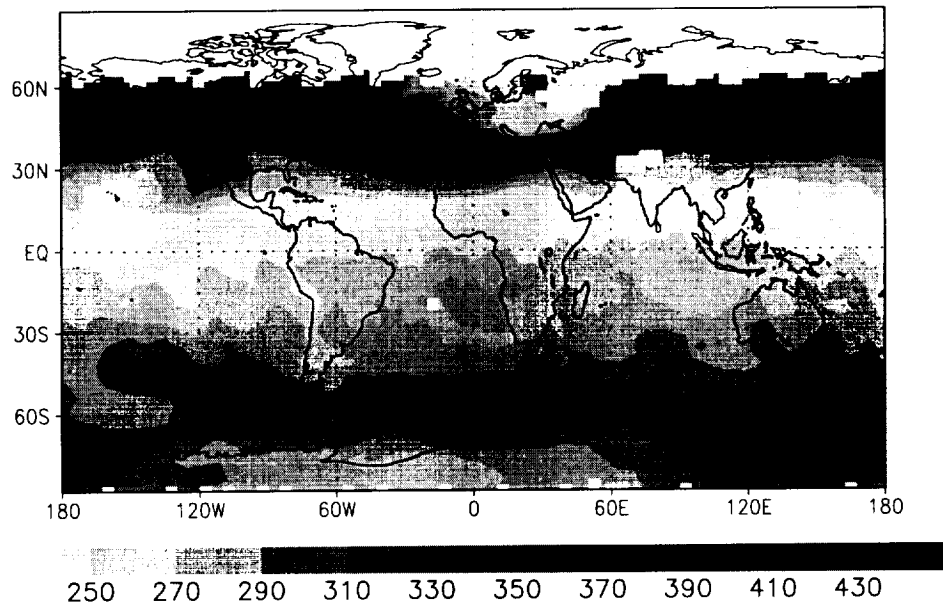
Figure 4: The analyzed total ozone field at noon (upper plate) and the daily TOMS level 3 data (lower plate) on February 15, 1992 are shown.

Global daily RMS of TOMS
observed–minus–forecast residuals
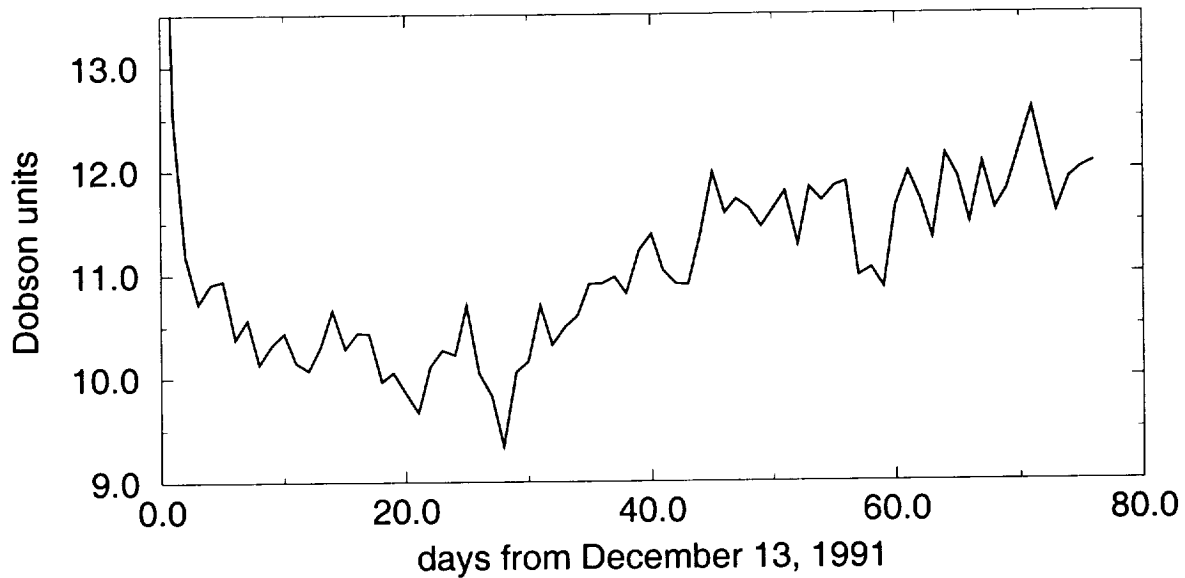
Figure 5: The evolution of daily global RMS of TOMS observed-minus-forecast residuals in the GEOS ozone DAS during the validation period in winter 1992 is shown.

