Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simulation for TOVS

By J. R. EYRE

Meteorological Office Unit, Robert Hooke Institute for Cooperative Atmospheric Research, Clarendon Laboratory, Oxford

(Received 14 December 1987; revised 6 March 1989).

SUMMARY

A new approach is proposed for inverting satellite sounding data for use in numerical weather prediction (NWP). It may be considered in two ways: either as a retrieval scheme which uses a forecast profile as a background or ‘first guess’ and the covariance of forecast error as a constraint in order to produce retrieved profiles compatible with the subsequent NWP analysis; or as a module within a NWP analysis scheme to perform the vertical analysis of satellite radiances at the observation points (i.e. to project the information in the radiances on to the NWP analysis levels in the vertical). The new scheme performs an inversion of the radiances to retrieve simultaneously the temperature and humidity profiles, the surface temperature and microwave emissivity, and the cloud-top pressure and amount. It employs an iterative method which finds the maximum probability solution to a nonlinear retrieval/analysis problem. It operates on ‘raw’, potentially cloud-affected radiances and thus bypasses the separate pre-processing and cloud-clearing stages necessary in many retrieval schemes. The technique is a fully ‘physical’ method, involving calculations for each profile of corresponding radiances (and of derivatives of radiance with respect to profile parameters). It is also a ‘statistical’ method, since it uses the covariance of forecast error as a constraint.

The paper develops the theory of the new approach and examines its convergence properties and inherent quality control characteristics. The application of the scheme to data from the TIROS Operational Vertical Sounder (TOVS) is then discussed. The theoretical error characteristics of the scheme are investigated, and the results of retrievals from simulated TOVS radiances are presented. The theoretical advantages of retrieving cloud parameters simultaneously with other variables are demonstrated, in terms of both the temperature retrieval accuracies in cloudy conditions and the skill with which the cloud parameters themselves may be obtained.

1. INTRODUCTION

Satellite soundings of atmospheric temperature and humidity are a vital component of the global data base used in numerical weather prediction (NWP). Global soundings from TIROS Operational Vertical Sounder (TOVS) data (Schwalb 1978) are provided operationally to the international community by NOAA/NESDIS in Washington, D.C. using a retrieval method which, at present, is broadly the same as that described by Smith et al. (1979). Many countries have investigated the production of high-resolution retrievals for their own regions, and some are already using them on a routine basis (see, for example, Kelly et al. 1983; Turner et al. 1985).

It is now widely recognized that satellite sounding products have a positive impact on NWP analyses in the southern hemisphere and in data-sparse areas of the northern hemisphere, and hence on subsequent forecasts (Halem et al. 1982; Bengtsson 1985; Kashiwagi 1987). However, there have been problems in showing how TOVS data can have a consistent positive impact on analyses and short-range forecasts in areas where the conventional observation network is relatively dense. This arises for a number of
reasons related to the information content of the data, deficiencies in data-processing and retrieval techniques, and weaknesses in analysis methods.

One of the main problems has been that of background or ‘first-guess’ dependence. The inversion of satellite radiances is a mathematically ill-posed problem unless additional information is used (see Rodgers 1976). This information generally takes the form of a ‘first-guess’ profile and constraints on the departures allowed away from this guess. These cause the retrieval to retain certain aspects of the guess. In conventional regression retrieval techniques, the implied first guess is a climatological mean profile. In data-dense areas, the background field for the NWP analysis (itself a short-range forecast) is usually quite accurate and can easily be degraded by the assimilation of (less accurate) climatological components in the retrievals. This problem is being addressed in several ways, all of which aim to provide a better first guess for the retrieval. Uddstrom and Wark (1985) have shown how the radiances themselves can be used to select an improved guess (and appropriate constraint) from a partitioned climatological data set. Chedin et al. (1985) have demonstrated the use of a large library of profiles and corresponding radiances from which an appropriate guess can be chosen. Other groups (e.g. Susskind et al. 1984; Smith et al. 1985; Eyre et al. 1986) use a first guess provided by a NWP model.

