JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 95, NO. D5, PAGES 5587–5595, APRIL 20, 1990

Characterization and Error Analysis of Profiles Retrieved From Remote Sounding Measurements

CLIVE D. RODGERS[1]

National Center for Atmospheric Research, Boulder, Colorado

The characterization and error analysis of profiles retrieved from remote measurements present conceptual problems, particularly concerning interlevel correlations between errors, the smoothing effect of remote sounding and the contribution of a priori information to profile. A formal analysis for profile retrieval is developed which is independent of the nature of the retrieval method, provided that the measurement process can be characterized adequately. The relationship between the retrieved and true profiles is expressed in terms of a smoothing function which can be straightforwardly calculated. The retrieval error separates naturally into three components, (1) random error due to measurement noise, (2) systematic error due to uncertain model parameters and inverse model bias, and (3) null-space error due to the inherent finite vertical resolution of the observing system. A recipe is given for evaluating each of the components in any particular case. Most of the error terms appear as covariance matrices, rather than simple error variances. These matrices can be interpreted in terms of "error patterns", which are statistically independent contributions to the total error. They are the multidimensional equivalent of "error bars". An approach is described which clarifies the relation of a priori data to the retrieved profile, and identifies a priori in cases where it is not an explicit part of the retrieval.

1. INTRODUCTION

A well understood approach to characterization and error analysis of profiles retrieved from remote measurements of the atmosphere is important for design of sounding systems and in guiding their calibration. However, these are topics that have presented conceptual problems for many workers in the subject, particularly in respect to the relationship between retrieved data and the real atmosphere, in presenting correlations between the errors at different levels, and in understanding the contribution of a priori information to the profile. As a result it is often difficult to make objective comparisons between different sounding systems, based on the published literature.

The purpose of this paper is to review the characterization and error analysis, put it on a reasonably straightforward but rigorous basis, and give an intuitive description of the nature of the errors. The analysis is carried out in such a way that it can be applied to any kind of approach to the retrieval problem, not only statistically optimum methods. Retrieval methods as such are not discussed; for a review of inverse theory in general, see, for example, *Menke* [1984]. For a review with an atmospheric bias, see *Rodgers* [1976].

The retrieval is characterized by explicitly evaluating the functional form of the relationship between the retrieved profile, the true atmosphere, and the various error sources. For small departures from a standard profile, the retrieval can be viewed as a smoothed version of the true profile, with a smoothing function that can be evaluated straightforwardly, leading to a natural interpretation of the concept of "resolution".

The total error separates into components due to (1) systematic errors in the forward model and the inverse model, (2) measurement noise in the instrument, (3) components in the profile that the observing system cannot see. A method is described whereby the effect of errors that are correlated between different levels can be readily appreciated.

The context of the discussion is in atmospheric remote sounding, but the concepts are sufficiently general that they may be applied to any type of profile retrieval from remote measurements.

2. DEFINITIONS

The "observing system" is the combination of the measuring instrument and the method of profile retrieval. The input to the observing system is an actual atmospheric profile, and its output is a retrieved profile.

The "state vector" $\mathbf{x}$ is a vector of unknowns to be estimated from the measurements, describing the state of the atmosphere. Usually, it will be a profile of some quantity, given at a finite number n of levels, where n is large enough to represent the possible atmospheric variations adequately. However, it may in principle comprise any set of relevant variables, such as coefficients for a representation of the profile; see section 8 below. Unless otherwise stated, the discussion below will assume that $\mathbf{x}$ is a profile.

The "measurement" $\mathbf{y}$ is a vector of m measured quantities. Often $m \ll n$, and the inverse problem is formally ill-posed. Measurements are made to a finite accuracy, with measurement error ϵ_y assumed to be normally distributed with mean zero and known error covariance $\mathbf{S}_\epsilon$.

The "forward model" F characterizes the measurement $\mathbf{y}$, describing how it depends on the state vector $\mathbf{x}$. It may be an algebraic or algorithmic description.

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

Paper number 89JD03063.
0148–0227/90/89JD-03063$05.00

$$\mathbf{y} = F(\mathbf{x}, \mathbf{b}) + \epsilon_{\mathbf{y}} \quad (1)$$

where $\mathbf{b}$ is a vector of n_b model parameters (such as spectral line data, calibration parameters, etc.) that are not perfectly known to the observer. They are a possible source of random or systematic differences between calculated and measured values of $\mathbf{y}$. The forward model should contain all of the physics of the measurement, to an accuracy far better than the measurement can be made to. If only an approximate model is available, then its accuracy should be quantifiable; if it is not, then the measurement is of no value. The error analysis requires an accurate forward model, which is not necessarily the same as any forward model that might be used as part of the inverse model. If the latter is approximate, its errors will be automatically allowed for by the method described here.

Thus we assume that the forward model represents the physics of the measurement accurately; forward model errors must be characterized only in terms of model parameters. Note that in some works on inverse theory, the term "model parameter" is used to denote the quantity called state vector here. In fact, the distinction between them is vague; the measurement depends on the model parameters, so they could be regarded as quantities to be retrieved. We categorize as model parameters those quantities which might be part of the state vector but are not being retrieved from the measurements.

For the purpose of error analysis we may linearize the forward model about some reference state $(\bar{\mathbf{x}}, \bar{\mathbf{b}})$ which may be the true state $\mathbf{x}$, the retrieved state $\hat{\mathbf{x}}$, an ensemble mean, or any arbitrary state, as required.

$$\mathbf{y} = F(\bar{\mathbf{x}}, \bar{\mathbf{b}}) + \frac{\partial F}{\partial \mathbf{x}}(\mathbf{x} - \bar{\mathbf{x}}) + \frac{\partial F}{\partial \mathbf{b}}(\mathbf{b} - \bar{\mathbf{b}}) + \epsilon_{\mathbf{y}} \quad (2)$$

We define the weighting function $\mathbf{K}_x$ as $\partial F/\partial \mathbf{x}$, so called because in some cases it is normalized to unit area, in which case $\mathbf{y} - \bar{\mathbf{y}}$ is a weighted mean of $\mathbf{x} - \bar{\mathbf{x}}$ with weights given by the rows of $\mathbf{K}_x$. Weighting functions give a broad idea of the information content of a set of measurements, in that they show the part of the profile that is represented by each measurement. However, even if they have well-defined peaks their width does not necessarily give a good indication of the vertical resolution, as this depends on both the width of the individual peaks and their spacing.

We also define $\mathbf{K}_b$ to be the sensitivity $\partial F/\partial \mathbf{b}$ of the measurement to the forward model parameters. If the derivatives in the linearization cannot be expressed algebraically, they may be evaluated numerically by perturbing the forward model.

The "inverse model" I describes how the retrieved state $\hat{\mathbf{x}}$ is obtained from the measurement:

$$\hat{\mathbf{x}} = I(\mathbf{y}, \mathbf{b}, \mathbf{c}) \quad (3)$$

where $\mathbf{c}$ is a vector of parameters used in the retrieval that do not appear in the forward model, i.e., are unconnected with the measurements, but may reasonably be varied, for example, a priori data. The inverse model may be an explicit algebraic function or may be defined algorithmically.

