AT721 Section 14:

Information Content

General references:
Rodgers, C.D., 2000; Inverse Methods for atmospheric sounding, *World Sci*, chapter 2.
Twomey, 1977;Introduction to the mathematics of inversion in remote sensing and indirect methods, chapter 8.
L'Ecuyer et al, 2004;
Labonnote and Stephens, 2004; A multi-sensor concept for retrieving absorbing aerosol optical parameters, Part I: Information Content Analysis.
Engelen and Stephens, 2004;Information content of infrared satellite sounding measurements with respect to CO2, *J.Appl. Meteorol.*

So far we have assumed that the measurements all provide information and further we have a clear idea about what state information $\mathbf{x}$ can be extracted from these measurements. However, for many modern observing systems that yield, for example, hundreds or thousands of spectral measurements of radiances, it is not always obvious how many of these measurements actually contribute to the retrieved state and how many are really redundant. Furthermore, it is often not even obvious that our predetermined notion of the state vector $\mathbf{x}$ is the optimal choice for a given observing system. These issues underly the topic of this chapter, namely the topic of information content.

14.1 Sensitivity Analysis

14.2 Measurement Redundancy

Another step toward an understanding of the information content inherent in any measurement lies in determining how many measurements add significantly to a retrieval of the state $\mathbf{x}$. This naturally leads to the posing of the following question:
given N measurements $y_i, i \ldots N$ with uncertainty $\epsilon_i, i \ldots N$ will additional information accrue with the additional of more measurements?
As we have seen throughout, the kernel and related measurement error are two quantities instrumental in characterizing the inverse problem and these two quantities are crucial for understanding measurement redundancy.

At first thought, we might suppose that we can get a clue at what measurements are unnecessary by examining, for example, the relationship between y_i and i (which might be thought of as wavelength λ_i).

Consider the following hypothetical data:

i	y_i
$\vdots$	$\vdots$
4.0	20.01
4.5	17.98
5.0	16.02
$\vdots$	$\vdots$

Provided we are only concerned with 3 figure accuracy, y_i is exactly linearly related to the index and there seems no point adding further measurements (at least between $i = 4.0$ and $i = 5.0$ as they can be easily predicted with equivalent accuracy provided our measurements cannot be made better than this 3 figure accuracy. We are however left with a quandry- is this linear relationship merely happen chance such that more measurements either outside or inside the range $4 <\subset< 5$ will actually deviate from the linear relationship implied from the 3 measurements given or is it robust. Suppose that the kernels are linearly related,

$$K(i = 4.5, x) = \frac{1}{2}K(i = 4, x) + \frac{1}{2}K(i = 5, x)$$

and that this relation applies over the entire interval of x then for all x, $y(i = 4.5) = \frac{1}{2}y(i = 4) + \frac{1}{2}y(i = 5)$ and the redundancy of the middle measurement at $i = 4.5$ is not accidental.

14.2.1 Interdependence of Kernels

The search for measurement redundancy can be thought of as looking for at the possibility of writing any kernel as a linear combination of others. By supposition we suppose that for the measurement at lth wavelength

$$K_l = \sum_{j \neq l} a_j K_j(x) \tag{14.1}$$

then

$$y_l = \sum_{j \neq l} a_j y_j(x) + \delta_l \tag{14.2}$$

where δ_l is the error associated with this representation of the measurement and, by supposition, is just the accumulation of all measurement errors

$$\delta_l = \sum_{j \neq l} a_j \epsilon_j(x) \tag{14.3}$$

. If the error of this 'synthetic measurement' is less than the error ϵ_l of an actual measurement then this actual measurement is totally redundant since it can be replace with a synthetic quantity written entirely as a linear combination of other measurements and thus is entirely ineffective in determining the retrieval state $\mathbf{x}$.

It would be useful to know the theoretical conditions under which the kernels are completely linearly dependent kernels as this determines a baseline to establish how many measurements might be independent. The condition

$$\sum_j a_j K_j(x) = 0 \tag{14.4}$$

obviously implies any kernel K_i is linearly related to all other kernels

$$K_i = a_i^{-1} \sum_{j \neq i} a_j K_j(x) \tag{14.5}$$

However, to avoid the trivial condition that

$$\sum_j a_j = 0 \tag{14.6}$$

some constraint must be applied to the a's and we lose no generality in choosing

$$\sum_j a_j^2 = 1 \tag{14.7}$$

which has the effect of placing a bound on the error

$$\mid \delta \mid^2 = \mid \sum_j a_j \epsilon_j(x) \mid^2 \leq \sum_j a_j^2 \sum_j \epsilon_j^2 \tag{14.8}$$

according to the Schwartz inequality (Twomey, 1977). Given (14.7), we obtain an upper bound $\mid \delta \mid^2 \leq \sum_j \epsilon_j^2$ which for independent randomly distributed errors, $\mid \delta \mid^2 \leq N \mid \epsilon \mid^2$.

In principle (14.4) is ideal and errors in the formulation will invariably effects how close the condition (14.4) really applies. For practical cases, we want to make (14.4) a minimum, i.e. determine that set of as such that $\sum_{min} a_j K_j(x)$ given the constraint $\sum_j a_j^2 = 1$. It turns out that the solution to this problem coincides with the smallest eigenvalue of the matrix $\mathbf{K}\mathbf{K}^T$ with the choice of a being the normalized eigenvector associated with this minimum eigenvalue.

Consider the general forward problem

$$y_l + \epsilon_l = \int K_l(x) f(x) dx \tag{14.9}$$

and suppose that

$$K_l = a_l^{-1} \sum_{j \neq l} a_j K_j(x) \tag{14.10}$$

where it is convenient to consider a_l as being the numerical largest of the as which means it must be at least of value $N^{-1/2}$ and this also avoids the problem of choosing a_l which has a zero value. The condition that $\sum_j a_j K_j(x) = 0$ is sufficient and necessary for a measurement y_l to be exactly predictable from other measurements, namely

$$y_i = -a_l^{-1} \sum_{j \neq l} a_j \int K_j(x) f(x) dx = -a_l^{-1} \sum_{j \neq l} a_j y_j \tag{14.11}$$

The vanishing of $\sum_j a_j K_j(x)$ to exactly zero rarely occurs but it often approaches a small value and the minimum eigenvalue approaches a small value and thus the prediction (14.11) has some (small) error

$$K_l = \sum_{j \neq l} (-a_j / a_l) K_j(x) + \delta(x) \tag{14.12}$$

Inserting (14.12) into (14.9) yields the following

$$y_l = \int \sum_{j \neq l} (-a_j/a_l) K_j(x) f(x) dx + \int \delta(x) f(x) dx - \epsilon_l \tag{4.13}$$

