AT721 Section 13:

Characterization and error analysis

General references:
Rodgers, C.D., 2000; Inverse Methods for atmospheric sounding, *World Sci*, chapter 3
L'Ecuyer and Stephens, 2002; An Uncertainty model Bayesian Monte carlo retrieval algorithms: application to the TRMM observing system, *Q.J.Roy. Meteorol. Soc.*, 128, 1713-1737.

In chapter 12 we hinted at how we might select the best or optimal solution from a large number of possible solutions of an ill-posed inverse problem. As a step in designing these optimal inverse methods it is important that the general characteristics of retrievals be understood in order to decide which properties are most appropriate to optimize. A retrieval method without characterization and error analyzes is of little value. By *characterization* we mean the sensitivity of the retrieved state to the true state and by *error* we mean the sensitivity of the retrieval to all sources of error associated with the transfer function introduced below.

13.1 Characterization of the retrieval problem

The basic function of an observing system can be described in terms of the observing system transfer function as presented in Fig 13.1. The transfer function determines the relationship between a given input of the system and the desired output of the system. The system connects the input signal $\mathbf{x}(t)$ that varies with time in some manner to the actual measured quantity $\mathbf{y}(t)$, satellite radiances for example, which are then connected to the output signal of the system $\hat{\mathbf{x}}(t)$ which, under ideal circumstances identically reproduces the input signal. Here we use the circumflex to indicate estimated quantities, differing in reality from actual quantities in ways that can be analyzed with methods described in this chapter. The input may be taken to be a given geophysical parameter whose measurement over some prescribed time T is the objective of the observing system.

There are two basic parts to the observing system that require characterization. The first is hidden from view in Fig. 13.1 and has to do with the space-time sampling characteristics of the observing system which introduce a class of errors to the measurements $\mathbf{y}(t)$ that too can vary in time as the sampling properties of the system vary. These sampling errors or errors of representation are discussed more fully in chapter 16 and are generally imposed on the system by the chosen orbit of the satellite platform.

The second basic component of the observing system deals with the two main elements of the transfer function as shown in Fig. 13.1. One element is the forward function as expressed by the forward model and the second is the inverse model or retrieval system. The remainder of this chapter is concerned with the characterization and error analysis of these two aspects of the transfer function.

13.1.1 The forward Model

We can express the general forward problem in the following heuristic way

$$\mathbf{y} = F(\mathbf{x}, \mathbf{b}) + \varepsilon \tag{13.1}$$

Fig. 13.1 A climate observing system presented in the form of a transfer function (outer box) and component transfer functions (the two inner boxes). This system produces an output Z' that differs from the input Z due to the internal parameters of the observing system. One of the goals of this system is to reconstruct the time variability of Z from the time variability of Z'.

where the forward function $F(\mathbf{x}, \mathbf{b})$ describes the complete physics of the measurements and for most observing systems, and especially those flown on satellites, this forward model represents various forms of radiative transfer theory as discussed in the first section of this book. Such theory relates the state $\mathbf{x}$ to the measured signal $\mathbf{y}$. The vector $\mathbf{b}$ are those parameters that influence the measurement but are not intended as quantities to be retrieved. The uncertainty attached to the latter parameters contribute to the total uncertainty. The error term ε is the measurement error and includes contributions such as detector noise and will be referred to this as 'measurement noise'.

13.1.2 The retrieval method

The inversion of (13.1) may also be expressed heuristically as

$$\hat{\mathbf{x}} = \mathbf{R}(\mathbf{y}, \hat{\mathbf{b}}, \mathbf{x}_a, \hat{\mathbf{c}}) \tag{13.2}$$

where again for emphasis the circumflex represents an estimated quantity. $\hat{\mathbf{b}}$ is our best estimate of $\mathbf{b}$. As mentioned above characterization of the observing system refers to the sensitivity of the retrieved state $\hat{\mathbf{x}}$ to the true state $\mathbf{x}$, i.e. to the quantity $\partial\hat{\mathbf{x}}/\mathbf{x}$ and by error analysis we mean the sensitivity of the retrieval to all sources of error. The vectors $\mathbf{x}_a$ and $\mathbf{c}$ comprise parameters that do not appear in the forward function but affect the retrieval. $\mathbf{x}_a$ is the *a priori* state vector and the vector $\mathbf{c}$ is a catch all for any parameters used in the retrieval (such as convergence criteria) and is referred to as the retrieval method parameter vector.

For many problems the *a priori* is the information that matters most yet for other problems $\hat{\mathbf{b}}$ dominate the retrieval process. Some methods use explicit *a priori*, while others do not and yet others claim not to use any such information where in reality such information is concealed in the method.

13.2 Linearization of the transfer function

Substituting (13.1) into (13.2) yields

$$\hat{\mathbf{x}} = \mathbf{R}(F(\mathbf{x}, \mathbf{b}) + \varepsilon, \hat{\mathbf{b}}, \mathbf{x}_a, \hat{\mathbf{c}}) \tag{13.3}$$

The forward model itself often represents a significant source of difficulty in retrievals yet many published retrieval studies overlook this source of error. The difficulty with the forward model may arise from the fact that the physics of the problem is too complex, such as the case of complexities introduced in the case of 3D radiative transfer, or that the physics is itself uncertain. For whatever reason, an approximate forward model f is invariably used, namely

$$F(\mathbf{x}, \mathbf{b}) \approx f(\mathbf{x}, \mathbf{b}, \mathbf{b}') \tag{13.4}$$

where $\mathbf{b}$ is decomposed into a set that contains $\mathbf{b}'$ that represent model parameters now required specifically by the approximate form of the forward model. Substituting into (12.3) yields

$$\hat{\mathbf{x}} = \mathbf{R}(F(\mathbf{x}, \mathbf{b}) + \Delta f(\mathbf{x}, \mathbf{b}, \mathbf{b}') + \varepsilon, \hat{\mathbf{b}}, \mathbf{x}_a, \hat{\mathbf{c}})$$

where

$$\Delta f(\mathbf{x}, \mathbf{b}, \mathbf{b}') = f(\mathbf{x}, \mathbf{b}, \mathbf{b}') - F(\mathbf{x}, \mathbf{b}) \tag{13.5}$$

is the model error.

