Tech Note: MIPAS L2 Error Assessment

Anu Dudhia
Atmospheric, Oceanic and Planetary Physics, Oxford University, UK

25 October 2019

1 Linearised Retrieval

A linearised retrieval can be modelled as

$$\mathbf{x} - \mathbf{x}_0 = \mathbf{G}\left(\mathbf{y} - \mathbf{y}_0\right) \tag{1}$$

where $\mathbf{x}$ is the vector of n quantities to be determined (e.g., a profile of concentration of some molecule), $\mathbf{y}$ is the set of m measurements used (e.g., infrared spectra). $\mathbf{x}_0$ represents the linearisation point, e.g., the initial guess as to the solution, and $\mathbf{y}_0$ the measurements expected (the 'forward model') if the linearisation point represented the actual atmosphere.

$\mathbf{G} = \mathrm{d}\mathbf{x}/\mathrm{d}\mathbf{y}$, is the $(n \times m)$ 'Gain Matrix' which, in the absence of significant constraints from regularisation or *a priori* information, approaches that of the weighted least squares fit (WLSF):

$$\mathbf{G} = \left(\mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K}\right)^{-1}\mathbf{K}^T\mathbf{S}_\epsilon^{-1} \tag{2}$$

where $\mathbf{S}_\epsilon$ is the measurement noise error covariance matrix (discussed later) and $\mathbf{K} = \mathrm{d}\mathbf{y}/\mathrm{d}\mathbf{x}$ is the $(m \times n)$ Jacobian matrix, defined by

$$\mathbf{y} - \mathbf{y}_0 = \mathbf{K}\left(\mathbf{x} - \mathbf{x}_0\right) \tag{3}$$

In practice, the problem is usually sufficiently nonlinear that an iterative approach is required. Nevertheless, the linearised version is useful for modelling the error propagation as a perturbation about the true state.

$$\delta\mathbf{x} = \mathbf{G}\delta\mathbf{y} \tag{4}$$

where $\delta\mathbf{y}$ represents an error in either the measurements ($\mathbf{y}$) or the forward model ($\mathbf{y}_0$), and $\delta\mathbf{x}$ the resulting error in the retrieval.

2 Covariance Matrices

While each element of the vector $\delta\mathbf{y}$ is assumed to have a mean value of zero (i.e., no bias, but see section 7) there may be correlations between errors in different elements. These can be represented by the $(m \times m)$ measurement *covariance* matrix, defined as

$$\mathbf{S}_y = \left\langle (\delta\mathbf{y})(\delta\mathbf{y})^T \right\rangle \tag{5}$$

where $\langle \ldots \rangle$ denotes expectation value over a large number of samples. In the absence of any correlation, $\mathbf{S}_y$ is just a diagonal matrix of the error variances of each measurement.

Given the Gain matrix, we can then construct the equivalent error covariance for the retrieval:

$$\begin{aligned} \mathbf{S}_x &= \left\langle (\delta\mathbf{x})(\delta\mathbf{x})^T \right\rangle & (6)\\ &= \left\langle (\mathbf{G}\delta\mathbf{y})(\mathbf{G}\delta\mathbf{y})^T \right\rangle & (7)\\ &= \mathbf{G}\mathbf{S}_y\mathbf{G}^T & (8) \end{aligned}$$

Theoretically, the Gain matrix (Eq. 2) should use the full error covariance matrix $\mathbf{S}_y$ but, since, many of the contributing error terms are not well-defined, in practice it is more common just to use the random noise component $\mathbf{S}_\epsilon$. For this special case of $\mathbf{S}_y \equiv \mathbf{S}_\epsilon$, Eq. (8) simplifies to

$$\mathbf{S}_x^{\mathrm{rnd}} = \left(\mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K}\right)^{-1} \tag{9}$$

The square roots of the diagonal elements of $\mathbf{S}_x^{\mathrm{rnd}}$ are the (random) error bars reported with the retrieved profiles.

3 Systematic Errors

The estimated radiance $\mathbf{y}_0$ (in Eq. 3) is calculated using a forward model $\mathbf{f}$

$$\mathbf{y}_0 \equiv \mathbf{f}(\mathbf{x}, \mathbf{b}) \tag{10}$$

which is a function of not only the retrieved state ($\mathbf{x}$) but also a set of additional parameters $\mathbf{b}$ (such as assumed concentration of interfering species), which are also subject to some uncertainty. Any forward model parameter error δb^i can be mapped into a corresponding measurement error

$$\delta\mathbf{y}^i = -\frac{\partial\mathbf{f}}{\partial b^i}\delta b^i \tag{11}$$

e.g., simply by perturbing the parameter by δb^i by its estimated 1σ uncertainty and re-running the forward model. The minus sign arises is since we are modelling the $\mathbf{y}_0$ component of $\delta\mathbf{y}$.

