Estimating the impact of small-scale variability in satellite measurement validation

L. C. Sparling, J. C. Wei, and L. M. Avallone

Received 1 December 2005; revised 12 March 2006; accepted 27 June 2006; published 27 October 2006.

[1] The necessity of validating satellite measurements of atmospheric chemical constituents with supplementary in situ measurements leads to problems with interpretation of the inevitable differences that arise because of measurement resolution or imperfect collocation, especially for highly heterogeneous fields. In this paper the contribution of small-scale structure to measurement differences is estimated from high-resolution aircraft measurements of atmospheric trace gases. The analysis uses the statistics of fractional differences in mixing ratio across a range of scales to estimate the contribution of real variability to differences in noncollocated measurements. The differences depend on the particular chemical tracer, location and season. We find a range of behavior: Differences of 50% across horizontal scales 100 km or less are fairly common for tropospheric water vapor under convective conditions, or carbon monoxide in regions influenced by biomass burning. Ozone varies by about 4–12% in the lower stratosphere and 15–25% in the upper/middle troposphere across scales of about 150 km. The effect of coarse satellite measurement resolution is also estimated by comparing point measurements to locally averaged measurements and is found to reduce the occurrence of large tracer differences. The choice of coincidence lengths should be based on both the scale-dependent variability and a priori estimates of satellite accuracy if real variability is to remain small relative to satellite measurement uncertainties. For satellite instrument uncertainties of about 10%, coincidence lengths for ozone should be less than 50 km in the upper troposphere and less than 100 km in the lower stratosphere.


I. Introduction

[2] The detection of climate change and its dependence on the chemical composition of the atmosphere requires long-term global monitoring with satellite-borne instruments. Measurements of the chemical composition of the atmosphere from space are derived from direct radiance measurements which are converted to chemical concentrations via retrieval algorithms. Satellite observations must be compared to independent measurements obtained with accurately calibrated instruments on aircraft, balloon or ground-based platforms (“correlative measurements”) soon after launch and periodically thereafter to ensure that any instrument drift or bias, which can lead to spurious trends in the state of the atmosphere, is well understood and can be corrected for. These ongoing calibration and validation activities constitute a crucial part of the evaluation of the global earth system database, and are essential in order to quantify uncertainties in the climate change signal [see, for example, King, 1999].

[3] In situ observations of chemical tracers in the UT (Upper Troposphere) and UT/LS (Upper Troposphere/Lower Stratosphere) show a high degree of spatial variability at scales that are well below the resolution of satellite measurements. This small-scale variability is problematic in that it contributes to differences between satellite measurements and nearby (but not strictly collocated) in situ or ground-based measurements obtained for the purpose of evaluating the quality of the satellite measurements. For purposes of direct comparison, correlative measurements should ideally sample the atmosphere at the exact times and locations of the satellite measurement, but this is not always possible and the number of such direct comparisons is typically small. Coincidence criteria must be relaxed to some degree, and this leads to the recurring problem of interpreting the inevitable differences in observations that are not exactly collocated. In addition, in situ correlative measurements are point measurements which must be compared to lower-resolution satellite observations leading to further uncertainties and problems of interpretation. The goal of validation is to understand measurement uncertainties related to viewing geometry, retrieval algorithms or instrumentation, but these can be overwhelmed by natural variability when the tracer field is highly heterogeneous.
[4] The strong gradients in constituent concentration fields pose special problems for retrievals from satellite instruments and other remote sensors, whose resolutions are low compared to the scale of the atmospheric structures and whose algorithms typically assume homogeneity along the line-of-sight. The satellite instruments currently measuring ozone profiles in the UT/LS (for example, EOS (Earth Observing System) MLS (Microwave Limb Sounder), GOME (Global Ozone Monitoring Experiment), SAGE III (Stratospheric Aerosol and Gas Experiment III), POAM III (Polar Ozone and Aerosol Measurement)) have horizontal resolutions ranging from 5 to 960 km and vertical resolutions of about 1 to 5 km [Hoogen et al., 1999; Lucke et al., 1999; Waters et al., 1999, 2006]. While coarse compared to observations obtained from aircraft, these are considerably finer than earlier generations of satellite instruments. As the horizontal and vertical resolutions of remote sensors have improved, the ability to account for and/or retrieve information about line-of-sight gradients has also improved [Worden et al., 2004; Livesey and Read, 2000]. However, constituent variations along the line-of-sight are still not adequately understood, nor are the newly developed methods validated against data of appropriate resolution. We acknowledge that the phrase “satellite resolution” is not generic; that with the widely different types of remote sensors in orbit, “resolution” can have many meanings. Spatial resolution will depend on satellite orbit, as well as on instrument field of view and even on spectral range. In addition, most remote sensors have different spatial resolutions in the two horizontal dimensions (e.g., along-track versus across-track); EOS MLS, for example, samples approximately 5 km across-track and 160 km along-track [Waters et al., 2006]. These differences among various types of satellite measurements make it difficult to say in general what the impact of small-scale variability will be on validation, since it will depend on the particular measurement under consideration. Another challenge to defining “coincidence criteria” is that the appropriate coincidence lengths (see section 3) will also depend on the expected uncertainty of the satellite measurement. It is important that natural variability on the scale of the coincidence length is less than the satellite measurement uncertainties, otherwise little is gained by the validation exercise. For the EOS instruments, 10% seems to be a typical value for satellite precision [Schoeberl, 2004] and is on the order of the estimated accuracy of several satellite-based trace gas measurements of ozone profiles such as SAGE III (5–10% [Brogan et al., 2002]) GOME (10% [Hoogen et al., 1999]), POAM III (5% [Randall et al., 2003]), and anticipated from the recently launched EOS Aura instruments MLS [Waters et al., 2006] and OMI. We will therefore use a nominal value of 10% to illustrate some of the points we make later.

[5] A number of validation strategies that incorporate modeling have been developed in an attempt to overcome the difficulties with noncoincident validation, including data assimilation [e.g., Lary et al., 2003] and Lagrangian modeling techniques. The latter use trajectories to accumulate satellite measurements over several days or to propagate them to new locations and times to increase the number of coincidences [e.g., Morris et al., 2000; Lu et al., 2000]. Others use Lagrangian air mass sampling to limit comparisons of in situ and satellite measurements to regions where the air masses have similar origins [Schwab et al., 1996] or to sample multiple times along the flow as in the Match technique [Rex et al., 1999]. Other approaches take the quasi-isentropic nature of stratospheric transport into account by using potential temperature and potential vorticity (PV) as flow-tracking coordinates, which may increase coincidences to the extent that it allows comparison of measurements separated by larger distances but with a common recent history [Schoeberl and Lait, 1991; Redaelli et al., 1994; Lary et al., 1995]. However, the advantage of these methods is not always clear [Morris et al., 2004; Lait et al., 2004], as they are not applicable at low latitudes, and the PV/tracer correlation is often too weak to be useful. While providing some additional assessments of satellite measurement quality, these different approaches to validation introduce some measure of uncertainty due to the meteorological analyses or Lagrangian computations themselves and so will likely continue to be used in conjunction with direct comparisons.