In thinking about the errors associated with this y_l, it is first convenient to scale the problem as follows:

$$\alpha y_l + \alpha \epsilon_l = \int \frac{\alpha}{\beta} K_l(x) \beta f(x) dx$$

where *alpha* and β are arbitrary scale factors. We select α so the measurements y_l are order unity so ϵ_l are percentage errors and if the rms relative error is e, then $\mid \epsilon \mid^2 = Ne^2$ We choose β so that $\mid f(x) \mid^2$ is order unity. Given this relative scaling, then we can relate the different components of error by considering the equivalent of (4.13) in scaled form

$$y_l = y_l' + \int \delta(x) f(x) dx - \sum_j (a_j/a_l) \epsilon_j \tag{14.14}$$

The interpretation of each term follows
•a component y_l' that is entirely predictable from other measurements
•a component $\int \delta(x) f(x) dx$ that is not predictable and depends on the unknown $f(x)$. As such this component identifies new information provided by making the measurement.
•The error term: resulting in our experimental determination of y.

If the second term exceeds the first term, then making the measurement offers new information otherwise we can calculate y_l' and obtain an estimate of y_l closer to than we can measure it. That independence implies that

$$\int \delta(x) f(x) dx > \sum_j (a_j/a_l) \epsilon_j$$

and when proper scaling is done, the minimum eigenvalue λ_{min} is the upper bound to the square norm of the left hand side, and $N^{-1} \mid \epsilon \mid^2$ is the upper-bound on the error term (Twomey, p 194) so that the important result emerges:
Provided the system is properly scaled, the independence of N measurements in the presence of a relative error of measurement $\mid \epsilon \mid$ *is assured if* $\lambda_{min} > N^{-1} \mid \epsilon \mid^2$. This is an important result and can be extended more generally owing to the orthogonality of the eigenvectors of KK^T. This general state is that if there are m eigenvalues that are less than $\mid \epsilon \mid^2$, then there are m redundant measurements that can be predicted as well as they can be measured.

14.2.2 A simple Example

Consider the example of Twomey

$$K_\ell(x) = xe^{-\ell x} \tag{14.15}$$

such that the maximum of function occurs at $x = \ell^{-1}$ ranges from 0 and 1. For each N, N kernels are constructed so the maxima uniformly spread between between $x = 0$ and 1, e.g.

$$K(x) = xe^{-x}$$

$$K(x) = xe^{-4x/3}$$
$$K(x) = xe^{-2x}$$
$$K(x) = xe^{-4x}$$

(and see Fig. 14.1) and are normalized to unit area under the curves so that

$$\int_0^1 xe^{-\ell x}dx = \ell^{-2}[1. - e^{-\ell} - \ell e^{-\ell}]$$

Fig. 14.1 Kernels for the 'standard problem' with the maxima distributed through the normalized range of x.

The eigenvalues of the N order covariance matrix

$$C = \| \int_0^1 K_i(x)K_j(x)dx \|$$

were then calculated as shown in Fig 14.2. The eigenvalues are shown as a function of the number of 'measurements' N which are introduced in this example by increasing the number of ℓ values over a fixed interval $0 < x < 1$. For smaller values of N all eigenvalues are plotted but for larger values of N some values are too small to appear in the figure. The number of pieces of useful information is given by the number of eigenvalues that exceed a fixed value that depends on the accuracy of the measurement. The

Fig. 14.2 The eigenvalues of the correlation matrix as a function of the number of measurements

dashed line shows the limit for a 1% measurement error. The figure illustrates that merely increasing the number of observations N does not achieve much increase in information.

Introduce and discuss AIRS water vapor results

14.3 Shannon's information content

Sensitivity analysis like that described above provides insight into what state variables most influence the measurements and thus are a measure of information. However, the ultimate utility of the observations, also depends on other factors such as instrumental characteristics that include signal-to-noise ratios and calibration accuracy. With this in mind, we now consider more specific and quantitative measures of information that take these additional factors into consideration. The approach to information theory used here is that according to Shannon and Weaver (1949) as applied to satellite radiance measurements by Rodgers (2000).

We will introduce three measures of information, each defined relative to a given state of prior knowledge: (i) the number of independent *measurements* made to better than measurement error - this is referred to as the degrees of freedom for signal; (ii) the singular values of a scaled weighting function matrix that tells us how many pieces of information about the *physical* system can be extracted over and above the inherent noise level of the system; and (iii) the Shannon information content of each independent measurement and the total Shannon information content which broadly characterizes the resolution of the observing system.

Fig. 14.3 The AIRS example.

14.2.1 Theory

The information content of a set of observations can be defined by the change in the logarithm (base 2) of the number of distinct possible states of the system being measured. First we define the possible states x of a system by a probability distribution function $P(x)$. In both the statistical and thermodynamical context, entropy is the logarithm of the number of distinct internal states of a system consistent with the measured (macro-)state and therefore is a measure of the information of that system. The entropy $S(P)$ of the system follows as

$$S(P) = -k \int P(x) \ln P(x) dx \tag{14.15}$$

If we assume that Gaussian statistics represent the probability distributions function of the (vector) state $\mathbf{x}$ of dimension n, namely

$$P(\mathbf{x}) = \frac{\mathbf{1}}{(\mathbf{2}\pi^{\frac{\mathbf{n}}{\mathbf{2}}} |\mathbf{S_x}|^{\frac{\mathbf{1}}{\mathbf{2}}}} \exp\{-\frac{\mathbf{1}}{\mathbf{2}}(\mathbf{x} - \overline{\mathbf{x}})^{\mathbf{T}} \mathbf{S_x^{-1}} (\mathbf{x} - \overline{\mathbf{x}})\} \tag{14.16}$$

$\mathbf{S_x}$ is the covariance matrix that characterizes our knowledge of the system relative to the true state $\overline{\mathbf{x}}$, then the entropy follows as

$$S(P) = \frac{1}{2} \ln |\mathbf{S_x}| \tag{14.17}$$

Thus the entropy is the logarithm of the volume of state space occupied by the probability density function defined by the covariance matrix.