Daily zonal RMS diff. of total ozone
TOMS observations and forecast
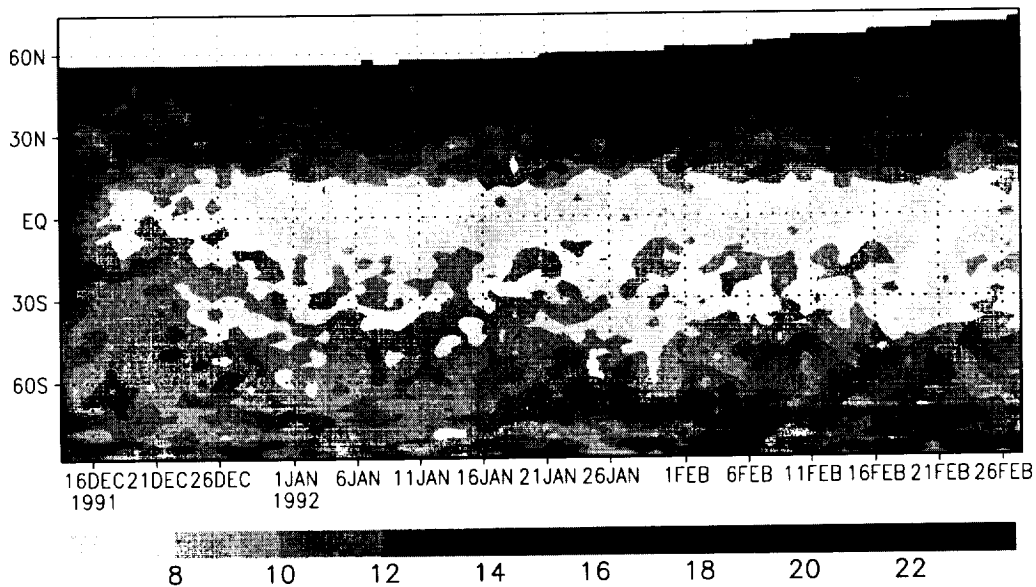
Figure 6: The evolution of daily zonal RMS of TOMS observed-minus-forecast residuals in the GEOS ozone DAS during the validation period in winter 1992 in shown.
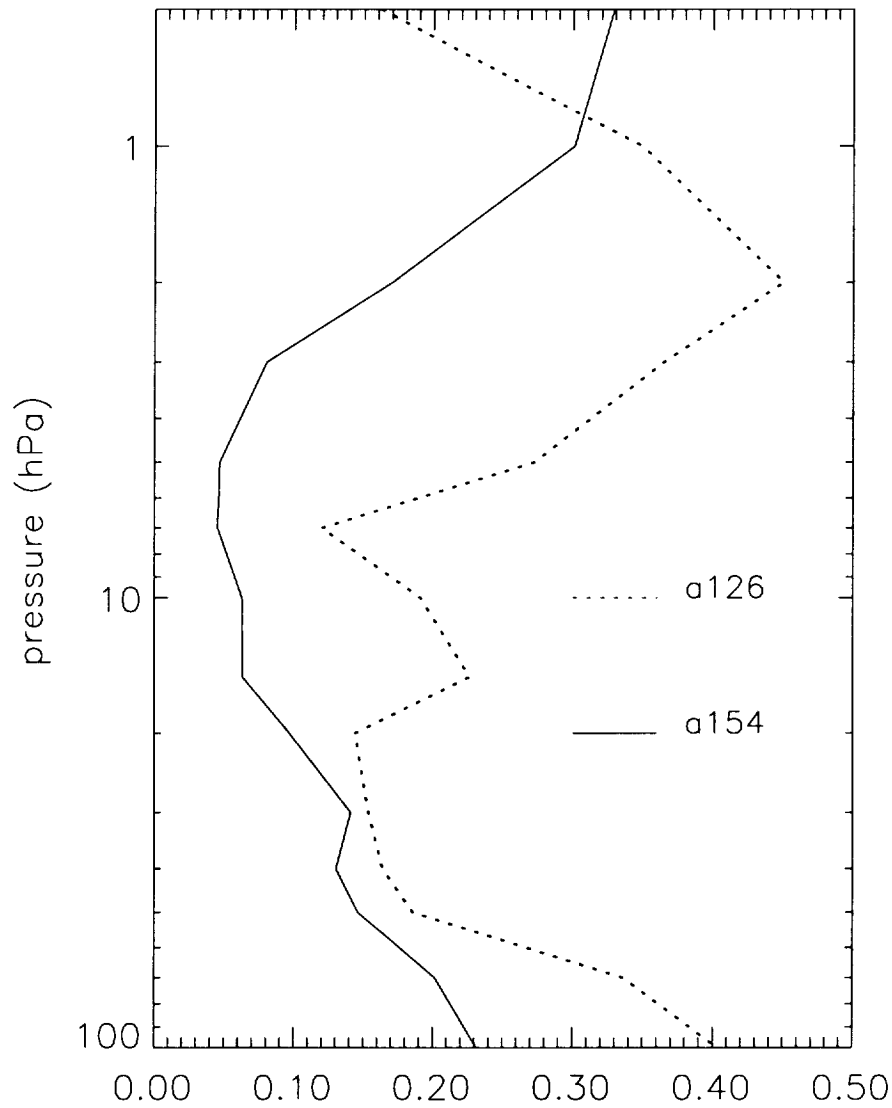
Figure 7: The RMS difference between SBUV observations and GEOS