JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 112, D13306, doi:10.1029/2006JD008192, 2007

Estimating biases and error variances through the comparison of coincident satellite measurements

M. Toohey[1] and K. Strong[1]

Received 27 October 2006; revised 13 February 2007; accepted 21 March 2007; published 7 July 2007.

[1] A framework for the statistical comparison of six coincident remote sounding measurements is presented, which distinguishes between additive and multiplicative biases. The relationship between multiplicative bias and error variance is explored, and three methods are proposed for producing sets of values for three comparison variables: the multiplicative bias, and the error variance for each of two instruments. We illustrate and compare the three methods through the comparison of coincident measurements of the relatively long-lived stratospheric species O_3, N_2O, and HNO_3 from two independent measurement sets: version 2.2 retrievals (with updated O_3) from the Atmospheric Chemistry Experiment-Fourier transform spectrometer onboard SCISAT-1, and version 1.51 retrievals from the Earth Observing System Microwave Limb Sounder onboard Aura. We find that multiplicative bias between the two measurement sets, compared on a common vertical grid, is significant at some heights for O_3 and N_2O, and for all heights tested for HNO_3. The most realistic estimates of measurement error are produced by a method which incorporates a third correlative data set into the analysis. Using this method, estimated error standard deviations (SDs) are comparable between the two instruments for O_3 measurements, and are less than 10% of the mean measurement value between approximately 100 and 1 hPa. ACE N_2O measurements are consistent with a 10% error SD at all heights tested, although the uncertainty of the estimates is large at heights above 5 hPa. Estimated MLS N_2O error SDs are comparable with those for ACE in the lower stratosphere, but increase steeply with height. For HNO_3, estimated error SDs are approximately 10% between 70 and 10 hPa for both instruments. At heights above 10 hPa and below 100 hPa, estimated ACE errors are significantly smaller than those for MLS.

Citation: 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.

1. Introduction

[2] Comparisons of atmospheric quantities retrieved from remote sounding instruments aim to characterize the nature of the differences between the measurement sets. The goal of such comparisons may be the rigorous validation of a new instrument against a standard, or simply a description of the differences between two measurement sets (especially if both sets are unvalidated).

[3] Commonly, comparisons focus on a data subset for which both instruments are understood to measure approximately the same true quantity. Differences between "coincident measurements" are assumed to be a combination of random noise and systematic bias.

[4] The characterization of bias is the primary goal of comparison studies. Validation of an instrument requires that no bias be present in comparison to a standard instrument. The detection of bias signals the presence of an inconsistency between the two measurement systems which should be identified and corrected. Bias may be detected through the use of statistical tests, which require a priori information regarding the random instrument errors [*Rodgers*, 1990].

[5] When bias is present, a description of its nature may aid in the identification of its source. If the bias is only additive in nature (i.e., a constant offset between data sets), its magnitude can be estimated by the mean difference of the two measurement sets. This procedure has the advantage that the random error is averaged out over a large enough data set, and so detailed knowledge of the errors is not needed for the bias estimate. But bias may be multiplicative (i.e., dependent upon the magnitude of the measurement) as well as additive [*Hocking et al.*, 2001; *von Clarmann*, 2006]. In section 2, we present a comparison model which includes a multiplicative bias, as well as the more typically assumed additive bias and instrument error, and show that the mean difference may fail to illuminate bias in the data.

[6] A secondary goal of comparisons is to verify that the measurements are consistent with the predicted pre-

[1]Department of Physics, University of Toronto, Toronto, Ontario, Canada.

0148-0227/07/2006JD008192$09.00

cision of the measurements. We find that when multiplicative bias is present, its effect is confounded with the effect of measurement error in an analysis of the variance and covariance of the data. In section 3 we present three methods for separating the confounded parameters and producing sets of values for three variables: the multiplicative bias, and the error variance for the two measurements. For methods 1 and 2 we assume values for one of the three variables, and calculate the other two. In method 3, we incorporate a third measurement to estimate the multiplicative bias, which then allows the estimation of the error variances. A variation of method 1 has been used previously by *Fioletov et al.* [2006] to calculate uncertainties in ozone profile measurements from satellite, ozonesonde, and ground-based observations.

[7] We apply the three methods to coincident measurements of three stratospheric trace gases from two satellite instruments: the Atmospheric Chemistry Experiment-Fourier transform spectrometer (hereinafter, ACE), and the Microwave Limb Sounder (MLS) onboard the satellites SCISAT-1 and Aura, respectively. Descriptions of the instruments, as well as the process used to compose a set of coincident measurements are included in section 4, and the results of the analysis are discussed in section 5.

2. Comparison Framework

[8] In an idealized coincident measurement event, two instruments, X and Y, observe the atmosphere and retrieve some atmospheric quantity for one particular location in space and time. For each coincidence (indexed by subscript i) we define the measurements x_i and y_i as the quantities to be compared: in other words, we define the measurement system as encapsulating all processes leading to the compared quantities, including observation, retrieval, and any standardization to a common coordinate system. For the purposes of comparison we then model the two measurements as simple linear functions of the true atmospheric state τ_i; a random, zero-expectation-value, instrument-dependent error for each measurement δ_i and ϵ_i; and two variables describing potential bias between the measurements, an additive bias α and a multiplicative bias β:

$$x_i = \tau_i + \delta_i \tag{1}$$

$$y_i = \alpha + \beta\tau_i + \epsilon_i. \tag{2}$$

While the truth τ_i remains unknown, any comparison can only describe the relative biases between instruments. Only in the case that one instrument (for example, X) is believed to be adequately validated can any estimated bias be understood to be a systematic error of the second instrument (Y).

[9] Since the underlying relationship between any remote sounding measurement and the truth is some complex nonlinear function of many parameters, it is quite certain that the relationship between two sets of measurements is of higher order than the simple model presented here. Nonetheless, any statistical analysis of data explicitly or implicitly assumes some model of the data, and even the model given by equations (1) and (2) represents a significant increase in complexity over that implicitly assumed in typical remote sounding comparisons.

[10] Because of the random error on individual measurements, an ensemble of measurements is required in order to determine measurement bias. Random errors "average out" when mean quantities are calculated, i.e. $\bar{x} = \bar{\tau}, \bar{y} = \alpha + \beta\bar{\tau}$, and so, bias detection techniques often involve the comparison of mean measurements. In terms of the present comparison model, the difference of measurement means is given by:

$$\bar{y} - \bar{x} = \alpha + \bar{\tau}(\beta - 1). \tag{3}$$

When there is no multiplicative bias ($\beta = 1$), the difference of means is a viable method for calculating the additive bias α. On the other hand, when both bias terms are present, the difference of means may be small even if the terms α and β are not. In other words, it is possible that the two forms of bias compensate for one another. By differentiating between additive and multiplicative bias, the comparison model presented here is able to uncover bias undetectable by a simple comparison of measurement means.

[11] Once either the additive or multiplicative bias is found, the other can be calculated from equation (3). For the remainder of this work we thus focus on estimating the multiplicative bias. In order to isolate β, one must look at higher order moments of the measurement data: specifically, we aim to compare the variance and covariance of the measurements with their mathematical expectations based on the comparison model of equations (1) and (2).

[12] The expected variances of the comparison model are simplified by some assumptions. We assume that the measurement errors are independent, with zero mean and that the two instruments' measurement errors are also uncorrelated with each other (i.e. $\mathrm{Cov}(\delta; \epsilon) = 0$). We treat each vertical level independently, thereby ignoring cross correlation of errors between different heights. We also assume that the errors are independent of τ_i. These assumptions, and the corresponding covariance terms that are ignored in the analysis to follow are the source of uncertainties in standard regression analysis.

[13] We describe the uncertainty or precision of each measurement set with the error variances of each set, i.e. $\sigma_\delta^2 = \mathrm{Var}(\delta_i)$ and $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_i)$, and we define $\sigma_\tau^2 = \mathrm{Var}(\tau_i)$ as the true natural atmospheric variance. From equations (1) and (2), and the assumptions stated above, we then obtain the following equations for the expected population variances for x and y, and the covariance of x and y:

$$\sigma_x^2 = \sigma_\tau^2 + \sigma_\delta^2 \tag{4}$$

$$\sigma_y^2 = \beta^2\sigma_\tau^2 + \sigma_\epsilon^2 \tag{5}$$

$$\sigma_{xy} = \beta\sigma_\tau^2. \tag{6}$$

Thus, the expected measurement variances (σ_x^2, σ_y^2, σ_{xy}) are functions of the true atmospheric variance (σ_τ^2), the error variances (σ_δ^2, σ_ϵ^2), and the multiplicative bias (β).

