# MOPITT (Measurements of Pollution in the Troposphere) Validated *Version 4* Product User's Guide

Merritt N. Deeter*

MOPITT Algorithm Development Team

Atmospheric Chemistry Division

National Center for Atmospheric Research

Boulder, CO 80307

Last Revised May 28, 2009

----

\* Correspond to mnd@ucar.edu

# **Table of Contents**

# 1  Introduction

This guide describes the MOPITT Version 4 (V4) Level 2 and Level 3 products for tropospheric carbon monoxide (CO) and discusses appropriate analysis methods. Users will find that the V4 product is significantly advanced compared to the Version 3 (V3) product, exploits the results of many recent instrument and data analyses, and is formatted differently.  Therefore, all users of the new V4 product are encouraged to familiarize themselves with this document.

# 2  Review of the MOPITT Version 3 Product

The MOPITT V3 product became available in 2000 and was the first satellite dataset for tropospheric CO featuring global coverage.  Characteristics of this product included

- Retrieval algorithm based on *optimal estimation* principles to combine measurements with statistical background information ('a priori')

- Retrievals expressed as profiles of CO volume mixing ratio (VMR) values on fixed 7-level pressure grid

- Retrievals all based on same "global" CO VMR a priori profile and covariance matrix

- Retrievals based on subset of MOPITT thermal-infrared (TIR) radiances

- Operational radiative transfer model (MOPFAS) based on pre-launch estimates and laboratory measurements of instrument parameters and training set mainly composed of low-to-moderate CO concentrations

Details of the V3 retrieval algorithm and underlying radiative transfer model were provided in [1] and [2], respectively.  V3 data processing algorithms were developed before launch and relied on projected in-orbit instrument parameters (such as modulation cell temperatures and pressures).  To clearly distinguish features in the actual measurements from a priori information, this product was based on a single "global" a priori profile and corresponding a priori covariance matrix.  The V3 product was extensively validated using in-situ CO profiles measured from aircraft over a variety of geographical regions [3,4].  As the longest and most thoroughly validated satellite dataset for CO, the V3 product became the *de facto* standard satellite product for validating CO retrieval products for the AIRS [5], TES [6], SCIAMACHY [7] and ACE-FTS [8] instruments.

# 3  Features of the MOPITT Version 4 Product

For V4, the retrieval algorithm incorporates a variety of features which should enhance its scientific value to most users, but may also require users of the V3 product to adapt their data analysis methods and software.  For example, because CO VMR variability is now modeled in terms of a log-normal distribution (in contrast to the V3 product) [9], random errors in the VMR retrievals due to instrument noise will also follow a log-normal distribution.  Thus, for proper data averaging, CO retrievals should first be converted to log(VMR) values, as explained further in Section 5.1.  Also, users familiar with specific diagnostics in the V3 product, such as "Percent A Priori," will not find the same diagnostics in the V4 product, but will find other more conventional diagnostics, including the retrieval averaging kernels and the Degrees of Freedom for Signal (i.e., the trace of the averaging kernel matrix).  (The

contents of the V3 and V4 Level 2 data products are compared in Appendix 7.1.) Similarly, users will notice that the V4 retrieval grid is finer than the V3 grid, and that some levels in the V3 grid are not present in the V4 grid.

As described in the following sections, the retrieval algorithm used to generate the V4 product benefits from significant advances in radiative transfer modeling, state vector representation, and a priori statistics.  Other aspects of the retrieval algorithm, such as the radiance basis set (i.e., the specific measurements actually used in the retrieval), ancillary data, and cloud detection methods, are largely unchanged from their V3 counterparts.   In the following sections, especially significant changes in the V4 product (with respect to both scientific content and format) are emphasized in bold text.

## 3.1  Radiative Transfer Modeling

The spectroscopic data underlying MOPFAS has been updated for V4. MOPFAS is now based on the HITRAN 2004 database [10], together with the 2006 HITRAN water vapor update. Changes in the forward model radiances (and MOPITT retrievals) due to these spectroscopic updates are generally small.

MOPFAS incorporates models of the physical states of the MOPITT Pressure Modulation Cells (PMCs) and Length Modulation Cells (LMCs) in order to calculate channel radiances. **For V4, both the PMC and LMC models have been updated for consistency with the actual cell pressure (P) and temperature (T) values observed during the mission.** Specifically, values of P and T used to model the LMCs are now time-mean values based on long telemetry timeseries from P and T sensors onboard the instrument. For "Phase 1" observations (from March 2, 2000 to May 7, 2001) acquired before the failure of one of MOPITT's two coolers, the time-mean LMC P and T are computed over the entire Phase 1 period. For "Phase 2" observations (beginning August, 2001 and continuing to the present), the LMC P and T time-means are calculated over 2006. The year 2006 was chosen because it represents the approximate midpoint of the Phase 2 period.  For the PMCs, the Phase 2 P and T cycles are based on 2006 telemetry/sweep data.  The shapes and relative phases of the PMC P and T cycles remain unchanged. For Phase 1, the PMC P and T values used in V4 are the same as those used in V3.

In the V3 retrieval product, MOPITT retrievals failed systematically in strong CO source regions. Subsequent analysis demonstrated that such retrievals were sometimes failing because they fell outside the bounds of the original MOPFAS training set.  MOPFAS is a regression-based forward model trained on a representative set of atmospheres derived from model estimates and observations. Application of MOPFAS to profiles with VMRs outside of the envelope defined by the training set profiles produces unreliable radiances and is therefore prohibited within MOPFAS. **For V4, the MOPFAS training set has been expanded in order to permit retrievals involving especially high CO VMRs.**  Specifically, the original training set composed of 58 profiles used in V3 has been expanded to 116 atmospheres. The added atmospheres were formed by scaling all of the original training set profiles by a factor of two. Retrieval simulations and limited case studies both indicate that this approach yields reliable retrievals of high CO profiles without degrading retrieval quality for low and moderate CO concentrations.