A second problem concerns the effects of cloud on the satellite measurements, principally on the infrared channels. Current operational methods attempt to identify cloud-free data and to produce ‘cloud-cleared’ radiances in partly cloudy areas. A great deal of effort has gone into the improvement of these schemes (e.g. McMillin and Dean 1982; Eyre and Watts 1987). Nevertheless, the cloud detection and cloud-clearing processes still are major contributors to the errors in the final products, particularly in the lower troposphere. Alternative approaches to the cloud problem have been demonstrated. Some schemes limit themselves to the detection of cloud and subsequently use measurements only from those channels judged to be unaffected by cloud (Smith et al. 1984; Chedin et al. 1985). Other schemes make direct use of the (potentially) cloud-affected radiances and attempt to retrieve cloud parameters along with (and consistent with) the retrieved temperature and humidity profiles (e.g. Susskind et al. 1984; Smith et al. 1985; Huang and Smith 1986).

Another, and less explored, class of problems concerns the incompatibility which often exists between satellite sounding retrieval techniques and the NWP analysis schemes which try to assimilate the retrievals. The first-guess dependence problem is one aspect of this, but it is more general. Initial efforts to make use of satellite soundings in NWP have concentrated on application of the data in the form of retrieved profiles. NWP analysis schemes have been developed and tuned to make good use of radiosonde data, and the obvious initial approach has been to consider satellite soundings in a similar manner. Because of the inherently poor vertical resolution from currently available sounding systems and other problems in the data processing, it is evident that satellite products should be given lower weight than radiosondes if they are used in this way. Therefore, high-quality radianc data have tended to be used as though they were poor-quality radiosondes, and the strength of the satellite system and the true information content of its data have not been properly exploited. Satellite retrievals have rather subtle error characteristics (Eyre 1987) which must be understood and allowed for if the data are to be used optimally.

The idea that radianc data might be used directly in the analysis of meteorological fields, thus bypassing an explicit ‘retrieval’ stage, has been discussed for some time. With a few exceptions (see Durand and Juwanon du Vachat 1986; Nathan et al. 1985; Eyre 1981; Hoffman 1983; and their references), this idea has not been carried very far. Nevertheless, satellite sounding retrieval and NWP analysis are similar inverse problems.
There would seem to be advantage in considering them together rather than as two separate problems, particularly since the primary objective is not to make good retrievals for their own sake, but to provide the best possible NWP analysis (and subsequent forecast).

In this paper all the problems discussed above are addressed, and a new technique is described for extracting information from satellite sounding data for use in NWP. The method may be thought of in two ways: either as a retrieval scheme which uses a profile from a NWP model as background and provides retrieved profiles suitable for assimilation by a NWP analysis scheme; or as a module within a NWP analysis scheme for performing the analysis of satellite sounding data in the vertical. The mathematical equivalence of these two approaches demonstrates the relationship between the retrieval problem and that of assimilating radiance information directly into the NWP system (see Lorenc 1986 and 1988).

Ideally, the method should be applied three-dimensionally, i.e. it should use multi-channel radiances over an area to analyse/retrieve three-dimensional atmospheric fields. At present, this is computationally prohibitive, and so the method has been developed in one dimension: multi-channel radiances for a single sounding location (field-of-view) are used to retrieve/analyse the profile at that location. This is compatible with a 'split' analysis scheme in which the vertical and horizontal analyses are separated. It assumes that the vertical analysis/retrieval described here is followed by a horizontal analysis stage which interpolates the information from sounding locations to NWP model grid points.

The new scheme performs a simultaneous retrieval of the temperature and humidity profiles, the surface temperature, pressure and microwave emissivity, and the cloud-top pressure and cloud amount. Other parameters, such as total ozone, could easily be added to the scheme.

The method operates on cloudy radiances and thus bypasses the cloud-clearing and other pre-processing procedures present in most retrieval schemes. These lead to error characteristics in the clear-column radiances which are difficult to handle and can cause serious quality control problems. Also, because the method operates on single fields-of-view, it avoids assumptions of uniformity between adjacent sounding locations (e.g. of cloud-top pressure) inherent in many cloud-clearing techniques. Studies using high-resolution image data in conjunction with infrared sounding data have shown that these assumptions are rarely valid (Eyre et al. 1986).

The method follows the Newtonian iteration approach to the nonlinear inversion problem described by Rodgers (1976), which is essentially the same as the nonlinear Bayesian analysis method proposed by Purser (1984). It is an extension of the forecast first-guess retrieval/analysis scheme proposed by Eyre et al. (1985) and Lorenc et al. (1986), which is a quasi-linear scheme operating on cloud-cleared radiances. Both this and the new scheme use the forecast error covariance as a constraint on the inversion and are based on optimality principles.