For the purpose of error analysis we also linearize the inverse model:

$$\hat{\mathbf{x}} = I(\bar{\mathbf{y}}, \bar{\mathbf{b}}, \bar{\mathbf{c}}) + \frac{\partial I}{\partial \mathbf{y}}(\mathbf{y} - \bar{\mathbf{y}}) + \frac{\partial I}{\partial \mathbf{b}}(\mathbf{b} - \bar{\mathbf{b}}) + \frac{\partial I}{\partial \mathbf{c}}(\mathbf{c} - \bar{\mathbf{c}}) \quad (4)$$

and define the contribution function $\mathbf{D}_y$ as $\partial I/\partial \mathbf{y}$, so called because each column of $\mathbf{D}$ is the contribution to the solution due to a unit change in the corresponding element of $\mathbf{y}$. We also define $\mathbf{D}_b$ and $\mathbf{D}_c$ to be the sensitivities $\partial I/\partial \mathbf{b}$ and $\partial I/\partial \mathbf{c}$ of the inverse model to the model parameters.

In the error analysis we must take care when assigning errors to the model parameters. For example, $\mathbf{b}$ takes its exact value as far as the observing system is concerned but takes an estimated value $\hat{\mathbf{b}}$ when it appears in the inverse model, or when the forward model is being evaluated by the observer. The identification of model parameters may not be straightforward. For example, calibration coefficients are forward model parameters, and the observing system knows their true values but the observer must use estimated values. This fact can easily be overlooked if the forward model is formulated to give scientific units, and the measurement $\mathbf{y}$ is precalibrated and also presented in scientific units.

3. Requirements for a Retrieval

The error analysis presented below does not require the inverse model to produce even a qualitatively "good" retrieval and can be applied to any arbitrary function of the measurements. However, it was felt worthwhile to discuss the criteria that ought to be met for $\hat{\mathbf{x}}$ to be an acceptable retrieval. In this section we discuss the the basic requirement that the measurement should be reproduced to some appropriate accuracy when $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$ are inserted in the forward model. We would also expect the profile to be consistent with whatever is known a priori about its physics and climatology (although a priori information can be wrong, of course).

What accuracy is appropriate? If we knew the true profile $\mathbf{x}$, then we could calculate a measurement $F(\mathbf{x}, \hat{\mathbf{b}})$ which would differ from the actual measurement $F(\mathbf{x}, \mathbf{b}) + \epsilon_y$ by $\epsilon_y - \mathbf{K}_b(\hat{\mathbf{b}} - \mathbf{b})$. This difference should have an expected value of zero, and a covariance $\mathbf{S}_\delta = \mathbf{S}_\epsilon + \mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T$, where $\mathbf{S}_b$ is the error covariance of $\hat{\mathbf{b}}$. It would be reasonable for the retrieved profile $\hat{\mathbf{x}}$ to give a corresponding $\hat{\mathbf{y}} = F(\hat{\mathbf{x}}, \hat{\mathbf{b}})$ which differs from the actual measurement by an amount consistent with $\mathbf{S}_\delta$. A suitable measurement of consistency is a χ^2 test. If we define $\delta\mathbf{y}$ as $\mathbf{y} - \hat{\mathbf{y}}$, then

$$\chi^2 = \delta\mathbf{y}^T\mathbf{S}_\delta^{-1}\delta\mathbf{y} \quad (5)$$

which must satisfy, for example, the 99% confidence limits for $m + n_b$ degrees of freedom. (It should be pointed out that this is not intended to be used as a convergence criterion in an iterative retrieval method.)

We can go further than this, by examining the distributions of $\delta\mathbf{y}$ and χ^2 over an ensemble of measurements and corresponding retrievals. It might be thought that χ^2 would form a χ^2 distribution, with $m + n_b$ degrees of freedom, and that $\delta\mathbf{y}$ would have a covariance $\mathbf{S}_\delta$. However this will only be the case if the ensemble allows both ϵ_y and $\epsilon_b = \hat{\mathbf{b}} - \mathbf{b}$ to vary randomly and independently. In practice, many components of ϵ_b will be constants, for example, spectral line data errors, and will not contribute

variability to the solution. In the extreme case when all the elements of ϵ_b are constants, the ensemble mean of $\delta\mathbf{y}$ may be nonzero, and the covariance about this mean should be approximately $\mathbf{S}_\epsilon$ rather than $\mathbf{S}_\delta$.

When due allowance is made for the actual variability of ϵ_b in the ensemble, the χ^2 can be used as a test of the inverse method. It should have a χ^2 distribution, with a mean value of an appropriate number of degrees of freedom. If $\overline{\chi^2}$ is too large, the inverse method is not doing a good enough job of fitting retrievals to the measurements, whilst if it is too small, then it is doing too good a job, and measurement error is being misinterpreted as real information.

In practice, it is difficult to avoid misinterpreting measurement error as real information. For example, an error in an absorption coefficient will lead to a compensating error in a retrieved absorber amount, and noise cannot be completely separated from signal, even by an optimal estimator. Therefore a proper error analysis should always be carried out, so that these effects are understood.

4. Characterization of the Retrieval

If we ignore for the moment the errors in measurements and model parameters, we can relate the retrieval to the unknown profile by a transfer function T:

$$\hat{\mathbf{x}} = I(F(\mathbf{x}, \mathbf{b}), \mathbf{b}, \mathbf{c}) = T(\mathbf{x}, \mathbf{b}, \mathbf{c}) \tag{6}$$

The way to understand the nature of the retrieved state vector is to study this transfer function. If we linearize with respect to $\mathbf{x}$ about a reference state $\bar{\mathbf{x}}$, and use the definitions of the $\mathbf{D}$ and $\mathbf{K}$ matrices, we obtain

$$\hat{\mathbf{x}} = T(\bar{\mathbf{x}}, \mathbf{b}, \mathbf{c}) + \mathbf{D}_y\mathbf{K}_x(\mathbf{x} - \bar{\mathbf{x}}) \tag{7}$$

Thus components of the solution corresponding to departures of the profile from $\bar{\mathbf{x}}$ are smoothed by the operator

$$\mathbf{A} = \frac{\partial T}{\partial \mathbf{x}} = \mathbf{D}_y\mathbf{K}_x \tag{8}$$

Following *Backus and Gilbert* [1970], we call rows of this square matrix the "averaging kernel". The retrieval at any location is an average of the whole profile weighted by this row. Columns may be thought of as the response of the observing system to a δ-function disturbance. A δ-function disturbance in the real profile will be reflected in the retrieval as the shape of a column of $\mathbf{A}$. For many retrieval methods, numerical simulation of the δ-function response may be the best way of evaluating $\mathbf{A}$.

For an "ideal" observing system, $\mathbf{A}$ will be a unit matrix, but this is often not the case. Normally, the rows and columns will represent peaked functions, with the width of the peak of either being a qualitative measure of the resolution of the observing system. *Menke* [1984] calls $\mathbf{A}$ the "model resolution matrix" for this reason. The shape of the averaging kernel is an important characteristic of an observing system, as it not only indicates resolution but can also be used to detect a misleading retrieval if the shape is other than a simple peak. See section 9 for an example of this.