## 3.2  State Vector Representation

The retrieval state vector specifies which quantities are derived from the measurement vector (i.e., the Level 1 radiances).  For both V3 and V4, the state vector includes the discretized CO profile, surface temperature, and surface emissivity [1].  There are significant differences, however, in the V4 representation of the CO profile.  First, whereas the V3 state vector represented the CO vertical profile as

a set of VMR values, **the V4 state vector represents the CO profile as a set of log(VMR) values**. The underlying motivation for this change and its consequences are detailed in [9]. For simplicity, the actual retrieval code relies on base-10 logarithms for log(VMR) calculations; however, the use of natural logarithms or any other logarithm base would produce identical retrieval results (i.e., identical retrieved VMR values). The change to log(VMR) has important implications with respect to appropriate data analysis methods, as described in Section 5. Second, **the V4 state vector expresses the CO profile on a ten-level pressure grid** instead of the seven-level grid used for V3. This change allows somewhat more vertical detail in the a priori profiles to be retained in the retrieved profiles and provides better representation of the MOPITT weighting functions. **The new V4 retrieval grid includes a floating surface level (as in V3) followed by nine uniformly-spaced levels from 900 to 100 hPa**. Employing a retrieval grid with uniform grid spacing (in pressure) substantially simplifies the physical interpretation of the retrieval averaging kernels [11].

## 3.3 Radiance Basis Set

Approximately one and a half years after launch, the MOPITT instrument experienced a cooler failure which disabled Channels 1-4. In response, the MOPITT instrument team revised the operating parameters for Channel 7 to boost signal-to-noise ratio. Since the initial V3 'Phase 1' retrieval product had been based on Channels 1, 3, and 7, the retrieval algorithm was reconfigured to only exploit Channels 5 and 7 [12]. Studies demonstrated that the V3 'Phase 2' product, based on the 5A, 5D, and 7D radiances, was comparable to the earlier Phase 1 product, and suffered only a small decrease in information content. **V4 Products for both Phase 1 and Phase 2 are respectively based on the same radiance sets employed for the V3 Product; these V4 products have been separately validated. Nevertheless, because of the different retrieval characteristics of Phase 1 and Phase 2 retrievals (as quantified by the averaging kernels), particular caution should be exercised by users exploiting both Phase 1 and Phase 2 MOPITT products.** (Incorporation of the 6A and 6D solar-reflectance radiances as well as the 7A thermal-infrared radiance are primary objectives in the development of the next MOPITT product.)

## 3.4 A Priori

The MOPITT retrieval algorithm is based on the Maximum a Posteriori solution [1], which depends explicitly on both the a priori state vector, which represents the statistically most probable state, and the a priori covariance matrix, which describes the statistical variability relative to that state. A priori information is required for all elements of the state vector, i.e., the CO profile, surface temperature, and surface emissivity.

### 3.4.1 CO

Results from the global chemical transport model MOZART-4 were used to create a monthly mean climatology for use as the V4 MOPITT a priori. MOZART-4 (Model for Ozone and Related chemical Tracers, version 4) simulates 100 chemical species with relatively detailed hydrocarbon chemistry [13] (also see http://gctm.acd.ucar.edu/). The simulation used for this climatology is for 1997-2004, with a horizontal resolution of T42 (2.8 x 2.8 deg) and 28 vertical levels, driven by meteorology from NCEP/NCAR reanalyses. The anthropogenic emissions were constant over the simulation, but biomass burning emissions were based on satellite fire counts for each month from the Global Fire Emissions

Database (GFED-v2) [14].  The climatology was created by making monthly means over the 8 years of simulation, retaining the horizontal resolution and interpolating in the vertical to the MOPITT Forward Model 35 levels.  For each retrieval, the climatology for the same month as the observation is spatially interpolated to the location of the observation.  **V4 a priori profiles, which vary geographically and temporally, are reported in the V4 Level 2 product for each retrieval.**

The CO a priori covariance matrix describes the expected variability and inter-level correlations of the CO profile; this matrix quantifies the applied constraint in the optimal estimation retrieval algorithm. Diagonal elements of this matrix quantify variances at particular pressure levels in the retrieval grid.  Off-diagonal covariances describe correlations of CO variability for each pair of levels.  In the V3 retrieval algorithm, where the retrieval state vector represented CO in terms of VMR, the CO a priori covariance matrix quantified VMR variability.  In contrast, **the V4 a priori covariance matrix describes variability of log(VMR)**.  Thus, because

$$\partial(ln(VMR)) = \partial(VMR)/VMR \tag{1}$$

it is clear that log(VMR) variances and covariances describe *fractional* VMR variability rather than *absolute* VMR variability.  As shown in [9], analysis of in-situ datasets for geographically diverse locations demonstrates that fractional VMR variability is more consistent (i.e., varies less from site to site) than absolute VMR variability.  For example, regions which are typically unpolluted (on average) exhibit smaller VMR variances than CO source regions which exhibit larger mean VMR values.  Thus, although log(VMR) variability statistics are not well established for all geographic regions and all seasons, the results of [9] support the use of a single "global" a priori covariance matrix for CO retrievals based on a log(VMR) state vector.

**The V4 a priori covariance matrix $C$ incorporates the same variance value for CO at all levels, with a constant correlation height $P_c$ (expressed in pressure units) defining the off-diagonal elements.** Thus,

$$C_{ii} = C_0 \tag{2}$$

and

$$C_{ij} = C_0 \, exp[-(p_i - p_j)^2/P_c^2] \tag{3}$$

where $p_i$ and $p_j$ are the pressures for retrieval levels $i$ and $j$, respectively.  **For V4, $C_0$ is set to $(0.30 \, log_{10}e)^2$, corresponding to a fractional VMR variability of 30%, and $P_c$ is set to 100 hPa**.  This relatively small value for $P_c$ reduces the projection (or extrapolation) of information from levels where the MOPITT weighting functions are relatively strong (e.g., the mid-troposphere) to levels where the weighting functions are relatively weak (e.g., the surface); this phenomenon was often apparent in the V3 product.  While approximate, these values for $C_0$ and $P_c$ are generally consistent with analyses of aircraft in-situ datasets at individual MOPITT validation sites.

### 3.4.2  Surface Parameters

A priori values and variances are also required for surface temperature and surface emissivity.  For surface temperature, both V3 and V4 a priori values are based on spatially and temporally interpolated surface air temperatures from NCEP (NOAA's National Centers for Environmental Prediction).  The assumed surface temperature a priori variance is 25 $K^2$, corresponding to an uncertainty of 5 K. **Unlike V3, however, V4 surface emissivity a priori values are based on an analysis of gridded MOPITT**

**radiances and corresponding MODIS surface temperatures [15]. The assumed surface emissivity a priori variance for V4 is 0.0025, corresponding to an uncertainty of 0.05.** This variance value is a factor of 10 smaller than the V3 value.

