ELSEVIER

Journal of Quantitative Spectroscopy & Radiative Transfer 109 (2008) 2244–2260

Journal of Quantitative Spectroscopy & Radiative Transfer

www.elsevier.com/locate/jqsrt

An extended intercomparison of simultaneous ground-based Fourier transform infrared spectrometer measurements at the Toronto Atmospheric Observatory

Jeffrey R. Taylor[a,*], Debra Wunch[a,b], Clive Midwinter[a], Aldona Wiacek[c], James R. Drummond[a,d], Kimberly Strong[a]

[a]*Department of Physics, University of Toronto, 60 St. George St., Toronto, ON, Canada M5S 1A7*
[b]*Division of Geological and Planetary Sciences, California Institute of Technology, 1200 E. California Blvd. Pasadena, CA 91125, USA*
[c]*Institute for Atmospheric and Climate Science, ETH Zürich, Universitätstrasse 16, CHN P16.1, CH-8092 Zürich, Switzerland*
[d]*Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, Canada B3H 3J9*

Received 14 December 2007; received in revised form 7 March 2008; accepted 11 March 2008

Abstract

Simultaneous measurements of O_3, HCl, N_2O, and CH_4 were recorded by two infrared Fourier transform spectrometers of differing resolution (0.004 and 0.02 cm^{-1}) over a period of four months in the summer of 2005. These coincident observations were made at the Toronto Atmospheric Observatory, a complementary site of the Network for the Detection of Atmospheric Composition Change, and provide one of the longest records of simultaneously recorded ground-based infrared spectra to date. Retrievals performed on the spectra utilized the SFIT2 optimal estimation algorithm with HITRAN 2004 spectroscopic parameters. The influence of instrument resolution was considered in relation to the respective averaging kernels, with the predicted influence of multiplicative bias agreeing well with the observed influence for the stratospheric species. The retrieved column amounts correlated well for the stratospheric gases ($R^2 > 0.6$) but poorer correlations were observed for the well-mixed tropospheric species that were investigated. The median column differences observed by the instruments are −1.7% and 2.7% in two different micro-windows of O_3, 2.2% for HCl, −0.36% for N_2O, and 3.7% for CH_4.

Keywords: FTIR; Intercomparison; Remote sensing; Atmospheric trace gases; Infrared spectroscopy

1. Introduction

Ground-based remote sensing of atmospheric trace gas constituents by Fourier transform infrared (FTIR) spectroscopy has been carried out under the auspices of the Network for the Detection of Atmospheric Composition Change (NDACC) (formerly known as the Network for the Detection of Stratospheric Change-NDSC), for almost 15 years [1]. The primary goals of this observing network are to characterize and monitor

*Corresponding author. Tel.: +1 416 946 7129; fax: +1 416 978 8905.
E-mail address: jeff@atmosp.physics.utoronto.ca (J.R. Taylor).

0022-4073/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jqsrt.2008.03.011

long-term changes in both stratospheric and tropospheric composition, as well as to aid in the validation of coincident ground-based and space-based measurements. In 2001, the Toronto Atmospheric Observatory was established, with its principal instrument being a high-resolution, ground-based Fourier transform spectrometer (TAO-FTS) [2]. As a result of an algorithm user intercomparison exercise, this was designated a complementary instrument of the NDACC in March, 2004. Since then, the TAO-FTS has taken part in both satellite validation activities [3–5] and scientific process studies [6].

During the NDACC complementary instrument certification process, the opportunity for a side-by-side instrument intercomparison between the TAO-FTS and another certified ground-based FTS was unavailable. In the past, validation via direct intercomparison has been a common requirement of certification although it was not feasible for this case. The goal of this study was thus to make an extended intercomparison of measurements with a lower-resolution FTS which would also be the first coincident ground-based measurements by the TAO-FTS and another instrument. This comparison took place from 26 May–2 September 2005 and provided one of the longest time series of coincident ground-based FTS observations to date. These have been used to assess the impacts of spectral resolution on the retrieval of vertical column abundances and to expand upon the work initially performed by Wunch et al. [7].

This second instrument, a moderate-resolution FTS, was acquired by the University of Toronto from Environment Canada in 2001 and refurbished in order to participate in the MANTRA (Middle Atmosphere Nitrogen TRend Assessment) 2004 balloon flight [8]. The MANTRA mission consisted of four high-altitude balloon flights, flown biennially from Vanscoy, Saskatchewan, Canada (52°01′N, 107°02′W, 511.0 m), since 1998 [9]. The primary objectives for the FTS (referred to as the U of T FTS) on the MANTRA balloon were to retrieve profiles of O_3 and HCl by measuring atmospheric absorption during solar occultation. The intercomparison of ground-based observations undertaken at the Toronto Atmospheric Observatory (TAO) allowed for the U of T FTS to be thoroughly tested in a controlled setting.

Previous intercomparisons of ground-based FTS observations have primarily focused upon the agreement of quantities retrieved with different analysis algorithms or addressed how the individual instrument performance impacts the retrieved vertical column concentrations [10–14]. Similarly, this intercomparison sought to address the question of agreement between the instruments' measurements, but because they were both located at the University of Toronto, extensive simultaneous measurements could be made over a duration of four months. These relatively large data sets allowed for a more detailed statistical comparison than previous ground-based campaigns. By employing a vertical profile retrieval technique, analytical calculations of averaging kernels could be made, allowing for a better characterization of the agreement between altitude sensitivity of the individual measurements. Consequently, the differences between the measurements were considered in terms of systematic biases, both additive and multiplicative.

The study presented here is an extension of that presented by Wunch et al. [7], which compared these two instruments with a third, lower-resolution FTS using data acquired over a two-week period. That study showed that the degraded spectral resolution had little impact on the relative differences in predominantly tropospheric gases (N_2O and CH_4), but larger differences between predominantly stratospheric gases (O_3 and HCl). It also demonstrated that it is imperative to include information about instrument line shape (ILS) characterization in order to adequately retrieve atmospheric trace gas concentrations. These results were confirmed using both simulated spectra and TAO spectra with degraded resolution. To investigate these differences further with the extended data set, this intercomparison focused on the same four gases. In addition to testing the repeatability of whether spectral resolution impacts the retrievals of all four gases, the cause of the differences between O_3 and HCl were determined by analysing differences between the instruments' total column averaging kernels.

2. Instruments

2.1. TAO-FTS

The TAO (43°40′N, 79°24′W, 174.0 m) was established in 2001 with the installation of a high-resolution, DA8-model infrared Fourier transform spectrometer (TAO-FTS) manufactured by ABB Analytical Business PRU (formerly Bomem Inc.), Québec, Canada. The optical design of the instrument consists of a vertically

oriented, linear Michelson interferometer with a maximum optical path difference (OPD) of 250 cm, providing a maximum apodized resolution of $0.004\,cm^{-1}$. This design incorporates a novel dynamic alignment that is described in detail by Wiacek et al. [2].

Infrared solar absorption spectra are regularly recorded with indium antimonide (InSb) and mercury cadmium tellurium (MCT) detectors using a potassium bromide (KBr) beamsplitter to cover the spectral range from 750 to $4400\,cm^{-1}$. All of the internal optics, including the liquid nitrogen-cooled detectors, are evacuated to approximately 0.06 Torr. The external optical components include a dedicated altitude-azimuth tracker (manufactured by AIM Controls Inc., California, USA) which actively tracks direct solar radiation throughout the day, as well as several flat mirrors and a collimating mirror used to direct the radiation into the interferometer with a full field of view of approximately 1.54 mrad (with the InSb detector). Clear sky conditions are necessary to acquire spectra, restricting the average number of observing days to approximately 90 per year over the first four years of operation.

Spectra are recorded with six different narrow-band optical interference filters, which are widely used within the NDACC InfraRed Working Group (IRWG). For the purposes of this campaign, only one of these filters was used with the InSb detector, reducing the spectral range to $2400–3100\,cm^{-1}$. To attain further gains in SNR, each interferogram consisted of four co-added, 250-cm optical path difference scans recorded in the forward direction, resulting in one spectrum being obtained over a period of approximately 20 min. Each interferogram was transformed into a spectrum using a boxcar apodization.

2.2. U of T FTS

The University of Toronto's Fourier Transform Spectrometer (U of T FTS) is an ABB Analytical DA5 instrument with an apodized resolution of $0.02\,cm^{-1}$ that records single-sided interferograms along a linear mirror path. The instrument measures with two photovoltaic detectors (InSb and MCT) simultaneously through the use of a beamsplitter. The U of T FTS has a spectral range spanning $1200–5000\,cm^{-1}$ that is constrained by the detectors, a CaF_2 beamsplitter and a germanium solar filter.

