Intercomparison of direct and indirect measurements: Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) versus sonde ozone profiles

Stefano Migliorini
Data Assimilation Research Centre, Department of Meteorology, University of Reading, Reading, UK

Chiara Piccolo and Clive D. Rodgers
Clarendon Laboratory, Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, UK

Received 5 May 2004; revised 13 July 2004; accepted 15 July 2004; published 14 October 2004.

[1] The theoretical characterization of the specific relation between a measurement and the physical quantities we want to determine, along with an assessment of the accuracy of the instrument, form the instrument model. In order to check whether the instrument model is correct, the measurements need to be compared with independent measurements whose characterization and error analysis is better understood. A remote sounding instrument provides measurements that are indirectly related to the atmospheric state. In this paper a general technique for comparing indirect with direct measurements is discussed, applicable to remote sounding measurements obtained using both optimal estimators and maximum likelihood methods. A specific application to validation of Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) ozone retrievals using ozonesondes during winter 2002–2003 is described. No significant bias between the two instruments and a confirmation of the theoretical MIPAS accuracy assessment, after a discussion of specific comparisons, were found. *INDEX TERMS:* 3360 Meteorology and Atmospheric Dynamics: Remote sensing; 0340 Atmospheric Composition and Structure: Middle atmosphere—composition and chemistry; 0394 Atmospheric Composition and Structure: Instruments and techniques; *KEYWORDS:* validation, sonde, satellite, accuracy, error, intercomparison

Citation: Migliorini, S., C. Piccolo, and C. D. Rodgers (2004), Intercomparison of direct and indirect measurements: Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) versus sonde ozone profiles, *J. Geophys. Res.*, *109*, D19316, doi:10.1029/2004JD004988.

1. Introduction

[2] Validation is the process of assessing the overall confidence in a reported measurement result. Confidence is often derived from intercomparisons between a measurement result, taking into account all its error components, and available reference measurements.

[3] In situ measurement campaigns take an essential part in the validation effort because they provide a direct estimate of the quantity to be investigated, even if they are representative of a rather localized area and relatively expensive.

[4] Satellite measurements, which imply remote sounding of the atmosphere, allow global coverage and geographic variability with limitations in both spatial resolution and short-term variability.

[5] The purpose of an intercomparison is to determine whether different observing systems agree within their known limitations. When independent measurements are compared, care is needed to take into account all their error components and, in the case of indirect measurements, the characterization will be expressed in terms of averaging kernels and error covariance [*Rodgers and Connor*, 2003].

[6] In this paper a comparison between satellite ozone retrievals from the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) and sonde ozone profiles is described.

[7] The sonde measures the ozone profile at high vertical resolution, better than that of the satellite. *Connor et al.* [1994] have proposed that this difference can be accommodated by applying the satellite averaging kernels to the sonde data before comparison. Unfortunately in the case of MIPAS, nonlinearity results in this approach causing significant errors, so we have used the more direct, but time consuming, method of simulating the MIPAS retrieval that would correspond to each sonde measurement.

2. Methodology

[8] To compare retrievals from a satellite instrument with direct measurements from in situ instruments, we need to allow for the several ways in which the characteristics of the two data sets differ. The main considerations are: (1) the relatively low vertical resolution of the satellite data, (2) the different representations and altitude grids used,

0148-0227/04/2004JD004988

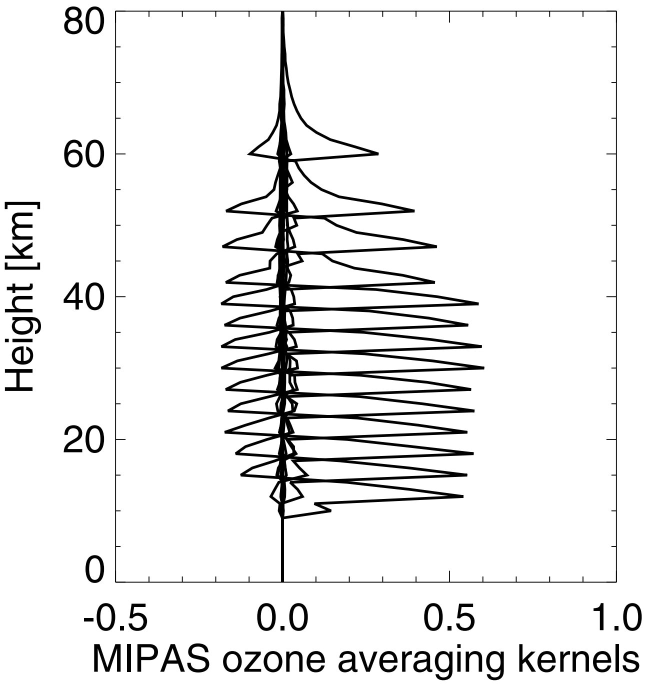

Figure 1. MIPAS ozone averaging kernels for the 14 nominal retrieval levels, calculated for atmospheric conditions typical of the Arctic region in January [from *Ceccherini and Ridolfi*, 2002].

(3) the range of altitudes, and (4) the different locations and horizontal sampling of the instruments.

[9] *Connor et al.* [1994] have proposed a method to compare a high-resolution profile with a low-resolution one. In this, the characterization of the satellite data in terms of an averaging kernel representation [*Rodgers*, 1990, 2000] is applied to the high spatial resolution profile of the in situ measurement to produce a simulation of the lower resolution retrieval that the satellite measurement would be expected to produce if all is well.

[10] We have found practical difficulties in implementing the approach of Connor et al. because of the nonlinearity of the forward model and because the MIPAS retrieval grid comprises the actual rather than nominal tangent levels at which the measurements were taken, which can vary from one profile to the next. As a result, the averaging kernels vary. However, the MIPAS operational retrieval software does not provide them routinely. The nominal ozone averaging kernels (Figure 1), while giving a good indication of the resolution of the retrievals, are not adequate for retrieval simulation.

[11] We have therefore used the more expensive but more accurate method of simulating the satellite measurement and retrieval, taking the in situ measurement as input, and have used the result to compare with the actual satellite retrieval. This has the advantage of dealing with both nonlinearity and the varying MIPAS retrieval grid, because the forward calculation can be carried out on a relatively fine grid, and the simulated retrieval on the actual MIPAS retrieval grid, which corresponds to the tangent altitudes at which the atmospheric spectra were measured, nominally spaced at 3 km intervals.