### 3.5 *Meteorological Data*

For each MOPITT observation (i.e., a single "pixel"), the retrieval algorithm requires profiles of both temperature and water vapor. For both the V3 and V4 products, these profiles are spatially and temporally interpolated from NCEP operational analysis products. Beginning on June 1, 2005, the MOPITT team observed a sharp global decrease in the number of successful retrievals in the V3 product. It was discovered that the problem was the result of a sudden increase in the occurrence of NCEP water vapor profiles including one or more anomalously small relative humidity (RH) values (exactly zero). In the original V3 processing software, such RH profiles were rejected and caused the entire observation to be discarded.

A corrective scheme was then developed to recover the MOPITT retrievals which were rejected by the original V3 retrieval algorithm; all V3 MOPITT retrievals after June 1, 2005 were processed with this correction applied. **For V4 processing, this corrective algorithm for water vapor is applied to the entire record of MOPITT observations.** As described below, the new scheme exploits the NCEP/NCAR Reanalysis long-term monthly-mean water vapor product (obtained at www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml) in cases where the standard NCEP operational water vapor profiles appear to be unphysical.

Water vapor profiles used in V4 retrieval processing are generated by the following method. First, for each MOPITT pixel, the latitude, longitude, and observation time are used to identify the eight NCEP water vapor profiles required for three-dimensional (3D) interpolation. For each of the eight NCEP profiles, all pressure levels where the RH values are exactly zero are then identified. If these anomalies occur at pressure levels of 250 mb or higher (i.e., the upper troposphere or lower stratosphere), the water vapor mixing ratio for that level is replaced by the mixing ratio value at the next lowest valid (non-zero) level in the profile. If these anomalies occur at pressure levels below the 250 mb level (i.e., the mid- or lower troposphere), the entire profile is replaced by the water vapor profile from the long-term monthly mean climatology for the same latitude and longitude. The final water vapor profile is produced by 3D linear interpolation on the resulting eight profiles. Generally, these eight profiles may include both operation NCEP profiles (possibly corrected in the upper troposphere or lower stratosphere) and climatology profiles.

**Users of the MOPITT V4 product who are interested in separating retrievals based purely on standard NCEP profiles from retrievals incorporating water vapor climatology can do so by exploiting a new retrieval diagnostic included in the Level 2 product.** The new diagnostic ("Water Vapor Climatology Content") is calculated during the 3D interpolation for each MOPITT pixel as the sum of the interpolation weighting factors for all climatology profiles. The Water Vapor Climatology Content varies from 0.0 (for retrievals based exclusively on operational NCEP water vapor profiles) to 1.0 (for retrievals based entirely on water vapor climatology).

Earlier comparisons of V3 retrievals processed with both the original and water vapor-corrected retrieval algorithms indicated some general differences. For example, for early June, 2005, the corrected product produced significantly better global coverage (by 20-25%) than the original V3 product. For MOPITT observations leading to successful retrievals in both versions, retrieval differences in the middle and

upper troposphere tended to be fairly small (usually less than 10 ppbv). However, in a few isolated regions where the influence of water vapor climatology was particularly strong (as indicated by Water Vapor Climatology Content values near unity), substantial differences in surface-level volume mixing ratios were sometimes observed. In these cases, differences in retrieved CO total column of up to 10-20% were observed.

### 3.6 Retrieval Convergence

**Retrieval convergence in the V4 Level 2 product is generally higher than in the V3 product.** For V3, failed retrievals commonly followed one of two paths, characterized by regions of exceptionally low and high VMR. At the low extreme, the assumption of a VMR-normal probability distribution function in V3 occasionally led to negative VMR values after inversion by the maximum a posteriori retrieval algorithm. Negative VMR values are unphysical (and incompatible with MOPFAS) and would immediately result in a failed retrieval. Conversely, in extremely polluted conditions, V3 retrievals would sometimes exceed the upper limit of MOPFAS (as described in Section 3.1) and fail. **In V4, these convergence problems are substantially reduced through (1) the use of log(VMR) in the retrieval state vector, which prevents negative VMR values, and (2) an expanded training set for MOPFAS, allowing retrievals with higher values than in V3 (up to a factor of two).**

However, even in the V4 product, a small fraction of retrievals may still fail. Over a full day, observed retrieval failure rates vary from a few tenths of a percent to a few percent. The majority of these retrievals fail to converge within the maximum allowed number of iterations (see next paragraph). Retrievals failing this way tend to occur in isolation, i.e., are not clustered spatially. A smaller number of retrievals fail because they yield retrieved profiles outside of the acceptable range of MOPFAS, the same problem that occurred in V3. Limited analysis suggests that this effect occurs in rare cases in daytime observations characterized by very high CO VMRs and yields small clusters of failed retrievals.

The criteria used to determine retrieval convergence in both V3 and V4 is based on the root-mean-square change (over the ten-level profile) of log(VMR) (base 10), in comparison with the previous iteration. If this value is 0.01 or less, the retrieval is considered to be converged and written to output. If the value exceeds 0.01, iterations continue. In practice, retrievals usually converge in four iterations or less. The maximum allowable number of iterations is set at 20.

### 3.7 Cloud Detection

Prior to the retrieval algorithm, the location of a particular MOPITT observation is determined to be either clear or cloudy. Observations determined to be cloudy are not processed further. This clear/cloudy determination is based on both the MOPITT radiances themselves and a "cloud mask" produced from near-simultaneous observations by the Terra/MODIS ("MODerate resolution Imaging Spectroradiometer") instrument. **The cloud detection module ("MOPCLD") for V4 is unchanged from V3.** Possible outcomes of the cloud detection module (with corresponding "Cloud Description" values in the L2 product in parentheses) include (1) "MOPCLD only clear, thermal only," (2) "MOPCLD and MODIS cloud mask agree on clear," (3) "MODIS cloud mask only clear," (4) "MOPCLD overriding MODIS cloud mask over low clouds" and (5) "MODIS cloud mask only, clear over polar regions."

## 4  Product Format

## 4.1  Level 1 Data

The format of the HDF files containing the MOPITT calibrated radiances, i.e., the Level 1 data product, is unchanged for V4.  However, the actual radiance values contained in these files have changed, as the result of changes in the V4 instrument model and operational radiative transfer model (MOPFAS).

## 4.2  Level 2 Data

Like the previous V3 product, the MOPITT V4 Level 2 data product is stored in HDF-4 format data files. **However, MOPITT data users will notice numerous differences between the V3 and V4 products, which are summarized in Appendix 7.1.** Some V4 data products and diagnostics, such as the "A Priori CO Mixing Ratio Profile" and "Retrieval Averaging Kernel Matrix," were absent in V3. Other parameters, such as the "Retrieved Surface Temperature" have been renamed for greater clarity, and to distinguish a priori values from retrieved parameters.

