Retrievals of carbon monoxide profiles from MOPITT observations using lognormal a priori statistics

M. N. Deeter,[1] D. P. Edwards,[1] and J. C. Gille[1]

Received 5 September 2006; revised 27 February 2007; accepted 14 March 2007; published 15 June 2007.

[1] Optimal estimation methods, such as the "maximum a posteriori" solution, are commonly employed for retrieving profiles of atmospheric trace gases from satellite observations. To complement the information actually contained in the measured radiances, such methods exploit a priori information describing the gases' variability characteristics. We show that in situ surface-based data sets for carbon monoxide (CO) volume mixing ratio (VMR) indicate that the variability of CO is more accurately modeled in terms of a "lognormal" probability distribution function (PDF) than a "VMR-normal" PDF. The VMR-normal PDF is particularly poor at describing CO variability in unpolluted conditions. We also compare retrievals of carbon monoxide (CO) vertical profiles based on Measurements of Pollution in the Troposphere (MOPITT) observations for 1 day using both VMR-normal and lognormal statistical models. Use of the lognormal model improves retrieval convergence and yields fewer profiles with unphysically small VMR values. Generally, these results highlight the importance of properly representing the variability of trace gas concentrations in optimal estimation-based retrieval algorithms.

Citation: Deeter, M. N., D. P. Edwards, and J. C. Gille (2007), Retrievals of carbon monoxide profiles from MOPITT observations using lognormal a priori statistics, *J. Geophys. Res.*, *112*, D11311, doi:10.1029/2006JD007999.

1. Introduction

[2] By themselves, radiance measurements from passive satellite instruments are not usually sufficient to determine trace gas atmospheric profiles; additional constraints are typically required. In retrieval methods based on the "maximum a posteriori" (MAP) solution, these external constraints are provided by the a priori profile and the a priori covariance matrix [*Rodgers*, 2000]. In the Measurements of Pollution in the Troposphere (MOPITT) version 3 (or "V3") product for carbon monoxide (CO), the state vector used in the MAP-based retrieval algorithm includes surface temperature, surface emissivity, and CO volume mixing ratio (VMR) values at 7 levels from the surface up to 150 mb [*Deeter et al.*, 2003]. To simplify interpretation of the MOPITT CO product, the V3 retrieval algorithm was designed to clearly distinguish information contained in the observations from a priori information. Thus a single "global" a priori profile and covariance matrix are employed in V3 retrievals. The V3 a priori profile and covariance matrix were obtained by analyzing a set of 525 CO profiles drawn from field experiments and fixed measurement sites around the world.

[3] Both the MOPITT calibrated radiances and CO retrieved profiles were validated globally using in situ profiles measured from aircraft [*Deeter et al.*, 2004a; *Emmons et al.*, 2004]. Analysis of MOPITT averaging kernels indicated that V3 retrievals are typically most sensitive to CO in the midtroposphere and contain less than two pieces of independent information [*Deeter et al.*, 2004b]. Nevertheless, the vertical resolution is sufficient to distinguish CO in the midtroposphere and upper troposphere, at least in tropical regions [*Deeter et al.*, 2004b; *Edwards et al.*, 2006]. Sensitivity to surface-level CO is typically lower than for the midtroposphere because of low thermal contrast between the Earth's surface and the boundary layer. Thus the effects of the assumed a priori are typically much stronger for retrieved surface-level CO concentrations than for higher levels.

[4] For simplicity and speed of convergence, MAP-based retrieval algorithms usually assume Gaussian ("normal") probability distribution functions (PDFs) to describe the a priori variability of the retrieval parameters contained in the retrieval state vector. In the V3 MOPITT product, CO concentrations are represented in the state vector in terms of VMR. The operational retrieval algorithm for the Microwave Limb Sounder (MLS) also employs a VMR-based state vector [*Livesey et al.*, 2006].

[5] Alternatively, the state vector in the MAP framework can be based on values of log(VMR) [*Engelen and Stephens*, 1999; *Rodgers*, 2000]; the VMR PDF is then assumed to be "lognormal." Like the simple Gaussian PDF, the lognormal PDF is characterized by two parameters which describe the peak and width of the distribution. Trace-gas profiles retrieved from tropospheric emission spectrometer (TES) observations are currently based on lognormal statistics

[1]Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, Colorado, USA.

0148-0227/07/2006JD007999

Table 1. GMD Surface Stations for Which VMR Histograms are Plotted in Figure 1

Station Name	Latitude	Longitude	Altitude, m	Observation Period	Number of Observations
American Samoa	14.24°S	170.57°W	42	1988–2004	2114
Ascension Is.	7.92°S	14.42°W	54	1989–2004	2207
Seychelles	4.67°S	55.17°E	7	1990–2004	1355
Mt. Waliguan	36.29°N	100.90°E	3810	1990–2004	1178
Niwot Ridge	40.05°N	105.58°W	3523	1988–2004	1487
Baltic Sea	55.42°N	17.07°E	28	1992–2004	2044

[*Bowman et al.*, 2006]. Lognormal statistics are often used to describe physical parameters which exhibit a skewed frequency distribution. This occurs, for example, when (1) the retrieved parameter is physically restricted to positive values and (2) the variability (for example, standard deviation) of the retrieval parameter is comparable to (or even larger than) the mean value.

[6] Relative to the VMR-normal PDF, the lognormal PDF decreases more sharply at low VMR and decreases more gradually at high VMR. Negative VMRs, which are unphysical, are permissible in the VMR-normal PDF but are disallowed by the lognormal PDF. As demonstrated in section 3, these differences produce fundamentally different constraints in VMR-based and log(VMR)-based retrievals. Differences are most apparent at very low and very high VMR. At low VMR values, retrievals based on log(VMR) are confined to a narrower range than VMR-based retrievals. This results in a stronger a priori constraint for the log(VMR)-based retrievals. Conversely, at high VMR, log(VMR)-based retrievals will naturally extend over a broader range of VMR values compared to VMR-based retrievals. This yields a weaker constraint for the log(VMR)-based retrievals. These differences in the underlying constraints imply that the relative weighting between information from the observed radiances and a priori information should be different in the two schemes.

