Validating the reported random errors of ACE-FTS measurements

M. Toohey,[1,2] K. Strong,[1] P. F. Bernath,[3] C. D. Boone,[4] K. A. Walker,[1] A. I. Jonsson,[1] and T. G. Shepherd[1]

Received 12 March 2010; revised 22 June 2010; accepted 6 July 2010; published 19 October 2010.

[1] In order to validate the reported precision of space-based atmospheric composition measurements, validation studies often focus on measurements in the tropical stratosphere, where natural variability is weak. The scatter in tropical measurements can then be used as an upper limit on single-profile measurement precision. Here we introduce a method of quantifying the scatter of tropical measurements which aims to minimize the effects of short-term atmospheric variability while maintaining large enough sample sizes that the results can be taken as representative of the full data set. We apply this technique to measurements of O_3, HNO_3, CO, H_2O, NO, NO_2, N_2O, CH_4, CCl_2F_2, and CCl_3F produced by the Atmospheric Chemistry Experiment–Fourier Transform Spectrometer (ACE-FTS). Tropical scatter in the ACE-FTS retrievals is found to be consistent with the reported random errors (RREs) for H_2O and CO at altitudes above 20 km, validating the RREs for these measurements. Tropical scatter in measurements of NO, NO_2, CCl_2F_2, and CCl_3F is roughly consistent with the RREs as long as the effect of outliers in the data set is reduced through the use of robust statistics. The scatter in measurements of O_3, HNO_3, CH_4, and N_2O in the stratosphere, while larger than the RREs, is shown to be consistent with the variability simulated in the Canadian Middle Atmosphere Model. This result implies that, for these species, stratospheric measurement scatter is dominated by natural variability, not random error, which provides added confidence in the scientific value of single-profile measurements.

Citation: Toohey, M., K. Strong, P. F. Bernath, C. D. Boone, K. A. Walker, A. I. Jonsson, and T. G. Shepherd (2010), Validating the reported random errors of ACE-FTS measurements, *J. Geophys. Res.*, *115*, D20304, doi:10.1029/2010JD014185.

1. Introduction

[2] To date, validation of the Atmospheric Chemistry Experiment–Fourier Transform Spectrometer (ACE-FTS) data set has focused primarily on the determination of systematic bias between the ACE-FTS retrieved trace gas profiles and those from numerous other instruments. These studies have shown that, in general, ACE-FTS retrieved profiles show good agreement with other space-based missions [*Clerbaux et al.*, 2008; *De Mazière et al.*, 2008; *Dupuy et al.*, 2009; *Kerzenmacher et al.*, 2008; *Mahieu et al.*, 2008; *Strong et al.*, 2008; *Wolff et al.*, 2008].

[3] There has, however, to this point been little discussion of the quality of the reported errors of the ACE-FTS measurements. Validation of the reported errors is desirable in order to gain confidence in the scientific use and interpretation of measurements in small sample sizes, such as in the use of observed chemical data in atmospheric data assimilation systems.

[4] Validating the reported errors amounts to showing consistency between those errors and the measurement data. While other methods have been described [e.g., *Rodgers*, 2000; *von Clarmann*, 2006; *Toohey and Strong*, 2007], most studies that address the issue of random error validation do so through a comparison of reported errors and the measured short-term scatter in some latitude band, usually the tropics, where natural variability is known or assumed to be small. When random errors can be assumed to be independent of latitude, a validation of the reported random errors of tropical measurements leads to confidence in the reported random errors over the full globe. For example, *Brühl et al.* [1996] show sample standard deviation (SD) profiles of Halogen Occultation Experiment (HALOE) O_3 measurements over two-day sample sizes, for summer 1992 and 1993, at low (10°, 14°N), middle (43°, 46°N) and high (76°N) latitudes, and show that O_3 SD profiles at low and midlatitudes are comparable in magnitude with each other, and with the ~5% random errors reported for HALOE, while the high latitudes show considerably higher variance. Similarly, *Abrams et al.* [1996] quote SDs of profiles measured by the Atmospheric Trace Molecule Spectroscopy

[1]Department of Physics, University of Toronto, Toronto, Ontario, Canada.
[2]Leibniz-Institute of Marine Sciences at University of Kiel (IFM-GEOMAR), Kiel, Germany.
[3]Department of Chemistry, University of York, York, UK.
[4]Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada.

0148-0227/10/2010JD014185

(ATMOS) experiment for a number of species in tropical zonal bands (latitudes not specified), and show their general agreement with the reported random error estimates. *Livesey et al.* [2005] use SDs of measurements between 10°S and 10°N in order to assess the realism of the reported Aura Microwave Limb Sounder (MLS) precisions for a number of species. *Santee et al.* [2007] show good agreement between the SD of tropical (10°S–10°N) Aura MLS HNO_3 measurements (for single days) and reported precision estimates.

[5] The spectra measured by ACE-FTS are typically of very good quality, with signal-to-noise ratio (SNR) in excess of 300 over most of the spectral band reported for early measurements [*Bernath et al.*, 2005]. In theory, high SNR should lead to small random errors in retrieved volume mixing ratios (VMRs), and in fact, the reported errors (as currently estimated) for many ACE-FTS retrievals reach minima as low as 1%. This work aims to determine whether or not the ACE-FTS retrievals are as precise as is currently reported. In section 2, we first introduce some background theory on random error validation, and use climatological data from HALOE to motivate the use of short-term measurement variability in the tropics as a proxy for random error. In section 3 we then perform a statistical comparison of the variability of ACE-FTS tropical retrievals with their reported random error. In section 4 we investigate the short-term tropical variability in a chemistry-climate model to help understand the role that natural short-term variability plays in our random error validation method. Finally, a summary of results and conclusions is included in section 5.

2. Random Error Validation

[6] The simplest measurement model describes a measurement x_i as being the sum of the true quantity being measured, τ_i, and some zero-mean random error ε_i:

$$x_i = \tau_i + \varepsilon_i. \tag{1}$$

In a general sense, the most straightforward way to assess random error is to examine the results of repeated measurements on a single static quantity. Due to random error, the measured values will vary, and any quantification of the scatter in repeated measurements represents a quantification of the random error, or precision of the measurement. Given enough measurements, the true value τ is well estimated by the mean value of the measurements, $\bar{x}$. The differences between x_i and $\bar{x}$, i.e., the deviations about $\bar{x}$, then represent estimates of the errors ε_i. In theory, any measure of the scatter of x about $\bar{x}$ represents a measure of the scatter in ε, and so can be used to quantify the random error. In practice, the scatter in a set of n measurements is most often quantified by the SD:

$$\sigma_x = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{2}$$

with σ_x serving as an estimate of σ_ε, the scatter in the random error term ε.

[7] In terms of the simple measurement model of equation (1), assuming that the errors ε_i are uncorrelated with the truth τ_i, the variance of any measurement set (repeated or not) is equal to the sum of the variances of the truth and the measurement error:

$$\sigma_x^2 = \sigma_\tau^2 + \sigma_\varepsilon^2. \tag{3}$$

Any set of trace gas measurements will exhibit scatter, or measurement variability (σ_x^2), due to a combination of both the natural variability of the true atmospheric state (σ_τ^2) and scatter due to random measurement errors (σ_ε^2). Nevertheless, it is reasonable to expect agreement between the reported random errors and the measurement scatter (i.e., $\sigma_x \approx \sigma_0$) in regions when and where the natural variability is significantly less than the random error.

[8] Validation of random measurement errors using the method of repeated measurements described above thus depends on the identification of regions where natural variability is small compared to the measurement errors. A number of satellite validation studies, including those for the solar occultation instruments HALOE [*Brühl et al.*, 1996] and ATMOS [*Abrams et al.*, 1996], have shown good agreement between reported errors and measured variability in the tropical region, implying minimal natural variability there.

[9] A long-term examination of natural trace gas variability within zonal bands can be extracted from a climatology of trace gas measurements from the HALOE instrument [*Grooß and Russell*, 2005]. In this climatology, monthly means and SDs of trace gas VMRs of O_3, H_2O, CH_4, NO_x, HCl, and HF measured by HALOE are calculated from all data spanning the years 1991–2002. The climatology is reported for 5° bins of latitude and equivalent latitude, where equivalent latitude is a dynamical coordinate based on the contours of potential vorticity [*Butchart and Remsberg*, 1986].

