Intercomparison of remote sounding instruments

Clive D. Rodgers[1] and Brian J. Connor

National Institute for Water and Atmospheric Research, Lauder, New Zealand

[1]   When intercomparing measurements made by remote sounders, it is necessary to make due allowance for the differing characteristics of the observing systems, particularly their averaging kernels and error covariances. We develop the methods required to do this, applicable to any kind of retrieval method, not only to optimal estimators. We show how profiles and derived quantities such as the total column of a constituent may be properly compared, yielding different averaging kernels. We find that the effect of different averaging kernels can be reduced if the retrieval or the derived quantity of one instrument is simulated using the retrieval of the other. We also show how combinations of measured signals can be found, which can be compared directly. To illustrate these methods, we apply them to two real instruments, calculating the expected amplitudes and variabilities of the diagnostics for a comparison of CO measurements made by a ground-based Fourier Transform spectrometer (FTIR) and the "measurement of pollution in the troposphere" instrument (MOPITT), which is mounted on the EOS Terra platform. The main conclusions for this case are the following: (1) Direct comparison of retrieved profiles is not satisfactory, because the expected standard deviation of the difference is around half of the expected natural variability of the true atmospheric profiles. (2) Comparison of the MOPITT profile retrieval with a simulation using FTIR is much more useful, though still not ideal, with expected standard deviation of differences of around 20% of the expected natural variability. (3) Direct comparison of total columns gives an expected standard deviation of about 9%, while comparison of MOPITT with a simulation derived from FTIR improved this to 8%. (4) There is only one combination of measured signals that can be usefully compared. The difference is expected to have a standard deviation of about 5.5% of the expected natural variability, which is mostly due to noise.   INDEX TERMS: 0394 Atmospheric Composition and Structure: Instruments and techniques; 3260 Mathematical Geophysics: Inverse theory; 3360 Meteorology and Atmospheric Dynamics: Remote sensing; 3394 Meteorology and Atmospheric Dynamics: Instruments and techniques; 9820 General or Miscellaneous: Techniques applicable in three or more fields; KEYWORDS: intercomparison, remote sensing, inverse methods, optimal estimation, profile retrieval

1.   Introduction

[2]   Remote-sounding methods are used to make a wide range of atmospheric measurements, both from satellites and from the surface. Because a remote sounder measures some more or less complicated function of the quantity of interest, a retrieval or inverse process is usually required to derive the final product. Consequently, profiles obtained from remote sounding are not usually simple measurements of the actual state of the atmosphere with independent errors, rather they are the best estimate the experimenter could make given the measurements and whatever prior knowledge about the state of the atmosphere is available, and are often an estimate of some smoothed function of the profile, with errors which are correlated between different altitudes. To validate such methods it is necessary to understand the relationship between the quantity measured and the true atmospheric state, and to carry out a detailed error analysis [e.g., Rodgers, 1990; Connor et al., 1995].

[3]   It is also necessary for validation purposes to compare measurements of the same quantity from different instruments. If an "ideal" instrument were available, one which directly measured the quantity of interest, at the time and location measured by the remote-sounding instrument, even though with some measurement error, then a comparison with a single remote sounder would be straightforward. The effect of the remote sounder could be simulated using the direct measurement, and compared with the actual remote

---

[1]On leave from Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK.

measurement, either as the original measurements or as the retrieval.

[4] The question of establishing appropriate criteria for adequate temporal and spatial coincidence for intercomparison will not be dealt with in this paper. This has been well covered in the literature, generally in terms of examining the variation of the statistics of differences as a function of spatiotemporal separation, for example see the discussion by [*Russell and Smit*, 1998]. Rather we will deal with the techniques needed for comparing remote sounders with different characteristics, a question which has received little attention so far. In comparing two remote sounders we try to validate both without knowing the actual value of the quantity being measured. The intercomparison of different remote sounding instruments is not a simple process, and the proper statistical methods are not often used. Not only must the effects of random and systematic measurement error be considered, but also the different observing system characteristics. The aim of this paper is to set these methods on a proper footing, illustrating the approach with comparisons of real remote-sounding instruments.

[5] In section 2 we describe the basic characteristics of a measurement and retrieval in the linear approximation, and outline some of the concepts that will be used. We consider ways of comparing both the measured signals and the retrieved profiles. In section 3 we present the formalism for comparing retrieved profiles, and in section 4 we extend it for derived quantities such as total column amounts. In section 5 we consider what can be compared directly without reference to the retrieval process, thus allowing us to distinguish effects that might be due to the retrieval process. After some comments on nonlinearity in section 6, we illustrate the approach in sections 7 and 8 by calculating the expected sizes of the diagnostics discussed for a comparison of CO measurements by a ground-based Fourier Transform spectrometer and the "measurement of pollution in the troposphere" instrument MOPITT, which flew on the EOS Terra platform.

[6] At the time of writing, flight data from MOPITT is not yet at a stage of development where these techniques can be usefully applied. We hope to hope to make formal comparisons of real data in a future publication.

1.1. Terminology and Notation

[7] We will use the term "measurement" generically to refer to both the quantity originally measured and the quantity retrieved. The quantity originally measured (perhaps after a simple conversion to appropriate scientific units), for example radiance or transmittance, will be described by the term "(measured) signal". The quantity finally required will be described as the "retrieved quantity" or a "retrieval". The combination of instrument and retrieval method will be called the "observing system".

[8] The theory will be developed in terms of the algebra of vectors and matrices. A matrix will be denoted by an upper case bold symbol, e.g., **A**, and a column vector by a lower case bold symbol, e.g., **x**. The atmospheric state to be retrieved by the two observing systems is denoted by a state vector **x**, of length n, containing for example mixing ratios on a set of pressure levels. The signal measured by instrument i is denoted by a measurement vector $\mathbf{y}_i$, of length m_i, containing such things as signals for each channel. The

instrument model includes a forward model $\mathbf{f}_i(\mathbf{x})$ and measured signal error $\boldsymbol{\epsilon}_i$, both random and systematic, such that:

$$\mathbf{y}_i = \mathbf{f}_i(\mathbf{x}) + \boldsymbol{\epsilon}_i \qquad (1)$$

The statistics of the measured signal error must be known, and will be described by a covariance matrix $\mathbf{S}_{\epsilon_i} = \mathcal{E}\{\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^T\}$.

1.2. Basic Requirements

[9] The purpose of an intercomparison is to determine whether the observing systems agree within their known limitations, i.e. the extent to which the forward models and the error covariances satisfactorily describe the instruments and their errors, and the retrieval methods reproduce the atmosphere. To this end the following considerations are relevant:

1. For simplicity, the observing systems should both retrieve the same target quantity. The definition of the state vector **x** includes the parameter retrieved, such as trace gas mixing ratio or concentration, but there may also be instrument-specific parameters in the state vectors that are not targets for comparison but need to be retrieved, such as surface reflectivity or instrumental offsets. The definition of **x** also includes the grid on which the targets for comparison are retrieved, and any interpolation rule implied. For the purpose of this paper we take all profiles to be represented on the same vertical grid, fine enough that representation error can be ignored. If a retrieval is carried out on a coarse grid, then an interpolation must be used before the methods described here are applied. It is possible to compare observing systems that retrieve different representations of the state, but this leads to extra complications in the comparison that we prefer to leave to a future publication.

2. We must specify an ensemble of states over which the comparison is to take place. For a theoretical assessment of instrumental capabilities, this can be chosen to be all-encompassing, but need not include states grossly outside the range of atmospheric possibilities. For an actual intercomparison over an ensemble of real atmospheres, it should describe the real ensemble as far as possible, but in many cases it is not well known, so we must be able to treat these cases too. It need not be the same as any a priori used by either or both retrieval methods, but it could be if appropriate. For convenience we will assume the comparison ensemble to be described by a Gaussian distribution, with mean $\mathbf{x}_c$ and covariance $\mathbf{S}_c$.

3. The intercomparison must allow for differences between a priori states used by different observing systems. In this analysis a priori is used primarily as the linearization point for the characterization of the retrieval process in terms of its averaging kernel; it is a profile which would be unchanged by measurement and retrieval in the absence of measured signal error.

[10] We will develop first the necessary theoretical methods for linear problems, including cases in which the forward model can be adequately linearized within the range of the comparison ensemble. In these cases the instrument model may be described by:

$$\mathbf{y}_i - \mathbf{y}_0 = \mathbf{K}_i(\mathbf{x} - \mathbf{x}_0) + \boldsymbol{\epsilon}_i \qquad (2)$$

where $\mathbf{K} = \partial\mathbf{f}/\partial\mathbf{x}$ is the $m \times n$ weighting function matrix, $\mathbf{x}_0$ is a linearization point, and $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$.

[11] A retrieval method may be described as optimal. By this term we will denote one which is, for example, a maximum a posteriori method, providing some characteristic of the Bayesian ensemble consistent with the measurement, its noise statistics and some a priori state. However, it is important to note that an optimal retrieval depends on both the measurement and the a priori, and an inverse method which is optimal with respect to its own a priori, $\mathbf{x}_a$, $\mathbf{S}_a$, will not be optimal with respect to a different one such as $\mathbf{x}_c$, $\mathbf{S}_c$. The intercomparison method must be able to treat observing systems without regard to whether the retrieval methods are optimal in any sense.

2. Characterization and Information Content of a Retrieval

[12] An intercomparison of two observing systems should start with a comparison of their theoretical capabilities. This will include such things as (1) the averaging kernels, to characterize the altitude resolution and vertical range of the retrievals (2) the retrieval noise covariance (systematic and random components), to characterize the accuracy and precision of the retrieval (3) the degrees of freedom for signal, i.e. the number of independent parameters defining the improvement of the profile over the a priori and (4) the Shannon information content, as an overall measure of signal-to-noise ratio relative to the a priori.

2.1. Averaging Kernels and Errors

[13] Following the approach of *Rodgers* [1990, 2000] we relate the retrieved quantity $\hat{\mathbf{x}}$ to the true quantity $\mathbf{x}$ and to any a priori used in its retrieval by

$$\hat{\mathbf{x}} - \mathbf{x}_a = \mathbf{A}(\mathbf{x} - \mathbf{x}_a) + \boldsymbol{\epsilon}_x \qquad (3)$$

where $\mathbf{A}$ is the averaging kernel matrix, ideally (but not usually) a unit matrix, and the error $\boldsymbol{\epsilon}_x$ in $\hat{\mathbf{x}}$ is due to both random and systematic errors in the measured signal and in the instrument's forward model. If $\mathbf{x}$ describes an altitude profile of some quantity, the l-th row $\mathbf{a}_l^T$ of $\mathbf{A}$ can be regarded as a smoothing function for the altitude corresponding to l. It should be a peaked function, the width of the peak qualitatively describing the vertical resolution of the retrieval. The range of altitudes over which the observing system is sensitive to the profile is indicated by the range of altitudes for which the area of the averaging kernel (the sum of its elements) is of order unity. Outside this range the area will tend toward zero as the retrieval tends toward $\mathbf{x}_a$ plus a measurement error component.