[7] In the following, we investigate the effects of representing CO variability in the MOPITT retrieval algorithm in terms of lognormal statistics. In section 2, arguments are presented which motivate the use of lognormal statistics in the MOPITT retrieval algorithm. We find that the lognormal model approximates statistics of observed in-situ measurements significantly better than the VMR-normal model. In section 3, we apply both the operational retrieval algorithm and an adapted lognormal version to one day of MOPITT observations in order to analyze differences in retrieval results associated with the use of lognormal statistics. Conclusions are offered in section 4.

2. The Lognormal Model

2.1. Observed Variability of Surface-Level CO

[8] Previously, studies based on in situ measurements of urban pollution indicated that pollutant concentrations (including CO) are well represented by the simple two-parameter lognormal distribution [*Bencala and Seinfeld*, 1976]. However, for a retrieval algorithm to be applied globally, the validity of the assumed PDF should be demonstrated generally. Within the National Oceanic and Atmospheric Administration (NOAA), the Global Monitoring Division (GMD, formerly the Climate Monitoring and Diagnostics Laboratory) conducts an ongoing program to monitor surface-level c rations of CO (and other trace gases) at observation sites around the world [*Novelli et al.*, 1998]. Histograms based on archived data from these sites are useful for evaluating PDF models for CO concentration.

[9] For defining long-term variability, especially at the extreme values of observed VMR, data sets with records lasting at least 10 years and including at least a thousand discrete observations are especially valuable. Below, VMR histograms are analyzed for the six GMD surface stations listed in Table 1. The list includes both oceanic and continental surface sites and sites in both the Northern and Southern Hemispheres. The list includes sites mainly affected by biomass burning (for example, Ascension Island [*Gregory et al.*, 1996; *Edwards et al.*, 2006]) and sites mainly affected by fossil fuel consumption (for example, the Baltic Sea [*Karlsdottir et al.*, 2000]). With respect to the general shape of the VMR histograms, the VMR statistics for these sites represent the majority of GMD surface stations with continuous records of a decade or more.

[10] VMR histograms for the locations and observation periods listed in Table 1 are shown in Figure 1. Histograms are shown for both VMR (using bins of fixed ΔVMR) and log(VMR) (using bins of fixed Δlog(VMR)). VMR and log(VMR) histograms are shown in blue and red, respectively. (Total areas for the VMR- and log(VMR)-based histograms are unequal because their bin sizes are necessarily unequal.) Within the same figure are shown normalized VMR and log(VMR) Gaussian functions (dashed curves) based on the mean and variance of the VMR and log(VMR) data.

[11] The similarity of the fitted Gaussian curves to their corresponding histograms indicates the validity of the PDF models. In all of the VMR histograms (blue) shown in Figure 1, the underlying PDF is clearly asymmetric. Relative to the peak of the histogram, the frequency of low CO values decreases much more sharply than the frequency of high CO values. This asymmetry characterizes VMR histograms for the vast majority of GMD surface stations. Consequently, the Gaussian fits to the VMR histogram (i.e., the dashed blue curves) tend to decrease too slowly toward low CO VMR values and too rapidly toward high CO VMR values. Conversely, in terms of log(VMR), the lognormal fits (the dashed red curves) reasonably approximate the actual histograms at both the low and high ends of the range of VMR values. When considered with previous findings [*Bencala and Seinfeld*, 1976], these results indicate that the lognormal distribution describes the variability of tropospheric CO over widely varying atmospheric conditions in diverse geographical settings. In particular, for describing the variability of CO in relatively unpolluted conditions (for example, VMRs less than 70–80 ppbv), the lognormal model is clearly superior to the VMR-normal model.