The instrument has had new electronics and software installed so that it can be used both on high-altitude balloon platforms and on the ground. The new package includes control software that allows automated measurements and access to housekeeping data (voltages, temperatures, etc.). The U of T FTS has also been fitted with a sun tracker with a small tracking range ($\pm 10^\circ$ in both zenith and azimuth). It is used on the balloon to track the sun during sunrise or sunset, where it performs fine azimuth pointing to complement the payload's main pointing system. The tracker for this intercomparison is used to align the U of T FTS with the TAO sun tracker. A more detailed description of the U of T FTS can be found in Wunch et al. [8].

For the purpose of this intercomparison campaign, only data from the MCT detector are shown. At two points during the summer of 2005, the detector alignment system for the InSb detector was upgraded (without affecting the MCT detector alignment), hence, a time series based on InSb observations would add uncertainty to the results. The spectral range of the U of T FTS MCT detector overlaps with the TAO-FTS InSb detector in all regions of interest for this study.

3. Observation strategy

The observation strategy for this intercomparison campaign was designed to determine the influence of the instruments on retrieved vertical column amounts by minimizing the impact of other factors. This was achieved by measuring simultaneously from the same location, in the same spectral range and using identical retrieval methods, *a priori* information, line parameters and model atmospheres.

To measure the same atmospheric path simultaneously with both instruments, the solar beam from the TAO sun tracker was shared by using a small pick-off mirror to deflect a portion of the light into the U of T FTS. Every effort was made to ensure that the TAO-FTS incurred a minimal loss of signal (the mean SNR decrease was less than 5% in comparison to the nominal SNR), while still providing sufficient signal for the U of T FTS. This arrangement ensured that the atmospheric conditions were identical for each instrument throughout the measurements. To further ensure simultaneity, the U of T FTS co-added interferograms to match the recording periods of the TAO instrument, which has a longer scan time. While the TAO-FTS

co-adds four interferograms to record one spectrum in approximately 20 min, the U of T FTS only requires 50 s to record an interferogram. Consequently, 15–20 scans from the U of T FTS were co-added to generate a spectrum that was coincident in time with that of the TAO-FTS. See Table 1 for more details. The effects of co-adding different numbers of interferograms over the same integration time is reflected in the SNR of the measurements as this is determined from the residuals of the spectral fits (see Section 5.1).

The instruments measured solar absorption by O_3, HCl, CH_4, and N_2O in the same spectral region. TAO-FTS measurements used in the intercomparison were carried out solely with NDACC filter 3, which spans 2400–3100 cm^{-1}. This was chosen for two reasons: first, the filter 3 region has the strongest signal, and so the slight loss from the pick-off mirror used to feed the U of T FTS was of little consequence to the SNR of the TAO-FTS observations, and second, filter 3 contains informative spectral features for the four species of interest.

By eliminating atmospheric differences between measurements, differences in line parameter characterization and differences in the retrieval methods, the bulk of any observed discrepancies could be attributed to the instruments themselves, including their variable spectral resolution and the SNR differences caused by it.

4. Analysis methods

4.1. Retrievals

The OEM is commonly employed in deriving vertical profiles of atmospheric trace gases from ground-based solar absorption spectra. Some of the basic formulae relevant to this intercomparison are discussed below, while a more rigorous derivation of the mathematical formalism can be found in Rodgers [15].

In OEM, the retrieved best estimate, $\hat{\mathbf{x}}$, can be regarded as a combination of an *a priori* estimate of the atmospheric state, $\mathbf{x_a}$, and the true atmospheric state, $\mathbf{x}$:

$$\hat{\mathbf{x}} = \mathbf{x_a} + \mathbf{A}(\mathbf{x} - \mathbf{x_a}), \tag{1}$$

where $\mathbf{A}$ is known as the averaging kernel matrix, which represents the sensitivity of the retrieved state to the true state

$$\mathbf{A} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}}. \tag{2}$$

The measurement vector, $\mathbf{y}$, which contains the intensities of the spectrum, is related to the underlying true atmospheric state by a nonlinear forward model, $\mathbf{F}$, which contains *a priori* model parameters, $\mathbf{b}$, and random spectral measurement noise, $\boldsymbol{\varepsilon}$

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\varepsilon}. \tag{3}$$

In order to determine the optimal solution to Eq. (1) with Newtonian Iteration, the forward model must be linearized about a reference state (in this case, the *a priori* state is used), to derive the weighting function matrix, $\mathbf{K}$. Computationally, the averaging kernel matrix in Eq. (2) is determined from this weighting function matrix, the *a priori* covariance matrix, $\mathbf{S}_a$, and the measurement covariance matrix, $\mathbf{S}_\varepsilon$. This ultimately yields an iterative solution of the form:

$$\hat{\mathbf{x}}_{\mathbf{i+1}} = \mathbf{x_a} + (\mathbf{S_a^{-1}} + \mathbf{K_i^T S_\varepsilon^{-1} K_i})^{-1} \mathbf{K_i^T S_\varepsilon^{-1}} [(\mathbf{y} - \mathbf{y_i}) - \mathbf{K_i}(\mathbf{x_a} - \mathbf{x_i})], \tag{4}$$

Table 1
Instrument parameters

Parameters	U of T FTS	TAO-FTS
Maximum optical path difference (cm)	50	250
Time for a single scan (s)	50	300
Number of interferograms added per spectrum	15–20	4
Spectral range (cm^{-1})	1200–5000	750–4400 F3 (2400–3100)
Measurement period	May 26–September 12, 2005	Year-round

where $\mathbf{K_i}$ and $\mathbf{y_i}$ denote $\mathbf{K}$ and $\mathbf{y}$ evaluated at the retrieved state from the previous iteration, $\mathbf{x_i}$. The vertical column amount is obtained by integrating the volume mixing ratio (VMR) profile weighted by the air mass density.

For the results shown here, the OEM was implemented semi-empirically by using the SFIT2 retrieval algorithm (v3.82beta3) [16]. With this algorithm, the volume mixing ratio profile of the trace gas of interest is retrieved on a 38-layer altitude grid designed for TAO's altitude above sea level. Interfering species are considered by determining the scaling factors which result in the best overall spectral fit when applied to every point in the *a priori* profile of each interfering gas. Previous ground-based intercomparison campaigns have relied on the SFIT1 algorithm [17], which treated all quantities in the state space as scaling factors for VMR profiles, and utilized a nonlinear least squares fitting routine as opposed to the OEM technique. Column concentrations retrieved with SFIT2 from ground-based spectra have been compared with like results from other retrieval algorithms and have been found to agree within 1% for matched retrieval constraints [13].

The nonlinear forward model of the SFIT2 algorithm also uses *FSCATM* v2.03, which relies on *a priori* knowledge of pressure and temperature to perform refractive ray tracing and calculations of the air mass distribution for a model atmosphere. The ray tracing code was originally developed by Gallery et al. [18] with recent improvements made specifically for use with SFIT2 [19]. The pressure and temperature profiles are obtained from NCEP analyses provided by the NASA Goddard Space Flight Centre automailer (URL hyperion.gsfc.nasa.gov/data_services/automailer/index.html). Spectroscopic absorption line parameters and their temperature and pressure dependencies are taken from the HITRAN 2004 spectral database [20] and are used by SFIT2 for calculating absorption features with a Voigt line shape. The *a priori* state estimates of trace gas VMR profiles were constructed from a combination of climatological estimates of HALOE v.19 satellite data [21] and mid-latitude daytime 2001 MIPAS reference profiles [22,23].

To reduce the computational burden, retrievals have traditionally been carried out in small "micro-window" intervals which contain the spectral absorption features of interest. Table 2 shows the micro-windows used for all of the retrievals in this intercomparison as well as the interfering gases fitted. The selection of the range and parameters of these micro-windows was based upon previous intercomparison campaigns [12] and on the available spectral range of the instruments. For each of the target gases, the *a priori* covariance matrices were constructed to represent 30% standard deviation in all 38 layers of the state space without any off-diagonal correlations so as to provide the most numerically stable retrievals [24]. The measurement covariance matrices were constructed using the SNRs of the spectral micro-windows and were also chosen to be uncorrelated (the SNR values were determined from the residuals of the spectral fits—see Section 5.1 for details). It should be noted that the SNR of methane is unexpectedly greater for the TAO-FTS. All of the methane fits contained systematic structure in the residuals (see [7] for details) due to poor spectroscopic parameterization [25,26]. The residuals of the U of T FTS were much larger than those of the TAO-FTS. This apparent enhancement in the SNR of the TAO-FTS over that of the U of T FTS is most likely a manifestation of the poor fitting of the U of T FTS spectra and should not be interpreted as greater measurement sensitivity.

