united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

Course on "Inverse Methods in Atmospheric Science"
1 - 12 October 2001

301/1332-14

"Spectral Channel Optimisation"

A. DUDHIA
University of Oxford
D. of Physics, Atmospheric, Oceanic & Planetary Physics
England

Spectral Channel Optimisation

Anu Dudhia
Atmospheric, Oceanic & Planetary Physics, University of Oxford
Course on 'Inverse Methods in Atmospheric Science', Trieste 1–12 October 2001

1 Introduction

'Inverse theory' usually concerns retrieving the optimal profile from a given set of measurements. However it is possible to extend the theory to include the determination of the optimal measurements.

Traditionally, selection of spectral filters for radiometers, or channels for multispectral instruments, has been done qualitatively. This describes a quantitative approach to the problem.

The aim is to identify a region of spectrum giving best compromise between

- Random errors ('Precision')
- Total error ('Accuracy'), including both random and systematic errors

Other factors such as computing cost can also be considered.

Using the simplified case of a column retrieval of carbon monoxide from a nadir-viewing instrument, three examples of spectral selection will be considered in detail

1. Selection of individual channels
2. Selection of a filter band
3. Selection of microwindows

Following that, there is a brief discussion of the extension to profile retrievals.

2 Channel Selection

Fig. 1 illustrates the radiance spectrum that might be measured by a downward-viewing instrument viewing at 1 cm^{-1} resolution.

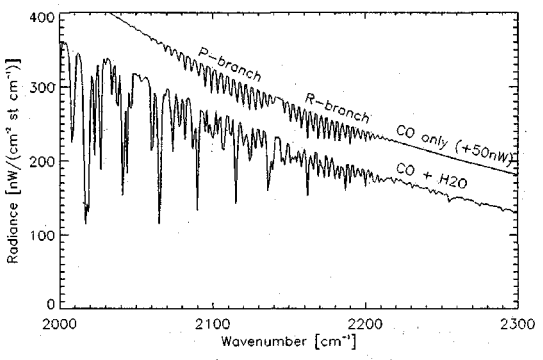

Figure 1: Nadir-viewing radiance spectrum at 1 cm^{-1} resolution for an atmosphere containing CO and H_2O (lower curve) above a black surface at 285 K. The upper curve, offset by 50 $nW/(cm^2\ st\ cm^{-1})$, shows the contribution from CO alone.

This spectrum covers the main CO infrared absorption band 2100–2200 cm^{-1} so would be the obvious region to investigate for retrieving CO. Since the surface is warmer than the atmosphere, the CO and H_2O features appear as lower radiances than transparent regions of the spectrum (opposite to limb-sounding where the background is cold-space).

Defining each 1 cm^{-1} element of this spectrum as a 'channel' (typical resolution obtainable from grating spectrometers), how do we determine the best channel(s) to use?

2.1 Single Channel: Precision

We can describe the radiance spectrum in Fig. 1 ($CO+H_2O$) as a set of measurements y_i each corresponding to a single 1 cm^{-1} channel. Each measurement is linearly related to the retrieved param-

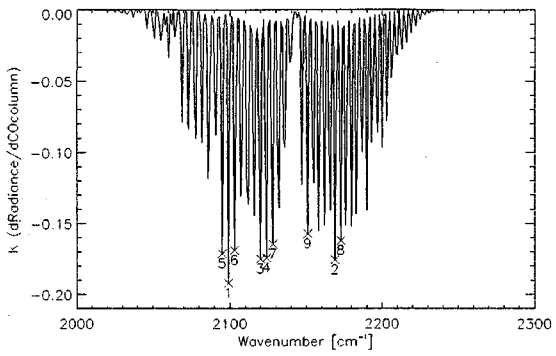

Figure 2: Jacobians of the CO column amount for each 1 cm^{-1} channel. Crosses and numbers indicate the channels with the 9 largest values.

eter x (i.e., the CO column amount) by

$$y_i = \left(\frac{\partial y_i}{\partial x}\right) x = k_i x \tag{1}$$

which defines the elements k_i of the Jacobian (or Weighting Function) vector $\mathbf{k}$. (The fact that relationship is usually non-linear is not relevant here since we are only considering the error analysis). This relationship can be simply inverted to give an estimate $\hat{x}$ of the CO column

$$\hat{x} = \frac{y_i}{k_i} \tag{2}$$

Assuming all measurements have the same random noise variance S_ϵ, the random error in the estimate $\hat{x}$ is given by

$$S_x^{\text{rnd}} = \frac{S_\epsilon}{(k_i)^2} \tag{3}$$

For a single channel, then, the random error is minimised by selecting the channel with the highest sensitivity $|k_i|$ to CO column (obviously!). Fig. 2 shows the Jacobian spectrum of $\mathbf{k}$, and the figure '1' marks the 'best' channel in the sense of minimising the random error.

2.2 Multiple Channels: Precision

A set of channels can be expressed as elements of a vector $\mathbf{y}$, with the corresponding Jacobians as vector $\mathbf{k}$

$$\mathbf{y} = \mathbf{k}x \tag{4}$$

Since $\mathbf{k}$ is not a square matrix, we cannot directly invert this to obtain $\hat{x}$, but, for example, can use a least-squares fit

$$\hat{x} = \left(\mathbf{k}^T\mathbf{k}\right)^{-1}\mathbf{k}^T\mathbf{y} \equiv \mathbf{g}^T\mathbf{y} \tag{5}$$

$$= \frac{\sum_i k_i y_i}{\sum_i (k_i)^2} \tag{6}$$

which defines the 'Gain Vector' (row vector) $\mathbf{g}^T$. If the noise on each measurement is constant S_ϵ and uncorrelated, this solution has random variance

$$S_x^{\text{rnd}} = \mathbf{g}^T\mathbf{S}_\epsilon\mathbf{g} \tag{7}$$

$$= \left(\mathbf{k}^T\mathbf{k}\right)^{-1} S_\epsilon \tag{8}$$

$$= \frac{S_\epsilon}{\sum_i (k_i)^2} \tag{9}$$

where $\mathbf{S}_\epsilon$ is the diagonal matrix $S_\epsilon\mathbf{I}$. It can be shown that the weights represented by g_i in Eq. 5 represent the minimum variance solution S_x^{rnd}.

