A Multi-Layer Technique for Microwindow Selection
PO-TN-OXF-GS-0013
Task 2.1, CCN4 11886/96/NL/GS

Anu Dudhia*

September 21st, 1999

1 Introduction

This document describes a method for constructing MIPAS microwindows considering a 'multi-layer' retrieval. As such, it represents a modification of the scheme developed under contract 12054 which constructed microwindows assuming a 'single-layer' retrieval.

The operational requirement is for microwindows which are usable over a range of tangent altitudes. With microwindows defined using the single (tangent) layer scheme, this requires an additional 'consolidation' step where microwindows optimised for a particular tangent altitude are extended, non-optimally, to cover other tangent altitudes. It is then necessary to define an occupation matrix algorithm to determine the combination of microwindows that gives the required accuracy for the retrieved profile.

Using a multi-layer retrieval allows microwindows boundaries to be optimised over all levels simultaneously, removing the consolidation step. It also allows microwindows to be selected sequentially so that each minimises the remaining profile retrieval error: the appropriate occupation matrix for n microwindows is then simply microwindows 1 ... n in the sequence. A further advantage is the capability of including inter-level correlations, both in the retrieval and in the systematic errors, which can have a significant effect on the apparent accuracy.

2 Theory

2.1 Contribution Function

Consider an existing estimate of a state vector (i.e. profile) $\mathbf{a}$ of n elements, with covariance $\mathbf{S}_a$, obtained from a set of m measurements $\mathbf{y}$ using a 'Global Fit' (GF) retrieval (Appendix A).

*Atmospheric, Oceanic and Planetary Physics, Oxford University, UK

It can be shown (Appendix C) that rerunning the GF retrieval with an *additional* measurement y, variance σ^2, would modify the retrieval $\mathbf{a} \rightarrow \mathbf{x}$:

$$\mathbf{x} = \mathbf{d}y + (\mathbf{I}_n - \mathbf{dk})\mathbf{a} \quad (1)$$

where $\mathbf{k}$ is the weighting function for y ($y = \mathbf{ka}$) and $\mathbf{d}$ is its contribution function:

$$\mathbf{d} = \mathbf{S}_a\mathbf{k}^T \left(\sigma^2 + \mathbf{kS}_a\mathbf{k}^T\right)^{-1} \quad (2)$$

Note that this equivalent to the Optimal Estimation (OE) contribution function, Eq. 24. This is only 'optimal' in the sense of combining the measurement y with a previous estimate $\mathbf{a}$ weighted by their random errors σ^2, $\mathbf{S}_a$, which are assumed uncorrelated. It is no longer optimal if additional or correlated (between y and $\mathbf{a}$) sources of error are present. However, since it is the weighting that will be used in the operational Global Fit retrieval, Eq. (2) is the appropriate contribution function to use in the total error analysis.

2.2 Error Covariances

Taking the covariance of Eq. (1) (see Appendix B) the additional measurement will modify the random error covariance $\mathbf{S}_a \rightarrow \mathbf{S}_x$:

$$\begin{aligned} \mathbf{S}_x &= (\mathbf{I}_n - \mathbf{dk})\mathbf{S}_a & (3) \\ &= \mathbf{S}_a - \mathbf{S}_a\mathbf{k}^T \left(\sigma^2 + \mathbf{kS}_a\mathbf{k}^T\right)^{-1} \mathbf{kS}_a & (4) \end{aligned}$$

However, there will also be systematic (non-random) error terms, e.g. due to underestimating a contaminating species. Suppose a particular systematic error source i contributes a forward model error δy^i to the measurement y. Since the previous estimate $\mathbf{a}$ is based on m similar measurements, it will also contain an error component $\delta\mathbf{a}^i$ representing the accumulation of forward model errors from the same source. These two error terms are completely correlated, so Eq. 1 gives the resulting retrieval error component $\delta\mathbf{x}^i$:

$$\delta\mathbf{x}^i = \mathbf{d}\delta y^i + (\mathbf{I}_n - \mathbf{dk})\,\delta\mathbf{a}^i \quad (5)$$

This error component will have a covariance $\mathbf{S}_x^i$:

$$\mathbf{S}_x^i = E\left\{(\delta\mathbf{x}^i)(\delta\mathbf{x}^i)^T\right\} = (\delta\mathbf{x}^i)(\delta\mathbf{x}^i)^T \quad (6)$$