[6] In regions such as the LS polar vortex edge or the UT/LS, assessing the impact of small-scale variability in tracer fields is particularly difficult because these are transition zones between chemically distinct air masses that are being stirred together by the winds. Stratosphere-troposphere exchange (STE) dynamical processes such as tropopause folds blend low water vapor/high ozone stratospheric air with high water/low ozone tropospheric air resulting in thin vertical layers and horizontal filaments [Orsolini et al., 1995; Wu et al., 1997]. Stretching and folding by the winds cascades large-scale tracer structure to small scales, resulting in a complex multiscale interleaving of air with different chemical properties as demonstrated in many numerical simulations and observed in satellite imagery [for example, Sutton et al., 1994; Appenzeller et al., 1996]. Certainly the concept of “sampling the same air mass” becomes muddy in regions where the boundary between air masses is highly convoluted over a wide range of scales.

[7] The goal of the work presented here is to provide some perspectives on statistical properties of small-scale variability in trace gases on length scales that are relevant for satellite data validation, and to provide some quantitative estimates of the contribution of natural variability to differences in correlative measurements under a range of dynamical conditions. We begin in section 2 with a discussion of data sources and statistical methods, and define statistical ensembles with respect to region, season and prevailing meteorology. It is impossible to consider all chemical species under all possible conditions, so this analysis emphasizes LS, UT/LS, and UT ozone, with some additional examples of variability in tropospheric water vapor and carbon monoxide (CO) under different source and meteorological conditions. Section 3 presents an overview of the variability in LS ozone across a range of scales as determined by a statistical analysis of the difference between pairs of measurements separated by a specified horizontal and vertical distance. In this section we also quantify the likelihood of observing differences that are larger than a given value, investigate the isotropy of the spatial variability to see whether coincidence criteria that differ for zonal and latitudinal separations are warranted, and address the problem of defining suitable coincidence
criteria. The impact of satellite resolutions is also estimated. A similar analysis is applied to the DC-8 ozone data in section 4, with a brief analysis of CO and water vapor for comparison. In section 5 we show that the statistics of spatial variability have a degree of universality that seems to hold (approximately) over a wide range of conditions. We conclude in section 6 with a summary and discussion.

2. Data and Methods

2.1. Aircraft Data

The emphasis in this study is on small-scale horizontal structure for which the high-resolution (2 km) and high-precision aircraft measurements are well suited. There is less information about vertical structure per se because the moving aircraft platform does not sample solely in the vertical direction over a wide enough range of scales for the type of scale analysis we wish to do here. Nevertheless, some aspects of vertical variability must be taken into account in the horizontal structure analysis as described below. The aircraft data used in this study were collected during several NASA-sponsored measurement campaigns over a period of about 15 years and are summarized in Table 1, which also includes some general remarks about sources and dynamical conditions. The LS analysis is based on ozone measurements taken with the dual-beam ultraviolet absorption instrument [Proffitt et al., 1989; Proffitt and McLaughlin, 1993] with a precision of about 0.5% (E. L. Richard, personal communication, 2005) on the NASA ER-2 aircraft at altitudes in the range 18–21 km. Because the objective of the winter ER-2 missions was to sample inside/outside the polar vortex, the high-latitude flight paths were overwhelmingly across the polar vortex jet, which could introduce a sampling bias. The analysis of anisotropy in a later section suggests that the impact of horizontal anisotropy, in the sense of N/S (North/South) versus E/W (East/West), is small on typical coincidence scales.

Results for the LS are presented for four statistical ensembles that encompass a range of variability in LS ozone: (1) the tropics (25°S–25°N); (2) the summer Northern Hemisphere (NH, May–November) and summer Southern Hemisphere (SH, December–May) extratropics; (3) the winter NH middle/high latitudes, outside the polar vortex and poleward of 30°; and (4) the interior of the NH/SH winter polar vortices. We define the interior of the polar vortex as the high-latitude winter region with potential vorticity >35 pvu (1 potential vorticity unit = 1.0 × 10⁻⁶ m² s⁻¹ K kg⁻¹), a value typical of the edge of the polar vortex at ER-2 altitudes.

The tropospheric analysis is based primarily on DC-8 ozone measurements as described by Gregory et al. [1987] which have a 1-s precision of 1%. The water vapor [Vay et al., 1998] and carbon monoxide (CO) [Sachse et al., 1987] measurements have 1-s precisions of 2% and 1%, respectively. We use the 10-s averaged data for which the precision should be even better. The high precision of the aircraft data is particularly important for this study which is based on the difference of pairs of measurements.

The analysis of the DC-8 data is restricted to the vertical region from 8 to 13 km, which has been more extensively sampled relative to the lower troposphere in this database. This vertical range corresponds to the middle/upper troposphere at lower latitudes and the UT/LS at middle/high latitudes. Results will be presented for the individual aircraft campaigns because the overall meteorological conditions and source variability for the tropospheric data tend to organize themselves according to the goals of the particular campaign. The emphasis is again on ozone, but for comparison we also show some results for CO in regions influenced by biomass burning and continental pollution. Tropospheric ozone and CO are also of interest in the application of remote sensing data to issues related to air quality, long-range transport in the middle/upper troposphere, and mixing and transport of chemical tracers in the tropopause region. In addition, some statistics are also presented for water vapor because it represents an upper bound on the variability of atmospheric trace gases, especially under conditions of strong convection.

We make the standard assumption that the statistics, if not the exact morphological details, of small-scale spatial structure based on aircraft measurements collocated to within 30 min or so are the same as if the measurements were made instantaneously. This “frozen turbulence” assumption is reasonable if a time span of 30 min is temporarily collocated as far as the statistics of two-point variability is concerned; this is plausible even if the spatial configuration of the tracer field evolves to some degree on this timescale.