[12] To deal with the question of differing ranges of altitudes, we only make the comparison over the common range. However, to carry out the forward calculation, we need to extend the sonde range upward, which we do by appending the MIPAS retrieval above the sonde height range and a climatology above the highest MIPAS retrieval level.

[13] The next step is to carry out an error analysis to determine what differences we might expect between the real MIPAS retrievals and the simulated ones.

[14] Sonde measurements are made on a sufficiently fine grid that we assume that structure on a finer scale can be ignored. We regard the measured profile $\hat{\mathbf{x}}_s$ as a direct estimate of the true state on the sonde grid $\mathbf{x}_s$, with measurement error ε_s:

$$\hat{\mathbf{x}}_s = \mathbf{x}_s + \varepsilon_s \tag{1}$$

The MIPAS forward model requires its input on a calculation grid with a spacing of 1 km from 0–120 km, so we define the atmospheric state vector $\mathbf{x}$ on the MIPAS calculation grid. The sonde measurement is transformed to this grid using an interpolation matrix $\mathbf{W}$, so that $\mathbf{x} = \mathbf{W}\mathbf{x}_s$.

[15] The simulated retrieval $\hat{\mathbf{z}}_s$ is related to the true state on the sonde grid according to

$$\hat{\mathbf{z}}_s = \mathbf{r}\left(\mathbf{F}\left(\mathbf{W}\mathbf{x}_s + \mathbf{W}\varepsilon_s, \hat{\mathbf{b}}\right), \hat{\mathbf{b}}, \mathbf{c}\right) \tag{2}$$

where $\mathbf{r}(\mathbf{y}, \mathbf{b}, \mathbf{c})$ is the retrieval method applied to a measurement $\mathbf{y}$ with forward model parameters $\mathbf{b}$, $\mathbf{F}$ is the forward model, $\hat{\mathbf{b}}$ is a vector of forward model parameter estimates, and $\mathbf{c}$ is a vector of inverse method parameters. Note that the MIPAS retrieval does not use an explicit a priori profile.

[16] We expect the MIPAS retrieval $\hat{\mathbf{z}}$ to be related to the true state by

$$\hat{\mathbf{z}} = \mathbf{r}\left(\mathbf{f}(\mathbf{x}_s, \mathbf{b}) + \varepsilon, \hat{\mathbf{b}}, \mathbf{c}\right) \tag{3}$$

where $\mathbf{f}$ represents the true forward physics, with true forward model parameters $\mathbf{b}$ and ε is the measurement noise. Defining the error in the forward model as

$$\Delta\mathbf{f} = \mathbf{F}(\mathbf{W}\mathbf{x}_s, \mathbf{b}) - \mathbf{f}(\mathbf{x}_s, \mathbf{b}) \tag{4}$$

and assuming that both the forward model and the retrieval method are linear within the error bounds of the comparison, we find the difference $\delta\hat{\mathbf{z}} = \hat{\mathbf{z}} - \hat{\mathbf{z}}_s$ to be

$$\delta\hat{\mathbf{z}} = -\mathbf{A}\mathbf{W}\varepsilon_s + \mathbf{G}\Delta\mathbf{f} + \mathbf{G}\left(\mathbf{K}_b\left(\mathbf{b} - \hat{\mathbf{b}}\right) + \varepsilon\right). \tag{5}$$

The averaging kernel $\mathbf{A}$ is $\partial\mathbf{r}/\partial\mathbf{x}$, the retrieval gain $\mathbf{G}$ is $\partial\mathbf{r}/\partial\mathbf{y}$, and $\mathbf{K}_b$ is $\partial\mathbf{F}/\partial\mathbf{b}$.

[17] The error involved in interpolating between the sonde grid and the MIPAS calculation grid is included in $\Delta\mathbf{f}$. The MIPAS forward model does not allow us to calculate radiances on the sonde grid, but we anticipate that because both grids are relatively fine compared with the MIPAS retrieval grid, any errors due to this source will be small compared with the smoothing error. We have estimated the size of this error for a few sample cases by using a different forward model, the MIPAS Reference Forward Model (RFM) [*Dudhia*, 1996], and found it to be much smaller than

the MIPAS measurement noise, which in the case of ozone is between 18.5 and 35 nW cm^{-2} sr^{-1} $(cm^{-1})^{-1}$. The maximum root mean square and the maximum value of the interpolation error for these cases are 0.054 and 0.275 nW cm^{-2} sr^{-1} $(cm^{-1})^{-1}$, respectively.

[18] From equation (5) we find that the covariance matrix $\mathbf{S}_{\delta\hat{\mathbf{z}}}$ of the difference $\delta\hat{\mathbf{z}} = \hat{\mathbf{z}} - \hat{\mathbf{z}}_s$ between the satellite retrieval and the simulated satellite retrieval using sonde data is given by:

$$\mathbf{S}_{\delta\hat{\mathbf{z}}} = \mathbf{AWS}_s\mathbf{W}^T\mathbf{A}^T + \mathbf{GRG}^T \quad (6)$$

where $\mathbf{S}_s$ is the sonde error covariance matrix, the first term on the right hand side is a contribution from sonde error, and the second term is from MIPAS measurement error, $\mathbf{R} = \mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T + \mathbf{S}_\varepsilon$.

[19] To perform the comparison between MIPAS and sonde ozone profiles we obtain $\hat{\mathbf{z}}_s$ from equation (2) and $\hat{\mathbf{z}}$ from the nearest MIPAS profile.

[20] We compare $\boldsymbol{\mu} = E\{\delta\hat{\mathbf{z}}\}$, the expected value over the ensemble of $\delta\hat{\mathbf{z}}$, which should be zero, with the mean $\mathbf{m}$ of available coincidences. In order to evaluate whether $\delta\hat{\mathbf{z}} - \mathbf{m}$ belongs to an unbiased Gaussian distribution with covariance $\mathbf{S}_{\delta\hat{\mathbf{z}}}$ we examine its χ^2:

$$\chi^2 = (\delta\hat{\mathbf{z}} - \mathbf{m})^T\mathbf{S}_{\delta\hat{\mathbf{z}}}^{-1}(\delta\hat{\mathbf{z}} - \mathbf{m}) \quad (7)$$

where the covariance of the difference $\mathbf{S}_{\delta\hat{\mathbf{z}}}$ is evaluated from equation (6) using the nominal ozone averaging kernel matrix.