ozone DAS analyzed fields relative to the mean SBUV measurements for the current version (solid curve) and the system described by Riishøjgaard *et al.* (2000) which included a biased chemistry model (dashed curve) are shown for January 20, 1992.

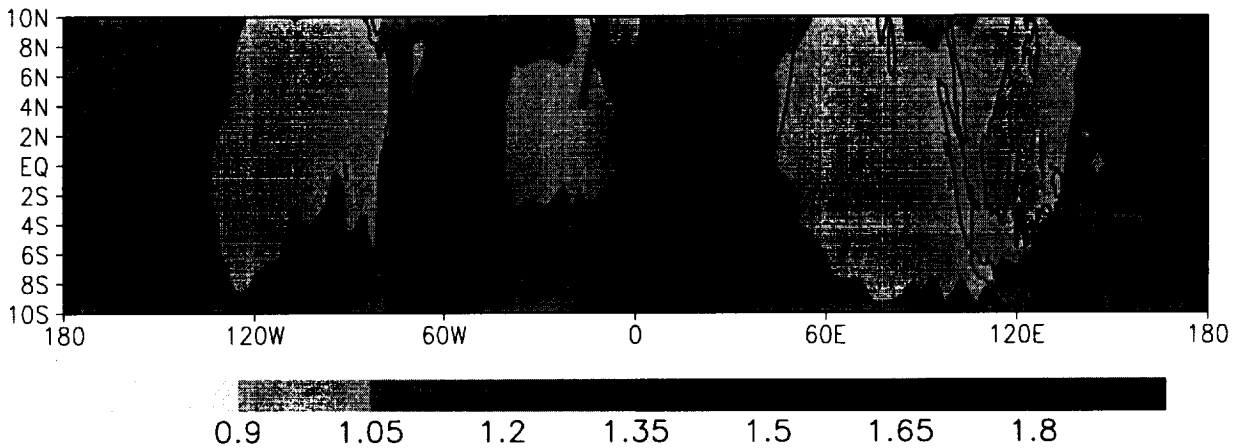# Mean analyzed ozone mixing ratio in Jan. 92 [ppmv] at 150 hPa