Table 2
Spectral micro-windows used for the retrieval of each trace gas as well as interfering species that were fitted simultaneously

Target gas	Micro-window(s) (cm^{-1})	Interfering gases	Typical SNR (TAO-FTS)	Typical SNR (U of T FTS)
O_3	3039.90–3040.60	H_2O, CH_4	205	365
O_3	2775.68–2776.30	CH_4, CO_2, HCl, N_2O		
	2778.85–2779.20	CH_4, HDO, N_2O	410	915
	2781.57–2782.06	CH_4, HDO, N_2O, CO_2		
HCl	2925.75–2926.05	H_2O, CH_4, NO_2, O_3	685	760
N_2O	2481.30–2482.60	CO_2, CH_4, O_3	465	610
CH_4	2859.83–2860.21	H_2O, HCl, O_3		
	2898.32–2898.98	Same as above	420	350
	2903.60–2904.16	Same as above		

4.2. Intercomparison methodology

Instrument intercomparison and validation campaigns often encounter the problem of only having a brief period of time in which to gather coincident data, thus limiting the number of measurements available for comparison. Previous campaigns have provided statistical comparisons of such data sets, but were constrained to comparing mean values [10,11] or considering systematic biases [12,14]. As well, previous campaigns have been restricted to using column retrieval algorithms in which the data products are less influenced by the altitude sensitivity of the measurements.

The advantage of having a relatively long-term intercomparison of coincident observations by two instruments is the opportunity not only to gather a large quantity of good data, but also to develop a more rigorous statistical analysis of the differences between the data sets. Furthermore, by having two observing systems make simultaneous observations with identical viewing geometry, similar instrumentation, and the same retrieval algorithms and constraints, many of the barriers that inhibit the direct comparison of observations, such as differing error covariances [24], *a priori* states [27], or different altitude grids [28], can be neglected. By utilizing the SFIT2 profile retrieval algorithm, an analysis of the vertical sensitivity of the observations can aid in better quantifying the cause of the differences. For this reason, initial comparisons were made between the two data sets by using statistical approaches that incorporate both additive and multiplicative biases [29].

If it is assumed that two sets of multiple measurements of a variable quantity, $\mathbf{y_1}$ and $\mathbf{y_2}$, made by the two instruments are each linear representations of a true quantity, $\boldsymbol{\tau}$, then the two data sets can be expressed as

$$\begin{aligned}\mathbf{y_1} &= \alpha + \beta\boldsymbol{\tau} + \mathbf{e_1},\\ \mathbf{y_2} &= \gamma + \delta\boldsymbol{\tau} + \mathbf{e_2}.\end{aligned} \tag{5}$$

The slopes of the equations, β and δ, represent the multiplicative biases for the first and second data set, respectively. Physically, this characterizes how sensitive an instrument is to changes in the truth. The intercepts of the equations, α and γ, represent the additive biases for the first and second data sets, respectively. Physically, this accounts for any consistent bias between the observations and the truth. The vectors $\mathbf{e_1}$ and $\mathbf{e_2}$ represent the errors in $\mathbf{y_1}$ and $\mathbf{y_2}$, respectively.

By substituting for the true quantity, τ, a single equation relating the two data sets can be derived:

$$\mathbf{y_2} = \frac{\delta}{\beta}\mathbf{y_1} + \left[\gamma + \mathbf{e_2} - \frac{\delta}{\beta}(\alpha + \mathbf{e_1})\right]. \tag{6}$$

Therefore, by plotting data sets of coincident observations against each other and solving for the line of best fit, the slope will yield the ratio of the relative multiplicative biases of the two instruments, while the intercept will be a combination of the multiplicative biases, additive biases, and errors.

It is important to note that this reasoning is predicated by our initial assumption of linearity between the measurements and the true state. While this is not necessarily true, it is often necessary to assume that the problem is nearly linear, that is, the problem is linear to the accuracy of the measurements within the normal range of variation of the truth, τ [15]. Because of this, it is possible that some of the variance of the data will be due to nonlinearity. Furthermore, these data sets should be treated as heteroskedastic because there is nothing requiring that the data have a constant variance. Consequently, a weighted least squares approach outlined by York et al. [30] was used to fit the data to Eq. (6).

By substituting the retrieved values from Eq. (1) into Eq. (6), a linear relation between the retrieved data from each instrument is derived. Thus, the line of best fit of a scatter plot of the two data sets can be expressed in terms of averaging kernels, *a priori* states, and errors as

$$\hat{\mathbf{x}}_2 = (\mathbf{A_2}\mathbf{A_1^{-1}})\hat{\mathbf{x}}_1 + (\mathbf{I} - \mathbf{A_2}\mathbf{A_1^{-1}})\mathbf{x_a} - \mathbf{A_2}\mathbf{A_1^{-1}}\boldsymbol{\varepsilon_1} + \boldsymbol{\varepsilon_2}. \tag{7}$$

Therefore, by plotting the retrieved results from each instrument against each other, the slope of the best fit represents the multiplicative biases in terms of the averaging kernels. For this comparison, the focus has been placed on total columns, which involves multiplying Eq. (1) by a total column operator, $\boldsymbol{\rho}$, composed of the air density profile. As with Eq. (7), the line of best fit for the resulting total columns, $\hat{c}_1$ and $\hat{c}_2$, can be expressed in

terms of averaging kernels, *a priori* states, and errors as

$$\hat{c}_2 = (\boldsymbol{\rho}\mathbf{A_2}\mathbf{A_1^{-1}}\boldsymbol{\rho}^{-1})\hat{c}_1 + \boldsymbol{\rho}(\mathbf{I} - \mathbf{A_2}\mathbf{A_1^{-1}})\mathbf{x_a} - \boldsymbol{\rho}\mathbf{A_2}\mathbf{A_1^{-1}}\boldsymbol{\varepsilon_1} + \boldsymbol{\rho}\boldsymbol{\varepsilon_2}. \tag{8}$$

By comparing Eqs. (8) and (6), the corresponding coefficients result in the slope of the total column scatter plots being given by

$$slope = \frac{\delta}{\beta} = \boldsymbol{\rho}\mathbf{A_2}(\boldsymbol{\rho}\mathbf{A_1})^{-1}. \tag{9}$$

5. Results and discussion

5.1. Quality control and treatment of errors

The quality of the spectra was determined by examining the signals of all recorded interferograms as well as the root-mean-square (RMS) of the fit residuals in a given micro-window. As the RMS of the fit residual is related to the SNR of the spectrum (in the absence of serious systematic fitting errors), this screening technique filtered any spectra that had an excessive amount of noise. The fits were quantitatively scrutinized by setting a threshold RMS value above which all fits were discarded, thus eliminating the likelihood of using cloud contaminated data. To ensure a high standard for this comparison, the threshold value was set to be 20% greater than the median value for a given micro-window. A more detailed discussion of the micro-window fits, including sample plots of fits and residuals, can be found in Wunch et al. [7].

Because spectra from both instruments were retrieved with the same algorithm and constraints, errors arising from systematic errors in the forward model or in the forward model parameters are neglected. It should be noted, however, that it is possible that errors arising from spectroscopic parameterization could be manifested in different ways due to the difference in instrument resolution. Simulations by Wunch et al. have indicated that this effect is most likely negligible [7]. Likewise, errors associated with cloud contamination and/or day-to-day sub-visible solar variability need not be considered as both spectrometers were measuring the collimated solar beam simultaneously. Total column errors were estimated from three sources: retrieval noise, smoothing error, and state space parameter interference error [15]. The error bars on all column amounts shown in figures represent these three errors combined together in quadrature.

The retrieval noise is an estimate of how the measurement error gets mapped onto the retrieved state space. The measurement error was estimated from the RMS noise of the fit residual for a given micro-window. For computational ease, it was assumed that the noise was uncorrelated.

In order to estimate the smoothing error, an ensemble of states that adequately represents the true state of the atmosphere must be chosen. Since the *a priori* states used in the retrievals were already constructed from climatologies derived from ensembles of observations (see Section 4.1), these same states were used for estimating the smoothing error. Once again, the standard deviations of these values were assumed to be uncorrelated between atmospheric layers. Although the smoothing error is the dominant source of error for both instruments, it is necessary to note that an *a priori* covariance of 30% for CH_4 and N_2O is higher than the expected variability of these gases in the troposphere, which may create larger than realistic error estimates.

Parameter interference error arises from retrieving a number of non-VMR elements in the state space. Not only is a volume mixing ratio profile being retrieved, but other parameters that are secondary to the comparison are also retrieved (such as wavenumber calibration, baseline curvature, etc.) Invariably, the retrieved VMR state has some correlation with all of these parameters and this must be considered by utilizing the averaging kernels [24].