[14] Given a sample of n coincidences, we calculate the unbiased sample variances s_{xx} and s_{yy} and covariance s_{xy}:

$$s_{xx} = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{7}$$

$$s_{yy} = \frac{1}{(n-1)} \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{8}$$

$$s_{xy} = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{9}$$

[15] Equating the sample statistics (s_{xx}, s_{yy}, s_{xy}) with their expectations (σ_x^2, σ_y^2, σ_{xy}), and eliminating σ_τ^2 in equations (4) and (5) using equation (6), we obtain two equations that relate the error variances and the multiplicative bias in terms of the sample statistics s_{xx}, s_{yy}, and s_{xy}:

$$\hat{\sigma}_\delta^2 = s_{xx} - \frac{1}{\hat{\beta}} s_{xy} \tag{10}$$

$$\hat{\sigma}_\epsilon^2 = s_{yy} - \hat{\beta} s_{xy} \tag{11}$$

[16] Such quantities derived from the data through the sample statistics are estimates, and as such will be denoted by a caret (^) above the variable.

[17] At this point we are faced with two equations for three unknowns. *Hocking et al.* [2001] presented equivalent expressions for the underdetermined relationship coupling bias and instrument error variances in the analysis of radar-measured stratospheric winds. As discussed by *Hocking et al.* [2001], and explored in great detail in a more general context by *Dunn* [1989], estimation of any of the three comparison variables requires one variable to be fixed through assumption, or through the incorporation of more information into the analysis. We present three methods for proceeding from equations (10) and (11) toward values for $\hat{\beta}$, $\hat{\sigma}_\delta^2$, and $\hat{\sigma}_\epsilon^2$.

3. Variable Estimation Methods

3.1. Method 1: Grubbs Estimators

[18] If there is no multiplicative bias between measurements ($\beta = 1$), then the instrument error estimates given in equations (10) and (11) simplify to:

$$\hat{\sigma}_{\delta 1}^2 = s_{xx} - s_{xy} \tag{12}$$

$$\hat{\sigma}_{\epsilon 1}^2 = s_{yy} - s_{xy} \tag{13}$$

(where the subscript "1" refers to the method used). These relations were derived by *Grubbs* [1948, 1973] and are therefore frequently referred to as Grubbs estimators in the statistics literature [*Dunn*, 1989]. *Fioletov et al.* [2006] implicitly assume $\beta = 1$ in deriving estimates of instrument error in comparisons of ozone profile measurements from satellite, ozonesonde, and ground-based observations, although they used an equivalent formulation using the variance of measurement differences rather than measurement covariances.

3.2. Method 2: Use of Predicted Error Variances

[19] If we are provided with a reliable estimate of the error variance of one instrument, we can then estimate the multiplicative bias and the other instrument error variance. Let predictions of statistical parameters based on reported quantities be denoted by a tilde ($\sim$) above the variable, in contrast to the estimates marked by a caret. Then, given a prediction of the error variance for instrument X, $\tilde{\sigma}_\delta^2$, we solve equations (10) and (11) for the estimates $\hat{\beta}$ and $\hat{\sigma}_\epsilon^2$

$$\hat{\beta}_2|_{\tilde{\sigma}_\delta^2} = \frac{s_{xy}}{s_{xx} - \tilde{\sigma}_\delta^2} \tag{14}$$

$$\hat{\sigma}_{\epsilon 2}^2 = s_{yy} - \frac{s_{xy}^2}{s_{xx} - \tilde{\sigma}_\delta^2} \tag{15}$$

[20] Retrieved quantities from satellite observations typically report an error derived from the spectral fitting residuals, and the propagation of this fitting error through the retrieval algorithm (or some similar procedure). We take each reported error for instrument X, d_i, to be a prediction of the absolute value of δ_i, the difference between the measurement and the truth (i.e., $d_i = |\delta_i|$, $\delta_i = x_i - \tau_i$). Recalling that the expected value of δ_i is zero, we write the predicted instrument X error population variance in terms of the reported error d_i:

$$\tilde{\sigma}_\delta^2 = \frac{1}{n} \sum_{i=1}^{n} (\delta_i - \bar{\delta})^2 = \frac{1}{n} \sum_{i=1}^{n} d_i^2, \tag{16}$$

[21] The predicted error variance is then simply the mean of the square of the individual measurement error predictions.

[22] If we choose instead to use the predicted error variance for instrument Y, $\tilde{\sigma}_\epsilon^2$, calculated as the mean of the square of the predicted instrument Y error terms e_i, then estimates of the multiplicative bias and instrument X error variance are given by:

$$\hat{\beta}_2|_{\tilde{\sigma}_\epsilon^2} = \frac{s_{yy} - \tilde{\sigma}_\epsilon^2}{s_{xy}} \tag{17}$$

$$\hat{\sigma}_{\delta 2}^2 = s_{xx} - \frac{s_{xy}^2}{s_{yy} - \tilde{\sigma}_\epsilon^2} \tag{18}$$

[23] It can be noted that in the unrealistic (but often assumed) case of zero error in measurement x, we have $\hat{\beta}_2 = s_{xy}/s_{xx}$, which is equivalent to an ordinary least squares fit of y to x. Equation (14) can easily be used to show that ordinary least squares fitting of y to x underestimates the true slope in magnitude when x is subject to error.

[24] Other options for incorporating a priori error information into the analysis, including assuming $\tilde{\sigma}_\epsilon^2 = \tilde{\sigma}_\delta^2$, or taking the ratio $\tilde{\sigma}_\delta^2/\tilde{\sigma}_\epsilon^2$ as a known quantity will not be explored here.

3.3. Method 3: Instrument Variable Method

[25] If a suitable third variable, z, is measured in addition to x and y, instrument model parameters may be estimated directly from the data, without making assumptions regarding any of the parameters. This third measurement, known as an instrument variable [*Dunn*, 1989], need only be correlated with τ, while uncorrelated with the measurement errors for x and y.

[26] We choose as our instrument variable a second pseudo-coincident observation by instrument Y. In this special case, this third coincident observation is modeled as:

$$z_i = \alpha + \beta(\tau_i + \eta_i) + \gamma_i \tag{19}$$

where the instrument model parameters α and β are common between the two observations by instrument Y. We assume that each measurement error γ_i is uncorrelated with δ_i or ϵ_i, and that $\sigma_\gamma^2 = \sigma_\epsilon^2$. We introduce the term η: a perturbation to the true state τ due to the fact that our second Y observation is not perfectly coincident in time and space with the X and first Y observation. We assume the noncoincidence parameter η is uncorrelated with the measurement errors, and define the variance of η over a set of measurements as σ_η^2.

[27] With a third measurement, we obtain three more equations for the expected population variances and covariances:

$$\sigma_z^2 = \beta^2\sigma_\tau^2 + \beta^2\sigma_\eta^2 + \sigma_\epsilon^2 \tag{20}$$

$$\sigma_{xz} = \beta\sigma_\tau^2 \tag{21}$$

$$\sigma_{yz} = \beta^2\sigma_\tau^2. \tag{22}$$

[28] We can then obtain estimates of the right-hand-sides of equations (20), (21), and (22) (in terms of the estimates $\hat{\beta}, \hat{\sigma}_\tau^2, \hat{\sigma}_\eta^2$, and $\hat{\sigma}_\epsilon^2$) if we replace the expected variances σ_z^2, σ_{xz}, and σ_{yz} with the sample variances s_{zz}, s_{xz}, and s_{yz}. The ratio s_{yz} / s_{xz} is known as the instrumental variable estimate of β [*Dunn*, 1989], i.e.:

$$\hat{\beta}_3 = \frac{s_{yz}}{s_{xz}}. \tag{23}$$

It is valid in general for any suitable choice of instrument variable, not only for the special case of a secondary measurement by one instrument explored here.