To obtain the basic sensitivities of the transfer function we now linearize our transfer function about the states $\mathbf{x} = \mathbf{x}_a, \mathbf{b} = \hat{\mathbf{b}}$

$$\hat{\mathbf{x}} = \mathbf{R}(F(\mathbf{x}_a, \hat{\mathbf{b}}) + \mathbf{K}(\mathbf{x} - \mathbf{x}_a) + \mathbf{K}_b(\mathbf{b} - \hat{\mathbf{b}}) + \Delta f(\mathbf{x}, \mathbf{b}, \mathbf{b}') + \varepsilon, \hat{\mathbf{b}}, \mathbf{x}_a, \hat{\mathbf{c}}) \tag{13.6}$$

where

$$\mathbf{K} = \frac{\partial F(\mathbf{x}, \mathbf{b})}{\mathbf{x}}$$

and

$$\mathbf{K}_b = \frac{\partial F(\mathbf{x}, \mathbf{b})}{\mathbf{b}}$$

and $\mathbf{K}$ is the Jacobian matrix. Note that this is the derivative of the forward model with respect to $\mathbf{x}$.

13.2 Error analysis

Next linearize (13.3) with respect to $\mathbf{y}$. We write (13.6) as

$$\hat{\mathbf{x}} = \mathbf{R}[F(\mathbf{x}_a, \hat{\mathbf{b}}) + \Delta\mathbf{y}, \hat{\mathbf{b}}, \mathbf{x}_a, \hat{\mathbf{c}}]$$

where

$$\Delta\mathbf{y} = \mathbf{K}(\mathbf{x} - \mathbf{x}_a) + \mathbf{K}_b(\mathbf{b} - \hat{\mathbf{b}}) + \Delta f(\mathbf{x}, \mathbf{b}, \mathbf{b}') + \varepsilon \tag{13.7}$$

such that

$$\hat{\mathbf{x}} = \mathbf{R}[F(\mathbf{x}_a, \hat{\mathbf{b}}, \hat{\mathbf{b}}, \mathbf{x}_a, \hat{\mathbf{c}}] + \frac{\partial R}{\partial \mathbf{y}}\Delta\mathbf{y}$$

and thus

$$\hat{\mathbf{x}} = \mathbf{R}[F(\mathbf{x}_a, \hat{\mathbf{b}}), \hat{\mathbf{b}}, \mathbf{x}_a, \hat{\mathbf{c}}] + \frac{\partial R}{\partial \mathbf{y}}[\mathbf{K}(\mathbf{x} - \mathbf{x}_a) + \mathbf{K}_b(\mathbf{b} - \hat{\mathbf{b}}) + \Delta f(\mathbf{x}, \mathbf{b}, \mathbf{b}') + \varepsilon]$$

with rearrangement becomes

$$\hat{\mathbf{x}} - \mathbf{x}_a = \mathbf{R}[f(\mathbf{x}_a, \hat{\mathbf{b}}), \hat{\mathbf{b}}, \mathbf{x}_a, \hat{\mathbf{c}}] - \mathbf{x}_a \qquad \text{bias error}$$

$$+\mathbf{A}(\mathbf{x}-\mathbf{x}_a) \quad \text{smoothing error}$$

$$+\mathbf{D}_y\epsilon_y \quad \text{retrieval error} \tag{13.8}$$

The first term is the bias error that results from a simulated retrieval using a simulated error free measurement of the a priori state computed with the forward model. This error can be analyzed via simulation using synthetic data and it should be zero by construction.

13.2.1 Smoothing Error

The second term represents the way the observing system smooths information about the state $\mathbf{x}$. We can better understand this by rearranging (13.8) in the form

$$\hat{\mathbf{x}}-\mathbf{x}_a = \mathbf{A}(\mathbf{x}-\mathbf{x}_a)+\mathbf{D}_y\epsilon_y \tag{13.9}$$

where we have introduced $\mathbf{D}_y = \partial\mathbf{R}/\partial\mathbf{y}$ and

$$\mathbf{A} = \mathbf{D}_y\mathbf{K} \tag{13.10}$$

and further write

$$\epsilon_y = \mathbf{K}_b(\mathbf{b}-\hat{\mathbf{b}})+\Delta f(\mathbf{x},\mathbf{b},\mathbf{b}')+\varepsilon) \tag{13.11}$$

Suppose for the sake of discussion that ϵ_y is negligibly small, then it follows that

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{x}+[\mathbf{E}-\mathbf{A}]\mathbf{x}_a \tag{13.11}$$

and thus for this ideal case, the retrieval is characterized by

$$\frac{\partial\hat{\mathbf{x}}}{\partial\mathbf{x}} = \mathbf{A} \tag{13.12}$$

Ideally the matrix $\mathbf{A}$ should be a unit matrix if the system were perfect. In reality, however, this matrix differs from the unit matrix and further examination of the properties of the matrix is required. To fix ideas, suppose our state vector $\mathbf{x}$ is a set of temperatures where each element represents a layer temperature and where the vector represents a piece-wise continuous temperature. Then the retrieved temperature for an individual layer follows as

$$\hat{x}_i = \sum_{j=1}^{N} A_{ij}x_j \tag{12.13}$$

where the retrieved layer temperature $\hat{x}_i$ are in a sense related not just to the corresponding actual layer temperature x_i but also to the entire profile weighted according to A_{ij}. Thus the rows of A indicated the extent that the real temperature information is spread across layers, and thus $\mathbf{A}$ acts to smooth the profile. This is highlighted in Fig 13.2 showing examples of the A matrix derived for an observing system defined by a forward model of IR radiances, an a priori temperature constraint of varying strengths and a temperature retrieval procedure based on (12.x). The precise details of the retrieval system do not matter for this illustration. What is highlighted in this figure is how the constraint broadens the structure of the

Fig. 13.2 Examples of the smoothing Kernels derived for the retrieval of temperature from IR radiances assuming different amounts of constraint. Note how adding constraints smears information but makes the matrix much smoother.