3. Data Set

3.1. Case Study

[21] The period under study is between mid-November 2002 and the beginning of February 2003, covering the northern hemisphere winter. The winter pole is characterized by the presence of the polar vortex, which forms as air cools and descends (diabatic circulation). When temperatures are below the nitric acid tetrahydrate (NAT) threshold inside the vortex, Polar Stratospheric Clouds (PSCs) can form and induce heterogeneous reactions that in presence of sunlight produce clorine atoms, which lead to ozone depletion [e.g., *Wayne*, 2000]. The Arctic vortex is less stable than the Antarctic because of a higher degree of wave activity. Nonetheless, Arctic ozone losses have been detected over the last few years. To investigate such issues, the European field campaign VINTERSOL (Validation of INTERnational Satellites and study of Ozone Loss) is taking place from late 2002 until mid 2004 (European Ozone Research Coordinating Unit (EORCU), Vintersol master document, 2002, available at http://www.ozone-sec.ch.cam.ac.uk/downloads/VINTERSOL.html).

[22] This choice provides a sample of profiles for which ozone is relatively variable, but because of the nature of the polar winter circulation and the existence of PSCs, it also provides a stringent test of the practicality of any approach to validation by intercomparison.

3.2. MIPAS Satellite Data

[23] The Michelson Interferometer for Passive Atmospheric Sounding, launched in March 2002 on board ESA's Envisat, is a Fourier transform spectrometer for the measurement of in limb emission spectra in the uppe ratmosphere [*European Space Agency*, 2000]. The operational configuration of MIPAS allows the determination of vertical profiles of pressure, temperature and volume mixing ratio of O_3, H_2O, HNO_3, CH_4, N_2O and NO_2 from the spectra, and provides global coverage and diurnal variation data sets.

[24] The retrieval scheme is based on a nonlinear least squares method without the use of explicit a priori information [*Ridolfi et al.*, 2000].

[25] Ozone profiles are currently retrieved on 14 nominal altitude levels, which cover the atmospheric range between 12 km and 60 km, vertically spaced by 3 km between 12 km and 42 km, with wider spacing above. Such vertical spacing corresponds approximately to the vertical resolution of each measurement as shown in Figure 1 where the MIPAS averaging kernels are represented.

[26] In this paper we have used profiles produced by using the version 4.53 of the level 2 processor which have been made available to the MIPAS Atmospheric Chemistry Validation Team since mid-November 2002.

3.3. Ozonesondes

[27] The in situ ozone measurements considered in this study are ozonesonde data that are available in the Envisat Validation Database, stored within the NILU Atmospheric Database for Interactive Retrieval (NADIR) at the Norwegian Institute for Air Research (NILU). Data consist of profiles of ozone partial pressure, pressure, temperature and height, up to a height of 30–35 km, with a vertical resolution of about 150 m.

[28] We considered all of the ozonesonde data within the period under investigation for which MIPAS data was available for the same day, and for which the ascent altitude increased monotonically and height, pressure and temperature information was given. We found 133 profiles.

[29] The observation locations and the number of observations considered are reported in Table 1.

3.4. Climatology

[30] The reference atmosphere profiles were taken from MIPAS reference atmospheres developed using the methodology of *Remedios* [1999].

[31] This compilation of climatological profiles is separated in the four seasons (January, April, July and October) and in six latitude bands (90S–65S, 65S–20S, 20S–0, 0–20N, 20N–65N, 65N–90N), and provides profiles of pressure, temperature and volume mixing ratio (VMR) of N_2, O_2, CO_2, O_3, H_2O, CH_4, N_2O, HNO_3, CO, NO_2, N_2O_5, ClO, HOCl, $ClONO_2$, NO, HNO_4, HCN, NH_3, F11, F12, F14, F22, CCl_4, COF_2, H_2O_2, C_2H_2, C_2H_6, OCS, SO_2, SF_6, in an altitude range 0–120 km with 1 km step.

[32] This reference atmosphere is based on observations, analyzed data and model integrations.

[33] The most important sources of observational and analyzed data are the UARS reference atmosphere project (URAP, http://hyperion.gsfc.nasa.gov/Analysis/UARS/urap/home.html) and ECMWF reanalysis data (1979–1993) [*Gibson et al.*, 1997], respectively.

[34] Chemistry transport models (CTMs) are particularly important for describing chemical processes which contribute to altering the concentration of atmospheric compounds. Stratospheric CTM data (M. Chipperfield, The SLIMCAT

Table 1. Ozonesondes

Location	Longitude, deg	Latitude, deg	Type[a]	Number of Coincidences[b]
De Bilt	5.18	52.10	ECC	2
Jokioinen	23.50	60.80	ECC	2
Legionowo	20.97	52.40	ECC	2
Ny-Ålesund	11.93	78.92	ECC	4 (13)
Payerne	6.95	46.82	ECC	2 (3)
Scoresbysund	−21.97	70.48	ECC	4 (5)
Sodankÿla	26.63	67.37	ECC	9
Thule	−68.74	76.53	ECC	5

[a]ECC stands for electrochemical concentration cell.
[b]Number in parentheses is the total number of coincidences, regardless of the result of the cloud detection check (see text).

offline transport model, UGAMP Internal Rep. 39, 1996, available at http://www.env.leeds.ac.uk/~martyn/slimcat.html) available at 7.5° latitude spacing (multiyear runs) were supplied by M. Chipperfield (private communication, 1998), while tropospheric CTM data (from MOZART [*Brasseur et al.*, 1998]) available at 3° latitude spacing were provided by D. Hauglustaine (private communication, 1998) from a 1-year run.

4. Combination of Measurements

4.1. Coincidence Criteria

[35] In order to perform a meaningful comparison between two instruments, suitable coincidence criteria must be defined. The time interval between the two measurements should be small with respect to the timescales characteristic of the photochemical processes that lead to ozone production or destruction. For the altitude range of ozone sondes, this is generally more than a day. Furthermore, the two measurement locations should not be separated by regions where the potential temperature profile undergoes large excursions and should lie within their temporal separation times the atmospheric wind magnitude. However, if criteria are too strict, we might end up without having enough data to perform a statistical analysis.

[36] A possible way to relax such requirements is to use the so-called trajectory mapping technique [*Morris et al.*, 2000], which makes use of calculated isentropic trajectories of air parcels using transport models and analyzed wind fields. The use of such a technique is very expensive and beyond the scope of the present paper. Possible conventional coincidence criteria [e.g., *Harris et al.*, 1998] are ±2 degrees latitude, ±12.5 degrees longitude, and 2 days time, which generally enable the comparison of a sufficient number of measurements.