[29] With the multiplicative bias estimated directly from the data, it becomes possible to write equations (10) and (11) strictly in terms of the measurement statistics:

$$\hat{\sigma}_{\delta 3}^2 = s_{xx} - \frac{s_{xy}s_{xz}}{s_{yz}} \tag{24}$$

$$\hat{\sigma}_{\epsilon 3}^2 = s_{yy} - \frac{s_{xy}s_{yz}}{s_{xz}}. \tag{25}$$

[30] We define σ_ν^2 as the total error variance for the secondary Y measurement, the sum of the noncoincidence error ($\beta^2\sigma_\eta^2$) and the Y measurement error variance (σ_ϵ^2), and estimate this quantity from the data in a procedure equivalent to the two error variances [equations (24) and (25)]:

$$\hat{\sigma}_{\nu 3}^2 = s_{zz} - \frac{s_{xy}s_{yz}}{s_{xz}}. \tag{26}$$

When the noncoincidence error variance σ_η^2 is small, we expect $\hat{\sigma}_{\nu 3}^2 \approx \hat{\sigma}_{\epsilon 3}^2$. It should be noted that a noncoincidence parameter comparable to η could have been used in the original expression for y [equation (2)], explicitly describing the imperfect coincidence of the X and the primary Y observations. Under this model formulation, the noncoincidence error variance becomes confounded with the Y error variance. As such, we maintain the model as presented, but stress that the estimated measurement error, $\hat{\sigma}_\epsilon^2$, calculated through all three methods, should be understood as the sum of the true measurement error variance and an implicit noncoincidence error variance.

3.4. Estimating Confidence Intervals for Estimates Through Bootstrapping

[31] Without invoking any assumptions concerning the distribution of the measurement errors (normal or otherwise), we can estimate the variance of the instrument variables estimated above through the use of bootstrapping [*Efron and Tibshirani*, 1994]. This method allows for the determination of the standard error in each estimate through the sensitivity of the estimate to repeated random resampling of the data. We use bootstrapping to estimate the 95% confidence intervals of $\hat{\beta}_3$, $\hat{\sigma}_{\delta 3}^2$, and $\hat{\sigma}_{\epsilon 3}^2$ based on 1000 resamplings.

4. Data

[32] Here we compare results from the satellite instruments MLS and ACE over the full year of 2005. We focus on the relatively long-lived stratospheric species O_3, N_2O, and HNO_3, in order to try to minimize the effect of noncoincidence error.

[33] ACE is a Fourier transform spectrometer operating at high spectral resolution in the infrared, measuring atmospheric extinction by solar occultation, from which profiles of temperature, pressure, and dozens of constituents are retrieved through a global fitting algorithm [*Bernath et al.*, 2005; *Boone et al.*, 2005]. ACE results are from the version 2.2 data set, with O_3 results from the version 2.2 ozone update.

[34] MLS is used to retrieve atmospheric temperature and more than a dozen atmospheric constituent profiles through the measurement of thermally emitted radiation from the Earth's limb [*Waters et al.*, 2006]. MLS version 1.51 retrieval results are used here, and are based on an optimal estimation method which includes use of a priori constraints [*Livesey et al.*, 2005a]. MLS makes limb scans in the forward direction along the satellite orbit track, hence, consecutive scans cover significantly overlapping regions of the atmosphere. The MLS retrieval technique takes advantage of this fact by dividing the collected radiance data into "chunks" of about ten vertical scans, and simultaneously retrieving a similar number of profiles of atmospheric temperature and composition from each chunk. Therefore the retrievals within each chunk are not independent.

[35] ACE and MLS observe the atmosphere with similar limb-viewing geometries. Consequently, the retrieved pro-

Table 1. Number of Coincidences (n) for Each Species, and Mean ($\pm 1\ \sigma$) Differences Between ACE and MLS Primary (Subscript 1) and Secondary (Subscript 2) Measurement Latitude (ϕ), Longitude (λ), and Local Solar Time (t)

Species	O_3	N_2O	HNO_3
n	664	590	701
$\overline{\Delta\phi_1}$ (°)	0.04 ± 0.47	0.03 ± 0.47	0.03 ± 0.47
$\overline{\Delta\phi_2}$ (°)	0.01 ± 0.46	−0.01 ± 0.47	−0.01 ± 0.46
$\overline{\Delta\lambda_1}$ (°)	0.27 ± 2.87	0.14 ± 2.9	0.18 ± 2.89
$\overline{\Delta\lambda_2}$ (°)	0.16 ± 2.92	0.31 ± 2.93	0.1 ± 2.91
$\overline{\Delta t_1}$ (hr)	3.37 ± 5.66	3.54 ± 5.84	3.75 ± 6.15
$\overline{\Delta t_2}$ (hr)	3.45 ± 10.28	3.76 ± 9.9	3.66 ± 10.08

files from both instruments have similar resolutions in the horizontal (∼500 km) and vertical (3–6 km, depending on species and altitude and ACE occultation viewing geometry).

[36] ACE measured spectra have a signal-to-noise ratio of greater than 300 over most of the spectral range. Uncertainties provided for the ACE mixing ratio results are one-sigma statistical errors from its global-fitting retrieval algorithm, and do not include systematic contributions [*Boone et al.*, 2005]. MLS retrieval precisions are calculated as a function of the measurement error and the a priori error covariance matrix [*Livesey et al.*, 2005a].

[37] Validation of the data products from ACE and MLS is presently ongoing [e.g., *Walker et al.*, 2005; *Froidevaux et al.*, 2006]. Version 1.5 MLS and ACE version 2.1 retrieved trace gas profiles have been compared by *Froidevaux et al.* [2006].

[38] MLS data used in this work is screened based on the precision, status, and quality fields of the MLS data files as described by *Livesey et al.* [2005b]. In addition, the MLS N_2O data is filtered according to flags provided by the MLS team, in order to attempt to remove a systematic high bias in low altitude vortex N_2O retrievals (as also discussed in *Livesey et al.* [2005b]). In addition, in order to remove some suspicious ACE profiles, an ad hoc filter has been implemented, excluding from consideration any ACE O_3 profiles for which the error exceeds 160 ppbv, and any N_2O profiles outside the range −10 to 800 ppbv.

[39] For the application of the instrument variable estimation methods described above, we aim to produce from the combined ACE and MLS data sets a subset of coincident measurements of stratospheric air at the same height, latitude, longitude, and time. The latter three requirement coordinates are dealt with by finding coincident observations of vertical profiles. In order to meet the data requirements of all three comparison methods, we produce a set of coincident profile measurements for each species based on a two-stage coincidence criterion. All coincidences are defined herein as those occurring within ±5° longitude (λ) and ±1° latitude (ϕ). Primary coincidences are defined as those occurring within ±6 hours. If multiple potential primary coincidences are found, we choose the coincidence for which the parameter $D = \Delta\phi(^\circ) + \Delta t(\text{hours})$ is minimized. Secondary coincidences (defining set z) are found within ±12 hours of the primary coincidence. In order to ensure independence between the MLS measurement errors ϵ and