Resolution has always been a vague concept with many possible definitions, the most appropriate one to use depending on the context. We could use the Rayleigh criterion, and define it as the separation at which two δ functions in the profile can just be distinguished, but δ functions are not typical of atmospheric phenomena, so this may not be helpful. We could define it as the wavelength of a sinusoidal disturbance that can just be detected, but resolution can vary with height, whilst the wavelength of a sinusoid does not. Further, the detection limit is hard to define.

Resolution could be defined as the width, in some sense, of the averaging kernel or the δ-function response. There are many ways of defining width, for example, as the full width at half height, or in terms of the second moment about the mean. In formalizing this, we note that resolution is only a valid concept for continuous functions, so $A(z, z')$ must be translated from the matrix form $\mathbf{A}$, using the appropriate profile representation. We first define $a(z) = \int A(z, z')dz'$, the area of the the averaging kernel for height z. Then the second moment width $w(z)$ is

$$w(z) = \left(\int_{-\infty}^{\infty} A(z, z')(z' - \bar{z}(z))^2 dz'/a(z)\right)^{\frac{1}{2}} \tag{9}$$

where

$$\bar{z}(z) = \int_{-\infty}^{\infty} z' A(z, z')dz'/a(z) \tag{10}$$

However, these definitions do not give satisfactory results with functions with negative excursions or broad and oscillatory wings, because it is possible for the integrals to be negative or zero. *Backus and Gilbert* [1970] used the following definition of the width of the averaging kernel as their definition of "spread", and developed a retrieval method that trades off spread against measurement noise:

$$s(z) = \frac{12}{a(z)^2}\int (z - z')^2 A(z, z')^2 dz' \tag{11}$$

The factor 12 ensures that a boxcar function has a spread equal to its width. The Backus-Gilbert spread has proved to be a satisfactory definition of resolution, notably for their inverse method which it is minimized.

5. Error Analysis

For the error analysis we must include measurement error ϵ_y and note that the inverse model uses our best estimates $\hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$, whilst the measuring system uses exact values of the model parameter $\mathbf{b}$, and does not use $\mathbf{c}$ at all:

$$\hat{\mathbf{x}} = I(F(\mathbf{x}, \mathbf{b}) + \epsilon_y, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \tag{12}$$

If we linearize about $(\bar{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$ and use the definitions of the $\mathbf{D}$ and $\mathbf{K}$ matrices, we obtain

$$\hat{\mathbf{x}} = T(\bar{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \mathbf{D}_y\mathbf{K}_x(\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{D}_y\mathbf{K}_b(\mathbf{b} - \hat{\mathbf{b}}) + \mathbf{D}_y\epsilon_y \tag{13}$$

Linearization about the true state $(\mathbf{x}, \mathbf{b}, \mathbf{c})$ might be more appropriate, but as it is not known, it would not be possible to evaluate anything there. We can rearrange equation (13) to give the total error in a retrieval:

$$\hat{\mathbf{x}} - \mathbf{x} = [T(\bar{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) - \bar{\mathbf{x}}] + (\mathbf{A} - \mathbf{I})(\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{D}_y \mathbf{K}_b \epsilon_b + \mathbf{D}_y \epsilon_y \quad (14)$$

where $\mathbf{I}$ is a unit matrix. We have written ϵ_b for $\mathbf{b} - \hat{\mathbf{b}}$, representing the uncertainty in our knowledge of these quantities. Uncertainty in our knowledge of $\mathbf{c}$ only appears in the term $T(\bar{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$ at this stage. This will be discussed later. Equation (14) gives some insight into the nature of the transfer function. Each of the terms makes its own contribution to the overall error budget of the retrieval as follows:

Null-Space Error

There are two ways of viewing the smoothing implied by the averaging kernel, we either accept that the retrieved profile is this particular function of the true profile, or we examine the statistics of the difference between the two. This difference we call the null-space error, as it corresponds to those portions of profile space that cannot be measured by the observing system. For any one profile, the null-space error is $(\mathbf{A} - \mathbf{I})(\mathbf{x} - \bar{\mathbf{x}})$. The statistics of this error can be estimated only if we have independent estimates of the statistics of $\mathbf{x} - \bar{\mathbf{x}}$, and in particular the covariance $\mathbf{S}_x$ of its variability. The null-space error covariance is then

$$\mathbf{S}_N = (\mathbf{A} - \mathbf{I})\mathbf{S}_x(\mathbf{A} - \mathbf{I})^T \quad (15)$$

In some cases, where there are many independent high-resolution measurements, it is possible to estimate $\mathbf{S}_x$, but in other cases it may be difficult to obtain a good value. Often a realistic estimate of the diagonal elements S_{zz} (the variances at each level) will be available, and if the off-diagonal elements (the interlevel covariances) are unknown, a conservative assumption is to set them to zero. (This gives the estimate of the ensemble probability density function which has maximum entropy, or minimum information.) If some idea of the length scale l of correlations is available, then an expression such as

$$S_{zz'} = (S_{zz} S_{z'z'})^{\frac{1}{2}} \exp(-(z - z')^2/l^2) \quad (16)$$

can be used. If an artificial covariance matrix is to be constructed in this way, it must be positive definite.

Measurement Error

As seen in equation (14), the contribution of measurement error ϵ_y directly to error in the retrieval is $\mathbf{D}_y \epsilon_y$. This has error covariance

$$\mathbf{S}_M = \mathbf{D}_y \mathbf{S}_\epsilon \mathbf{D}_y^T \quad (17)$$

Often $\mathbf{S}_\epsilon$ is diagonal, if the elements of $\mathbf{y}$ are measured independently. In any case it should be possible to calculate it on the basis of the design and laboratory calibration of the instrument.

Model Parameter Errors

Errors ϵ_b in the forward model parameters may be random or systematic, or more generally they may vary on any time scale. Nevertheless, $\hat{\mathbf{b}}$ is a best estimate, with an error covariance $\mathbf{S}_b$ which should be known. This translates into a contribution to the retrieval error with covariance given by

$$\mathbf{S}_S = \mathbf{A}_b \mathbf{S}_b \mathbf{A}_b^T \quad (18)$$

Systematic error due to bias in the inverse model relative to the forward model is given by $T(\bar{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) - \bar{\mathbf{x}}$. This will clearly not necessarily be zero for any particular $\bar{\mathbf{x}}$, but the inverse model can be redefined if appropriate by subtracting this quantity from the solution.

The contribution due to uncertainty in the inverse model parameters $\hat{\mathbf{c}}$ will be discussed in section 7.

The overall error in the retrieval thus comprises three elements, (1) null-space error, (2) measurement error and (3) model error. To compare different observing systems, we must compare all these and note especially that the total random error is the sum of the null-space and measurement errors, plus any random components in the model error.