[14] Equation (3) may be used as the basis for comparing real measurements to other measurements or model calculations with much higher vertical resolution and much less dependence on their a priori. As pointed out by [*Connor et al.*, 1994], we may then ignore the influence of the averaging kernels and the a priori on the high resolution profile, $\mathbf{x}_h$, considering it as if it were an "ideal" measurement, and substitute it for $\mathbf{x}$ in (3). We then obtain

$$\mathbf{x}_s = \mathbf{x}_a + \mathbf{A}(\mathbf{x}_h - \mathbf{x}_a) \qquad (4)$$

where $\mathbf{x}_s$ is a smoothed version of $\mathbf{x}_h$. In particular, $\mathbf{x}_s$ is the profile which would be retrieved by our lower resolution measurement in the absence of retrieval error if the high resolution profile $\mathbf{x}_h$ were the true atmospheric profile. By comparing $\hat{\mathbf{x}}$ to $\mathbf{x}_s$ rather than to $\mathbf{x}_h$ we eliminate the effects of the lower resolution measurement's smoothing. The more general case, where the smoothing of both instruments must be considered, is addressed in section 3.

[15] When intercomparing real observing systems, another complication arises that there may be extra elements in the state vectors which are not common to the two observing systems. We wish to compare only the parts of the two state vectors that are common and refer to the atmospheric profile. In the case to be examined in this paper, the FTIR state vector contains a constituent profile to be compared, plus elements such as those describing the instrument characteristics and amounts of other constituents, which are not to be compared. The MOPITT state vector contains the constituent profile to be compared, plus the surface temperature and emissivity. Hence the averaging kernel description of the observing systems, equation (3), must be generalized.

[16] Let us write the full state vector for an instrument as $\mathbf{s}$, made up of $\mathbf{x}$, the part to be compared, and $\mathbf{e}$, the extra elements not to be compared. Then equation (3) becomes

$$\hat{\mathbf{s}} - \mathbf{s}_a = \begin{pmatrix} \hat{\mathbf{x}} - \mathbf{x}_a \\ \hat{\mathbf{e}} - \mathbf{e}_a \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{xx} & \mathbf{A}_{xe} \\ \mathbf{A}_{ex} & \mathbf{A}_{ee} \end{pmatrix} \begin{pmatrix} \mathbf{x} - \mathbf{x}_a \\ \mathbf{e} - \mathbf{e}_a \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_x \\ \boldsymbol{\epsilon}_e \end{pmatrix} \qquad (5)$$

so that the component for $\mathbf{x}$ alone is

$$\hat{\mathbf{x}} - \mathbf{x}_a = \mathbf{A}_{xx}(\mathbf{x} - \mathbf{x}_a) + \mathbf{A}_{xe}(\mathbf{e} - \mathbf{e}_a) + \boldsymbol{\epsilon}_x \qquad (6)$$

giving the error in $\hat{\mathbf{x}}$ as

$$\hat{\mathbf{x}} - \mathbf{x} = (\mathbf{A}_{xx} - \mathbf{I})(\mathbf{x} - \mathbf{x}_a) + \mathbf{A}_{xe}(\mathbf{e} - \mathbf{e}_a) + \boldsymbol{\epsilon}_x \qquad (7)$$

where $\mathbf{I}$ is a unit matrix. The error in $\hat{\mathbf{x}}$ now includes an extra term $\mathbf{A}_{xe}(\mathbf{e} - \mathbf{e}_a)$ due to interference between the profile part and the rest of the state vector. The covariance of all error terms other than that due to smoothing of $\mathbf{x}$ by the averaging kernel is, taking $\mathbf{x}$ and $\mathbf{e}$ to be uncorrelated,

$$\mathbf{S}_x = \mathbf{A}_{xe}\mathbf{S}_{ea}\mathbf{A}_{xe}^T + \mathbf{S}_{\epsilon_x} \qquad (8)$$

where $\mathbf{S}_{ea}$ is the a priori covariance of $\mathbf{e}$. The contribution $\mathbf{A}_{xe}(\mathbf{e} - \mathbf{e}_a)$ and its covariance may be described by the term "interference error".

[17] The term $\boldsymbol{\epsilon}_x$ is due only to errors (noise plus systematic errors) in the measured signal and the forward model. We will call this "measurement error". Measurement error plus interference error is the error in the profile as smoothed by the averaging kernel. We will call this "retrieval error". The total error $\hat{\mathbf{x}} - \mathbf{x}$ is the retrieval error plus the smoothing error.

[18] The measurement error $\boldsymbol{\epsilon}_x$ is not usually independent between altitude levels. Its covariance matrix $\mathbf{S}_{\epsilon_x}$ is not diagonal. If the measured signal error $\boldsymbol{\epsilon}$ has covariance $\mathbf{S}_\epsilon$, often diagonal, then

$$\mathbf{S}_{\epsilon_x} = \mathbf{G}\mathbf{S}_\epsilon\mathbf{G}^T \qquad (9)$$

where $\mathbf{G} = \partial\hat{\mathbf{x}}/\partial\mathbf{y}$ is the retrieval gain, or contribution function matrix. We may characterize the measurement error, or any error covariance contribution, by its diagonal elements or more generally by its eigenvectors and values. The diagonal elements are the traditional variances of the retrieval at each level, and give an indication of the error in the retrieval. However, because of the correlations between levels this is not the whole story. A more general approach to understanding correlated errors is to consider the "error patterns", or eigenvectors of $\mathbf{S}_{\epsilon_x}$ scaled by the square roots of the eigenvalues. These are independent quantities with unit variance [*Rodgers*, 1990].

2.2.   Information Content and Degrees of Freedom for Signal

[19]   The Shannon information content H of the measurement and its degrees of freedom for signal d_s are closely related. They both characterize the change in the knowledge of the state as a result of making a measurement, so that they must be calculated with respect to a specified a priori. For a particular optimal estimator, take the a priori covariance $\mathbf{S}_a$ to describe the quality of knowledge of the state before the measurement is made, and the total retrieval covariance $\hat{\mathbf{S}}_x$ to describe the knowledge afterward. We can define a generalization of the noise to signal variance ratio (i.e. the reciprocal of the s.n.r. squared) by the symmetric matrix $\mathbf{R} = \mathbf{S}_a^{-1/2}\hat{\mathbf{S}}_x\mathbf{S}_a^{-1/2}$. The information content is then $H = -\frac{1}{2}\log_2 |\mathbf{R}|$ bits, and the degrees of freedom for signal is $d_s = \text{trace}(\mathbf{I} - \mathbf{R})$.

[20]   Information content may also be defined for the original measured signal in terms of a noise-to-signal matrix defined as $\mathbf{R}_y = \mathbf{S}_{cy}^{-1/2}\mathbf{S}_{\epsilon}\mathbf{S}_{cy}^{-1/2}$ where $\mathbf{S}_{cy}$ is the covariance of the signal over the comparison ensemble. For a retrieval optimal with respect to $\mathbf{S}_{cy}$, the information content of the signal would be the same as that for the retrieval, but for a general (nonoptimal) retrieval, this may not be the case.

[21]   Information content of a retrieval depends on the a priori covariance matrix. With respect to a different a priori covariance, e.g., $\mathbf{S}_c$, a particular retrieval may be nonoptimal, and contain less (even negative) information.

3.   Comparison of Retrieved Profiles

[22]   A comparison of the retrieved profiles is an end-to-end test which depends on both the forward model and the inverse method being correct. A comparison must be applicable to any retrieval method, not just "optimal" ones, so is based on the general linearized description of an observing system given by equation (3).

[23]   In general the two retrievals will have been computed with different a priori, both different from the comparison ensemble. To compare them directly it is simpler to first transform the characterization equation (3) so that both are expressed in terms of departures from the comparison ensemble mean. We can put

$$\hat{\mathbf{x}}_i - \mathbf{x}_c + (\mathbf{A}_i - \mathbf{I})(\mathbf{x}_{ai} - \mathbf{x}_c) = \mathbf{A}_i(\mathbf{x} - \mathbf{x}_c) + \epsilon_{xi} \quad (10)$$

so that adding the term $(\mathbf{A}_i - \mathbf{I})(\mathbf{x}_{ai} - \mathbf{x}_c)$ to each retrieval will adjust them for different a priori. Remember that even if the retrievals are optimal with respect to their own a priori, they will not normally be optimal with respect to the

comparison ensemble. For the rest of this section we will assume that this adjustment has been carried out and the linearization point for the characterization is $\mathbf{x}_c$.

[24]   When comparing two independent scalar measurements, x_1 and x_2, with variances σ_1^2 and σ_2^2, the test to see whether the error bars overlap involves $\chi^2 = (x_1 - x_2)^2/(\sigma_1^2 + \sigma_2^2)$. Similarly, a test for two independent vector measurements, $\hat{\mathbf{x}}_1, \hat{\mathbf{S}}_1$ and $\hat{\mathbf{x}}_2, \hat{\mathbf{S}}_2$ would use the following χ^2:

$$\chi^2 = (\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2)^T(\hat{\mathbf{S}}_1 + \hat{\mathbf{S}}_2)^{-1}(\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2) \quad (11)$$

However this is not appropriate here because the retrievals may not be independent. If the averaging kernels are not unit matrices, the total errors will be correlated because both have a component which depends on the state.

[25]   The difference between the retrievals is

$$\boldsymbol{\delta} = \hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2 = (\mathbf{A}_1 - \mathbf{A}_2)(\mathbf{x} - \mathbf{x}_c) + \epsilon_{x1} - \epsilon_{x2}, \quad (12)$$

a random variable with covariance

$$\mathbf{S}_\delta = (\mathbf{A}_1 - \mathbf{A}_2)^T\mathbf{S}_c(\mathbf{A}_1 - \mathbf{A}_2) + \mathbf{S}_{x1} + \mathbf{S}_{x2}. \quad (13)$$

For particular comparisons, given $\mathbf{S}_\delta$, we might then do χ^2 tests on $\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2$ using

$$\chi^2 = (\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2)^T\mathbf{S}_\delta^{-1}(\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2). \quad (14)$$

This should indicate whether something is wrong, but of course would not tell you what. Unfortunately, it is possible for $\mathbf{S}_\delta$ to be singular, making χ^2 impossible to calculate in this manner. This will happen if the two measurements have some null-space in common. For example $\mathbf{S}_\delta$ will certainly be singular if the total number of $\mathbf{y}$ elements, $m_1 + m_2$, is smaller than the number of elements n in $\mathbf{x}$. For an underconstrained measurement, the components which are not measured are provided by the a priori, and the same components will be provided to both retrievals. The cure is to calculate χ^2 only in the subspaces which are actually measured. A method of doing this using an eigenvector expansion of $\mathbf{S}_\delta$ is detailed in Appendix A.

[26]   Clearly the direct comparison of retrievals is not as straightforward as it seems at first sight. We therefore look at other quantities that might be compared.

4.   Comparison of Derived Quantities

[27]   One particularly useful comparison is that of a total column from one instrument with a column or profile from another. Other useful derived quantities include such things as such as thicknesses in the case of temperature profiles, or layer amounts for constituent profiles. Derived quantities, like profile retrievals, are not necessarily perfect estimates of the desired quantity plus noise, but will have some imperfect averaging kernel. In order to compare functions of profiles we will first show how they are best derived from the measurements, and then consider how these estimates should be compared.