More obviously, perhaps, this is the same solution obtained by combining the estimates $\hat{x}$ obtained from each channel individually (Eq. 2) weighted by their variances (Eq. 3).

The minimum (random) variance solution for m channels is obtained simply by combining the measurements with the m largest Jacobians $|k_i|$ (Fig. 2), which are the same as the m channels giving the most precise individual estimates of $\hat{x}$.

Taking a noise variance of S_ϵ=(1.0 nW/(cm^2 st cm^{-1})^2, Fig. 3 shows the improvement in precision as channels are added in decreasing order of sensitivity $|k_i|$. Also shown is the curve $\sim \sqrt{m}$ corresponding to the improvement expected if all channels were as equally sensitive as the first channel selected. Out of a potential 301 channels, the minimum variance is approached after about 50 channels are selected, corresponding to those which cover the strongest absorption lines.

2.3 Single Channel: Accuracy

So far it has been assumed that the random noise S_ϵ is the only source of uncertainty in the retrieval. As illustrated in Fig. 1, water vapour is also a significant absorber in this spectral region. If the water vapour is not accurately known, this can contribute to the uncertainty in modelling the radiance spectrum. If we assume a ±5% uncertainty in the estimated total H_2O column amount, this contributes

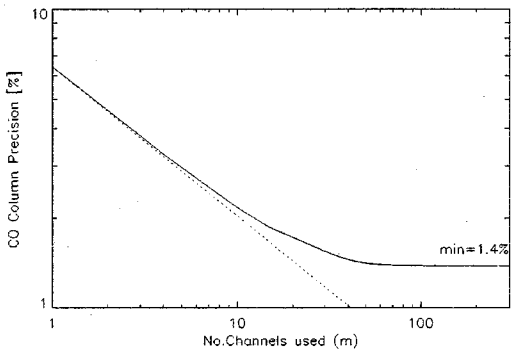

Figure 3: Reduction in random error as channels are combined in order of decreasing sensitivity. The dotted line shows the $\sqrt{m}$ curve through the first point for comparison.

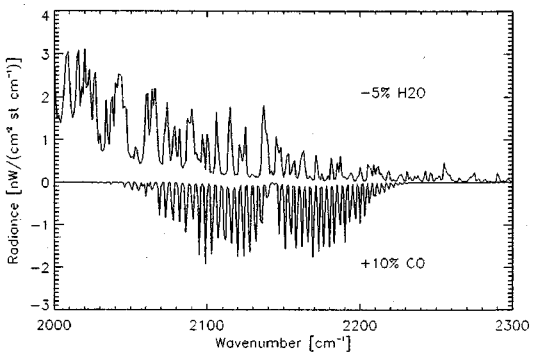

Figure 4: Changes in nadir-viewing radiance spectrum for a 5% decrease in H_2O (top curve) or a 10% increase in CO (bottom curve).

an error *spectrum* comparable in magnitude (Fig. 4) to a 10% variation in CO column, and affects the P-branch more than the R-branch.

For a single-channel retrieval (Eq. 2) $\hat{x} = y_i/k_i$ so if δy_i represents the value of the H_2O error spectrum for measurement y_i, this will contribute a retrieval error variance of

$$S_x^{\text{sys}} = \frac{(\delta y_i)^2}{(k_i)^2} \tag{10}$$

Combining this with the noise contribution (Eq. 3), the total retrieval error will then be

$$S_x^{\text{tot}} = S_x^{\text{rnd}} + S_x^{\text{sys}} = \frac{1}{(k_i)^2}\left(S_\epsilon + (\delta y_i)^2\right) \tag{11}$$

So the best single channel would be the one which produces the minimum value of S_x^{tot} which now depends not only on having a large sensitivity to CO (large $|k_i|$), but also on a low sensitivity to H_2O variability $|\delta y_i|$ compared to the noise S_ϵ. In this case it turns out to be the same as the first channel selected on the basis of minimising the random error alone.

2.4 Combining Systematic Errors

Although for a single channel we can effectively regard the contribution from H_2O uncertainty as an additional noise variance $(\delta y_i)^2$, when combining two or more channels it has to be treated separately. This is because the H_2O errors are correlated: an underestimate of the water-vapour column by 5% will lead to a positive increase in radiance across the whole spectrum.

Assuming that the retrieval weights each measurement as

$$\hat{x} = \mathbf{g}^T\mathbf{y} \tag{12}$$

where $\mathbf{g}^T$ may or may not be the weights used in Eq. 5, the systematic component of the retrieval error due to water-vapour is

$$\delta x = \mathbf{g}^T\delta\mathbf{y} \tag{13}$$

where $\delta\mathbf{y}$ is a random vector with shape given by the 5% H_2O error spectrum but multiplied by an unknown random scalar with mean 0 and standard deviation 1.

The total error on $\hat{x}$ is then

$$S_x^{\text{tot}} = \mathbf{g}^T\mathbf{S}_\epsilon\mathbf{g} + \left(\mathbf{g}^T\delta\mathbf{y}\right)^2 \tag{14}$$

2.5 Systematic Error Sources

By 'systematic' we mean any source of error which is fully correlated between measurements and thus can be represented as an error spectrum.