In the context of an observing system, if $P_1(x)$ describes our knowledge of the system before a measurement is made (the prior distribution and characterized by the *a priori* error covariance matrix $\mathbf{S_a}$), and $P_2(x)$ describes knowledge after the measurement (the *posteriori* distribution as represented by the retrieval error covariance matrix $\mathbf{S_x}$), then the information content provided by the measurements may be formally defined as the reduction of entropy:

$$H = S(P_1) - S(P_2) \tag{14.18}$$

which becomes

$$H = \frac{1}{2} \ln |\mathbf{S_a S_x^{-1}}| \tag{14.19}$$

In this way, we interpret H as a reduction in volume of the prior probability function after making the observations. If we consider the 'atmospheric state' as defined by the retrieval vector $\mathbf{x}$ then of all the possible atmospheric states within the space defined by the *a priori* covariance matrix a number of 2^H states can actually be distinguished by the observations. Thus H provides some measure of the resolving power of the observing system. For example, and in the context of the present study, $H = 1$ implies a very coarse observing system that is only able to identify two different states over and above our prior knowledge - that is it can only differentiate (say) between a state that is optically thinner or thicker than the a priori state.

Another useful measure of information is the degrees of freedom for signal. While the total degrees of freedom of a set of observations is equal to the number of observations, only a selection of these total degrees of freedom is independent and significant with respect to the measurement noise. The degrees of freedom for signal are therefore defined as the number of independent pieces of information in a set of measurements that can be observed above the noise of the observations. This does not equate to a practical measure of how many parameters might be extracted from the measurements, it merely tells us about the number of useful measurements. Using the above covariance definitions, we can write the degrees of freedom for signal, d, as the trace of the same matrix product as is used in the definition of the Shannon information content namely:

$$d = Tr(\mathbf{I_n} - \mathbf{S_a S_x^{-1}}) \tag{14.20}$$

where $\mathbf{I_n}$ is a n×n identity matrix.

The third piece of information about the observing system can be conveniently introduced considering a general observing system expressed in a linear in form,

$$\mathbf{y} = \mathbf{K}\mathbf{x} \tag{14.21}$$

where $\mathbf{K}$ is the weighting function matrix constructed from the sensitivities of the kid discussed in the previous sections. Given a constraint in the form of an a priori covariance matrix $\mathbf{S_a}$, we can write the retrieval error covariance matrix as (see Rodgers, 2000; also Part II)

$$\mathbf{S_x} = (\mathbf{S_a^{-1}} + \mathbf{K^T S_y^{-1} K})^{-1} \tag{14.22}$$

where $\mathbf{S_y}$ is the measurement plus model error covariance matrix. Here it is useful to introduce the scaled weighting function matrix

$$\tilde{\mathbf{K}} = \mathbf{S_y^{-1/2} K S_a^{1/2}} \tag{14.23}$$

such that the singular values λ_i of this matrix are a direct measure of the signal-to-noise ratio taking into account our prior knowledge of the atmospheric state. Thus, singular vectors with singular values exceeding

unity contain usable information about the atmospheric state, while singular vectors with a singular values smaller than unity are more likely dominated by the measurement noise. Therefore the number of singular vectors with associated singular values exceeding unity equates to the number of different parameters we might usefully extract from the measurements to define the state $\mathbf{x}$. If we find there are four pieces of information about the state that can be so determined, this may or may not mean that our predetermined set of 4 parameters, such as τ, ϖ_o, m_r and r_e, or combinations of them, optimally define the state. Closer study of the singular vectors gives some sense for which combination of parameters best represents the information in the measurements.

Rodgers (2000) shows that the singular values can be used to calculate the degrees of freedom for signal

$$d = \sum_i \frac{\lambda_i^2}{(1+\lambda_i^2)} \tag{14.24}$$

and the information content

$$H = \frac{1}{2}\sum_i \ln(1+\lambda_i^2) \tag{14.25}$$

Therefore, by calculating the singular values of $\tilde{\mathbf{K}}$ we are able to determine the two other pieces of information. We now consider these measures of information first in relation to a simple model of sunlight reflected by aerosol and then in application to the A-Train observing system.

14.3 Example 1: The CO2 retrieval

A simple example can show the use of the above-defined measures of information in an observation. Assume we want to retrieve the true value of the CO_2 concentration (x) from a direct flask observation (y) given some a priori guess of x_a. If the observation includes an error ε , we have the following relation between x and y:

$$y = x + \varepsilon$$

and therefore $K = 1$. We can then calculate the retrieval error from (14.22) given the measurement error and the a priori error from

$$\sigma_x^2 = [\sigma_a^2 + \sigma_y^2]^{-1}$$

where standard deviations are used to represent the errors.

The information content and degrees of freedom for signal can then be calculated from (x) and (y), respectively. Table 14.1 shows the results for 2 cases: (i) a small measurement error case (measurement error is 0.25 ppmv and a priori error is 4 ppmv), and (ii) a large measurement error case (measurement error is 4 ppmv and a priori error is 0.25 ppmv).

Although the retrieval error σ_x is the same for both retrievals, the degrees of freedom and the information content are very different. In the case with small measurement error, the degrees of freedom for signal is almost 1, while in the case with large measurement error the degrees of freedom for signal is almost zero. As expected, the information content of the low noise measurement is much larger than the information content of the high noise measurement. The low observational noise retrieval can distinguish $2^4 = 16$ values within the a priori variance of 4 ppmv, which in this scalar case is equal to the signal-to-noise ratio defined by σ_a/σ_a. In other words, while the retrieval error does not distinguish between the low and high noise case, the degrees of freedom and the information content differentiate between the two cases and identify the better measurement system.

14.3.1 Application to HIRS and AIRS observing systems

Engelen and Stephens (2004) calculated the information content with respect to atmospheric CO2 for observations by TOVS/HIRS and AIRS observing systems. Models of the measurements (brightness temperatures) of HIRS and AIRS were developed (Engelen et al. 2001). For AIRS, a spectral resolution of 1 cm-1 for the band between 500 cm-1 and 2500 cm-1 was assumed, whereas HIRS spectral channels 1 - 7 and 15 - 17 (e.g. Smith et al., 1979) were modelled with a half width of the instrumental response functions, which is about 15 cm-1 for the long-wave channels and about 25 cm-1 for the shorter-wave channels. The measurement covariance matrix for both instruments was specified as a diagonal matrix with standard deviations of 0.5 K on the diagonal elements. This error includes uncertainties in the temperature profile, which acts in this simple set-up as an input for the radiative transfer model. These temperatures could come from the Advanced Microwave Sounding Unit (AMSU) or a weather forecast model. The above assumptions for the measurement covariance matrix are optimistic, especially because errors in the assumed temperature profiles will introduce correlations. Also, the value of 0.5 K is rather small. Therefore, we will also use a value of 1.0 K in one of the experiments. The a priori covariance matrix has, in our first example, diagonal elements of 16 ppmv^2 (i.e. a standard deviation of 4 ppmv) and off-diagonal elements specified as follows:

$$S_{ij} = \sigma_a^2 \exp(-|z_i - z_j|/H)$$

where the linear scale height H is set to 25 km and where the minimum vertical correlation is set to 0.5. The lowest 2 km, which represent the boundary layer, was decoupled from the rest of the atmosphere by setting the correlations to zero. This covariance matrix setup, including the uncertainty estimate of 4 ppmv, was based on hourly output of CO2 profiles from a GCM simulation. Although the uncertainty at individual levels is 4 ppmv, the uncertainty in the column averaged mixing ratio is 3.1 ppmv (using equation (12)) due to the correlations between the levels.