2.2. Statistical Methods

The statistical analysis is based on the computation of increments in the tracer mixing ratio across a given distance. Let \( \chi(s,z) \) be a tracer measurement at horizontal location \( s \) and height \( z \) and \( \chi(s + r, z + \delta z) \) be another measurement separated from the first by a horizontal distance \( r \) and a vertical distance \( h = |\delta z| \). Scale-dependent increments in the tracer field are defined by the absolute value of the fractional difference (in percent) of the two measurements

\[
\Delta_{r,h} = \frac{\chi(s + r, z + h) - \chi(s, z)}{\frac{1}{2}(\chi(s + r, z + h) + \chi(s, z))} \times 100
\]

The emphasis in this paper on fractional rather than absolute differences enables us to synthesize results across a wide range of conditions and widely varying tracer concentrations. The horizontal separation \( r \), vertical separation \( h \), and tracer increment \( \Delta_{r,h} \) are computed for each pair. Taking into account the vertical separation allows us to isolate the contribution from horizontal structure alone, which is our main concern here; the aircraft data set does not allow a statistically robust analysis of vertical profiles for which an extensive sonde data set is more appropriate. Tracer differences due to horizontal structure alone will be inferred from the limit \( \Delta_r = \lim_{h \to 0} \Delta_{r,h} \), i.e., from the limiting statistical behavior of the tracer increments as the vertical separation becomes very small. The two-point statistics of “horizontal” structure reported in this paper are derived from pairs of measurements whose vertical separation is small in this sense. For the cases studied here, we have found that \( \Delta_{r,h} \) is nearly independent of \( h \) for a sampling aspect ratio \( h):r < 1:500 \) and \( \mu_r = \langle \Delta_r \rangle \) (note that these are the means of the absolute
Table 1. Summary of Aircraft Data Used in This Study

<table>
<thead>
<tr>
<th>DC8 Mission</th>
<th>Time</th>
<th>Region/Conditions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AASE2</td>
<td>Oct-Dec 1991, Jan-Mar 1992</td>
<td>NH mid-high latitude, polar vortex, surf zone, STE</td>
<td>Anderson and Toon [1993], <a href="http://cloud1.arc.nasa.gov/aase2/">http://cloud1.arc.nasa.gov/aase2/</a></td>
</tr>
<tr>
<td>PEM-West B</td>
<td>Feb-Mar 1994</td>
<td>NW Pacific, Asian continent outflow, convection</td>
<td>Talbot et al. [1997], <a href="http://www.gte.larc.nasa.gov/gte_fld.htm#PEM">http://www.gte.larc.nasa.gov/gte_fld.htm#PEM</a></td>
</tr>
<tr>
<td>SUCCESS</td>
<td>Apr-May 1996</td>
<td>midwest USA, continental, convection, subsonic aircraft emissions</td>
<td><a href="http://cloud1.arc.nasa.gov/success">http://cloud1.arc.nasa.gov/success</a></td>
</tr>
<tr>
<td>TRACE-P</td>
<td>Feb-Apr 2001</td>
<td>western Pacific, Asian outflow, pollution</td>
<td>Jacob et al. [2003], <a href="http://www.gte.larc.nasa.gov/gte_fld.htm#TRACE">http://www.gte.larc.nasa.gov/gte_fld.htm#TRACE</a></td>
</tr>
<tr>
<td>SOLVE</td>
<td>Nov 1999, Jan-Mar 2000</td>
<td>Arctic high latitude, Arctic polar vortex, STE</td>
<td>Newman et al. [2002], <a href="http://cloud1.arc.nasa.gov/solve/">http://cloud1.arc.nasa.gov/solve/</a></td>
</tr>
<tr>
<td>SOLVE2</td>
<td>Dec 2002, Jan-Feb 2003</td>
<td>Arctic high latitude, Arctic polar vortex, surf zone variability, STE</td>
<td><a href="http://cloud1.arc.nasa.gov/solveII/">http://cloud1.arc.nasa.gov/solveII/</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ER2 Mission</th>
<th>Time</th>
<th>Region/Conditions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAOE</td>
<td>Aug-Sep 1987</td>
<td>SH LS high latitude, polar vortex</td>
<td>Tuck et al. [1989], <a href="http://cloud1.arc.nasa.gov/aaoe/">http://cloud1.arc.nasa.gov/aaoe/</a></td>
</tr>
<tr>
<td>AASE1</td>
<td>Dec 1988, Jan-Feb 1989</td>
<td>NH LS mid/high latitude, polar vortex</td>
<td>Turco et al. [1990], <a href="http://cloud1.arc.nasa.gov/aae/">http://cloud1.arc.nasa.gov/aae/</a></td>
</tr>
<tr>
<td>ASHOE/MAESA</td>
<td>Mar-Nov 1994</td>
<td>SH LS tropics-polar, western Pacific, mixing and transport, filaments, STE</td>
<td>Tuck et al. [1997], <a href="http://cloud1.arc.nasa.gov/ashoe_maesa/">http://cloud1.arc.nasa.gov/ashoe_maesa/</a></td>
</tr>
<tr>
<td>SPADE</td>
<td>Nov 1992, May 1993</td>
<td>LS transport/chemistry</td>
<td>Wofsy et al. [1994], <a href="http://cloud1.arc.nasa.gov/spade">http://cloud1.arc.nasa.gov/spade</a></td>
</tr>
<tr>
<td>STRAT</td>
<td>May 1995 to Feb 1996</td>
<td>NH LS tropics-mid latitude transport, STE</td>
<td><a href="http://cloud1.arc.nasa.gov/strat/">http://cloud1.arc.nasa.gov/strat/</a></td>
</tr>
<tr>
<td>POLARIS</td>
<td>Apr-Jul and Sep 1997</td>
<td>NH mid/high latitude, transport</td>
<td>Newman et al. [1999], <a href="http://cloud1.arc.nasa.gov/polaris">http://cloud1.arc.nasa.gov/polaris</a></td>
</tr>
<tr>
<td>SOLVE</td>
<td>Nov 1999, Jan-Mar 2000</td>
<td>NH mid/high latitude, polar vortex</td>
<td>Newman et al. [2002], <a href="http://cloud1.arc.nasa.gov/solve/">http://cloud1.arc.nasa.gov/solve/</a></td>
</tr>
<tr>
<td>SOLVE2</td>
<td>Dec 2002, Jan-Feb 2003</td>
<td>NH mid/high latitude, polar vortex</td>
<td><a href="http://cloud1.arc.nasa.gov/solveII/">http://cloud1.arc.nasa.gov/solveII/</a></td>
</tr>
</tbody>
</table>

*ER2 altitudes 18–21 km; DC8, 8–13 km. MT, middle troposphere; UT, upper troposphere; LS, lower stratosphere; STE, stratosphere-troposphere exchange; LRT, long-range transport.
value of the tracer difference), and the standard deviation
\[ \sigma_r = \sqrt{\langle \Delta_r^2 \rangle} \]
of the signed (i.e., not absolute value) tracer difference which has zero mean. For the purposes of data validation, and partly because of sampling characteristics, our analysis is restricted to horizontal scales less than 400 km in the LS and 200 km in the UT and UT/LS.