[10] The statistics of this overall climatology are calculated from multiple years of HALOE measurements for each spatial bin and calendar month. The SDs of the climatology thus contain the effects of both short-term and interannual natural variability (in addition to random measurement errors). We have derived an estimate of the typical short-term (intramonthly) natural variability by averaging the SDs of the individual months of HALOE observations.

[11] Figure 1 shows equivalent latitude-pressure slices of the overall average intramonthly SD for CH_4 and O_3, i.e., a mean of all monthly SDs from the 11 year HALOE data set. The trace gases CH_4 and O_3 are chosen here as two representative but contrasting examples of stratospheric variability. CH_4 is a long-lived species with a tropospheric source and horizontal gradients typical of other long-lived species, and is often used as a tracer of dynamical transport. O_3 is also relatively long-lived in the lower stratosphere, but is short-lived in the upper stratosphere. Its horizontal gradients are significantly different from those of CH_4, due to their different sources.

[12] The minimum percent SD for any pressure surface is found (for both species) in a vertical band centered on the equator. CH_4 variability is less than 10% through a large region of the lower tropical stratosphere. The measured variability increases with height, reaching a local maximum of about 12% at approximately 1 hPa, and decreases above. For O_3, tropical variability is less than 10% through all but the lower stratosphere.

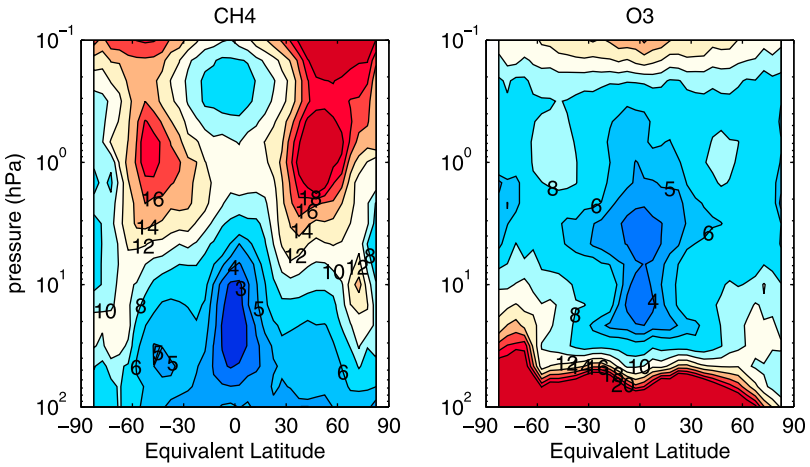

Figure 1. Average intramonthly percent SD of HALOE (left) CH_4 and (right) O_3 measurements, i.e., the mean of all monthly SDs in 5° equivalent latitude bins from the 11 year HALOE data set.

[13] An examination of the temporal evolution of variability in the HALOE climatology [*Toohey*, 2009] shows that the tropics exhibit weak variability for both CH_4 and O_3 throughout the year.

[14] As a final note regarding the short-term SD climatology from HALOE, it is notable that the minimum short-term variability reported by HALOE for O_3 and CH_4 in the tropics is comparable in magnitude to the reported random error of its measurements (~5% for both species). We can see from the preceeding analysis that natural variability is minimal in the tropics, but our knowledge of the magnitude of that natural variability is limited by the precision of the HALOE measurements. A better description of the natural variability in the tropics will be a byproduct of the efforts toward the validation of ACE-FTS RREs described in the following sections.

3. ACE-FTS Tropical Scatter Versus Reported Random Errors

3.1. Data

[15] The Atmospheric Chemistry Experiment–Fourier Transform Spectrometer (ACE-FTS) onboard the SCISAT-1 satellite, launched 12 August 2003 into a low-Earth circular orbit (altitude 650 km, inclination 74°), collects high-resolution (0.02 cm^{-1}) infrared (2.2–13.3 μm, 750–4400 cm^{-1}) spectra, measuring atmospheric extinction by solar occultation [*Bernath et al.*, 2005]. ACE-FTS performs approximately 15 sunrise and 15 sunset occultations per day, with a latitudinal coverage that depends strongly on time of year. Over a full year, the latitudinal coverage of ACE-FTS covers approximately 85°N to 85°S [*Bernath et al.*, 2005].

[16] VMR profiles as a function of altitude for pressure, temperature, and over 30 trace gases are retrieved from these spectra. The details of ACE-FTS processing are described in the work of *Boone et al.* [2005]. The altitude spacing of the FTS measurements, controlled by the scan time and the orbit of the satellite, is typically 3–4 km. ACE-FTS retrievals used here are from the version 2.2 data set, with O_3 from the v2.2 O_3 update, reported on a 1 km vertical grid. The results of a number of v2.2 validation studies are collected in a special issue of Atmospheric Chemistry and Physics (available at http://www.atmos-chem-phys.net/special_issue114.html), including articles validating measurements of the species used in this study, including O_3 [*Dupuy et al.*, 2009], N_2O [*Strong et al.*, 2008], HNO_3 [*Wolff et al.*, 2008], CH_4 [*De Mazière et al.*, 2008], and NO and NO_2 [*Kerzenmacher et al.*, 2008], and the CFC species CCl_3F and CCl_2F_2 [*Mahieu et al.*, 2008].

[17] Each individual VMR profile retrieved from an ACE-FTS limb sequence of spectra is reported along with a statistical 1σ random error. These errors are calculated from the square roots of the diagonal elements of the covariance matrix in the least squares fitting process [*Boone et al.*, 2005].

[18] In addition to random errors from the VMR retrievals themselves, there is the potential for random error contribution from an initial step of the retrieval process. Information on pressure, temperature, and measurement tangent heights are derived prior to performing VMR retrievals through fitting of CO_2 lines in each spectrum, with CO_2 VMR fixed to an assumed profile. Thus, components of random error in pressure, temperature, and tangent height information will propagate forward, and compound the random errors from the VMR retrievals. The random errors associated with the pressure/temperature retrievals are, however, small by design. CO_2 lines employed in the analysis are situated in regions of the spectrum with very high SNR (typically 300), and many CO_2 lines (more than 100 microwindows) are used to reduce the effects of spectral noise on the CO_2 lines, the primary source of the random errors. It is therefore assumed that random errors carried forward from the pressure/temperature retrievals can be neglected in comparison to the dominant source of random errors: the statistical fitting error in the VMR retrievals, henceforth referred to as the reported random error (RRE). Validation of the RREs is therefore a validation of this assumption.

[19] It is important to point out that these RREs do not include any estimate of systematic errors. The systematic errors carried forward from the pressure/temperature retrievals would not be negligible. Only random errors are considered here.

Table 1. The Number of Retrieved ACE-FTS Profiles Between the Latitudes 10°S and 10°N as a Function of Year and Month[a]

	Feb	Apr	Aug	Oct
2004	0	27	33	18
2005	13	53	55	29
2006	29	21	15	18
2007	22	21	22	15
2008	29	8	19	0

[a]The lack of measurements in February 2004 reflects the fact that the first ACE-FTS measurements were made near the end of this month. The lack of measurements in October 2008 reflects the fact that this analysis was done before retrievals for this month were completed.

[20] Since the ACE-FTS RREs are related only to the SNR of the spectral measurements, they do not in general vary systematically with latitude or season. The only exception is that due to the variation of tropopause height with latitude: at tangent heights below ~20 km, the relative RREs of species with small tropospheric VMRs can be much larger in the tropics than in the extratropics.

[21] The following analysis of ACE-FTS data will focus on the measured scatter of trace gas species in the 10°S–10°N tropical region. The orbit geometry of the ACE-FTS satellite platform allows sampling of the tropical latitudes four times per year, in February, April, August and October. The period of each tropical observation window is short: ACE-FTS samples latitudes between 10°S and 10°N during approximately eight days in each month of coverage. The sampling is split into periods of sunrise and sunset occultations, each of approximately four days length, with sunsets measured in the first half of each month, and sunrises in the second, with approximately 13 days separating the sunset and sunrise measurement periods.