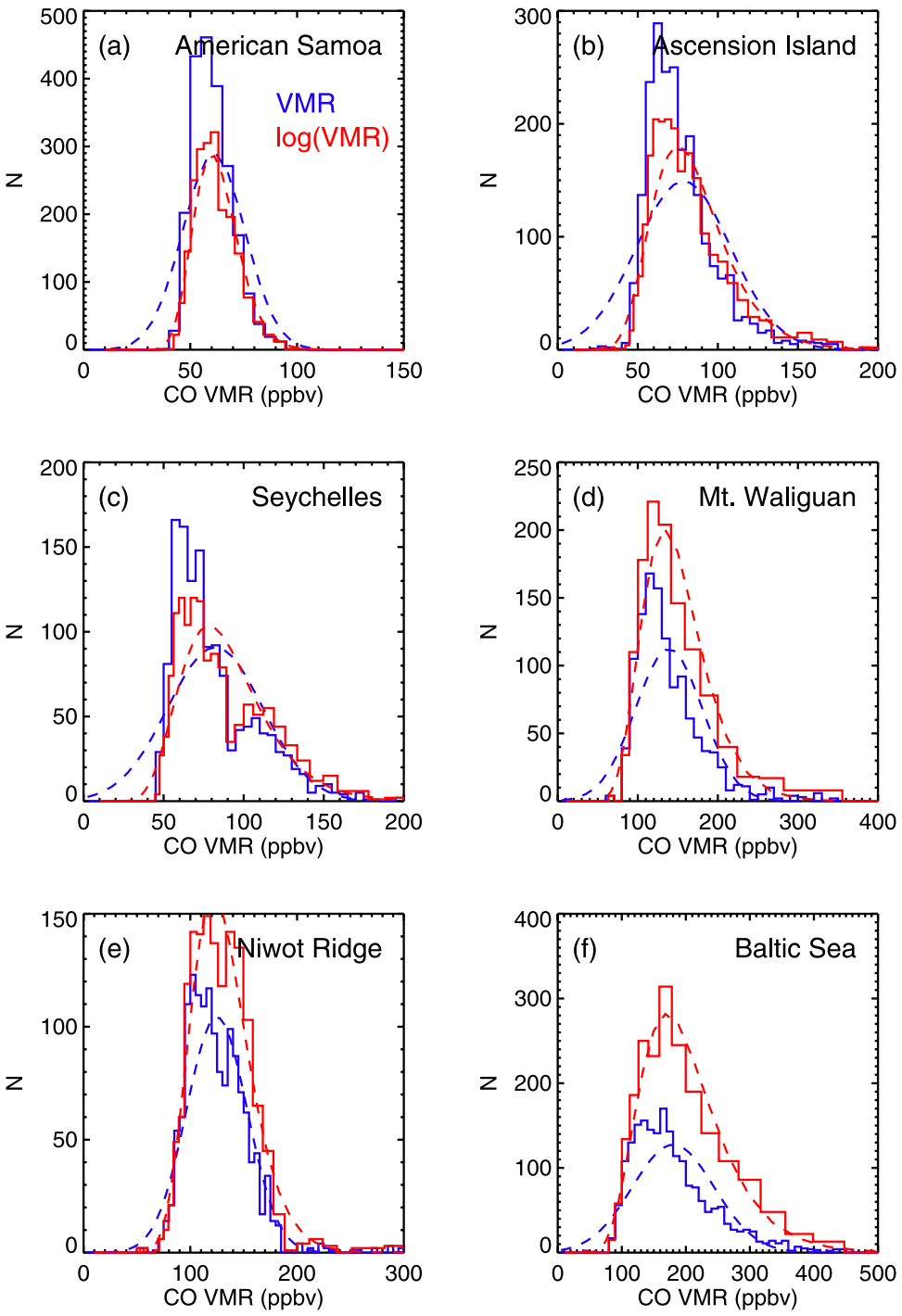

Figure 1. Histograms of CO VMR measurements at GMD monitoring sites listed in Table 1. Blue plots show histograms and Gaussian fits (dashed) for VMR statistics. Red plots show corresponding results based on log(VMR) statistics. Areas of VMR and log(VMR) histograms are unequal because of unequal bin sizes. For all sites, the log(VMR) Gaussian fit (in red) approximates the actual observations much better than the VMR Gaussian fit (in blue).

[12] Physically, the asymmetry in the VMR histograms might be linked with characteristics of CO sources. For example, methane oxidation produces a global "background" concentration of approximately 40–50 ppbv [e.g., *Hauglustaine et al.*, 1998]. Since methane is relatively well mixed in the troposphere, this source of CO exhibits relatively weak temporal and spatial variability. In contrast, surface-based emissions of CO (such as biomass burning, fossil fuel consumption, etc.) are characterized by high degrees of temporal and spatial variability. Thus CO variability should be small in relatively clean conditions (where methane oxidation dominates) and large in relatively pol-

luted conditions (where surface-based emissions dominate). The asymmetry in the VMR histograms might also be linked with wind speed statistics [*Bencala and Seinfeld*, 1976].

2.2. Limitations of the Mopitt V3 Product

[13] The MOPITT global a priori used in V3 retrieval processing includes a surface-level CO VMR of 119 ppbv and a surface-level standard deviation (calculated as the square-root of the diagonal element of the covariance matrix) of 94 ppbv. Under the assumption of a VMR-normal PDF (as in V3), these values would suggest that a VMR of 25 ppbv is entirely realistic, since such a value would fall only one standard deviation below the mean (a priori) value. In fact, however, such low values are never actually observed in in situ data sets. The lowest CO VMR values generally observed in the troposphere are in the range of 40–50 ppbv and are determined by methane oxidation. As described above, this contradiction is a natural consequence of the assumption that CO VMR variability obeys a normal (Gaussian) PDF; this PDF fails to properly represent the sharp dropoff in observed VMRs at the low end. With respect to the MOPITT product, this problem allows retrieved VMR values to fall below physically realistic limits. Conversely, the same problem might unnecessarily suppress very high VMR values, for which there is no well-defined physical maximum.

[14] In the MOPITT retrieval algorithm, the retrieved profile is derived iteratively [*Deeter et al.*, 2003]. Due to the nonlinear nature of the radiative transfer model, several iterations (typically three to five) are required before the profile for a particular observation converges. After each iteration, the new profile is fed to the operational forward radiative transfer model [*Edwards et al.*, 1999] to calculate new theoretical radiances. This forward model requires that the input CO profile be consistent with the "training-set" profiles used in its development (which, like the a priori, are based on in situ data sets). Model-calculated radiances for profiles outside the envelope of training-set profiles would be characterized by unacceptably large radiance errors. Therefore if a profile containing unphysically small (or large) CO VMR values is fed to the model, the retrieval processing for that observation will cease. We refer to such retrievals as "unconverged." Thus if VMR variability is substantially overestimated (as described in the preceding paragraph), a natural consequence would be a significant number of unconverged retrievals. In fact, up to about 10% of MOPITT retrievals fail in this manner in particularly unpolluted regions. As described in section 3.2.3, this effect can effectively bias ensembles of retrieved profiles.

2.3. Log(VMR)-Based Weighting Functions

[15] In the MOPITT retrieval algorithm, expressing the CO profile in the state vector in terms of log(VMR) rather than VMR will modify fundamental properties of the MOPITT weighting functions and, by association, the MOPITT averaging kernels. For MAP-based retrieval algorithms, the relationship between the retrieved profile x_{rtv}, the true profile x_{true}, and a priori profile x_{a} can be expressed through the averaging kernel matrix A [*Rodgers*, 2000]. Specifically,

$$x_{\mathrm{rtv}} \approx A x_{\mathrm{true}} + (I - A) x_{\mathrm{a}} \tag{1}$$

where I is the identity matrix. Thus A defines the relative weights of the true profile and a priori profile in the retrieved profile. A is dependent on the radiance error covariance matrix C_{e}, a priori covariance matrix C_{x}, and weighting function matrix K through the relation

$$A = \left(K^T C_{\mathrm{e}}^{-1} K + C_{\mathrm{x}}^{-1}\right)^{-1} K^T C_{\mathrm{e}}^{-1} K \tag{2}$$

Ultimately, therefore, the relative weight of the true profile in the retrieved profile depends strongly on K. The weighting function matrix K quantifies the sensitivity of the observed radiances to perturbations in each element of the state vector, i.e.,

$$K_{ij} = \frac{\partial R_i}{\partial x_j} \tag{3}$$

where R_i is the ith element of the measurement vector and x_j is the jth element of the state vector.