In addition to the two-point statistics, we consider the satellite measurement resolution in an approximate way by averaging over a segment of the in situ data and computing the statistics of the difference between the center of the averaged measurement and an in situ measurement some distance away. In the following, we will refer to these local averages as “satellite” measurements. Each such segment has a mean location \((\langle x \rangle, \langle y \rangle)\), and we define the satellite “footprint” \(L\) as the average distance from the mean location:
\[ L = \langle (x - \langle x \rangle)^2 + (y - \langle y \rangle)^2 \rangle^{1/2} \]
which is a measure of the horizontal scale over which the time average takes place. When the flight paths have long straight-line segments, a time average over an interval \(\Delta t\) for a constant aircraft speed \(v\) corresponds to an average over a horizontal distance \(r = v\Delta t\). This is not true in general, particularly for the more complicated flight paths in some of the DC-8 data. By varying the time-averaging window, we generate a range of values for \(L\), and \(L\)-dependent statistics are obtained by conditioning on \(L\). The mean tracer mixing ratio in the segment is \(\langle \chi \rangle_L\), and the tracer difference now depends on \(L\),
\[ \Delta \chi_L = \left\| \frac{\chi(s + r, z + h) - \langle \chi(s, z) \rangle_L}{\langle \chi(s + r, z + h) + \langle \chi(s, z) \rangle_L \rangle} \right\| \times 100 \]  
\[ \]  
[15] The horizontal separation is now defined to be the distance from the point measurement to the center of the satellite measurement and we will be concerned only with the mean \(\mu_r = \langle \Delta \chi_r \rangle\) for sufficiently small \(h\). Note that for nonzero \(L\), the limit \(r \to 0\) describes the deviation of a point measurement from the average in a surrounding neighborhood of scale \(L\), and therefore is a measure of the degree of local representativeness.

[16] For certain questions it is necessary to go beyond the first few moments and consider the entire distribution of increments. The PDF (Probability Distribution Function) \(P(\Delta_r)\) [see, for example, Sparling and Bacmeister, 2001; Hu and Pierrehumbert, 2001] is the fraction of increments in a bin centered on \(\Delta_r\). An important question for satellite data validation is how often geophysical variability would be expected to overwhelm instrumental or retrieval uncertainties, whose estimation is the purpose of the validation exercise. Insight can be gained by considering the cumulative distribution function (CDF) \(C_r(d) = \int_{-\infty}^{d} \! P(\Delta_r) \, d\Delta_r\). \(C_r(d)\) is the fraction of measurement pairs, separated by a given value of \(r\), for which \(\Delta_r\) is greater than a specified value \(d\). We will also consider the median \(\mu_r\), defined by \(C_r(\mu_r) = 1/2\), and because \(P(\Delta_r)\) is normalized, we also have \(C_r(0) = 1\). The corresponding CDF of the satellite/in situ differences will be denoted by \(C_r(d)\).

3. Horizontal Variability of LS Ozone

[17] In this section we present an overview of the spatial variability of ozone in the LS by looking at how the fractional difference of two measurements depends on their horizontal and vertical separation. It is important to separate the contributions from vertical and horizontal structure because the atmosphere is not isotropic in three dimensions, to establish horizontal coincidence criteria where possible, and to interpret differences between satellite instruments with different horizontal and vertical resolutions.

[18] In order to characterize the statistical properties of horizontal structure, we must first remove the contribution from the large-scale vertical gradient or strong vertical layering. One way to do this is to look at the behavior of \(\mu_r\) as a function of vertical separation \(h\) for fixed horizontal separation \(r\). The example in Figure 1 shows that \(\mu_r\) decreases with decreasing \(h\) and as discussed in section 2, asymptotes to the limiting value \(\mu_h\) as \(h \to 0\). For this example, \(r = 100\) km and across this scale \(\mu_h\) ranges from 4–5% inside the vortex to 8–10% in the NH high latitudes. Figure 1 also points out that horizontal collocation and vertical resolution cannot be considered independently. This is especially true for the tropical curve which indicates that when \(r = 100\) km, the large vertical gradient in the tropics overwhelms differences due to horizontal structure if the vertical separation exceeds about 200 m. In such cases, the interpretation of differences between satellite and in situ measurements will depend less on the degree of horizontal collocation and more on the vertical resolution/weighting function of the satellite measurement.

[19] Before discussing the dependence of the tracer increments on horizontal scale, it is necessary to consider the shape of the PDF, as it has some bearing on the way in which the scale dependence is best characterized. The stretching and folding of atmospheric tracers leads to a distribution that is non-Gaussian; an example can be seen in Figure 2 which shows \(P(\Delta_r)\) for the NH winter ensemble for pairs of measurements with a horizontal separation \(r = 50\) km (thick solid line). Here we have retained the sign of the tracer difference in order to highlight the non-Gaussian shape of the PDF (note that \(\Delta_r\) was defined as an absolute value). The generic long-tailed shape is characteristic of filamentary structure of long-lived tracers in
which chemically dissimilar air masses are stirred together by the large-scale winds. Indicated by vertical lines are \( \bar{\mu}_r, \bar{\mu}_s, \), and \( \sigma_r \). The shaded curve is a Gaussian with a standard deviation of 3% which is a reasonable fit to the distribution near its maximum; in this case the median of the full distribution is close to the standard deviation of the Gaussian core. Also overlapped (thin solid line) is a Gaussian with variance \( \sigma_r^2 \), clearly a poor description of \( P(\Delta r) \) in that it overemphasizes the large increments. We do not necessarily expect the distribution of the fractional difference to be Gaussian, even if the distribution of the differences themselves is Gaussian, although this could occur in well-mixed regions where the mean tracer concentration is much larger than the standard deviation. The point here is just to show that while the variance of a short-tailed distribution such as a Gaussian is a reasonable measure of typical variability, it may not be so when the distribution has long tails.

[20] It is worth noting that long tails are also observed for the absolute (not fractional) tracer differences as shown for example by Sparling and Bacmeister [2001] and theoretically by Hu and Pierrehumbert [2001]. The tails do not arise here because of a small denominator (e.g., in the case of two independent Gaussian random variables whose ratio has a long-tailed Cauchy distribution). This is easy to see from the definition of \( \Delta r \), as a quantity proportional to the difference of two positive numbers divided by their sum. The largest possible value of the fractional difference defined in equations (1) and (2) is 200%, which occurs in the limit when the two values are very different.

[21] Figure 3 compares \( \bar{\mu}_r, \mu_r, \sigma_r \) versus horizontal separation \( r \) for the different ensembles, and shows that \( \sigma_r \) is substantially larger than \( \mu_r \), especially at small scales. For example, at \( r = 50 \text{ km} \), \( \mu_r/\sigma_r \approx 0.5 \), while for a zero-mean Gaussian random variable the two measures are more similar, \( \mu_r/\sigma_r \approx 0.8 \). These distinctions are especially important for long-tailed distributions, because they can lead to different conclusions about the degree to which geophysical variability contributes to differences in observations. For example, with the exception of NH winter, the median in Figure 3a suggests that real variability for coincidence scales up to 300 km would be less than a nominal satellite accuracy of 10%, while the standard deviation in Figure 3c suggests that coincidence lengths should be well below 100 km if natural variability is to remain less than 10%. With the exception of NH winter, the mean deviation suggests that coincidence lengths should be less than about 200 km. The median may underestimate the impact of small scales on validation, while the standard deviation overestimates it and, in addition, may be strongly affected by undersampling. As a compromise, we will focus on the mean \( \mu_r \) in most of the following analysis.