A matrix deviating from a diagonal form more characteristic of the ideal case. Without the constraint, we can see how oscillations appear in the matrix symptomatic of instabilities discussed previously.

By contrast, the columns of A give the response of the retrieval to a delta function perturbation in the state vector. This column thus represents ho variation in the real state at a given level is spread across the retrieved state vector by the retrieval process.

The magnitude of the elements of the **A** matrix contain yet further information about the observing system that helps us characterize it. Consider (13.8) in the form

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{x} + [\mathbf{E} - \mathbf{A}]\mathbf{x}_a + \mathbf{D}_y\epsilon_y \tag{13.14}$$

and again suppose for now that ϵ_y is small, then the closer are the magnitudes of the diagonal elements to unity, the less the a priori information creeps into the solution and the more the actual observations contribute to the retrieval.

Figure 13.3 offers an example of the diagonal values of the **A** matrix taken from the work of Engelen and Stephens (1999) as it applies to the retrieval of water vapor.

Fig. 13.3 The diagonal values of the for the retrieval of layer water vapor from the surface-700mb and 500-300 mb.

discuss

13.2.2 Retrieval Error

Further re-arranging of () and () leads to the error in $\mathbf{x}$ of the form

$$\begin{aligned}\hat{\mathbf{x}} - \mathbf{x} = [\mathbf{A} - \mathbf{E}](\mathbf{x} - \mathbf{x}_a) \quad & \text{smoothing error} \\ +\mathbf{G}_y\mathbf{K}_b(\mathbf{b} - \hat{\mathbf{b}}) \quad & \text{model paramter error} \\ +\mathbf{G}_y\Delta f(\mathbf{x}, \mathbf{b}, \mathbf{b}') \quad & \text{forward parameter error} \\ +\mathbf{D}_y\epsilon_y \quad & \text{retrieval noise}\end{aligned} \tag{13.15}$$

Because the true state $\mathbf{x}$ is not usually known, we cannot estimate the actual smoothing error. What is needed is the statistics of the error and this may be deduced from ensembles of states and could be described by $\mathbf{S}_a$

The total retrieval error, as given by covariance matrix (12.x) is

$$\mathbf{S}_x^{-1} = \mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_a^{-1} \tag{13.16}$$

which can also be written as

$$\mathbf{S}_x = \mathbf{G}_y\mathbf{S}_y\mathbf{G}_y^T + \mathbf{G}_y\mathbf{S}_a\mathbf{G}_y \tag{13.17}$$

where the first term of the rhs represents the combined errors of models and measurements, and the second term is the error that is projected into the solution by uncertainties that are attached to the a priori state. The first term can be further written

$$\mathbf{G}_y\mathbf{S}_y\mathbf{G}_y^T = \mathbf{G}_y[\mathbf{S}_b + \mathbf{S}_f + \mathbf{S}_\epsilon]\mathbf{G}_y^T \tag{13.18}$$

where we write $\mathbf{S}_b$ for the error covariance matrix that represents the model parameter error term $[\mathbf{K}_b(\mathbf{b} - \hat{\mathbf{b}})]$, $\mathbf{S}_f$ for the forward model error term $[\Delta f(\mathbf{x}, \mathbf{b}, \mathbf{b}')]$ and $\mathbf{S}_\epsilon$ for the measurement noise ϵ

13.3 Examples of Retrieval Error

1. Water vapor retrieval
2. Lidar example

12.2.1 Measurement Noise

Random - discuss the different forms o f measurement noise viz-a-viz Philip's notes

13.4 Representing Covariances

Errors of vector $\mathbf{x}$ are expressed by the covariance matrix $\mathbf{S}_x$. The diagonal elements of $\mathbf{S}_x$ are the familiar error variances of the elements of the x and the off diagonal elements indicate the degree of correlation of errors between elements of $\mathbf{x}$. The error covariance matrix is defined as

$$S_x(i,j) = \frac{1}{N}\sum_{k=1}^{N}[\varepsilon_i - \langle\varepsilon\rangle]_k[\varepsilon_j - \langle\varepsilon\rangle]_k \tag{13.19}$$

where $\varepsilon_i = \hat{\mathbf{x}}_i - \mathbf{x}_i$ and $\langle\rangle]$ is the average error or bias in the data and the summation is over N data sets.

13.4.1 Microwave rainfall example

L'Ecuyer and Stephens (2002) apply this to define the model error covariance associated with the forward model simulations of the TMI microwave radiances. They define this covariance error for three rain rate regimes, $0 < R < 5$ mmhr^{-1}, $5 < R < 20$ mmhr^{-1} and $R > 20$ mmhr^{-1}. They separate this into two independent sources of error - a surface emission error ($\mathbf{S}^R_{SFC}$) based on the difference between simulated clear sky microwave radiances and measured radiances under clear sky conditions and radiative transfer model errors inferred from comparisons with other models under raining conditions ($\mathbf{S}^R_{RT}$). They combine these errors

$$\mathbf{S}^R = \mathbf{S}^R_{SFC} + \mathbf{S}^R_{RT}$$

and provide examples of these covariance matrices, expressed in the form correlation matrices

$$C_{ij} = \frac{S^R_{ij}}{\sigma_i \sigma_j}$$

as shown in Figs. 13.4a and b. Clearly the model errors of different microwave radiometer channels are correlated, especially as the rain rate increases but the errors are much less correlated for simulations of emission (P) and scattering (S) indices defined as

$$P = \frac{T_{bV} - T_{bH}}{T_{bV,0} - T_{bH,0}}$$

$$S = PT_{bV0} + (1 - P)T_c - T_{bV}$$

where $T_c = 273K$. T_{bV}, T_{bH} are the horizontally and vertically polarized brightness temperatures and the subscript '0' are the respective brightness temperatures in the absence of clouds or precipitation.