The new method has aspects in common with the physical retrieval schemes of Susskind et al. (1984) and Huang and Smith (1986). The former retrieves in an iterative manner the full range of atmospheric profile, surface and cloud parameters. However, it is not a simultaneous technique, retrieving consecutively clear-column radiances, surface parameters, temperature profile and humidity profile, and then iterating the sequence. Nor does it seek to be statistically optimal. The latter is a modification of the simultaneous inversion scheme of Smith et al. (1985), extended to include cloud parameters. It is not based on optimality principles but constrains the solution using a limited set of empirically chosen basis functions.
In the remainder of this paper, the theoretical background to the new method is considered, and its applicability to analysis/retrieval using TOVS data is discussed. The performance of the scheme is explored, both in terms of its theoretical error characteristics and through studies of retrievals made from simulated TOVS data.

2. Theory

(a) The maximum probability solution

Let \( \mathbf{x} \) be a vector representing the atmospheric state (or a NWP model's representation of it) and \( \mathbf{y}^m \) be a vector of satellite sounding measurements (radiiances or brightness temperatures). Our purpose in making a 'retrieval' or 'analysis' is to find the most probable value of the atmospheric state \( \mathbf{x} \) given the measurements \( \mathbf{y}^m \), i.e. to maximize the conditional probability of \( \mathbf{x} \) given \( \mathbf{y}^m \):

\[
P(\mathbf{x}|\mathbf{y}^m) = \text{maximum}. \tag{1}
\]

Lorenc (1986) applies Bayesian estimation theory to the problem of NWP analysis and shows that, in the case of Gaussian error distributions, the most probable solution is that which minimizes a 'cost' or 'penalty' function

\[
J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^m - \mathbf{y}(\mathbf{x}))^T \cdot \mathbf{E}^{-1} \cdot (\mathbf{y}^m - \mathbf{y}(\mathbf{x})) \tag{2}
\]

where \( \mathbf{x}^b \) is the forecast background; \( \mathbf{C} \) is its expected error covariance; \( \mathbf{y}(\mathbf{x}) \) is the radiative transfer model or 'forward model', an operator for calculating the error-free measurements corresponding to the state \( \mathbf{x} \); \( \mathbf{E} \) is the expected covariance of the combined measurement and forward model errors; and superscripts \( ^T \) and \( ^{-1} \) denote matrix transpose and inverse respectively.

To find the most probable value of \( \mathbf{x} \), we may either minimize Eq. (2) or, assuming there are no multiple minima, find where the gradient of \( J(\mathbf{x}) \) with respect to \( \mathbf{x} \) is zero:

\[
J'(\mathbf{x}) = \mathbf{C}^{-1} \cdot (\mathbf{x} - \mathbf{x}^b) - \mathbf{K}(\mathbf{x}) \cdot \mathbf{E}^{-1} \cdot (\mathbf{y}^m - \mathbf{y}(\mathbf{x})) = 0 \tag{3}
\]

where \( \mathbf{K}(\mathbf{x}) \) contains the partial derivatives of \( \mathbf{y}(\mathbf{x}) \) with respect to the elements of \( \mathbf{x} \).

In this work we consider only the one-dimensional (vertical) retrieval/analysis problem. However, the theory is quite general and may, in principle, be applied to the three-dimensional analysis problem and also extended to four-dimensional data assimilation (see Lorenc 1988).

(b) The linear case

For the linear problem, \( \mathbf{K} \) is independent of \( \mathbf{x} \) and we may write

\[
\mathbf{y}(\mathbf{x}) = \mathbf{y}(\mathbf{x}^b) + \mathbf{K} \cdot (\mathbf{x} - \mathbf{x}^b). \tag{4}
\]

In the weakly nonlinear case, in which \( \mathbf{K} \) is a function of \( \mathbf{x} \), but varies only a little over the domain which represents all reasonable departures of \( \mathbf{x} \) from \( \mathbf{x}^b \), we may also use Eq. (4) in which \( \mathbf{K} = \mathbf{K}(\mathbf{x}^b) \).