As described in Section 3.2, V4 retrieved profiles are expressed on a ten-level grid.  Retrieved profiles are separated into (1) the "Retrieved CO Surface Mixing Ratio" and (2) the "Retrieved CO Mixing Ratio Profile," formed by the retrieved CO VMR at the nine fixed-level pressures (beginning with the level closest to the surface). **Corresponding to each ten-level retrieved profile is a ten-by-ten "Retrieval Averaging Kernel Matrix."  Users should note that V4 averaging kernels describe the sensitivity of retrieved log(VMR) to actual log(VMR), and are thus not exactly equivalent to V3 VMR-based averaging kernels.  Nevertheless, as explained in Section 5.2, the provided averaging kernels are the correct averaging kernels to use when comparing MOPITT retrievals to in-situ measurements or model output.** The reported averaging kernels, unlike the retrieved profile, are not divided into separate surface and fixed-level components.  Thus, the first row of the reported averaging kernel matrix describes the sensitivity of retrieved log(VMR) values at the surface to actual log(VMR) changes at all ten levels in the retrieval grid (beginning with the surface level).

For MOPITT retrievals where the surface pressure is less than 900 hPa, missing values will appear in both the "Retrieved CO Mixing Ratio Profile" and "Retrieval Averaging Kernel Matrix" at all unrealized levels. For example, retrievals where the surface pressure is between 600 and 700 hPa will exhibit missing values at 900, 800, and 700 hPa. Users must reformat these products to extract the physically meaningful elements of the L2 products.

**In addition to the averaging kernels, new diagnostics in the V4 L2 product also include "Degrees of Freedom for Signal" (an index for retrieval information content), "Water Vapor Climatology Content" (described in Section 3.5), and "Retrieval Iterations."  The new "Surface Index" parameter replaces the "Surface Indicator" parameter in V3, and has been simplified. Surface Index values of 0, 1, and 2 correspond to water, land, and mixed (e.g., coastline), respectively.**

Dimensions of multi-dimensional L2 products listed in Appendix 7.1 are ordered to be consistent with the IDL programming language, as demonstrated in Appendix 7.3.  Other tools and programming languages may reverse the ordering.  For example, IDL commands to read the L2 field 'Retrieved CO Mixing Ratio Profile' will yield a three-dimensional array with the dimensions (2,9,ntime) where ntime is the number of  retrievals in a given Level 2 HDF file. In contrast, the HDFview utility will list the dimensions of  this field as (ntime,9,2).  In general, the order of the dimensions will be self-evident, except for the 'Retrieval Averaging Kernel Matrix' field.  Since this matrix has the same number of rows and columns, users should be particularly careful about observing the correct row/column ordering.  Specifically, when read by IDL, the L2 'Retrieval Averaging Kernel Matrix' field must be interpreted as having the dimensions

(nrow, ncolumn, ntime).  Users of the V4 product who have questions regarding array dimensioning using particular operating systems and analysis tools are encouraged to contact the MOPITT/NCAR team.

## 4.3  Level 3 Data

The Level 3 Product is created by gridding MOPITT Level 2 data on a 1 degree latitude/longitude grid on both daily and monthly time scales.  MOPITT Level 3 data files are produced in HDF format for both individual days and individual months, providing daily-mean and monthly-mean CO distributions at a resolution of 1 degree (latitude and longitude) respectively. Daytime and nighttime MOPITT observations are processed separately in the Level 3 processing (based on solar zenith angle), and lead to separate daytime and nighttime products within each Level 3 HDF file.  Retrieval sensitivity is generally greater for daytime overpasses than for nighttime overpasses, particularly over land.

**Level 3 products are listed in Appendix 7.2.** Generally, there exists a corresponding L3 product for each data field in the Level 2 Product. In addition, the number of Level 2 retrievals (or 'Number of Pixels') used as the basis of each Level 3 gridded value is also provided.  Moreover, for each retrieved parameter, additional fields provide (1) the mean uncertainty of the Level 2 values and (2) the variability of the Level 2 values as represented by the standard deviation.

In rare special situations, quality-control considerations prevent inclusion of all the Level 2 data within a particular grid cell for producing Level 3 gridded values. For example, for grid cells overlapping coastline, Level 2 data within that grid cell will often include more than one surface type.  Averaging Level 2 data characterized by significantly different averaging kernels, as would occur in this case, should generally be avoided.  Other special cases are listed below, together with descriptions of the methods by which such data are processed.

If any single surface type constitutes at least 75% of the Level 2 retrievals in a particular grid cell, only retrievals with that surface type are used as the basis of the Level 3 gridded value.  If no single surface type constitutes 75% of the available Level 2 retrievals, all of the Level 2 data are retained in the L3 gridded value, and the surface index is set to 2 ('Mixed').

If the L2 retrievals in a particular L3 grid cell contain varying numbers of valid levels (usually because of varying topography), it is determined which case (i.e., number of valid levels) occurs most frequently. Only this subset of L2 retrievals are retained as the basis of the L3 gridded value.

As for the L2 products listed in Appendix 7.1, dimensions of multi-dimensional L3 products listed in Appendix 7.2 are ordered to be consistent with the IDL programming language.  Other tools and programming languages may reverse the ordering.  In general, the order of the dimensions will be self-evident, except for the 'Retrieval Averaging Kernel Matrix' field.  Since this matrix has the same number of rows and columns, users should be particularly careful about observing the correct row/column ordering. **Previous users of the V3 Level 3 Product should also note that averaging kernel matrix rows and columns in the V4 Level 3 Product are switched relative to the V3 convention.  For V4, the dimensions of the four-dimensional Retrieval Averaging Kernel Matrix Day and Retrieval Averaging Kernel Matrix Night products are ordered in IDL as (nrow, ncolumn, nlon, nlat).  For V3, the ordering was (ncolumn, nrow, nlon, nlat).** Users of the V4 product who have questions regarding array dimensioning using particular operating systems and analysis tools are encouraged to contact the MOPITT/NCAR team.

# 5 MOPITT Data Analysis

## 5.1 Data Averaging

Because of instrument noise, individual MOPITT retrieved profiles do not provide a strong basis for scientific analysis. Instead, users should employ statistical methods applied to ensembles of retrievals. Simple data averaging is the most common statistical method for reducing the effects of random instrument noise. The tradeoff to this benefit is reduced spatial and/or temporal resolution. When averaging over an ensemble, retrieval errors associated with random instrument noise decrease as the square root of the number of retrievals in the ensemble.