Here we consider only one source of systematic error, but additional independent sources of systematic error j can be included in Eq. 14 as addi-

tional terms:

$$S_x^{\rm tot} = \mathbf{g}^T \mathbf{S}_\epsilon \mathbf{g} + \sum_j \left(\mathbf{g}^T \delta\mathbf{y}^j\right)^2 \qquad (15)$$

Effectively, any source of error in the measurement or forward model that can be represented as a 1-σ error spectrum and is independent of other error sources can be included.(If error sources are correlated they can, in principle, be decoupled into independent components). Examples are

- Uncertainties in interfering species
- Calibration uncertainties (both radiometric and spectral)
- Instrumental uncertainties (e.g., pointing)
- Forward model deficiencies (e.g., omitting non-LTE)

2.6 Covariance Matrices

The systematic error component in the measurements and the retrieval has been described in terms of an error vector $\delta\mathbf{y}^j$ although the measurement total error covariance matrix $\mathbf{S}_y$ is a more natural method of containing this information. The two are related by:

$$\mathbf{S}_y = \mathbf{S}_\epsilon + \sum_j E\{(\delta\mathbf{y}^j)(\delta\mathbf{y}^j)^T\} \qquad (16)$$

$$= \mathbf{S}_\epsilon + \sum_j (\delta\mathbf{y}^j)(\delta\mathbf{y}^j)^T \qquad (17)$$

where $\mathbf{S}_\epsilon$ is the random noise covariance and $E\{\ldots\}$ denotes 'expectation value'.

The retrieval total error is given by

$$S_x^{\rm tot} = \mathbf{g}^T \mathbf{S}_y \mathbf{g} \qquad (18)$$

While this is more elegant than Eq. 15, it is necessary to maintain the systematic errors as separate vectors $\delta\mathbf{y}^j$ so we shall continue to use the form in Eq. 15.

2.7 Gain matrix

In Eqs. 13–18 we have not specified the elements of $\mathbf{g}^T$ that map each measurement into the retrieval.

We want to select channels to minimise the total retrieval error $S_x^{\rm tot}$, but there are two possibilities according to whether or not the gain matrix is specified.

If we are at liberty to choose our own gain matrix then, in principle one could redefine the weights g_i that minimise $S_x^{\rm tot}$.

However, if we are selecting channels for some operational retrieval that is defined to weight measurements by random noise then we need to use the same form. This will be assumed in the following, although see §4.4 on spectral masks.

2.8 Multiple Channels: Accuracy

We can now evaluate the total error for any combination of channels, but how do we determine which channels to use? There are several possibilities

1. Select channels in sequence of increasing random error (decreasing $|k_i|$) as in §2.2
2. Determine total error of each channel regarded as a single channel retrieval (Eq. 11), and select in sequence of increasing value
3. Determine total error of each channel, as in 2, but after selecting channel with the smallest value, evaluate *increase* in total error for each channel, select whichever minimises total error, reevaluate increase, and so on.

For the first method, the total error together with the random and systematic error components are plotted in Fig. 5. This method reaches a minimum of 2.2% accuracy after about 20 channels and then the systematic errors start to accumulate, increasing the total error.

The results are compared in with the other two methods in Fig. 6.

The first three selected points are the same for all methods, but the other two methods remain close for the first 10 points before the systematic errors start to dominate the second method, but the third method is clearly the best for any number of selected channels. Between about 30–200 channels points the accuracy is constant, reflecting the fact that points with zero CO sensitivity (and therefore zero weight) are introduced in preference to points which improve the CO precision but at the expense of accuracy due to H_2O inteference. The selection would stop well before this, but if all methods are forced to use all 301 points then the accuracies necessarily converge once again. The first 9 selected

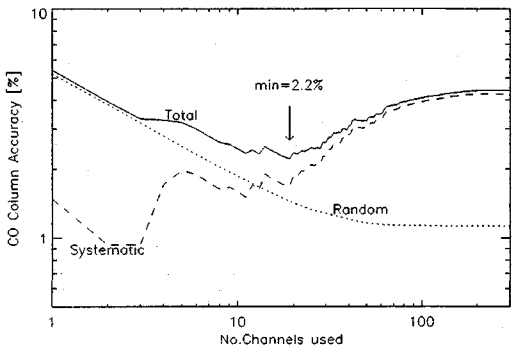

Figure 5: The change in accuracy as channels are selected in the sequence which minimises random error. Also plotted are the random (same as in Fig. 3) and systematic components of the total error.

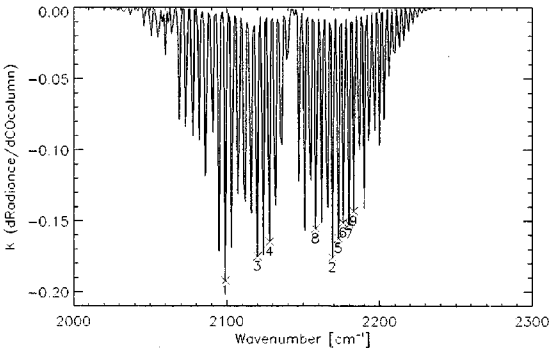

Figure 7: Jacobians of the CO column amount for each 1 cm^{-1} channel, as before. Crosses and numbers indicate the best 9 channels (selected by Method 3) allowing for a 5% uncertainty in water vapour.