Substituting Eq. (4) into (3), we can find an analytic solution

\[
\mathbf{x} = \mathbf{x}^b + (\mathbf{C}^{-1} + \mathbf{K}^T \cdot \mathbf{E}^{-1} \cdot \mathbf{K})^{-1} \cdot \mathbf{K}^T \cdot \mathbf{E}^{-1} \cdot (\mathbf{y}^m - \mathbf{y}(\mathbf{x}^b)). \tag{5}
\]

Matrix manipulation gives an equivalent formula which is computationally more efficient when the dimension of \( \mathbf{x} \) exceeds that of \( \mathbf{y}^m \):

\[
\mathbf{x} = \mathbf{x}^b + \mathbf{C} \cdot \mathbf{K}^T \cdot (\mathbf{K} \cdot \mathbf{C} \cdot \mathbf{K}^T + \mathbf{E})^{-1} \cdot (\mathbf{y}^m - \mathbf{y}(\mathbf{x}^b)). \tag{6}
\]

This is the familiar minimum variance solution which is also the maximum probability solution in the case of Gaussian statistics (see, for example, Rodgers 1976).
When performing a retrieval from cloud-cleared radiances, the problem is usually sufficiently linear for this form of the solution to be adequate. Indeed, it is possible to pre-compute the inverse matrix, \( W = C \cdot K^T \cdot (K \cdot C \cdot K^T + E)^{-1} \), for a small number of atmospheric and surface conditions, scan angles, etc., and then to choose a value of \( W \) appropriate to \( \mathbf{x}^2 \) by ‘nearest neighbour’ methods or by interpolation. However, experience with TOVS retrievals from cloud-cleared radiances has revealed other problems with the measurements after pre-processing and cloud-clearing; the cleared radiances often contain large errors and errors which are non-Gaussian (see Watts and Eyre 1986) and correlated between channels (Watts and McNally 1988). In general, it is clear that the error characteristics after cloud-clearing are complicated and are not accommodated properly by the simple formulation of Eq. (6).

(c) The nonlinear case

Even when using cloud-cleared radiances the problem is nonlinear, particularly with respect to the retrieval of humidity, and ideally should be treated as such. When attempting to retrieve from cloudy radiances, the problem is highly nonlinear, i.e. \( K \) changes rapidly as a function of some of the elements of \( x \). We must, therefore, seek an alternative approach to the solution of Eq. (3). Newtonian iteration yields the following solution: if we guess a vector \( x_n \), then the solution is found through the iteration

\[
x_{n+1} = x_n - J'(x_n)^{-1} \cdot J'(x_n)
\]  

(7)

where, by differentiation of Eq. (3),

\[
J'(x)^{-1} = S_n = (C^{-1} + K_n^T \cdot E^{-1} \cdot K_n)^{-1}
\]

(8)

and \( K_n = K(x_n) \). The matrix \( S_n \) is interesting in its own right, as it represents the error covariance of the retrieval/analysis (see Rodgers 1976). When the measurement vector is shorter than the profile vector, \( S_n \) is more efficiently computed in its equivalent form:

\[
S_n = C - C \cdot K_n^T \cdot (K_n \cdot C \cdot K_n^T + E)^{-1} \cdot K_n \cdot C.
\]

(9)

Iteration of these equations proceeds until convergence, i.e. until the increment \( (x_{n+1} - x_n) \) is acceptably small.

By matrix manipulation, we arrive at another formulation which is computationally more efficient:

\[
x_{n+1} = x_n + (x^b - x_n) + W_n \cdot \{y^m - y(x_n) - K_n \cdot (x^b - x_n)\}
\]

(10)

where

\[
W_n = C \cdot K_n^T \cdot (K_n \cdot C \cdot K_n^T + E)^{-1}.
\]

(11)

(d) Stability and convergence

Do the iterations represented by the above equations converge to a stable solution? Experiments have shown that, even when a reasonable first guess is used, the iteration will often oscillate away from the required value unless we ’damp’ the iterative descent to the minimum of Eq. (2). This can be achieved by replacing \( C \) by \( C' \) in Eqs. (8) and (9), where \( C'_{ij} = C_{ij} \), except for elements representing parameters causing the instability. These must be considerably reduced. Care is taken to use the correct value of \( C \) in the calculation of \( J'(x_n) \) (Eq. (3)) in order that the final value of \( x \) solves the appropriate equation. When \( J'(x_n) \) calculated using \( C \) and \( J'(x_n) \) calculated using \( C' \) are substituted into Eq. (7), the equivalent of Eq. (10) is

\[
x_{n+1} = x_n + J' \cdot (x^b - x_n) + W_n \cdot \{y^m - y(x_n) - K_n \cdot J' \cdot (x^b - x_n)\}
\]

(12)
where $\mathbf{I} = \mathbf{C}^\prime \cdot \mathbf{C}^{-1}$ and $\mathbf{W}_n^i = \mathbf{C}^\prime \cdot \mathbf{K}_n^\prime \cdot (\mathbf{K}_n \cdot \mathbf{C}^\prime \cdot \mathbf{K}_n^\prime + \mathbf{E})^{-1}$. If the error in each damped parameter is uncorrelated with other errors, then $\mathbf{I}$ is a diagonal matrix in which $I_{ii} = C_{ii}/C_{ii}$ for the damped parameter elements and $I_{ii} = 1$ for other elements.