Figure 14.x shows the singular vectors and their corresponding singular values for both HIRS and AIRS with an a priori standard deviation of 4 ppmv. Only singular values that are larger than 1 are significant with respect to the measurement error. HIRS has no significant vectors, while AIRS has two significant vectors. The first two AIRS singular vectors represent broad vertical patterns without much vertical resolution. The third AIRS singular vector adds some vertical resolution, but is not significant. The degrees of freedom and the Shannon information content for this setup are shown in Table 2. The total degrees of freedom for HIRS is only 0.22. The total Shannon information content is 0.18, which means that only different atmospheric states can be detected. This means that within the a priori uncertainty of 4 ppmv less than two different atmospheric states can be distinguished. For AIRS, however, the total degrees of freedom is 1.6 and the total Shannon information content is 2.3, which translates into five distinguishable atmospheric states within the a priori uncertainty. The retrieval error of the column averaged mixing ratio is 2.8 ppmv for HIRS and 1.2 ppmv for AIRS. Clearly, while there is some CO2 signal in the HIRS radiances, it is not enough to observe atmospheric CO2 concentrations better than 4 ppmv. On the other hand, radiances observed by AIRS will be capable of providing significant atmospheric CO2 information, especially total column values as is shown by the first two singular vectors. What is also quite interesting to note is that the singular vectors for HIRS and AIRS are very similar. Apparently, there is not a great difference in what structures both instruments can observe; the difference is in the signal to noise ratio reflected by the information content. If we increase our a priori uncertainty to 10 ppmv (with a column averaged uncertainty of 7.6 ppmv), which is close to the seasonal amplitude of atmospheric CO2 concentrations, the HIRS radiances have a clearer signal, as shown in Table 3. The singular vectors are the same as in Figure 1, but the singular values, and therefore the degrees of freedom, and the Shannon information content have changed. HIRS now has one significant singular vector, and the

degrees of freedom have increased to almost 1. The total Shannon information content is now 0.78, which represents almost two different atmospheric states. This means that HIRS is able to estimate a column averaged CO2 concentration as represented by the first singular vector when our priori knowledge is on the order of 10 ppmv. AIRS still has two significant singular vectors. The total amount of atmospheric states that can be detected by AIRS has increased to 32. The column averaged retrieval errors are now 5.1 ppmv for HIRS and 2.2 ppmv for AIRS. This shows that AIRS can significantly improve over the a priori estimate, while HIRS does not reach an uncertainty that would be small enough to have a significant effect in CO2 inversions. Our estimate of the effect of errors in the assumed temperature profile on the forward radiative transfer model is quite conservative. We assumed an error of 0.5 K, but it could easily be as large as 1 - 2 K. Table 4 shows the retrieval statistics for an a priori error of 4 ppmv and a total measurement and forward model error of 1 K for both HIRS and AIRS. As before, HIRS does not have any significant singular values and its degrees of freedom dropped to 0.06 with 1.0 distinguishable atmospheric state. The retrieval error of the column-averaged volume mixing ratio is 3.0 ppmv. The AIRS retrievals also degrade, but there is still one significant singular value. The total degrees of freedom is now 0.95 and the number of distinguishable atmospheric states is 2.1. The retrieval error of the column-averaged volume mixing ratio is 1.8 ppmv. For the case with a 10 ppmv a priori error specification, similar degradation results are obtained. Although our specification of the a priori covariance matrix is an estimate based on model output, we now show the role of well-specified vertical correlations in the covariance matrix. Figure 2 and Table 5 show the results of the first experiment, but now with a diagonal a priori covariance matrix that contains no vertical correlations at all. Because the levels are completely uncorrelated, errors at different levels start to compensate when we calculate the column-averaged uncertainty. For the diagonal a priori covariance matrix here specified the column-averaged uncertainty is 0.98 ppmv. Both HIRS and AIRS show singular vectors with more vertical structure, but the singular values have decreased significantly. HIRS does not have any significant singular vectors at all, while the number of significant singular vectors for AIRS is also reduced to zero. The number of distinguishable atmospheric states is now 1.0 for HIRS and 1.4 for AIRS. The column-averaged uncertainty for HIRS is 0.97 ppmv, which is basically equal to the a priori uncertainty. The column-averaged uncertainty for AIRS is 0.87 ppmv. So, although the non-diagonal a priori covariance matrix seems to constrain the retrieval more than the diagonal a priori covariance matrix, the amount of information that can be retrieved from the observations is actually higher for the non-diagonal matrix. The reason for this is that the weighting functions see only large-scale structure. The non-diagonal covariance matrix constrains the small-scale structure, but has a larger variance at the larger scale, therefore allowing retrieving more information about the large-scale structure than the pure diagonal covariance matrix. All analyses in this section have been carried out for individual profiles. However, to reduce the retrieval error spatial and temporal averaging could be applied rendering CO2 distributions on spatial and temporal scales useful for current inversion studies. Most recent CO2 inversion studies have used monthly mean observations and a transport model grid of the order of 5 - 10 (e.g., Gurney et al. 2002; Kaminski et al. 2002; Rdenbeck et al. 2003). Most areas with significant cloudiness would allow the averaging of at least several satellite observations on these space and time scales. However, although averaging will reduce the random component of the observation error, any systematic errors in the retrieved values will remain in the averaged product. These systematic errors arise from biases in the a priori estimates, biases in the radiative transfer modeling, and biases in the temperature field. Furthermore, most of these biases are spatially heterogeneous, which will create errors in the horizontal gradients of the averaged CO2 fields. Therefore, a proper characterization of especially these systematic errors is crucial for a correct interpretation of the results.