[37] In this work we have had access to a large enough ozonesonde data set (see Table 1), so that we can afford to specify a more stringent criterion, namely a 6-hour time interval and 400 km mutual distance. The distance corresponds approximately to spatial resolution of a limb sounder along the line of sight. We have required that there be a vertical range of at least four MIPAS levels containing data from both MIPAS and sonde. By using this criterion, 41 coincidences can be investigated, as summarized in Table 1.

4.2. Vertical Merge

[38] In order to be able to calculate the expression in equation 5, a measur $\hat{\mathbf{x}}_s$ of the atmospheric state vector is needed. The MIPAS forward model requires input on the 1 km spacing grid up to 120 km, including temperature, pressure and constituent mixing ratios. Ozonesondes provide temperature, pressure, and ozone only up to altitudes of 30–35 km, therefore extra sources of information must be included in $\hat{\mathbf{x}}_s$. MIPAS is the appropriate choice for extending the temperature, pressure and ozone profiles in the stratosphere provided the comparison is only carried out on that part of the profile covered by the ozonesonde. As the averaging kernel of MIPAS retrievals has a width comparable with the grid spacing, this procedure will affect at most only the top level of the intercomparison.

[39] Starting from the temperature and pressure at the highest altitude measured by the colocated sonde, height information is derived by integrating the MIPAS temperature and pressure profile upward using the hydrostatic equation.

[40] In order to extend the temperature, pressure and ozone profiles above the highest MIPAS tangent heights we made use of the climatological information described in section 3. It is advisable to avoid discontinuities at the junction level, so the climatologies were smoothly scaled by a cosine function over the 20 km above the join to make the temperature and ozone profiles continuous. The pressure information was determined by integrating the hydrostatic equation upward, making use of the climatological temperature information, the highest MIPAS pressure and temperature measurements and the height of the highest tangent level.

[41] Examples of the vertical merging are shown in Figures 2 and 3.

5. Error Budget

5.1. Ozonesonde Errors

[42] The ozonesondes available for the comparison with MIPAS profiles use an ECC (electrochemical concentration cell) sensing device, carried on a standard balloon-borne meteorological radiosonde. Ozone partial pressure (mPa) is measured from the reaction of ozone with potassium iodide in aqueous solution:

$$p_{O3} = 0.04307\eta_p \frac{T_p}{\Phi_p}(I_M - I_B) \tag{8}$$

where η_p is the pumping efficiency correction, T_p(K) is the air temperature of the air sampling pump, Φ_p is the airflow rate in the pump (cm^3 s^{-1}), I_M and I_B are the output and the background current in μA, respectively, the latter being the output current for ozone-free air.

[43] Precision of the measurements is given by the propagation of errors affecting the quantities in equation (8), which are a function of height. Ozone volume mixing ratio is calculated from the pressure profile measured by the accompanying radiosonde. Possible sources of systematic errors include the use of a normalization factor from an independent ozone column measurement, atmospheric pollution, the method used in treating sensor background current, air pump efficiency uncertainties, errors due to ECC sensor sensing solution loss due to evaporation, finite response time to ozone and pump temperature measurement

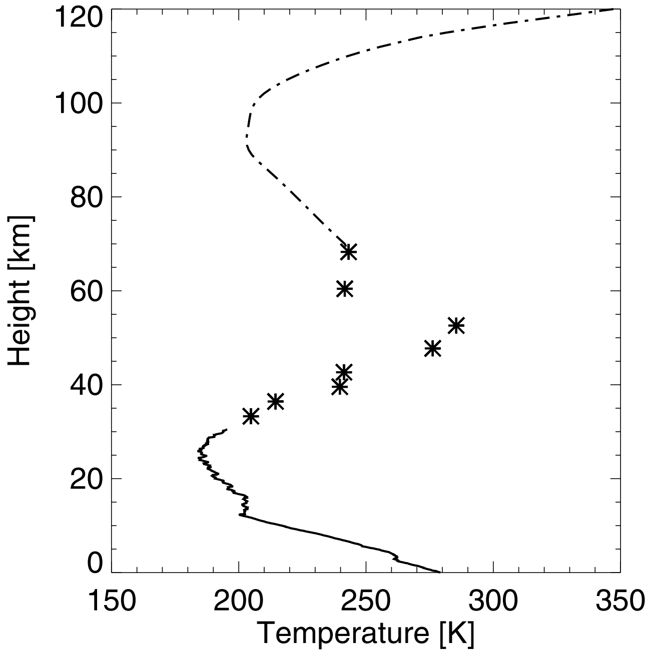

Figure 2. Temperature profile from 0 to 120 km as measured by a radiosonde in Ny-Ålesund on 7 December 2002 1200:51 UTC (solid line) and by MIPAS on 7 December 2002 1012:58 UTC from a distance of 234 km (asterisks). Temperature information in the upper part of the atmosphere (dash-dotted line) is taken from a climatology of the Arctic region in January, which was smoothly joined to the MIPAS retrieval (see text).

bias [*Komhyr et al.*, 1995] [*Harris et al.*, 1998]. Relative importance of systematic errors varies with height.

[44] In this paper we interpolate the vertical profile of the measurement accuracy $\sigma_s(z)$ from data reported in *Komhyr et al.* [1995] as ±6% near the ground, ±17% at 200 hPa, ±5% between 100 and 10 hPa, ±14% at 4 hPa. Systematic errors are spatially and/or temporally correlated, so that an appropriate correlation should be included in $\mathbf{S}_s$. Also, correlation prevents systematic errors to be averaged out by the averaging kernels. A simple first-order autoregressive correlation model was used, resulting in $\mathbf{S}_s(z_1, z_2) = \sigma_s(z_1)\,\sigma_s(z_2)\exp(-|z_1 - z_2|/H)$, where a value of H = 10 km was chosen in order to reflect the slowly varying nature in the vertical of the sources of systematic errors.

5.2. MIPAS Errors

[45] The specified performance requirements for the near-real-time retrieval scheme of MIPAS were: (i) noise error on temperature within 2 K, (ii) noise error on tangent pressure within 3%, and (iii) noise error on VMR of the target species within 5%, at all altitudes covered by the typical MIPAS scan [*Ridolfi et al.*, 2000].

[46] The main error sources that affect the accuracy of the MIPAS retrieved profiles are the noise error, due to the mapping of instrument radiometric noise in the retrieved profiles, the retrieval errors (i.e., forward model error and convergence error) and the systematic error, due to incorrect input parameters.