(e) Quality control and subsequent horizontal analysis

Iteration of Eq. (12) proceeds until convergence, i.e. until the increment $(\mathbf{x}_{n+1} - \mathbf{x}_n)$ is acceptably small. At this point, we should also find that, if we substitute the profile $\mathbf{x}$ back into the radiative transfer equation, the departures $(\mathbf{y}_n - \mathbf{y}(\mathbf{x}))$ are of the order of the measurement error in all channels. If this is not the case, it suggests a problem with either the measurements or the forward model, such as gross errors in the radiances or the presence of more complex atmospheric conditions than the forward model allows for. This offers a natural and powerful mechanism for quality control. In principle, large errors in the background profile could also have the same effect, leading to the rejection of good measurements by the retrieval/analysis system. In practice, such errors would have to be extreme, and such cases present a general problem for NWP analysis which is not specific to satellite sounding data. Moreover, studies of the spatial and temporal characteristics of the rejected data should provide a useful mechanism for diagnosing any such problems with the NWP system.

The final profile $\mathbf{x}$ may also be substituted into Eq. (9) to give the final value of $\mathbf{S}$, which is an estimate of the expected retrieval error covariance. $\mathbf{S}$, together with the background error covariance $\mathbf{C}$, may then be used to specify the weight to be given to the retrieval (or vertical analysis) when it is used in the subsequent horizontal analysis. The problem of appropriate weights is not trivial, since the use of the background field in the vertical analysis gives rise to a correlation between background and retrieval errors. This problem is inherent in some form in all retrieval methods because of the inevitable background dependence. However, within this approach, there exists a formalism for treating it. The correlated error problem is analogous to that arising in the ‘super-observation’ analysis procedure described by Lorenc (1981): the solution is to amplify the observation–background increments and to reduce their weights. Lorenc et al. (1986) have discussed how this approach may be applied in an approximate manner to the present problem of correlated retrieval error. The critical parameter in determining how the retrieval increments should be adjusted and weighted at level $i$ is the ratio $S_{ii}/C_{ii}$. When this ratio is much less than one, the retrieved value is almost independent of the background and can be treated as an independent observation with full weight. As the ratio tends to one, the retrievals are adding no information to the background and should be given zero weight.

3. Application to TOVS data

(a) The retrieval scheme

TOVS consists of three radiometers: the High-resolution Infrared Radiation Sounder (HIRS/2, hereafter called HIRS), the Microwave Sounding Unit (MSU), and the Stratospheric Sounding Unit (SSU). See Smith et al. (1979) for a description of all the TOVS instruments and the characteristics of their channels. In this work only HIRS and MSU data have been considered.

In previous work, the linear form of the method described above was applied to cloud-cleared TOVS data (Eyre et al. 1986). Also, the data used were ‘corrected’ for scan angle effects and for microwave surface emissivity, which simplifies the problem
and reduces the number of pre-computed $W$ matrices required. However, this approach suffers from the inevitable, and sometimes large, errors introduced by these correction procedures and the cloud-clearing.

Here we have developed a method intended for application to TOVS data which have undergone no cloud-clearing or pre-processing (other than the mapping of MSU data to HIRS fields of view). The profile vector includes all the variables required for an adequate representation of the radiative transfer problem, i.e. temperature profile (surface to 0.1 mb), humidity profile (surface to 300 mb), surface air and skin temperatures, surface pressure, microwave surface emissivity, cloud-top pressure and cloud amount. It could probably be extended profitably to include total ozone amount and parameters representing the reflection of solar radiation in the short-wave channels. Forty pressure levels from 1000 to 0.1 mb have been used to represent the temperature profile. In this way, the sharply-peaking weighting functions which arise in the presence of cloud may be represented quite accurately. Humidity is expressed as the logarithm of mixing ratio, since forecast errors in this quantity are more constant than in mixing ratio itself.

Experiments to date have used a measurement vector containing HIRS channels 1–8 and 10–15 and MSU channels 1–4. (See Smith et al. (1979) for a description of TOVS channels.) HIRS channel 16 would be added for satellites other than NOAA-9 (on which it is defective), and channels 18 and 19 could be used without problem during the night. We have chosen to work with radiances rather than brightness temperatures. This improves the linearity of the problem in certain respects (e.g. radiance is linear in cloud amount) although it degrades the linearity of other aspects.