As described above, the V4 state vector represents CO VMR variability as a log-normal quantity. This has important implications for data averaging. For MOPITT data subsets where the main source of retrieval variability is instrument noise, retrieved VMR values will follow a log-normal distribution. **In such cases, the appropriate method for reducing the effects of instrument noise is to (1) convert individual VMR profiles into log(VMR) profiles, (2) compute the mean, and (3) convert the mean log(VMR) profile back to a VMR profile.** Because of the asymmetry of the log-normal distribution [9], directly computing the mean VMR profile from a set of VMR profiles in such cases will produce a positive bias relative to "truth." In contrast, when CO variability is the dominant source of retrieval variability (such as when averaging over large regions and temporal intervals), directly calculated mean VMR values are most meaningful.

## 5.2 Data Quality and Filtering

A variety of effects besides instrument noise influence MOPITT data quality. Some of these effects are well understood and some are not. For example, the relative weight of the a priori profile $x_a$ in the retrieval $x_{rtv}$, which ideally should be as small as possible, is quantified by the retrieval averaging kernel matrix according to

$$x_{rtv} = x_a + A(x_{true} - x_a) = Ax_{true} + (I - A)x_a \qquad (4)$$

where $x_{true}$ is the true profile, $A$ is the averaging kernel matrix, and $I$ is the identity matrix. Inclusion of a priori information is an effect associated with all remote sensing methods based on optimal estimation. For MOPITT retrievals based on upwelling thermal-infrared radiation, the weighting functions typically yield poor sensitivity near the surface (except in regions of strong thermal contrast). The optimal estimation method exploits a priori information to "fill-in" those parts of the retrieved profile where the actual observations are insensitive. This effect is well understood. **An important consequence of this effect is that, when comparing MOPITT retrieved profiles against in-situ data or model results, users must transform these comparison datasets using Eq. 4, so that the comparison data exhibit the same degree of smoothing and a priori dependence as the MOPITT product [3]. For the V4 product, however, users must make this transformation in terms of log(VMR). An example of IDL code used to extract and apply the V4 averaging kernels is provided in Appendix 7.3.**

Other effects have less quantifiable influence on retrieval data quality. Examples of such effects include forward model error, errors in assumed meteorological quantities, and instrumental variability. In such cases, the unquantified accuracy of the model/product prevents robust estimates of the associated retrieval error. Quantifying such errors is also complicated because they may be highly variable in space and time.

Because of the potential variability of MOPITT data quality, as discussed above, users may wish to exclude particular subsets of the data in their analyses. This is a reasonable strategy, *unless such methods could impose retrieval bias*. For example, in the V4 product, users might consider filtering MOPITT data using the "Degrees of Freedom for Signal" (DFS) parameter, since this diagnostic quantifies information content; larger values of DFS indicate smaller weighting of a priori profile information. However, DFS is derived from the averaging kernel matrix, which, in the optimal estimation retrieval algorithm, depends on the weighting function matrix. Radiative transfer modeling shows that the magnitude of the MOPITT weighting functions varies significantly as the CO concentration changes. Thus, the DFS diagnostic is not independent of the CO profile. <u>Filtering on DFS, or other characteristics of the retrieval averaging kernels, has the potential to impose retrieval bias.</u> If such filters are employed, efforts should first be made to understand the underlying source of DFS variability, and verify that CO variability is not the primary cause.

Alternatively, filters based on geophysical criteria which are known to provide better conditions for remote sensing retrievals are often quite appropriate. For example, over land, daytime conditions typically provide better thermal contrast conditions for thermal infrared-based retrievals than nighttime conditions [11]. Thus, a filter based on solar zenith angle to exclude nighttime MOPITT data should effectively enhance data quality without biasing the resulting subset. Users considering more sophisticated filtering methods (including methods based on analysis of the retrieval averaging kernels) are encouraged to first consult with members of the MOPITT Algorithm Development Team.

## 5.3 *Analysis of Retrieved CO Total Column Values*

For users interested in quantitatively comparing MOPITT retrieved CO total column values with other datasets, special methods are required to obtain the averaging kernel for the retrieved CO total column. These methods are described in Appendix 7.4.

Special diagnostics related to the retrieved CO total column are provided in the L2 field 'Retrieved CO Total Column Diagnostics'. For each retrieval, this two-element vector contains both the estimated smoothing error and measurement error contributions to the total retrieval error (in that order). Smoothing error represents the uncertainty in the retrieved total column due to the departure of the actual total column averaging kernel from the ideal total column averaging kernel. Measurement error describes the uncertainty due to errors in the measured radiances.

# 6 References

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

[2] "Radiative transfer modeling for the EOS Terra satellite Measurement of Pollution in the Troposphere (MOPITT) instrument," D. P. Edwards, C. M. Halvorson, and J. C. Gille. *J. Geophys. Res.,* 104, doi: 10.1029/1999JD900167 (1999).

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

[4] "Measurements of Pollution in the Troposphere (MOPITT) validation exercises during summer 2004 field campaigns over North America," L. K. Emmons, et al., *J. Geophys. Res.*, 112, doi: 10.1029/2006JD007833 (2007).

[5] "A comparison of satellite tropospheric carbon monoxide measurements from AIRS and MOPITT during INTEX-A," J. Warner, et al., *J. Geophys. Res.*, 112, doi:10.1029/2006JD007925 (2007).

[6] "Comparison of carbon monoxide measurements by TES and MOPITT: Influence of a priori data and instrument characteristics on nadir atmospheric species retrievals," M. Luo, et al., *J. Geophys. Res.*, 112, doi:10.1029/2006JD007663 (2007).

[7] "Three years of global carbon monoxide from SCIAMACHY: comparison with MOPITT and first results related to the detection of enhanced CO over cities," M. Buchwitz, et al., *Atmos. Chem. Phys.*, 7:2399-2411, May 2007.

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

[9] "Retrievals of carbon monoxide profiles from MOPITT observations using log-normal a priori statistics," M. N. Deeter, D. P. Edwards, and J. C. Gille, *J. Geophys. Res.*, 112, doi: 10.1029/2006JD007999 (2007).

[10] "The HITRAN 2004 molecular spectroscopic database," L. S. Rothman, et al., *J. Quant. Spectros. and Radiative Transfer*, 96, p. 139-204 (2005).

[11] "Sensitivity of MOPITT Observations to Carbon Monoxide in the Lower Troposphere," M. N. Deeter, et al., *J. Geophys. Res.*, 112, doi:10.1029/2007JD008929 (2007).