5.2. Tropospheric species

Measuring CH_4 columns in the infrared has previously been identified as problematic for both ground-based [25] and satellite-based measurements [26]. These problems are primarily rooted in the current lack of accurate knowledge of the spectroscopic line parameters. However, for the purposes of this comparison, these

do not contribute to any discrepancy between the retrieved columns. The retrieved CH_4 total columns have a median difference $(200 \times (\text{U of T} - \text{TAO})/(\text{U of T} + \text{TAO}) \pm \sigma)$ of $3.7 \pm 2.5\%$ with the U of T FTS observations showing consistently elevated values compared to those of the TAO-FTS (see Fig. 1). However, the scatter plot of the columns (Fig. 2) shows that the R^2 goodness of fit estimator is negative for the line of best fit. This estimate is calculated by using the standard formula: $R^2 = 1 - SSE/SST$, where SSE and SST are the sum of the squared error and total sum of squares respectively. Therefore, a negative value indicates that this fit does not adequately represent the data (it is actually worse than fitting a horizontal line), and should be discarded. This is not unexpected for this well-mixed gas as the observations sample a very small range of the methane parameter space, resulting in a plot with no discernable slope. In the comparison carried

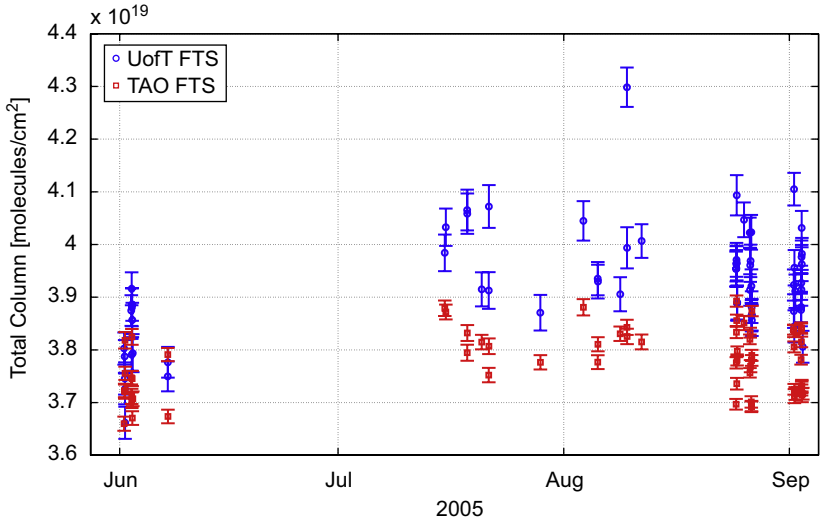

Fig. 1. Time series of coincident CH_4 total column observations.

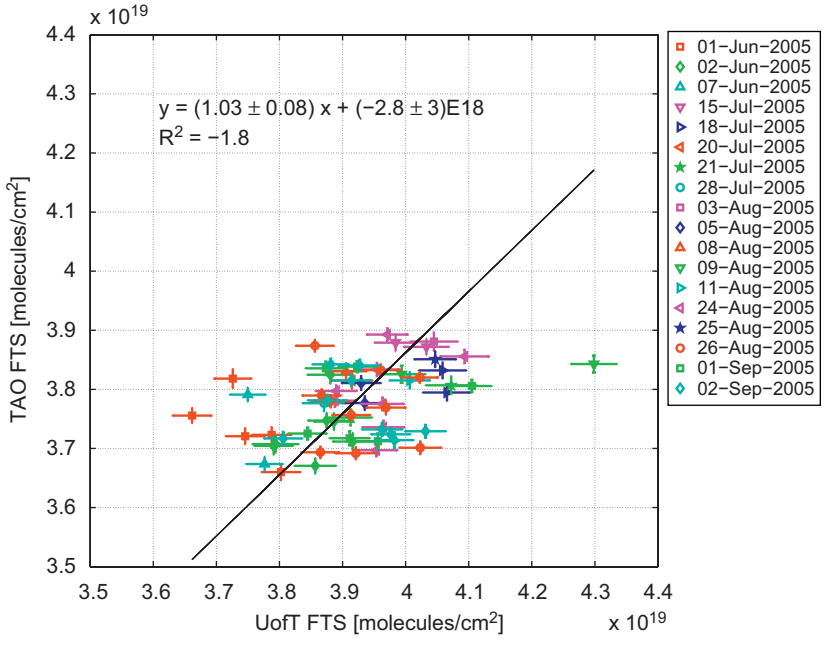

Fig. 2. Scatter plot and fit of coincident CH_4 total column observations.

out by Wunch et al. [7], the instruments' column values yielded a median difference of 2.3%. Although this extended comparison incorporates 63 paired data points (compared to only 19 in the other comparison [7]), it appears that this disagreement is enhanced. The cause of this enhanced scatter is most likely due to a known problem with the solar tracker that occurs near local noon [7]. These results are summarized in Table 3.

The N_2O observations are in better agreement, with a median total column difference of less than $-0.4\pm1.8\%$ (see Fig. 3). Unlike the CH_4 observations, there appears to be no significant additive bias between the two instruments. However, similar to CH_4, there is very little variation in the day-to-day values of N_2O, resulting in a scatter plot of concentrated data (see Fig. 4). Consequently, fitting with the linear model again results in a negative R^2 value and does not adequately represent the data. The previous intercomparison of N_2O total column values showed the same level of agreement (0.4%) [7]. This suggests that the broadened line features of N_2O are readily recorded in the lower-resolution spectra of the U of T FTS and the results compare well with those of the TAO-FTS.

5.3. Stratospheric species

As the study conducted by Wunch et al. [7] concluded that the largest disagreements were for the stratospheric gases, emphasis has been placed on understanding how well these concentrations agree. In order to compare results from micro-windows with different height sensitivities, two different O_3 retrievals were carried out in the vicinity of 3000 cm^{-1} (see Table 2). Figs. 5 and 6 show the time series of the retrieved total

Table 3
Median total column differences: 200 * (U of T − TAO)/(U of T + TAO), standard deviation of total column differences, theoretical slopes of linear trends, fitted slopes of linear trends, R^2 value of the fits, and the total number of data points for each of the four species investigated

Target gas and micro-window (cm^{-1})	Mean % Diff. (from [7])	Median % Diff.	% Std. dev.	Expected slope	Fitted slope	R^2	N
O_3-3040	0.9–4.3	2.7	3.7	0.94	0.89±0.03	0.76	75
O_3-2775	0.7–2.8	−1.7	3.7	0.79	0.78±0.03	0.77	70
HCl-2925	1.7–4.5	2.2	4.8	1.22	1.18±0.05	0.64	50
N_2O-2482	0.4–0.8	−0.36	1.8	1.02	1.27±0.1	−2.7	58
CH_4-2859	0.5–2.3	3.7	2.5	1.02	1.03±0.08	−1.8	63

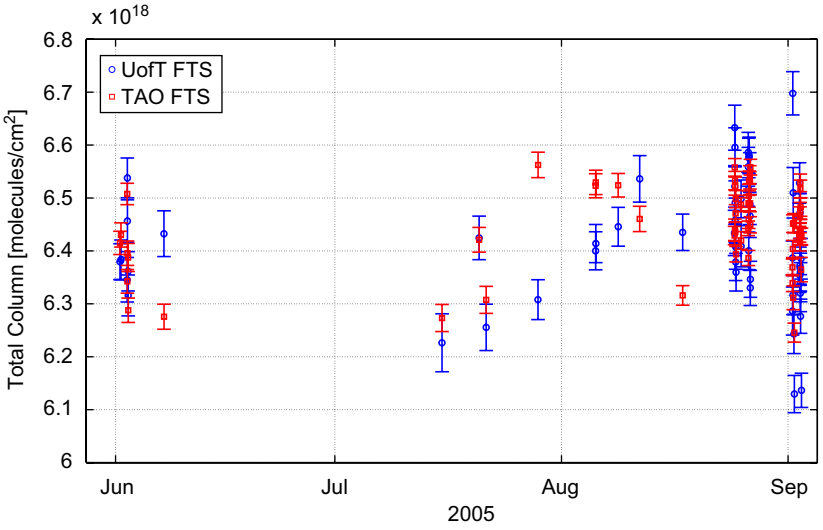

Fig. 3. Time series of coincident N_2O total column observations.