[16] In MOPITT V3 retrievals, K^{V3} is mathematically defined as the radiative sensitivity to absolute VMR changes (for example, in units of parts per billion by volume). Thus

$$K_{ij}^{V3} = \frac{\partial R_i}{\partial\left(VMR_j\right)} \tag{4}$$

In the log(VMR) retrieval scheme, however, K^{Log} quantifies the radiative response to perturbations in log(VMR), i.e., fractionally based perturbations. In this case,

$$K_{ij}^{\mathrm{Log}} = \frac{\partial R_i}{\partial\left(\log_{10}\left(VMR_j\right)\right)} = \left(\log_{10} e\right) VMR_j K_{ij}^{V3} \tag{5}$$

This relationship between K^{Log} and K^{V3} suggests that fundamental differences should be expected between VMR-based and log(VMR)-based retrievals. Specifically, because of the proportionality of K^{Log} and VMR, log(VMR)-based retrievals should be characterized by weighting functions which increase in magnitude along with increasing VMR. As K increases or decreases, equation (2) dictates that A must also increase or decrease. For example, as $|K|$ approaches 0, so does A. As A decreases, so does the degrees of freedom for signal (DFS), which is an index for information content [*Rodgers*, 2000; *Deeter et al.*, 2004b]. DFS is simply the trace of the averaging kernel matrix A.

[17] This effect may be demonstrated through numerical simulations using the MOPITT operational forward radiative transfer model [*Edwards et al.*, 1999]. Calculated weighting functions for the MOPITT Channel 1 Difference ("1D") signal for VMR-based and log(VMR)-based state vectors are compared in the two panels in Figure 2. In each panel, weighting functions are shown for four "base" CO profiles produced by applying one of four profile scaling

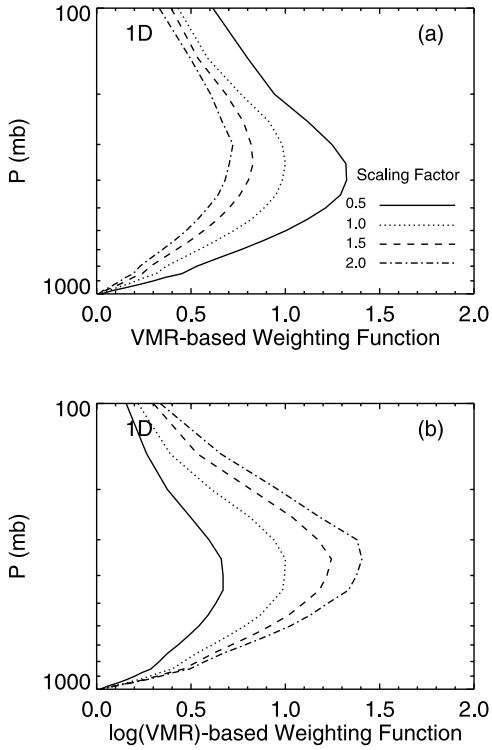

Figure 2. Comparison of normalized weighting functions for MOPITT 1D signal for state vectors based on CO VMR (top panel) and log(VMR) (bottom panel). Response of weighting function to increasing CO concentration (as produced through a profile scaling factor between 0.5 and 2.0) indicates opposite trends for VMR-based and log(VMR)-based state vectors. (Actual weighting functions are negative; normalized weighting functions do not indicate this.)

factors (0.5, 1.0, 1.5, and 2.0) to the global a priori profile used in V3 processing. Weighting functions in each panel were normalized to the peak value for the a priori profile weighting function (scaling factor of 1.0). Normalizing the weighting functions this way results in strictly positive values; the actual weighting functions K are negative.

[18] For the VMR-based weighting functions shown in Figure 2a, higher concentrations of CO decrease $|K|$. This decrease indicates the asymptotic behavior of the dependence of thermal channel radiances on CO concentration; as CO concentrations increase without bound, $|K|$ must approach 0. In contrast, the log(VMR) weighting functions shown in Figure 2b increase with increasing CO concentration as indicated by equation (5). These results suggest that log(VMR)-based retrievals should yield averaging kernels which tend to be smaller in regions of low CO concentrations, and larger in regions of high CO concentrations. These trends will apply to DFS also, since DFS is simply the trace of A. Therefore the retrieval information content (as indicated by the DFS) should generally be different in VMR-based and log(VMR)-based retrievals. Physically, however, this difference is associated with the different underlying constraints associated with the assumed PDF (as described in section 1) and is not related to the sensitivity of the radiances to absolute changes in CO concentration.

Table 2. A Priori Mean Profile Based on log(VMR), Base 10

Surface	850 mb	700 mb	500 mb	350 mb	250 mb	150 mb
−7.0115	−7.0438	−7.0668	−7.0892	−7.0931	−7.1301	−7.2445

3. Retrieval Comparisons

3.1. Implementation

[19] Differences between VMR-based and log(VMR)-based retrievals were analyzed after processing one day of MOPITT observations with both the V3 algorithm and an adapted algorithm based on log(VMR) statistics. The selected day, 18 March 2001, was chosen arbitrarily, but exhibits features (for example, large hemispheric differences in zonal-mean CO) commonly observed in MOPITT retrievals. Details of the MOPITT V3 retrieval algorithm were described previously [*Deeter et al.*, 2003]. Several modifications were made to that algorithm to enable the use of lognormal statistics. First, the global a priori profile x_a and a priori covariance matrix C_x used in V3 were recalculated after converting all in situ profile data from profiles of VMR to profiles of log(VMR), base 10. (Base 10 logarithms were used because they can be more readily interpreted than natural logarithms; for example, a log(VMR) value of −7.0 corresponds exactly to a VMR value of 100 ppbv. Physically, the choice of logarithm base is irrelevant.) The resulting "global" log(VMR) a priori profile and a priori covariance matrix are listed in Tables 2 and 3. In addition, all references to the MOPITT state vector x in the retrieval code were converted from VMR to log(VMR).