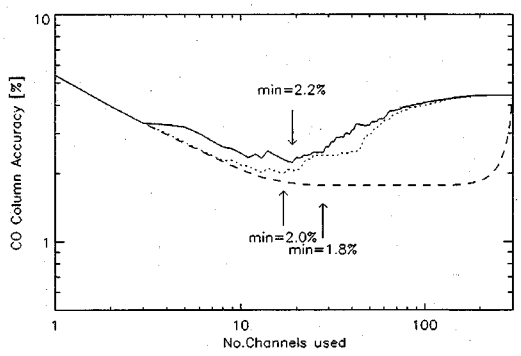

Figure 6: CO column retrieval accuracy as a function of number of channels used for 3 different selection schemes: (1) select to minimise random error (solid); (2) select by minimum initial total error (dotted); (3) select by minimum iterated total error (dashed).

points for the third method are plotted in Fig. 7 — note the shift to higher wavenumbers (smaller H_2O contamination) compared to the points selected for method 1 (Fig. 2).

2.9 Optimal Selection?

For m channels, the 'Optimal Selection' is the set of channels which produces the minimum error.

Considering random errors alone this is straightforwards: the random error contribution from each possible channel is evaluated, and the best m channels correspond to the smallest m random retrieval errors.

Including systematic errors it is more complicated. We can evaluate total error for each channel, and the best single channel is the channel with the smallest total error. However, for two channels we cannot assume that the next best channel is the one with the second smallest total error (Method#2), or even that it is the one with the smallest increase (Method#3) — there may be two completely different channels which combine to give a smaller total error.

The only method to be sure is to calculate the total error for all possible combinations of m channels — usually too large a problem to be practical.

3 Filter Selection

A filter can be regarded as a sequence of adjacent channels, all of which will be equally weighted (assuming a 'boxcar' filter response), and defined by the two cut-off points.

For a filter spanning m channels $i = j+1 \ldots j+m$, the measurement $\overline{y}$ and Jacobian $\overline{k}$ are just the averages over the component channels

$$\overline{y} = \frac{1}{m} \sum_{i=j+1}^{j+m} y_i \tag{19}$$

$$\overline{k} = \frac{1}{m} \sum_{i=j+1}^{j+m} k_i \tag{20}$$

Assuming that the noise variance is the sum of the constant noise variances S_ϵ of each channel:

$$\overline{S}_\epsilon = \frac{1}{m^2} \sum_{i=j+1}^{j+m} S_\epsilon = \frac{S_\epsilon}{m} \tag{21}$$

giving S/N $\sim \sqrt{m}$ for constant y_i, k_i.

3.1 Filter: Precision

Considering just the random error, the retrieval error can be expressed as

$$S_x^{\mathrm{rnd}} = \frac{\overline{S}_\epsilon}{\overline{k}^2} \tag{22}$$

So, starting with the single channel i with maximum $|k_i|$ we can compare the improvement if adjacent channels are added, and move the filter edge to include whichever leads to the greatest reduction in S_x^{rnd}. The expansion of the filter edges is plotted in Fig. 8

The first selected point is the same as the first channel selected in Fig. 2, corresponding to the centre of the strongest P-branch line near 2100 cm^{-1}. After that, the upper edge first moves to 2140 cm^{-1} then the lower edge to 2075 cm^{-1} then the upper edge again to encompass most of the R-branch, etc.

The change in retrieval precision as the filter is expanded is plotted in Fig. 9.

The precision initially gets worse as the filter expands from 1–5 cm^{-1} width. While the filter expands across the gaps between CO lines (spacing

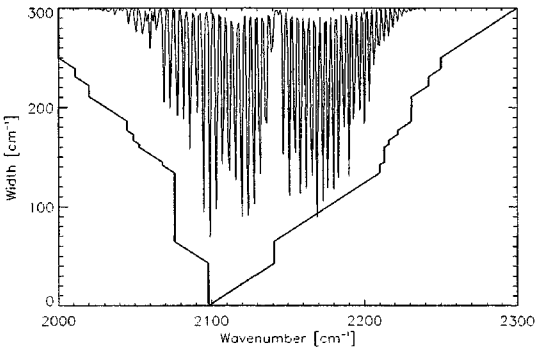

Figure 8: Growth of filter for CO column retrieval considering random errors alone.

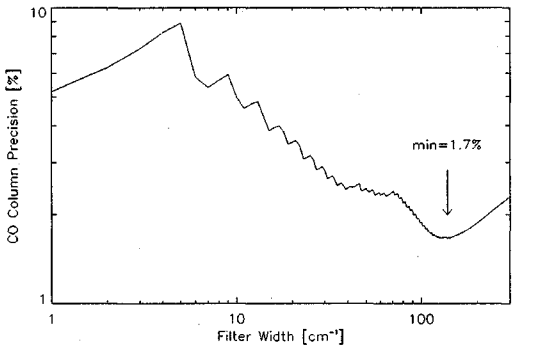

Figure 9: Change in precision for CO column as a function of filter size in Fig. 8.

~5 cm^{-1}), the signal contribution from CO is reduced but the noise continues to increase in proportion to the square root of the filter width. The ripple effect is apparent as more CO lines are included. Finally a minimum (1.7%) is reached at a width of around 140 cm^{-1}. At this point (Fig. 9) most of the strong lines in the CO band are included and any further expansion in either direction introduces more noise than signal.

Note: this finite limit does not occur with selection of individual channels (Fig. 3) since g_i (Eq. 5) reduces both signal and noise contributions for channels with low signal contribution.