Since the measurement error vector $\delta\mathbf{y}^i$ originates from a single scalar error b^i, it follows that all components of $\delta\mathbf{y}^i$ are fully correlated, i.e., a systematic error.

Although we treat all systematic errors as associated with the simulated radiances $\mathbf{y}_0$, some errors, such as uncertainty in radiometric gain, are more accurately described as errors in the measurements themselves, $\mathbf{y}$. However, the difference is only in the sign of δb^i.

Thus, provided the set of systematic errors i are independent, we can construct the total systematic error covariance of the retrieval as the sum of covariances of the individual components (equivalent to summing variances in the scalar case)

$$\mathbf{S}_y^i = \left(\delta\mathbf{y}^i\right)\left(\delta\mathbf{y}^i\right)^T \quad (12)$$

$$\mathbf{S}_x^i = \mathbf{G}\mathbf{S}_y^i\mathbf{G}^T \quad (13)$$

$$\mathbf{S}_x^{\mathrm{sys}} = \sum_i \mathbf{S}_x^i \quad (14)$$

In practice, rather than create separate covariance matrices for each systematic error, it is more convenient to use just the error vectors $\delta\mathbf{x}^i$:

$$\mathbf{S}_x^i = \left(\delta\mathbf{x}^i\right)\left(\delta\mathbf{x}^i\right)^T \quad (15)$$

$$\delta\mathbf{x}^i = \mathbf{G}\delta\mathbf{y}^i \quad (16)$$

where the magnitudes of the elements of $\delta\mathbf{x}^i$ provide the value of systematic error i at each level of the retrieved profile.

As an aside, although we treat random (noise) and systematic errors differently, this is just a matter of convenience. The random measurement noise could also be represented as an additional set of m (sparse) error vectors $\delta\mathbf{y}^i$ and propagated the same was as systematic errors.

4 Optimal Estimation

For each MIPAS retrieval the total number of measurements used, $m \sim 1000s$. To avoid the computation of $m \times m$ covariance matrices the error analysis uses a sequential estimation approach where the retrieval $\mathbf{x}$ is progressively updated using measurements in smaller blocks of $m \sim 100$ corresponding to a single microwindow at a single tangent height.

This is mathematically equivalent to considering all the measurements simultaneously, although the actual handling of systematic errors through the sequential updating becomes more complicated. However, one consequence of this approach is that an *a priori* estimate is implicitly required in order that the retrieval using just the first block of measurements is adequately constrained.

This modifies the Gain matrix (Eq. 2):

$$\mathbf{G} = \left(\mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1}\mathbf{K}^T\mathbf{S}_\epsilon^{-1} \quad (17)$$

where $\mathbf{S}_a$ is the *a priori* covariance matrix. It is assumed $\mathbf{S}_a$ is diagonal with elements corresponding to $(10\ \mathrm{K})^2$ for the temperature retrieval, $(10\ \%)^2$ for the pressure retrieval, and $(100\ \%)^2$ for VMR (volume mixing ratio, i.e., molecular concentration) trievals.

The *a priori* estimate is assumed to have only a random error component, no systematic errors. Where the retrieval is well-constrained by the measurements (i.e., good S/N) the Gain matrix approaches the WLSF of Eq. (2). Elsewhere it limits the random error to a maximum value given by $\mathbf{S}_a$ (so, 100% in the case of the VMR retrievals).

The actual MIPAS L2 retrievals use a mixture of *a priori* and regularisation constraints, depending on species, so the Gain matrix used in the error analysis is only an approximation which becomes better the smaller the random error compared to the *a priori* error.

5 pT Sys Error Propagation

MIPAS retrievals are performed sequentially, starting with the pressure-temperature (pT) retrieval, then H_2O, O_3 etc, with the results of previous retrievals also used in the modelling of $\mathbf{y}_0$. This means that the same systematic error can contribute both directly and indirectly through errors in the previously retrieved parameters.

For the error analysis, the only indirect propagation that is considered is through the pT retrieval. The reasoning is that the pressure and temperature are always required for retrieving VMR, but the microwindow selection aims to minimise spectral interference between different molecules.