14.4 Example 2: Aerosol retrieval

For this problem, we seek to retrieve aerosol optical depth τ from a direct observation of reflected sunlight given some *a priori* guess of τ_a. A simple model of this reflectance, and one that provides the basis of a number of different operational aerosol algorithms, is

$$I = I_{surf} + I_{atm} \tag{14.26}$$

where I_{surf} is the contribution by surface reflection α_{sfc} according to

$$I_{surf} = \alpha_{sfc}\frac{F_\odot}{\pi}\mu_\odot(1 - m\tau) \tag{14.27}$$

and I_{atm} is the atmospheric scattering term given by

$$I_{atm} = \frac{F_\odot}{4\pi\mu}\varpi_o P(\mu_\odot, \mu)\tau \tag{14.28}$$

which implies only single scatter and that the aerosol optical depth $\tau < 1$. $P(\mu_\odot, \mu)$ is the scattering phase function. The airmass factor $m = \mu^{-1} + \mu_\odot^{-1}$ and $F_\odot$ is the (normal) solar flux incident on the aerosol. When expressed in terms of reflectance $R = \pi I/\mu_\odot F_\odot$, (21) follows as

$$R = \alpha_{sfc} + (b - m\alpha_{sfc})\tau + \varepsilon \tag{14.29}$$

where

$$b = \frac{1}{4\mu\mu_\odot}\varpi_0 P(\mu_\odot, \mu) \tag{14.30}$$

and where we now include an observation error ε. Assuming this error is Gaussian with a variance σ_ε, and further assuming errors in the model parameter b and α_{sfc} defined by a variance σ_b and an a priori error σ_a, then the retrieval error follows

$$\sigma_x^2 = (\sigma_a^{-2} + K^2\sigma_y^{-2})^{-1} \tag{14.31}$$

where K is the weighting function parameter

$$K = \frac{dR}{d\tau} = (b - m\alpha_{sfc}) \tag{14.32}$$

and σ_y^2 contains both the forward model plus measurement errors

$$\sigma_y^2 = \sigma_\varepsilon^2 + \sigma_b^2 \tag{14.33}$$

The information content and degrees of freedom for signal can then be calculated from (14) and (15), respectively given $S_a = \sigma_a^2$ and $S_x = \sigma_x^2$.

Table 14.1 presents results of this analysis applied to this simple model. The parameters used to create these results are as follows: the phase function is defined using the parameters $r_e = 0.2\mu m$, $v_e = 0.4$, and $m_r = 1.4$, and the other parameters are $\tau = 0.2$, $\alpha_{sfc} = 0.1$ and 0.3 and $\mu = 1$ and $\mu_\odot = 0.866$. Results are presented for several cases: (i) a small measurement error case (measurement error is 5%, $\sigma_\varepsilon = 0.05 \times R_T$), a large a priori error (100% error, $\sigma_a = 1 \times \tau_a$) and no forward model error ($\sigma_b = 0$). (ii) A large measurement error case (measurement error is 100%, $\sigma_\varepsilon = 1 \times R$), a small a priori error (5%, $\sigma_a = 0.05 \times \tau_a$), and no forward model error ($\sigma_b = 0$). (iii) same as case (i) but we introduce a forward

model error ($\sigma_b \neq 0$). (iv) same as case (i) but we consider ten measurements corresponding to 10 different geometries in the solar plane ($\theta = 80^o$ to -10^o in steps of 10^o).

The degrees of freedom and information content for both cases (i) and (ii) are very different as expected. For the case of small measurement error, the degrees of freedom is almost 1, whereas in the case of large measurement error the degree of freedom is almost zero. As expected the information content is highest when the measurement error is least. Case (i) distinguishes $2^{3.8} \simeq 14$ values of τ within the a-priori variance of 0.3, which in this scalar case is equal to the signal-to-noise ratio $R(\sigma_a)/\sigma_y$, whereas case (ii) distinguishes only one value being the a-priori value. The retrieval error is smaller for case (ii) than case (i), serving to demonstrate that this error does not distinguish between the low and large noise cases and thus does not necessarily provide a measure of the best observing system. The degrees of freedom and information content together provide a means of differentiating which is the better observing system.

Table 14.1 also shows the effect of adding forward model error on the information content analysis (case (iii)). For this example we considered a more realistic 20% error on both surface albedo and phase function, although it could be argued that these phase function errors are too optimistic. As expected the information content and degrees of freedom decrease and the retrieval system is now degraded. The observing system now distinguishes only $2^{2.6} \simeq 6$ states. Also noteworthy is the increase of the information content as the number of observations from the multi-viewing instruments increases (case (iv)). For a system that provides measurements at ten different viewing geometries, the number of distinguishable states is $2^{5.5} \simeq 45$, highlighting the obvious advantage of multi-viewing measurements, like MISR or POLDER, over methods that retrieve using only a single view.

The analysis of cases (i) and (ii) was repeated and the results are shown as a function of varying α_{sfc} in Fig. 9 This figure presents an unexpected result. For small surface albedo, H decreases as α_{sfc} increases and then increases as α_{sfc} increases beyond a value of 0.05. The decrease in H reflects the fact that for dark surfaces, the information about the aerosol optical depth derives from the atmospheric scattering term. As α_{sfc} increases from zero, this source of information is systematically reduced being replaced by the surface term that adds information via the process of two-way attenuation.