4.1.   Best Estimate of a Function of the State Vector

[28]   Is it optimal to determine a function of a profile by calculating the same function of a retrieved profile, or is

there some way to obtain it more accurately directly from the measured signals? Is the total column best estimated by integrating a retrieval, or is there a better way?

[29] Let us consider estimating some scalar-valued linear function $z(\mathbf{x}) = z_c + \mathbf{g}^T (\mathbf{x} - \mathbf{x}_c)$ of the underlying state vector. We might expect the best estimate to be $\hat{z} = z_c + \mathbf{g}^T(\hat{\mathbf{x}} - \mathbf{x}_c)$ where $\hat{\mathbf{x}}$ is the optimal retrieved state. The following simple argument, based on the Bayesian approach to retrieval, shows that it is, taking the retrieval to be the expectation of $\mathbf{x}$ given the measurement. Let the p.d.f. of the state given the measurements be $P(\mathbf{x}|\mathbf{y})$. The expected value of the function $z(\mathbf{x})$ is

$$\hat{z} = \int P(\mathbf{x}|\mathbf{y}) \big[z_c + \mathbf{g}^T (\mathbf{x} - \mathbf{x}_c) \big] \mathrm{d}\mathbf{x} \qquad (15)$$

where $\mathrm{d}\mathbf{x}$ is a volume element of state-space. Provided $\mathbf{g}$ does not depend on $\mathbf{x}$, this can be rewritten as

$$\hat{z} = z_c + \mathbf{g}^T \int P(\mathbf{x}|\mathbf{y})(\mathbf{x} - \mathbf{x}_c)\mathrm{d}\mathbf{x}$$
$$= z_c + \mathbf{g}^T(\hat{\mathbf{x}} - \mathbf{x}_c) \qquad (16)$$

[30] It is easy to show that the error covariance of $\hat{z}$ is $\mathbf{g}^T\mathbf{S}\mathbf{g}$, its measurement error is $\mathbf{g}^T\mathbf{S}_x\mathbf{g}$, and its derivative with respect to the true state $\partial z / \partial \mathbf{x}$ is $\mathbf{g}^T\mathbf{A}$. The latter is the same kind of quantity as a row of the averaging kernel matrix, $\mathbf{A} = \partial\hat{\mathbf{x}}/\partial\mathbf{x}$, though conceptually different because z is not a state vector element. Nevertheless we will describe $\mathbf{a}^T = \mathbf{g}^T\mathbf{A}$ as an averaging kernel, mainly from a desire not to invent yet more terminology.

4.2. Best Estimate of a Function Given a Retrieval

[31] This is a slightly different question from that of the previous section. Given an arbitrary retrieved profile, not necessarily an optimal one, what is the best estimate of a function of the state, such as the total column? Any retrieval may be written in terms of an averaging kernel (equation (3)):

$$\hat{\mathbf{x}} - \mathbf{x}_c = \mathbf{A}(\mathbf{x} - \mathbf{x}_c) + \boldsymbol{\epsilon}_x, \qquad (17)$$

after adjusting $\hat{\mathbf{x}}$ to the comparison ensemble as in section 3. We wish to find, for the comparison ensemble, the best estimate $\hat{z}$ of $z = z_c + \mathbf{g}^T(\mathbf{x} - \mathbf{x}_c)$. Clearly we could simply calculate $z_c + \mathbf{g}^T(\hat{\mathbf{x}} - \mathbf{x}_c)$, i.e. $z_c + \mathbf{g}^T(\mathbf{A}(\mathbf{x} - \mathbf{x}_c) + \boldsymbol{\epsilon}_x)$, but $\mathbf{g}^T\mathbf{A}$ is not the same as $\mathbf{g}^T$, and there may be a better estimate—though not in the case of the optimal estimator, as shown by section 4.1.

[32] We note that equation (17) is of exactly the same algebraic form as the instrument model, equation (1), so we use it as an instrument model in a linear optimal retrieval:

$$\tilde{\mathbf{x}} = \mathbf{x}_c + \mathbf{S}_c\mathbf{A}^T \big(\mathbf{A}\mathbf{S}_c\mathbf{A}^T + \mathbf{S}_{\epsilon_x} \big)^{-1}(\hat{\mathbf{x}} - \mathbf{x}_c) \qquad (18)$$

where $\tilde{\mathbf{x}}$ gives a retrieval optimized for the comparison ensemble, insofar as the information content of the measured signal is still present in the nonoptimal retrieval. The matrix to be inverted should never be singular, as the error

covariance of the original retrieval should never be singular. We can now use the result of section 4.1 to obtain

$$\hat{z} = z_c + \mathbf{g}^T(\tilde{\mathbf{x}} - \mathbf{x}_c)$$
$$= z_c + \mathbf{g}^T\mathbf{S}_c\mathbf{A}^T \big(\mathbf{A}\mathbf{S}_c\mathbf{A}^T + \mathbf{S}_{\epsilon_x} \big)^{-1}(\hat{\mathbf{x}} - \mathbf{x}_c) \qquad (19)$$

The averaging kernel and retrieval error for this process can be obtained by substituting equation (17), giving the averaging kernel for $\mathbf{g}$ as

$$\mathbf{a}^T = \mathbf{g}^T\mathbf{S}_c\mathbf{A}^T \big(\mathbf{A}\mathbf{S}_c\mathbf{A}^T + \mathbf{S}_{\epsilon_x} \big)^{-1}\mathbf{A} \qquad (20)$$

and retrieval error variance

$$\sigma^2 = \mathbf{g}^T\mathbf{S}_c\mathbf{A}^T \big(\mathbf{A}\mathbf{S}_c\mathbf{A}^T + \mathbf{S}_{\epsilon_x} \big)^{-1}\mathbf{S}_{\epsilon_x}\big(\mathbf{A}\mathbf{S}_c\mathbf{A}^T + \mathbf{S}_{\epsilon_x} \big)^{-1}\mathbf{A}\mathbf{S}_c\mathbf{g} \quad (21)$$

Incidentally, equation (18) provides a means of converting a retrieval which is not optimal for the ensemble being considered into one which is, provided all of the information in the original measurements has survived the original retrieval method.

4.3. Comparison of Two Estimates of the Total Column

[33] For instrument i the retrieved total column $\hat{c}_i$ is related to the true state by:

$$\hat{c}_i - c_c = \mathbf{a}_i^T(\mathbf{x} - \mathbf{x}_c) + \epsilon_{ci} \qquad (22)$$

where $\mathbf{a}_i$ is the total column averaging kernel, and ϵ_{ci} is the measurement error of the column retrieval. Remember that the ideal $\mathbf{a}_i$ will not necessarily be a vector containing all ones—it depends on the state vector representation. The difference between total column estimates for the two observing systems is:

$$\hat{c}_1 - \hat{c}_2 = (\mathbf{a}_1 - \mathbf{a}_2)^T(\mathbf{x} - \mathbf{x}_c) + \epsilon_{c1} - \epsilon_{c2} \qquad (23)$$

and the variance of the difference is

$$\sigma^2(\hat{c}_1 - \hat{c}_2) = (\mathbf{a}_1 - \mathbf{a}_2)^T\mathbf{S}_c(\mathbf{a}_1 - \mathbf{a}_2) + \sigma_{c1}^2 + \sigma_{c2}^2 \qquad (24)$$

which is of the same form as the covariance of the difference of retrieved profiles, equation (13), the first term being the contribution of the different averaging kernels (the smoothing error), and the other terms the independent measurement error contributions.

[34] We may find that for one or both observing systems the total column averaging kernel is not as close to the total column operator as might be desired, and consequently the retrieved total column does not well represent the true total column. In such a case a comparison with the other observing system will be poor (unless, of course, it has the same averaging kernel). As an alternative, we can use the second observing system to try to retrieve what the first system claims to produce as its retrieved total column, using the method described in section 4.1 or section 4.2. If retrieval 2 is optimal with respect to the comparison ensemble, the best estimate is

$$\hat{c}_{12} = c_c + \mathbf{a}_1^T(\hat{\mathbf{x}}_2 - \mathbf{x}_c) \qquad (25)$$

and its averaging kernel is $\mathbf{a}_1^T \mathbf{A}_2$. The smoothing error of the comparison of $\hat{c}_{12}$ with $\hat{c}_1$ should be smaller than that of the direct comparison of $\hat{c}_1$ and $\hat{c}_2$. The difference between these two estimates is:

$$\hat{c}_1 - \hat{c}_{12} = \mathbf{a}_1^T(\mathbf{I} - \mathbf{A}_2)(\mathbf{x} - \mathbf{x}_c) + \epsilon_{c1} - \mathbf{a}_1^T \epsilon_{x_2} \qquad (26)$$

and the variance of the difference is

$$\sigma^2(\hat{c}_1 - \hat{c}_{12}) = \mathbf{a}_1^T(\mathbf{I} - \mathbf{A}_2)\mathbf{S}_c(\mathbf{I} - \mathbf{A}_2)^T\mathbf{a}_1 + \sigma_{c1}^2 + \mathbf{a}_1^T \mathbf{S}_{x_2}\mathbf{a}_1 \quad (27)$$

If retrieval 2 is not optimal with respect to the comparison ensemble, the algebra is more complicated, following section 4.2 rather than section 4.1, but is otherwise straightforward.

4.4. Simulating One Profile Retrieval With the Other

[35] We can apply the same philosophy as in the previous section to the profile retrieval itself. To reduce smoothing error in the comparison, that is to compare things which are more alike, we can use the method of section 4.1 or section 4.2 to attempt to reproduce the (imperfect) profile retrieval of instrument 1 by modifying the retrieval of instrument 2. If retrieval 2 is optimal with respect to the comparison ensemble, the best estimate is

$$\hat{\mathbf{x}}_{12} = \mathbf{x}_c + \mathbf{A}_1(\hat{\mathbf{x}}_2 - \mathbf{x}_c) \qquad (28)$$

and its averaging kernel is $\mathbf{A}_1\mathbf{A}_2$, so that it is simply retrieval 2 smoothed with the averaging kernel of retrieval 1. The smoothing error of the comparison of $\hat{\mathbf{x}}_{12}$ with $\hat{\mathbf{x}}_1$ should be smaller than that of the direct comparison of $\hat{\mathbf{x}}_2$ and $\hat{\mathbf{x}}_1$. Note that this comparison is not symmetrical, in that one order of comparison would be expected to give a more satisfactory result than the other, because the higher resolution measurement would be expected to reproduce the lower resolution measurement better than the other way round. As above, the difference between these two estimates is:

$$\begin{aligned}\delta_{12} &= \hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_{12} \\ &= (\mathbf{A}_1 - \mathbf{A}_1\mathbf{A}_2)(\mathbf{x} - \mathbf{x}_c) + \epsilon_{x1} - \mathbf{A}_1\epsilon_{x2}\end{aligned} \qquad (29)$$

(compare equation (12)) and the covariance of the difference is

$$\mathbf{S}_{\delta_{12}} = (\mathbf{A}_1 - \mathbf{A}_1\mathbf{A}_2)\mathbf{S}_c(\mathbf{A}_1 - \mathbf{A}_1\mathbf{A}_2)^T + \mathbf{S}_{x_1} + \mathbf{A}_1\mathbf{S}_{x_2}\mathbf{A}_1^T \quad (30)$$

As in the case of total columns, if retrieval 2 is not optimal with respect to the comparison ensemble, then equation (18) should be used to convert it, and to provide an improved averaging kernel matrix.