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

[13] "Impact of Mexico City emissions on regional air quality from MOZART-4 simulations," L. K. Emmons, et al., submitted to *Atmos. Chem. Physics*, MILAGRO/INTEXB special section (2008).

[14] "Interannual variability in global biomass burning emissions from 1997 to 2004," G. R. van der Werf, et al., *Atmos. Chem. Physics, 6*, 3423-3441 (2006).

[15] "Estimates of 4.7 um surface emissivity and their impact on the retrieval of tropospheric carbon monoxide by Measurements of Pollution in the Troposphere (MOPITT)," S. P. Ho, et al., *J. Geophys. Res.,* 110, doi:10.1029/2005JD005946 (2005).

# 7  Appendices

## 7.1  Comparison of V3 and V4 Level 2 Product Content

| V3 | | V4 | |
|---|---|---|---|
| Label (units) | Dimensions [1] | Label (units) | Dimensions [1] |
| Seconds in Day (s) | ntime | Seconds in Day (s) | ntime |
| Pressure Grid (hPa) | nprs | Pressure Grid (hPa) | nprs |
| Solar Zenith Angle (deg) | ntime | Solar Zenith Angle (deg) | ntime |
| Satellite Zenith Angle (deg) | ntime | Satellite Zenith Angle (deg) | ntime |
| Ancillary Surface Emissivity | ntwo,nw,ntime | A Priori Surface Emissivity | ntwo,ntime |
| Retrieval Bottom Pressure (hPa) | ntwo,ntime | Surface Pressure (hPa) | ntime |
| Derived Effective Retrieval Bottom Temperature (K) | ntwo,ntime | Retrieved Surface Temperature (K) | ntwo,ntime |
| Derived Effective Retrieval Bottom Emissivity | ntwo,nw,ntime | Retrieved Surface Emissivity | ntwo,ntime |
| Surface Indicator [2] | ntime | Surface Index [2] | ntime |
| DEM Altitude (m) | ntime | DEM Altitude (m) | ntime |
| Cloud Description [2] | ntime | Cloud Description [2] | ntime |
| CO Total Column (mol/cm$^2$) | ntwo,ntime | Retrieved CO Total Column (mol/cm$^2$) | ntwo,ntime |
| Retrieval Error Covariance Matrix (ppbv$^2$) | ntwo,np2,ntime | N/A | |
| CO Mixing Ratio (ppbv) | ntwo,nprs,ntime | Retrieved CO Mixing Ratio Profile (ppbv) | ntwo,nprs,ntime |
| Retrieval Bottom CO Mixing Ratio (ppbv) | ntwo,ntime | Retrieved CO Surface Mixing Ratio (ppbv) | ntwo,ntime |
| CO Mixing Ratio Percent Apriori | nprs,ntime | N/A | |
| Retrieval Bottom CO Mixing Ratio Percent Apriori | ntime | N/A | |
| N/A | | A Priori Surface Temperature (K) | ntwo,ntime |
| N/A | | A Priori CO Mixing Ratio Profile (ppbv) | ntwo,nprs,ntime |
| N/A | | A Priori CO Surface Mixing Ratio (ppbv) | ntwo,ntime |
| | | Retrieved CO Total Column Diagnostics | ntwo, ntime |
| N/A | | Retrieval Averaging Kernel Matrix | nprs2,nprs2,ntime [3] |
| N/A | | Degrees of Freedom for Signal | ntime |
| N/A | | Water Vapor Climatology Content | ntime |
| N/A | | Retrieval Iterations [2] | ntime |

[1] Array indices ordered according to IDL convention (see Section 4.2). ntime = number of retrievals (varies), ordered chronologically; nprs = number of fixed pressure levels (not including surface; nprs = 6 for V3 and 9 for V4), starting at fixed level closest to surface; ntwo = 2 (parameter and standard deviation/uncertainty); nprs2 = nprs + 1; obsolete dimensions (used in V3 only): nw, np2

[2] Integer-valued product (all others floating point)

[3] Dimensions of Retrieval Averaging Kernel Matrix are ordered (nrow, ncolumn, ntime)

## 7.2  V4 Level 3 Product Content

| V4 Level 3 Products | |
|---|---|
| Label (units) | Dimensions [1] |
| Latitude (deg) | nlat |
| Longitude (deg) | nlon |
| Pressure Grid (hPa) | nprs |
| Pressure Grid2 (hPa) | nprs2 |
| Solar Zenith Angle Day (deg) | nlon,nlat |
| Solar Zenith Angle Night (deg) | nlon,nlat |
| Satellite Zenith Angle Day (deg) | nlon,nlat |
| Satellite Zenith Angle Night (deg) | nlon,nlat |
| Surface Pressure Day (hPa) | nlon,nlat |
| Surface Pressure Night (hPa) | nlon,nlat |
| Retrieved Surface Temperature Day (K) | nlon,nlat |
| Retrieved Surface Temperature Night (K) | nlon,nlat |
| Retrieved Surface Temperature Mean Uncertainty Day (K) | nlon,nlat |
| Retrieved Surface Temperature Mean Uncertainty Night (K) | nlon,nlat |
| Retrieved Surface Temperature Variability Day (K) | nlon,nlat |
| Retrieved Surface Temperature Variability Night (K) | nlon,nlat |
| Retrieved Surface Emissivity Day | nlon,nlat |
| Retrieved Surface Emissivity Night | nlon,nlat |
| Retrieved Surface Emissivity Mean Uncertainty Day | nlon,nlat |
| Retrieved Surface Emissivity Mean Uncertainty Night | nlon,nlat |
| Retrieved Surface Emissivity Variability Day | nlon,nlat |
| Retrieved Surface Emissivity Variability Night | nlon,nlat |
| Retrieved CO Mixing Ratio Profile Day (ppbv) | nprs,nlon,nlat |
| Retrieved CO Mixing Ratio Profile Night (ppbv) | nprs,nlon,nlat |
| Retrieved CO Mixing Ratio Profile Mean Uncertainty Day (ppbv) | nprs,nlon,nlat |