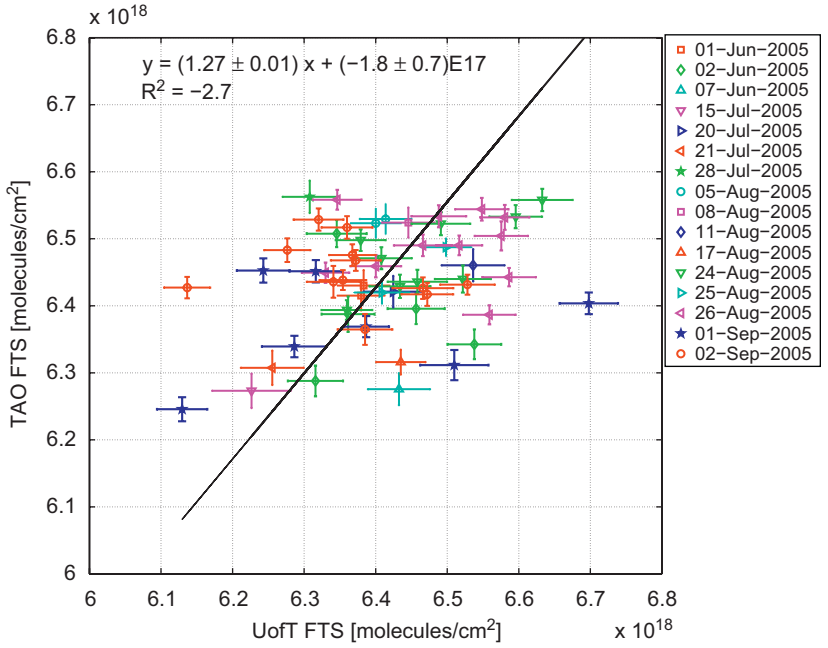

Fig. 4. Scatter plot and fit of coincident N_2O total column observations.

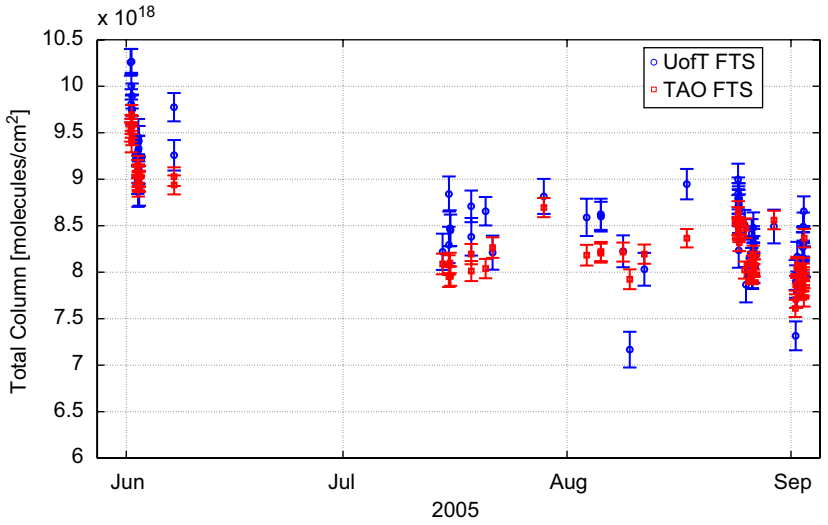

Fig. 5. Time series of coincident O_3 total column observations retrieved from the 3040 cm^{-1} micro-window.

column O_3 observed by each instrument from the $3\nu_3$ band (3040 cm^{-1}) and the $2\nu_1 + \nu_2$ band (2775 cm^{-1}) micro-windows, respectively. The overall qualitative agreement between the O_3 observations is, in fact, better for the 2775 cm^{-1} micro-window with a median difference of $-1.7 \pm 3.7\%$. The 3040 cm^{-1} micro-window columns have a median difference of $2.7 \pm 3.7\%$. These findings are consistent with those of the previous study, which found that the 2775 cm^{-1} micro-window had a difference of 0.7% and the 3040 cm^{-1} micro-window had a difference of 3.3% [7].

Analyzing the differences in the retrieved column concentrations is the traditional way in which to quantify an intercomparison of this nature [10–12,14]. However, by merely considering the average column difference, only the additive differences in the retrieved values are being investigated. A simple result from this

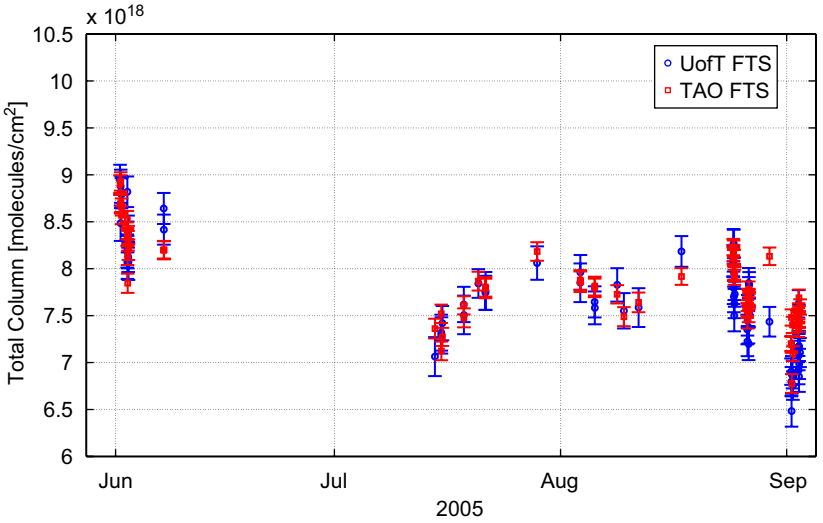

Fig. 6. Time series of coincident O_3 total column observations retrieved from the 2775 cm^{-1} micro-window.

comparison would be inconclusive as it would show that the O_3 values observed by the TAO-FTS are generally greater than those of the U of T FTS when retrieving from the 2775 cm^{-1} micro-window, while the opposite trend is found when retrieving from the 3040 cm^{-1} micro-window. Hence, it is necessary to consider other factors.

From the linear relation between the observation and the truth (Eq. (6)), it can be seen that it is important to also consider the influence of the multiplicative biases. This is of particular concern for the campaign results as the long time period over which measurements were recorded captures some seasonal variation in total column concentrations (particularly for the stratospheric gases). As average comparisons aimed at identifying additive differences will have a large variance, agreement between instruments must also be assessed in terms of multiplicative bias. By using the weighted least squares approach outlined in Section 4.2, these data can be fit and the ratio of the multiplicative biases given by Eq. (6) can be found.

Figs. 7 and 8 show the scatter plots for the retrievals from the 3040 and 2775 cm^{-1} micro-windows, respectively. The slopes of the lines of best fit are both less than 1 (0.89 and 0.78, respectively), indicating that there is a definite difference in multiplicative bias, with the U of T FTS having a larger bias.

As the averaging kernels of both of these micro-windows are well conditioned and invertible, it is possible to directly calculate the expected value of these slopes. Using Eq. (9), the calculated value of the slope for the 3040 cm^{-1} micro-window is 0.94, and the slope for the 2775 cm^{-1} micro-window is 0.79. The latter value agrees with the empirically fitted slope (0.78 ± 0.03), while the 3040 cm^{-1} calculation is slightly larger than the fitted slope (0.89 ± 0.03).

Although both the theory and the results indicate that the multiplicative bias of the U of T FTS is greater than that of the TAO FTS for O_3 observations, it is not directly obvious how this is manifested in the total column comparison. The median differences indicate that the TAO-FTS recorded generally higher concentrations from the 2775 cm^{-1} micro-window, while the U of T FTS recorded generally higher concentrations from the 3040 cm^{-1} micro-window. If one were to rely entirely on the multiplicative biases, it would be expected that the U of T FTS should report higher column amounts from both micro-windows. The reasons for this discrepancy are discussed below.

By considering the averaging kernels (see Eq. (2)), this apparent difference in retrieved total columns can be explained. Figs. 9 and 10 show the instruments' total column averaging kernels for the 3040 and 2775 cm^{-1} micro-windows, respectively, as well as the ratio of the U of T FTS total column averaging kernel to that of the TAO-FTS, for each. It is interesting to note that the U of T FTS shows a consistent increase in sensitivity below 20 km relative to the TAO-FTS, however, it is markedly different in magnitude for each micro-window.

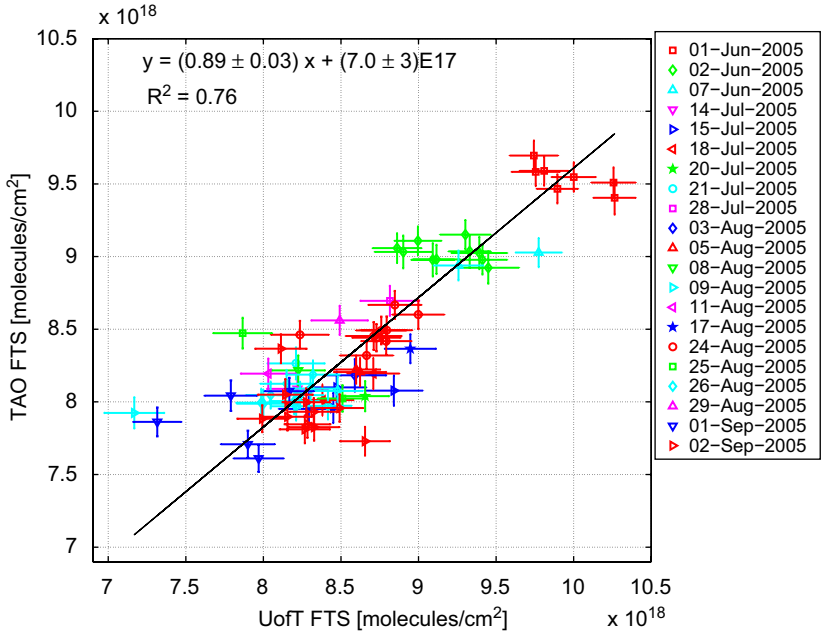

Fig. 7. Scatter plot and linear fit of coincident O_3 total column observations retrieved from the 3040 cm^{-1} micro-window.