3.2. Potential Systematic Retrieval Differences

[20] A variety of effects are likely to contribute to systematic differences between VMR-based and log(VMR)-based retrieval results. These differences should be largest where the influence of a priori on the retrieved profile is greatest. Thus VMR-based and log(VMR)-based retrievals should exhibit the closest agreement in the middle-troposphere grid levels (for example, 500 and 700 mb), and should exhibit the largest differences at the surface. At least four distinct effects might contribute to systematic differences between VMR- and log(VMR)-based retrievals.

3.2.1. The A Priori Profile Effect

[21] First, as indicated by equation (1), retrieval differences will result directly from the different a priori profiles x_a in the two methods. For example, at the surface, the VMR-based a priori value is 119 ppbv [*Deeter et al.*, 2003], whereas the log(VMR)-based a priori value corresponds to a

Table 3. A Priori Covariance Matrix C_a Based on log(VMR), Base 10[a]

	Surface	850 mb	700 mb	500 mb	350 mb	250 mb	150 mb
Surface	0.0685	0.0508	0.0349	0.0224	0.0185	0.0165	0.0139
850 mb		0.0505	0.0362	0.0234	0.0187	0.0160	0.0138
700 mb			0.0344	0.0208	0.0167	0.0129	0.0101
500 mb				0.0199	0.0149	0.0101	0.0067
350 mb					0.0159	0.0118	0.0091
250 mb						0.0135	0.0124
150 mb							0.0173

[a]Elements below matrix diagonal are not shown, since C_a is symmetric.

VMR value of 97 ppbv. This effect will tend to produce log(VMR)-based retrievals with smaller VMR values compared to the V3 product. The magnitude of this effect depends on the weighting of the a priori profile in the retrieved profile [as defined in equation (1)]. Differences associated with this effect should be most evident in situations where a priori influence is particularly strong, such as nighttime scenes over land [*Deeter et al.*, 2003]. As demonstrated in section 3.3, this effect can be quantified without interference from the other three effects described immediately below. Specifically, VMR-based and log(VMR)-based retrieval results may be obtained using the same a priori profile (for example, the log(VMR)-based a priori profile) and compared.

3.2.2. The PDF Shape Effect

[22] Second, the fundamentally different shapes of the VMR-normal PDF and lognormal PDF produce different effects depending on the context. In very clean regions, VMR-based retrievals extend to unphysically small VMR values, as noted above. In the log(VMR)-based method, these same retrievals should shift to higher CO concentrations, thereby raising the mean retrieved value. Likewise, in regions of very high CO concentrations, the lognormal PDF decays slower than the VMR-normal PDF, potentially increasing the mean VMR in those regions as well. Regions with moderate CO concentrations should not change as significantly.

3.2.3. The Convergence Effect

[23] Third, in especially clean regions, a small but significant fraction (usually less than 10%) of V3 retrievals fails when the profile falls outside of boundaries of minimum physical VMR imposed by the operational forward model (as described in section 2.2). In log(VMR)-based retrievals, these observations are much more likely to result in converged retrievals but will probably produce VMR values well below the local mean value. The absence of these low-VMR retrievals in the VMR-based product may therefore act as a source of positive bias which is eliminated in log(VMR)-based retrievals.

3.2.4. The DFS Effect

[24] Finally, as described in section 2.3, differences in the underlying constraint associated with the assumed PDF lead to different dependences of DFS on VMR for VMR-based and log(VMR)-based retrieval methods. In particularly clean regions, this effect should be manifested as a reduction in DFS (compared to the V3 product). Conversely, in regions of relatively high CO, the same effect could produce log(VMR)-based retrievals with apparently higher DFS values. As DFS increases, retrievals are less constrained by the a priori, and relatively more sensitive to the true CO profile.

[25] Depending on the relative magnitudes of these four effects, log(VMR)-based retrievals could either be systematically smaller or larger than corresponding VMR-based retrievals. As indicated by DFS, information content could also be affected. These effects have been quantitatively studied both globally and regionally after processing one day of MOPITT data with two versions of the retrieval algorithm.

3.3. Zonal-Mean Comparisons

[26] Zonal-mean VMR values calculated for 10°-wide latitudinal bins for 1 rch 2001 are compared in Figure 3. Retrieval results over both ocean and land were included. For clarity, only results for the surface, 700 and 350 mb levels are plotted. Results are shown for (1) the V3 product (VMR-based algorithm with the VMR-based a priori profile, shown in blue), (2) the log(VMR)-based algorithm with the V3 a priori profile (red), and (3) the log(VMR) algorithm with the log(VMR)-based a priori profile (green). For log(VMR)-based results presented in both this section and the following section, means were calculated in "log-space," and then converted to VMR. Differences between the three products are much larger for the surface level VMR than for the two higher levels for two reasons. First, because the MOPITT weighting functions are often weaker near the surface than at higher levels, the influence of a priori is generally larger at the surface than at either of the other two levels. Second, as indicated by both the VMR-based a priori covariance matrix (listed in the work of *Deeter et al.* [2003]) and the log(VMR)-based a priori covariance matrix (listed in Table 3), a priori variance values peak sharply at the surface. Generally, as a priori variance values increase, so does the relative weight of the a priori in the retrieved profile.

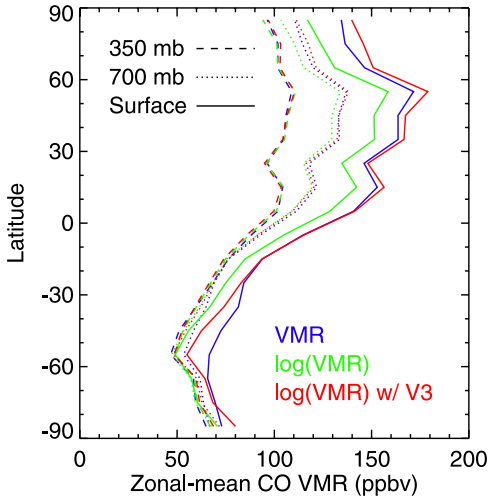

Figure 3. Comparisons of zonal-mean retrieval results for one day (18 March 2001). Results for the standard VMR-based algorithm are shown in blue. Results produced by the log(VMR)-based algorithm with the log(VMR)-based a priori profile are plotted in green. Results for the log(VMR)-based algorithm with the VMR-based a priori profile are plotted in red. See text in section 3.3.