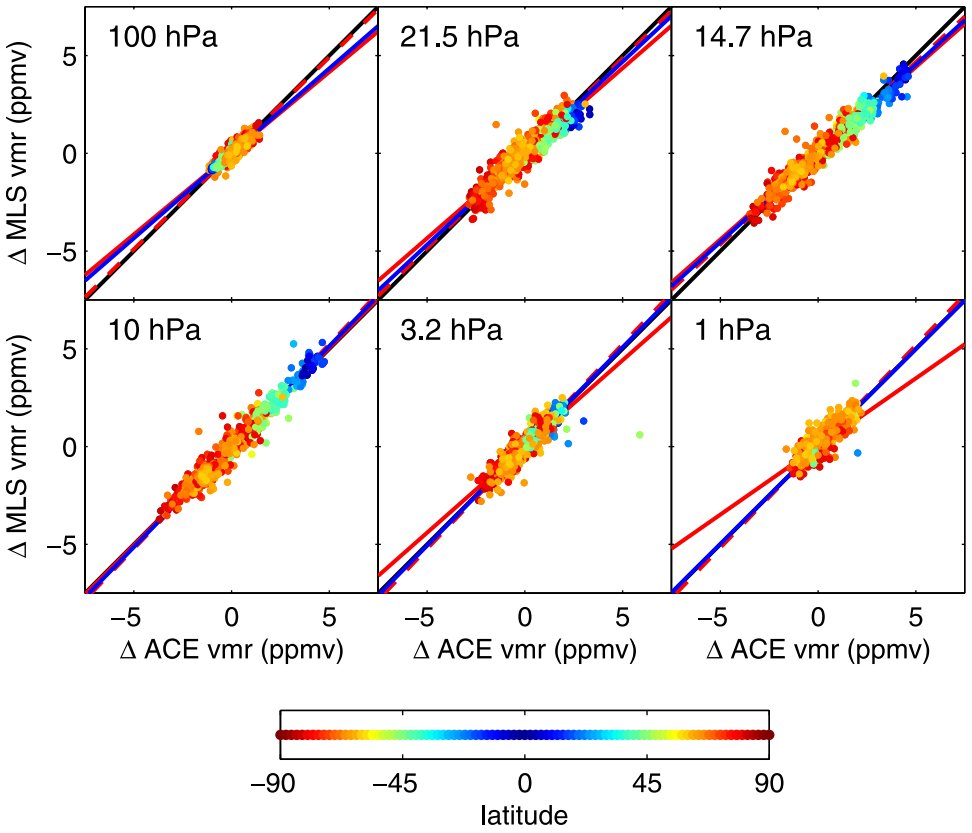

Figure 1. Scatterplots of MLS versus ACE-measured O_3 anomalies for selected pressure surfaces. Lines have slope: $\beta = 1$ (black), $\hat{\beta}_2|_{\hat{\sigma}^2_\delta}$ (red), $\hat{\beta}_2|_{\hat{\sigma}^2_\epsilon}$ (red dashed), and $\hat{\beta}_3$ (blue). Color coding of points is based on the absolute value of latitude, with equatorial measurements in blue and polar measurements in red.

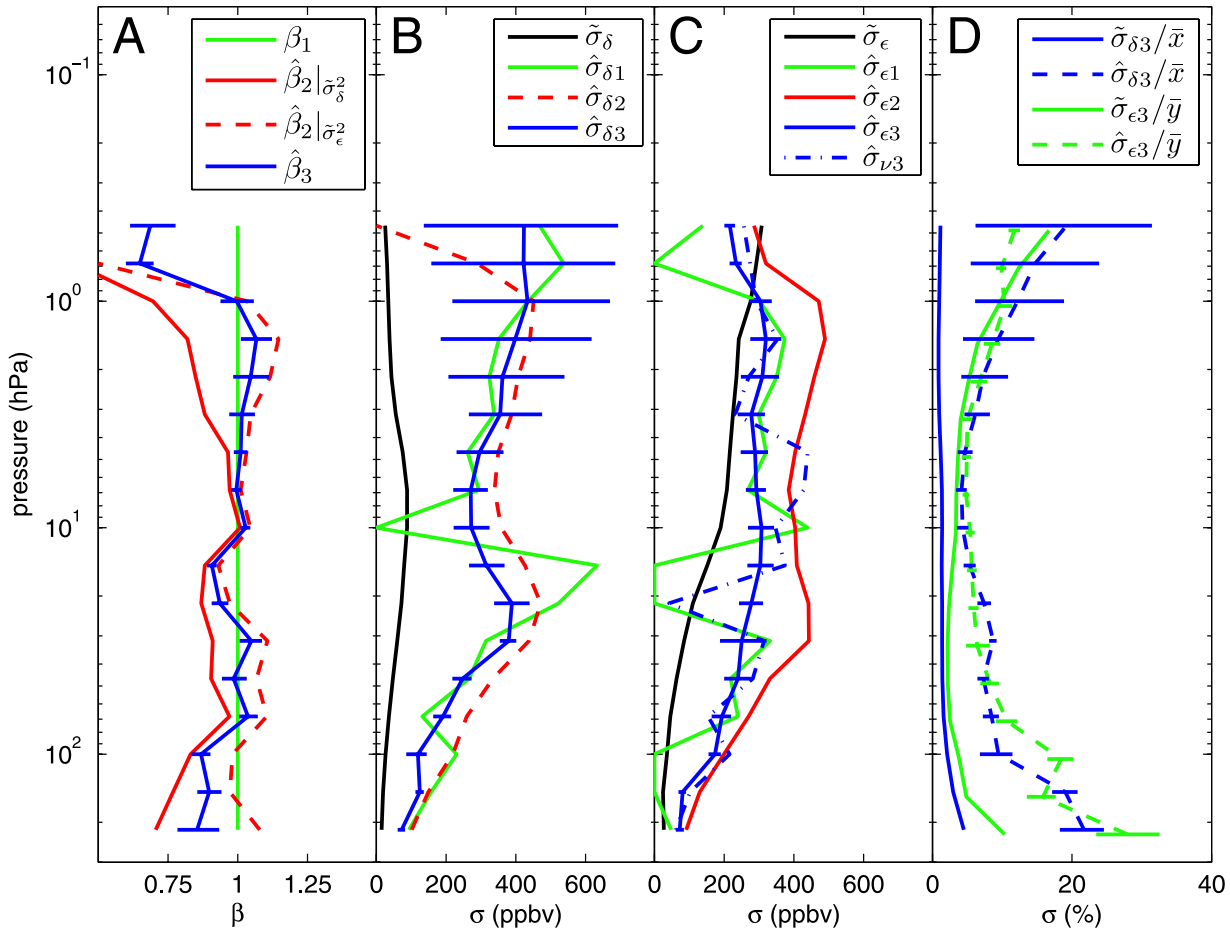

Figure 2. O_3 biases and errors: (a) Three estimates of the multiplicative bias [$\hat{\beta}_2|_{\tilde{\sigma}^2_\delta}$ derived from predicted ACE error variances, $\hat{\beta}_2|_{\tilde{\sigma}^2_\epsilon}$ derived from predicted MLS error variances, and $\hat{\beta}_3$; equations (14), (17) and (23)] compared to $\beta_1 = 1$. (b) Predicted and estimated ACE error standard deviation (SD) profiles [equations (12), (16), (18), and (24)]. (c) Predicted and estimated MLS error SD profiles [equations (13), (15), (16), (25), and (26)]. (d) Predicted ($\tilde{\sigma}$) and method 3 estimated ($\hat{\sigma}$) measurement error SD profiles for ACE (blue) and MLS (green), in percent of mean measurement. 95% confidence intervals for all quantities estimated by method 3 are shown with error bars.

γ, we require that the secondary coincidence be from a different retrieval "chunk" from the primary coincidence (in practice we conservatively require a difference of at least three "chunks"). Again, if multiple secondary coincidences are found, the parameter D defined above is minimized to find the best secondary coincidence. If no secondary coincidence is found, the primary coincidence is thrown out. Table 1 gives some statistics for the number of coincidences found and the space-time proximity of the coincidences.

[40] ACE and MLS measurements must then be mapped onto a common vertical grid. A fair comparison should take into account the differing characteristics of the observing systems. Comparison methods incorporating differences described by the instrument averaging kernels and error covariances have been described [e.g., *Rodgers and Connor*, 2003]. Many comparisons [e.g., *Froidevaux et al.*, 2006] proceed by the simpler route of interpolating to a common vertical grid. When the vertical resolutions of the two instruments are similar (as is the case for ACE and MLS), it is generally assumed that this simpler procedure will not adversely affect the statistics of the comparison. Also, this technique is applicable when averaging kernels and error covariances are unavailable for the data. In this comparison, retrieved ACE mixing ratio and uncertainty profiles, originally reported on a geometric height vertical grid of 1-km resolution, are linearly interpolated to the MLS pressure grid of six surfaces per decade change in pressure, using the ACE retrieved pressure profile for each measurement. This process adds an unquantifiable error to the resultant ACE mixing ratio profiles due to an unquantified error in the ACE retrieved pressure (C. D., Boone, personal communication, 2006). Furthermore, we expect that interpolation introduces its own unquantifiable sampling error.