5.   Comparing Other Functions of the Measurements
5.1.   Functions of Measured Signals

[36] A comparison of the measured signals, if it were possible, would not involve the retrieval process, so would

test only the understanding of the instrument behavior, i.e. the forward models. However, two different instruments typically measure different quantities in different units, and have different weighting functions measuring different parts of state-space, so cannot be compared directly. It is likely that they do not measure any subspace of state-space in common. Nevertheless, there may be combinations of their weighting functions that are similar enough for a useful comparison to be made for states within the comparison ensemble. To identify these, we look for linear combinations of the measured signals whose differences are expected to be less than noise for states comprising the comparison ensemble. The linear combinations themselves should have large variability compared with noise. This process is closely related to the time series analysis technique known as "maximum covariance analysis" [*Bretherton et al.*, 1992].

[37] Let both instruments be described by equation (2):

$$\mathbf{y}_1 - \mathbf{y}_{c1} = \mathbf{K}_1(\mathbf{x} - \mathbf{x}_c) + \epsilon_1 \qquad (31)$$

$$\mathbf{y}_2 - \mathbf{y}_{c2} = \mathbf{K}_2(\mathbf{x} - \mathbf{x}_c) + \epsilon_2 \qquad (32)$$

with $\mathbf{x}_c$, the mean of the comparison ensemble, as the linearization point. If the original instrument description uses a different linearization point, $\mathbf{x}_0$, $\mathbf{y}_0$, it can be rewritten in terms of $\mathbf{x} - \mathbf{x}_c$ using

$$\mathbf{y}_i - [\mathbf{y}_0 + \mathbf{K}_i(\mathbf{x}_c - \mathbf{x}_0)] = \mathbf{K}_i(\mathbf{x} - \mathbf{x}_c) + \epsilon_i \qquad (33)$$

so that $\mathbf{y}_{ci} = [\mathbf{y}_0 + \mathbf{K}_i(\mathbf{x}_c - \mathbf{x}_0)]$. Any instrument-specific extra elements $\mathbf{e}_i$ of the state vectors can be included by taking both instruments to use the same full state vector, including the extra instrument-dependent terms, i.e. $(\mathbf{x}, \mathbf{e}_1, \mathbf{e}_2)$, but with zero elements in the weighting function matrix for each instrument corresponding to the extra elements relating to the other instrument. For simplicity, in the rest of this section we will denote this full state vector by just $\mathbf{x}$.

[38] We look for linear functions $\mathbf{l}_i$, $i = 1, 2$, with the corresponding combination of signals

$$z_i = \mathbf{l}_i^T(\mathbf{y}_i - \mathbf{y}_{ci}), \qquad (34)$$

such that the expected value over the comparison ensemble of the squared difference, $\mathcal{E}\{(z_1 - z_2)^2\}$, is as small as possible. To avoid the trivial solution $z_i = 0$ we will take $\mathcal{E}\{z_i^2\} = 1$, with the consequence that minimizing $\mathcal{E}\{(z_1 - z_2)^2\}$ is the same as maximizing $\mathcal{E}\{z_1 z_2\}$. Using the linear instrument models for $\mathbf{y}_i - \mathbf{y}_{ci}$ we find the problem becomes one of maximizing

$$\begin{aligned}&\mathcal{E}\{\mathbf{l}_1^T(\mathbf{K}_1(\mathbf{x} - \mathbf{x}_c) + \epsilon_1)(\mathbf{K}_2(\mathbf{x} - \mathbf{x}_c) + \epsilon_2)^T \mathbf{l}_2\} \\ &= \mathbf{l}_1^T \mathbf{K}_1 \mathbf{S}_c \mathbf{K}_2^T \mathbf{l}_2\end{aligned} \qquad (35)$$

subject to the constraint

$$\begin{aligned}&\mathcal{E}\{\mathbf{l}_i^T(\mathbf{K}_i(\mathbf{x} - \mathbf{x}_c) + \epsilon_i)(\mathbf{K}_i(\mathbf{x} - \mathbf{x}_c) + \epsilon_i)^T \mathbf{l}_i\} \\ &= \mathbf{l}_i^T(\mathbf{K}_i \mathbf{S}_c \mathbf{K}_i^T + \mathbf{S}_{\epsilon_i})\mathbf{l}_i = 1\end{aligned} \qquad (36)$$

Using the transformation $\mathbf{l}_i = \left(\mathbf{K}_i \mathbf{S}_c \mathbf{K}_i^T + \mathbf{S}_{\epsilon_i}\right)^{-\frac{1}{2}} \mathbf{m}_i$ so that the constraint becomes $\mathbf{m}_i^T \mathbf{m}_I = 1$ is easy to see that $\mathbf{m}_1$ and $\mathbf{m}_2$ are the left and right singular vectors respectively of

$$\left(\mathbf{K}_1 \mathbf{S}_c \mathbf{K}_1^T + \mathbf{S}_{\epsilon_1}\right)^{-\frac{1}{2}} \mathbf{K}_1 \mathbf{S}_c \mathbf{K}_2^T \left(\mathbf{K}_2 \mathbf{S}_c \mathbf{K}_2^T + \mathbf{S}_{\epsilon_2}\right)^{-\frac{1}{2}} \quad (37)$$

and the corresponding combined weighting functions are

$$\breve{\mathbf{K}}_i = \mathbf{m}_i^T \left(\mathbf{K}_i \mathbf{S}_c \mathbf{K}_i^T + \mathbf{S}_{\epsilon_i}\right)^{-\frac{1}{2}} \mathbf{K}_i \quad (38)$$

The singular value λ_j corresponding to the j-th pair of singular vectors indicates the closeness of the fit, because $\mathcal{E}\{(z_1 - z_2)^2\} = 2 - 2\lambda_j$. Consequently the singular values should lie between zero and unity, and the closer to unity the better the fit. For any particular singular vector, the variance of $z_1 - z_2$ can also be expressed as the sum of terms corresponding to noise and smoothing error:

$$\mathcal{E}\left\{(z_1 - z_2)^2\right\} = \mathbf{l}_1^T \mathbf{S}_{\epsilon_1} \mathbf{l}_1 + \mathbf{l}_2^T \mathbf{S}_{\epsilon_2} \mathbf{l}_2$$
$$+ (\mathbf{l}_1 \mathbf{K}_1 - \mathbf{l}_2 \mathbf{K}_2) \mathbf{S}_c (\mathbf{l}_1 \mathbf{K}_1 - \mathbf{l}_2 \mathbf{K}_2)^T. \quad (39)$$

[39] For instruments with large numbers of channels, the matrices to be diagonalized are large, and the evaluation of e.g., $\left(\mathbf{K}_i \mathbf{S}_c \mathbf{K}_i^T + \mathbf{S}_{\epsilon_i}\right)^{-\frac{1}{2}}$ may be difficult. A method of dealing with this by means of a singular vector decomposition of $\tilde{\mathbf{K}} = \mathbf{S}_{\epsilon_i}^{-\frac{1}{2}} \mathbf{K}_i \mathbf{S}_c^{\frac{1}{2}}$, thus reducing the effective number of channels to not more than the length of the state vector, is given in Appendix B.

5.2. Functions of Retrieved Profiles

[40] Because the averaging kernels are different, the retrievals from two different observing systems are different functions of the atmospheric state. As with the comparison of measured signals, it is pertinent to ask whether there are any functions of the retrievals that can be properly compared. Are there any functions of the retrieved profiles that can be compared with negligible contribution from smoothing error?

[41] The retrieval characterization, equation (3), is of the same algebraic form as the linearized instrument model, equation (2), so the same formalism as described for direct measurements can in principle also be used for retrievals. We can therefore follow the approach of section 5.1, with $\mathbf{A}$ replacing $\mathbf{K}$ and $\mathbf{S}_{\epsilon_x}$ replacing $\mathbf{S}_\epsilon$.

[42] It is often felt desirable to plot profiles from two different instruments on the same diagram in order to compare them. If the averaging kernels are significantly different, this can be misleading. An alternative which is less so is to use this technique to identify which components of the retrieved profiles can be properly compared, and which cannot. Reconstructed profiles using only the linearization point and the properly comparable components can then be plotted for comparison. The omitted components which should not be compared can be plotted separately.

6. Nonlinearity

[43] We have developed the theory for linear or weakly nonlinear systems so far. However, most real observing systems are nonlinear, so any intercomparison must be able to treat this aspect. We distinguish between moderately nonlinear, meaning that a linear expansion is adequate to carry out an error analysis, even if it is not adequate to solve the inverse problem, and grossly nonlinear, for cases when nonlinearity must be treated within the bounds of the errors. The latter case is sufficiently difficult that Monte-Carlo techniques would be a suitable approach. We will only treat moderately nonlinear cases here.

[44] Moderately nonlinear retrievals can be compared using the linear formalism because we are only interested in the behavior of differences within the error bars. If differences are larger than that, then a linear analysis should indicate a problem, but the size of the problem may be wrongly estimated. We linearize about some point near the retrievals being compared, for example about $\mathbf{x}_0 = (\hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2)/2$. Let the nonlinear relation between the true profile and the retrieval be expressed in terms of a transfer function $\mathbf{T}_i$ for instrument i:

$$\hat{\mathbf{x}}_i = \mathbf{T}_i(\mathbf{x}, \mathbf{x}_{ia}) + \boldsymbol{\epsilon}_{x_i} \quad (40)$$

An expansion about $\mathbf{x} = \mathbf{x}_0$ will be valid provided that $\mathbf{x}_0$ is more or less within the error bars of both retrievals:

$$\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_{i0} + \mathbf{A}_{i0}(\mathbf{x} - \mathbf{x}_0) + \boldsymbol{\epsilon}_{x_i} \quad (41)$$

where $\hat{\mathbf{x}}_{i0} = \mathbf{T}_i(\mathbf{x}_0, \mathbf{x}_{ia})$ is the retrieved state corresponding to the linearization point, and the averaging kernel matrix $\mathbf{A}_{i0}$ is $\partial \mathbf{T}_i / \partial \mathbf{x}$ evaluated at $\mathbf{x} = \mathbf{x}_0$. Express this in terms of departures from the comparison ensemble mean:

$$\hat{\mathbf{x}}_i - \mathbf{x}_c = \hat{\mathbf{x}}_{i0} - \mathbf{x}_c - \mathbf{A}_{i0}(\mathbf{x}_0 - \mathbf{x}_c) + \mathbf{A}_{i0}(\mathbf{x} - \mathbf{x}_c) + \boldsymbol{\epsilon}_{x_i} \quad (42)$$

If, for the purpose of comparison, we replace $\hat{\mathbf{x}}_i$ by

$$\hat{\mathbf{x}}_i' = \hat{\mathbf{x}}_i - [\hat{\mathbf{x}}_{i0} - \mathbf{x}_c - \mathbf{A}_{i0}(\mathbf{x}_0 - \mathbf{x}_c)] \quad (43)$$

then we have the retrieval characterization in the original form but with the averaging kernel evaluated at or near the retrieval, and any of the methods discussed can be applied, provided that they do not require averaging or sampling over the whole comparison ensemble. For example the profile comparison χ^2 test of section 3 is still valid.