| V4 Level 3 Products | |
|---|---|
| Retrieved CO Mixing Ratio Profile Mean Uncertainty Night (ppbv) | nprs,nlon,nlat |
| Retrieved CO Mixing Ratio Profile Variability Day (ppbv) | nprs,nlon,nlat |
| Retrieved CO Mixing Ratio Profile Variability Night (ppbv) | nprs,nlon,nlat |
| Retrieved CO Surface Mixing Ratio Day (ppbv) | nlon,nlat |
| Retrieved CO Surface Mixing Ratio Night (ppbv) | nlon,nlat |
| Retrieved CO Surface Mixing Ratio Mean Uncertainty Day (ppbv) | nlon,nlat |
| Retrieved CO Surface Mixing Ratio Mean Uncertainty Night (ppbv) | nlon,nlat |
| Retrieved CO Surface Mixing Ratio Variability Day (ppbv) | nlon,nlat |
| Retrieved CO Surface Mixing Ratio Variability Night (ppbv) | nlon,nlat |
| Retrieved CO Total Column Day (mol/cm$^2$) | nlon,nlat |
| Retrieved CO Total Column Night (mol/cm$^2$) | nlon,nlat |
| Retrieved CO Total Column Mean Uncertainty Day (mol/cm$^2$) | nlon,nlat |
| Retrieved CO Total Column Mean Uncertainty Night (mol/cm$^2$) | nlon,nlat |
| Retrieved CO Total Column Variability Day (mol/cm$^2$) | nlon,nlat |
| Retrieved CO Total Column Variability Night (mol/cm$^2$) | nlon,nlat |
| Retrieval Averaging Kernel Matrix Day | nprs2,nprs2,nlon,nlat [3] |
| Retrieval Averaging Kernel Matrix Night | nprs2,nprs2,nlon,nlat [3] |
| A Priori Surface Temperature Day (K) | nlon,nlat |
| A Priori Surface Temperature Night (K) | nlon,nlat |
| A Priori Surface Emissivity Day | nlon,nlat |
| A Priori Surface Emissivity Night | nlon,nlat |
| A Priori CO Mixing Ratio Profile Day (ppbv) | nprs,nlon,nlat |
| A Priori CO Mixing Ratio Profile Night (ppbv) | nprs,nlon,nlat |
| A Priori CO Surface Mixing Ratio Day (ppbv) | nlon,nlat |
| A Priori CO Surface Mixing Ratio Night (ppbv) | nlon,nlat |
| Degrees of Freedom for Signal Day | nlon,nlat |
| Degrees of Freedom for Signal Night | nlon,nlat |
| Surface Index Day [2] | nlon,nlat |
| Surface Index Night [2] | nlon,nlat |
| Water Vapor Climatology Content Day | nlon,nlat |
| Water Vapor Climatology Content Night | nlon,nlat |
| Number of Pixels Day [2] | nlon,nlat |
| Number of Pixels Night [2] | nlon,nlat |

[1] Array indices ordered according to IDL convention (see Section 4.3).  Standard values for nlon, nlat, nprs and nprs2 are 360, 180, 9, and 10, respectively.

[2] Integer-valued product (all others floating point)

[3] **Note regarding Level 3 Averaging Kernel Matrices**: Previous users of the V3 Level 3 Product should note that averaging kernel matrix rows and columns in the V4 Level 3 Product are switched relative to the V3 convention.  So, for V4, the dimensions of the four-dimensional Retrieval Averaging Kernel Matrix Day and Retrieval Averaging Kernel Matrix Night products are ordered in IDL as (nrow, ncolumn, nlon, nlat). For V3, the ordering was (ncolumn, nrow, nlon, nlat).

## 7.3 Sample IDL code for Applying V4 Averaging Kernels

The following IDL code demonstrates an appropriate method for extracting and applying the MOPITT Level 2 retrieved profiles and averaging kernels, in the context of comparing a single MOPITT retrieval with a single in-situ profile. Note that this example only treats the case of a 'complete' 10-level retrieval. Code revisions would be necessary to treat retrievals containing missing levels (as described in Section 4.2).

```
pro sample_idl_code

; sample idl code demonstrating methods for extracting and applying
; MOPITT V4 retrievals and averaging kernels, and comparing retrieved
; profiles with 'simulated' retrievals based on in-situ profile data
;
; user-specified variables are:
;
; l2file = filename of MOPITT V4 Level-2 hdf file
; inearest = index of matching retrieval in l2file (determined by user based on lat/lon, etc.)
; valvmr = validation in-situ profile in ppb
; valprs = pressure grid corresponding to valvmr (in hPa, not necessarily consistent with MOPITT retrieval grid)
;

l2file = 'MOP02-20020922-L2V8.0.2.val.hdf'

; inearest is closest MOPITT pixel in l2file to 21.2 N, 158.9 W
inearest = 155899

; coarse in-situ CO profile (ppb) observed at 21.2 N, 158.9 W on 9/22/02
;
valvmr = [ 101.92, 101.92, 96.28, 117.72, 86.94, 64.52, 65.69, 83.14, 83.14, 73.40, 53.93]
valprs = [ 1020.,  1000.,  900.,   800.,  700.,  600.,  500.,  400.,  300.,  200.,  100.]

; extract a priori profiles (fixed-level elements and surface value)

approfl = get_sd(l2file,'A Priori CO Mixing Ratio Profile')
apsfcmr = get_sd(l2file,'A Priori CO Surface Mixing Ratio')

; extract retrieved profiles (fixed-level elements and surface value)

rtvprofl = get_sd(l2file,'Retrieved CO Mixing Ratio Profile')
rtvsfcmr = get_sd(l2file,'Retrieved CO Surface Mixing Ratio')
rtvcolm  = get_sd(l2file,'Retrieved CO Total Column')

; extract averaging kernel matrices

avkrn = get_sd(l2file,'Retrieval Averaging Kernel Matrix')

; extract fixed-level pressure grid

file_id = hdf_open(l2file, /read)
prs = get_vd(file_id, 'Pressure Grid')
psfc = get_vd(file_id, 'Surface Pressure')
hdf_close, file_id

; construct 1-D vectors for pressure grid, a priori log(vmr) profile, and retrieved log(vmr) profile

prtv = [psfc(inearest),prs(0:8)]
ap1 = alog10([apsfcmr(0,inearest),reform(approfl(0,0:8,inearest))])
rtv1 = alog10([rtvsfcmr(0,inearest),reform(rtvprofl(0,0:8,inearest))])

; construct 2-D matrix for single averaging kernel matrix

; need to transpose ak1 since FORTRAN stores arrays in 'column major' format, but IDL is 'row major'

ak1 = transpose(reform(avkrn(0:9,0:9,inearest),10,10))

; interpolate validation in-situ profile to MOPITT retrieval grid

valvmrint = interpol(valvmr,valprs,prtv)
logvalvmr = alog10(valvmrint)
```

```
; calculate 'simulated' retrieval; first convert (logvalvmr - ap1) to column vector

simprof = ap1 + ak1 ## transpose(logvalvmr - ap1)

; print out both retrieved log(vmr) profile and 'simulated' log(vmr) retrieval based on in-situ profile

print, 'retrieved log(vmr) profile = '
print, rtv1
print, 'simulated log(vmr) profile = '
print, simprof

; print out both retrieved vmr profile and 'simulated' vmr retrieval based on in-situ profile

print, 'retrieved vmr profile (ppb) = '
print, [rtvsfcmr(0,inearest),reform(rtvprofl(0,0:8,inearest))]
print, 'simulated vmr profile (ppb) = '
print, 10.^(simprof)

return
end

;-------------------------------------------------------------------------
;   Get the Scientific Data information
;-------------------------------------------------------------------------

Function get_sd, filename, name
  sd_id = hdf_sd_start(filename,/read)
  index = hdf_sd_nametoindex(sd_id, name)
  sds_id = hdf_sd_select(sd_id,index)
  hdf_sd_getdata, sds_id, var
  hdf_sd_endaccess, sds_id
  hdf_sd_end, sd_id
  return, var
end

;-------------------------------------------------------------------------
;   Get the Vdata information
;-------------------------------------------------------------------------

Function get_vd,file_id,name
  vd_ref = hdf_vd_find(file_id, strtrim(name,2))
  vdata=hdf_vd_attach(file_id,vd_ref)
  nread=hdf_vd_read(vdata,var)
  hdf_vd_detach, vdata
  return,var
end
```