[47] We examined the amplitude of noise error by performing test retrievals with observations generated from assumed reference profi d perturbed with random noise of amplitude consistent with MIPAS noise specification. The noise error obtained in these test retrievals is consistent with the accuracy requirements at most altitudes covered by the standard MIPAS scan.

[48] The retrieval errors were validated by performing retrievals from spectra generated using the MIPAS Reference Forward Model [*Dudhia*, 1996] and results indicated that both forward model error and convergence error were much smaller than the measurement error.

[49] Systematic errors, such as spectroscopic errors or errors due to imperfect knowledge of the VMR profiles of nontarget species and of pressure and temperature profiles, are taken into account in the definition of the optimum size of each microwindow and in the selection of the optimal set of microwindows [*Dudhia et al.*, 2002]. All of these sources of error are taken into account in the computation of the retrieval error covariance. Preliminary results indicated that temperature error may be a significant component of the error budget.

[50] Another important source of uncertainty is due to thick clouds, which can seriously affect retrieved values below 20 km. Apparent ozone values of an order of magnitude higher than would be expected from the reference atmosphere have been retrieved in presence of clouds [*Remedios and Spang*, 2003].

6. Results

[51] The level 2 MIPAS data used in this paper do not include any cloud flags so, given the extent of the effects of

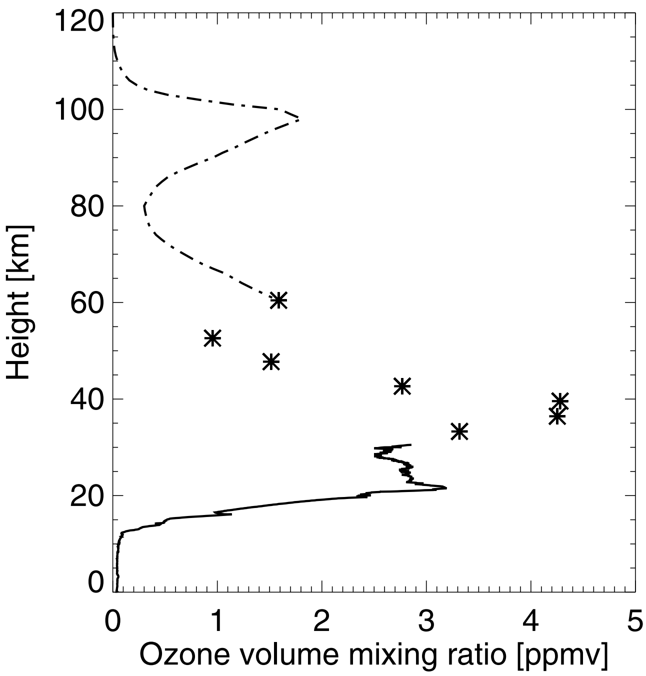

Figure 3. Ozone profile from 0 to 120 km as measured by an ozonesonde in Ny-Ålesund on 7 December 2002 1200:51 UTC (solid line) and by MIPAS on 7 December 2002 1012:58 UTC from a distance of 234 km (asterisks). Ozone information in the upper part of the atmosphere (dash-dotted line) is taken from a climatology of the Arctic region in January, which was smoothly joined to the MIPAS retrieval (see text).

clouds on trace gas retrievals, we have applied the cloud detection technique described in [*Remedios and Spang*, 2003] to the Level 1 data.

[52] Clouds are detected when the ratio between the integrated radiance in two mesowindows is less than a threshold value chosen so as to minimize effects on MIPAS retrievals of trace gases.

[53] Unfortunately it was not possible to have access to all the Level 1 files needed, so we could check for cloud contamination for only 31 MIPAS ozone retrievals out of the 41 coincidences reported in Table 1. Eleven retrievals were discarded because of the presence of clouds.

[54] Sample comparison for three cases out of the case study as in Table 1 are shown in Figures 4a, 4b, and 4c. The sonde profile is indicated by dots, and the MIPAS retrieval by open diamonds. Asterisks indicate the simulated retrieval. Statistics are computed only for that part of the profile that is well below the top sonde height, that is, by at least half MIPAS nominal vertical resolution and only these points are shown in Figures 4a, 4b, and 4c.

[55] Figures showing comparisons for the whole data set can be found at http://www.met.reading.ac.uk/~stefano/research/mipas_o3_valid/index.html, or obtained from the authors.

[56] Figure 5 shows the differences between the MIPAS retrievals and the simulated retrievals for the whole data set (without the cloud-affected cases), each with an error bar. Also shown is an estimate of the bias, given by a straight line fit to the mean difference, with $\pm 1\sigma$ error bounds. This indicates that the bias is zero within the error bounds of the fit.

[57] The χ^2 statistic for each comparison, and its distribution, are useful indications of the quality of the error analysis for the two data sources, and allow us to select dubious cases for further examination. As the top altitude of the sondes varies from profile to profile, the number of degrees of freedom, that is, the number of levels taking part in each comparison, also varies. Therefore we have scaled each χ^2 by χ_f^2, corresponding to a significance level f for the appropriate number of degrees of freedom. Thus the ratio χ^2/χ_f^2 should exceed unity with probability f.

[58] In Figures 6a and 6b are shown histograms of the ratio χ^2/χ_f^2 for $f = 0.05$ and $f = 0.01$, respectively, for the whole case study. In our ensemble, eight comparisons exceed $\chi_{0.05}^2$ (26.7%) and four exceed $\chi_{0.01}^2$ (13.3%). This means that the fraction of profiles with the ratio greater than unity for both thresholds is too large. In order to understand the possible reasons why, we examined each of the eight in detail.