Fig. 13.4 covariance matrices

13.4.2 Error correlations in the temperature sounding problem

In general we can find a basis for which the errors are independent by diagonalizing the covariance matrix - which as we have seen previously require finding the eigen-values and eigen-vectors of $\mathbf{S}_x$, according to

$$\mathbf{S}_x \mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{13.20}$$

Since the covariance matrix is symmetric, the eigenvectors are orthogonal and these in a sense can be thought of as representing the 'error patterns' in the sense that the total error can be expressed as a sum of the individual patterns

$$\epsilon_x = \sum_{i=1}^{N} a_i \mathbf{u}_i$$

where the relative weighting of these patterns is determined by the magnitude of the respective eigenvalues.

Consider the example introduces by Rodgers (ch 3) where he proposes

$$S_{a,ij} = \sigma_a^2 \exp(- \mid i - j \mid \delta z/H) \tag{13.21}$$

as an approximate form for the *a priori* covariance error matrix characteristic of temperature sounding problem. This form broadens the matrix about the diagonal in a way representing the broadening of the error matrix due to the effects of the spreading weighting function. The values of $\delta z/h$ are chosen to characterize the width of typical temperature weighting functions.

Fig. 13.1 An example of the ten most significant error patterns (eigenvectors) of the *a priori* covariance matrix (13.21).