[27] Comparing surface-level zonal means for the V3 product (blue) and results based on the log(VMR) algorithm and V3 a priori profile (red) in Figure 3 is useful because these retrievals share the same a priori profile. Any differences in these retrievals should therefore be the result of the effects described in sections 3.2.2, 3.2.3, and 3.2.4. Surface-level zonal means for these two are typically within about 10 ppbv, except in SH midlatitudes. In the NH, the slightly larger zonal mean values produced by the log(VMR)-based retrievals are consistent with the effects described in sections 3.2.2 and 3.2.4. In relatively polluted conditions, both the PDF shape effect and increased DFS values should yield larger retrieved VMR values in the log(VMR)-based algorithm. In SH midlatitudes, the smaller surface-level zonal

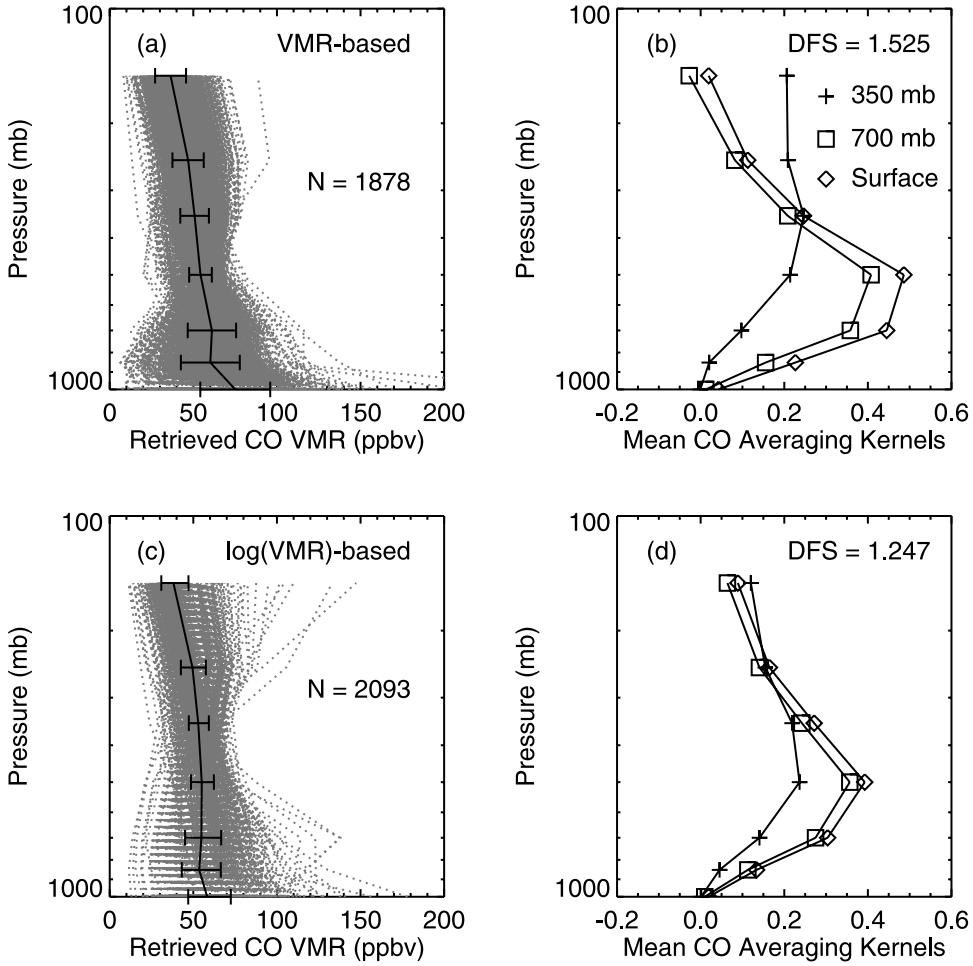

Figure 4. VMR-based daytime retrieved profiles and mean averaging kernels for Southern Hemisphere oceans (between 50°S and 40°S) for 18 March 2001. (a) Top left panel and (c) bottom left panel show all retrieved profiles (in grey) for operational (VMR-based) and log(VMR)-based retrievals, respectively. Mean and standard deviation values at each level are indicated by black line and error bars. (b) Top right panel and (d) bottom right panel show mean averaging kernels for same region for retrieved VMR, and retrieved log(VMR), respectively. Averaging kernels are shown for surface (diamonds), 700 mb (squares), and 350 mb (plus symbol). "*N*" and "DFS" refer to the total number of retrievals and the mean degrees of freedom for signal, respectively.

means for the log(VMR)-based retrievals are consistent with increased retrieval convergence as described in section 3.2.3. Evidently, the convergence effect dominates the PDF shape effect and the DFS effect which should both yield higher retrieved VMR in log(VMR)-based retrievals in unpolluted conditions.

[28] The effect of the a priori profile described in section 3.2.1 is demonstrated in Figure 3 by comparing zonal means for the log(VMR)-based algorithm (green) and log(VMR)-based algorithm with V3 a priori profile (red). At the surface, the smaller VMR associated with the log(VMR)-based a priori profile results in zonal means smaller by about 10 to 25 ppbv. Zonal mean differences associated with the a priori profile appear to scale with the zonal mean VMR. Fractionally, the smaller VMR values associated with the log(VMR)-based a priori profile result in surface level zonal means smaller by about 15%. The magnitude of these differences is roughly comparable to the difference in the surface-level VMRs indicated by the V3 and log(VMR)-based a priori profiles (22 ppbv).

3.4. Regional Comparisons

[29] As observed in Figure 1, differences between VMR-based and log(VMR)-based PDFs are most apparent at the extremes of CO VMR. Similarly, differences between retrievals based on VMR and log(VMR) state vectors should be most apparent in especially clean regions and polluted regions. Comparisons of the retrieval results, both in terms of the retrieved profiles and associated averaging kernels were analyzed for both one typically clean region and one typically polluted region.