| 0.06 | 0.07 | 0.08 | 0.09 | 0.1 | 0.11 | 0.12 |

at 50 hPa

| 0.9 | 1.05 | 1.2 | 1.35 | 1.5 | 1.65 | 1.8 |
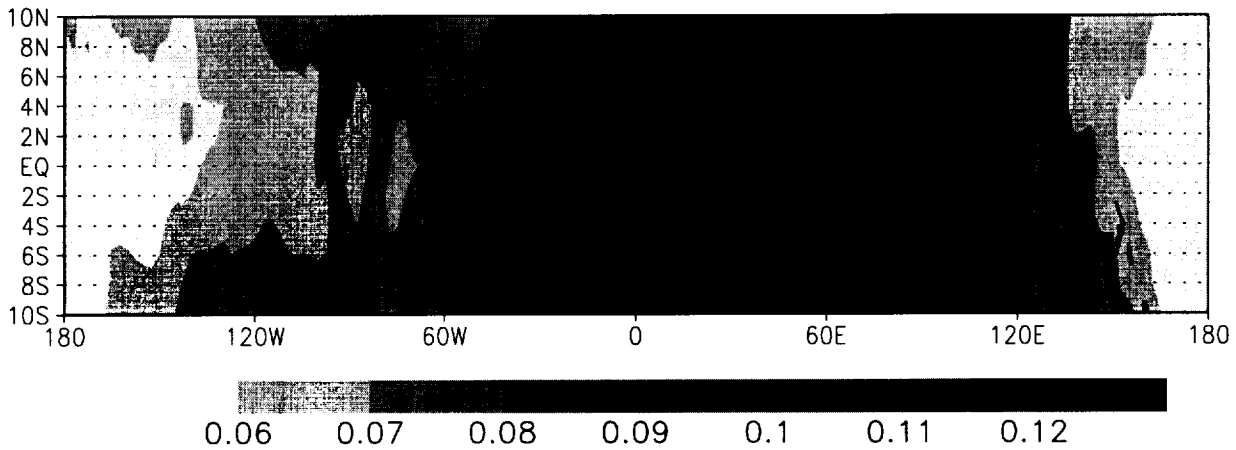
Figure 8: Mean January 1992 GEOS ozone DAS analyzed fields are shown. At 150 hPa high values are over Africa, the Indian Ocean and the southern Atlantic while low values are over Pacific (upper plate). Less of the zonal variability is present at 50 hPa (lower plate).