[22] Due to restrictions on data downlinking and measurement frequency, the number of retrieved profiles is significantly less than what would be expected based on the orbit of the satellite platform. Table 1 lists the number of retrieved profiles produced as a function of month and year for the years 2004–2008. The number of retrieved profiles per month ranges from 8 to 55, with an average of approximately 25. The full data set of ACE-FTS tropical measurements, to be analyzed in the following, is made up of 432 retrieved profiles.

3.2. Method

[23] Two main difficulties arise in the attempt to properly quantify the scatter in tropical ACE-FTS measurements: the presence of gross outliers in the tropical data set, and the small data set resulting from the sparse coverage of measurements in the tropics.

[24] The presence of outliers in a data set poses a challenge for random error validation. On the one hand, unless data can be excluded on independent grounds, such as knowledge of anomalous conditions during the measurement, all data should be used in the precision validation, since the results of the validation should apply to the full data set. On the other hand, commonly used measures of scatter, like SD, are heavily influenced by outliers. A small number of outliers can drastically affect the SD, pulling it away from a value representative of the majority of the data.

[25] Statistics that are relatively insensitive to the presence of outliers in data are known as robust. One of the most robust estimates of scale (or scatter) is the median absolute deviation (MAD) [*Huber*, 2004]. The MAD was first promoted by *Hampel* [1974], who attributed it to Gauss. The following description is based on that given by *Maronna et al.* [2006]. Given a sample $x = (x_1, \ldots, x_n)$, the MAD is defined as:

$$\mathrm{MAD} = \mathrm{med}(|x - \mathrm{med}(x)|), \tag{4}$$

where $\mathrm{med}(x)$ denotes the sample median of x. This estimator uses the sample median twice, first to get an estimate of the center of the data in order to form the set of absolute residuals about the sample median, $(|x - \mathrm{med}(x)|)$, and then to compute the sample median of these absolute residuals.

[26] The MAD represents the interval around the median that contains 50% of the data [*Rousseeuw and Croux*, 1993]. As such, the MAD ignores the values of 50% of the data outside this interval. The "robustness" of the MAD is defined by this property: up to 50% of the data can be composed of extreme outliers, and the MAD will still give a value representative of the scatter of the central 50% of the data distribution. Throughout the following presentation of results, both SDs and MADs will be shown to help illuminate the behavior of the central portion (MAD), and the full set (SD) of the ACE-FTS measurement distributions.

[27] When examining the scatter of the tropical data set, time restriction is important in order to minimize any variance due to temporal natural variability (such as the seasonal cycle or interannual variability). On the other hand, temporal restriction must be chosen so as to result in sample sizes that lead to statistically significant results. The calculation of empirical random error estimates from the ACE-FTS data must then strike a balance between reducing natural variability and reducing sampling error: tighter temporal sampling bounds decreases the natural variability, but reduces the sample size, thus increasing the sampling error.

[28] In order to investigate the effects of reducing the temporal bounds of data taken to quantify the scatter in tropical measurements, *Toohey* [2009] partitioned the full ACE-FTS tropical data set based on a selection of temporal bounds, and calculated scatter statistics (SD and MAD) for each of the resulting subsets.

[29] The temporal partitions used were:

[30] 1. *All*: data from all months (February, April, August and October) and years (2004–2008).

[31] 2. *Months*/*All Years*: the full data set partitioned by calendar month. The partitioning thus removes the mean annual cycle of variability from the full data set, but retains any interannual variability for each month. This level of partitioning is thus equivalent to that of the HALOE SD climatology produced by *Grooß and Russell* [2005].

[32] 3. *Months:* data partitioned by year and month, thus removing both interannual and seasonal variability from the full data set. This level of partitioning is thus approximately equivalent to that produced by averaging the monthly SDs of HALOE measurements given by *Grooß and Russell* [2005], as used as a climatology of short-term variability in section 2.

[33] 4. *SR*/*SS:* data partitioned by occultation type within each month of each year. Given the tropical sampling pattern of the ACE-FTS occultations, this amounts to temporal partitions of approximately four days in length. Separating

Table 2. The Effect of Temporal Partitioning on the Sample Sizes of the Created Subsets[a]

	n_{tot}	m	n_{tot}/m
All	432	1	432
Month/all years	432	4	108
Month	432	18	24
SR/SS	378	26	14.5
Day	77	8	9.6

[a]At each level of partitioning, the total number of retrievals n_{tot} is partitioned into m subsets. n_{tot}/m then gives the average sample size of each subset. n_{tot} may change since any subsets with less than eight samples are excluded.

sunrise occultations from sunsets removes the effect of any diurnal variability.

[34] 5. *Day:* data partitioned by year, month, and finally into single days. Since sunrise and sunset occultations do not occur in the same day in the tropics, Day partitioning also removes diurnal variability. (While not true of the global ACE-FTS occultation set, the local solar time (LST, see equation (5)) of tropical occultations corresponds with expectations of typical sunrise and sunset times, with sunrises occurring between approximately 5:00 and 6:15, and sunsets between 17:00 and 18:15.)

[35] At each level of temporal partitioning, the sample sizes of each subset become progressively smaller. Table 2 details the relationship between the number of partitioned subsets (m), and the average sample size of each subset (n_{tot}/m). An arbitrary threshold of eight samples has been used as a minimum for each subset: when there are less than eight samples within a subset, it is excluded from the analysis. This sample size criterion leads to the exclusion of a handful of subsets for the SR/SS partitioning, taking the total number of measurements used from 432 to 378. A more drastic example is seen for Day partitioning, for which 82% of the retrieved profiles are excluded based on the fact that there are only 8 single days with 8 or more tropical measurements.

[36] For each level of temporal partitioning, deviations are produced by subtracting from each measurement the mean (or median) of its respective subset. For example, when partitioning by Month, deviations were produced based on the difference between each profile and its respective monthly mean. The RMS of all deviations is then used as a composite measure of the overall scatter. The composite RMS is equivalent to the square root of a weighted mean of the subset variances, with weights equal to the ratio of the sample size of each subset to the total sample size. Based on this interpretation, and for the sake of simplicity, the RMS of the full set of deviations from the subset means will be referred to in the following as a composite standard deviation, denoted SD*. Note that the SD* for the All partition is equal to the SD of the full data set. A composite MAD, or MAD*, is produced analogously by calculating the median absolute value of the deviation of each profile from its respective median. The advantage of a composite measure of scatter, such as the SD* or MAD* described above, is that it reflects the measurement variability using all of the data, and thus avoids the philosophical issue of trying to validate the RREs of a large data set based on the scatter of only an extremely small subset of measurements. The disadvantage of the composite scatter statistic is that it retains the effect of outliers in the data set, which can be removed through selection when only using small subsets of measurements.

[37] The process of partitioning the data into subsets based on temporal bounds, and calculating the SD* (and MAD*) based on deviations from the mean (median) of each subset, is performed for each of the partition bounds described above. Figure 2 shows the scatter statistics (SD* Figure 2 (top), MAD* Figure 2 (bottom)) calculated on ACE-FTS tropical measurements of O_3 at the different levels

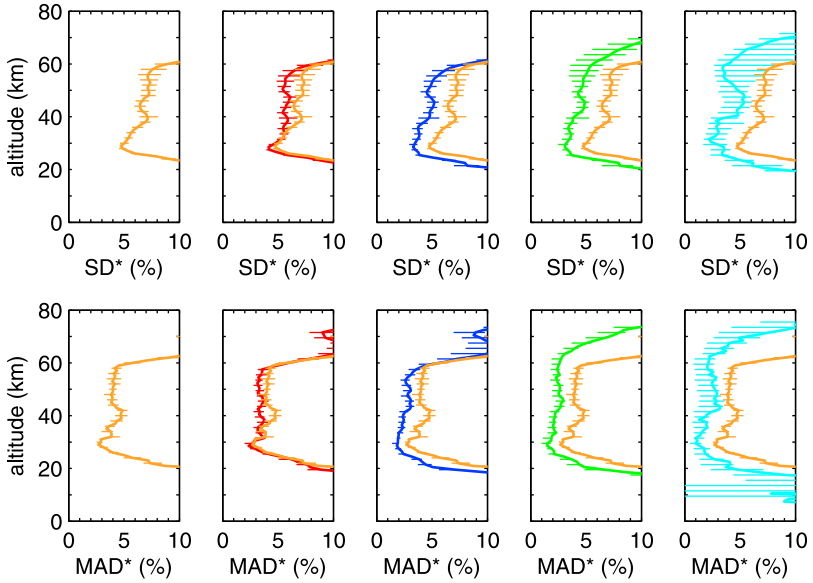

Figure 2. Tropical ACE-FTS O_3 measurement scatter quantified by the statistics (top) SD* and (bottom) MAD* as defined in the text, shown for All (orange), Month/All Years (red), Month (blue), SR/SS (green), and Day (cyan) partitioning on separate plots, with 95% confidence intervals shown by horizontal lines. The statistics for All are repeated in all plots to aid comparison.