[30] To evaluate retrieval differences in particularly clean conditions, all daytime oceanic retrievals for 18 March 2001 were extracted between 40°S and 50°S. Retrieved profiles and mean averaging kernels for the VMR-based and log(VMR)-based algorithms are shown in Figure 4. Retrieved profiles (plotted as dotted grey lines), mean

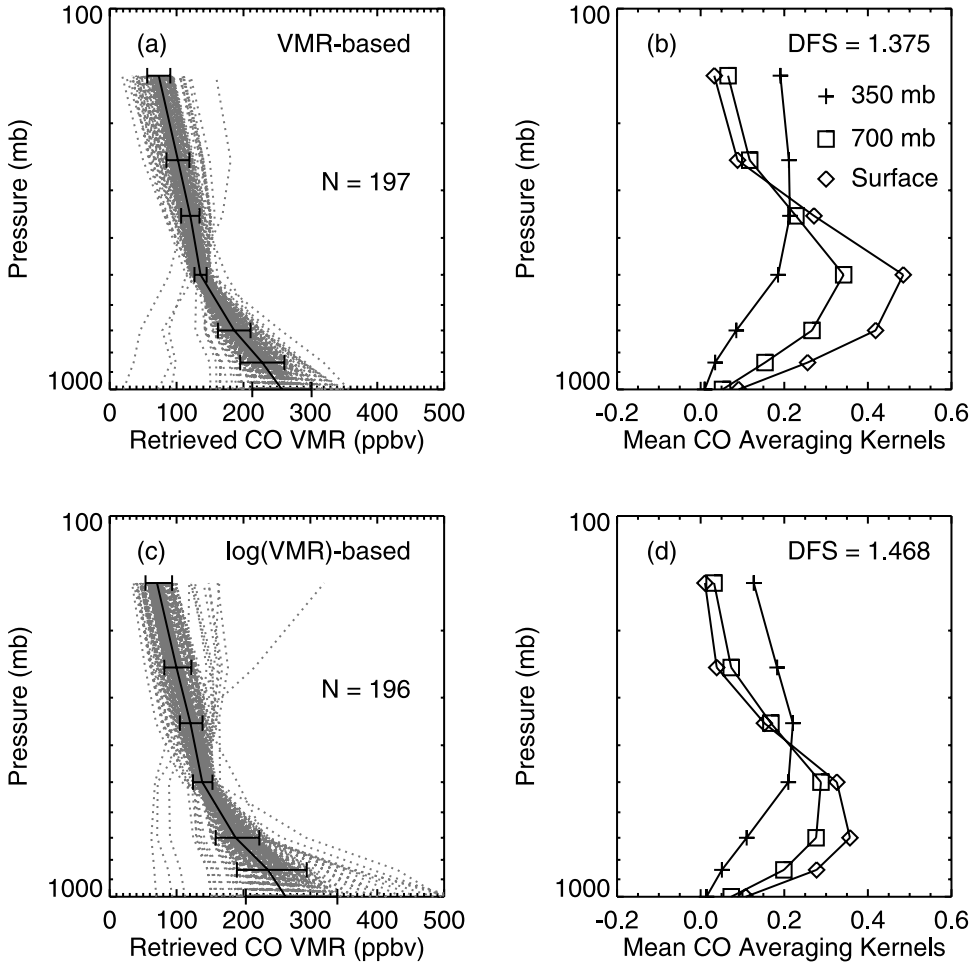

Figure 5. VMR-based and log(VMR)-based daytime retrieved profiles and mean averaging kernels for E. China (between 25°N, 35°N, 110°E, and 120°E) for 18 March 2001. See caption of Figure 4.

profiles (solid black lines), and standard deviations are plotted in Figures 4a and 4c. Mean averaging kernels for the surface (diamonds), 700 mb (squares), and 350 mb (plus symbols) levels are plotted in Figures 4b and 4d. The total number of converged retrievals *N* and mean DFS values are also indicated in the figures.

[31] For the selected region in the Southern Hemisphere, the retrieval yields (i.e., convergence rates) for the VMR-based and log(VMR)-based algorithms are 89.0 and 99.2%, respectively. Moreover, the fraction of unphysical converged retrievals (defined by the fraction of converged retrievals exhibiting a minimum VMR in the lower troposphere of less than 40 ppbv) for the two algorithms are 16.5 and 7.7%, respectively. The higher yield and lower fraction of unphysical retrieved profiles for the log(VMR)-based algorithm are both the result of better representation of CO variability in clean regions, as predicted in section 2.2.

[32] The mean surface-level VMR for the log(VMR)-based retrievals (60 ppbv) is substantially smaller than for the corresponding VMR-based retrievals (75 ppbv). As indicated by the results in section 3.3, both the higher convergence rate for the log(VMR)-based algorithm and the difference in a priori profiles contribute to this observed difference. Differences in retrieval statistics for the levels above 500 mb are substantially smaller than for the lower levels. For the lower levels, log(VMR)-based retrievals indicate weaker vertical gradients compared to VMR-based retrievals. This more accurately depicts observed vertical gradients in unpolluted SH oceanic regions [e.g., *Pougatchev et al.*, 1999]. The difference in DFS values supports the prediction in section 2.3 that log(VMR)-based retrievals should exhibit weaker averaging kernels in relatively clean regions compared to VMR-based retrievals.

[33] To compare retrieval results in particularly polluted conditions, all daytime retrievals were extracted for a region of eastern China bounded by 25°N, 35°N, 110°E, and 120°E. For this scene, retrieval yields for both algorithms exceed 99%. Neither algorithm produced significant numbers of unphysical converged retrievals. Retrieved profiles and mean averaging kernels for this subset are compared in Figure 5. Results for both algorithms clearly indicate the sharp vertical gradient in VMR in the lower troposphere which is typical for this region [e.g., *Wang et al.*, 2004]. For the surface level, mean VMR and variability are both substantially larger for log(VMR)-based retrievals. In contrast to the results for the SH oceans, mean DFS in this region is slightly higher for the log(VMR)-based retrievals than for the VMR-based retrievals. This again supports the

idea that log(VMR)-based averaging kernels depend on the VMR, as discussed in section 2.3.