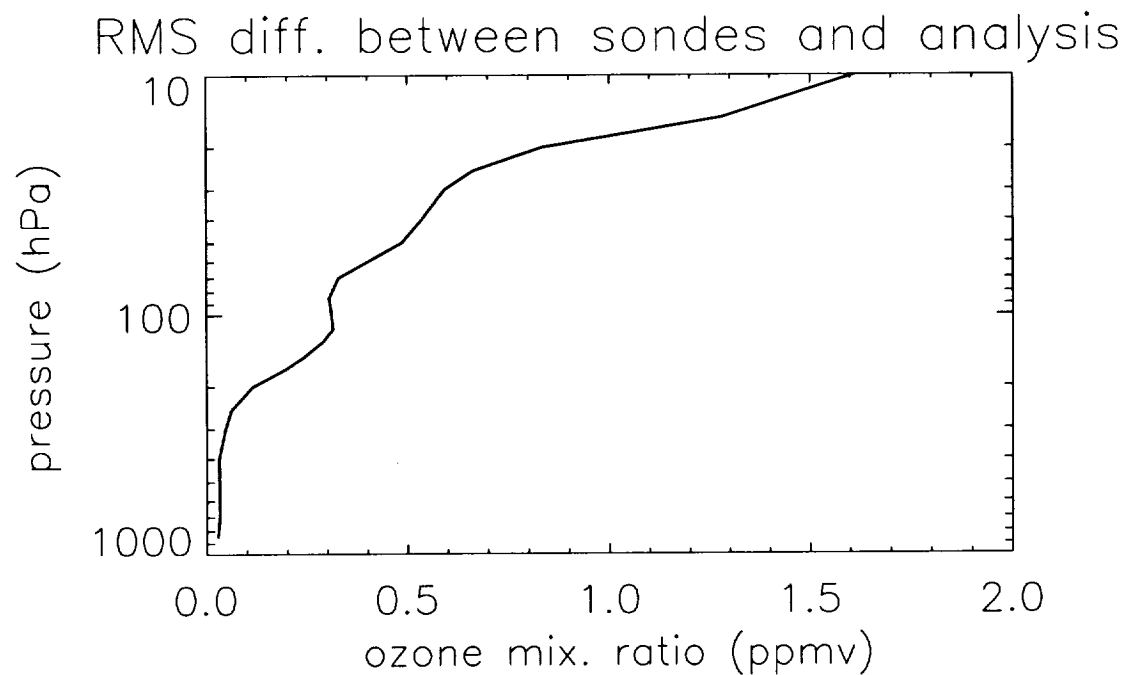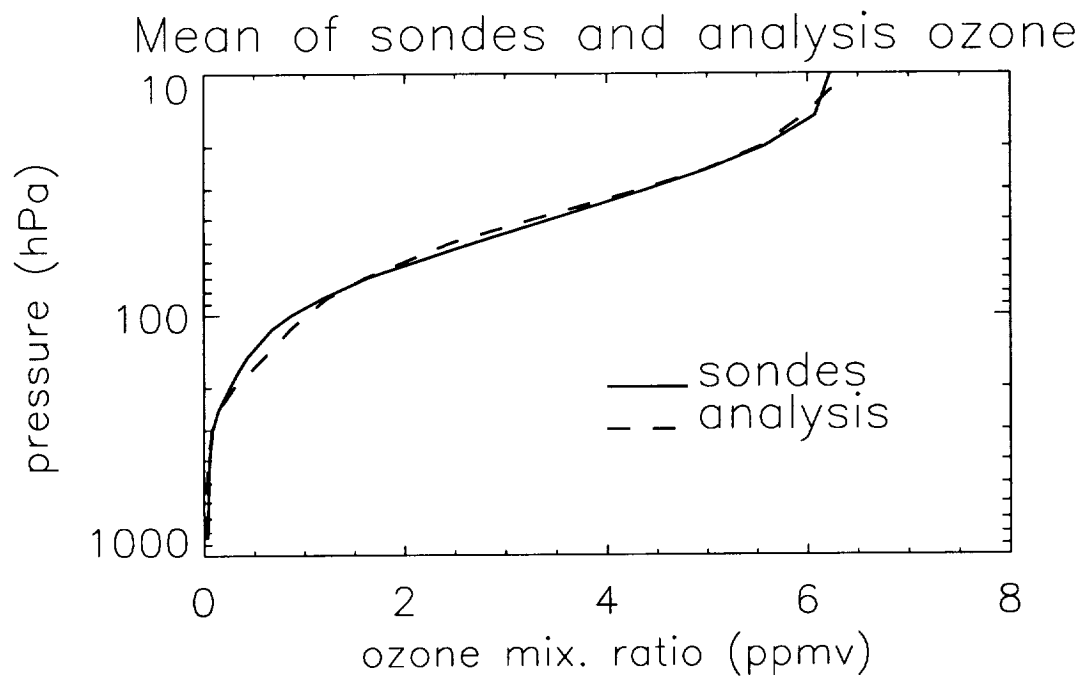
Figure 9: The means of ozone sonde observations (solid curve) and analyzed fields (dashed curve) between December 15, 1991 and February 27, 1992 are shown in the upper plate. The RMS difference between ozone sonde observations and analyzed fields is shown in the lower plate.