In previous work (Eyre et al. 1985, 1986), $K$ has been calculated by a 'brute force' method: the radiances are calculated for a given profile, and then each element of the profile vector is perturbed by one unit (i.e. 1 K for temperatures, 1 mb for pressures, etc.) in turn. The radiances calculated for the perturbed profiles minus those for the unperturbed profile yield all the elements of $K$. This method is accurate and adequately fast for off-line calculations, but it is too slow for real-time applications. Therefore, the TOVSRAD model (Eyre 1984) has been adapted to calculate all the elements of $K$ in parallel with the evaluation of the radiances $y(x)$ as described in appendix A. This involves only a little additional computation time.

The forward model also includes the presence of cloud: the overcast radiances for cloud at standard pressure levels close to the current value of cloud-top pressure are calculated, and the overcast radiance at this pressure level is then found by interpolation. The cloudy radiance is then given by

$$R = (1 - N_c) R^c + N_c R^o(p_c)$$

where $R^c$ is the clear radiance, $R^o(p_c)$ the overcast radiance for cloud-top $p_c$, and $N_c$ the 'effective' cloud amount. For opaque cloud, $N_c$ will represent the true cloud amount, while for semi-transparent cloud it will be the product of the fractional cloud cover and the cloud 'emissivity' (i.e. one minus cloud transmittance). It is assumed here that cloud emissivity is independent of wavelength and the reflection of solar radiation is negligible. It is also assumed that only one level of cloud is present.

Cloud is taken to have no effect on the calculated microwave radiances. This approximation is not always valid, as cloud liquid water often has a measurable influence on MSU channel 1. Since we do not have sufficient microwave channels to separate the effects of cloud liquid water from changes in surface emissivity, we can interpret the retrieved 'surface emissivity' as being an effective value which compensates also for cloud effects.
The $E$ matrix represents the 'measurement error' but must include not only radiometric errors (instrument noise) but also any uncorrected errors in the forward model. $E$ is specified in terms of radiance error variances for which the radiometric component is constant (for a given satellite). Forward model error originates from a number of sources including spectroscopic uncertainties and inhomogeneous scenes. Experience suggests that, for a simple representation, the error variance of brightness temperature calculated by the forward model will be roughly independent of the actual brightness temperature. Since radiance and brightness temperature are related nonlinearly, the error variance of the calculated radiance will vary with brightness temperature. Consequently, the $E$ matrix is not fixed but is calculated for each retrieval, with the forward model error component being a function of the measured brightness temperature. The values used in this study are given in Table 1. All errors have been assumed uncorrelated between channels, leading to a diagonal $E$ matrix. The values assumed for forward model error have been kept reasonably small, since it is intended that the aspects of quality control inherent in the method will be used to screen out cases where the forward model is grossly inadequate in representing the true atmospheric conditions. The correct specification of the forward model error is a difficult problem. In future it is hoped to refine this aspect both in terms of the size of the errors and the inter-channel correlations.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Radiometric error, $e'$ (HIRS in mW m$^{-2}$ sr$^{-1}$ (cm$^{-1}$)$^{-1}$)</th>
<th>Forward model error, $e''$, in brightness temperature (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIRS-1</td>
<td>0.62, 0.13, 0.11, 0.068, 0.048, 0.056, 0.040, 0.019, 0.023, 0.026, 0.040, 0.030, 0.0010, 0.00092, 0.00080, 0.00055, 0.00063, 0.00037, 0.00020</td>
<td>0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2</td>
</tr>
<tr>
<td>MSU-1</td>
<td>0.17, 0.15, 0.18, 0.18</td>
<td>0.2, 0.2, 0.2, 0.2</td>
</tr>
</tbody>
</table>