Assuming that each of the error sources i is uncorrelated with the other error sources, the *total* error covariance of the retrieval $\bar{\mathbf{S}}_x$ is then the sum of the random and correlated error covariances:

$$\bar{\mathbf{S}}_x = \mathbf{S}_x + \sum_i \mathbf{S}_x{}^i \quad (7)$$

2.3 Information Content

If the previous estimate is known with an accuracy represented by the total error covariance matrix $\bar{\mathbf{S}}_a$, and after adding measurement y this is modified to $\bar{\mathbf{S}}_x$, the information content H of the retrieval (i.e. of the additional measurement) can be defined as:

$$H = -\log_2 \left|\bar{\mathbf{S}}_x \bar{\mathbf{S}}_a^{-1}\right| \quad (8)$$

where H is measured in bits. This properly accounts for any off-diagonal elements in both covariance matrices. For $H > 0$, the measurement has had a positive contribution. Note that if the retrieval were performed 'optimally', allowing for all error terms in the previous estimate and the measurement, together with their correlations, the information content would never be negative. However, since the operational retrieval only allows for random noise components, an additional measurement can lead to $H < 0$ in which case the measurement should be excluded.

For a single parameter retrieval this expression reduces to the ratio of the total variances, $H > 0$ being equivalent to $\bar{S}_x < \bar{S}_a$, which is the test used in the single-layer algorithm.

2.4 Updating

If a measurement is accepted as part of the retrieval (i.e. contributes 'positive' information) the error components are updated according to Eqs. (3)–(7) and then used to redefine the 'previous estimate' when considering the next measurement:

$$(\sigma^2, \mathbf{S}_a) \rightarrow \mathbf{S}_x = \mathbf{S}_a{}' \quad (9)$$

$$(\delta y^i, \delta\mathbf{a}^i) \rightarrow \delta\mathbf{x}^i = \delta\mathbf{a}^{i\prime} \quad (10)$$

$$(\mathbf{S}_x, \delta\mathbf{x}^i) \rightarrow \bar{\mathbf{S}}_x = \bar{\mathbf{S}}_a' \quad (11)$$

where $'$ denotes the new 'previous estimate' error terms now incorporating $m + 1$ measurements.

2.5 Initialisation

It is not possible to perform a Global Fit retrieval (Eq. 20) of n parameters (e.g. VMR at 16 levels) using $m < n$ measurements. However, Eq. 1 *can* be applied even for $m = 0$ provided that an initial random *a priori* covariance $\mathbf{S}_a^0$ is defined, e.g. some large 'climatological' uncertainty. This makes it possible to define the information content of a single measurement rather than have to preselect $m = n$ measurements before being able to calculate the contribution of the $n + 1$th measurement.

Effectively the algorithm starts off as an 'Optimal Estimation' retrieval and, once a sufficient number of measurements are included, converges towards a 'Global Fit' retrieval as the error covariance converges to that of the measurements alone.

3 A Practical Scheme

3.1 *A Priori* Covariance

The form chosen for the *a priori* covariance $\bar{\mathbf{S}}_a^0$ will influence the selection sequence of the first few microwindows.

The suggestion is to initialise with large variances at all altitudes, e.g. a diagonal matrix with variances corresponding to a factor 100 uncertainty in the climatological value. This should then lead to a 'natural' microwindow selection sequence where the maximum information is obtained over the whole profile.

An alternative is to select a small initial variance in a certain altitude range. Any microwindows which only retrieve in this altitude range will then appear to contribute little additional information, so will not be selected. This can be used to bias the selection away from certain altitude ranges, e.g. if the 'natural' selection produces too many high altitude microwindows.

3.2 State Vector

The state vector $\mathbf{x}$ is represented by 25 elements containing:

$x_1 - x_{16}$ Volume Mixing Ratio at 8, 11, ..., 53 km
$x_{17} - x_{24}$ Continuum at 8, 11, ..., 29 km.
x_{25} Radiometric Offset

The continuum and offset terms are specific to each microwindow, and assumed fully independent from one microwindow to the next. Therefore the full state vector need only represent these components for the current microwindow. The full state vector is used to calculate the total error covariances, but only VMR elements (1–16) are used to test information content since these are the output parameters.

3.3 Weighting Functions

The weighting functions k_{ij} for the sensitivity of measurement y_i to the profile at tangent height j can be calculated in one of three ways (in order of increasing complexity):

1. Assume the measurement is only sensitive to the atmosphere at the tangent point (k_{jj} precomputed, $k_{ij} = 0$ if $i \neq j$)
2. Calculate sensitivity to higher altitudes by scaling tangent points values for those altitudes by airmass ($k_{ij} \propto k_{jj}$).
3. Calculate weighting functions k_{ij} explicitly for each measurement.