14.5 Optimizing an observing system

14.5.1 Atmospheric soundings from spectrometer measurements

14.5.2 Cloud property retrievals

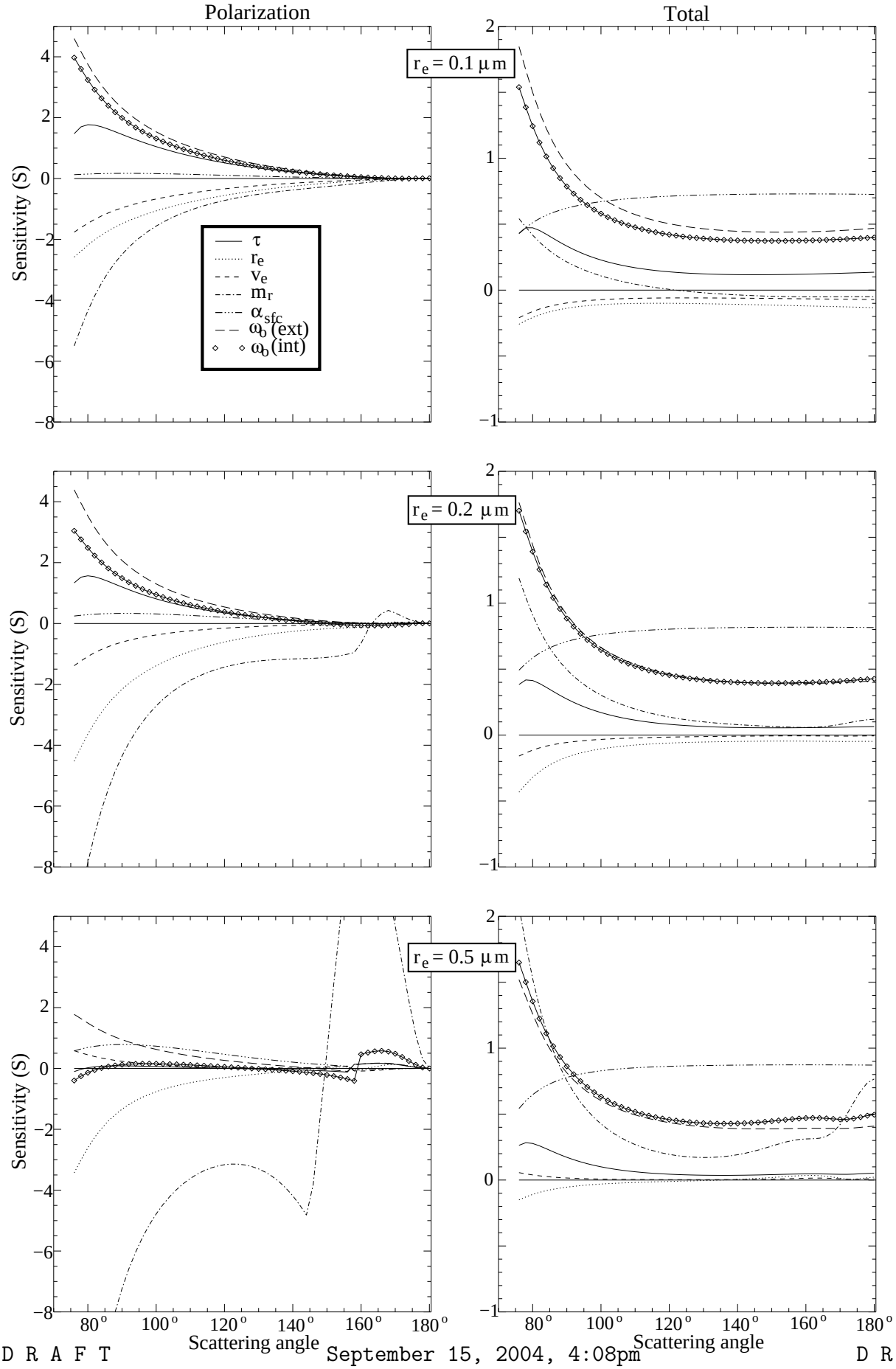
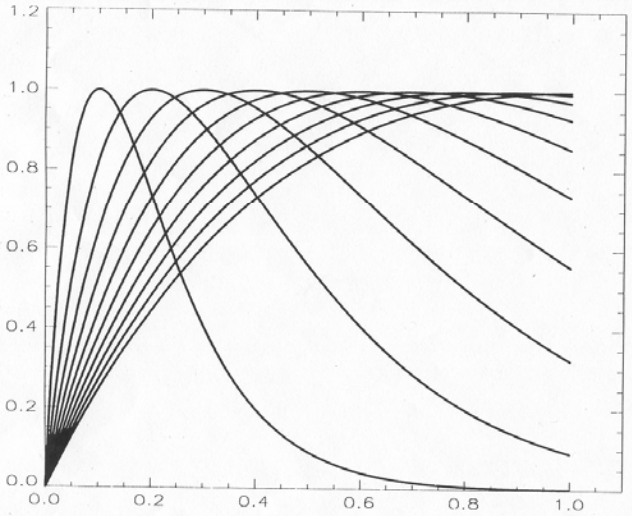


Figure 6. Sensitivity of the polarized reflectance (left panel) and the total reflectance (right panel), to aerosol microphysical and optical parameters, as a function of the scattering angle. The sensitivities correspond to an aerosol layer composed of small particles ($r_e = 0.1\mu m$, top

Twomey's Example (pp 196-197)



i	λ_i
1	6.28
2	0.04
3	4.2e-5
4	1.8e-8
5	3.7e-12
6	< 1.0e-14
7	< 1.0e-14
8	< 1.0e-14
9	< 1.0e-14
10	< 1.0e-14

Fig. 14.1 (TBD)

Number of useful weighting functions for Twomey's example

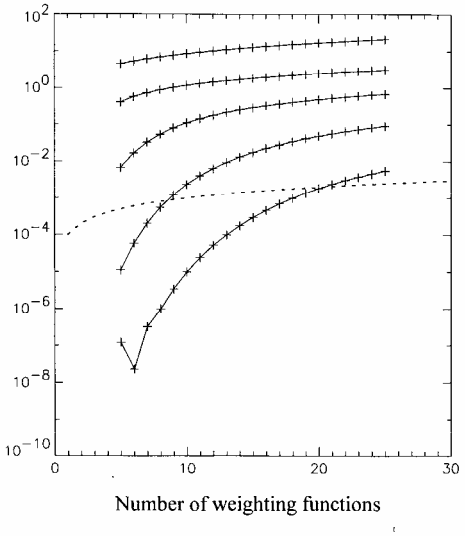


Fig. 14.2

Number of useful weighting functions for the AIRS instrument

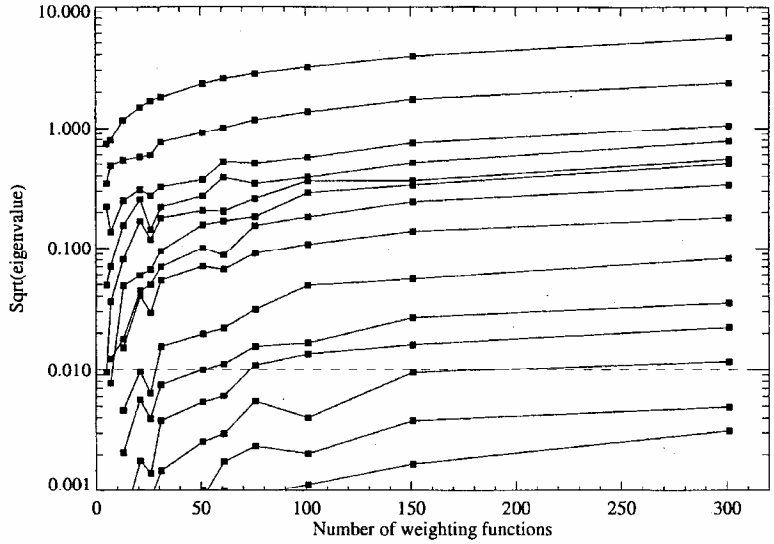
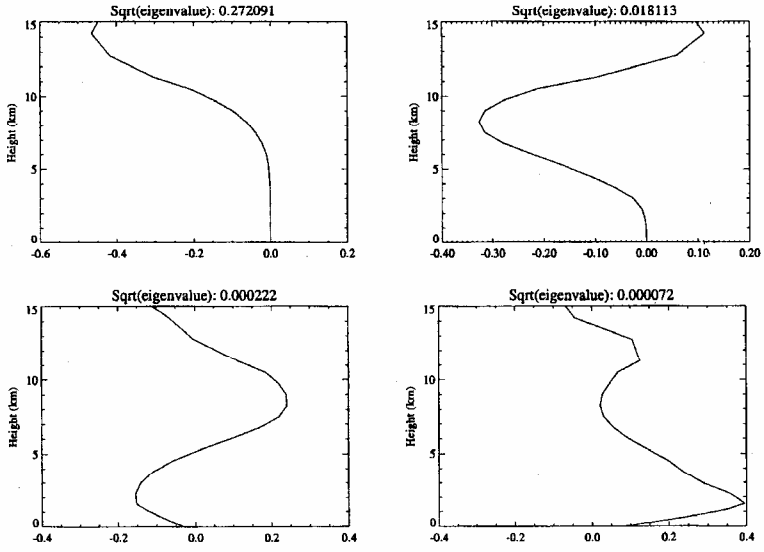