7. FTIR and MOPITT Observing Systems
7.1. FTIR Ground-Based Sounder
7.1.1. Instrument Characteristics

[45] Solar absorption spectra in the midinfrared are routinely recorded at a number of stations world-wide, especially in conjunction with the Network for Detection of Stratospheric Change [Kurylo and Zander, 2001]. CO and a variety of other gases have strong, well-resolved spectral lines in this region, which can be used for retrieval of column densities or low-resolution profiles. Fourier transform spectrometers are used for these measurements, hence we refer to them as "FTIR".

[46] In the current context we will focus on a measurement of the R3, P7, and P10 CO lines in the fundamental vibrational band, near 2100 cm^{-1} (4.7 μm). These lines were chosen for analysis because they are relatively free of
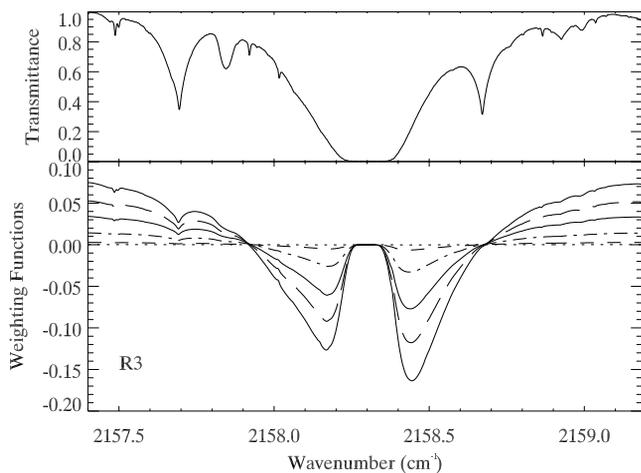
Figure 1. Measured R3 spectrum (upper) and weighting functions (lower) for the FTIR measurement. Each curve in the lower panel is the weighting function for CO at a selected altitude as a function of wave number, computed for the a priori state. 2 km: ———; 4 km: — —; 6 km: – · · ·; 10 km: – · – – ·; 14 km: – – – and 20 km: · · ·.

interference, close enough together for convenient simultaneous observation, and span a large range of opacities. This last feature provides good sensitivity from the ground into the lower stratosphere. A high quality spectrum measured at the NDSC site in Lauder, New Zealand is shown in the upper panels of Figures 1 and 2. The saturated absorption near the center of the region in Figure 1 is the CO R3 line. The strong absorptions to either side are due to N_2O. There are also weaker features from H_2O, CO_2, O_3, and CO in the solar atmosphere. In Figure 2, the slope in the P7 region is the wing of a CO_2 feature. The P10 line is blended with and ozone line of approximately equal strength. All other significant absorptions in the P7 and P10 windows are due to ozone.

[47] For the purpose of the current illustration we will assume that the basic measurement signal-to-noise is 300:1 and the maximum optical path difference is 130 cm (resolution of 0.006 cm^{-1}). While it is possible to achieve a measurement signal-to-noise of 300:1, there are regions where the spectroscopic data is poor, and systematic errors in the forward model will be the dominant source of error. In the R3 window shown in Figure 1 the vicinity of the solar CO lines and of a weak isotopic CO line at 2158.6 cm^{-1} are such regions. These effects introduce off-diagonal elements into the large measurement covariance matrix $\mathbf{S}_\epsilon$, and hence introduce computational difficulties into the retrieval. To prevent the systematic errors in the forward model from being mapped into the retrieval, we have made the simplifying assumption that the signal-to-noise in the vicinity of those lines is much lower, in the range 10–100, but uncorrelated between spectral elements so that $S\bar\epsilon$ can be taken as diagonal. Consequently the retrieval will not be strictly optimal.

7.1.2. Retrieval Method

[48] We use a retrieval based on [*Rodgers*, 1976]. The state vector includes the mean CO mixing ratio for 2 km layers from 0 to 26 km, scaled by the a priori mixing ratio, plus several extra terms making up the vector **e** discussed in

section 2.1: a wavelength shift, a background slope, column densities for the four interfering species, and parameters for a solar CO model. The CO column density is derived by integrating the retrieved profile, as described in section 4.2.

[49] The profile part of the a priori covariance matrix for FTIR is assumed to be a unit matrix, that is, the standard deviation of CO at each layer is assumed to be 100% and to be uncorrelated from layer to layer. This admittedly unrealistic assumption is one adopted by many investigators in an attempt to stabilize the retrieval without too much influence from the a priori information. With the state vector used, this corresponds to an a priori standard deviation of the total column of 43%. The a priori profile used is the same as that used for MOPITT, see section 7.2.2.

[50] The weighting functions for the FTIR measurement all peak at the surface and decrease monotonically with height. Therefore it is more useful in this case to plot the profile part of the weighting function, $K(\nu, z)$, as a function of ν for each altitude z, rather than as a function of z for each ν. These are shown in Figures 1 and 2, evaluated for the a priori profile. The structure of the spectrum can clearly be seen in the weighting functions.

7.1.3. Retrieval Characteristics

[51] The Shannon information content and the number of degrees of freedom for signal can be calculated as described in section 2.2 and are found to be 12.0 bits and 2.8 respectively for the profile part of the state vector. There are a further 116 bits of information and 14 degrees of freedom associated with the nonprofile elements of the state vector.

[52] The profile retrieval averaging kernels for FTIR are shown in Figure 3. These peak at or near their nominal altitude from 1 to 13 km, with near unit area in this range. There is an apparent resolution (full width at half height) of about 2 km near the surface and about 6 km between 5 and 11 km altitude.

[53] The r.m.s. error of the retrieved profile (square root of the error variance) and its components are shown in Figure 4. Remember that this is not a complete description of the errors, because they will be correlated between different heights, but it does provide an indication of the expected precision of the retrieval. The standard deviation of the
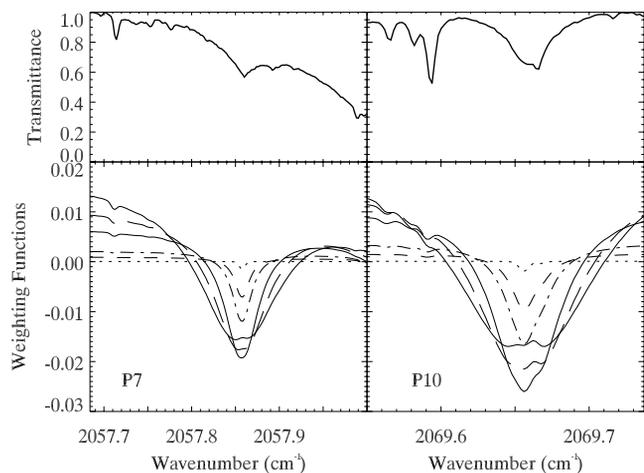
Figure 2. As Figure 1, but for the P7 and P10 CO lines. 2 km: ———; 6 km: — —; 10 km: – · · ·; 14 km: – · – – ·; 16 km: – – – and 30 km: · · ·.
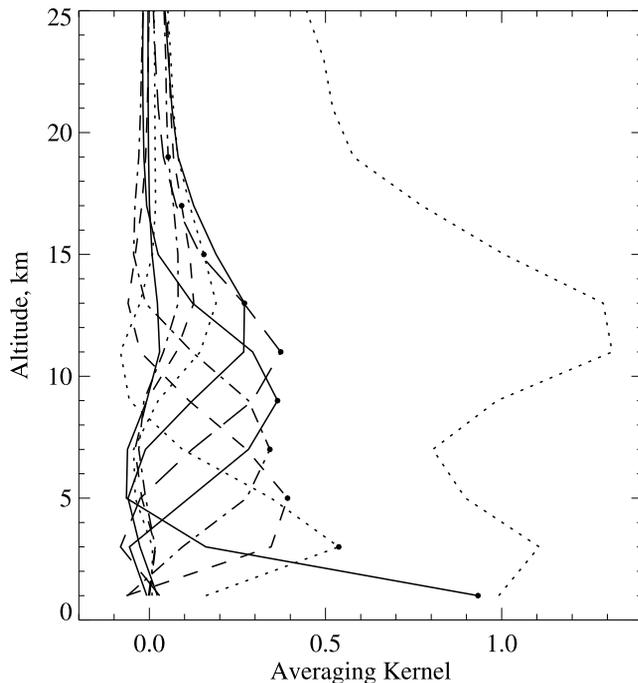
Figure 3. Averaging kernels for the FTIR retrieval for CO from 1–19 km. The dotted line to the right is the area of the kernels as a function of height. The nominal height of each kernel is marked by a filled circle.
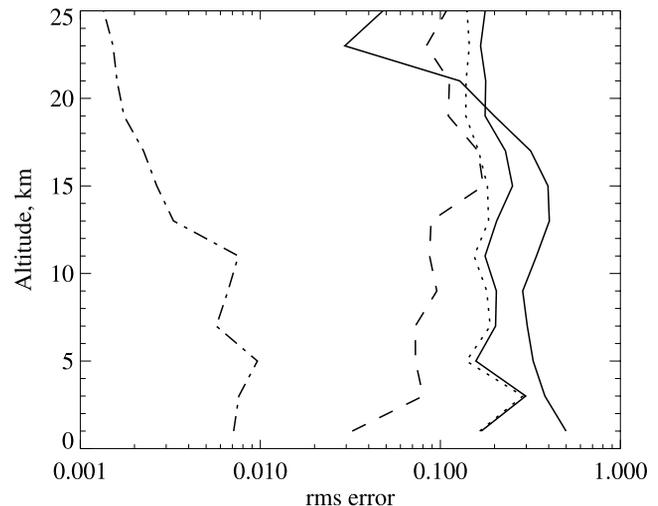
Figure 4. Standard deviations of the FTIR retrieval error and its components. Comparison ensemble (——); total (– · · ·); smoothing (– –); measurement error (· · · ·); interference error (– · – ·).

comparison ensemble is plotted for reference. The three components of the total error covariance are retrieval noise (the largest), smoothing error (somewhat smaller except around 11–13 km), and interference error (the smallest). It can be seen that retrieval noise comprises most of the total error in this case. The total error is greater than the comparison ensemble standard deviation above about 20 km because the FTIR retrieval is not optimal for this ensemble.

7.2. MOPITT Satellite-Based Sounder

7.2.1. Instrument Characteristics

[54] MOPITT is a nadir-viewing satellite-borne instrument which is part of the Earth Observing System's Terra platform. It uses a gas correlation technique to observe upwelling radiation in the 4.7μm and 2.3μm bands, where the fundamental and first overtone vibrational bands of CO are located [*Drummond*, 1992; *Drummond and Mann*, 1996]. It also observes the 2.2μm CH_4 band, but that does not concern us here. In addition to gas profiles, the surface emissivity and temperature are retrieved from the measured radiances, so the state vector of interest contains those two quantities along with the CO mixing ratios. The relevant radiative transfer in the atmosphere and instrument are discussed by [*Pan et al.*, 1995]. The MOPITT CO retrieval algorithm is described by [*Pan et al.*, 1998].