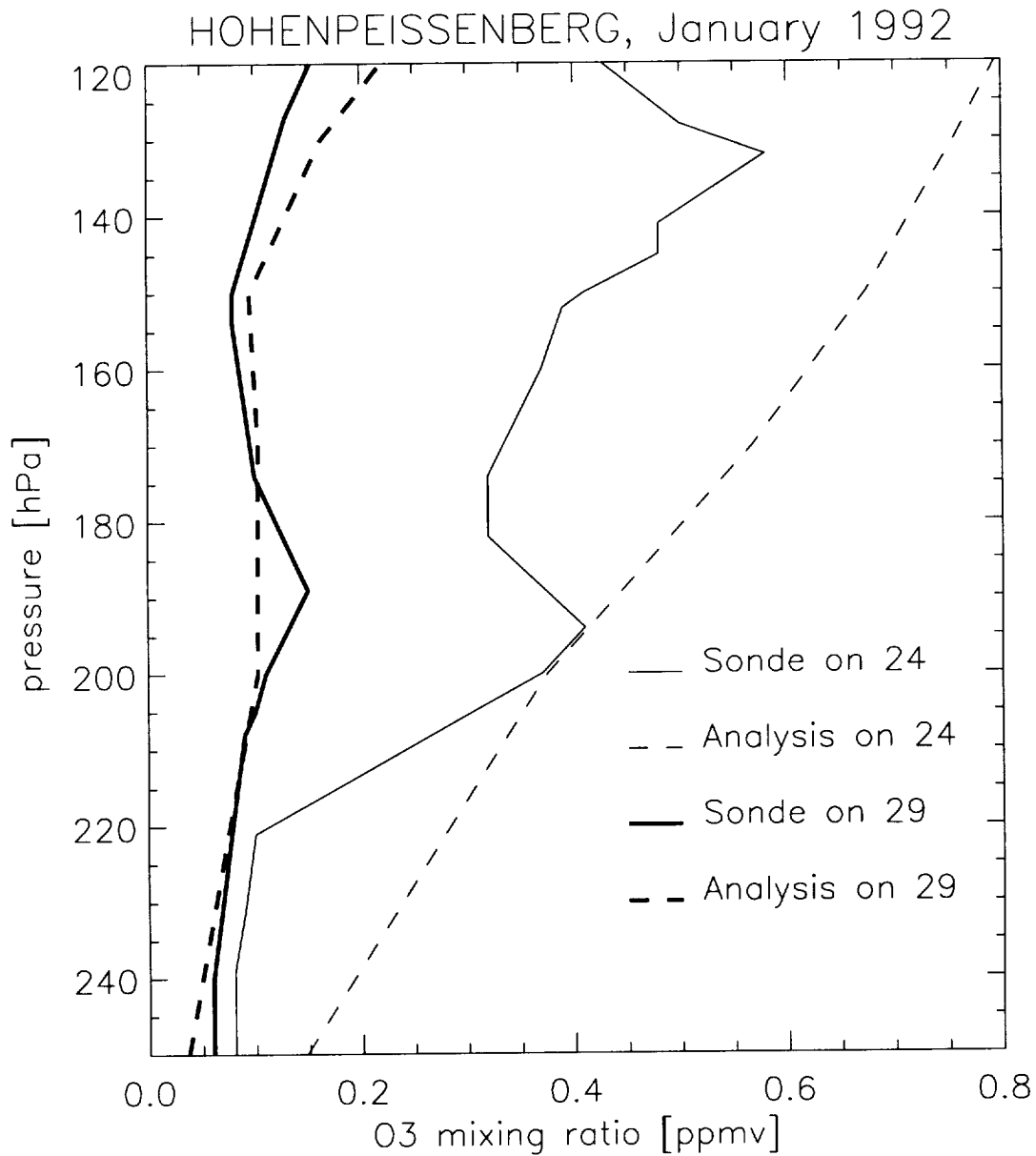
Figure 10: Partial profile of ozone sonde measurements and GEOS ozone DAS analyzed ozone fields at Hohenpeissenberg (47.8° lat., 11.02° lon.) before (Jan. 24, 1992) and during (Jan. 29, 1992) the intrusion of the tropical tropospheric air are shown.