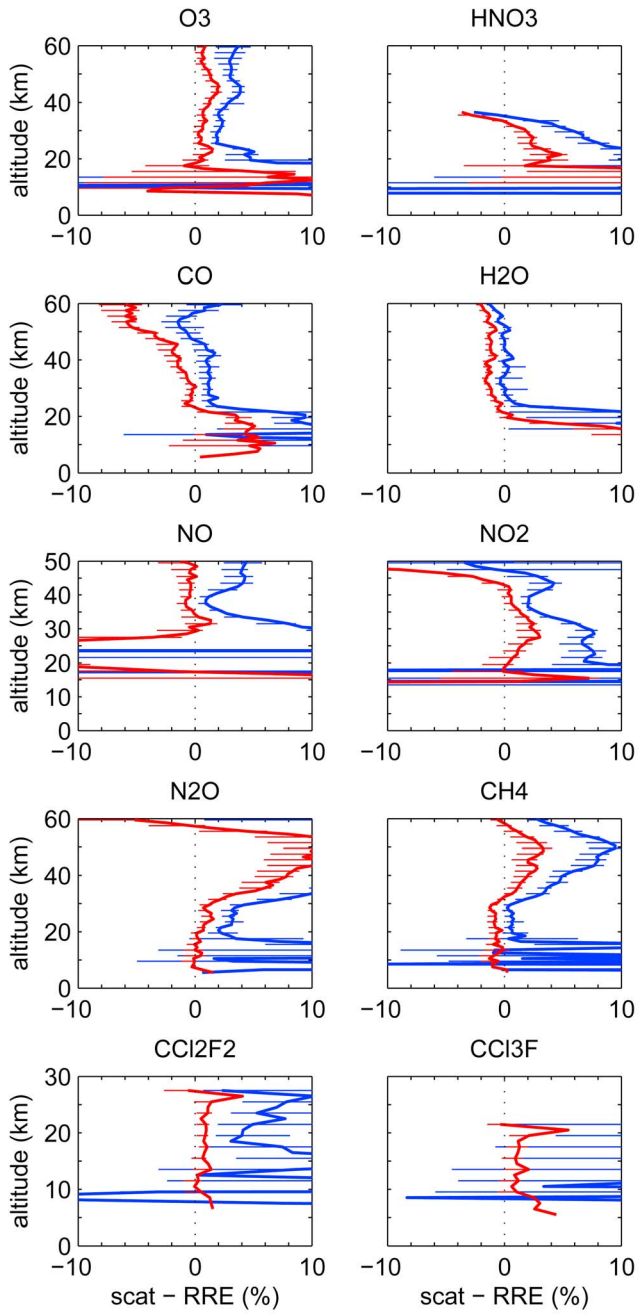

Figure 3. Differences between percent short-term scatter measured by ACE-FTS and the percent RREs. Each plot shows the SR/SS partition SD* - RMS(RRE) (blue) and MAD* - med(RRE) (red). Confidence intervals, based on the confidence intervals calculated for the scatter statistics through bootstrapping are shown as horizontal bars every 2 km.

of partitioning. To aid comparison, each panel also contains the All partition statistic. Confidence intervals at the 95% level are produced through the bootstrapping technique [*Efron and Tibshirani*, 1994], wherein the calculation used to produce a statistic is performed on a large number of random resamplings of the full data set, producing an ensemble of statistic estimates. The width of the resulting distribution of statistic estimates can then be used to define a confidence interval for the statistic. Here we produce confidence intervals based on 1000 resamplings.

[38] The scatter in deviations steadily decreases with tighter temporal partitioning, as seen in Figure 2. It should be noted that the SD* of the full O_3 data set, which includes variability on time scales of interannual, seasonal, and short term, is only a few percent larger than that based on Day partitioning, although the 95% confidence intervals imply that these differences are significant. All levels of partitioning give SD*s that are below 8% between the altitudes of roughly 25 and 55 km. Partitioning the data, and in so doing removing different scales of temporal variability, reduces the scatter by the order of a few percent. Notable reductions in scatter occur between 25 and 60 km due to the removal of seasonal and interannual variability and above 55 km due to the removal of diurnal variability. The reduction in scatter produced by partitioning by SR/SS subsets, in so doing removing the effect of diurnal variations, is displayed most clearly by the MAD* statistic. It should also be noted that the SR/SS MAD* profile is consistent with that produced by Day partitioning (with a much smaller sample size) for all but a few altitudes. The SR/SS MAD* profile is also quite close to the Month MAD* profile below 55 km. It would appear that there is only a small reduction in scatter produced by moving from Month to tighter levels of partitioning below 55 km where diurnal variations are insignificant.

[39] The confidence intervals shown for O_3 are of relatively equal size for All, Month/All Years, Month and SR/SS partitioning for altitudes between 20 and 50 km. For SD*s, the confidence intervals have widths ranging from approximately 1 to 3% between 22 and 60 km for these levels of partitioning. The Day SD* has somewhat larger confidence intervals, ranging in width from 2 to 5% between 22 and 60 km, due to the smaller total sample size resulting from the minimum subset size threshold. Confidence intervals for the MAD* are in general smaller than those for the SD*s, as would be expected given the robust nature of the MAD*, with widths ranging from 0.6 to 1% between 22 and 60 km for all but Day partitioning. At low altitudes for the Day MAD*, very small sample sizes lead to confidence intervals that have a minimum of zero.

[40] The confidence intervals show that while the Month/All Year SD* is not significantly different from the All SD*, the SD*s for the other levels of partitioning are (at the vertical levels shown).

[41] The scatter in tropical ACE-FTS measurements, quantified by SD* and MAD* statistics, has been calculated for a number of trace gas species measured by ACE-FTS. In general, the short-term partitioned statistics, SR/SS and Day, show the smallest values. The SR/SS and Day statistics are for the most part not significantly different from each other (i.e., their 95% confidence intervals overlap), although typically the SR/SS confidence intervals are significantly smaller than those for the Day statistics. Therefore, in order to compare the measured scatter with the RRE, SR/SS statistics will be used as a best measure of short-term scatter.

3.3. Results

[42] Agreement between SR/SS scatter statistics and RREs is examined in Figure 3, with each panel showing the dif-

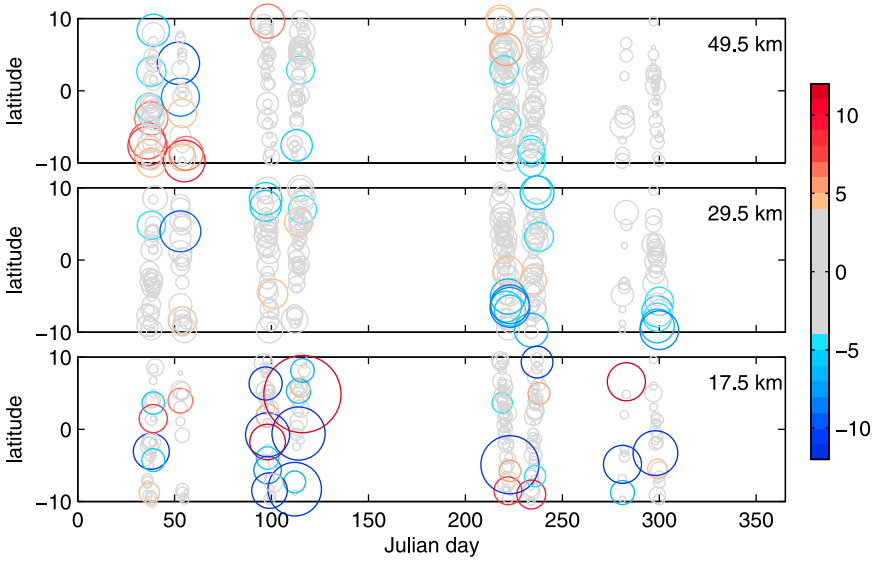

Figure 4. ACE-FTS measured CH_4 deviations from SR/SS medians, at 49.5, 29.5, and 17.5 km, shown as a function of measurement Julian day and latitude. Deviations are normalized by the MAD*. The symbol size and color are used to convey the magnitude and sign of the deviations, with red colors for positive deviations and blue for negative.

ference between scatter and RRE (i.e., SD* - RMS(RRE) and MAD* - med(RRE)) for ten sample trace gas species measured by ACE-FTS. Confidence intervals at the 95% level, based on the confidence intervals calculated for the SR/SS scatter statistics, are shown as horizontal lines.