Fig. 14.3

TOVS water vapor retrieval



AIRS water vapor retrieval

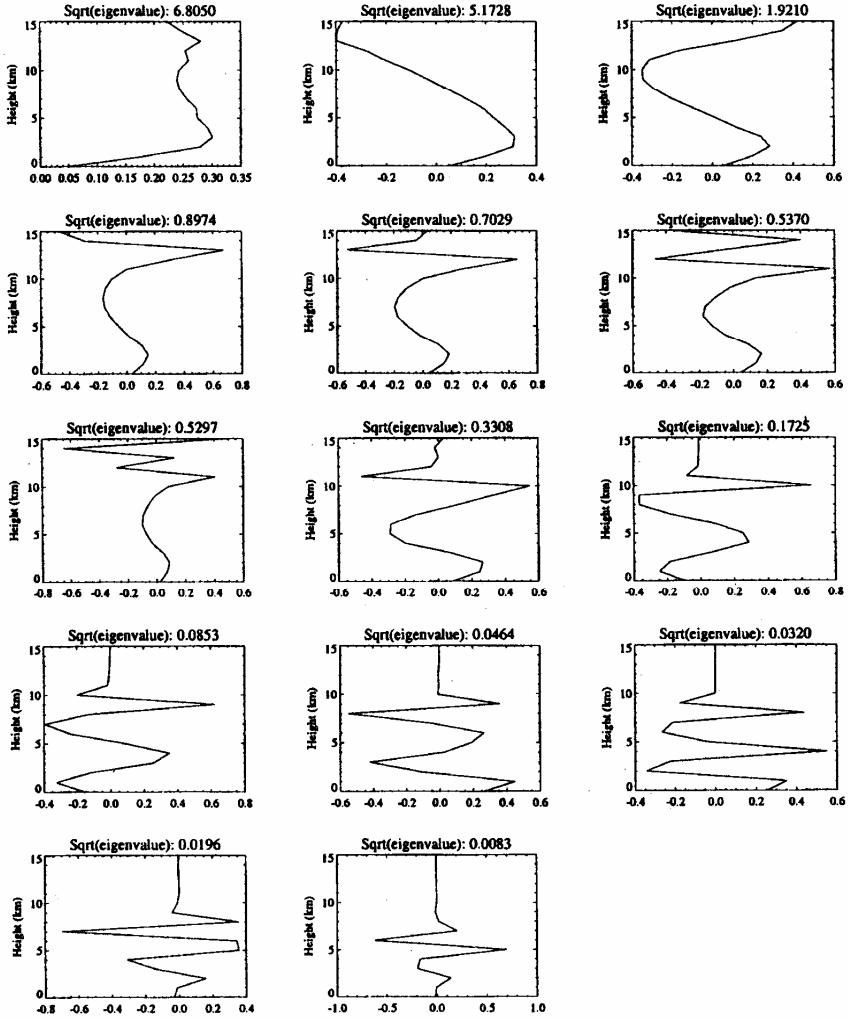


Table 1: Retrieval error, degrees of freedom for signal, and information content for a simple measurement.

	Small measurement error	Large measurement error
σ_y (ppmv)	0.25	4.0
σ_a (ppmv)	4.0	0.25
σ_x (ppmv)	0.235	0.235
d_s	0.94	0.06
H	4.09	0.089

Table 2: Singular values and the contribution of each singular vector to the degrees of freedom and information content for HIRS and AIRS with a priori CO_2 errors of 4 ppmv and observation errors of 0.5 K. The total degrees of freedom and the total information content are shown in the bottom row.

	HIRS			**AIRS**		
i	**λ_i**	**d_s**	**H**	**λ_i**	**d_s**	**H**
1	0.4937	0.1960	0.1573	3.0561	0.9033	1.6850
2	0.1573	0.0242	0.0176	1.0212	0.5105	0.5153
3	0.0386	0.0015	0.0011	0.2643	0.0653	0.0487
4	0.0206	0.0004	0.0003	0.2643	0.0653	0.0487
Total		0.2220	0.1763		1.5677	2.3157

Table 3: Singular values and the contribution of each singular vector to the degrees of freedom and information content for HIRS and AIRS with a priori CO_2 errors of 10 ppmv and observation errors of 0.5 K. The total degrees of freedom and the total information content are shown in the bottom row.

	HIRS			AIRS		
i	**λ_i**	**d_s**	**H**	**λ_i**	**d_s**	**H**
1	1.2342	0.6037	0.6676	7.6401	0.9832	2.9459
2	0.3933	0.1339	0.1037	2.5531	0.8670	1.4552
3	0.0964	0.0092	0.0067	0.7529	0.3618	0.3240
4	0.0516	0.0027	0.0019	0.6606	0.3038	0.2612
Total		0.7496	0.7800		2.5496	5.0109

Table 4: Singular values and the contribution of each singular vector to the degrees of freedom and information content for HIRS and AIRS with a priori CO_2 errors of 4 ppmv and a observation error of 1K. The total degrees of freedom and the total information content are shown in the bottom row.

	HIRS			AIRS		
i	**λ_i**	**d_s**	**H**	**λ_i**	**d_s**	**H**
1	0.2468	0.0574	0.0427	1.5280	0.7001	0.8688
2	0.0787	0.0062	0.0045	0.5106	0.2068	0.1671
3	0.0193	0.0004	0.0003	0.1506	0.0222	0.0162
4	0.0103	0.0001	0.0001	0.1321	0.0172	0.0125
Total		0.0641	0.0408		0.9477	1.0657

Table 5: Singular values and the contribution of each singular vector to the degrees of freedom and information content for HIRS and AIRS with a priori CO_2 errors of 4 ppmv and observation error of 0.5 K, but for a diagonal a priori covariance matrix without any vertical correlations. The total degrees of freedom and the total information content are shown in the bottom row.