**Notes**
1. Radiometric errors are typical of NOAA-7 instruments.
2. $E$ matrix is obtained as follows:

$$E_y = (e')^2 + \left(\frac{dB_y(T)}{dT} \cdot e''\right)^2$$

where the Planck function derivative is evaluated at the measured brightness temperature. Off-diagonal elements of $E$ are set to zero.
The C matrix has been obtained from forecast error covariance data supplied by the Forecasting Research branch of the Meteorological Office, as described by Eyre et al. (1986). Log(mixing ratio) error covariances are calculated from the temperature and relative humidity covariances as described in appendix B. The resulting errors at each level and the inter-level correlation of errors are given in Table 2. Surface emissivity and cloud error covariances are not taken from this source. They are allowed to be large and uncorrelated with other variables, with current values as shown in Table 3. This means that the retrieved values of these parameters are effectively unconstrained by their background values. Surface pressure errors have been taken from forecast error statistics. These strictly refer to mean sea level pressure. Over the ocean they are appropriate for surface pressure errors, but over land an additional contribution should be added to represent the uncertainty in the mean surface elevation at the measurement point. Also over land, errors in surface skin temperature should be raised to represent the looser relationship between the variable provided by the NWP model and that which determines the radiance emitted by the surface. Cloud errors and humidity errors are assumed uncorrelated. This may seem unrealistic, but it is necessary to remember that the cloud parameters must represent the structure within a single HIRS field-of-view whereas the background values of humidity resolve only larger scales.

Since a forecast model can in practice provide a background profile at only a limited number of pressure levels, a method is required for specifying the values at the intermediate levels (and the stratospheric levels above the range of the model) used in the radiative transfer calculation. In this study we have simulated a situation in which all the extra levels are found by regression from the levels provided, given a vector $x_1^b$ at the provided subset of levels. The vector for the extra levels $x_2^b$ is given by

$$ x_2^b = D \cdot x_1^b $$

(14)

where $D$ is a matrix of regression coefficients. These are obtained from a large representative set of profiles in which data are specified at all levels. If we write the full profile vector as

$$ x^b = \begin{pmatrix} x_1^b \\ x_2^b \end{pmatrix} $$

(15)

then its corresponding error covariance $C$ is given by

$$ C = \begin{pmatrix} C_1 & C_1 \cdot D^T \\ D \cdot C_1 & D \cdot C_1 \cdot D^T + G \end{pmatrix} $$

(16)

where $G$ is the residual error covariance in the regression which provides $D$.

The use of a single $C$ implies that forecast errors are homogeneous. However, we might expect them to vary significantly with geographical location and meteorological situation. This is a problem common to many methods of NWP analysis, affecting the relative weights given to forecast and observations (including conventional retrievals). As more information becomes available on the characteristics of forecast error, then the method presented here can easily be adapted to use it.