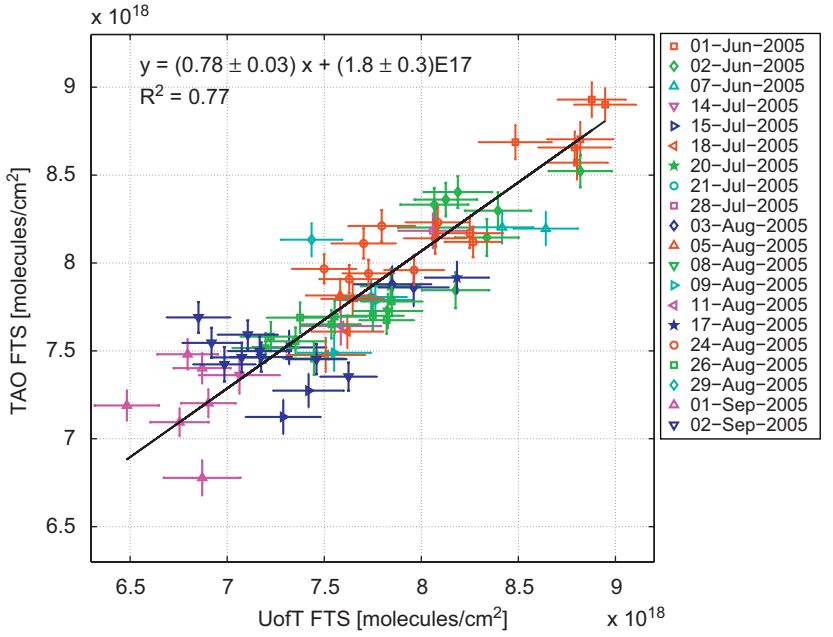

Fig. 8. Scatter plot and linear fit of coincident O_3 total column observations retrieved from the 2775 cm^{-1} micro-window.

For the 3040 cm^{-1} micro-window, the sensitivity of the U of T FTS retrieval constantly increases relative to that of the TAO-FTS retrieval from about 0.9 times that of the TAO-FTS at 20 km to about 1.5 times that of the TAO-FTS at the surface. Above 20 km, the U of T FTS shows a constant deficit of 10% relative to the TAO-FTS sensitivity. Since the median O_3 concentration in the atmosphere peaks near 25 km at the latitude of Toronto [31], it is likely that the U of T FTS is essentially underestimating the O_3 observations above 20 km

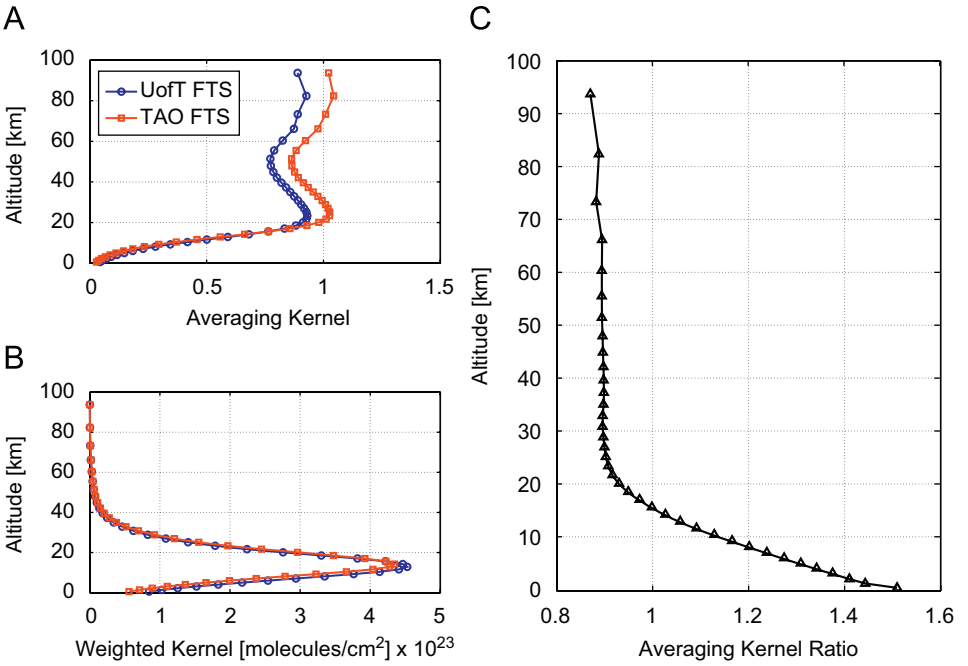

Fig. 9. O_3 averaging kernels for the 3040 cm^{-1} micro-window: (A) total column averaging kernels, (B) density weighted total column averaging kernels, and (C) ratio of the U of T FTS total column averaging kernel to that of the TAO-FTS.

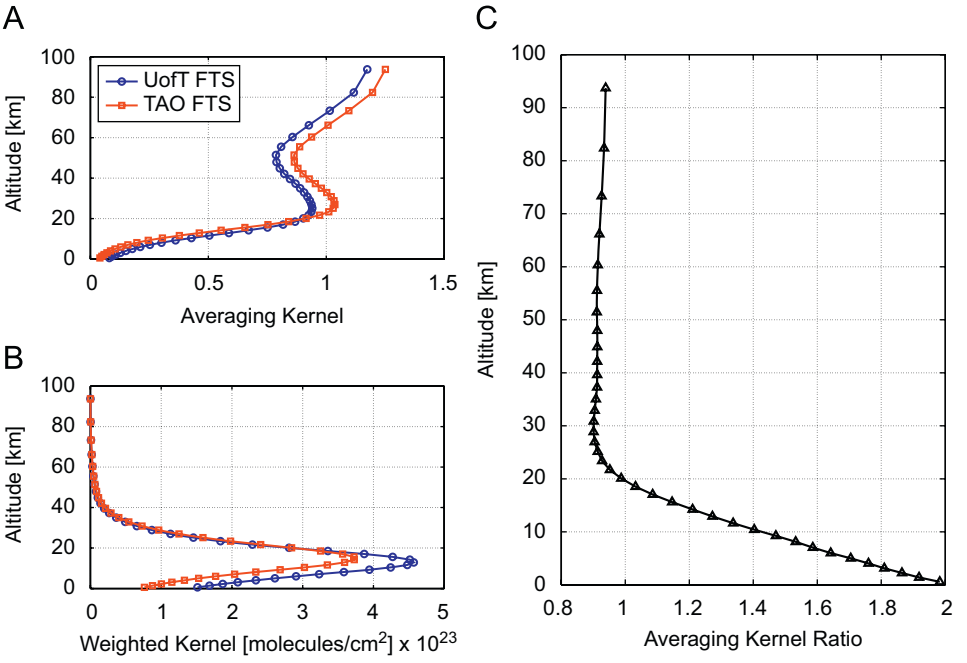

Fig. 10. O_3 averaging kernels for the 2775 cm^{-1} micro-window: (A) total column averaging kernels, (B) density weighted total column averaging kernels, and (C) ratio of the U of T FTS total column averaging kernel to that of the TAO-FTS.

compared to the TAO-FTS, and overestimating below 20 km compared to the TAO-FTS, yielding total columns that are in relatively close agreement. As the overestimation below 20 km is more pronounced, it is responsible for the derived multiplicative bias enhancement for the U of T FTS O_3 columns. This difference in

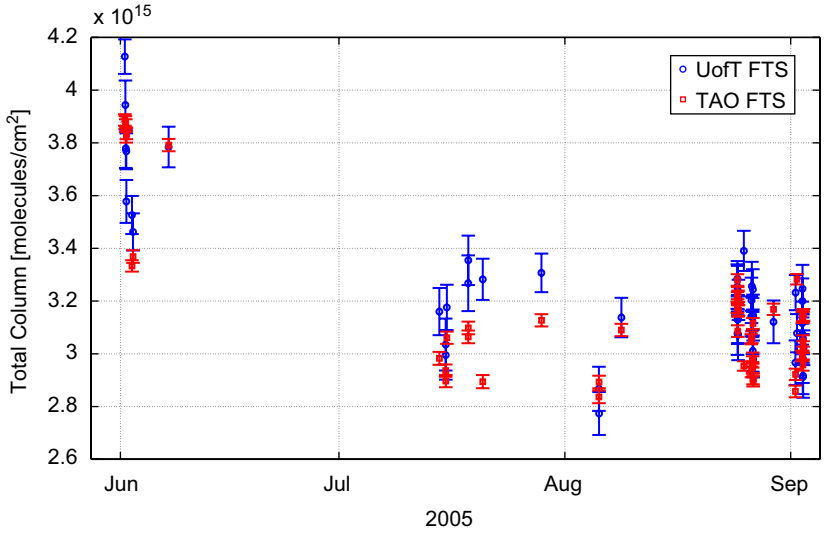

Fig. 11. Time series of coincident HCl total column observations.

altitude sensitivity is most likely a direct consequence of apparent broadening of the spectral lines due to the differences in spectral resolution of the instruments and related differences in SNR.