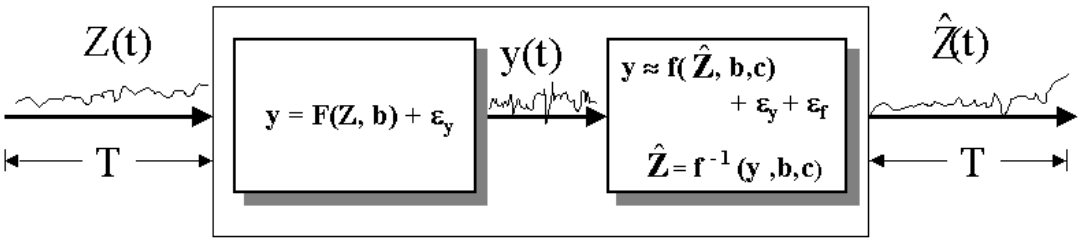


Fig. 13.1

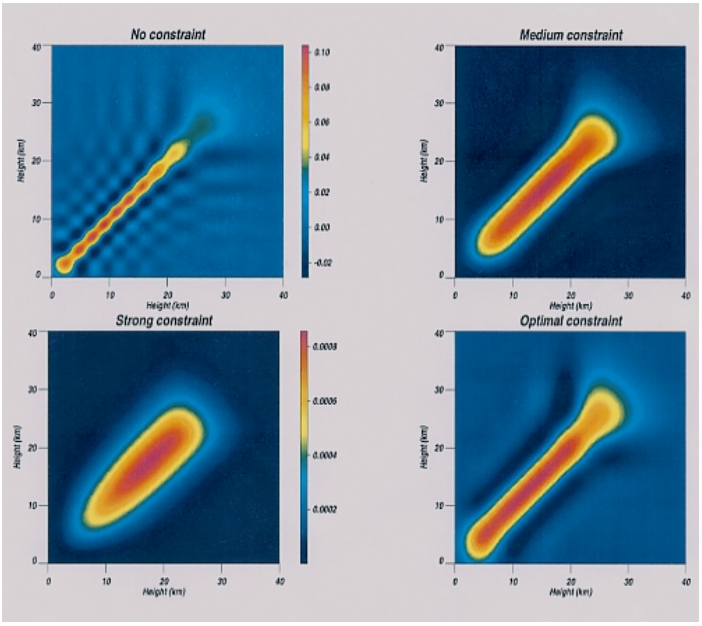


Fig. 13.2

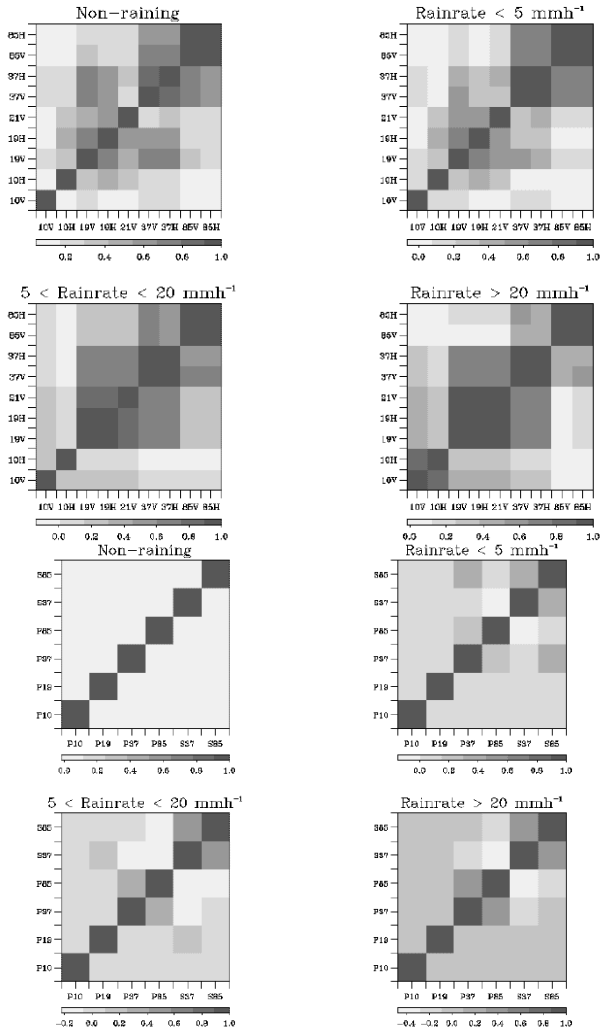


Fig. 13.3

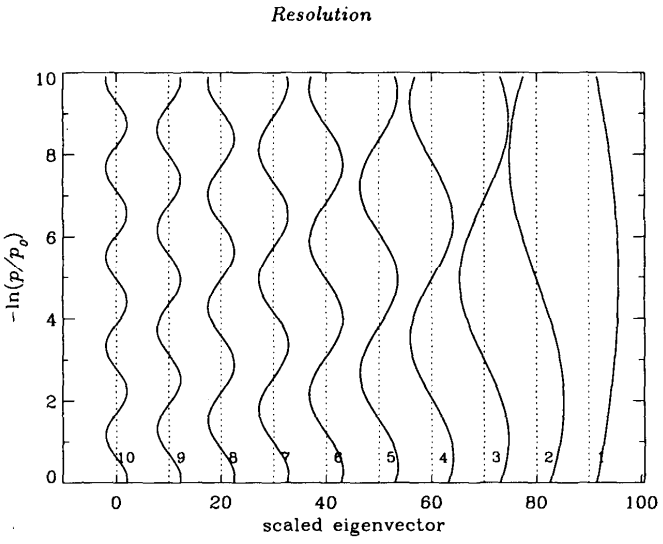


Fig. 13.4

An example of a Lidar observing system in tenuous media

$$\ln(C \times P(r)) \approx \ln(\beta_m(r) + k\sigma(r)) - 2\int_0^r \eta(r)\sigma(r')dr'$$