3.2 Filter: Accuracy

The 5% uncertainty in H_2O column contributes an error $\overline{\delta y}$ to the measurement $\overline{y}$ given by:

$$\overline{\delta y} = \frac{1}{m} \sum_{i=j+1}^{j+m} \delta y_i \tag{23}$$

Analogously with the single channel case (Eq. 11), the total retrieval error is:

$$S_x^{\mathrm{tot}} = S_x^{\mathrm{rnd}} + S_x^{\mathrm{sys}} \tag{24}$$

$$= \frac{1}{(\overline{k})^2} \left(\overline{S}_\epsilon + (\overline{\delta y})^2 \right) \tag{25}$$

Growing the filter edgewise as before, except this time minimising S_x^{tot} rather than just S_x^{rnd}, the results are shown in Figs. 10.

The first selected point is the same as before, but this time the initial growth is generally towards lower wavenumber rather than higher wavenumbers. However, Fig. 11 shows that the total error never improves after the first point. It is apparent that this is due to the increasing systematic error component: having included many CO lines with significant H_2O contributions while the random error is dominant, it is not then possible to exclude them so as to reduce the systematic error contribution.

This is another example of the difficulty in finding 'optimal' solutions when considering systematic errors: the initial selection makes no allowance for the ultimate accuracy.

The 'best' filter can be obtained by considering the error of all possible filter positions (involving

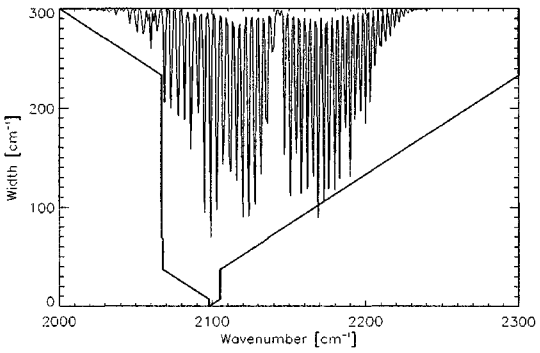

Figure 10: Growth of filter for CO column retrieval considering total error.

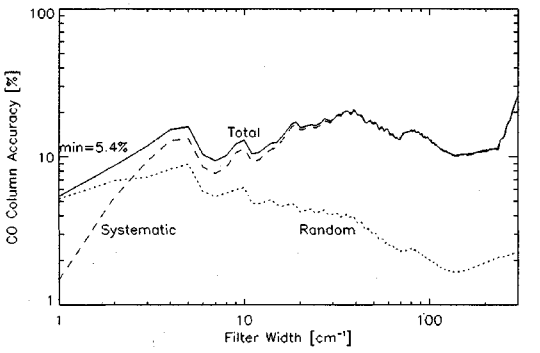

Figure 11: Growth of filter for CO column retrieval considering total error.

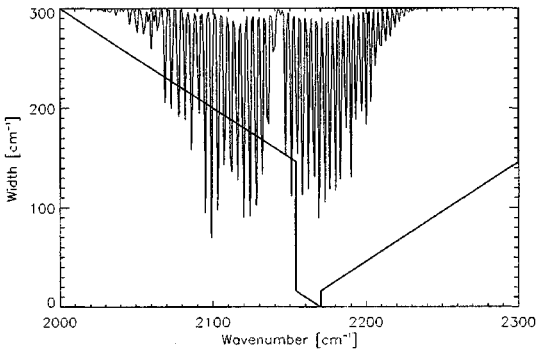

Figure 12: Growth of filter for CO column retrieval considering total error, as in Fig. 10, except starting at the second best point (around 2170 cm^{-1})

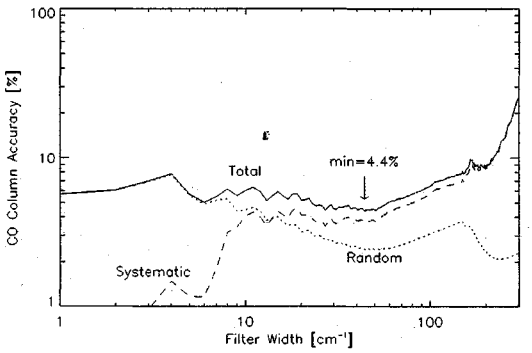

Figure 13: Accuracy as a function of filter width for filter grown in Fig. 12.

300×300 trials in this limited example). More practically, one can try 'growing' a number of filters from different starting positions and select the best.

Figs. 12 and 13 show the results of repeating the selection but this time starting at the point with the second lowest value of S_x^{tot}. Since this happens to be in the less-contaminated R-branch, the systematic errors are generally lower to start with and as a result the total error reaches a lower minimum of 4.4% accuracy for a filter width of around 40 cm^{-1} encompassing most of the R-branch.

As well as growing a number of 'trial' filters, one can also weight the initial selection against systematic errors by growing towards the side which, instead of minimising S_x^{tot}, minimises

$$S_x^{\mathrm{rnd}} + \alpha S_x^{\mathrm{sys}} \tag{26}$$

for $\alpha > 1$. Using a value $\alpha = 3$ promotes the R-branch channel from second to first selected point and the filter subsequently grows in a similar fashion, although moving the higher rather than lower wavenumber boundary more to start with (Fig. 14).

Fig. 15 shows a slightly lower minimum (4.1%), achieved at 20 cm^{-1} width, but the curves are all similar to the previous example starting at the same point. Given sufficient number of trials with different starting points, further improvements may still be found.

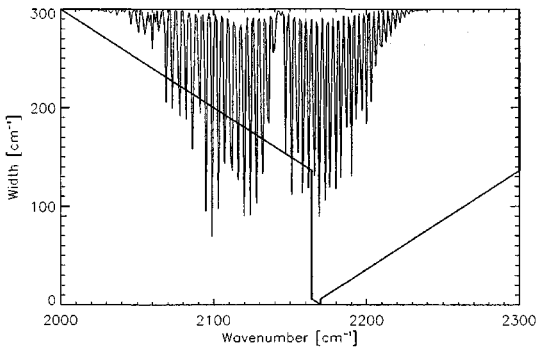

Figure 14: Growth of filter for CO column retrieval considering total error, as in Fig. 10, except penalising systematic errors.