5. Results

5.1. O_3

[41] In a comparison of global mean coincident profiles from MLS and other instruments, *Froidevaux et al.* [2006] showed that MLS retrieved O_3 values tend to be, in general, slightly higher than the comparison results in the lower

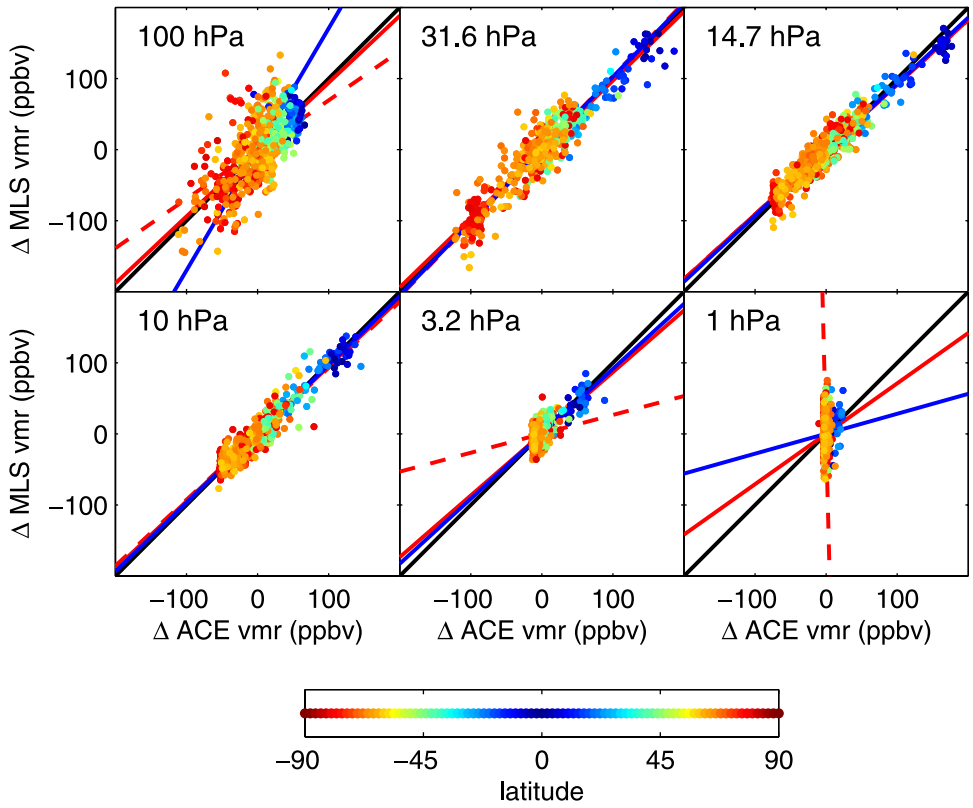

Figure 3. Scatterplots of MLS versus ACE measured N_2O anomalies for selected pressure surfaces. Lines have slope: $\beta = 1$ (black), $\hat{\beta}_2|_{\tilde{\sigma}^2_\delta}$ (red), $\hat{\beta}_2|_{\tilde{\sigma}^2_\varepsilon}$ (red dashed), and $\hat{\beta}_3$ (blue). Color coding of points is based on the absolute value of latitude, with equatorial measurements in blue and polar measurements in red.

stratosphere, and slightly lower in the upper stratosphere. The degree of "tilt" in the slope of the average differences was shown to depend on the measurement compared to: MLS was shown to agree with HALOE results to within 5% over essentially the whole range from 100 to 1 hPa, and similarly well with SAGE II results, except in the region near 1 hPa, where MLS was lower by 10 to 15%. ACE version 2.2 O_3 update results represent a significant improvement upon the results compared to MLS in the same study.

[42] Figure 1 shows MLS versus ACE O_3 volume mixing ratio (VMR) anomalies (means subtracted) for some selected pressure surfaces. The points of the scatterplots are color-coded based on the absolute value of latitude, with polar values in red and equatorial values in blue. The widest range of O_3 mixing ratio values is seen between 21.54 and 10 hPa, i.e., the absolute natural variance σ^2_τ is maximum at these heights.

[43] In the absence of multiplicative bias (i.e., under the assumptions of method 1), the measurements should lie along a 1:1 line with slope of one (shown in black on the scatterplots), with scatter about the line due to the error variance of each measurement. Any deviation of slope from the 1:1 line is evidence of a multiplicative bias in the data. Lines with slopes corresponding to the multiplicative bias estimates $\hat{\beta}_2|_{\tilde{\sigma}^2_\delta}$, $\hat{\beta}_2|_{\tilde{\sigma}^2_\varepsilon}$ and $\hat{\beta}_3$ (calculated by equations (14), (17), and (23)] are plotted on each scatterplot. Multiplicative bias estimate $\hat{\beta}_2|_{\tilde{\sigma}^2_\delta}$ is visibly less than one for all but one (100 hPa) pressure surface shown here, while $\hat{\beta}_2|_{\tilde{\sigma}^2_\varepsilon}$ is quite close to one for all plots. Estimated multiplicative bias $\hat{\beta}_3$ is less than one below 10 hPa, and quite close to one otherwise.

[44] Multiplicative bias estimates $\hat{\beta}_2|_{\tilde{\sigma}^2_\delta}$, $\hat{\beta}_2|_{\tilde{\sigma}^2_\varepsilon}$ and $\hat{\beta}_3$ are plotted in Figure 2a as a function of the MLS retrieval pressure surfaces. $\hat{\beta}_3$ is generally consistent with a slope of one, except above 1 hPa, and below 100 hPa, and at 14.68 and 23.54 hPa. Estimates calculated via method 2 bracket $\hat{\beta}_3$, and generally follow its vertical structure.

[45] Figures 2b and 2c show the measurement error estimates corresponding to each multiplicative bias profile in Figure 2a. Error variance estimates calculated by methods 1, 2, and 3, are converted to standard deviation (SD) estimates by taking the square root. The error SD profiles are in units of parts per billion. Also shown are the predicted error SD profiles, calculated as the square root of equation (16).

[46] Method 1 leads to negative error variance estimates at a number of heights. The error SD is undefined for such cases, but is set to zero for the profile plots (for example, at 10 hPa for ACE; 0.7, 14, 23, and 100 hPa for MLS). Negative variance estimates can result from a mis-specified model or an insufficient sample size [*Dunn*, 1989]. In this case, comparing Figures 2a–2c shows that the negative variances calculated by method 1 for MLS occur at those heights for which the multiplicative bias estimates are furthest from a value of 1. Hence at these heights the $\beta = 1$ assumption of method 1 is most suspect, and the method leads to erroneous estimates. To be more specific, the form of equation (11) shows that if the true multiplicative bias is less than 1, then assuming $\beta_1 = 1$ gives too much weight to the second term on the right-hand-side, leading to a negative value for the error variance. Similarly, the form of equation (10) shows that if the true multiplicative bias is