The total column averaging kernels from the 2775 cm^{-1} micro-window show similar behavior, although the U of T FTS averaging kernel sensitivity is more pronounced at altitudes below 20 km, attaining double the sensitivity of the TAO-FTS averaging kernel at the surface. The same decrease in sensitivity above 20 km is present as in the 3040 cm^{-1} micro-window, resulting in the TAO-FTS being approximately 10% more sensitive at higher altitudes. Although this micro-window shows a more extreme difference in lower altitude sensitivities, the total column values are still in relatively close agreement between the two instruments. This large difference at lower altitudes causes the U of T FTS to have a larger multiplicative bias value, but because there is a relatively small amount of O_3 in this portion of the total column, it does not cause the retrieved values to differ significantly. Without the use of a profile retrieval algorithm and the subsequently derived averaging kernels, the multiplicative bias ratios (from Figs. 7 and 8) would have incorrectly indicated that the U of T FTS total columns should have been more sensitive to the truth.

Following from this, the same comparison was carried out for HCl retrievals. The coincident HCl total column observations for each instrument are shown in Fig. 11. The median total column difference is approximately $+2.2 \pm 4.8\%$. The scatter plot of the columns shows that the multiplicative bias of the TAO-FTS is about 18% larger than that of the U of T FTS (see Fig. 12). This is explained by the averaging kernels shown in Fig. 13. The TAO-FTS retrieval appears to be more sensitive to HCl concentrations for all altitudes, with the largest difference being about 22% near 12 km. As with O_3, HCl is predominantly a stratospheric species and any enhanced sensitivity to concentrations below about 20 km will have a negligible impact on the total column concentration. Consequently, the total columns remain in relatively close agreement between the two instruments. The fitted slope value of 1.18 ± 0.05 agrees well with the theoretically derived value of 1.22, as computed from Eq. (9). These results are summarized in Table 3.

6. Conclusion

Simultaneous measurements of O_3, HCl, N_2O, and CH_4 were recorded by two infrared FTSs of differing resolution (0.004 and 0.02 cm^{-1}) over a period of four months in the summer of 2005. Comparisons between the retrievals of stratospheric species (O_3 and HCl) show different multiplicative biases for each instrument. O_3 retrievals in two micro-windows have shown that the lower-resolution observations by the U of T FTS are more sensitive to changes in the lower atmosphere, although this has little impact on the total column comparison. The difference in total column O_3 retrieved in each micro-window is also noticeable. On average,

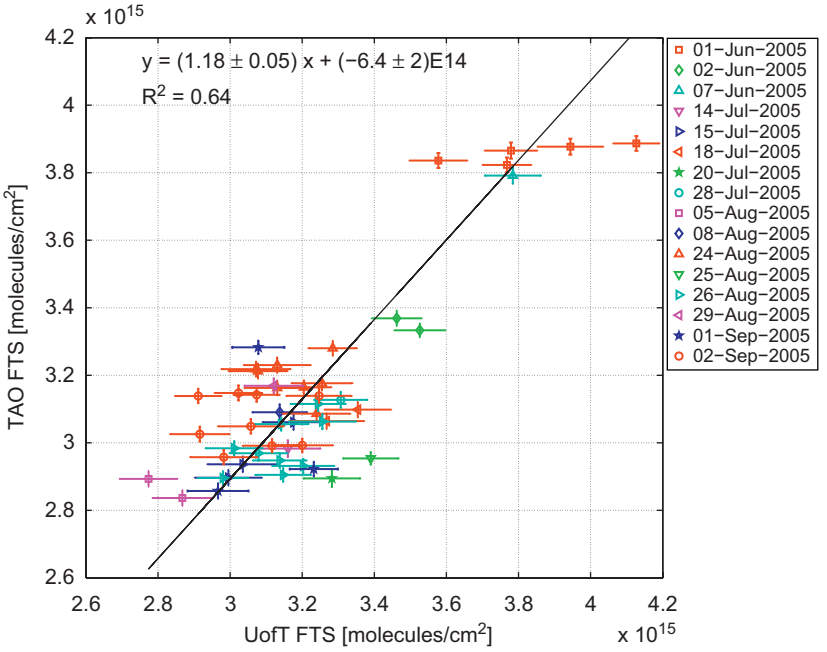

Fig. 12. Scatter plot and fit of coincident HCl total column observations.

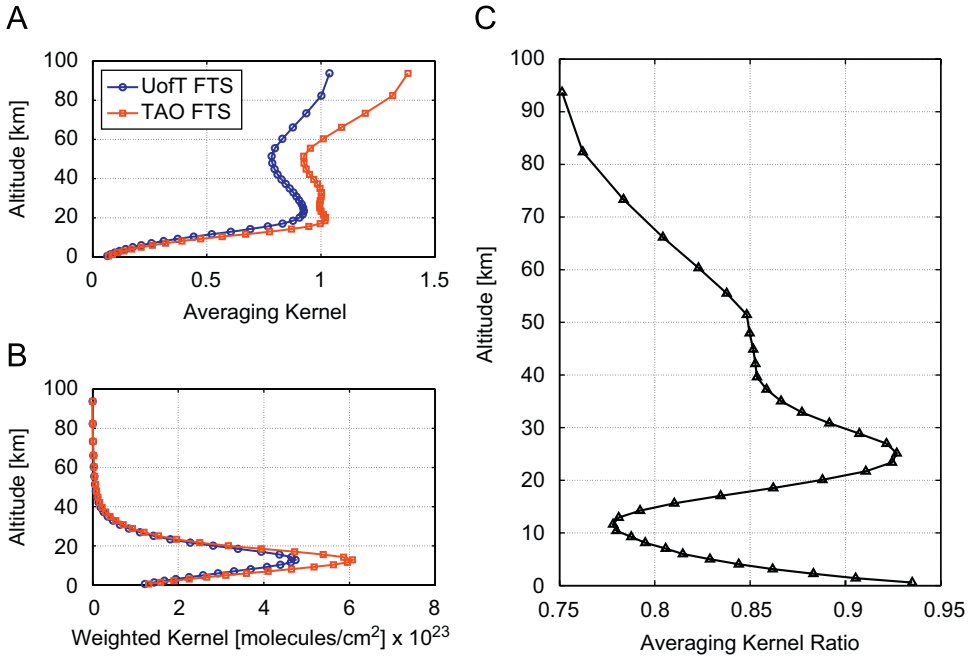

Fig. 13. HCl averaging kernels retrieved from the 2859 cm^{-1} micro-window: (A) total column averaging kernels, (B) density weighted total column averaging kernels, and (C) ratio of the U of T FTS total column averaging kernel to that of the TAO-FTS.

O_3 columns retrieved in the 3040 cm^{-1} micro-window are 5% higher than those from the 2775 cm^{-1} micro-window, for both instruments. By considering the density-weighted averaging kernels for each micro-window, it may be seen that there is enhanced sensitivity to O_3 concentrations near 20 km for both instruments in the

3040 cm^{-1} micro-window (although it is much more pronounced for the TAO-FTS). For both micro-windows, the U of T FTS has a fitted multiplicative bias of 0.89 ± 0.03 and 0.78 ± 0.03 times that of the TAO-FTS for the 3040 and 2775 cm^{-1} micro-windows, respectively. The median column differences (U of T-TAO) were $2.7 \pm 3.7\%$ and $-1.7 \pm 3.7\%$ for the 3040 and 2775 cm^{-1} micro-windows, respectively.

The TAO-FTS observations are 5% to 25% more sensitive to HCl concentrations than the U of T FTS, resulting in a total column multiplicative bias ratio of 1.18 ± 0.05. This resulting median difference in column measurements was $2.2 \pm 4.8\%$. Compared to the tropospheric retrievals, the stratospheric species show qualitatively less scatter from the linear fit and are better correlated. The difference in multiplicative bias between the two instruments arises from the lack of sensitivity to pressure broadening for HCl (as well as for O_3) by the U of T FTS, resulting in narrower spectroscopic features that are not easily captured by lower-resolution instruments. This highlights the need to be aware of instrument biases in total column measurements arising from lower-resolution infrared observations and their averaging kernels.