6. Characterization of Correlated Errors

Error bars on scalar quantities are easy to understand, but when there are correlated errors in our knowledge of different quantities such as the elements of a vector $\mathbf{x}$, as expressed by an error covariance matrix $\mathbf{S}$, it becomes more difficult to understand the implications. The diagonal elements of $\mathbf{S}$ are the error variances of the elements of $\mathbf{x}$, but if the off-diagonal elements show that the errors are highly correlated, then we have more information about $\mathbf{x}$ than is the case if the off-diagonal elements are small. One way of conceptualizing the error covariance matrix is to diagonalize it, i.e. find its eigenvalues λ_i and eigenvectors $\mathbf{l}_i$, such that $\mathbf{S}\mathbf{l}_i = \lambda_i \mathbf{l}_i$. In this case $\mathbf{S}$ is symmetric, and the matrix of eigenvectors $\mathbf{L}$ can be normalized so that $\mathbf{L}^{-1} = \mathbf{L}^T$. Consequently, $\mathbf{S}$ can be decomposed as

$$\mathbf{S} = \sum_i \lambda_i \mathbf{l}_i \mathbf{l}_i^T = \sum_i \mathbf{e}_i \mathbf{e}_i^T \quad (19)$$

The orthogonal quantities $\mathbf{e}_i = \lambda_i^{\frac{1}{2}} \mathbf{l}_i$ can be thought of as "error patterns", the multidimensional equivalent of error bars. The error in $\mathbf{x}$ is a sum of these error patterns, each multiplied by a random factor a_i having unit variance

$$\epsilon_x = \sum_{i=1}^{n} a_i \mathbf{e}_i \quad (20)$$

When applied to the total error covariance $\mathbf{S}_S + \mathbf{S}_N + \mathbf{S}_M$ this analysis never seems to clarify the nature of the errors as much as one might wish, but if it is applied to the individual components separately, it can give a useful insight into the error characteristics. Some examples are given in section 9.

7. Effect of A Priori Data

Particular concern is often expressed about the contribution of a priori data to the retrieval, so a few words are in order here to illustrate how this may be understood. A priori is taken to mean many things, for example, the starting point of an ad hoc iteration, which may or may not affect the outcome, a constraint expressed by a representation, or an independent estimate of the state obtained from sources other than the direct measurement (a "virtual" measurement). For the purpose of the error analysis, we only consider the latter type. These may be

thought of as what the inverse model takes the profile to be in the absence of any real measurements. Thus if measurements corresponding to the a priori are used in the inverse model, the output should be the a priori profile, unchanged.

If we take the linearization point $\bar{\mathbf{x}}$ in equations (13) and (14) to be the a priori profile $\mathbf{x}^a$, then we obtain

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{x}^a + \mathbf{A}(\mathbf{x} - \mathbf{x}^a) + \mathbf{D}_y\mathbf{K}_b\epsilon_b + \mathbf{D}_y\epsilon_y \\ &= \mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A})\mathbf{x}^a + \mathbf{D}_y\mathbf{K}_b\epsilon_b + \mathbf{D}_y\epsilon_y \end{aligned} \quad (21)$$

because, according to our assumption about the nature of a priori, $T(\mathbf{x}^a, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = \mathbf{x}^a$. Thus we can express the retrieval as a linear combination of the true profile and the a priori, with matrix weights $\mathbf{A}$ and $\mathbf{I} - \mathbf{A}$, respectively, plus contributions from experimental error in $\mathbf{b}$ and $\mathbf{y}$. The sensitivity of the retrieval to changes in the a priori is given by

$$\mathbf{D}_a = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}^a} = \mathbf{I} - \mathbf{A} \quad (22)$$

Thus errors in a priori data are multiplied by the matrix factor $\mathbf{I} - \mathbf{A}$ in the retrieved profile.

Scalar weights are easy to understand, matrix weights are less so. One way to understand how matrix weights affect the solution is to use an eigenvector decomposition, as we did before. If the matrix of eigenvectors of $\mathbf{A}$ is $\mathbf{U}$, so that $\mathbf{AU} = \mathbf{U}\Lambda$ and $\mathbf{U}^{-1}\mathbf{A} = \Lambda\mathbf{U}^{-1}$, then premultiplying equation (21) by $\mathbf{U}^{-1}$, and ignoring the terms involving random error, gives

$$\mathbf{U}^{-1}\hat{\mathbf{x}} = \Lambda\mathbf{U}^{-1}\mathbf{x} + (\mathbf{I} - \Lambda)\mathbf{U}^{-1}\mathbf{x}^a \quad (23)$$

Let us think of the columns of $\mathbf{U}$, i.e., the eigenvectors of $\mathbf{A}$, as a representation of $\mathbf{x}$, i.e., $\mathbf{x} = \mathbf{U}\mathbf{u}$, where the vector $\mathbf{u}$ forms the representation coefficients. Thus $\mathbf{u} = \mathbf{U}^{-1}\mathbf{x}$, and the equation simplifies to $\hat{\mathbf{u}} = \Lambda\mathbf{u} + (\mathbf{I} - \Lambda)\bar{\mathbf{u}}$. This can be separated into components, showing that the elements of $\hat{\mathbf{u}}$ are scalar weighted means of the corresponding elements of $\mathbf{u}$ and $\mathbf{u}^a$:

$$\hat{u}_i = \lambda_i u_i + (1 - \lambda_i)u_i^a \quad (24)$$

Thus we can decompose the state vector into patterns, some of which (those with $\lambda_i \sim 1$) will be well reproduced by the measurement system, and others (with $\lambda_i \sim 0$) that will come mainly from the a priori vector.

As mentioned in section 2, a priori data can be regarded as one component of $\mathbf{c}$, the inverse model parameters, because they affect the retrieval without entering the forward model. Let the a priori state vector estimate be $\mathbf{x}^a$. Separate out the a priori from $\mathbf{c}$ explicitly in equation (12) and rewrite it as

$$\hat{\mathbf{x}} = I(F(\mathbf{x}, \mathbf{b}) + \epsilon_y, \hat{\mathbf{b}}, \mathbf{x}^a, \hat{\mathbf{c}})$$

where $\hat{\mathbf{c}}$ now represents any inverse model parameters that are not included in $\mathbf{x}^a$. The sensitivity of the retrieval to a priori is $\mathbf{D}_a = \partial I/\partial \mathbf{x}^a$, which must be equal to $\mathbf{I} - \mathbf{A}$, as shown above. If we assume that the 'correct' value of $\mathbf{x}^a$ is $\mathbf{x}$, then the contribution to the total error from this source is $\mathbf{D}_a(\mathbf{x}^a - \mathbf{x})$. The error covariance due to a priori error is therefore $\mathbf{D}_a\mathbf{S}_a\mathbf{D}_a^T$ where $\mathbf{S}_a$ is the covariance of $\mathbf{x}$ about $\mathbf{x}^a$. This is not a new source of error, it is identical to the null-space error (compare with equation (15)). Normally, $\mathbf{x}_a$ and $\mathbf{S}_a$ will be the same as the $\bar{\mathbf{x}}$ and $\mathbf{S}_x$ referred to in the discussion of null-space error.

Another conceptual approach is to consider a priori as a direct measurement with weighting functions equal to a unit matrix and with error covariance $\mathbf{S}_a$ and use it in an inverse model which has no explicit a priori. It is straightforward to show that the error covariance attributable to $\mathbf{x}^a$ obtained in this case would also be $\mathbf{D}_a\mathbf{S}_a\mathbf{D}_a^T$.

For those inverse methods that do not appear to use a priori data explicitly, it may be helpful to try to identify an implicit a priori. For ill-posed inverse problems, there must always be some implicit a priori to provide profile information in the null space. For well-posed problems, there may not be one. The implicit a priori is any profile (or class of profiles) that is unchanged by the transfer function. Its value can be constructed by solving $\mathbf{x}^a = T(\mathbf{x}^a, \hat{\mathbf{b}}, \hat{\mathbf{c}})$, if it has a solution.