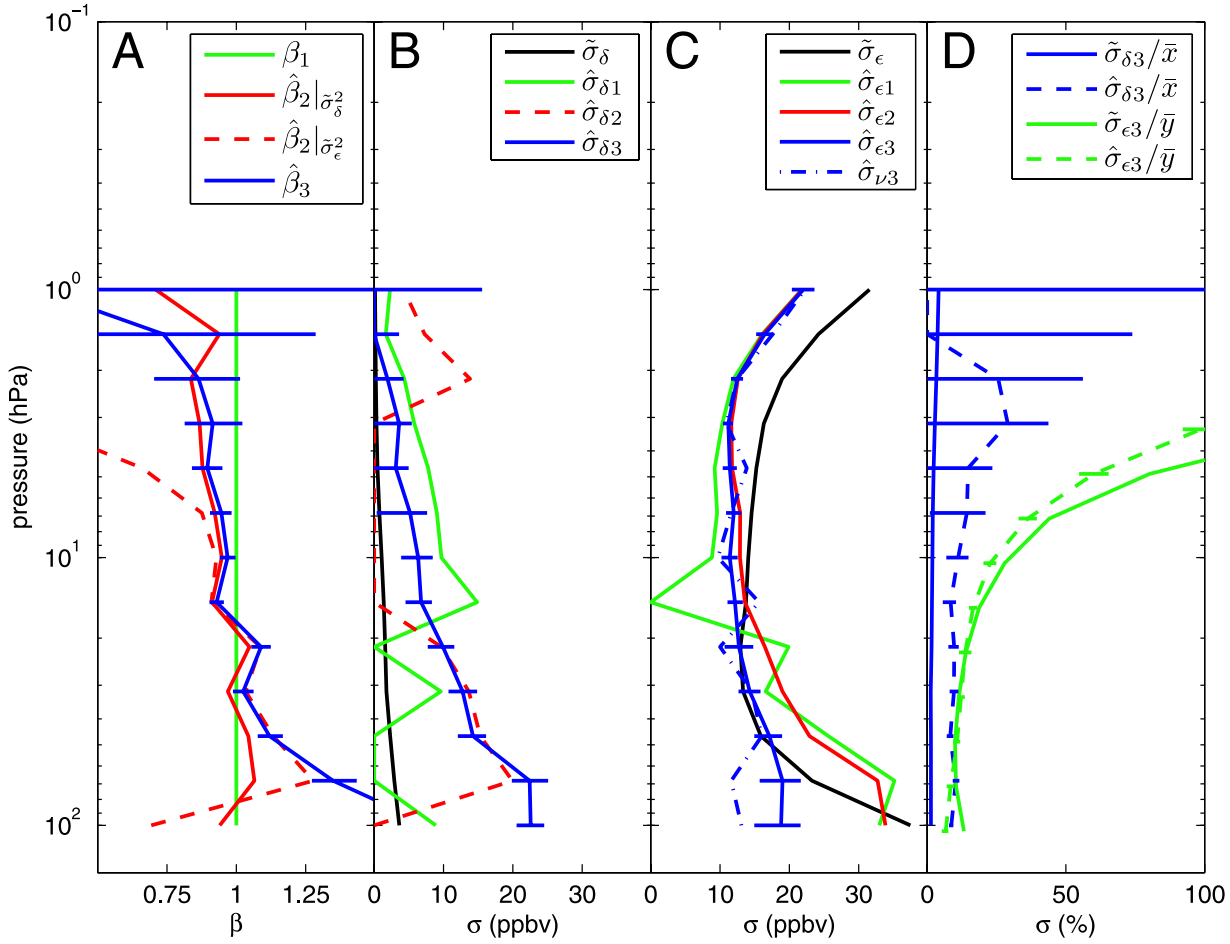

Figure 4. N_2O biases and errors: (a) Three estimates of the multiplicative bias [$\hat{\beta}_2|_{\tilde{\sigma}^2_\delta}$ derived from predicted ACE error variances, $\hat{\beta}_2|_{\tilde{\sigma}^2}$ derived from predicted MLS error variances, and $\hat{\beta}_3$; equations (14), (17) and (23)] compared to $\beta_1 = 1$. (b) Predicted and estimated ACE error standard deviation (SD) profiles [equations (12), (16), (18), and (24)]. (c) Predicted and estimated MLS error SD profiles [equations (13), (15), (16), (25), and (26)]. (d) Predicted ($\tilde{\sigma}$) and method 3 estimated ($\hat{\sigma}$) measurement error SD profiles for ACE (blue) and MLS (green), in percent of mean measurement. 95% confidence intervals for all quantities estimated by method 3 are shown with error bars.

greater than 1, then assuming $\beta = 1$ will lead to negative values for the ACE error variance. This fact occurs at 10 hPa (Figure 2b). Such unphysical estimates of the error variances motivate the use of methods 2 and 3.

[47] In method 2, the error estimate for each instrument is calculated based on the predicted error ($\tilde{\sigma}$) for the other instrument. It should be noted that by assuming the predicted error of one instrument is correct, we attribute any noncoincidence error to the other instrument. Thus while the estimated MLS error variance contains the implicit noncoincidence error like the other methods, the method 2 estimated ACE error variance contains noncoincidence error as well. This fact explains why the method 2 estimated ACE error variance is generally slightly larger than the estimate given by methods 1 and 3. However, these estimates are all much larger than the predicted ACE error variance, which explains why the method 2 estimated MLS error variance, based on the ACE error predictions, is significantly larger than the other estimates and the prediction.

[48] Method 3 error SD estimates generally agree with the results of method 1, except for the problem heights discussed above, for which method 3 leads to more realistic values. Comparison of $\hat{\sigma}_{\epsilon 3}$ and $\hat{\sigma}_{\nu 3}$ in Figure 2c shows that the effect of noncoincidence error is negligible below 30 hPa. Above 20 hPa the diurnal variability of O_3 likely produces significant noncoincidence error variance. The 95% confidence intervals calculated for method 3 estimates exclude the predicted error SD profiles in all cases.

[49] Figure 2d shows predicted error SDs and those estimated by method 3 in terms of percent of the mean O_3 profile measured by each instrument. Below 100 hPa, the estimates for both instruments are up to 20% higher than the predictions. Between 100 and 4 hPa, MLS estimates are ~1–8% larger than the predictions, while ACE estimates are ~3–8% greater than the predictions. Between 10 and 1 hPa, MLS estimates are within 2% of the predictions, and ACE estimates are 3–10% greater than the predictions, although the 95% confidence intervals at these higher altitudes imply the ACE error SD estimates could easily be within 5% of the predictions.

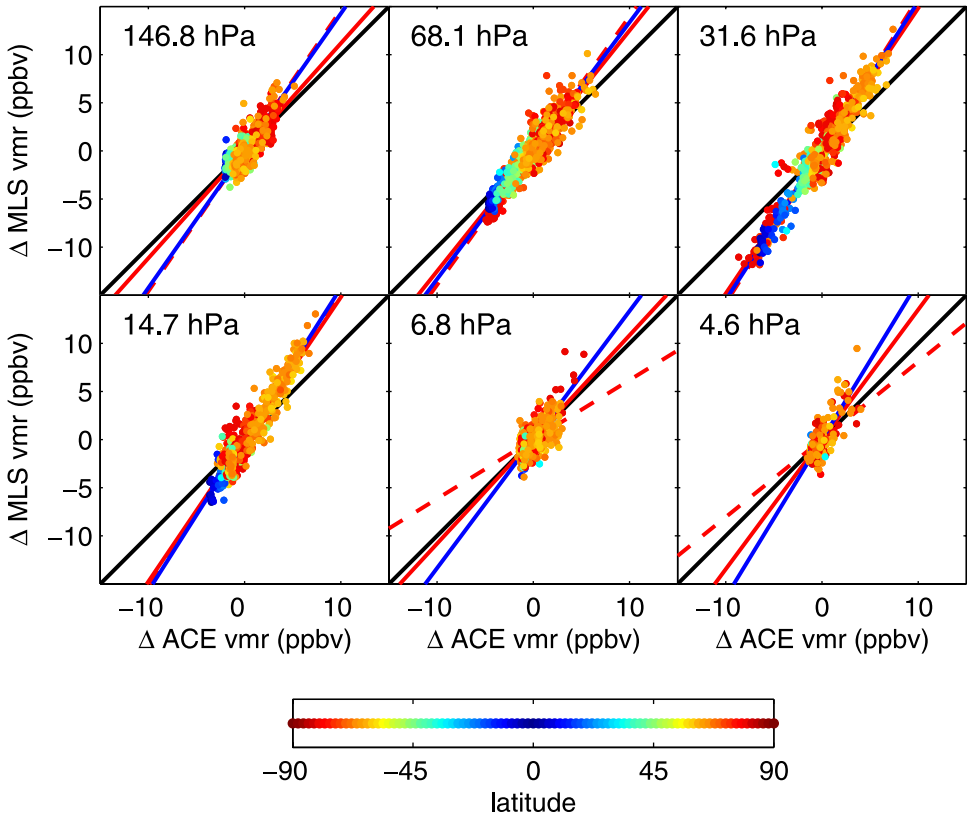

Figure 5. Scatterplots of MLS versus ACE measured HNO_3 anomalies for selected pressure surfaces. Lines have slope: $\beta = 1$ (black), $\hat{\beta}_2|_{\tilde{\sigma}_\delta^2}$ (red), $\hat{\beta}_2|_{\tilde{\sigma}_\epsilon^2}$ (red dashed), and $\hat{\beta}_3$ (blue). Color coding of points is based on the absolute value of latitude, with equatorial measurements in blue and polar measurements in red.

5.2. N_2O

[50] Figure 3 shows MLS versus ACE N_2O VMR anomalies for selected pressure surfaces. At high altitudes (for example, 1 hPa), the absolute variance of MLS anomalies is much larger than that for ACE. This discrepancy in variance is likely more significantly due to noisy MLS retrievals (for which significant averaging is suggested for useful signals [*Livesey et al.*, 2005b]) rather than insensitivity in the ACE retrievals. As will be discussed below, the reported MLS N_2O error variances used as predictions of the error variance in method 2, are larger than the true random error in the measurements. This leads to anomalously small (for example, at 3.2 hPa) and even negative (at 1 hPa) estimates of multiplicative bias given by method 2 when using these predicted errors.

[51] Throughout the middle stratosphere, the natural absolute variance of N_2O is larger than at high altitudes, and the scatter lies roughly along the 1:1 line. At 100 hPa, the effects of the reported noisy retrievals at high latitudes is apparent, as the mask used to filter the N_2O data is seen to filter most, but not all of the anomalously noisy data.