[59] First, it is worth noting that the outlier with $\chi^2/\chi_{0.05}^2$ of about 4 relates to a coincidence at Thule on 14 January 2003 when no MIPAS Level 1 file is available for applying the cloud-screening algorithm. Also, the sonde ozone profile has a feature above 80 hPa, where the volume mixing ratio decreases by 1 ppmv in about 2.5 km and then recovers. This could be either an indication of problems in the sonde measurement or of the possibility that MIPAS and the sonde are measuring different air masses. The shape of the feature is reminiscent of the kind of depletion seen in ozone hole profiles, so we investigated the latter possibility by examin-

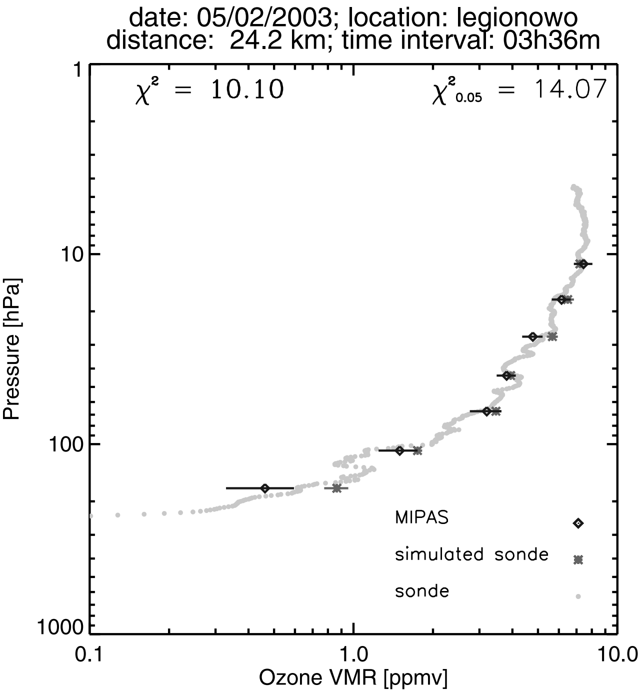

Figure 4b. Same as Figure 4a but with data acquired at Legionowo on 5 February 2003 at 24.2 km distance and 3 hours 36 min time interval. Value of χ^2 with respect to the 5% threshold $\chi_{0.05}{}^2$ for 7 degrees of freedom shows that MIPAS retrievals in the comparison are validated.

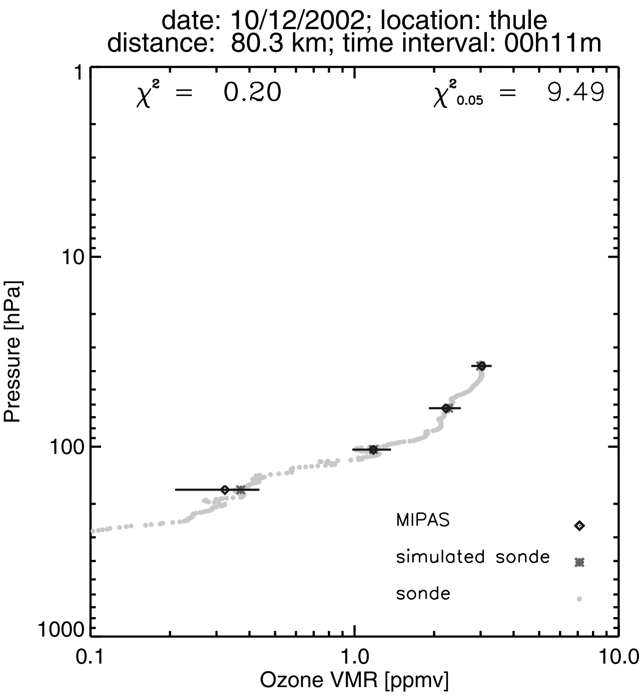

Figure 4a. Comparison between MIPAS ozone retrievals (open diamonds) and MIPAS ozone simulated retrievals (asterisks) using sonde ozone profiles (dots). Data were acquired at Thule on 10 December 2003 at 80.3 km distance and 0 hours 11 min time interval. Value of χ^2 with respect to the 5% threshold $\chi_{0.05}^2$ for 4 degrees of freedom shows that MIPAS retrievals i comparison are validated.

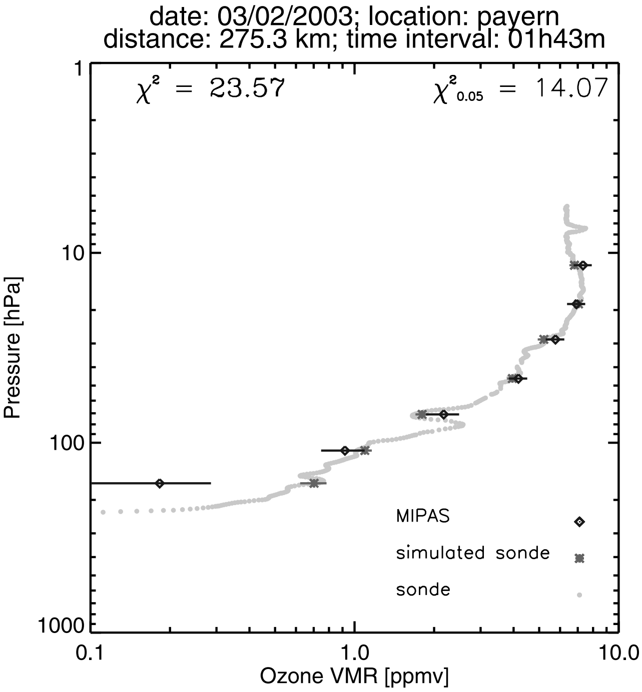

Figure 4c. Same as Figure 4a but with data acquired at Payerne on 3 February 2003 at 275.3 km distance and 1 hour 43 min time interval. Value of χ^2 with respect to the 5% threshold ${\chi_{0.05}}^2$ for 7 degrees of freedom shows that MIPAS retrievals in the comparison are not validated.

ing operational potential vorticity maps from the Met Office analysis for 14 January 2003 at 1200 UTC, obtained from the British Atmospheric Data Centre. At 68.13 hPa the locations of the two measurements seem to be at the edge of a streamer of low PV, that is, in a region of high potential vorticity gradient. This is only tentative, being the resolution of the analysis (3.75 deg longitude × 2.5 deg latitude) coarser than the spatial separation of the two measurements (79.6 km). For the reasons outlined, our outlier can be removed from the comparison ensemble.

[60] Three comparisons – in Sodankÿla on 17 December 2002, in Scoresbysund on 17 January 2003 and in Payerne on 3 February 2003 – exceed their threshold because of a discrepancy on a single level, which is the lowermost (two cases out of three) or the second lowermost. This could be due to inadequately characterized aerosol (or other absorber), or undetected cloud. For the comparison in Payerne we do not have cloud information from satellite, but visual observation showed thick clouds during all day (R. Stübi (MeteoSwiss), private communication, 2004), suggesting that the mismatch is due to undetected clouds.