k = Backscatter to extinction

η = multiple scattering

β_m = Rayleigh scattering

C = calibration

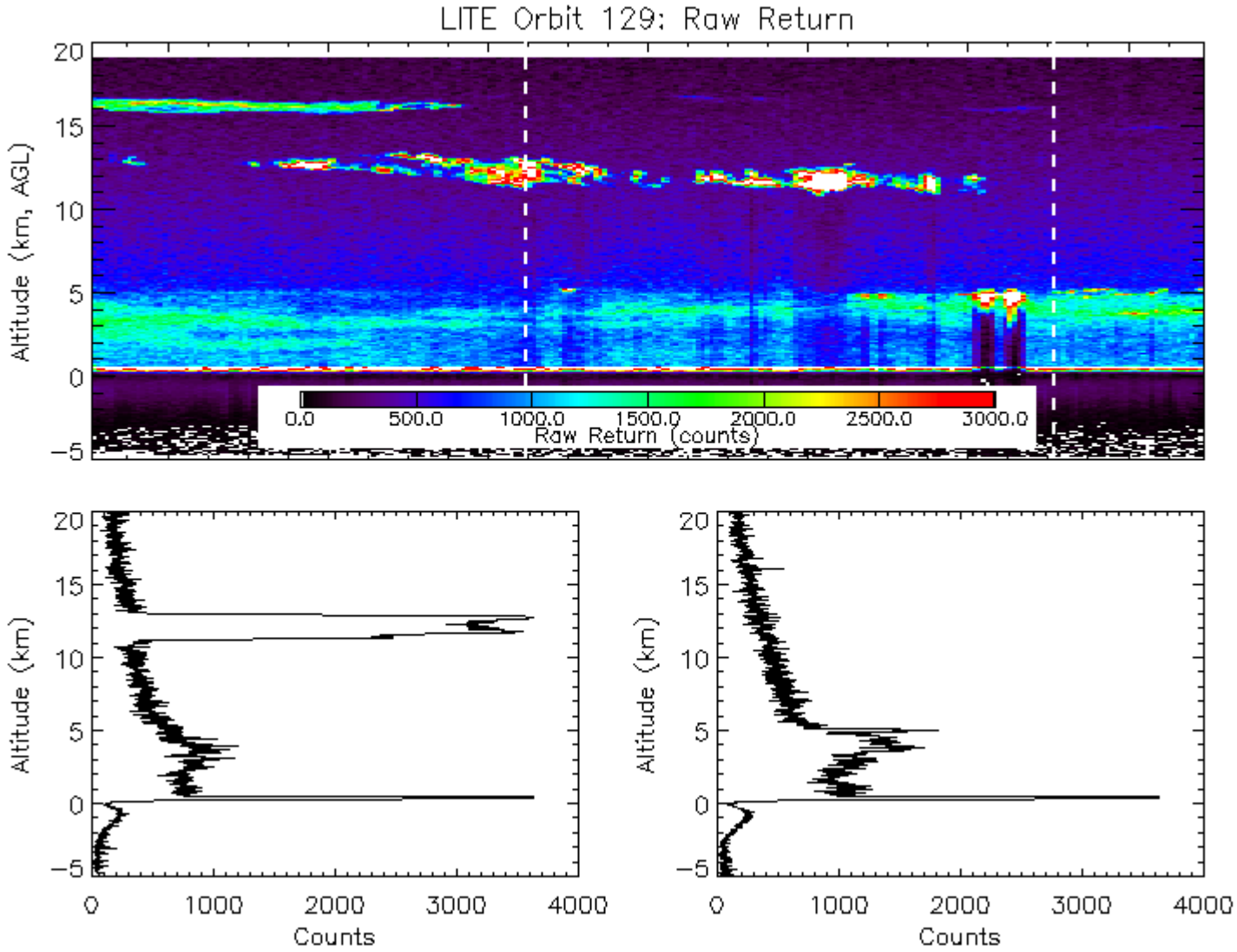
} Parameters b and associated error

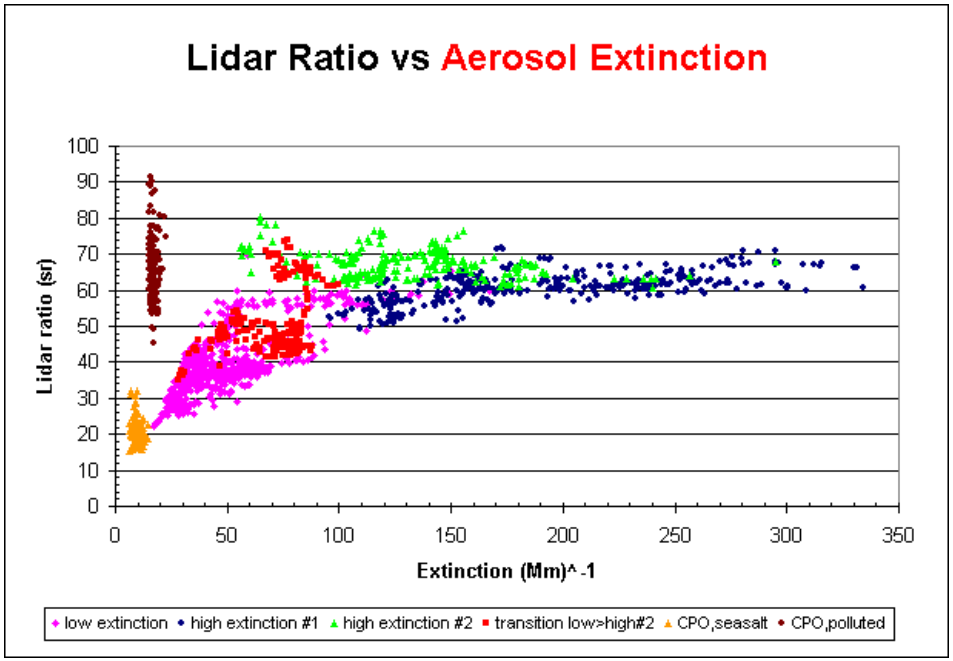
$$y(r) = \ln(C \times P(r))$$

$$f(x) = \ln(\beta_m(r) + kx(r)) - 2\int_0^r \eta(r)x(r')dr'$$

Measurement error

Impacts of other contraints: optical depth (Σx_i)


LITE Orbit 129: Raw Return
Altitude (km, AGL)
20
15
10
5
0
−5
0.0
500.0
1000.0
1500.0
2000.0
2500.0
3000.0
Raw Return (counts)
Altitude (km)
0
1000
2000
3000
4000
Counts



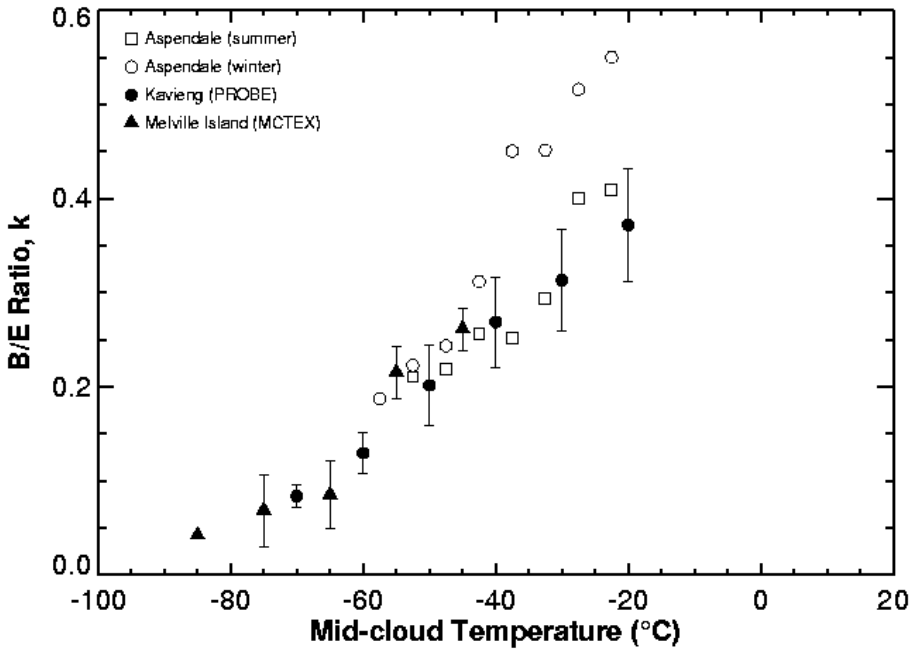
Courtesy, Andersen et al

Backscatter-to-extinction (k)

single value (large uncertainty)

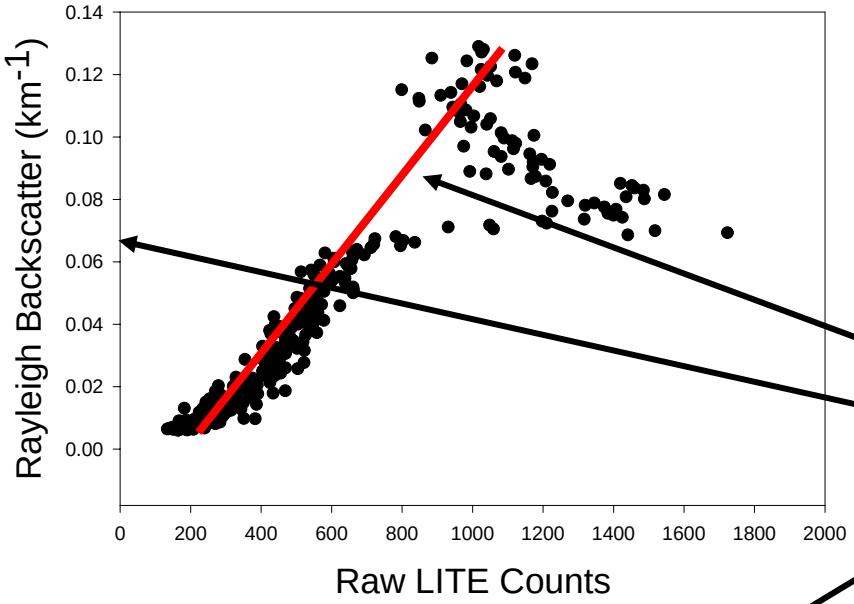
or introduce yet further parameters into the observing system (e.g. T as in the case of cirrus) to reduce uncertainty