[52] Figure 4a shows the multiplicative bias estimates. At high altitudes, the small natural variability, and the large MLS variance leads to large uncertainties in the slope estimates, and the uncertainties in the estimate from method 3 include a slope of one.

[53] Moving down in altitude, between ~4 and 14 hPa, the estimated bias is slightly, but significantly less than one. At heights below 30 hPa, $\hat{\beta}_2|_{\tilde{\sigma}_\epsilon^2}$ and $\hat{\beta}_3$ values are anomalously large due to the noisy MLS polar measurements, while $\hat{\beta}_2|_{\tilde{\sigma}_\delta^2}$ remains in close agreement with $\beta = 1$.

[54] The error SD estimates corresponding to the bias estimates in Figure 4a are shown in Figures 4b and 4c. As was the case for O_3, method 1 leads to a number of negative error variance estimates (as shown by plotted error SD values of zero) at heights where multiplicative bias is significant.

[55] Focussing on the MLS error SD estimates of Figure 4c, at low altitudes the estimates based on methods 1 and 2 agree closely and are ~30% larger than the predicted values, as might be expected due to the anomalous polar retrievals. Method 3 apparently underestimates the error SD at these low altitudes, as the slope estimate $\hat{\beta}_3$ is seen instead to compensate for the noisy polar measurements. Between 14 and 68 hPa, the 95% confidence intervals of $\hat{\sigma}_{\epsilon 3}$ include the predicted error SD profile.

[56] At heights above 10 hPa, all three methods lead to MLS error SD estimates that converge as the altitude increases. At these heights, the large variance of the MLS measurements (s_{yy}) dominates all other terms in the expressions for MLS error variance. The MLS error SDs above 10 hPa are smaller than the predicted values. This result is consistent with simulation results discussed by *Livesey et al.* [2005b] in which the variance of quantities retrieved from simulated, noisy spectra was less than the error variance (or precision) predicted by the retrieval algorithm. This occurs when the effect of "retrieval smoothing" artificially relaxes retrievals toward a mean or an a priori value.

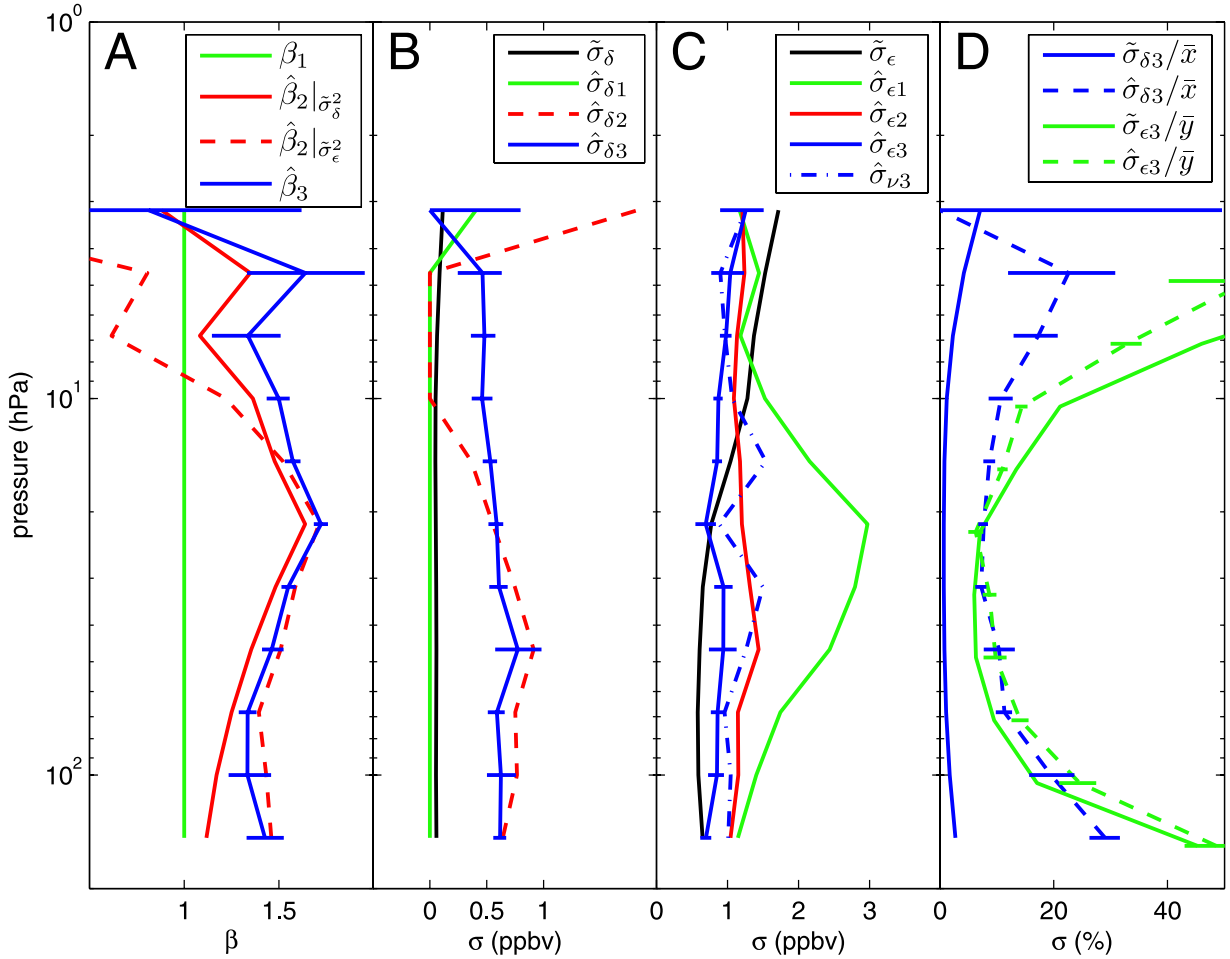

Figure 6. HNO_3 biases and errors: (a) Three estimates of the multiplicative bias [$\hat{\beta}_2|_{\tilde{\sigma}^2_\delta}$ derived from predicted ACE error variances, $\hat{\beta}_2|_{\tilde{\sigma}^2_\epsilon}$ derived from predicted MLS error variances, and $\hat{\beta}_3$; Equations (14), (17), and (23)] compared to $\beta_1 = 1$. (b) Predicted and estimated ACE error standard deviation (SD) profiles [Equations (16), (12), (18), and (24)]. (c) Predicted and estimated MLS error SD profiles Equations (16), (13), (15), (25), and (26). (d) Predicted ($\tilde{\sigma}$) and Method 3 estimated ($\hat{\sigma}$) measurement error SD profiles for ACE (blue) and MLS (green), in percent of mean measurement. 95% confidence intervals for all quantities estimated by Method 3 are shown with error bars.

[57] Where the estimated MLS error SDs are smaller than the predicted values, the estimation of ACE error SD based on the MLS predictions via method 2 leads to anomalously negative variance values. Only method 3 leads to physically acceptable error SD profiles for ACE over the vertical measurement range.

[58] Figure 4d shows the estimated and predicted error SD profiles in percent of the mean measured profile. The ACE error SD estimates are roughly 8% higher than the predictions between 100 and 5 hPa. Above 5 hPa, the ACE error SD 95% confidence intervals become exceedingly large. The estimated MLS percent-error SD profile is in good agreement with the predicted profile, although the estimate is smaller than the prediction at heights above 10 hPa, as discussed above.

5.3. HNO_3

[59] HNO_3 provides an interesting test case as the scatterplots in Figure 5 are markedly sloped, signaling the definite presence of multiplicative bias. The slopes estimated via methods 2 and 3 (Figure 5a) peak in magnitude at 21 hPa, where the absolute natural variance is largest.

[60] By erroneously ignoring multiplicative bias, method 1 leads to negative error variance estimates for ACE (Figure 5b) and simultaneously, MLS error SD estimates (Figure 5c) up to five times the predicted error SD values over the vertical measurement range.