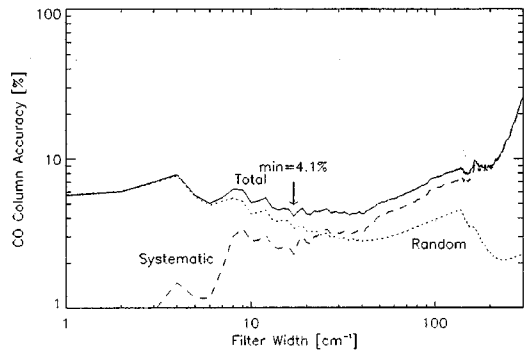

Figure 15: Accuracy as a function of filter width for filter grown in Fig. 14.

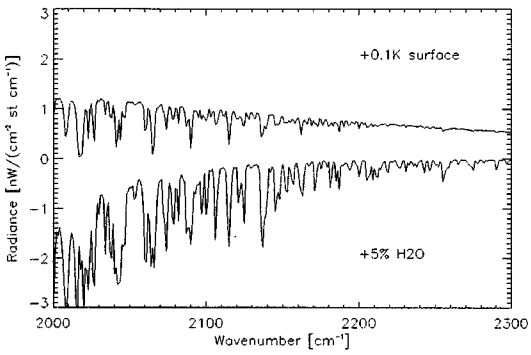

Figure 16: Changes in nadir-viewing radiance spectrum for a 0.1K increase in surface temperature (top curve) compared with a 5 % increase in H_2O column (bottom curve).

4 Microwindow Selection

A microwindow can be regarded as a set of adjacent channels. However, unlike filters, individual weights can be applied to each channel. The advantage of selecting adjacent channels is that an additional component can be fitted within each microwindow, e.g., surface radiance (nadir sounding) or atmospheric continuum (limb sounding) — anything whose source can be assumed spectrally flat over the width of the microwindow. This is useful for eliminating smoothly varying but unknown terms which affect the absolute radiance.

Fig. 16 shows the Jacobian spectrum for the surface emission. Although the surface emission is spectrally smooth (Planck function), its Jacobian has significant structure corresponding to the positions of strong H_2O lines.

4.1 Single Microwindow: Precision

To select the first microwindow, there are now two unknowns: the CO column x_1 and the surface emission x_2 in vector $\mathbf{x}$. To solve this we can consider retrieving from pairs of adjacent spectral points $\mathbf{y} = (y_i, y_{i+1})$:

$$\mathbf{G} = \left(\mathbf{K}^T\mathbf{K}\right)^{-1}\mathbf{K}^T \quad (27)$$

$$\hat{\mathbf{x}} = \mathbf{G}\mathbf{y} \quad (28)$$

$$\mathbf{S}_x^{\mathrm{rnd}} = \mathbf{G}\mathbf{S}_\epsilon\mathbf{G}^T \quad (29)$$

where $\mathbf{K}$ is initially a 2×2 matrix.

We are only interested in the precision with which the CO column can be retrieved — not the surface term — so we need only consider minimising the first diagonal element of $\mathbf{S}_x^{\mathrm{rnd}}$ (corresponding to the random variance of the CO retrieval).

Having found the best 2-element microwindow, we then investigate expanding either edge in turn by adding an extra element to either end of the $\mathbf{y}$ vector in Eq. 28, i.e., additional rows to the matrix $\mathbf{K}$ and expand the microwindow in whichever direction gives the best improvement or some maximum specified width is reached. Note that $\mathbf{G}$ is initially $\mathbf{K}^{-1}$ (2×2) but, unlike $\mathbf{K}^{-1}$, $\mathbf{G}$ remains well-defined — the least squares fit solution — as the first dimension of $\mathbf{K}$ is increased as more measurements are added to the microwindow.

The maximum width is the range over which we believe our assumption of a spectrally flat surface term is valid, in this case taken as 10 cm^{-1}. Fig. 17 shows the first microwindow is, not surprisingly, selected around the strong P-branch line at 2100 cm^{-1}. Selecting for precision, this will always grow to the maximum width.

4.2 Multiple Microwindows: Precision

Having selected the first microwindow, selection of the second and subsequent microwindows is not quite so straightforwards since every time we

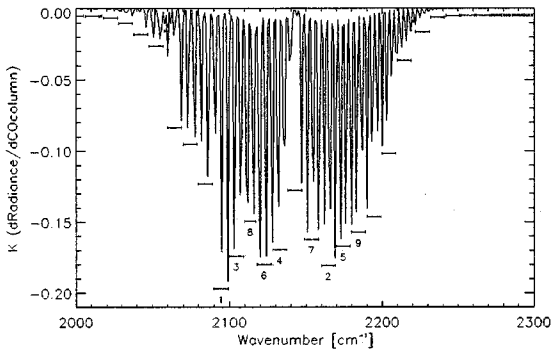

Figure 17: Selected microwindows for CO column retrieval minimising random error. Numbers indicate the first 9 microwindows selected.