If, for the VMR retrievals, we separate out the pT profile as a vector $\mathbf{t}$ (i.e., n pressure profile elements plus n temperature profile elements), then we modify Eqs. (10) and (11):

$$\mathbf{y}_0 \equiv \mathbf{f}(\mathbf{x}, \mathbf{b}, \mathbf{t}(\mathbf{b})) \quad (18)$$

$$\delta\mathbf{y}^i = -\left[\frac{\partial\mathbf{f}}{\partial b^i} + \frac{\partial\mathbf{f}}{\partial\mathbf{t}}\frac{\partial\mathbf{t}}{\partial b^i}\right]\delta b^i \quad (19)$$

$$= \delta\mathbf{y}^{i'} - \mathbf{J}\delta\mathbf{t}^i \quad (20)$$

where $\delta\mathbf{y}^{i'}$ is the original calculation (Eq. 11), $\mathbf{J}$ is the Jacobian of $\mathbf{f}$ (or $\mathbf{y}$) with respect to $\mathbf{t}$, and $\delta\mathbf{t}^i$ is the error in the pT retrieval due to error source i.

For example, if there is some error which leads to an underestimate of modelled radiance $\mathbf{y}_0$ (so $\delta\mathbf{y}^{i'}$ is positive), it is also likely to result in an overestimate of temperature (so $\delta\mathbf{t}^i$ is positive), and since $\mathbf{J}$

($=\partial\mathbf{y}/\partial\mathbf{t}$) is usually positive, the effect of pT error propagation term is to offset some of the impact of the error in $\delta\mathbf{y}^i$.

6 pT Rnd Error Propagation

The impact of the random (noise) component of the pT retrieval also has to be considered for subsequent retrievals. This is expressed as a covariance matrix $\mathbf{S}_t^{\mathrm{rnd}}$ of dimension $2n$ set by the pT retrieval vector, i.e., twice the number of sweeps n.

Since the microwindows used for the pT retrieval are different from those used for the VMR retrievals, we do not need to consider correlations within the forward model.

Formally, the required transform is first to map the pT error into the measurement space used by the VMR retrieval $\mathbf{S}_y^{pT}$ via the pT Jacobian $\mathbf{J}$, after which it can be treated as any other independent error covariance component and mapped into a retrieval error using the Gain matrix $\mathbf{G}$

$$\mathbf{S}_y^{pT} = \mathbf{J}\mathbf{S}_t^{\mathrm{rnd}}\mathbf{J}^T \quad (21)$$

$$\mathbf{S}_x^{pT} = \mathbf{G}\mathbf{S}_y^{pT}\mathbf{G}^T \quad (22)$$

The practical problem with this is that it requires an explicit calculation of $\mathbf{S}_y^{pT}$ which, like any covariance in measurement space, can be large when ∼1000s of measurements are used.

Instead, we decompose $\mathbf{S}_t^{\mathrm{rnd}}$ into a set of $2n$ EOFs (Empirical Orthogonal Functions) $\mathbf{e}$

$$\mathbf{S}_t^{\mathrm{rnd}} \equiv \mathbf{e}\mathbf{e}^T \quad (23)$$

This then allows the pT random error to be propagated as a set of $2n$ systematic error vectors which can then be summed to obtain the VMR error vector due to the propagation of the random noise component of the pT retrieval error:

$$\delta\mathbf{x}^{pT} = \sum_{j=1}^{2n} \mathbf{G}\mathbf{J}\mathbf{e}^j \quad (24)$$

7 Bias errors

So far it has been assumed that all forward model parameter errors b^i have unknown sign, essentially that we use are best guess of the value of these parameters so the error could be either positive or negative (i.e., unbiased)

However, there are two forward model errors which represent the neglect of certain physical processes and therefore the sign of the error is known: these are the errors due to the neglect of non-LTE effects and CO_2 line mixing (the error spectra are actually calculated by running more complex radiative transfer calculations with these effects included).

In this case, strictly, the two error vectors should be combined before a covariance is taken

$$\delta\mathbf{y}^i = \delta\mathbf{y}^{\mathrm{nte}} + \delta\mathbf{y}^{\mathrm{mix}} \quad (25)$$

however since the line-mixing error is confined to just a few Q-branches in the CO_2 spectrum (the microwindow selection process is 'aware' of these) the simpler approach is chosen to regard each systematic error source as independent.