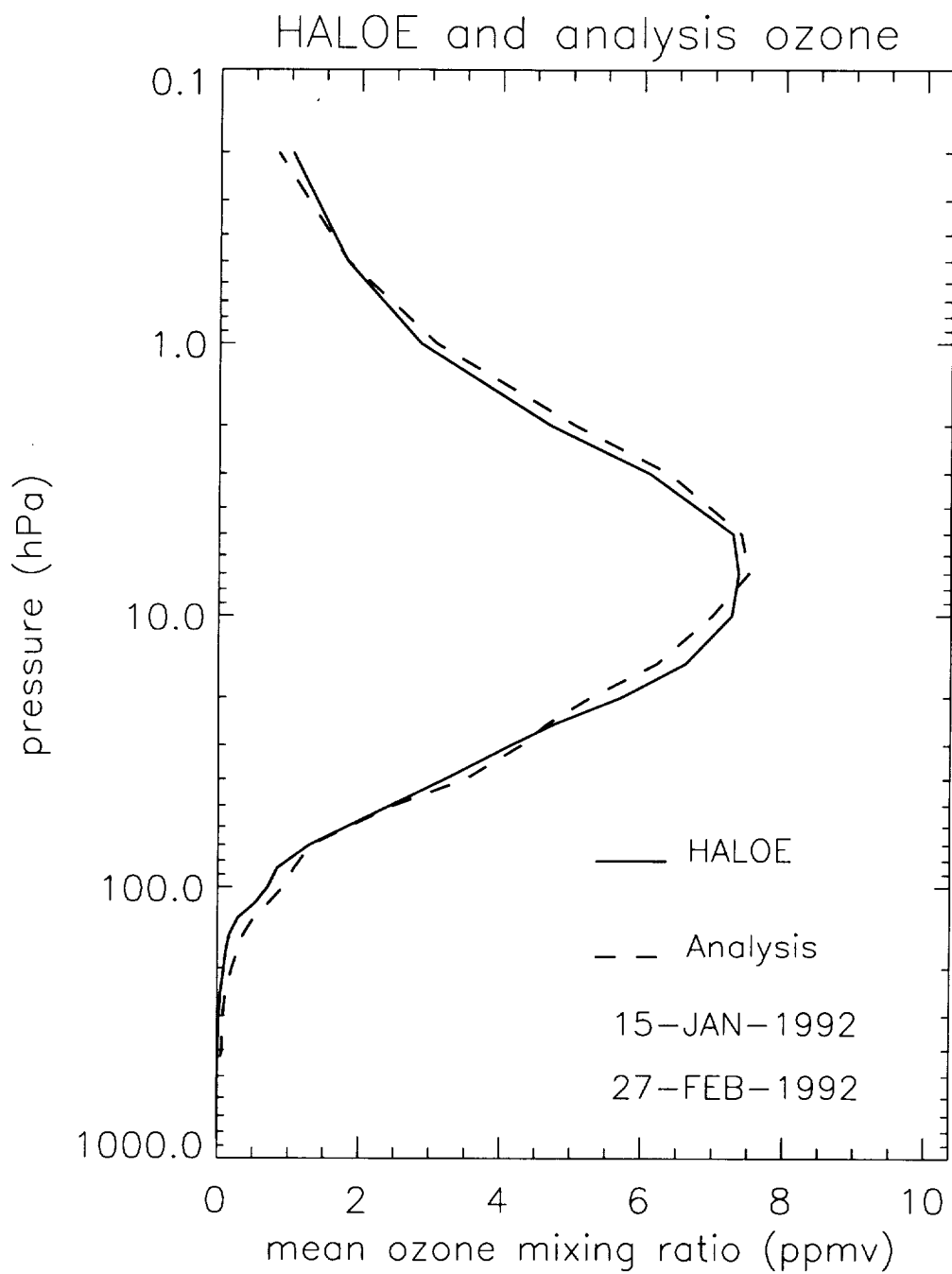
Figure 11: The mean ozone measured by HALOE at sunrise and the analyzed ozone for the period between January 15 and February 27, 1992 are shown.