8. Representations

The state vector $\mathbf{x}$ can in principle describe all the unknowns, at infinite resolution where relevant. In practice, we must use only a finite number of unknowns, so continuous profiles must be represented by, for example, an expansion in terms of some finite set of functions such as linear splines, polynomials, Fourier series, eigenvectors, etc. The error analysis must account for the error involved in an inadequate representation, or provide a rationale for choosing a representation that is "good enough". Let us regard the state vector $\mathbf{x}$ as comprising the coefficients of such a representation, rather than the original continuous representation. In this case the state vector may be shorter or longer than the the measurement vector, $m < n$ or $m > n$, and the problem may or may not be formally ill-posed.

The use of a representation creates its own null space, comprising all functions which cannot be represented by it. In the case of a linear representation it is all functions orthogonal to the basis vectors $\mathbf{R}$ of the representation.

Let $\mathbf{p}$ be a vector of n elements, representing a profile at sufficiently high resolution that differences from a continuous profile may safely be neglected. Let the coefficients $\mathbf{x}$ of a limited representation of $\mathbf{p}$ be written $\mathbf{x} = R(\mathbf{p})$, with $\mathbf{x}$ having r elements, $r < n$. Let the profile reconstructed from $\mathbf{x}$ be $\mathbf{p}_R = R^I(\mathbf{x})$. Most representations will be linear, but if not we can always linearize for the purpose of error analysis and write $\mathbf{x} = \mathbf{R}\mathbf{p}$ and $\mathbf{p}_R = \mathbf{R}^I\mathbf{x} = \mathbf{R}^I\mathbf{R}\mathbf{p} = \mathbf{A}_R\mathbf{p}$. In general, the representation will not be perfect, so $\mathbf{A}_R \neq \mathbf{I}$ and $\mathbf{p} \neq \mathbf{p}_R$. To estimate the accuracy of the representation, we need to know the statistics of $\mathbf{p} - \bar{\mathbf{p}}$. The error covariance for a representation of an ensemble of real profiles will be

$$\mathbf{S}_{real} = (\mathbf{A}_R - \mathbf{I})\mathbf{S}_p(\mathbf{A}_R - \mathbf{I})^T \quad (25)$$

This is the error in the representation of a real profile and is different from the error in the retrieved profile due to this source. The retrieval error is likely to be smaller, as the representation error will normally be in the fine structure not represented in the retrieval. In principle we can only estimate the extent to which a representation causes retrieval error if the forward model is formulated in terms of the high resolution profile, so we can define a transfer function between $\hat{\mathbf{x}}$ and $\mathbf{p}$:

$$\hat{\mathbf{x}} = T_{xp}(\mathbf{p}, \mathbf{b}, \mathbf{c}) = I(F(\mathbf{p}, \mathbf{b}), \mathbf{b}, \mathbf{c}) \quad (26)$$

which can be linearized as before:

$$\hat{\mathbf{x}} - \bar{\mathbf{x}} = \mathbf{A}_{xp}(\mathbf{p} - \bar{\mathbf{p}}) + \ \ldots \quad (27)$$

The averaging kernel matrix for $\mathbf{x}$ is $\mathbf{A} = \mathbf{A}_{xp}\mathbf{R}^I$, and for $\mathbf{p}$ is $\mathbf{A}_{pp} = \mathbf{R}\mathbf{A}_{xp}$. Unfortunately, the only way to quantify the difference between the retrieval being analyzed, and a hypothetical full resolution retrieval, is to construct the full resolution retrieval. The error in $\mathbf{x}$ due to assuming that $\mathbf{p}$ is of the form required by the representation is $\mathbf{A}_{xp}(\mathbf{I} - \mathbf{A}_r)\mathbf{p}$, with covariance $\mathbf{A}_{xp}(\mathbf{I} - \mathbf{A}_r)\mathbf{S}_p(\mathbf{I} - \mathbf{A}_r^T)\mathbf{A}_{xp}^T$.

9. Examples

As an illustration, the above concepts have been applied to the Solar Backscatter Ultraviolet (SBUV) ozone sounding instrument on Nimbus 7 [*Heath et al.*, 1975]. A simplified retrieval method has been chosen for the the purpose of illustration, rather than that used by NASA for operational retrievals. Examples of the application to operational algorithms will be found in chapter 3 of the report of the NASA/WMO Ozone Trends Panel [*World Meteorological Organisation*, 1990], where the analyses have been carried out by the data processing teams of a variety of ozone sounding instruments (SBUV, Umkehr, the Stratospheric Aerosol and Gas Experiment, the Solar Mesosphere Explorer and the Limb Infrared Monitor of the Stratosphere). *Connor and Rodgers* [1989] have also used the approach to compare the onion peeling and optimal estimation methods for retrieval of profiles from a limb sounder.

A set of weighting functions for SBUV is shown in Figure 1, which depend on an assumed mean ozone profile as a linearization point. Only this one linearization point has been studied. The state vector $\mathbf{x}$ consists of logarithms of layer amounts of ozone in a set of closely spaced layers, and the measurement vector $\mathbf{y}$ consists of the logarithm of the measured albedo in 10 spectral bands, together with a measurement of total ozone that is obtained from a separate nonlinear retrieval from three spectral bands. Curve A is the weighting function for the total ozone measurement.

The retrieval approach analyzed is similar to, but not the same in detail as, that used by the SBUV team. It is a simple-minded maximum likelihood retrieval [*Rodgers*, 1976, equation (21)], chosen for algebraic convenience. It is assumed that the measurement noise covariance $\mathbf{S}_\epsilon = 10^{-4}\mathbf{I}$, i.e., 1% in all channels with no interchannel correlation, and that the a priori covariance $\mathbf{S}_a$ is of the form given by equation (16) with $l = 5$ km and S_{zz}=0.04, i.e., a 20% variance in ozone layer amount. This form is qualitatively reasonable, but it has been chosen for illustration only, and should not be used for serious retrievals without validation.

Sensitivity to model error has not been included in this illustration, as the full SBUV forward model was not easily available. This will be found in the Ozone Trends Panel report.

The contribution functions $\mathbf{D}_y$ for this retrieval may be written [*Rodgers*, 1976] as

$$\begin{aligned}\mathbf{D}_y &= \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1} \\ &= (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1}\mathbf{K}^T\mathbf{S}_\epsilon^{-1}\end{aligned} \quad (28)$$

where the subscript x has been dropped from $\mathbf{K}_x$ for clarity. They are given in Figure 2, showing that each channel contributes in a complicated way to the overall profile, although there is a slight tendency for the information to be put into the profile in the same general