[55] There are 12 MOPITT radiance measurements pertinent to CO, eight thermal radiances in the 4.7μm band and four solar reflection signals in the 2.3μm band. The solar signals are preprocessed by dividing the gas correlation (difference) signal by the wideband (average) signal in order to eliminate the surface reflectivity, resulting in 10 signals as input to the retrieval [*Pan et al.*, 1995]. The weighting functions were kindly supplied by Dr Liwen Pan; the parts

corresponding to the CO profile are shown in Figure 5. The 4.7μm band is dominated by thermal emission from the atmosphere, its weighting functions peaking in the upper troposphere. By contrast the 2.3μm band is dominated by the reflection of solar radiation from both the surface and the atmosphere, and its weighting functions peak at the surface. This difference between the two wavelength bands is the primary source of height information in the MOPITT measurements. In addition, some further height discrimination is achieved by use of different gas correlation cell pressures.

7.2.2. Retrieval Method

[56] For the purpose of this paper we have not used the current MOPITT operational retrieval, as this uses a different vertical grid from FTIR, leading to complications that are beyond the scope of this paper. We have used an optimal estimator, with the same CO component of the state vector as FTIR, i.e. the mean CO mixing ratio for 2 km layers from
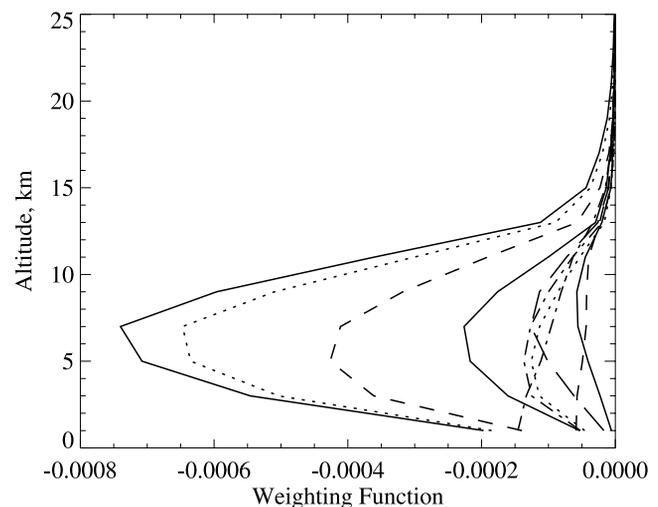
Figure 5. Weighting functions for the ten MOPITT CO channels, showing the contribution of CO as a function of altitude to the signals in the given MOPITT channel.
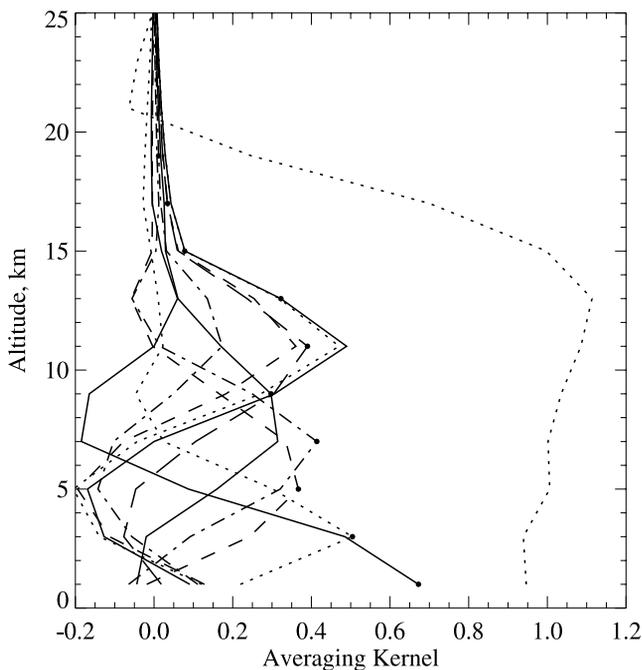
Figure 6. Averaging kernels for the MOPITT CO retrieval for 1–19 km. The dotted line to the right is the area of the kernels as a function of height. The nominal height of each kernel is marked by a filled circle.

0 to 26 km, scaled by the a priori mixing ratio, plus surface temperature and emissivity as extra terms. The a priori profile is based on an ensemble of aircraft CO measurements and the covariance matrix is calculated from the statistics of the ensemble. At altitudes above aircraft range a set of model output profiles was used [*Pan et al.*, 1998]. The measured signal errors are assumed to be uncorrelated noise (i.e. S_ϵ is diagonal), with values based on one orbit of level 1 flight data. The total column is obtained by integrating the retrieved profile.

7.2.3. Retrieval Characteristics

[57] The MOPITT profile retrieval averaging kernels are shown in Figure 6. Relative to the FTIR kernels (Figure 3) there is less resolution in the lowest 1–2 kilometers, and they are less well shaped than the FTIR averaging kernels, with peaks at or near their nominal altitude only up to 11 km. However, they have an area near unity from 0 km to about 15 km, with a resolution of about 5 km. The Shannon information content is 12.0 bits for the profile and 13.7 bits for the surface parameters. There are 3.1 degrees of freedom for signal for the profile and 2.0 for the surface parameters.

[58] The r.m.s. error of the retrieved profile and its components are shown in Figure 7. The standard deviation of the comparison ensemble is plotted for reference. The three components of the total error covariance are smoothing error (the largest), retrieval noise, and interference error (the smallest). It can be seen that smoothing error comprises most of the total error in this case.

8.   Comparisons: Agreement to be Expected

[59] To assess the level of agreement we would expect between FTIR and MOPITT, we have taken the comparison
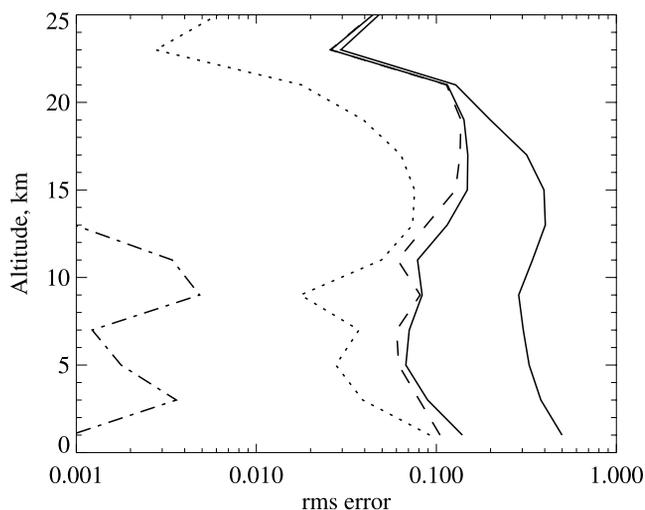
Figure 7. Standard deviations of the MOPITT retrieval error and its components. Comparison ensemble (———); total (− · · ·); smoothing (− −); measurement error (· · · ·); interference error (− · − ·).

ensemble to be described by the MOPITT a priori. For this ensemble, the FTIR retrieval is not an optimal estimate because it uses a different a priori. Note that when carrying out actual comparisons, the real ensemble is likely to be unknown, so this assessment can only be a guide.

8.1.   Profile Retrievals

[60] For a direct intercomparison of MOPITT and FTIR profiles, we expect to find differences in accordance with the statistics calculated by the method developed in section 3. The standard deviations of the errors to be expected from each of the three terms in equation (13) are given in Figure 8, i.e., the square roots of the diagonals of the matrices. The standard deviations of the comparison ensemble and the total
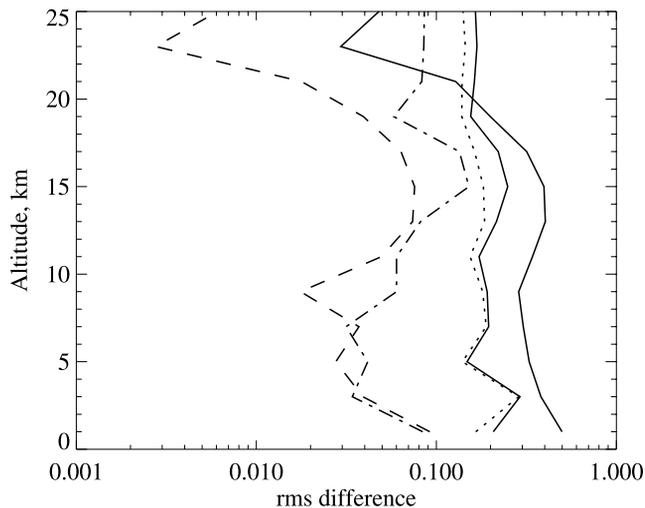
Figure 8. Expected standard deviations of the MOPITT and FTIR retrieval difference, and its components according to equation (13). Comparison ensemble (———); MOPITT retrieval noise (− −); FTIR retrieval noise (· · · ·); smoothing error due to difference in averaging kernels (− · − ·); total comparison error (− · · ·).
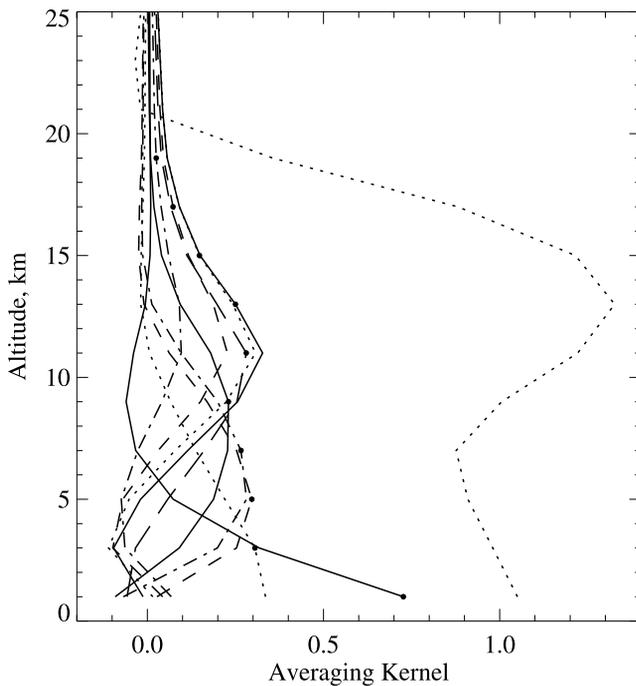
Figure 9. A simulation by FTIR of the MOPITT CO Averaging kernels for 1–19 km. The nominal height of each kernel is marked by a filled circle.

difference covariance $\mathbf{S}_\delta$ are also shown for reference. We see that the difference is dominated by the FTIR noise term $(\cdots)$, with MOPITT noise $(--)$ being somewhat smaller, and the smoothing error term $(-\cdot-\cdot)$ also being smaller except around 15–17 km. However, the total $(-\cdots)$ is large enough compared with $\mathbf{S}_c$ that single profile comparisons of these FTIR retrievals with MOPITT may not be very useful. Note that this conclusion is mainly a consequence of using a unit matrix as a priori covariance for FTIR, but illustrates the effect of a retrieval which is not optimal for the comparison ensemble.