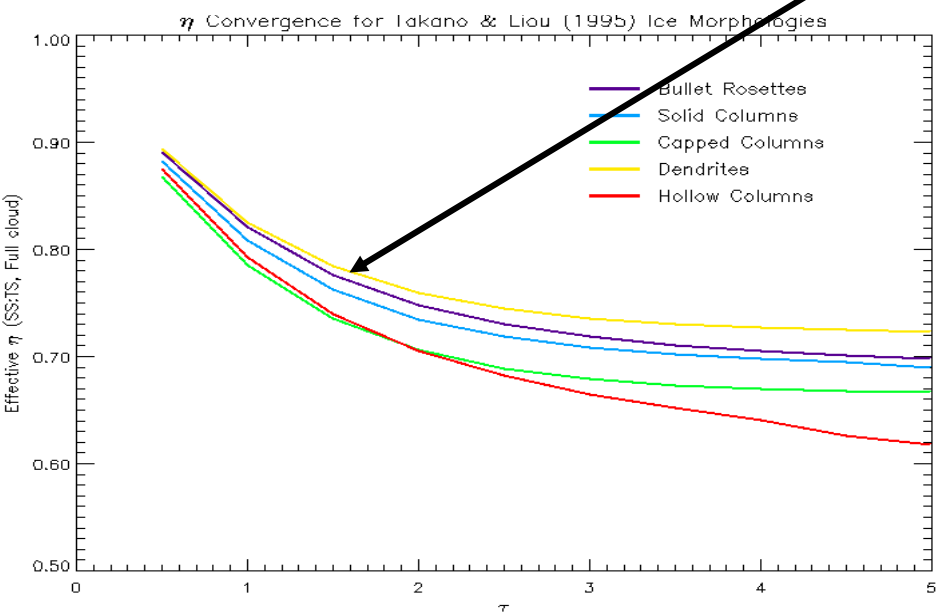
estimate uncertainty σ_k based on $\partial f / \partial k$ and Δk



Courtesy, Platt and Austin

Rayleigh Calibration





Calibration and multiple scattering

estimate uncertainties

σ_y ,

σ_b ,

σ_η

which are determined from Δb and related quantities $\partial f(b) / \partial b$

Illustrative example of retrieval

Onion-peeling optimization method

$$\Phi = (\hat{\mathbf{x}} - \mathbf{x_a})^T \mathbf{S_a}^{-1} (\hat{\mathbf{x}} - \mathbf{x_a}) + (\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}))^T \mathbf{S_y}^{-1} (\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}})) + \mathbf{C}(\tau - \Sigma \hat{\mathbf{x}}_j)$$

$$\mathbf{S_y} \rightarrow \sigma^2{}_y + \sigma^2{}_\eta + \sigma^2{}_k + \sigma^2{}_b$$

$$\mathbf{S_a} \rightarrow \sigma^2{}_a$$

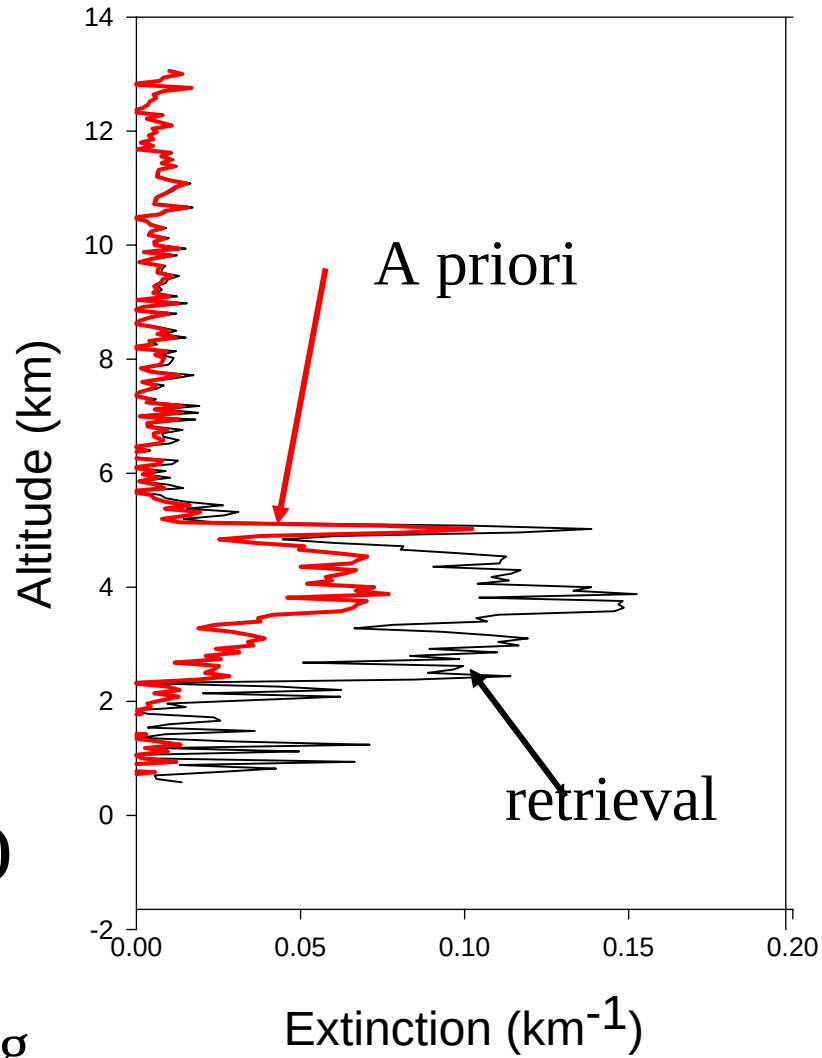
$$\mathbf{C} \rightarrow \sigma^2{}_\tau$$

Solution is iterative

$$\mathbf{x}^{i+1} = \mathbf{S}_X(\mathbf{S_a}^{-1}\mathbf{x_a} + \mathbf{K_i}^T \mathbf{S}_{y,i}^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x_i}) + \mathbf{K_i}\mathbf{x_i}))$$

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$$

plus term involving optical depth constraint (e.g Engelen & Stephens, 1999)