For the particular parameters (l2file, inearest, etc.) specified in the example, the correct output is

```
IDL> sample_idl_code
% Loaded DLM: HDF.
% Compiled module: INTERPOL.
retrieved log(vmr) profile =
      1.90526      1.97632      2.00192      1.97044      1.89730      1.78686      1.67100      1.60461      1.62432
      1.63398
simulated log(vmr) profile =
      1.89039      1.92080      1.93258      1.92881      1.92111      1.90850      1.89422      1.87761      1.85298
      1.75106
retrieved vmr profile (ppb) =
      80.4011      94.6945      100.443      93.4202      78.9405      61.2155      46.8815      40.2359      42.1033
      43.0509
simulated vmr profile (ppb) =
      77.6938      83.3297      85.6213      84.8806      83.3897      81.0035      78.3828      75.4415      71.2812
      56.3718
```

## 7.4 Calculation of CO Total Column Averaging Kernels

The averaging kernel for the retrieved CO total column is not included in the standard Level 2 product, but may be determined by the user from the retrieved profile averaging kernel matrix. Sample IDL code at the end of this section demonstrates the method in a specific example.

If $C_{rtv}$ is the retrieved CO total column, and x is the CO state vector comprised of log(VMR) values, the total column averaging kernel a is the vector defined by

$$a_j = \partial C_{rtv} / \partial x_{true,j} \qquad (5)$$

Since it can be shown that C is related to the profile of VMR values by the relation

$$C = K \sum \Delta p_i \, VMR_i \qquad (6)$$

where $K \sim 2.12 \cdot 10^{13}$ (mol/cm$^2$)/(hPa ppb) and $\Delta p_i$ is the pressure layer thickness associated with level i, it follows from Eqs. 5 and 6 that

$$a_j = (K / log_{10}e) \sum \Delta p_i \, VMR_{rtv,i} \, A_{ij} \qquad (7)$$

where A is the standard log(VMR)-based averaging kernel matrix defined by

$$A_{ij} = \partial \, log(VMR_{rtv,i}) \, / \, \partial \, log(VMR_{true,j}) \qquad (8)$$

Thus, according to Eq. 7, each element of a is determined by summing the elements of the corresponding column of A, weighted by the absorber amounts.

Then, the total column averaging kernel vector can then be used to 'simulate' MOPITT total column retrievals by the equation

$$C_{rtv} = C_a + a(x_{true} - x_a) \qquad (9)$$

where $C_a$ is the a priori total column value corresponding to the a priori profile $x_a$ (calculated by the user according to Eq. 6).

Determination of the $\Delta p_i$ values needed in Eq. 7 must be made individually for each retrieval because of the variability of the surface pressure. The boundaries of the imaginary layer associated with each level are located at the pressure midpoints between the levels in the grid. If the index of the surface level (the first valid retrieval level) is isfc, then

$$
\begin{aligned}
\Delta p_i &= (p_{sfc} - p_{isfc+1})/2 & &(\text{for } i = isfc) \\
&= (p_{sfc} - p_{isfc+1})/2 + 50 \, hPa & &(\text{for } i = isfc + 1) \qquad (10)\\
&= 100 \, hPa & &(\text{for } isfc + 1 < i \le 10)
\end{aligned}
$$

For example, if $p_{sfc}$ is 850 hPa, isfc would be 2, the retrieval pressure grid would be (850, 800, 700, 600, 500, 400, 300, 200, 100) hPa, and the corresponding $\Delta p$ values would be (25, 75, 100, 100, 100, 100, 100, 100, 100) hPa.

The following section of IDL code demonstrates the calculation and application of the total column averaging kernel using the specific example presented in Appendix 7.2 (for which isfc = 1) as a starting point. Code revisions would be necessary to treat retrievals containing missing levels (as described in Section 4.2).

```
; calculate simulated total column (after calc'g total colm averaging kernel)

; example assumes psfc > 900 hPa, and that ap1, rtv1, and and ak1
; have been previously defined

dp = fltarr(10)
dp(0) = (psfc(inearest) - 900.)/2.
dp(1) = (psfc(inearest) - 900.)/2. + 50.
dp(2:9) = 100.

log10e = alog10(exp(1.))

; calc a priori total column

colm_ap = 2.12e13 * ( dp ## transpose(10.^ap1) )

; calculate and apply total column averaging kernel

ak_colm = (2.12e13/log10e) * ( (dp * (10.^rtv1)) ## ak1 )

colm_sim = colm_ap + ( ak_colm ## transpose(logvalvmr - ap1) )

; print out both retrieved total colm and 'simulated' total colm based on in-situ profile

print, 'retrieved total column (mol/cm2) = '
print, rtvcolm(0,inearest)
print, 'simulated total column (mol/cm2) = '
print, colm_sim
```

When this code is added to the end of the IDL procedure listed in Appendix 7.2 (just before the first 'return' statement) and executed, the proper output is

```
retrieved total column (mol/cm2) =
  1.39876e+18
simulated total column (mol/cm2) =
  1.60962e+18
```