[22] While a proper choice of coincidence length is unfortunately dependent on an a priori estimate of the uncertainty of the satellite measurement, an upper bound can sometimes be established if the tracer field has a well-defined correlation length. Under conditions of statistical homogeneity, a correlation length is defined as the scale \( r \) beyond which the autocorrelation \( \langle \chi(s + r)\chi(s) \rangle \) vanishes. Assuming homogeneity, we then have \( \langle \chi(s + r) - \chi(s) \rangle^2 = 2\langle \chi^2 \rangle - 2\langle \chi(s)\chi(s) \rangle \approx 2\langle \chi^2 \rangle \) which is independent of \( r \) when \( r \) exceeds the correlation length. Two measurements

---

**Figure 2.** PDF of tracer increments across horizontal scale \( r = 50 \text{ km} \) for NH winter LS ozone (thick solid line). The median is 3% and is characteristic of the width of the Gaussian core (thick shaded line). The mean \( \mu_r \) is 6%, and the standard deviation \( \sigma_r \) is 10%. The thin solid line is a Gaussian with variance \( \sigma_r^2 \) which can be seen to be a poor representation of the PDF.

**Figure 3.** Comparison of single-parameter measures of horizontal variability in LS ozone and their dependence on horizontal scale \( r \). (a) Median, (b) mean \( \mu_r \), (c) \( \sigma_r \), and (d) mean \( \mu_r \) as in Figure 3b but on a log scale showing that

\[ \mu_r \sim r^{1/2} \text{ for } r \geq 10 \text{ km}. \]
separated by a distance larger than this are uncorrelated and cannot reasonably be compared. More generally, an upper bound for the coincidence length is the scale beyond which the increment statistics become independent of scale. Figure 3 suggests that this scale is 300–400 km, in agreement with Schoeberl et al. [2002], except for the NH winter ensemble where the mean increment shows no sign of leveling off as the separation distance increases. This occurs because the increments become dominated by the large-scale latitudinal gradient in ozone during this time of year.

Figure 3d shows the same data as in Figure 3b, except plotted on a log scale to illustrate a scaling regime described to a good approximation by the power law $\mu_r \sim r^p$ where $p \sim 0.5$; the precise value of the exponent is of no concern to us here [see Tuck and Hovde, 1999]. Because $p < 1$, the mean increment is more sensitive to horizontal separation at the smaller scales. In NH winter, the power law extends to at least several hundred km, with no obvious upper bound on the coincidence length over this range of scales. In general, and for the scale-invariant case in particular, a coincidence length should be chosen so that the geophysical variability does not dominate the expected accuracy of the satellite measurement.

At the smallest scales, the tracer increments become more dependent on instrument precision. For $r = 2$ km, Figure 3d shows that $\mu_r$ inside the polar vortex is on the order of the precision of the ozone measurement. At larger scales the measurement precision is small relative to the real geophysical variability.

Figure 4 presents a summary of the latitudinal and seasonal dependence of $\mu_r$ for horizontal scales $r = 50$ and $r = 200$ km. There are some gaps in coverage, for example the summer SH is not well represented in this database. The mean increments are large in winter and early spring at high latitudes, most likely because of the breakup of the polar vortex. Low values occur throughout the NH in summer and fall. Overall, $\mu_r$ ranges from 1–10% for $r = 50$ km and 2–17% for $r = 200$ km. Figure 4 (bottom) shows that, for an estimated satellite uncertainty of 10%, coincidence lengths up to 200 km may be appropriate in NH summer and fall, but are too large for validation during NH winter and spring.

### 3.1. Isotropy and Zonal Versus Meridional Coincidence Criteria

It is common for coincidence criteria to be defined differently in the zonal and meridional directions. To see if this is justified it is necessary to investigate whether the tracer increments depend on the orientation of the separation between the two measurements. Long-lived LS tracers often have large-scale meridional gradients, but the issue here is whether the spatial structure in LS ozone is horizontally anisotropic on the smaller scales relevant for data validation.

The question of isotropy of the small-scale variability is best addressed by considering the high-latitude winter variability where we might expect the largest anisotropy. Small-scale structure might be anisotropic in the NH winter, for example, because vortex filaments tend to elongate in the direction of the principal strain axis and thin in the transverse direction. For each pair of measurements separated by horizontal distance $r$, the orientation of the separation vector connecting the two points is also computed. In order to look at NS/EW (“geographic”) isotropy, we define an orientation angle $\psi$ by $\sin^2 \psi = r_y \Delta \phi / r$ where $r_y$ is the earth radius and $\Delta \phi$ is the latitude separation. The angle $\psi$ is a measure of the orientation of the separation vector; $\psi = 0^\circ$ ($90^\circ$) if the two points are separated E/W (N/S), and anisotropy is defined as a dependence of the two-point statistics on $\psi$.

Figure 5 is a plot of the conditional average $\langle \Delta \mu | \psi \rangle$ versus $\psi$ for horizontal separation $r = 200$ km in the winter mid/high latitudes, conditions under which the anisotropy would be expected to be large. Shown for comparison are the results for the remainder of the ER-2 data. In the winter extratropical SH, where the structure is more nearly zonal, the mean increment shows a small overall increase as the orientation changes from E/W to N/S. In the NH on the other hand, the mean increment is largest when $\psi$ is between $30–60^\circ$. This is partly due to the fact that many of the NH winter flights crossed the edge of the polar vortex when it was distorted or off the pole. Overall, anisotropy effects appear to be small, at least on the smaller scales relevant for data validation, and do not support substantially different coincidence criteria for meridional versus zonal separations. It is important to note, however, that the ER-2 flight paths were overwhelmingly N/S oriented, and it is possible that large contrasts in the E/W direction were not adequately sampled. We also note that horizontal anisotropy would perhaps be better defined in terms of the large-scale PV.
For the polar vortex ensemble, spatial averaging sometimes degrades resolution is a reduction in the sensitivity of the measurement differences to the horizontal separation \( r \). As expected, the averaging reduces the occurrence of large differences, especially for \( r = 50 \) km, but the overall effect of resolution on the cumulative distribution is rather small. This is the effect of intermittency; a small number of extreme values can have a large effect on the moments, but little effect on the cumulative distribution except in the tails. The median values, where the CDF = 0.5, are fairly insensitive to averaging and are rather small (2–5%) in all cases for \( r = 50 \) km, increasing to 4–8% for \( r = 200 \) km. The cumulative distributions in Figure 7 demonstrate that about 40% of the measurement pairs separated by 200 km have differences larger than 10% in the NH winter middle/high latitudes. Instrumental or retrieval uncertainties that are about 10% would then be masked by geophysical variability in about 40% of the comparisons, unless the coincidence length is well below 200 km. When the separation distance is reduced to \( r = 50 \) km, only 15% of the measurement pairs have differences exceeding 10%. Figure 7 also indicates that for a coincidence length of 50 km, it is highly improbable when \( r < L \).