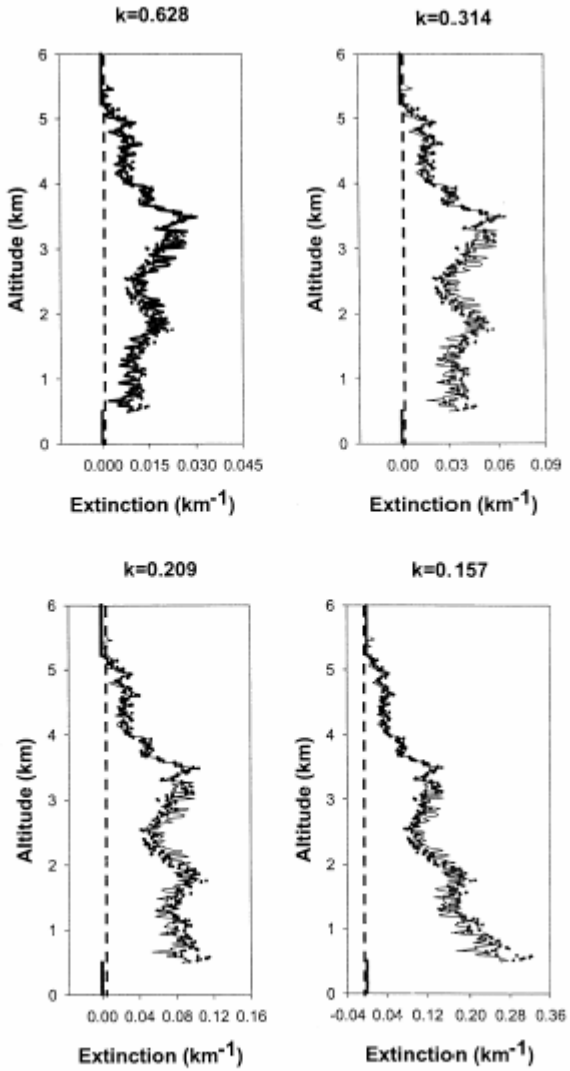


Figure 4. Extinction profiles derived from the backscatter profile located along orbit 129 at the position indicated by the arrow in Plate 1. The retrievals shown correspond to the four values of k noted. The solid line represents the retrieval result obtained using the method introduced in this paper. The profile defined by the solid circles is the equivalent extinction profile derived from an alternative algorithm based on the iterative procedure of *Alvarez and Vaughan* [1994].

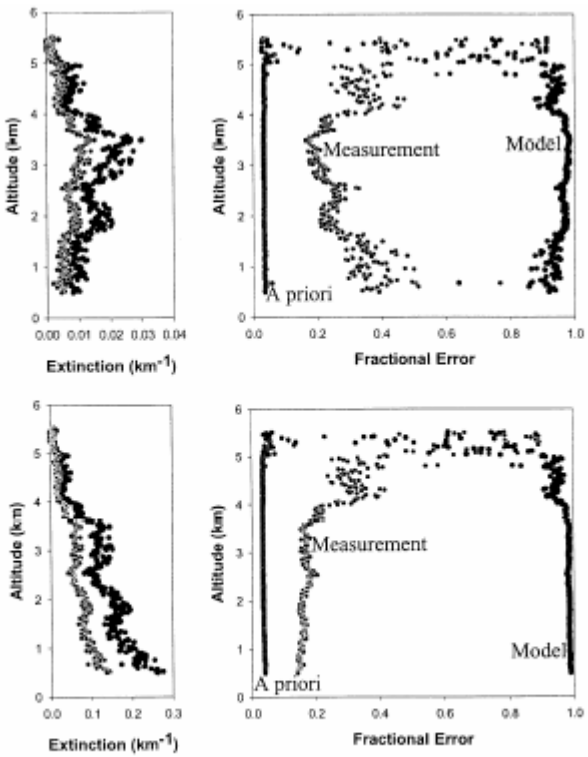


Figure 5. Error characteristics corresponding to the (a) $k = 0.628$ and (b) $k = 0.157$ profiles of Figure 4. The respective extinction profiles and total error is shown in the left-most panel (the grey curve is the total extinction error) and the breakdown of this total error into a priori, measurement and model components.

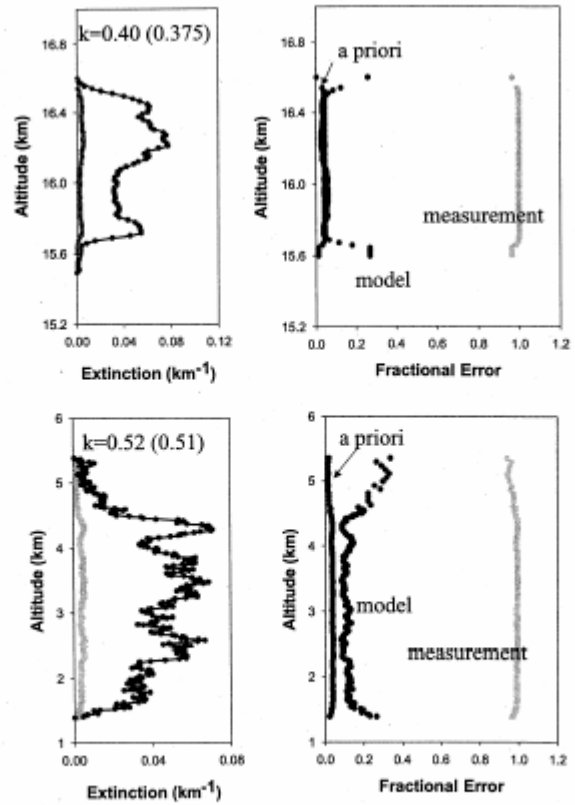


Figure 6. (a) Cirrus cloud extinction profile retrieved for that portion of orbit 129 noted in Plate 1. As for Figure 5, the retrieved profiles derived from the two different methods are shown in the left panel. Also shown is the total error, and the retrieved value of k is quoted in the figure. The right-hand panel presents the errors as for Figure 5a and 5b but for the aerosol layer of orbit 147.

Retrieval with b=k

Retrieval with x_{N+1}=k