	HIRS			AIRS		
i	**λ_i**	**d_s**	**H**	**λ_i**	**d_s**	**H**
1	0.1420	0.0198	0.0144	0.8497	0.4193	0.3921
2	0.0529	0.0028	0.0020	0.3360	0.1014	0.0772
3	0.0286	0.0008	0.0006	0.2260	0.0486	0.0359
4	0.0127	0.0002	0.0001	0.1425	0.0199	0.0145
Total		0.0236	0.0171		0.5983	0.5262

Figure 1. The constellation of satellites forming the Aqua-train.

Table 1. Microphysical specification of aerosol models used in our simulation

model	1	2	3	4	5	6	7	8	9	10	11	12
r_e (μm)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2
v_e	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
m_r	1.4	1.4	1.4	1.4	1.55	1.55	1.55	1.55	1.4	1.4	1.4	1.4
m_i	0.02	0.007	0.0	0.0	0.02	0.007	0.0	0.0	0.02	0.007	0.0	0.0
ϖ_0	0.773	0.908	0.85	0.95	0.856	0.944	0.85	0.95	0.865	0.949	0.85	0.95
k	0.282	0.335	0.316	0.353	0.305	0.345	0.315	0.352	0.147	0.176	0.167	0.187
S (sr)	44.6	37.5	39.8	35.6	41.2	36.4	39.9	35.7	85.5	71.4	75.2	67.2
model	13	14	15	16	17	18	19	20	21	22	23	24
r_e (μm)	0.2	0.2	0.2	0.2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
v_e	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
m_r	1.55	1.55	1.55	1.55	1.4	1.4	1.4	1.4	1.55	1.55	1.55	1.55
m_i	0.02	0.007	0.0	0.0	0.02	0.007	0.0	0.0	0.02	0.007	0.0	0.0
ϖ_0	0.90	0.962	0.85	0.95	0.873	0.95	0.85	0.95	0.876	0.95	0.85	0.95
k	0.198	0.248	0.242	0.271	0.102	0.149	0.163	0.182	0.341	0.539	0.607	0.678
S (sr)	63.5	50.7	51.9	46.4	123.2	84.3	77.1	69.0	36.8	23.3	20.7	18.5

Table 2. Retrieval error, degrees of freedom for signal and information content for a simple measurement case

	Small measurement error				**Large measurement error**	
σ_a	100%				5%	
σ_ε	5%				100%	
α_{sfc}	0.1			0.3	0.1	0.3
Number of geometry	1		10	1	1	1
σ_b	0%	20%	0%	0%	0%	0%
H	3.8	2.6	5.5	4.25	0.0003	0.0006
d	0.995	0.975	0.999	0.997	0.0004	0.0008
σ_τ	0.021	0.05	0.007	0.016	0.015	0.015

Table 3. Singular values of $\tilde{\mathbf{K}}$, together with contribution of each vector to the degrees of freedom and information content for the case of a more elaborate multi-viewing polarimeter.

	$r_e = 0.1\mu m$			$r_e = 0.2\mu m$			$r_e = 0.5\mu m$		
i	λ_i	d_i	H_i (bits)	λ_i	d_i	H_i (bits)	λ_i	d_i	H_i (bits)
1	240.1	0.999	7.91	194.7	0.999	7.60	139.8	0.999	7.13
2	22.9	0.998	4.52	9.8	0.989	3.30	7.0	0.980	2.82
3	2.9	0.899	1.65	3.6	0.929	1.91	1.8	0.776	1.08
totals		2.897	14.08		2.918	12.81		2.756	11.03

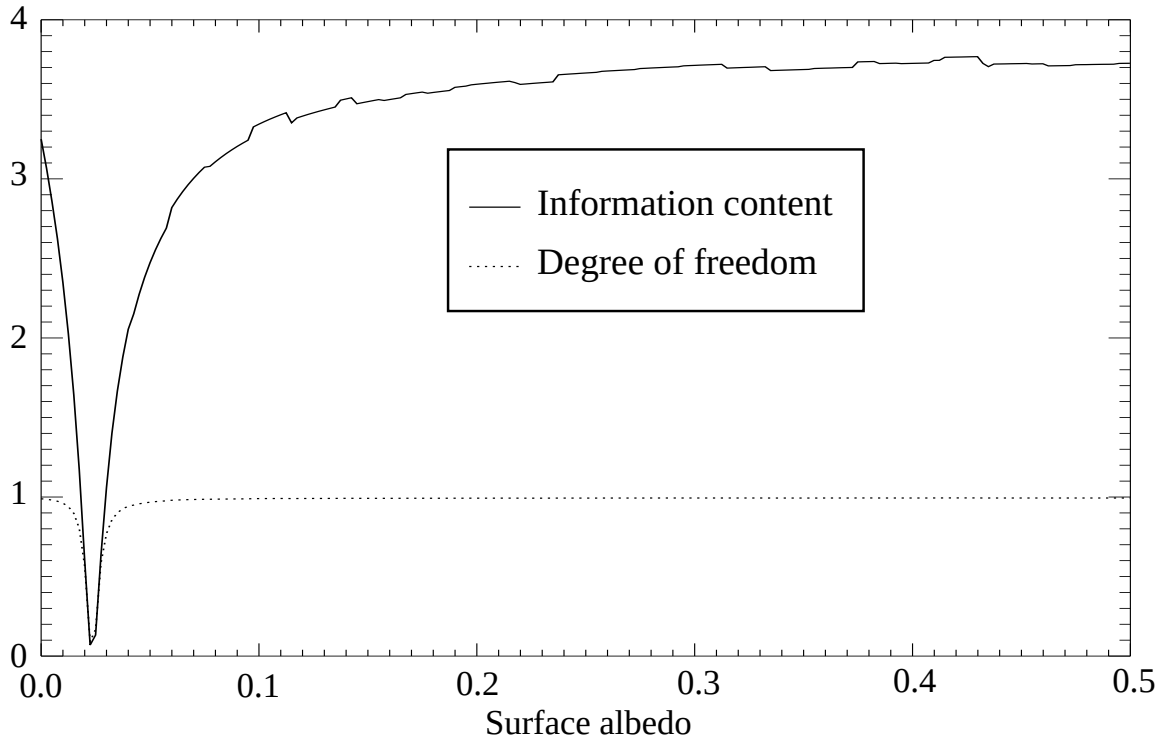


Figure 9. Information content computed as a function of surface albedo for the simple model described in the text.