3.3. Cumulative Distribution Functions

[31] The mean deviation \( \mu_r \) provides a useful summary of the spatial variability, but gives no information about how often geophysical variability might overwhelm instrumental or retrieval uncertainties. We can gain quantitative insight into this important problem from the CDF\( C_r(d) \), the fraction of tracer increments at scale \( r \) that exceed \( d \) (%). \( C_r(d) \) is shown in Figure 7 for separations \( r = 50 \) and 200 km, and \( C_r(d) \) is overplotted for the same cases with a satellite footprint \( L = 200 \) km (thin lines). As expected, the averaging reduces the occurrence of large differences, especially for \( r = 50 \) km, but the overall effect of resolution on the cumulative distribution is rather small. This is the effect of intermittency; a small number of extreme values can have a large effect on the moments, but little effect on the cumulative distribution except in the tails. The median values, where the CDF = 0.5, are fairly insensitive to averaging and are rather small (2–5%) in all cases for \( r = 50 \) km, increasing to 4–8% for \( r = 200 \) km. The cumulative distributions in Figure 7 demonstrate that about 40% of the measurement pairs separated by 200 km have differences larger than 10% in the NH winter middle/high latitudes. Instrumental or retrieval uncertainties that are about 10% would then be masked by geophysical variability in about 40% of the comparisons, unless the coincidence length is well below 200 km. When the separation distance is reduced to \( r = 50 \) km, only 15% of the measurement pairs have differences exceeding 10%. Figure 7 also indicates that for a coincidence length of 50 km, it is highly improbable

Figure 5. Isotropy of horizontal structure in LS ozone, shown by the conditional mean \( \langle \Delta_r, \psi \rangle \) for increments across horizontal distance \( r = 200 \) km in the NH and SH winter at middle/high latitudes and for all other cases. \( \psi \) is the orientation angle of the separation; \( \psi = 0^\circ \) (E/W), \( \psi = 90^\circ \) (N/S).

Figure 6. Mean difference \( \mu_r^L \) between a simulated satellite measurement of LS ozone with footprint \( L = 200 \) km and a point measurement as a function of their horizontal separation \( r \).
that tracer differences greater than about 25% could be attributed to real spatial variability.

4. Horizontal Variability of Ozone, CO, and Water Vapor in the Middle and Upper Troposphere

[32] We now turn to an analysis of the DC-8 data in order to characterize horizontal variability in the middle and upper troposphere. As before for the LS ozone analysis, we first consider tracer increments across a fixed horizontal scale, and look at the dependence of $\mu_{r,h}$ on vertical separation in the limit $h \to 0$ to isolate contributions from horizontal and vertical structure. Again, the flight paths have less information about vertical profiles, as there are far fewer pairs of measurements for which $r$ is very small and $h$ is large.

[33] A plot of $\mu_{r,h}$ versus vertical separation $h$ at fixed horizontal scale $r = 150$ km is shown in Figure 8b for the middle-/high-latitude UT/LS measurements of ozone, CO and water vapor. The plots for the different chemical species are similar in shape, differing mainly in the overall magnitude of the tracer increments, which are smallest for CO and largest for water vapor. This reflects the common origin of the tracer variability in the middle-/high-latitude UT/LS, which includes stratosphere-troposphere exchange (STE) processes acting on the large-scale differences in the mixing ratios characteristic of stratospheric and tropospheric air masses. The magnitude of the contrast across small scales ultimately comes from the large-scale gradients across air masses when they are stirred together. The fractional difference between the most probable tracer values in the altitude range 4–6 km and those from 10 to 12 km is about 100% for CO, 150% for ozone and 200% for water vapor (the maximum possible value): this 1:1.5:2 scaling is reflected in the overall scaling of the curves in Figure 8b.

[34] In addition, a feature not seen in the corresponding LS plots is the structure at $h < 200$ m, which is also evident in data from the individual missions. Analysis of simulated data with randomly placed Gaussian structures showed that very thin layers with strong contrasts across about 100–200 m produced similar results and could be the origin of this structure; this is also supported by the scale analysis of thin vertical layers during PEM-West A [Newell et al., 1996]. There is an intermediate plateau in which $\mu_{r,h}$ is insensitive to vertical separation, while at larger scales of about 600–700 m, the large-scale vertical gradient of the background atmosphere becomes apparent. Small-scale vertical structure is also seen in the lower-latitude middle/upper troposphere data (Figure 8a), especially for water vapor. The large-scale vertical gradient in water vapor becomes apparent for $h > 700$ m, but there is little dependence on vertical separation for CO and ozone. At smaller horizontal scales ($r = 50$ km, not shown) the results are similar, except that $\mu_{r,h}$ is smaller by about a factor of 2.
The asymptotic behavior in the limit $h \to 0$ is not as clear in the UT as in the LS, thus our ability to separate the contributions from vertical structure is more uncertain for the tropospheric data. To minimize the impact of vertical structure, the following analysis of horizontal variability is based only on measurements that have a vertical separation of less than 50 m. We recognize that the values reported for horizontal variability in the UT/LS could be overestimates by a few percent due to contributions from very thin vertical layers. This is unavoidable, because a cutoff value of $h$ that is too small leaves too few measurements for a proper analysis.

Figure 9 shows the dependence of $\mu_r$ on horizontal scale for UT and UTLS ozone based on data from various DC-8 missions. The inset is the same, but plotted on log axes, and shown also for comparison is $y = r^{1/2}$.

There is substantial variation in the statistics among the individual flights for each mission. The mean and median increments in ozone across horizontal separation scales $r = 40$ and 120 km are plotted in Figure 10, where each symbol represents a single flight. There is some overlap in that increments across scales of 40 km on some days are comparable to those across 120 km on other days. By definition, half of the measurement differences are larger than the median, and as shown later, about 1/3 are larger than the mean. Mean increments larger than 10% are found often and over a wide range of conditions even for small (40 km) separations. Taking an optimistic view, the median plot shows that for the majority of the individual flights, half of the measurement pairs differ by less than 10% if the horizontal separation is 40 km. Overall, Figure 10 argues that a coincidence distance of about 100 km would be too large for ozone, mean increments across scale $r = 150$ km are in the range 13–25% and increments of 10% occur across scales in the range 15–70 km. The results for CO and water vapor (not shown) are similar, except that the increments for CO are about half those of ozone, and for water vapor they are more than twice as large. These results suggest that the horizontal coincidence length for validation should be less than 150 km in the middle and upper troposphere, if the satellite measurement uncertainty is greater than 25%. If the satellite uncertainty is on the order of 10%, however, coincidence lengths should be less than 50 km if geophysical variability is to remain small relative to satellite measurement uncertainties. The inset in Figure 9 is a log-log plot of the same data, with the line $y = r^{1/2}$ overplotted for reference. This shows that, as in the LS, $\mu_r$ varies approximately as $r^{1/2}$.