Values of k_{ii}, k_{jj} are obtained from the existing Jacobian spectra (one per band, per tangent height, per absorbing species). Option 3 would require an associated forward model to calculate $\mathbf{k}$ explicitly as the microwindow construction proceeds since it is impractical to precalculate k_{ij} for all i, j for full spectra.

3.4 Error Sources

A separate error source is defined for:

- each contaminant profile
- temperature for each altitude
- pressure for each altitude
- gain for each microwindow
- HITRAN for each microwindow
- line position for each microwindow

The corresponding measurement errors δy^i are calculated, as in the single-layer scheme, using the existing jacobian spectra. These are combined with the stored previous estimate errors $\delta \mathbf{a}^i$ using Eq. (5), to produce retrieval errors $\delta \mathbf{x}^i$ which are then converted into to covariances $\mathbf{S}_x^i$ (Eq. 6) for the information content analysis. If the measurement is to be included in the microwindow, the updated error vectors are saved $\delta \mathbf{x}^i \rightarrow \delta \mathbf{a}^{i'}$, otherwise the previous error vectors $\delta \mathbf{a}^i$ are kept.

3.5 MW-Specific Errors

When each microwindow is completed, it is assumed that microwindow-specific error sources (e.g. gain error) will not be correlated with any further measurements. Rather than continue to maintain separate error vectors $\delta \mathbf{x}^i$ for these components, it may be computationally more convenient to combine these into an *uncorrelated* component of *a priori* covariance $\mathbf{S}_a^u$ for the start of the next microwindow. This will then map into subsequent retrievals in the same way as the random uncorrelated error $\mathbf{S}_a$ (Eq. 3), giving a modified version of Eq. (7):

$$\bar{\mathbf{S}}_x = \mathbf{S}_x + \mathbf{S}_x^u + \sum_i \mathbf{S}_x^i \qquad (12)$$

where

$$\mathbf{S}_x^u = (\mathbf{I}_n - \mathbf{dk})\mathbf{S}_a^u(\mathbf{I}_n - \mathbf{dk})^T \qquad (13)$$

Note that it is necessary to maintain $\mathbf{S}_a^u$ and $\mathbf{S}_a$ separately since only $\mathbf{S}_a$ is used to determine the contribution function (Eq. 2).

3.6 Starting Point

Every possible measurement (10^5 spectral grid points $\times$ 16 altitudes) is assessed for its information contribution. However, simply taking the point with the 'highest' information content as a starting point for defining a microwindow may lead to high-priority microwindows which are in fact very limited in altitude and spectral coverage. A better approach might be to define some sort of averaging function to determine the *region* (both spectral and altitude) containing the greatest concentration of high-information points.

One such function could be to sum positive values of H over all altitudes for a particular grid point, then applying a triangular filter in the spectral domain (full width=3cm^{-1}). The microwindow selection would then start from the spectral grid point at which this function was maximised.

3.7 Procedure

The suggested procedure for construction of MWs is as follows.

0. Initialise with suitable *a priori* covariance (§ 3.1).
1. Survey the entire spectrum to determine grid point with maximum information contribution over all altitudes (§ 3.6).
2. For a spectral grid point, evaluating information contribution tangent height by tangent height, setting masks for negative or negligible contributions.
3. Move to adjacent grid point and repeat from 2. Continue until either the additional grid point has no contribution at any altitudes or the 3 cm^{-1} maximum allowed width is reached.
4. A microwindow is now fully defined. Use the total covariance for this microwindow as the *a priori* and repeat from 1 for the next microwindow. Continue until retrieval covariance profile reaches desired value at all altitudes.

This need not be restricted to a single sequence of microwindows. A different set of microwindows would result if the first microwindow was constructed about the grid point with, say, the second highest information content. This would give an alternative sequence of microwindows in case there were a reason for excluding the first sequence.

4 Conclusion

Two basic ideas have been suggested for improving the MW database selection: the use of a multi-layer retrieval and the analysis of information content. The multi-layer retrieval avoids the necessity of a consolidation step, while the information content leads to a systematic method both for constructing individual microwindows and sequences of microwindows. The advantage of using both concepts together is that the inter-layer error correlations are also correctly modelled.