The tropospheric species did not allow for as detailed a comparison as the stratospheric gases. The median total column differences were $3.7 \pm 2.5\%$ and $-0.36 \pm 1.8\%$ for CH_4 and N_2O, respectively. The pressure-induced line broadening of these tropospheric species generates broader spectroscopic features that are captured by both the low and high-resolution instruments. Consequently, the averaging kernels for both CH_4 and N_2O contain information throughout the troposphere and, as a result, are non-invertible. This inhibits the calculation of multiplicative and additive biases. Unfortunately, the scatter plots for each of these species do not show representative linear relationships that can elucidate these biases further. A campaign of longer duration, albeit unpractical, would better capture the seasonal variation of the tropospheric gases and allow for better comparison of the scatter plotted data. The time series of data acquired in this study indicate that both the low-resolution and high-resolution FTS are capable of retrieving the same total columns of predominantly tropospheric species.

Acknowledgments

The authors wish to thank Environment Canada for providing in-kind instrument support and D.B.A. Jones and K. Sung for helpful discussions. We also wish to thank M. Jensen, T. Kerzenmacher, O. Mikhailov, and K. MacQuarrie for their technical support and help with measurements. Funding for this work was provided by the Natural Science and Engineering Research Council, the Canadian Space Agency, the Canadian Foundation for Climate and Atmospheric Science, ABB Bomem, the Canadian Foundation for Innovation, the Ontario Research and Development Challenge Fund, the Premier's Research Excellent Award and the University of Toronto.

References

[1] Kurylo MJ, Zander RJ. The NDSC—its status after ten years of operation. In: Proceedings of the 19th quadrennial ozone symposium, 2000.

[2] Wiacek A, Taylor JR, Strong K, Saari R, Kerzenmacher TE, Jones N, et al. Ground-based solar absorption FTIR spectroscopy: a novel optical design instrument at a new NDSC complementary station, characterization of retrievals and first results. J Atmos Oceanic Technol 2007;24(3):432–48.

[3] Taylor JR, Strong K, McLinden C, Degenstein DA, Haley CS. Comparison of OSIRIS stratospheric O_3 and NO_2 measurements with ground-based Fourier transform spectrometer measurements at the Toronto Atmospheric Observatory. Can J Phys 2007;85(11): 1301–16.

[4] Mahieu E, Zander R, Duchatelet P, Hannigan JW, Coffey MT, Mikuteit S, et al. Comparisons between ACE-FTS and ground-based measurements of stratospheric HCl and $ClONO_2$ loadings at northern latitudes. Geophys Res Lett 2005;32(L15S08).

[5] Dils B, De Mazière M, Blumenstock T, Buchwitz M, de Beek R, Demoulin P, et al. Comparisons between SCIAMACHY and ground-based FTIR data for total columns of CO, CH_4, CO_2 and N_2O. Atmos Chem Phys 2006;6(7):1953–76.

[6] Wiacek A, Jones NB, Strong K, Taylor JR, Mittermeier RL, Fast H. First detection of meso-thermospheric nitric oxide NO by ground-based FTIR solar absorption spectroscopy. Geophys Res Lett 2006;33(L03811).

[7] Wunch D, Taylor JR, Fu D, Bernath P, Drummond JR, Midwinter C, et al. Simultaneous ground-based observations of O_3, HCl, N_2O and CH_4 over Toronto, Canada by three Fourier transform spectrometers with different resolutions. Atmos Chem Phys 2007;7: 1275–92.

[8] Wunch D, Midwinter C, Drummond JR, McElroy CT, Bagès AF. The Univeristy of Toronto's balloon borne Fourier transform spectrometer. Rev Sci Instrum 2006;77(093104).
[9] Strong K, Bailak G, Barton D, Bassford MR, Blatherwick RD, Brown S, et al. MANTRA—a balloon mission to study the odd-nitrogen budget of the stratosphere. Atmosphere-Ocean 2005;43(4):289–99.
[10] Goldman A, Paton-Walsh C, Bell W, Toon GC, Blavier J-F, Sen B, et al. Network for the detection of stratospheric change fourier transform infrared intercomparison at table mountain facility. J Geophys Res 1999;104:30481–503.
[11] Griffith DWT, Jones NB, McNamara B, Paton-Walsh C, Bell W, Bernardo C. Intercomparison of NDSC ground-based solar FTIR measurements of atmospheric gases at Lauder, New Zealand. J Atmos Oceanic Technol 2003;20:1138–53.
[12] Meier A, Paton-Walsh C, Bell W, Blumenstock T, Hase F, Goldman A, et al. Evidence of reduced measurement uncertainties from an FTIR instrument intercomparison at Kiruna, Sweden. JQSRT 2005;96:75–84.
[13] Hase F, Hannigan JW, Coffey MT, Goldman A, Höpfner M, Jones NB, et al. Intercomparison of retrieval codes used for the analysis of high-resolution ground-based FTIR measurements. JQSRT 2004;87:25–52.
[14] Paton-Walsh C, Bell W, Gardiner T, Swann N, Woods P, Notholt J, et al. An uncertainty budget for ground-based Fourier transform infrared column measurements of HCl, HF, N_2O, HNO_3 deduced from results of side-by-side instrument comparisons. J Geophys Res 1997;102:8867–73.
[15] Rodgers C. Inverse methods for atmospheric sounding: theory and practice. World Scientific; 2004.
[16] Rinsland CP, Jones NB, Connor BJ, Logan JA, Goldman A, Murcray FJ, et al. Northern and southern hemisphere ground-based infrared spectroscopic measurements of tropospheric carbon monoxide and methane. J Geophys Res 1998;101:28197–217.
[17] Rinsland CP, Smith MAH, Rinsland PL, Goldman A, Brault JW, Stokes GM. Ground-based infrared spectroscopic measurements of atmospheric hydrogen cyanide. J Geophys Res 1982;87:11119–25.
[18] Gallery WO, Kneizys FX, Clough SA. Air mass computer program for atmospheric transmittance/radiance calculation: FSCATM. Environmental research paper ERP-828 = AFGL-TR-83-0065, Air Force Geophysical Laboratory, Hanscom Air Force Base, MA, USA, 1983.
[19] Meier A, Goldman A, Manning PS, Stephen TM, Rinsland CP, Jones NB, et al. Improvements to air mass calculations for ground-based infrared measurements. JQSRT 2004;83:109–13.
[20] Rothman LS, Jacquemart D, Barbe A, Benner DC, Birk M, Brown LR, et al. The HITRAN 2004 molecular spectroscopic database. JQSRT 2005;96:139–204.
[21] Russell JM, Gordley LL, Deaver LE, Thompson RE, Park JH. An overview of the halogen occultation experiment (HALOE) and preliminary results. Adv Space Res 1994;14:9–13.
[22] Carli B, Alpaslan D, Carlotti M, Castelli E, Ceccherini S, Dinelli BM, et al. First results of MIPAS/ENVISAT with operational level 2 code. Adv Space Res 2004;33:1012–9.
[23] Wiacek A. First trace gas measurements using Fourier transform infrared solar absorption spectroscopy at the University of Toronto Atmospheric Observatory. PhD thesis, University of Toronto, 2006.
[24] Rodgers C, Connor BJ. Intercomparison of remote sounding instruments. J Geophys Res 2003;108(D3).
[25] De Mazière M, Barret B, Vigouroux C, Blumenstock T, Hase F, Kramer I, et al. Ground-based FTIR measurements of O3- and climate-related gases in the free troposphere and lower stratosphere. In Quadrennial Ozone Symposium, Kos, Greece, June 2004.
[26] Worden J, Kulawik SS, Shephard MW, Clough SA, Worden H, Bowman K, et al. Predicted errors of tropospheric emission spectrometer nadir retrievals from spectral window selection. J Geophys Res 2004;109(D09308).
[27] von Clarmann T, Grabowski U. Elimination of hidden a priori information from remotely sensed profile data. Atmos Chem Phys 2007;7(2):397–408.
[28] Calisesi Y, Soebijanta VT, van Oss R. Regridding of remote soundings: formulation and application to ozone profile comparison. J Geophys Res 2005;110(D23306).
[29] Dunn G. Statistical evaluation of measurement errors: design and analysis of reliability studies. 2nd ed. Arnold; 2004.
[30] York D, Evensen N, López Martínez M, De Basab Delgado J. Unified equations for the slope, intercept, and standard errors of the best straight line. Am J Phys 2004;72(3):367–75.
[31] Brasseur G, Solomon S. Aeronomy of the middle atmosphere. 2nd ed. Dordrecht: D. Reidel Publishing Company; 1986.