[43] The results of the comparisons of scatter with RREs can be grouped into three cases. In "Case 1," differences between scatter and RREs are less than or approximately equal to zero (within error bars) for both the SD* and MAD* comparisons. Such is the case for H_2O at altitudes above 25 km, and for CH_4 between 20 and 30 km. For CO, differences between scatter and RREs are less than 1% above 20 km. For these species and altitude regions, this comparison amounts to a successful validation of the ACE-FTS RREs, in a manner similar to previous validation studies for other missions. The random measurement error can be no larger than the short-term measurement variability observed, thus it can be concluded that the true random measurement error for these species and regions is equal to, or less than the RREs.

[44] The second comparison result, "Case 2," describes instances where differences between the MAD* and median RRE are equivalent to or less than zero, while the SD* is significantly larger than the RMS RRE. Such is the case for O_3 between 20 and 40 km, NO between 30 and 50 km, NO_2 in two bands centered at 20 and 40 km, N_2O between 20 and 30 km, CH_4 below 20 km, and for most of the full range of measurements for the CFC species. Since the MAD* ignores the outlying 50% of the input data, agreement between the MAD*s and the median RREs shows that the width of the central 50% of the deviation distribution is in good agreement with the RREs. In other words, the RREs are representative of the scatter in 50% or more of the data. The larger differences between the SD*s and the RMS RREs are then understood to be a result of outliers in the data set, whose deviations are larger than their respective RREs.

[45] The third and final comparison result, "Case 3", describes instances where both the SD* or MAD* are significantly larger than the RREs. Such is the case for O_3 between 40 and 50 km, for HNO_3 over all but the very highest retrieved altitudes, for CO below 20 km, for NO_2 between 25 and 35 km, for both N_2O and CH_4 above 30 km, and for H_2O below 25 km.

[46] It is clear from the differences in results based on SD* and MAD* that outliers are present in the tropical ACE data set. Figure 4 shows CH_4 deviations from the SR/SS medians, as a function of Julian day and latitude for the full ACE-FTS tropical data set. These deviations are normalized by the MAD*, and the symbol size and color are used to convey the magnitude and sign of the deviation.

[47] At 17.5 km, there exist many large deviations, with absolute deviations greater than 10 MAD*. It is known that the quality of ACE-FTS retrievals below ~20 km suffers from the presence of high cirrus clouds. This is especially an issue in the tropical measurements. As a result of this a priori knowledge, an ACE-FTS data user could conceivably choose to discard ACE-FTS measurements below 20 km that showed large normalized deviations from the MAD.

[48] At 29.5 km, CH_4 deviations are predominantly negative, and located mostly in the winter/spring hemisphere. Since CH_4 mixing ratios decrease with latitude away from the tropics at this height, it is conceivable that these negative deviations are due to the intrusion of extratropical air into the 10°S–10°N latitude band due to vigorous mixing episodes in the surf zone. Instances of negative CH_4 deviations are well correlated with negative N_2O, and positive NO and NO_2 deviations (not shown), which supports this hypothesis. At 49.5 km, the picture is similar: negative deviations are clustered at the winter/spring edges of the latitude band, but there are also positive deviations at the summer edges. Inspection of the HALOE CH_4 climatology [*Grooß and Russell*, 2005] shows that above 10 hPa (~30 km) the mean CH_4 mixing ratios are sloped across the equatorial

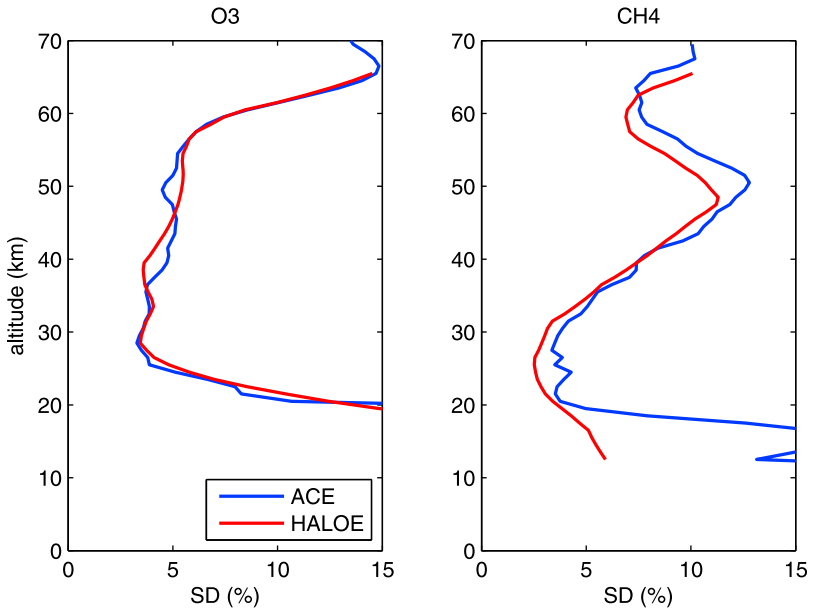

Figure 5. Variability in tropical ACE-FTS O_3 and CH_4 retrievals, compared to that from the HALOE climatology.

region, with higher VMR values in the summer/fall hemisphere. Thus, the deviations shown in Figure 4 are consistent with a strong gradient in VMR across the equatorial band, and perhaps mixing or sloshing across the 10°S and 10°N latitudes.

[49] In the light of the preceding characterization of ACE-FTS measurement scatter, it appears likely that discrepancies between scatter and RREs, in the form of either Case 2 or Case 3 described above, may be produced, at least in part, by short-term natural variability. The variability of chemical fields simulated in a chemistry-climate model will be used in the next section in order to investigate the possible extent of tropical natural variability.

[50] Before doing so, however, it is instructive to compare the tropical scatter observed by ACE-FTS to that reported in the HALOE climatology. In section 2, it was postulated that the minimum SD values of HALOE measurements in the tropics may have been representative of the random error in those measurements, and that the true natural variability may have been less. If so, and if the random errors of ACE-FTS measurements are smaller than those for HALOE (as reported), one would expect the scatter in ACE-FTS measurements to be smaller than those from HALOE. Figure 5 compares O_3 and CH_4 SD*s from ACE-FTS measurements, partitioned by Month, to the average SD profile from the HALOE time series of monthly SDs for the four months of ACE-FTS tropical coverage. There is excellent agreement between the scatter statistics for both instruments. This close agreement supports the idea that natural variability, not random measurement error, is the dominant source of the scatter in both measurement sets.

4. Tropical Variability in the CMAM

[51] In this section, the variability of tropical chemical species simulated by the Canadian Middle Atmosphere Model (CMAM) is compared to the scatter measured by ACE-FTS, and the reported random errors of the ACE-FTS measurements. By making the assumption that the CMAM accurately reproduces, in a statistical sense, the natural variability of the true atmosphere, the variability of CMAM chemical fields can be used as a lower bound on the variability of any measurement set, itself subject also to random measurement error. We focus specifically on species (O_3, HNO_3, N_2O, and CH_4) that showed significant discrepancies between measured scatter and reported random errors (i.e., those classified as "Case 3" in section 3) at altitudes above 20 km, and which are long-lived (since the temporal resolution of model output makes model-measurement comparisons of rapidly varying species difficult). Modeled and measured scatter will be compared at the level of SR/SS partitioning, since this level of partitioning is the shortest time span for which the confidence intervals of the measurement scatter remain small.