Figure 10.

Figure 9. Mean tracer increment $\mu_r$ versus horizontal separation $r$ for UT and UTLS ozone based on data from various DC-8 missions. The inset is the same, but plotted on log axes, and shown also for comparison is $y = r^{1/2}$.

Figure 10. (top) Mean $\mu_r$ and (bottom) median tracer increments across horizontal scales $r = 40$ and 120 km. Each plotted symbol corresponds to an individual flight.
large to validate tropospheric ozone under most conditions if the satellite measurement accuracy is 10–25%; in this case, the coincidence distance should be less than 40 km.

[38] The distribution \( P(\Delta_r) \) (not shown) for the DC-8 data has the generic long-tailed shape, similar to that seen for the LS ozone data. The cumulative two-point distributions \( C_r(d) \) for ozone (Figure 11, top), CO (Figure 11, middle) and water vapor (Figure 11, bottom) are shown in Figure 11 (left) for a measurement separation \( r = 120 \) km for several representative missions that span the range of variability. For ozone, the plots show that the fraction of measurement pairs whose fractional difference exceeds 10% ranges from about 0.4 in the UT during Trace-A (TRA) to 0.6 during PEP Tropics-A (PTA), with intermediate values of about 0.5 in the winter high-latitude UT/LS. For CO, the largest differences were found for the TRACE-P (TRP) mission, which sampled highly polluted air that was close to the source, off the east coast of China [Jacob et al., 2003]. The smallest increments were observed during PEP-Tropics B (PTB) for which CO levels were high, but sampled further away from the biomass burning sources in South America. In this case, small fractional differences could arise from large mean values in the denominator of equation (1) but weaker small-scale gradients (e.g., greater homogeneity) due to a longer exposure to small-scale mixing. It is not surprising that the plots for UT water vapor indicate a high degree of variability; 0.9 of the measurement pairs in the UT were different by more than 10% under conditions of strong convection during PEP-Tropics B. In the high-latitude UT/LS (SOLVE), about 0.45 differed by more than 10%.

[39] To estimate the effect of the coarse spatial resolution of a satellite measurement, we again compute the CDFs of fractional differences between a spatially averaged segment of data and a point measurement. The cumulative distributions \( C_r^L(d) \) for ozone, CO and water vapor for a separation \( r = 120 \) km and satellite footprint \( L = 120 \) km are shown for the same cases in Figure 11 (right). We expect spatial averaging to reduce the fractional differences, and examination of the corresponding left and right plots shows this to be the case. This is particularly true for the TRACE-P ozone and CO data, which show strong small-scale structure due to the proximity to sources.

[40] The effect of the averaging is most pronounced in the tails of the distributions, reducing the frequency of large fractional differences. The large reduction in the tails of the CDF for the SUCCESS water vapor data is due in part to the “near field” sampling of aircraft emissions [Toon and Miake-Lye, 1998] with strong small-scale inhomogeneities. The effect of resolution is smaller, however, on the differences that are more typically observed and hence more relevant for satellite data validation. With spatial averaging, median two-point increments of 8–14% are reduced to 5–13% for ozone, from 6–11% to 3–7% for CO and from 10–50% to 8–30% for water vapor. We have also found that there is little difference between \( C_r(d) \) and \( C_r^L(d) \) when the satellite footprint is reduced to 50 km (not shown). It should be emphasized again, however, that simple averaging over a flight path segment could be a poor approximation to averaging over the actual satellite measurement volume.

5. CDFs of Scaled Increments

[41] The CDFs \( C_r(d) \) quantify the likelihood of observing increments larger than an amount \( d \) over a horizontal scale \( r \), an important quantity in assessing the impact of natural variability, and also a means to estimate the amount of correlative data required for adequate validation. The variability in the CDFs \( C_r(d) \) in LS ozone (Figure 7) and middle and upper tropospheric ozone, CO, and water vapor (Figure 11) is not surprising considering the fact that this data set samples a range of sources and meteorological conditions from the middle troposphere to the lower stratosphere. Despite this variability, the CDFs of tracer increments normalized by the mean \( \mu_r \) appear to have a similar form, allowing for a simpler synthesis of the statistical properties of trace gases. We define the CDF of the scaled increments by \( C_r(\delta) = \text{fraction of measurements for which } \Delta_r/\mu_r \geq \delta \). The normalization takes into account the different source strengths, for example, the large-scale gradients in tracers across the tropopause or the intensity of convective, biomass burning or pollution sources.

[42] Figure 12 overplots \( C_r(\delta) \) for the individual LS ozone ensembles as in Figure 7, together with the CO and water vapor curves in Figure 11 (left). Curves are shown for \( r = 120 \) km, but the behavior for other values of \( r \) in the range 30–150 km is similar. The rescaling does not remove all of the differences in the individual curves, especially for water vapor; this is most likely due to the shorter atmospheric residence time of water vapor relative to that of ozone and...
over a wide range of conditions, in different parts of the atmosphere. Scaled increments on the order of $10^2$–$10^3$ are below the a priori estimate of satellite measurement errors. The coincidence length, when it exists, can be used to define an upper bound for the coincidence length. For LS ozone, the correlation length is $350–400$ km, except in the NH winter surf zone (mid latitude regions stirred by Rossby waves), where the lack of a correlation length indicates scale invariance over the range of scales considered ($0–400$ km). In the middle and upper troposphere the scaling range is smaller and we find a correlation length on the order of $100–150$ km under a wide range of conditions. For distances less than the correlation length, $\mu_r$ scales approximately as $r^{1/2}$ in both the stratosphere and troposphere. Across horizontal scales of $150$ km, $\mu_r$ ranges from $4$ to $12\%$ for LS ozone and $15$ to $25\%$ for MT/UT ozone. The variability can be substantial even at smaller scales; mean ozone increments of $9–16\%$ occur across scales of $50$ km in the middle and upper troposphere, and $2–7\%$ in the LS. Increments larger than $30\%$ across $100$ km routinely occur for the shorter-lived water vapor, with median increments as high as $50\%$ in some cases.

The analysis presented here can be used to set some realistic bounds on coincidence length based on the way the differences between two measurements depend on scale. A correlation length, when it exists, can be used to define an upper bound for the coincidence length. For LS ozone, the correlation length is $350–400$ km, except in the NH winter surf zone (mid latitude regions stirred by Rossby waves), where the lack of a correlation length indicates scale invariance over the range of scales considered ($0–400$ km). In the middle and upper troposphere the scaling range is smaller and we find a correlation length on the order of $100–150$ km under a wide range of conditions. For distances less than the correlation length, $\mu_r$ scales approximately as $r^{1/2}$ in both the stratosphere and troposphere. Across horizontal scales of $150$ km, $\mu_r$ ranges from $4$ to $12\%$ for LS ozone and $15$ to $25\%$ for MT/UT ozone. The variability can be substantial even at smaller scales; mean ozone increments of $9–16\%$ occur across scales of $50$ km in the middle and upper troposphere, and $2–7\%$ in the LS. Increments larger than $30\%$ across $100$ km routinely occur for the shorter-lived water vapor, with median increments as high as $50\%$ in some cases.