[61] Method 2 leads to error SD estimates generally more realistic than method 1. The method 2 MLS error SD estimate (Figure 5c) is smaller than the prediction above 10 hPa, which is again consistent with the simulation results discussed by *Livesey et al.* [2005b], wherein HNO_3 scatter was found to be smaller than the predicted variance due to retrieval smoothing at these heights. Using predicted error variances for MLS that are greater than the observed measurement variances in method 2 again leads to anomalously small $\hat{\beta}_2|_{\tilde{\sigma}^2_\epsilon}$ estimates (Figure 5a) and correspondingly negative error variance estimates for ACE between 10 and 4 hPa.

[62] Figure 5d shows the estimated and predicted error SD profiles in percent of the mean measured profile. Between 10 and 70 hPa, method 3 error SD estimates for ACE are roughly constant with height, and are approximately 10% (of the mean measurement value) larger than the predictions. The

method 3 error SD estimates for MLS are within ±5% of the predicted values between 10 and 147 hPa. While the predicted ACE error SD profile is a factor of 10–25% smaller than that for MLS, we find the estimated percent-error SD profiles are of comparable magnitude between 10 and 100 hPa. Only at high (above 10 hPa) and very low (146 hPa) heights does the comparison confirm the prediction of greater precision for the ACE measurements, as shown by the separation of the estimated percent-error profiles.

[63] In contrast with the error SD estimates for N_2O and O_3 (below heights of 10 hPa), the secondary coincidence error SD ($\hat{\sigma}_{\nu 3}$) profile lies outside the 95% confidence interval of the $\hat{\sigma}_{\epsilon 3}$ estimate. This means that the noncoincidence variance σ_η^2 is significant in this case, which suggests a tighter coincidence criterion may be required for HNO_3 compared to the other species.

6. Conclusions

[64] In this work we have presented a framework for the comparison of coincident measurement sets which distinguishes between additive and multiplicative bias. We find multiplicative bias is a significant source of differences between coincident measurements of O_3, N_2O, and HNO_3 by the satellite instruments ACE and MLS.

[65] Differentiation between bias types promises to aid the identification of the underlying source of inconsistencies between measurement systems. Multiplicative bias is significant at some, but not all heights for O_3 and N_2O. This suggests that the source of the multiplicative bias within the respective measurement systems is due to height-dependent retrieval processes or parameters. Conversely, the multiplicative bias estimated for HNO_3 is significant for all heights tested. In this case, it seems likely that the source of the multiplicative bias is some aspect(s) of the measurement systems affecting all heights, such as spectroscopic parameters. Conceivably, future validation work could proceed by producing multiplicative and additive bias profiles through comparison to a standard instrument, and comparing these to profiles of the sensitivity of the retrieval system to the known component error sources produced through a forward modeling simulation. In this way the underlying source of the bias might be uniquely identified.

[66] We have shown that multiplicative bias and measurement error are analytically coupled: any estimated or prescribed value of multiplicative bias leads directly to error variance values for the two instruments.

[67] The existence of significant multiplicative bias leads to unphysically negative estimates of measurement error variances when an estimation technique that assumes no multiplicative bias (method 1) is used. Consequently, multiplicative bias between two data sets should be thoroughly proven negligible if such a method is to be used to estimate measurement error variances.

[68] Knowledge of the measurement errors can in theory be used to estimate the multiplicative bias (method 2). We find though that the measurement errors reported with the measurement data used here are generally unsatisfactory for this purpose. MLS reported errors are in some cases dominated by smoothing error, rather than the strictly random error assumed in our model. ACE reported errors are understood to be lower limits based on the spectral fitting residuals, and do not take into account error introduced by the interpolation to a standard 1-km grid, or the subsequent interpolation to the MLS pressure grid.

[69] We have presented an analysis technique which does not rely upon assumptions concerning the magnitude of the multiplicative bias or the error variances. The ''instrument variable'' technique (method 3) relies upon a third variable (such as the secondary pseudocoincident measurement used here) which is correlated with the two measurements. We find that the instrument variable method leads to the most realistic error variance profiles for all three species studied. Qualitatively allowing for discrepancies at heights for which smoothing error is significant in the MLS error SD prediction, MLS error SD estimates are generally in good agreement with the predictions. For O_3, at heights below 10 hPa, a significant and unexplained difference between MLS error SD estimates and predictions is seen: the estimates are roughly 5–15% greater than the predictions. HNO_3 error SD estimates are also on the order of 5% greater than the predictions for MLS at heights below 20 hPa. For ACE, the estimated error SD profiles are in almost all cases larger than the predictions. Estimated ACE error SDs are less than or approximately equal to 10% of the mean measurement value: for O_3 between 100 and 2 hPa, for N_2O between 100 ad 10 hPa, and for HNO_3 between 30 and 10 hPa. ACE error SDs are significantly smaller than those for MLS at the highest and lowest heights compared for HNO_3, and over most of the vertical range for N_2O. On the other hand, O_3 error SDs are comparable between the two instruments.

[70] We propose that the instrument variable method of statistical comparison of coincident observations shows promise for future comparison and validation exercises, as it leads to an assumption-free estimate of multiplicative bias, and hence allows the direct estimation of random measurement errors.

[71] **Acknowledgments.** We thank the ACE and MLS teams for making their measurement data available for use. For helpful comments and discussions we thank P. Bernath, C. Boone, V. Fioletov, A. Fraser, T. Kerzenmacher, N. Livesey, J. Taylor, and K. Walker. In addition to the works cited, we acknowledge the helpful guidance of an unpublished manuscript by L. Flynn. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT-1, is a Canadian led mission mainly supported by the Canadian Space Agency (CSA) and the Natural Sciences and Engineering Research Council (NSERC) of Canada. MLS is funded by the National Aeronautics and Space Administration. This work was carried out with the aid of a grant from the CSA. M. T. gratefully acknowledges the scholarship support of NSERC and the CSA.

References

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.

Dunn, G. (1989), *Design and Analysis of Reliability Studies: The Statistical Evaluation of Measurement Errors*, Oxford Univ. Press, New York.

Efron, B., and R. J. Tibshirani (1994), An Introduction to the Bootstrap. CRC Press, Boca Raton, Fla.

Fioletov, V. E., D. W. Tarasick, and I. Petropavlovskikh (2006), Estimating ozone variability and instrument uncertainties from SBUV (/2), ozonesonde, Umkehr, and SAGE II measurements: Short-term variations, *J. Geophys. Res.*, *111*, D02305, doi:10.1029/2005JD006340.

Froidevaux, L., et al. (2006), Early validation analyses of atmospheric profiles from EOS MLS on the Aura satellite, *IEEE Trans. Geosci. Remote Sens.*, *44*(6), 1106–1121, doi:10.1109/TGRS.2006.864366.

Grubbs, F. E. (1948), On estimating precision of measuring instruments and product variability, *J. Am. Stat. Assoc.*, *43*(242), 243–264.

Grubbs, F. E. (1973), Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments, *Technometrics*, *15*(1), 53–66.

Hocking, W. K., T. Thayaparan, and S. J. Franke (2001), Method for statistical comparison of geophysical data by multiple instruments which have differing accuracies, *Adv. Space Res.*, *27*(6), 1089–1098, doi:10.1016/S0273-1177 (01)00143-0.

Livesey, N. J., W. V. Snyder, W. G. Read, and P. A. Wagner (2005a), Retrieval algorithms for the EOS Microwave Limb Sounder (MLS), *IEEE Trans. Geosci. Remote Sens.*, 1144–1155, doi:10.1109/TGRS.2006.872327.

Livesey, N. J., et al. (2005b), Version 1.5 Level 2 data quality and description document, Tech. Rep. D-32381, JPL.

Rodgers, C. D. (1990), Characterization and error analysis of profiles retrieved from remote sounding measurements, *J. Geophys. Res.*, *95*(D5), 5587–5595.

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

von Clarmann, T. (2006), Validation of remotely sensed profiles of atmospheric state variables: Strategies and terminology, *Atmos. Chem. Phys. Disc.*, *6*, 4973–4994.

Walker, K. A., C. E. Randall, C. R. Trepte, C. D. Boone, and P. F. Bernath (2005), Initial validation comparisons for the Atmospheric Chemistry Experiment (ACE-FTS), *Geophys. Res. Lett.*, *32*, L16S04, doi:10.1029/2005GL022388.

Waters, J. W., et al. (2006), The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1075–1092, doi:10.1109/TGRS.2006.873771.

K. Strong and M. Toohey, Department of Physics, University of Toronto, Toronto, ON, Canada. (mtoohey@atmosp.physics.utoronto.ca)