A Global Fit

The 'Global Fit' (GF) retrieval is a weighted least-squares fit which determines the state vector $\mathbf{x}$ of n elements, which fits a set of m measurements $\mathbf{y}$ so as to minimise a cost-function J (actually the χ^2 statistic):

$$J = (\mathbf{y} - \mathbf{Kx})^T \mathbf{S}_y{}^{-1} (\mathbf{y} - \mathbf{Kx}) \quad (14)$$

where $\mathbf{K}$ is the matrix of weighting functions: $K_{ij} = \partial y_i / \partial x_j$. This has the solution:

$$\begin{aligned} \mathbf{x} &= \left(\mathbf{K}^T \mathbf{S}_y{}^{-1} \mathbf{K}\right)^{-1} \mathbf{K}^T \mathbf{S}_y{}^{-1} \mathbf{y} &(15) \\ &= \mathbf{D}^{\mathrm{G}} \mathbf{y} &(16) \end{aligned}$$

where $\mathbf{D}^{\mathrm{G}}$ is the GF contribution function

$$\mathbf{D}^{\mathrm{G}} = \left(\mathbf{K}^T \mathbf{S}_y{}^{-1} \mathbf{K}\right)^{-1} \mathbf{K}^T \mathbf{S}_y{}^{-1} \quad (17)$$

Taking the covariance of Eq. 16 gives:

$$\mathbf{S}_x = \mathbf{D}^{\mathrm{G}} \mathbf{S}_y \mathbf{D}^{\mathrm{G}T} \quad (18)$$

A convenient simplification (for the algebra) is to assume that the measurement covariance matrix $\mathbf{S}_y$ is diagonal and of constant amplitude σ^2, so that $\mathbf{S}_y = \sigma^2 \mathbf{I}_m$. In this case the contribution function, retrieval and covariance reduce to:

$$\begin{aligned} \mathbf{D}^{\mathrm{G}} &= \left(\mathbf{K}^T \mathbf{K}\right)^{-1} \mathbf{K}^T &(19) \\ \mathbf{x} &= \mathbf{D}^{\mathrm{G}} \mathbf{y} = \left(\mathbf{K}^T \mathbf{K}\right)^{-1} \mathbf{K}^T \mathbf{y} &(20) \\ \mathbf{S}_x &= \mathbf{D}^{\mathrm{G}} \sigma^2 \mathbf{I}_m \mathbf{D}^{\mathrm{G}T} = \sigma^2 \left(\mathbf{K}^T \mathbf{K}\right)^{-1} &(21) \end{aligned}$$

As in any unconstrained least-squares fit, it is necessary (but not sufficient) for there to be at least as many measurements m in vector $\mathbf{y}$ as there are retrieved quantities n in vector $\mathbf{x}$ in order that the matrix inversion can be performed.

B Optimal Estimation

An 'Optimal Estimation' (OE) retrieval determines the state vector which best fits both a set of measurements and an *a priori* constraint (vector $\mathbf{a}$, covariance $\mathbf{S}_a$, of same dimensions as $\mathbf{x}$, $\mathbf{S}_x$), achieved by the minimisation of a modified cost-function:

$$\begin{aligned} J &= (\mathbf{y} - \mathbf{Kx})^T \mathbf{S}_y{}^{-1} (\mathbf{y} - \mathbf{Kx}) \\ &\quad + (\mathbf{a} - \mathbf{x})^T \mathbf{S}_a{}^{-1} (\mathbf{a} - \mathbf{x}) &(22) \end{aligned}$$

Implicit in the form of the cost-function (separated covariance matrices) is that the measurement errors and the *a priori* errors are uncorrelated with each other. This has the solution:

$$\mathbf{x} = \mathbf{a} + \mathbf{S}_a \mathbf{K}^T \left(\mathbf{K} \mathbf{S}_a \mathbf{K}^T + \mathbf{S}_y\right)^{-1} (\mathbf{y} - \mathbf{Ka}) \quad (23)$$

Defining an OE contribution function $\mathbf{D}$:

$$\mathbf{D} = \mathbf{S}_a \mathbf{K}^T \left(\mathbf{K} \mathbf{S}_a \mathbf{K}^T + \mathbf{S}_y\right)^{-1} \quad (24)$$

the retrieval can be expressed as:

$$\begin{aligned} \mathbf{x} &= \mathbf{a} + \mathbf{D} (\mathbf{y} - \mathbf{Ka}) &(25) \\ &= \mathbf{Dy} + (\mathbf{I}_n - \mathbf{DK}) \mathbf{a} &(26) \end{aligned}$$