[52] The CMAM is an extended version of the Canadian Centre for Climate Modelling and Analysis spectral General Circulation Model (GCM). The dynamical core and chemistry scheme are described by *Beagley et al.* [1997] and *de Grandpré et al.* [1997], respectively. The distributions of chemical species in the CMAM have been seen to generally compare well with observations [e.g., *de Grandpré et al.*, 2000; *Farahani et al.*, 2007; *Hegglin and Shepherd*, 2007; *Jin et al.*, 2005; *Jin et al.*, 2009; *Melo et al.*, 2008]. Simulated chemical fields from the last ten years (1995–2004) of the CMAM REF1 simulation described by *Eyring et al.* [2006] are used here. The specifications of the simulations follow or are similar to the "reference simulation 1" (REF1) of CCMVal [*Eyring et al.*, 2005] and include anthropogenic and natural forcings based on changes in sea surface temperatures, trace gases, and aerosol effects from major volcanic eruptions. While this version of CMAM does not simulate the quasi-biennial oscillation and thus underestimates interannual variability in the tropics, the intramonth variability appears to be of realistic magnitude [see *SPARC*

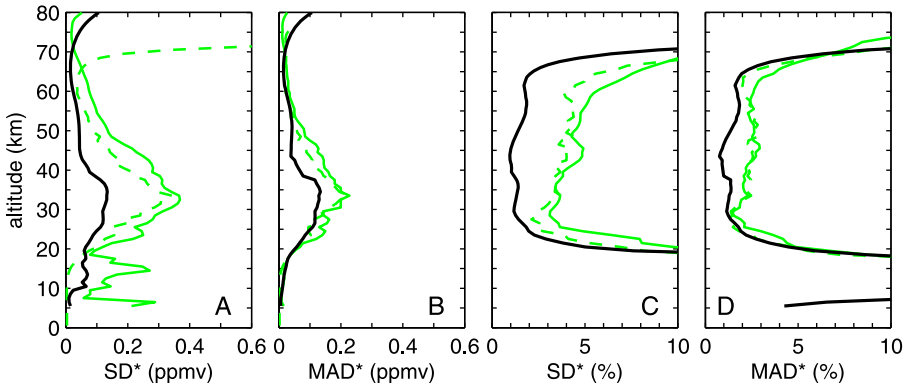

Figure 6. Tropical variability in CMAM O_3 (green dashed lines), compared to the scatter measured by ACE-FTS (green solid), both for SR/SS partitioning, and the ACE-FTS RREs (black) for (a) absolute SD*, (b) absolute MAD*, (c) percent SD*, and (d) percent MAD*.

CCMVal, 2010, Chapter 7]. The chemical fields are available for every model grid point, in intervals of 18 h, with this high sampling frequency (i.e., compared to monthly means) allowing the calculation of short-term variability.

[53] Quantification of atmospheric variability depends on resolution: the larger the air volumes associated with each sampling point are, the smaller the variability will be (and vice versa). When comparing variability between different sources (observations and/or models) it is therefore important that the resolutions of the compared fields are similar. ACE-FTS observations have a horizontal resolution of ~500 km, while CMAM has a horizontal resolution of roughly 3.75°, or 400 km. Tropical ACE-FTS retrievals have a vertical resolution of ~3 km, while the CMAM vertical resolution in the tropopause region is around 900 m, coarsening to around 2 km in the upper stratosphere. Since stratospheric variability is dominated by processes (mixing and transport) that occur on quasi-horizontal surfaces, we expect that the similar horizontal resolution of ACE-FTS and CMAM justifies a direct comparison between the two. In order to facilitate the comparisons, CMAM chemical fields have been interpolated from pressure coordinates onto the ACE-FTS 1 km altitude grid using the model geopotential field.

[54] In order to best reproduce the sampling of the tropical ACE-FTS measurements, the CMAM data is sampled based on the local solar time (LST), defined as:

$$\mathrm{LST} = \mathrm{UTC} + 24\left(\frac{\lambda}{360^\circ}\right), \tag{5}$$

where UTC represents the universal (Greenwich Mean) time, and λ the longitude. In the tropics, the LSTs of ACE-FTS occultations are generally consistent, with sunrises occurring between approximately 5:00 and 6:15, and sunsets between 17:00 and 18:15. In the following analysis, only CMAM fields with LSTs within the bounds corresponding to ACE-FTS sunrises and sunsets are used. Furthermore, only the months February, April, August, and October, corresponding to the months of ACE-FTS tropical coverage, are used. Lastly, CMAM SR/SS partitions are constrained to a length of four days, in order to match the length of ACE-FTS SR/SS partitions.

[55] Scatter statistics for tropical O_3 modeled by CMAM and measured by ACE-FTS are shown in Figure 6, along with the RMS RRE of the ACE-FTS measurements. Both the CMAM O_3 variability and the ACE-FTS measurement scatter are generally larger than the ACE-FTS RMS RRE. The CMAM variability is comparable to the ACE-FTS measured scatter in vertical shape, with a maximum in absolute variability (Figure 6a) at approximately 35 km, and percent variability (Figure 6c) that increases relatively linearly between 30 and 60 km. Below 20 km, CMAM O_3 variability decreases to negligible amounts, due to the lack of any tropospheric chemistry in the model. The ACE-FTS SD* here is significantly larger than the RMS RRE, due to outliers resulting from clouds. The MAD* below 20 km is, however, consistent with the RREs, suggesting that the precision of retrievals not affected by clouds is well described by the RREs. There is very good agreement between the CMAM and ACE-FTS MAD*s throughout the stratosphere, especially in terms of percent. The large increase in CMAM variability above 65 km compared to ACE-FTS variability is due to differences in sampling. Whereas ACE-FTS samples the diurnally varying O_3 at exactly sunrise and sunset, i.e., at solar zenith angles (SZAs) of 90°, the CMAM has been sampled based on the window of local solar times determined by the ACE-FTS tropical data set. Therefore, in terms of SZA, the CMAM O_3 values are sampled from a finite window which includes 90°, and since O_3 VMRs change very rapidly as a function of SZA in the mesosphere [*Allen et al.*, 1984], the CMAM fields display a larger variance.

[56] Comparisons of scatter statistics for modeled and measured HNO_3 are shown in Figure 7 (top). The CMAM variability between 20 and 35 km agrees well with the scatter measured by ACE-FTS, especially in terms of the MAD* statistic. In terms of percent, the CMAM variability is larger than the ACE-FTS scatter: this is a result of a low bias in CMAM HNO_3. Below 20 km, where CMAM variability is small, the ACE-FTS statistics are dominated by random error. Once again, the effect of clouds is apparent: the difference between the ACE-FTS SD* and the RMS RRE is relatively large, while the median RRE is within the 95% confidence intervals (not shown) of the absolute MAD*.