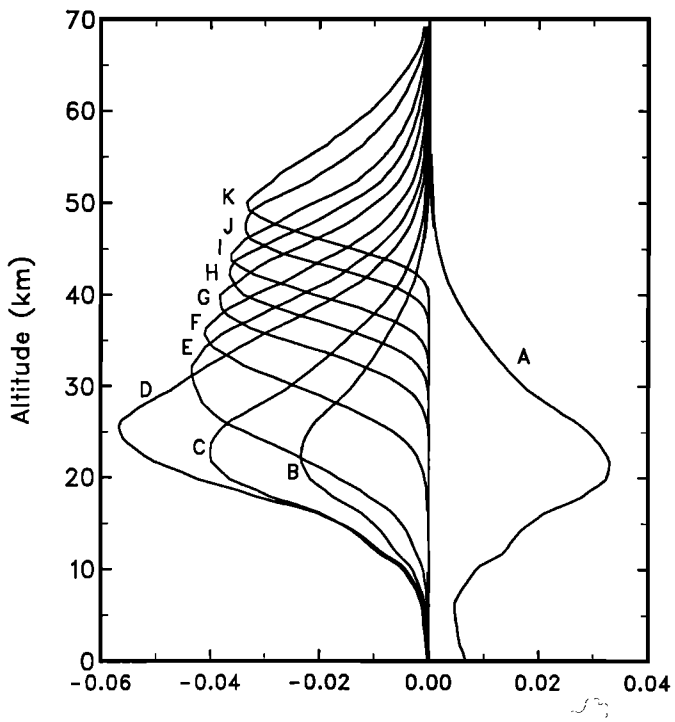

Fig. 1. Weighting functions for the SBUV ozone sounding instrument.

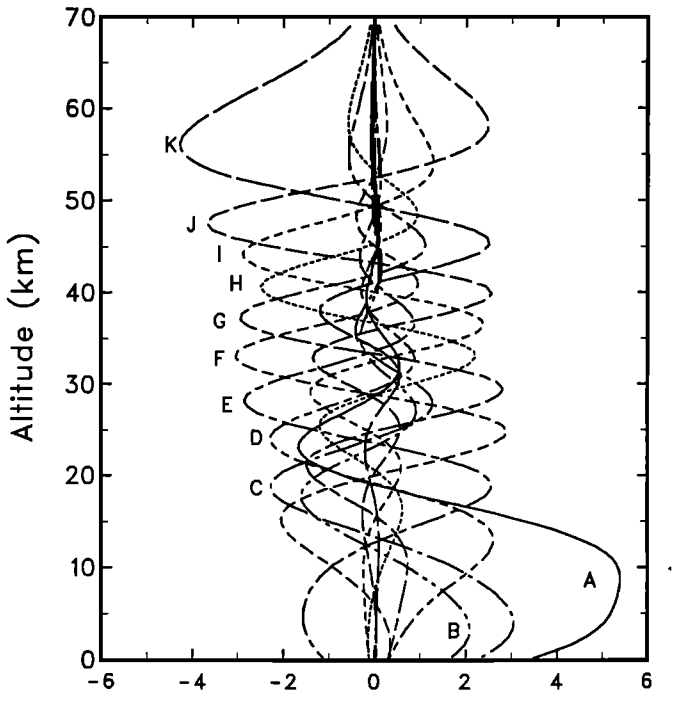

Fig. 2. Contribution functions for a maximum likelihood retrieval. Alphabetical labeling corresponds to Figure 1.

altitude region as the peak of the corresponding weighting function.

Figure 3 shows the averaging kernels, $\mathbf{A} = \mathbf{D}_y\mathbf{K}$, illustrating how the retrieval smooths the true profile. The retrieved state vector element at the altitude indicated is a weighted mean of the true state vector, with the averaging kernal as a weight. For clarity, these are not plotted at every level. The width of the primary peak of each curve is a measure of the resolution of the observing system. Note that the retrieval at the top two levels shown (for 62 and 69 km) does not correspond to the true profile around these levels as there is little information in the measurements here. Similarly, the retrieval below about 20 km (curves for 6 and 13 km) does not correspond well to the true ozone below 20 km for the same reason.

From equation (15) and the above expression for $\mathbf{D}_y$, the null-space error covariance is

$$\mathbf{S}_N = (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1}\mathbf{S}_a^{-1}(\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1} \quad (29)$$

Figure 4 shows the nine largest null-space error patterns. These are proportional to the eigenvectors of $\mathbf{S}_N$, showing orthogonal structures that the observing system cannot see. Errors in the profiles due to this source, as with the measurement noise error patterns, will be a linear combination of these functions, with random coefficients having unit variance. These vectors tend to have larger values at the top and bottom of the plot, where there is little information about the true profile in the measurements. Further, they all have a fairly short scale vertical structure, somewhat shorter than the width of the averaging kernels, but not very short. The a priori covariance matrix chosen allows little variability on scales much shorter than 5 km, so very fine scales do not appear.

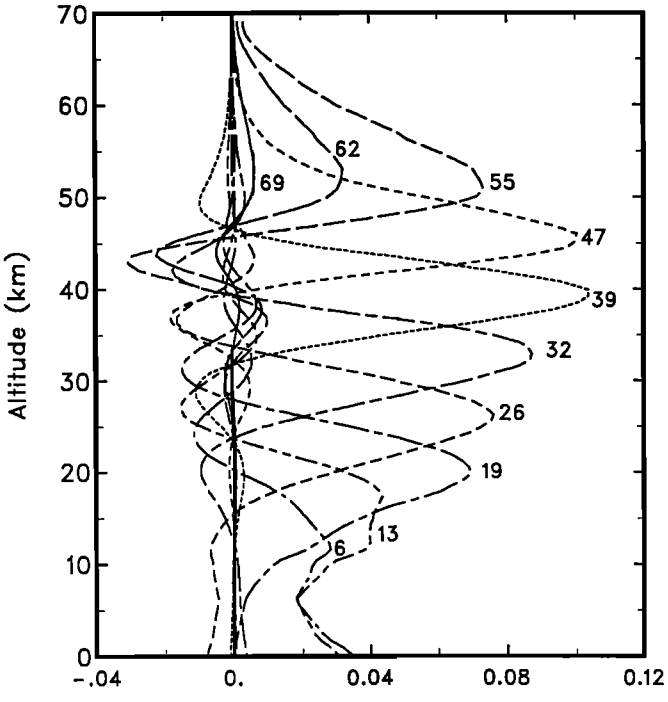

Fig. 3. Averaging kernels for selected altitudes. Curves are labeled with the altitude in Kilometers.

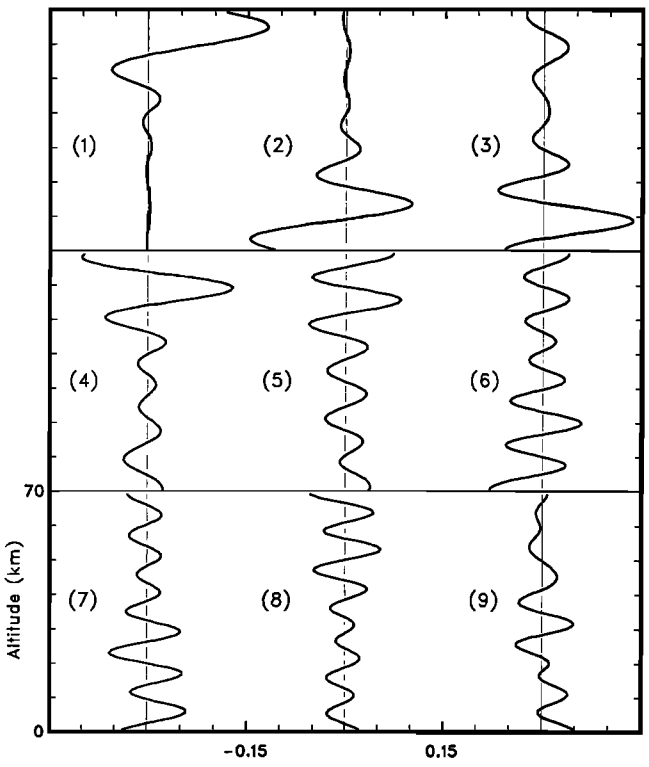

Fig. 4. The nine largest null-space error patterns.