[61] An alternative approach is to use the FTIR to simulate a MOPITT retrieval, or vice versa, according to the method described in section 4.4. As the FTIR retrieval is not optimal for the comparison ensemble, it is necessary to use the method of section 4.2. The resulting averaging kernels are shown in Figure 9. Qualitatively this appears to be a good simulation in comparison with Figure 6.

[62] The expected size of the differences between MOPITT retrievals and the FTIR simulations are shown in Figure 10. The square roots of the diagonals of the three terms in equation (30) are plotted, to be compared with the corresponding terms in Figure 8. We see that the FTIR retrieval simulation noise $(\cdots)$ is significantly smaller than the original FTIR retrieval noise (Figure 4), largely because the fine structure has been eliminated by the MOPITT averaging kernels, and the smoothing error $(-\cdot-\cdot)$ is much smaller, because the averaging kernels are much more alike. Consequently the total error in the comparison is now usefully smaller than the comparison ensemble standard deviation, so this kind of comparison is worthwhile.

[63] Further improvement is possible by tuning the FTIR a priori constraint. Reducing it to 0.1 times a unit matrix

(not shown) reduces the total error estimate still further in this case. However, simulating FTIR retrieval using MOPITT is not productive. We find that the total error in this case is little different from the direct comparison, mainly because the dominant FTIR retrieval noise is not reduced by this means, and the MOPITT averaging kernels are unable to reproduce the FTIR averaging kernels as well as the other way round.

8.2. Total Column Retrievals

[64] Comparison of the MOPITT and FTIR total columns will probably be the most direct and important use of FTIR measurements for MOPITT validation. The total column operator and the instrument total column averaging kernels are shown in Figure 11. These functions are to be multiplied by the scaled mixing ratio profile and integrated to derive the total column. The total column operator gives the true value, the instrument averaging kernels give approximations to it. The standard deviations of the error components for the total column estimates and the comparisons are given in Table 1.

[65] The FTIR averaging kernel is so close an approximation to the true total column operator that it cannot be distinguished on this plot. The MOPITT kernel is less so, having limited sensitivity near the surface and oscillating about the true operator through the troposphere. However, both measurements give good estimates of the total column: the MOPITT smoothing error is 2.9%, and the FTIR 0.06%, to be compared with the ensemble variability of 32%.

[66] Also shown in Figure 11 is a combination of FTIR averaging kernels derived using the procedure of section 4.2 to give the best possible reproduction of the MOPITT kernel. The coefficients of this combination can be applied to a profile retrieved by FTIR to give an estimate of the total column as retrieved by MOPITT, as a way of reducing the smoothing error in the comparison. For the case at hand the
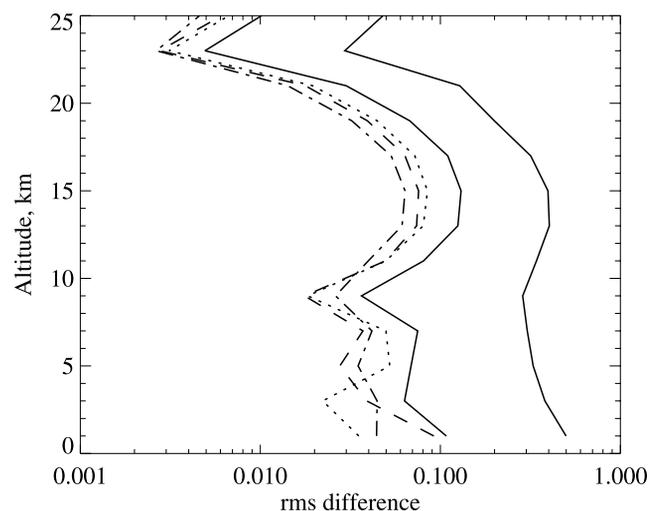
Figure 10. Expected standard deviations of the difference between the MOPITT and a simulation using FTIR, and its components according to equation (13). Comparison ensemble (———); MOPITT retrieval noise $(--)$; FTIR retrieval simulation noise $(\cdots)$; smoothing error due to difference in averaging kernels $(-\cdot-\cdot)$; total comparison error $(-\cdots)$.
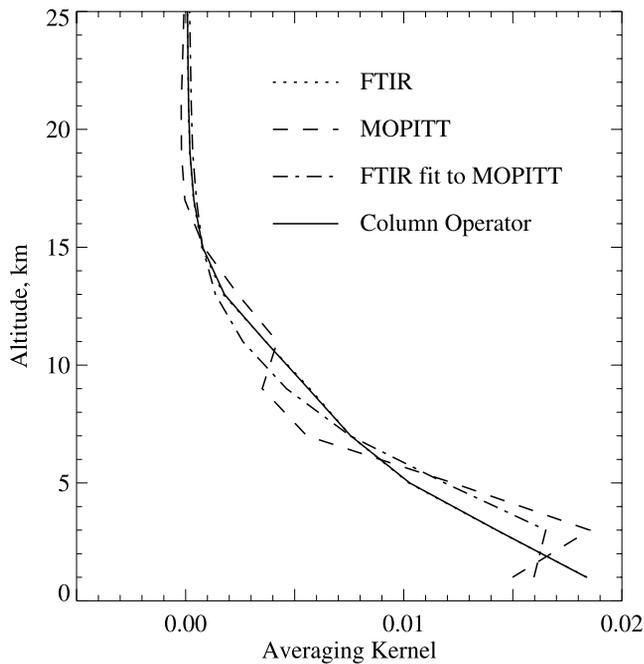
Figure 11. Comparison of total column operators. Note that the FTIR operator is coincident with the total column operator on this scale.
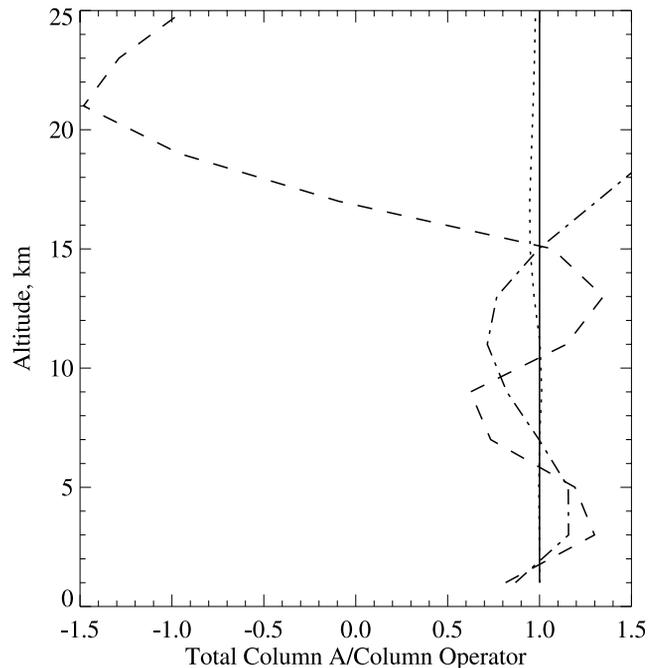
Figure 12. The ratio of the operators in Figure 11 to the total column operator. MOPITT ($\cdots$); FTIR (——); FTIR approximation to MOPITT ($- \cdot - \cdot$).

smoothing error to be expected in a simple difference of the MOPITT and FTIR total columns is 1.7%. If the FTIR estimate of the MOPITT kernel is used, this is reduced to 1.1%.

[67] Figure 12 shows the ratio of the averaging kernels in Figure 11 to the total column operator (thus an "ideal" averaging kernel would have a value of one at all altitudes). The FTIR kernel has values near unity at all altitudes. The MOPITT kernel is oscillatory in the troposphere but is reasonably close to unity up to about 15 km. At higher altitudes it has too small values, but in practice the amount of CO at those altitudes is insignificant, so the shape here is of little concern. The dash-dotted line in Figure 12 is the FTIR estimate of the MOPITT kernel, which shows that FTIR is able to reproduce the broad features of the MOPITT kernel, but not its oscillations, up to 15 km.

8.3. Measured Signals

[68] Applying the formulation of section 5.1 to the MOPITT and FTIR weighting functions leads to the conclusion that there are three linear combinations of measurements with singular values near unity, and which may hence show significant correlation, given the range of atmospheric

Table 1. Expected Error Components for Total Column Estimates and Comparisons[a]

	Noise, %	Smoothing Error, %	Total, %
MOPITT	2.4	1.7	2.9
FTIR	0.33	0.06	0.34
MOPITT−FTIR	2.4	1.7	2.9
MOPITT−simulation by FTIR	2.4	1.1	2.6

[a]Percentages of the total CO, to be compared with the ensemble variability of 32%.

variability embodied in the MOPITT a priori covariance. These are shown in Figure 13, and the r.m.s differences between the corresponding combinations of measured signals are given in Table 2, together with a breakdown into the components corresponding to instrument noise and smooth-
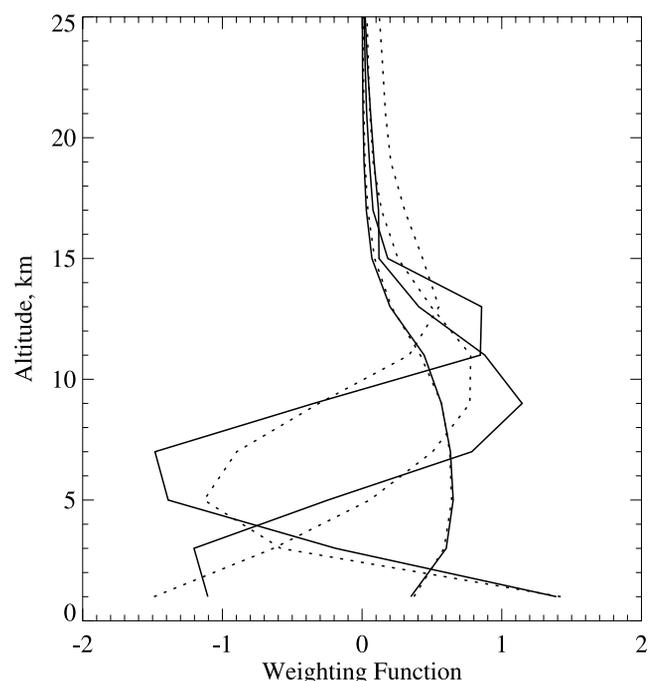
Figure 13. Combinations of the MOPITT and FTIR weighting functions which approximately match. The solid lines are for MOPITT and the dotted ones for FTIR.

Table 2. Expected r.m.s. Differences Between the Three Best Matching Combinations of Measured Signals, and its Components, as a Fraction of the Total Variance of Signals of One Instrument for the Comparison Ensemble

	First Combination	Second Combination	Third Combination
Singular value	0.9985	0.9774	0.8145
Total r.m.s.	0.0556	0.213	0.609
Smoothing error	0.0148	0.143	0.169
Total noise	0.0536	0.158	0.585
FTIR noise	0.0387	0.088	0.479
MOPITT noise	0.0371	0.131	0.337

ing error due to the mismatch between the weighting function combinations. These are all scaled so that the total variance of the combinations of the measured signals is unity for the comparison ensemble. Thus if the signals were uncorrelated the r.m.s. difference would be $2^{\frac{1}{2}}$.