[61] To check whether some bad comparisons are due to measurements of different air masses we compared qualitatively MIPAS temperature retrievals with sonde temperature profiles. In two cases – in Scoresbysund on 6 December 2002 and on 7 February 2003 – the temperature difference exceeds the MIPAS 1σ temperature random plus systematic error. To confirm that, we also looked to the Met Office temperature analyses for the same days, at a pressure where the MIPAS temperature retrieval and the sonde temperature profile differ most, na t 146.80 hPa on 6 December

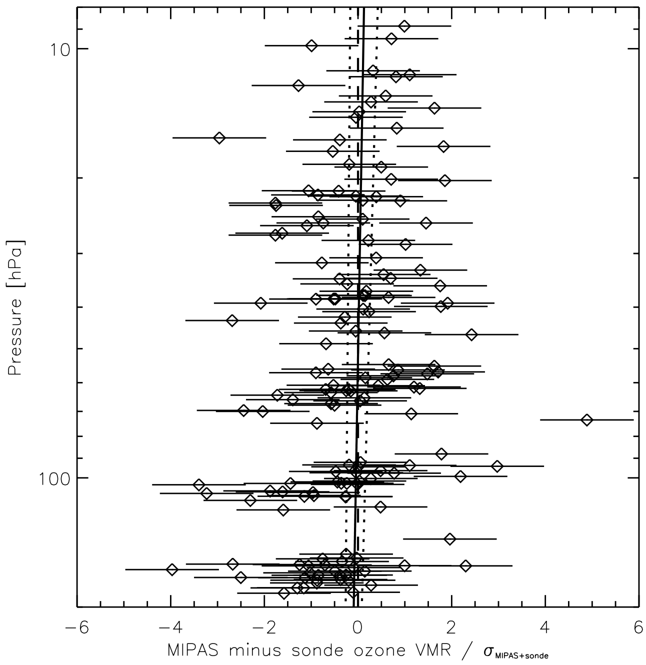

Figure 5. Differences between the MIPAS retrievals and the simulated retrievals for the whole data set, normalized with the standard deviations calculated from the diagonal of the comparison covariance matrix as in equation (6). The solid line is a fit of the normalized differences as a function of logarithmic pressure, while dotted lines highlight the 1σ uncertainty region of the fit, and the dashed line is a reference for zero bias. The plot shows that the distribution of the differences is unbiased within the error bounds of the fit.

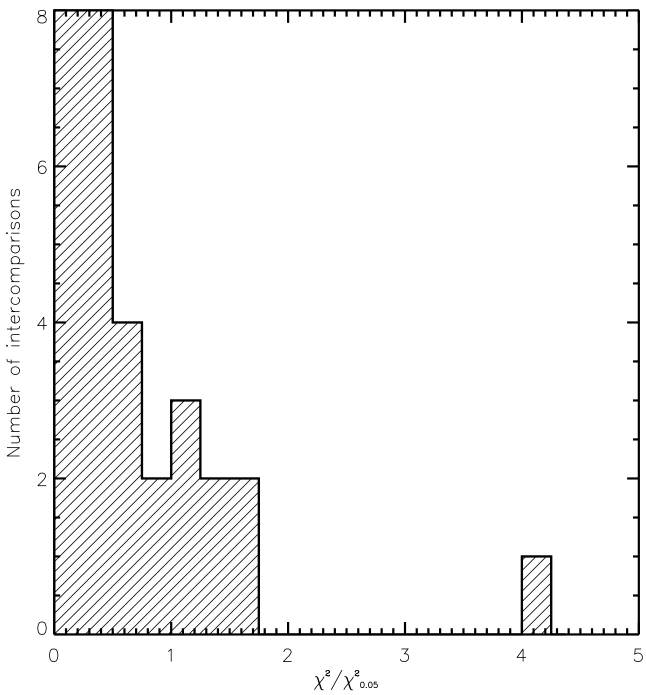

Figure 6a. Histogram of the ratio χ^2/χ_f^2 for $f = 0.05$ (5% significance level) for the whole case study.

2002 and 46.42 hPa on 7 February 2003. In both cases, we found that the location of the sonde is in a region of high-temperature gradient.

[62] The overall relevance of the spatial error component, which arises when the satellite measurement is not representative of the same portion of the atmosphere as the sonde, can also be investigated by comparing results using more stringent coincidence criteria.

[63] In Table 2, the root mean square difference between MIPAS retrievals and simulated retrievals, and the average scaled χ^2 are shown for the whole data set and different coincidence criteria. It appears that no definite improvement is achieved when closer coincidences are considered.

[64] The remaining two out of eight bad comparisons are in Thule on 23 January 2003 and Sodankÿla on 16 December 2002. The first one exceeds the 5% threshold for a very small amount ($\chi^2/\chi^2_{0.05}$ = 1.03) and is well within the 1% threshold ($\chi^2/\chi^2_{0.01}$ = 0.73). The second one is characterized by a MIPAS ozone profile that oscillates about the sonde profile, denoting possible problems in converging to the retrieval.

7. Summary and Conclusions

[65] In this paper we have described a method for comparing direct with indirect measurements, which is applicable to optimal estimators as well as maximum likelihood solutions.

[66] The presented technique – more expensive than the one described by *Connor et al.* [1994] – is best suited for avoiding inconsistencies due to different vertical resolution and influence of the a priori (when present explicitly), when the averaging kernel is not appropriate for a specific retrieval.

[67] Great care was taken in the error analysis of the comparison, by including both random and systematic components. Also, the importance of representativeness error both in the vertical and in the horizontal was investigated.

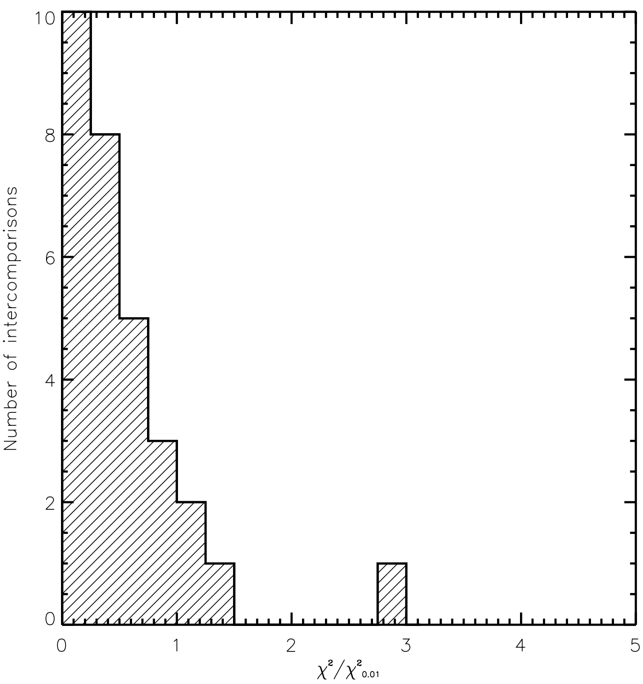

Figure 6b. Histogram of the ratio χ^2/χ_f^2 for f = 0.01 (1% significance level) e whole case study.