Using equations (13) and (27), the measurement error covariance is

$$\mathbf{S}_M = (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1}\mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K}(\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1} \quad (30)$$

Error patterns for the contribution of measurement noise to the overall error are shown in Figure 5. Patterns with broader scale structure generally correspond to smaller errors, as there is information in the measurements to determine these components well. Patterns with short vertical scale also give small error, as the inverse model does not attempt to determine these, information on this scale coming from the a priori profile. Larger errors are associated with intermediate scale patterns, on a scale comparable with the weighting function or averaging kernel width.

The errors of the retrieval at each level are given in Figure 6. Curve B is the total residual standard deviation, or the total error ignoring interlevel correlations. This is simply the square root of the diagonal of the total random error

$$\mathbf{S}_M + \mathbf{S}_N = (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1} \quad (31)$$

Curve A gives the component due to measurement noise, $\mathbf{S}_M$. These curves are simply the root-sum-squares of the corresponding error patterns, which clearly give more detailed information.

An ensemble of retrieved profiles will not have the same covariance as that of the corresponding ensemble of real profiles. From equations (8) and (9), and ignoring model errors, the solution ensemble covariance is

$$\mathbf{A}\mathbf{S}_a\mathbf{A}^T + \mathbf{D}_y\mathbf{S}_\epsilon\mathbf{D}_y = \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1}\mathbf{K}\mathbf{S}_a \quad (32)$$

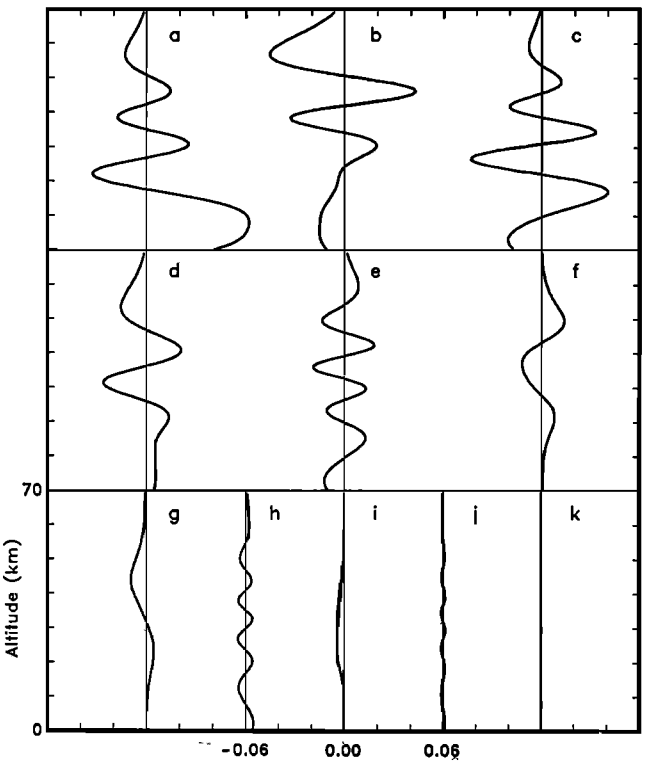

Fig. 5. Error patterns due to instrumental noise.

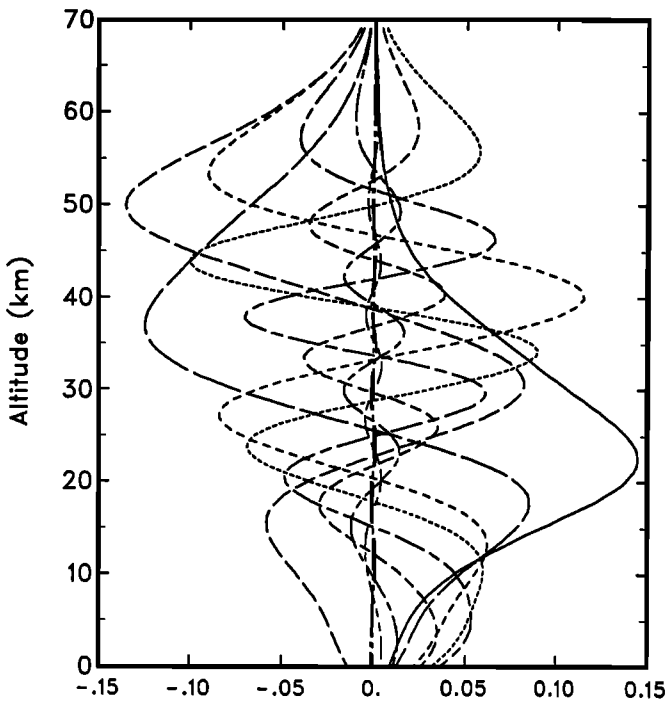

Fig. 7. Solution ensemble eigenvectors scaled by the square root of the eigenvalue.

where the covariance of the ensemble of real profiles has been taken to be the same as $\mathbf{S}_a$. Figure 7 shows the eigenvectors of the covariance of an ensemble of solutions, multiplied by the square root of the corresponding eigenvector. These are similar to the error patterns, but describe the ensemble rather than the error, a retrieved profile can be thought of as being made up of the mean state vector $\hat{\mathbf{x}}$ plus a linear combination of these patterns, with coefficients that will have a rms value of unity. This indicates the kind of structure that can be measured with the SBUV instrument and this retrieval approach. Incidentally, these patterns can be used to set up an efficient profile representation, with only eight or nine coefficients, by using only the patterns which have a significant amplitude.

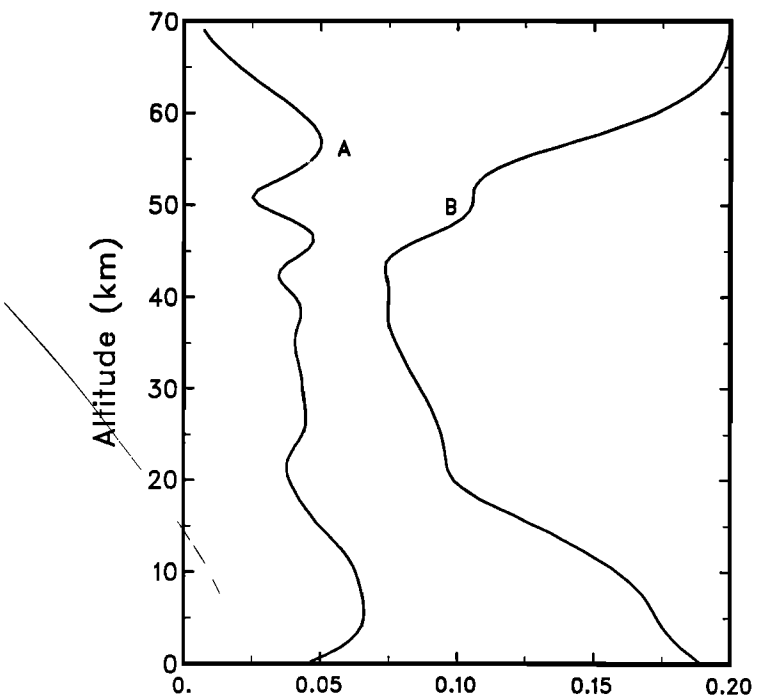

Fig. 6. The rms error due to instrumental noise (curve A) and the total rms error (curve B).