Taking the covariance of Eq. 26, again assuming no cross-correlation between the errors in vectors $\mathbf{y}$ and $\mathbf{a}$, gives the covariance of the solution associated with optimal estimation:

$$\begin{aligned} \mathbf{S}_x &= \mathbf{D} \mathbf{S}_y \mathbf{D}^T + (\mathbf{I}_n - \mathbf{DK}) \mathbf{S}_a (\mathbf{I}_n - \mathbf{DK})^T &(27) \\ &= \mathbf{D} (\mathbf{S}_y + \mathbf{K} \mathbf{S}_a \mathbf{K}^T) \mathbf{D}^T \\ &\quad - \mathbf{DK} \mathbf{S}_a - \mathbf{S}_a \mathbf{K}^T \mathbf{D}^T + \mathbf{S}_a &(28) \\ &= \mathbf{DK} \mathbf{S}_a - \mathbf{DK} \mathbf{S}_a - \mathbf{DK} \mathbf{S}_a + \mathbf{S}_a &(29) \\ &= (\mathbf{I}_n - \mathbf{DK}) \mathbf{S}_a &(30) \\ &= \mathbf{S}_a - \mathbf{S}_a \mathbf{K}^T \left(\mathbf{S}_y + \mathbf{K} \mathbf{S}_a \mathbf{K}^T\right)^{-1} \mathbf{K} \mathbf{S}_a &(31) \end{aligned}$$

C Contribution Function

This demonstrates that the Optimal Estimation equation is the appropriate form to use when updating a Global Fit retrieval with an additional measurement (as assumed in Eq. 1).

Assuming the simplified variance forms (Eqs. 20, 21), let the GF retrieval using a set of m measurements give a state vector $\mathbf{a}$ and covariance $\mathbf{S}_a$:

$$\begin{aligned} \mathbf{a} &= \left(\mathbf{K}^T \mathbf{K}\right)^{-1} \mathbf{K}^T \mathbf{y} &(32) \\ \mathbf{S}_a &= \sigma^2 \left(\mathbf{K}^T \mathbf{K}\right)^{-1} &(33) \end{aligned}$$

Suppose now the retrieval is repeated with an additional *independent* measurement y, with (vector) weighting function $\mathbf{k}$. Modifying the GF equation for $m+1$ measurements, the new solution $\mathbf{x}$ is given by:

$$\begin{aligned}
\mathbf{x} &= \left(\mathbf{K}^T\mathbf{K} + \mathbf{k}^T\mathbf{k}\right)^{-1}\left(\mathbf{K}^T\mathbf{y} + \mathbf{k}^T y\right) && (34)\\
&= \mathbf{a} + \left(\mathbf{K}^T\mathbf{K} + \mathbf{k}^T\mathbf{k}\right)^{-1}\left(\mathbf{k}^T y - \mathbf{k}^T\mathbf{k}\mathbf{a}\right) && (35)\\
&= \mathbf{a} + \mathbf{S}_a\left(\sigma^2\mathbf{I}_n + \mathbf{k}^T\mathbf{k}\mathbf{S}_a\right)^{-1}\mathbf{k}^T(y - \mathbf{k}\mathbf{a}) && (36)\\
&= \mathbf{a} + \mathbf{S}_a\mathbf{k}^T\left(\sigma^2 + \mathbf{k}\mathbf{S}_a\mathbf{k}^T\right)^{-1}(y - \mathbf{k}\mathbf{a}) && (37)
\end{aligned}$$

where the last step follows from considering the alternative factorisations of $\sigma^2\mathbf{k}^T + \mathbf{k}^T\mathbf{k}\mathbf{S}_a\mathbf{k}^T$:

$$\mathbf{k}^T(\sigma^2 + \mathbf{k}\mathbf{S}_a\mathbf{k}^T) = (\sigma^2\mathbf{I}_n + \mathbf{k}^T\mathbf{k}\mathbf{S}_a)\mathbf{k}^T \quad (38)$$

$$(\sigma^2\mathbf{I}_n + \mathbf{k}^T\mathbf{k}\mathbf{S}_a)^{-1}\mathbf{k}^T = \mathbf{k}^T(\sigma^2 + \mathbf{k}\mathbf{S}_a\mathbf{k}^T)^{-1} \quad (39)$$

Comparing Eq. (37) to Eq. (23), it can be seen that this is just the optimal estimation equation for a single measurement, with the GF solution for m measurements acting as the *a priori*.