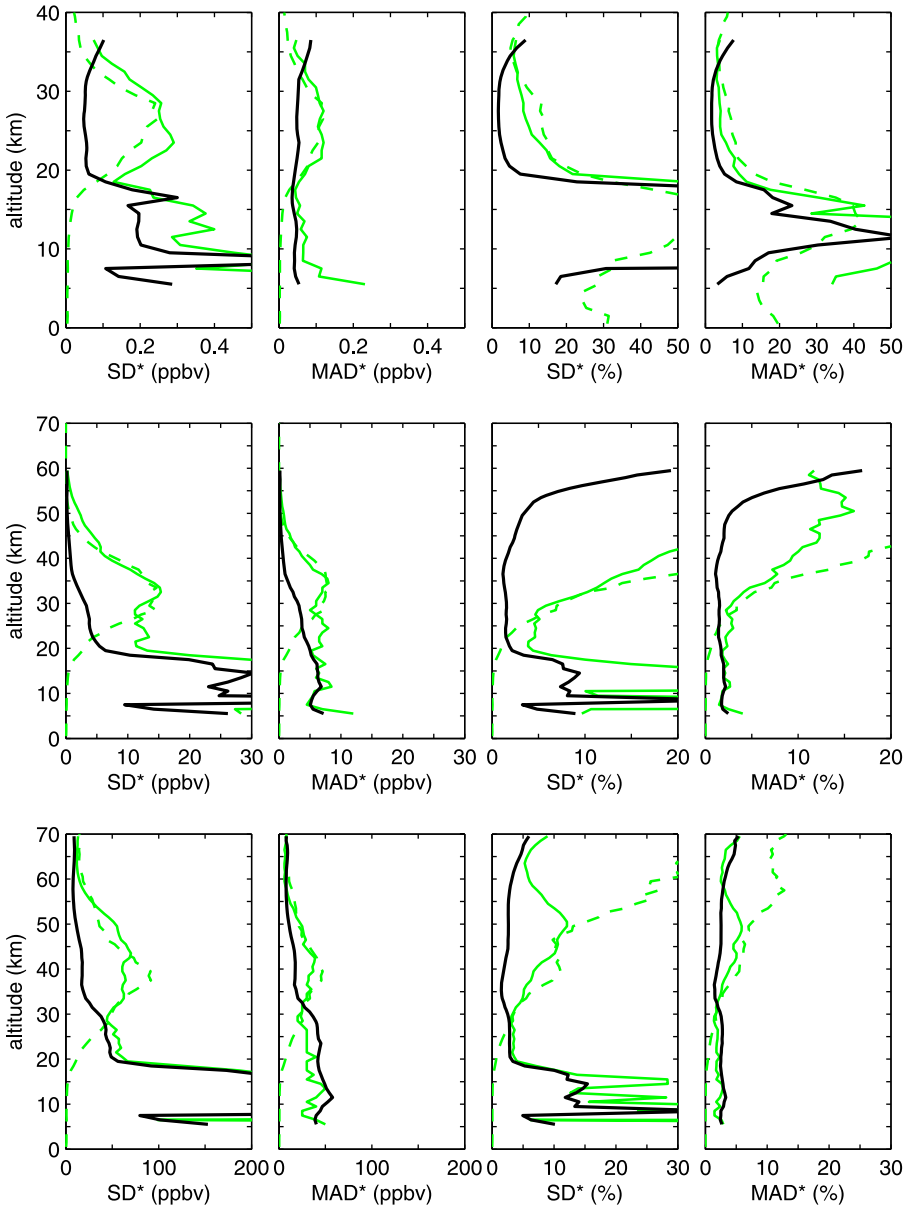

Figure 7. As in Figure 6 for (top) HNO_3, (middle) N_2O, and (bottom) CH_4.

[57] Comparisons of scatter statistics for modeled and measured N_2O are shown in Figure 7 (middle). The variability of the CMAM simulated N_2O agrees well with the scatter in ACE-FTS measurements between 25 and 50 km, especially for the absolute scatter statistic comparison. The relative scatter statistic comparison is affected by a slight low bias in the model mean fields at heights above 40 km. Below 20 km, where the CMAM variability decreases to zero, the ACE-FTS MAD* show excellent agreement with the RREs, and the differences between SD* and RMS RRE are often smaller than the 95% confidence intervals (see Figure 3). The results for the CH_4 comparison (Figure 7, bottom) are similar.

[58] In summary, the variability in tropical chemical fields simulated by the CMAM is in good agreement with the measured scatter in tropical stratospheric (~25–60 km) ACE-FTS measurements. Based on the CMAM variability results, it would appear likely that differences between the short-term scatter in ACE-FTS measurements and the RREs, at altitudes above ~25 km, shown in Figure 3 are due to the presence of real short-term natural variability in the tropical stratosphere.

[59] For contrast, we turn now to examine the variability of CMAM H_2O, as comparisons for H_2O showed consistency between ACE-FTS scatter and RRE. CMAM variability, ACE-FTS scatter, and the ACE-FTS RRE for H_2O are shown in Figure 8. In terms of the SD* statistic, the CMAM variability is comparable to (20–40 km), or smaller than (above 40 km) the ACE RMS RRE. The ACE-FTS SD* is seen to agree well with the RMS RRE, except below 25 km. The ACE-FTS MAD*, on the other hand, is significantly smaller than the median RRE, and is in closer agreement with the CMAM variability than with the RRE. In other words, while the scatter of the full set of ACE-FTS

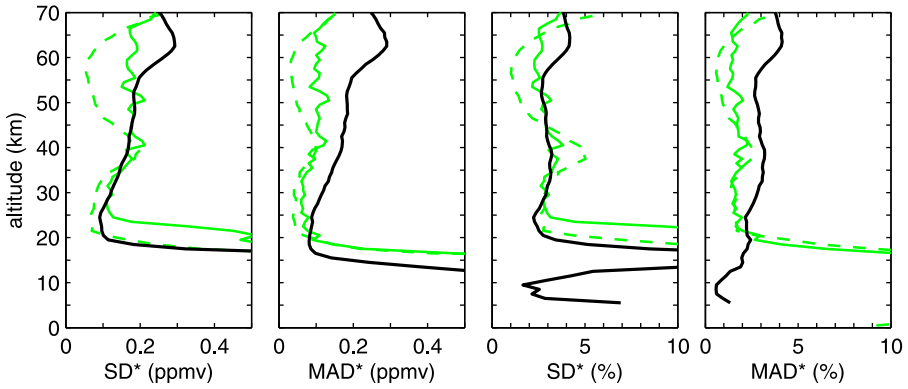

Figure 8. As in Figure 6 for H_2O.

measurements (i.e., the SD*) agrees well with the RREs, the scatter of the central portion of the measurements (i.e., the best 50%, quantified by the MAD*) is significantly less than the RREs for this portion of data. It would appear that the RREs of the "well behaved" retrievals may actually be overestimated.

5. Conclusions

[60] This work has focused on validating the reported random errors (RREs) of the ACE-FTS trace gas measurements through a comparison of RREs with the variability (or scatter) of measurements in the tropical latitude band 10°S–10°N, where natural variability is minimal. The scatter in tropical ACE-FTS measurements has been quantified based on a series of successively tighter temporal bounds. The scatter present in short-term (4-day) subsets of the data is compared directly with the ACE-FTS reported random errors.

[61] The measured scatter in CO and H_2O is found to be consistent with the RREs in terms of the standard deviation (SD), and to be smaller than the RREs in terms of the median absolute deviation (MAD). For these two species, this represents a validation of the random measurement errors in the manner of previous satellite validation studies, in the sense that the RREs are not underestimated. However, the results suggest that if outliers are excluded through the use of the MAD, then the CO and H_2O measurements may be more precise than their RREs would indicate. This hypothesis is supported in the case of H_2O by the fact that the measured scatter of H_2O is consistent with the variability simulated by the CMAM, implying that the measured scatter is dominated by real natural variability.

[62] Tropical scatter in measurements of NO, NO_2, CCl_2F_2 and CCl_3F is roughly consistent with the RREs as long as the effect of outliers in the data set is reduced through the use of the robust scatter statistic MAD rather than the SD. This result encourages the use of caution in the scientific use of the measurements in small sample sizes, but also suggests that the majority of the data is consistent with the RREs. The MAD, or other robust statistics may be quite useful in identifying outliers in this data set, and thereby screening the data.

[63] The scatter in measurements of O_3, HNO_3, CH_4 and N_2O, while significantly larger than the RREs in some regions of the stratosphere, is shown to be consistent with the variability simulated in the CMAM. This result strongly supports the hypothesis that the scatter in the stratospheric ACE-FTS measurements of these species is due to real natural variability, not random error. The implications of this result can be taken in two ways. Firstly, if the true random errors of the measurements are smaller than the minimum variability of the atmosphere, then the RREs cannot be validated through the method used here. On the other hand, this result should increase confidence in the utility of the measurements, since it implies that differences between measurements are dominated by differences in the true atmospheric state.