The retrieved profile is related to the true profile and the a priori by

$$\hat{\mathbf{x}} = \mathbf{x}_a + \mathbf{D}_y\mathbf{K}(\mathbf{x} - \mathbf{x}_a) = \mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A})\mathbf{x}_a \qquad (33)$$

which is of the same form as equation (21). As discussed in section 7, the contribution of a priori to the retrieval can be described in terms of the eigenvectors of the averaging kernel matrix $\mathbf{A}$. The largest nine eigenvalues of $\mathbf{A}$ and the corresponding eigenvectors are shown in Figure 8. The first four eigenvalues are essentially unity, implying that the a priori data contribute nothing to components of the solution with this vertical structure. The next three or four vectors correspond to vertical structure where both the measurements and the a priori contribute. At scales shorter than this only the a priori contributes, the measurements have nothing to say.

10. Summary

A formal error analysis for profile retrieval can be established, which is independent of the nature of the retrieval method, provided that the measurement process can be modeled adequately. The error separates naturally into three components, (1) random error due to measurement

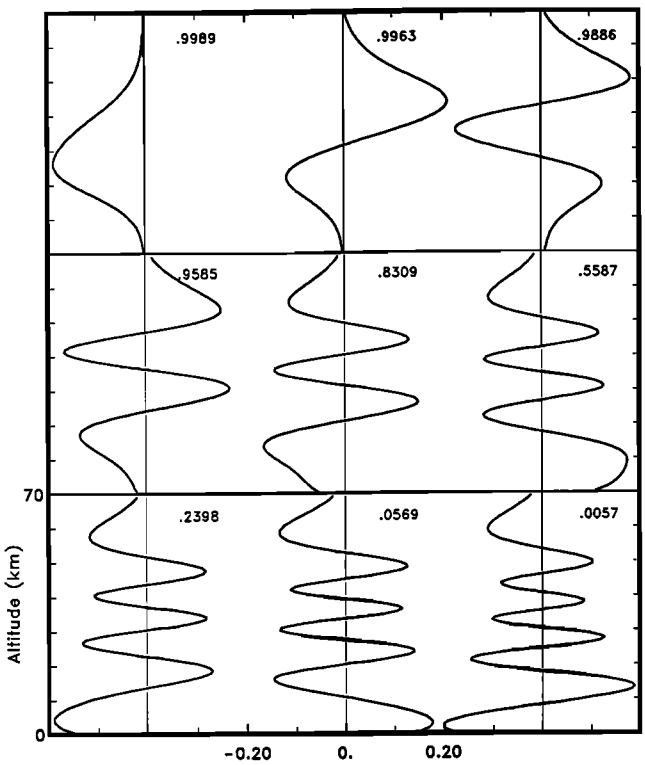

Fig. 8. The first nine eigenvectors and eigenvalues of the averaging kernel matrix.

noise, (2) model error due to uncertain model parameters and inverse model bias, and (3) null-space error due to the inherent finite vertical resolution of the observing system, and lack of information outside the range of the weighting functions. A recipe is given for evaluating each of the components in any particular case.

Most of the error terms appear as covariance matrices, rather than simple error variances. These matrices can be interpreted in terms of "error patterns", which are statistically independent contributions to the total error. They are the multidimensional equivalent of error bars.

A full error analysis of a remote sounding observing system, in the neighborhood of some state $\bar{\mathbf{x}}$, includes the following calculations:

1. Obtain the contribution function matrix $\mathbf{D}_y$, algebraically or by perturbing the inverse model.

2. Obtain the averaging kernel matrix $\mathbf{A}$, algebraically or by perturbing the transfer function. Spatial resolution of the observing system is described by the rows or the columns of $\mathbf{A}$.

3. Obtain the sensitivities to systematic errors $\mathbf{A}_b$ and $\mathbf{A}_c$ by perturbing the transfer function with respect to the model parameters $\mathbf{b}$ and $\mathbf{c}$.

4. Calculate the error covariance due to measurement error $\mathbf{D}_y\mathbf{S}_\epsilon\mathbf{D}_y^T$.

5. Calculate the bias of the inverse model $T(\bar{\mathbf{x}}, \mathbf{b}, \mathbf{c}) - \bar{\mathbf{x}}$ and the systematic error covariances $\mathbf{A}_b\mathbf{S}_b\mathbf{A}_b^T$ and $\mathbf{A}_c\mathbf{S}_c\mathbf{A}_c^T$.

6. If possible, construct an ensemble covariance $\mathbf{S}_x$ describing the possible variation of the profile and calculate the null-space error covariance $(\mathbf{A} - \mathbf{I})\mathbf{S}_x(\mathbf{A} - \mathbf{I})$ describing the components of profile variability that the observing system cannot see.

7. If required for understanding, compute the error patterns of each of the error covariance matrices, and the eigenvectors of $\mathbf{A}$ to characterize the contribution of a priori data to the retrieval.

The basic set of diagnostics for characterizing errors will consist of (1) the averaging kernels, (2) the measurement noise error patterns and the level-by-level rms error, (3) the inverse model bias, and (4) the model parameter error patterns. This set characterizes all the terms in the expression for the total error (equation (14)).

Acknowledgments. I would like to thank Bob Hudson of Goddard Space Flight Center for providing the weighting functions for the SBUV, Howard Roscoe, Paul Bailey, Jim Coakley and Brian Connor for reading the early drafts and making many helpful suggestions, and the NASA/WMO Ozone Trends Panel for providing the motivation for writing up this error analysis. I would also like to thank the Atmospheric Chemistry Division of NCAR for support, and the opportunity to carry out this work. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

REFERENCES

Backus, G. E., and J. F. Gilbert, Uniqueness in the inversion of inaccurate gross Earth data, *Philos. Trans. R. Soc. London, Ser. A, 266*, 123–192, 1970.

Connor, B. J., and C. D. Rodgers, Optimal estimation, onion peeling, and a combination of the two, in *RSRM '87: Advances in Remote Sensing Retrieval Methods*, edited by A. Deepak, H. E. Fleming and J. S. Theon, pp. 271–281, A. Deepak Publishing, Hampton, Va., 1989.

Heath, D. F., A. J. Kruger, H. R. Roeder, and B. D. Henderson, The Solar Backscatter Ultraviolet and Total Ozone Mapping Spectrometer (SBUV/TOMS) for Nimbus G, *Opt. Eng., 14*, 323–331, 1975.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Academic, San Diego, Calif., 1984.

World Meteorological Organisation, *Report of the International Ozone Trends Panel - 1988, WMO Report No. 18*, World Meteorological Organisation, Geneva, 1990.

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

C. D. Rodgers, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, England.

(Received March 8, 1989;
revised September 28, 1989;
accepted September 28 1989.)