The analysis presented here can be used to set some realistic bounds on coincidence length based on the way the differences between two measurements depend on scale. A correlation length, when it exists, can be used to define an upper bound for the coincidence length. For LS ozone, the correlation length is $350–400$ km, except in the NH winter surf zone (mid latitude regions stirred by Rossby waves), where the lack of a correlation length indicates scale invariance over the range of scales considered ($0–400$ km). In the middle and upper troposphere the scaling range is smaller and we find a correlation length on the order of $100–150$ km under a wide range of conditions. For distances less than the correlation length, $\mu_r$ scales approximately as $r^{1/2}$ in both the stratosphere and troposphere. Across horizontal scales of $150$ km, $\mu_r$ ranges from $4$ to $12\%$ for LS ozone and $15$ to $25\%$ for MT/UT ozone. The variability can be substantial even at smaller scales; mean ozone increments of $9–16\%$ occur across scales of $50$ km in the middle and upper troposphere, and $2–7\%$ in the LS. Increments larger than $30\%$ across $100$ km routinely occur for the shorter-lived water vapor, with median increments as high as $50\%$ in some cases.

The impact of small-scale variability on satellite data validation depends on the degree of heterogeneity in the observed field relative to the instrumental or retrieval measurement errors. The choice of optimal conditions for validation, for example coincidence criteria, requires some prior estimate of the accuracy and precision of the satellite measurement. The coincidence length should be smaller than the scale at which geophysical variability begins to dominate the estimated satellite measurement uncertainty, and a reasonable choice for the coincidence length $r_c$ is the scale at which $\mu_r$ is below the a priori estimate of satellite uncertainty. Our results show that for tropospheric ozone, validation of satellite measurements by direct comparison with noncoincident measurements is likely to be a problem when the measurements are separated by more than $50$ km and the estimated satellite uncertainties are on the order of $10\%$ or less. For similar satellite uncertainties, the coincidence length should be less than $200$ km for LS ozone except in the NH mid/high latitudes where it should be less than $100$ km. We note that in situ measurement uncertainties will also contribute to measurement differences; however they should be much smaller than those associated with satellite measurements if they are to be useful for validation purposes. Our results suggest that the suggested $6$ hr/100 km coincidence criteria for validation of tropospheric species, and the half-day/200 km(lat)/1000 km(lon) criteria for stratospheric species [Froidevaux and Douglass, 2001] may not be sufficient to insure that real geophysical variability does not dominate measurement uncertainties. In addition, vastly different coincidence lengths in the zonal and meridional directions do not seem warranted, but the possibility remains that along-jet and across-jet variability could be different and may have to be taken into account.

In order to estimate the effect of coarse satellite resolution, the two-point analysis was modified by comparing the difference between a neighborhood average around one point to a second point some distance away. As expected, the occurrence of large tracer differences was
reduced somewhat for a satellite footprint of 200 km (LS) and 120 km (UTLS/UT) relative to the two-point case. Using the two-point statistics to estimate the difference between a lower-resolution satellite measurement and a point measurement can therefore sometimes overestimate the contribution of real variability and in general should be regarded as an upper bound. Additional studies (not shown) indicate that the statistics of satellite/in situ differences are very similar to the two-point statistics when the satellite footprint is 50 km or smaller. Our attempt to take into account the satellite resolution in this way is constrained by the 1-D sampling characteristics of the aircraft platform. The average was based on temporally nearby measurements, but the spatial distribution of these measurements depends on the geometry of the flight path segment and may be a poor approximation to the actual volume sampled by the satellite measurement. Given these uncertainties, the two-point differences provide a less ambiguous, if somewhat overestimated, assessment of the contribution of real variability to measurement comparisons. A better approximation would require averaging the two-point increments over the actual satellite footprint, which would, in turn, require sampling designed for that purpose. It may also be possible to use the two-point PDF to appropriately average over the actual satellite footprint. Our results suggest that the two-point differences are a reasonable approximation for the typical variability if the satellite footprint is smaller than about 100 km. Validation of satellite measurements with a broad vertical weighting function remains problematic especially in situations where strong vertical layering below the satellite resolution occurs. The focus in this paper was on horizontal variability, but the methods could be applied to analysis of vertical structure using, for example, airborne or ground-based lidar data.

There were a few statistical features that were found to be common to the various data sets considered here, and that could perhaps lead to some new ways to think about noncoincident satellite data validation. The mean tracer increments show an approximate \( r^{1/2} \) dependence for separation distances less than the correlation length, and the likelihood that a measurement will exceed the mean \( \mu_r \) is about 1/3, a fact that follows from the nearly universal form of the CDF of the normalized differences \( \delta = \Delta_x / \mu_x \). These results hold for a wide range of conditions, and for different tracers in the middle and upper troposphere as well as LS ozone, suggesting that they follow from the properties of the underlying dynamics for the longer-lived species. In an aircraft validation campaign, one can compute the tracer increments from the aircraft data to characterize the local conditions. The CDFs could then provide some perspective on the likelihood that the observed differences have a strong geophysical component. Another way in which these methods might be utilized in a validation campaign would be to compute differences between satellite and correlative measurements as a function of separation and to look at their asymptotic properties in the limit as \( r \to 0 \). In the limit of perfect measurements, where the geophysical variability dominates, one should find these differences approaching zero as \( r \to 0 \), with a behavior proportional to \( r^{1/2} \) if the two measurements are coincident to within 30 min or so. The differences will asymptote to some nonzero value in this limit if there is a bias, and the asymptotic value represents the uncertainty in the measurement, assuming that the instrumental error for the in situ measurement is well below that of the satellite measurement. This method removes the requirement of exact spatial coincidence, utilizing instead the asymptotic properties of the differences in the limit of strict collocation. Of course another way to detect bias is to compare the peak (most probable) value of the satellite and correlative one-point PDFs. While this requires a large number of both satellite and in situ measurements at about the same place and time, strict collocation is not necessary. Finally, when evaluating the quality of the satellite measurement under conditions of strong variability, the non-Gaussian signature of the two-point statistics of real tracer variability may be of some use in deciding whether observed differences look geophysical, especially if the differences due to instrumentation or retrieval are Gaussian-distributed.

Acknowledgment. The authors would like to acknowledge the EOS Interdisciplinary program and the NASA/ACMAP program for support of this work.

References