[64] **Acknowledgments.** The authors would like to thank the Canadian Space Agency (CSA), the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS). The Atmospheric Chemistry Experiment (ACE), also known as SCISAT-1, is a Canadian-led mission mainly supported by the CSA and NSERC. The CMAM is supported by the CFCAS and the CSA. This work was also carried out with the aid of a grant from the CSA. The authors would also like to thank Dylan B. A. Jones, Michelle L. Santee, and three anonymous reviewers for their helpful comments on this work.

References

Abrams, M. C., et al. (1996), On the assessment and uncertainty of atmospheric trace gas burden measurements with high resolution infrared solar occultation spectra from space by the ATMOS experiment, *Geophys. Res. Lett.*, *23*(17), 2337–2340.

Allen, M., J. I. Lunine, and Y. L. Yung (1984), The vertical distribution of ozone in the mesosphere and lower thermosphere, *J. Geophys. Res.*, *89*, 4841–4872, doi:10.1029/JD089iD03p04841.

Beagley, S. R., J. de Grandpré, J. N. Koshyk, N. A. McFarlane, and T. G. Shepherd (1997), Radiative-dynamical climatology of the first-generation Canadian Middle Atmosphere Model, *Atmos. Ocean*, *35*(3), 293–331.

Bernath, P. F., et al. (2005), Atmospheric Chemistry Experiment (ACE): Mission overview, *Geophys. Res. Lett.*, *32*, L15S01, doi:10.1029/2005GL022386.

Boone, C. D., R. Nassar, K. A. Walker, Y. Rochon, S. D. McLeod, C. P. Rinsland, and P. F. Bernath (2005), Retrievals for the atmospheric chemistry experiment Fourier-transform spectrometer, *Appl. Opt.*, *44*(33), 7218–7231.

Brühl, C., et al. (1996), Halogen Occultation Experiment ozone channel validation, *J. Geophys. Res.*, *101*, 10,217–10,240, doi:10.1029/95JD02031.

Butchart, N., and E. E. Remsberg (1986), The area of the stratospheric polar vortex as a diagnostic for tracer transport on an isentropic surface, *J. Atmos. Sci.*, *43*, 1319–1339, doi:10.1175/1520-0469(1986)043.

Clerbaux, C., et al. (2008), CO measurements from the ACE-FTS satellite instrument: Data analysis and validation using ground-based, airborne and spaceborne observations, *Atmos. Chem. Phys.*, *8*(9), 2569–2594.

de Grandpré, J., J. W. Sandilands, J. C. McConnell, S. R. Beagley, P. C. Croteau, and M. Y. Danilin (1997), Canadian Middle Atmosphere Model: Preliminary results from the chemical transport module, *Atmos. Ocean*, *35*(4), 385–431.

de Grandpré, J., S. R. Beagley, V. I. Fomichev, E. Griffioen, J. C. McConnell, A. S. Medvedev, and T. G. Shepherd (2000), Ozone climatology using interactive chemistry: Results from the Canadian Middle Atmosphere Model, *J. Geophys. Res.*, *105*, 26,475–26,492, doi:10.1029/2000JD900427.

De Mazière, M., et al. (2008), Validation of ACE-FTS v2.2 methane profiles from the upper troposphere to the lower mesosphere, *Atmos. Chem. Phys.*, *8*(9), 2421–2435.

Dupuy, E., et al. (2009), Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), *Atmos. Chem. Phys.*, *9*(2), 287–343.

Efron, B., and R. J. Tibshirani (1994), *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Eyring, V., D. E. Kinnison, and T. G. Shepherd (2005), Overview of planned coupled chemistry-climate simulations to support upcoming ozone and climate assessments, *SPARC Newsl.*, *25*, 11–17.

Eyring, V., et al. (2006), Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past, *J. Geophys. Res.*, *111*, D22308, doi:10.1029/2006JD007327.

Farahani, E., et al. (2007), Nitric acid measurements at Eureka obtained in winter 2001–2002 using solar and lunar Fourier transform infrared absorption spectroscopy: Comparisons with observations at Thule and Kiruna and with results from three-dimensional models, *J. Geophys. Res.*, *112*, D01305, doi:10.1029/2006JD007096.

Grooß, J.-U., and J. M. Russell III (2005), Technical note: A stratospheric climatology for O_3, H_2O, CH_4, NO_x, HCl and HF derived from HALOE measurements, *Atmos. Chem. Phys.*, *5*(10), 2797–2807.

Hampel, F. R. (1974), The influence curve and its role in robust estimation, *J. Am. Stat. Assoc.*, *69*(346), 383–393.

Hegglin, M. I., and T. G. Shepherd (2007), O_3-N_2O correlations from the Atmospheric Chemistry Experiment: Revisiting a diagnostic of transport and chemistry in the stratosphere, *J. Geophys. Res.*, *112*(D11), D19301, doi:10.1029/2006JD008281.

Huber, P. (2004), *Robust Statistics*, Wiley, New York.

Jin, J. J., et al. (2005), Co-located ACE-FTS and Odin/SMR stratospheric-mesospheric CO 2004 measurements and comparison with a GCM, *Geophys. Res. Lett.*, *32*, L15S03, doi:10.1029/2005GL022433.

Jin, J. J., et al. (2009), Comparison of CMAM simulations of carbon monoxide (CO), nitrous oxide (N_2O), and methane (CH_4) with observations from Odin/SMR, ACE-FTS, and Aura/MLS, *Atmos. Chem. Phys.*, *9*(10), 3233–3252.

Kerzenmacher, T., et al. (2008), Validation of NO_2 and NO from the Atmospheric Chemistry Experiment (ACE), *Atmos. Chem. Phys.*, *8*(19), 5801–5841.

Livesey, N. J., et al. (2005), Version 1.5 Level 2 data quality and description document, *Tech. Rep. D-32381*, Jet Propulsion Lab., Pasadena, Calif.

Mahieu, E., et al. (2008), Validation of ACE-FTS v2.2 measurements of HCl, HF, CCl_3F and CCl_2F_2 using space-, balloon- and ground-based instrument observations, *Atmos. Chem. Phys.*, *8*(20), 6199–6221.

Maronna, R. A., R. D. Martin, and V. J. Yohai (2006), *Robust Statistics: Theory and Methods*, John Wiley, New York.

Melo, S. M. L., et al. (2008), Summertime stratospheric processes at northern mid-latitudes: Comparisons between MANTRA balloon measurements and the Canadian Middle Atmosphere Model, *Atmos. Chem. Phys.*, *8*(7), 2057–2071.

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

Rousseeuw, P. J., and C. Croux (1993), Alternatives to the median absolute deviation, *J. Am. Stat. Assoc.*, *88*(424), 1273–1283.

Santee, M. L., et al. (2007), Validation of the Aura Microwave Limb Sounder HNO_3 measurements, *J. Geophys. Res.*, *112*, D24S40, doi:10.1029/2007JD008721.

SPARC CCMVal (2010), SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models, edited by V. Eyring, T. G. Shepherd, and D. W. Waugh, *SPARC Rep. No. 5, WCRP-132, WMO/TD1526*. (available at http://www.atmosp.physics.utoronto.ca/SPARC)

Strong, K., et al. (2008), Validation of ACE-FTS N_2O measurements, *Atmos. Chem. Phys.*, *8*(16), 4759–4786.

Toohey, M. (2009), Comparing remote sounding measurements of a variable stratosphere, Ph.D. thesis, Univ. of Toronto.

Toohey, M., and K. Strong (2007), Estimating biases and error variances through the comparison of coincident satellite measurements, *J. Geophys. Res.*, *112*, D13306, doi:10.1029/2006JD008192.

von Clarmann, T. (2006), Validation of remotely sensed profiles of atmospheric state variables: Strategies and terminology, *Atmos. Chem. Phys.*, *6*(12), 4311–4320.

Wolff, M. A., et al. (2008), Validation of HNO_3, $ClONO_2$, and N_2O_5 from the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), *Atmos. Chem. Phys.*, *8*(13), 3529–3562.

P. F. Bernath, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK.

C. D. Boone, Department of Chemistry, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

A. I. Jonsson, T. G. Shepherd, K. Strong, M. Toohey, and K. A. Walker, Department of Physics, University of Toronto, 60 St. George St., Toronto, ON M5S 1A7, Canada.