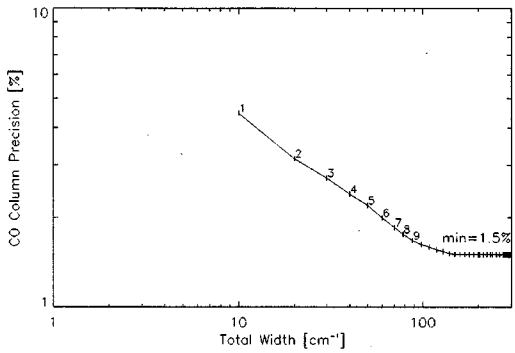

Figure 18: Precision of CO column retrieval as microwindows are added. The vertical ticks indicate the cumulative width for each new microwindow, the numbers indicate the first 9 microwindows.

add a microwindow we add an extra retrieved parameter x_j corresponding to the surface term in microwindow#$(j-1)$, so that the Jacobian matrix $\mathbf{K}$ becomes unwieldy (previously this was only a vector $\mathbf{k}$).

Fortunately, 'Sequential Estimation' provides a convenient method of updating the random covariance considering only the Jacobian matrix for microwindow at a time. Given $\mathbf{S}_x^{\mathrm{rnd}}$ from the first microwindow, start by setting an *a priori* estimate for the next microwindow covariance $\mathbf{S}_a^{\mathrm{rnd}}$

$$\mathbf{S}_a^{\mathrm{rnd}} = \begin{pmatrix} (S_x^{\mathrm{rnd}})_{11} & 0 \\ 0 & \infty \end{pmatrix} \qquad (30)$$

where the second diagonal element is a large number, corresponding to no prior knowledge of the surface term in the next microwindow. As measurements for the next microwindow are added the covariance is updated as follows:

$$\mathbf{G} = \mathbf{S}_a^{\mathrm{rnd}}\mathbf{K}^T\left(\mathbf{S}_\epsilon + \mathbf{K}\mathbf{S}_a^{\mathrm{rnd}}\mathbf{K}^T\right)^{-1} \qquad (31)$$

$$\mathbf{S}_x^{\mathrm{rnd}} = (\mathbf{I} - \mathbf{G}\mathbf{K})\mathbf{S}_a^{\mathrm{rnd}} \qquad (32)$$

where $\mathbf{K}$ only contains the measurements being considered for the current microwindow. Having completed the microwindow, $\mathbf{S}_a^{\mathrm{rnd}}$ is updated as before, ready for the next microwindow. Of course, this could also be used to select the first microwindow with $(S_a)_{11}$ also set large.

Applying this, the additional microwindows shown in Fig. 17 are selected. As expected the first 9 microwindows are concentrated around the strong CO lines and most grow to their maximum allowed width of 10 cm^{-1}. The only limitation on reaching maximum width is the overlap with points already selected for earlier microwindows.

The improvement in precision is plotted in Fig. 18. Note that this is not quite as good as the channel-by-channel selection, reaching 1.5% rather than 1.4% (Fig. 3). This is because some information from the measurements is required to retrieve the surface term for each microwindow.

4.3 Multiple Microwindows: Accuracy

For a single microwindow, the total error is given by

$$\mathbf{S}_x^{\mathrm{tot}} = \mathbf{G}\mathbf{S}_\epsilon\mathbf{G}^T + (\mathbf{G}\delta\mathbf{y})\,(\mathbf{G}\delta\mathbf{y})^T \qquad (33)$$

Where $\mathbf{G}$ is as defined in Eq. 27. This only differs from the multiple-channel case (Eq. 14) in that instead of the single row vector $\mathbf{g}^T$ there is now the 2-row matrix $\mathbf{G}$ to accommodate the two retrieved parameters. Again, only the first diagonal element of $\mathbf{S}_x^{\mathrm{tot}}$ needs to be considered. This shifts the first selected microwindow from the P-branch to the strongest line in the R-branch (Fig. 19).

For the second and subsequent microwindows we can use the 'Sequential Estimation' equations (Eqs. 31 and 32) once again, although it is necessary to introduce modifications to track the sys-

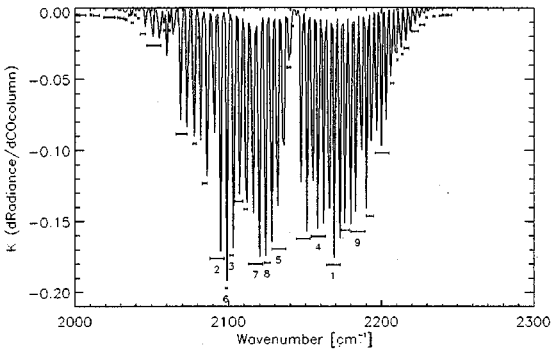

Figure 19: Selected microwindows for CO column retrieval minimising total error.

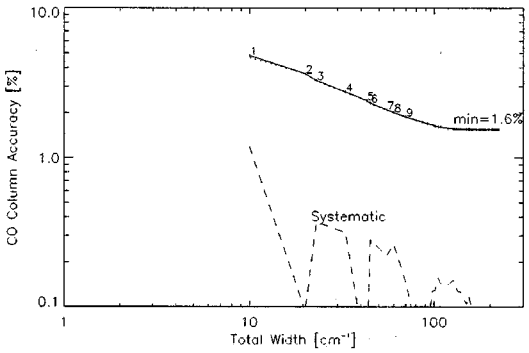

Figure 20: Accuracy of CO column retrieval as microwindows are added.

tematic errors through the *a priori* estimate $\delta\mathbf{a}$ (initially set to 0)

$$\delta\mathbf{x}^i = \mathbf{G}\delta\mathbf{y} + (\mathbf{I} - \mathbf{GK})\,\delta\mathbf{a} \quad (34)$$

$$\mathbf{S}_x^{\text{tot}} = \mathbf{S}_x^{\text{rnd}} + (\delta\mathbf{x})\,(\delta\mathbf{x})^T \quad (35)$$

Applying this, the selected microwindows are shown in Fig. 19. Compared to the previous case minimising only the random error (Fig. 17) the first few microwindows are still concentrated around the strong CO lines but in this case they often tend to be smaller than the maximum allowed width.