Table 2. Spatial Statistics

Coincidence Criteria	Number of Coincidences	RMS Difference $\delta\hat{\mathbf{z}}$, ppmv	Average $\chi^2/\chi_{0.05}{}^2$
400 km, 6 hours	30	0.387	0.720
300 km, 6 hours	25	0.348	0.690
200 km, 6 hours	18	0.366	0.797

[68] A validation of MIPAS ozone profile retrievals for the 2002–2003 northern hemisphere winter period was performed by evaluating the bias over an ensemble of 30 comparisons and by testing whether each retrieval belonged to a Gaussian distribution with covariance given by equation (6).

[69] Results can be summarized as follows:

[70] • No significant bias between MIPAS ozone retrievals and sonde ozone profiles was found.

[71] • The distributions of scaled χ^2 for both 5% and 1% confidence level threshold have a reasonable exponential decay, but too many comparisons exceed the thresholds.

[72] • A detailed investigation of the eight out of 30 comparisons exceeding the 5% threshold was performed. It resulted that the most likely explanations for six mismatches are that MIPAS and the sondes are actually measuring different air masses or that there may be problems in the two lowermost MIPAS retrievals. The latter might be a result of undetected low-level clouds or aerosol. Another case is only marginally discrepant and the last one is characterized by an oscillatory vertical pattern of the MIPAS retrieval, which might indicate a lack of convergence of the iterative retrieval algorithm.

[73] • We conclude that above the lowest two levels, the MIPAS retrievals are consistent with the sondes, but that further work (already in progress) is required to identify and deal with cloud.

[74] **Acknowledgments.** This work was funded by the NERC Data Assimilation Research Centre. The authors want to thank the European Space Agency for providing the MIPAS data and M. Allaart (KNMI), S. Bech Andersen (DMI), E. Kyrö (FMI), Z. Litynska (IMWM), R. Stübi (MeteoSwiss), P. von der Gathen (AWI), and P. Berg and G. Braathen (NILU) for the ozonesonde measurements used in this paper.

References

Brasseur, G. P., D. A. Hauglustaine, S. Walters, P. J. Rasch, J.-F. Müller, C. Granier, and X. X. Tie (1998), MOZART, a global chemical transport model for ozone and related chemical tracers: 1. Model description, *J. Geophys. Res.*, *103*, 28,265–28,289.

Ceccherini, S., and M. Ridolfi (2002), Averaging kernels for MIPAS near-real-time level 2 retrievals, *ESA Contract 11717/95/CN-CCN5*, Eur. Space Res. and Technol. Cent., Noordwijk, Netherlands.

Connor, B. J., D. E. Siskind, J. J. Tsou, A. Parrish, and E. E. Remsberg (1994), Ground-based microwave observations of ozone in the upper stratosphere and mesosphere, *J. Geophys. Res.*, *99*, 16,757–16,770.

Dudhia, A. (1996), High level algorithm definition, *ESA Contract 11886/96/NL/GS*, Eur. Space Res. and Technol. Cent., Noordwijk, Netherlands.

Dudhia, A., V. L. Jay, and C. D. Rodgers (2002), Microwindow selection for high spectral resolution sounders, *Appl. Opt.*, *41*, 3665–3673.

European Space Agency (2000), Envisat MIPAS: An instrument for atmospheric chemistry and climate research, *ESA Rep. SP-1229*, Eur. Space Agency, Paris.

Gibson, J. K., P. Kållberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano (1997), ERA description: Re-Analysis (ERA) project report, *Ser. 1*, Eur. Cent. For Medium-Range Weather Forecasts, Reading, U. K., July.

Harris, N., R. Hudson, and C. Phillips (Eds.) (1998), Assessment of trends in the vertical distribution of ozone, *SPARC Rep. 1, WMO-Ozone Res. Monit. Proj. Rep. 43*, Stratospheric Processes and Their Role in Clim., Geneva.

Komhyr, W. D., R. A. Barnes, G. B. Brothers, J. A. Lathrop, and D. P. Opperman (1995), Electrochemical concentration cell ozonesonde performance evaluation during STOIC 1989, *J. Geophys. Res.*, *100*, 9231–9244.

Morris, G. A., J. Ziemke, J. Gleason, and M. R. Schoeberl (2000), Trajectory mapping: A tool for validation of trace gas observations, *J. Geophys. Res.*, *105*, 17,875–17,894.

Remedios, J. J. (1999), Extreme atmospheric constituent profiles for MIPAS, paper presented at European Symposium on Atmospheric Measurements From Space, Eur. Space Res. and Technol. Cent., Nordwijk, Netherlands, 20–22 Jan.

Remedios, J. J., and R. Spang (2003), MIPAS observations of clouds and their effects on level 2 trace gas products, in *Proceeding of Envisat Validation Workshop, Frascati, Italy, 9–13 December 2002*, *ESA Spec. Publ. 531*, Eur. Space Agency, Paris.

Ridolfi, M., et al. (2000), Optimised forward model and retrieval scheme for MIPAS near-real-time data processing, *Appl. Opt.*, *39*, 1323–1340.

Rodgers, C. D. (1990), Characterization and error analysis of profiles retrieved from remote sounding measurements, *J. Geophys. Res.*, *95*, 5587–5595.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Sci., River Edge, N. J.

Rodgers, C. D., and B. J. Connor (2003), Intercomparison of remote sounding instruments, *J. Geophys. Res.*, *108*(D3), 4116, doi:10.1029/2002JD002299.

Wayne, R. P. (2000), *Chemistry of Atmospheres: An Introduction to the Chemistry of the Atmospheres of Earth, the Planets, and Their Satellites*, 3rd ed., Oxford Univ. Press, New York.

S. Migliorini, Data Assimilation Research Centre, Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading RG6 6BB, UK. (stefano@met.reading.ac.uk)

C. Piccolo and C. D. Rodgers, Clarendon Laboratory, Atmospheric, Oceanic and Planetary Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK. (piccolo@atm.ox.ac.uk; rodgers@atm.ox.ac.uk)