4. Conclusions

[34] Long-term observations in diverse settings indicate that the lognormal model better represents the variability of tropospheric CO than the VMR-normal model. Such an analysis of in situ observations is useful for deciding how to best represent trace gas concentrations in optimal estimation-based retrieval algorithms. In these algorithms, VMR-based and log(VMR)-based state vectors imply fundamentally different a priori constraints and can lead to retrievals with substantially different characteristics (for example, DFS). Benefits of the log(VMR)-based model for CO profile retrievals were verified in a comparison of VMR-based and log(VMR)-based algorithms applied to one day of MOPITT observations. In relatively unpolluted regions, the comparison indicated higher retrieval convergence for the log(VMR)-based algorithm and fewer converged retrievals with unphysically small VMR values.

[35] Even greater benefits are anticipated if the lognormal model is applied in parallel with other algorithm enhancements. For example, future MOPITT products may exploit spatially and/or temporally variable a priori profiles instead of the "global" a priori profile used in the V3 product. Consequently, retrieved profiles will probably exhibit greater variability than in the V3 product. In regions of low VMR (as indicated by the a priori profiles), the convergence problems associated with VMR-based retrievals (discussed in section 2.2) would probably be exacerbated. In contrast, the convergence of log(VMR)-based retrievals (which is greater than 99% when applied with the V3 uniform a priori profile) appears to be much less sensitive to the VMR.

[36] Future MOPITT products may also exploit larger subsets of the observed MOPITT radiances in order to increase retrieval information content. (In the V3 product, radiance bias problems prevented the incorporation of all MOPITT channels into the retrieval product [*Deeter et al.*, 2004a].) As the relative weight of information from the radiances in the retrieved profile increases, the relative weight of the a priori profile in the retrieval decreases. This enhancement, like variable a priori, will tend to increase retrieval variability. Again, the benefits of lognormal statistics should become more evident as the variability of the retrievals increases.

[37] **Acknowledgments.** Archived surface-level measurements of CO concentrations were obtained from the National Oceanic and Atmospheric Administration's Global Monitoring Division. The authors would like to express their thanks to Paul Novelli and colleagues for making these data available. The authors also acknowledge valuable comments made by three anonymous reviewers.

References

Bencala, K. E., and J. H. Seinfeld (1976), On frequency distributions of air pollutant concentrations, *Atmos. Environ.*, *10*, 941–950.

Bowman, K. W., et al. (2006), Tropospheric emission spectrometer: Retrieval method and error analysis, *IEEE Trans. Geosci. Remote Sens.*, *44*, 1297–1307.

Deeter, M. N., et al. (2003), Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, *J. Geophys. Res.*, *108*(D14), 4399, doi:10.1029/2002JD003186.

Deeter, M. N., et al. (2004a), Evaluation of operational radiances for the Measurements of Pollution in the Troposphere (MOPITT) instrument thermal band channels, *J. Geophys. Res.*, *109*, D03308, doi:10.1029/2003JD003970.

Deeter, M. N., et al. (2004b), Vertical resolution and information content of CO profiles retrieved by MOPITT, *Geophys. Res. Lett.*, *31*, L15112, doi:10.1029/2004GL020235.

Edwards, D. P., C. M. Halvorson, and J. C. Gille (1999), Radiative transfer modeling for the EOS Terra satellite Measurements of Pollution in the Troposphere (MOPITT) instrument, *J. Geophys. Res.*, *104*, 16,755–16,775.

Edwards, D. P., et al. (2006), Satellite-observed pollution from Southern Hemisphere biomass burning, *J. Geophys. Res.*, *111*, D14312, doi:10.1029/2005JD006655.

Emmons, L. K., et al. (2004), Validation of Measurements of Pollution in the Troposphere (MOPITT) CO retrievals with aircraft in situ profiles, *J. Geophys. Res.*, *109*, D03309, doi:10.1029/2003JD004101.

Engelen, R. J., and G. L. Stephens (1999), Characterization of water-vapour retrievals from TOVS/HIRS and SSM/T-2 measurements, *Q.J.R. Meteorol. Soc.*, *125*, 331–351.

Gregory, G. L., H. E. Fuelberg, S. P. Longmore, B. E. Anderson, J. E. Collins, and D. R. Blake (1996), Chemical characteristics of tropospheric air over the tropical South Atlantic Ocean: Relationship to trajectory history, *J. Geophys. Res.*, *101*, 23,957–23,972.

Hauglustaine, D. A., et al. (1998), MOZART, a global chemical transport model for ozone and related tracers: 2. Model results and evaluation, *J. Geophys. Res.*, *103*, 28,291–28,335.

Karlsdottir, S., I. S. A. Isaksen, G. Myhre, and T. K. Berntsen (2000), Trend analysis of O3 and CO in the period 1980–1996: A three-dimensional model study, *J. Geophys. Res.*, *105*, 28,907–28,934.

Livesey, N. J., W. Van Snyder, W. G. Read, and P. A. Wagner (2006), Retrieval algorithms for the EOS Microwave Limb Sounder, *IEEE Trans. Geosci. Remote Sens.*, *44*, 1144–1155.

Novelli, P. C., et al. (1998), An internally consistent set of globally distributed atmospheric carbon monoxide mixing ratios developed using results from an intercomparison of measurements, *J. Geophys. Res.*, *103*, 19,285–19,293.

Pougatchev, N. S., G. W. Sachse, H. E. Fuelberg, C. P. Rinsland, R. B. Chatfield, V. S. Connors, N. B. Jones, J. Notholt, P. C. Novelli, and H. G. Reichle Jr. (1999), Pacific Exploratory Mission-Tropics carbon monoxide measurements in historical context, *J. Geophys. Res.*, *104*, 26,195–26,207.

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

Wang, Y. X., M. B. McElroy, T. Wang, and P. I. Palmer (2004), Asian emissions of CO and NO sub (x): Constraints from aircraft and Chinese station data, *J. Geophys. Res.*, *109*, D24304, doi:10.1029/2004JD005250.

M. N. Deeter, D. P. Edwards, and J. C. Gille, Atmospheric Chemistry Division, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA. (mnd@ucar.edu)