[69] In the case of the first weighting function combinations a good match is obtained, but at the expense of relatively large noise, equally from both instruments. This indicates that it is quite difficult to find a good match with the basic weighting functions being so different, and it is necessary to take rather extreme combinations, so that noise is amplified. Nevertheless this is the best option - any other combination with less noise would have a larger smoothing error.

[70] With regard to the other combinations, it is clear by inspection that the match between the functions is much poorer, and indeed the variance of the difference between the combined signals much larger. Thus only the first combination of weighting functions will provide a strong test of the consistency of the instruments. With an r.m.s. error of 5.5% of the variability, it is a better test than the profile or total column comparisons.

9. Summary and Conclusions

[71] In this paper we have provided a description of methods which are needed to properly intercompare measurements and retrievals from remote sounders with different weighting functions, averaging kernels and errors. We have shown how to compare measured signals, retrieved profiles, and derived quantities such as total columns of constituents making due allowance for the differing characteristics of the observing systems. The differences in the characteristic imply that the comparison is not straightforward; the following considerations are relevant:

1. Observing systems to be compared should retrieve the same quantity, both in its representation and the grid on which it is obtained. For example it is not easy to make a proper statistical comparison of concentrations on a height grid with mixing ratios on a pressure grid. However, we hope to relax this requirement in a future publication.

2. The comparison must be general, and be able to deal with any kind of retrieval method, not only optimal estimators.

3. Profile comparisons must allow for the difference in averaging kernels as well as retrieval noise.

4. Comparison of derived quantities such as total column must also allow for the difference of averaging kernels.

5. We find that the effect of different averaging kernels can be reduced if the retrieval or derived quantity of one instrument is simulated from the retrieval of the other. This also has the effect of reducing the noise component of the second instrument.

6. We have also shown how combinations of measured signals can be found which can be compared directly.

[72] We illustrated the techniques by reference to the expected differences between a satellite instrument MOPITT and a ground based instrument FTIR, both measuring CO. FTIR instruments are being used in the validation campaign for MOPITT, so proper statistical intercomparison is required. In this case we find:

1. Direct comparison of retrieved profiles is not satisfactory, because the expected standard deviation of the difference is a large fraction of the ensemble standard deviation, mostly due to noise.

2. Comparison of the MOPITT profile retrieval with a simulation using FTIR is much more useful, though still not ideal. The expected standard deviation is of order 20% of the ensemble standard deviation in the troposphere.

3. Direct comparison of total columns gives an expected standard deviation of about 9% of the ensemble standard deviation. Comparison of MOPITT with a simulation derived from FTIR gives an expected standard deviation of 8%.

4. There is only one combination of measured signals that can be usefully compared. The difference is expected to have a standard deviation of about 5.5% of the ensemble variability, mostly due to noise. This quantity has the smallest expected error of all those whose comparison we have considered.

[73] At the time of writing, flight data from MOPITT is not yet at a stage of development where these techniques can be usefully applied. We hope to hope to make formal comparisons of real data in a future publication.

Appendix A: Treatment of χ^2 for a Singular Covariance

[74] This can be done by using a pseudo-inverse based on an eigenvector expansion of $\mathbf{S}_\delta$, ignoring vectors with zero eigenvalue, which correspond to components of $\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2$ which are exactly zero. If we expand:

$$\mathbf{S}_\delta = \tilde{\mathbf{L}}^T \tilde{\mathbf{\Lambda}} \tilde{\mathbf{L}} \qquad (A1)$$

where $\tilde{\mathbf{\Lambda}}$ is a $p \times p$ diagonal matrix containing only the p nonzero eigenvalues, and $\tilde{\mathbf{L}}$ is the corresponding $p \times n$ matrix of eigenvectors, then

$$\chi^2 = (\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2)^T \tilde{\mathbf{L}}^T \tilde{\mathbf{\Lambda}}^{-1} \tilde{\mathbf{L}}(\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2) = \mathbf{w}^T \tilde{\mathbf{\Lambda}}^{-1} \mathbf{w} \qquad (A2)$$

is the appropriate χ^2 with p degrees of freedom. Each element w_j of $\mathbf{w} = \tilde{\mathbf{L}}(\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2)$ should be independently distributed with variance $\tilde{\lambda}_j$, the corresponding diagonal element of $\tilde{\mathbf{\Lambda}}^{-1}$. Even if $\mathbf{S}_\delta$ is numerically nonsingular, its eigenvalues should be checked in case the nonsingularity is due to rounding errors. It is also worth noting that components with small eigenvalues may be subject to numerical error, and may not be useful.

Appendix B: Comparison of Measured Signals: Large Numbers of Channels

[75] By scaling the measurement vector by $\mathbf{S}_\epsilon^{-1/2}$ and the state vector by $\mathbf{S}_c^{-1/2}$ we can write

$$\tilde{\mathbf{y}} = \mathbf{S}_\epsilon^{-\frac{1}{2}}(\mathbf{y} - \mathbf{y}_c)$$
$$\tilde{\epsilon} = \mathbf{S}_\epsilon^{-\frac{1}{2}}\epsilon \qquad \tilde{\mathbf{x}} = \mathbf{S}_c^{-\frac{1}{2}}(\mathbf{x} - \mathbf{x}_c) \tag{B1}$$

thus reducing $\mathbf{x}$ and $\mathbf{y}$ to nondimensional forms with $\tilde{\epsilon}$ and both having unit covariances. The corresponding linearized forward model for instrument i is

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{K}}_i \tilde{\mathbf{x}} + \tilde{\epsilon}_i \tag{B2}$$

where $\tilde{\mathbf{K}}_i = \mathbf{S}_{\epsilon_i}^{-1/2}\mathbf{K}_i\mathbf{S}_c^{1/2}$. Replacing $\tilde{\mathbf{K}}_i$ by its singular vector decomposition $\mathbf{U}_i\mathbf{\Lambda}_i\mathbf{V}_i^T$, and premultiplying both sides by $\mathbf{U}^T$ gives

$$\mathbf{y}_i' = \mathbf{U}_i^T\tilde{\mathbf{y}}_i = \mathbf{\Lambda}_i\mathbf{V}_i^T\tilde{\mathbf{x}} + \mathbf{U}_i^T\tilde{\epsilon}_i = \mathbf{K}_i'(\mathbf{x} - \mathbf{x}_c) + \epsilon_i', \tag{B3}$$

thus defining $\mathbf{y}_i'$ and ϵ_i', and where $\mathbf{K}_I' = \mathbf{\Lambda}_i\mathbf{V}_i^T\mathbf{S}_c^{-\frac{1}{2}}$. Note that $\epsilon' = \mathbf{U}_i^T\tilde{\epsilon}_i$ has unit covariance, while $\mathbf{\Lambda}_i\mathbf{V}_i^T\tilde{\mathbf{x}}$ has covariance $\mathbf{\Lambda}_i^2$, so elements of $\mathbf{y}_i'$ corresponding to singular values much smaller than unity can be ignored without loss of information. Therefore we consider only the p_i singular vectors for which the singular value is not insignificant. Thus p_i is the effective rank of the weighting function matrix. Note that whereas $\mathbf{y}_i$ and $\tilde{\mathbf{y}}_i$ are vectors of length m_i, very large in the case of FTIR, $\mathbf{y}_i'$ is of length p_i, at most equal to the smaller of m_i and n.

[76] Following the analysis of section 5.1, but with this transformed version of the forward model, the vectors $\mathbf{m}_1$ and $\mathbf{m}_2$ are obtained as singular vectors of

$$\left(\mathbf{\Lambda}_1^2 + \mathbf{I}\right)^{-\frac{1}{2}}\mathbf{\Lambda}_1\mathbf{V}_1^T\mathbf{V}_2\mathbf{\Lambda}_2\left(\mathbf{\Lambda}_2^2 + \mathbf{I}\right)^{-\frac{1}{2}} \tag{B4}$$

and the combined weighting functions are

$$\check{\mathbf{K}} = \mathbf{m}_i^T\left(\mathbf{\Lambda}_i^2 + \mathbf{I}\right)^{-\frac{1}{2}}\mathbf{U}_i^T\mathbf{S}_{\epsilon_i}^{-1/2}\mathbf{K}_i \tag{B5}$$

References

Bretherton, C. S., C. Smith, and J. M. Wallace, An intercomparison of methods for finding coupled patterns in climate data, *J. Clim.*, *5*, 541, 1992.

Connor, B. J., D. E. Siskind, J. J. Tsou, A. Parrish, and E. E. Remsberg, Ground-based microwave observations of ozone in the upper stratosphere and mesosphere, *J. Geophys. Res.*, *99*, 16,757–16,770, 1994.

Connor, B. J., A. Parrish, J.-J. Tsou, and M. P. McCormick, Error analysis for the ground-based microwave ozone measurements during STOIC, *J. Geophys. Res.*, *100*, 9283–9292, 1995.

Drummond, J. R., Measurements of Pollution in the Troposphere (MO-PITT), in *The Use of EOS for Studies of Atmospheric Physics*, edited by J. C. Gille and G. Visconti, pp. 77–101, North-Holland, New York, 1992.

Drummond, J. R., and G. S. Mand, The Measurements of Pollution in the Troposphere (MOPITT) instrument: Overall performance and calibration requirements, *J. Atmos. Oceanic Technol.*, *13*, 314–320, 1996.

Kurylo, M. J., and R. J. Zander, The NDSC—Its status after ten years of operation, in *Proceedings of the Quadrennial Ozone Symposium, Sapporo 2000*, pp. 167–168, Hokkaido Univ., Sapporo, Japan, 2001.

Pan, L., D. P. Edwards, J. C. Gille, M. W. Smith, and J. R. Drummond, Satellite remote sensing of tropospheric CO and CH_4: Forward model studies of the MOPITT instrument, *Appl. Opt.*, *34*, 6976, 1995.

Pan, L., J. C. Gille, D. P. Edwards, P. L. Bailey, and C. D. Rodgers, Retrieval of tropospheric carbon monoxide for the MOPITT experiment, *J. Geophys. Res.*, *103*, 32,277–32,290, 1998.

Rodgers, C. D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Rev. Geophys. Space Phys.*, *14*, 609–624, 1976.

Rodgers, C. D., Characterization and error analysis of profiles retrieved from remote sounding measurements, *J. Geophys. Res.*, *95*, 5587–5595, 1990.

Rodgers, C. D., *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Sci., River Edge, N. J., 2000.

Russell, J. M., and H. G. J. Smit, Data quality, in SPARC/IOC/GAW assessment of trends in the vertical distribution of ozone, SPARC report 1, *World Meteorol Organ. Ozone Res. and Monit. Proj. Rep. 43*, edited by N. Harris, R. Hudson, and C. Phillips, chap. 2, Geneva, 1998.

———————————

B. J. Connor, National Institute of Water and Atmospheric Research, PB 50061, Omakau, Lauder, Otago 9182, New Zealand. (b.connor@niwa.cri.nz)

C. D. Rodgers, Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK. (c.rodgers@physics.ox.ac.uk)