The accuracy is plotted as a function of selection in Fig. 20. This shows remarkably small systematic errors, and a total error which is almost entirely dominated by random error. Using all the selected microwindows, an accuracy of 1.6% is obtained compared to 1.8% for individual channels (Fig. 6, Method#3) and 4.1% with a filter (Fig. 15).

There are three explanations for this

1. The algorithm has selected microwindows where the H_2O error spectrum is absorbed into the surface emission retrieval rather than the CO retrieval.

2. Instead of there always being a negative correlation (overestimate $H_2O \Rightarrow$ underestimate CO), a dual CO-surface retrieval allows both signs of correlation between H_2O and CO

3. In this particular example, there is a strong anticorrelation between the surface emission Jacobian and the H_2O error spectrum (Fig. 16) which assists the removal of H_2O errors (effectively performing an H_2O retrieval)

The first two points are generally true, but to achieve the second effect it is necessary to have a fixed sequence of microwindows, not just an arbitrary selection from a set of individually 'good' microwindows. Such a selection can probably only be made by computer.

4.4 Spectral Masks

Selecting for precision, microwindow growth is limited only by the maximum width allowed, or by encountering channels previously selected for other microwindows. This is because the Gain Matrix $\mathbf{G}$ (Eqs. 27 or 31) optimally weights each channel by its random noise and Jacobian so as to maximise precision.

However, selecting for accuracy using the same Gain Matrix, it is possible that while including the channel reduces the random error, the total error increases due to the larger systematic errors associated with the channel: the channel effectively contributes 'negative information', so the microwindow growth stops rather than include it.

If we are not allowed to change the Gain Matrix of the retrieval, there is still an option to 'mask' the channel, i.e., remove it from the measurement vector completely, and continue to grow the microwindow up to its maximum width. This effectively applies zero weight to the channel instead of the Gain Matrix weight.

5 Profile Retrievals

When retrieving a profile rather than a simple scalar, there are two problems

1. Finding a retrieval model which works for any number of unknowns
2. Finding a single scalar parameter to minimise

5.1 Retrieval Model

Up until now the starting point for each selection has been based on inverting an equation of the form:

$$\hat{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{y} \tag{36}$$

where $\mathbf{K}$ is either a scalar or (for microwindow selection) a 2×2 matrix, and the number of measurements in $\mathbf{y}$ is the same as the number of retrieved parameters in $\mathbf{x}$.

To retrieve a profile where $\mathbf{x}$ has dimension $n > 10$, it is probably impractical simply to form $\mathbf{y}$ from sets of n consecutive measurements and expect that $\mathbf{K}$ is non-singular (i.e., that the n measurements each contain some unique information on each of the n retrieved parameters).

'Optimal Estimation' turns out to be a convenient solution: start with some *a priori* estimate of the (random) covariance $\mathbf{S}_a$, then apply the 'Sequential Estimation' equations Eqs 31,32 and 34,35. This allows selection to start with a single channel for any number of retrieved parameters.

5.2 Figure of Merit

Assuming that we can calculate the retrieval covariance $\mathbf{S}_x$ how do we compare two alternative covariances (each containing $n \times n$ elements) and decide which is better?

For example, consider three covariance matrices that could describe a 2-level profile retrieval:

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \mathbf{C} = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$$

Using the trace (sum of diagonal elements), $\mathbf{A}$ appears best (=5), followed by $\mathbf{B}$ (=6) then $\mathbf{C}$ (=7).

Using the product of diagonal elements, $\mathbf{B}$ is best (=5), followed by $\mathbf{A}$ (=6) and $\mathbf{C}$ (=10).

Using the determinant, $\mathbf{C}$ is best (=1), followed by $\mathbf{B}$ (=5) and $\mathbf{A}$ (=6).

Unlike with the scalar retrieval, there is not a unique choice.

5.3 Information Content

The Shannon information content of a retrieval is a reasonable figure of merit to apply. It is defined by

$$H = -\frac{1}{2}\log_2\left(\frac{|\mathbf{S}_x|}{|\mathbf{S}_a|}\right) \tag{37}$$

where $|\ldots|$ signifies determinant and $\mathbf{S}_a$ represents the covariance of the *a priori* information, i.e., how well the profile is known before the retrieval. For example, if we have 10-level temperature profile and we knew the temperature at each level to ±10 K before the retrieval, and after the retrieval we knew the temperature at each level to ±5 K, this would yield a value $H = 10$. H is measured in 'bits', reducing the uncertainty at each profile level by a factor 2 corresponding to 1 'bit' of information.

Really, this is just weighting by the determinant.

6 Bibliography

Rodgers, C. D., "Information content and optimisation of high spectral resolution remote measurements," *Adv. Space Research* **21**, 361–367 (1998).

Clarmann, T. von and G. Echle, "Selection of optimized microwindows for atmospheric spectroscopy," *App. Optics* **37**, 7661–7660 (1998).

Echle, G., T. von Clarmann, A. Dudhia, J.-M. Flaud, B. Funke, N. Glatthor, B. J. Kerridge, M. López-Puertas, F. J. Martín-Torres and G. P. Stiller, "Optimized spectral microwindows for MIPAS-Envisat data analysis," *App. Optics* **39**, 5531–5540 (2000).

Rabier, F., N. Fourrié, D. Chafaï and P. Prunet, "Channel selection methods for infrared atmospheric sounding interferometer radiances," *Q. J. R. Meteorol. Soc.* **128**, 1–15 (2001).

Dudhia, A., V. L. Jay and C. D. Rodgers, "Microwindow Selection for High Spectral Resolution Sounders" submitted to *App. Optics*, August 2001.