In this study, inversion through iteration of Eq. (12) has been used. With the exception of the cloud parameters, the iteration is started with an initial profile equal to the background profile. This is not essential—any reasonably close profile should be adequate, but an accurate guess will speed the convergence. It is necessary to distinguish clearly between the role of the background profile and the role of the initial-guess profile.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Level (mb)</th>
<th>Standard deviation of error</th>
<th>Inter-level correlations of error (×100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>50</td>
<td>2.03</td>
<td>100</td>
</tr>
<tr>
<td>T</td>
<td>70</td>
<td>2.08</td>
<td>88 100</td>
</tr>
<tr>
<td>T</td>
<td>100</td>
<td>1.90</td>
<td>61 100</td>
</tr>
<tr>
<td>T</td>
<td>150</td>
<td>1.72</td>
<td>26 74 100</td>
</tr>
<tr>
<td>T</td>
<td>200</td>
<td>1.99</td>
<td>10 4 50 100</td>
</tr>
<tr>
<td>T</td>
<td>250</td>
<td>2.69</td>
<td>-1 3 0 0 28 100</td>
</tr>
<tr>
<td>T</td>
<td>300</td>
<td>1.90</td>
<td>-9 -11 -19 -26 -12 37 100</td>
</tr>
<tr>
<td>T</td>
<td>400</td>
<td>2.03</td>
<td>-10 -13 -23 -32 -29 -14 17 100</td>
</tr>
<tr>
<td>T</td>
<td>500</td>
<td>1.75</td>
<td>1 0 8 7 23 23 17 7 33 100</td>
</tr>
<tr>
<td>T</td>
<td>700</td>
<td>1.90</td>
<td>7 8 9 2 -6 -8 -6 0 15 100</td>
</tr>
<tr>
<td>T</td>
<td>850</td>
<td>2.15</td>
<td>6 2 6 -2 -5 -6 5 1 31 100</td>
</tr>
<tr>
<td>T</td>
<td>950</td>
<td>2.50</td>
<td>8 7 7 4 -1 -3 -3 -2 1 38 100</td>
</tr>
<tr>
<td>T</td>
<td>1000</td>
<td>2.53</td>
<td>6 6 4 2 -1 -2 -1 0 2 17 49 85 100</td>
</tr>
<tr>
<td>TS-AIR</td>
<td></td>
<td>2.34</td>
<td>14</td>
</tr>
<tr>
<td>TS-SKIN</td>
<td></td>
<td>3.67</td>
<td>15</td>
</tr>
<tr>
<td>LNC</td>
<td>300</td>
<td>0.54</td>
<td>-2 -3 -6 -2 3 1 1 0 -1 1 7 9 20 100</td>
</tr>
<tr>
<td>LNC</td>
<td>400</td>
<td>0.59</td>
<td>-2 -2 -4 -11 -16 -3 28 12 6 1 0 0 0 0 0 100</td>
</tr>
<tr>
<td>LNC</td>
<td>500</td>
<td>0.53</td>
<td>-4 -4 -5 -10 -15 -14 0 28 15 1 0 0 0 1 37 100</td>
</tr>
<tr>
<td>LNC</td>
<td>700</td>
<td>0.46</td>
<td>-2 -2 -3 -5 -8 -10 -1 8 20 4 2 2 1 1 1 1 16 46 100</td>
</tr>
<tr>
<td>LNC</td>
<td>850</td>
<td>0.37</td>
<td>2 3 1 -1 -3 -3 0 1 5 14 18 11 8 9 1 5 11 23 100</td>
</tr>
<tr>
<td>LNC</td>
<td>950</td>
<td>0.32</td>
<td>3 4 4 2 0 -3 -1 -1 3 15 31 29 23 21 3 3 5 9 35 100</td>
</tr>
<tr>
<td>LNC</td>
<td>1000</td>
<td>0.30</td>
<td>1 1 0 -1 0 -1 0 2 13 37 48 47 41 7 2 3 6 21 48 100</td>
</tr>
<tr>
<td>LNCS</td>
<td>0.31</td>
<td>23</td>
<td>3 2 0 0 -1 0 -1 1 14 41 59 61 47 7 1 2 5 17 36 69 95 100</td>
</tr>
<tr>
<td>PRESS</td>
<td>3.42</td>
<td>24</td>
<td>4 1 2 3 1 -2 -11 -15 -16 -25 -27 -26 -24 -15 -3 -6 -7 -6 -11 -18 -22 -22 -21 100</td>
</tr>
</tbody>
</table>

T = temperature in K
LNC = ln (mixing ratio in g/kg)
TS-AIR = surface air temperature in K
TS-SKIN = surface skin temperature in K
LNCS = surface ln (mixing ratio in g/kg)
PRESS = surface pressure in mb
The iteration will converge to a profile which satisfies Eq. (3). Unless multiple minima are present (see section 2(a)), this is independent of the initial profile. However, it is always a function of the background profile.

It has been found that a reasonable initial guess of the cloud parameter is useful in speeding the convergence, and indeed it is often essential for obtaining convergence. The method for obtaining initial cloud parameters using the guess profile and measured radiances in HIRS channels 7 and 8 is described by Eyre and Menzel (1989). This is a special case of the cloud retrieval method presented by Susskind et al. (1987) in which the cloud parameters which give the best fit between measured and calculated radiances are found. The initial cloud parameters are sensitive to errors in the guess profile. It is beneficial to use the best available guess profile in this step, and so the following procedure has been adopted for the iteration.

a) On the first step in the iteration, the guess profile is set equal to the background profile and only MSU channels are used to update the profile. (This improves the estimates of the temperature profile and the surface emissivity.)

b) The new profile and the measured radiances in HIRS channels 7 and 8 are used to calculate the initial cloud-top pressure and cloud amount.

c) The second and subsequent iterations use all the channels to update the profile, including the elements containing cloud information.

As explained in section 2(d), it is necessary to damp the adjustment of some parameters during the iteration. This is found to be required only for those parameters which make the problem highly nonlinear—in this case, the cloud parameters. Standard deviations of 25 mb and 0.05 for cloud-top pressure and cloud amount errors respectively are found to be satisfactory. These values control the iterative descent only; the values providing the constraint on the solution of Eq. (3) are much larger (see Table 3).

At each step in the iteration, the new profile is checked and, if necessary, corrected to be physically reasonable in the following ways. The cloud amount is constrained between 0 and 1, and the cloud-top pressure is held smaller than the station level pressure. The mixing ratio is checked for supersaturation.

The convergence criterion is somewhat arbitrary. We have chosen to stop the iteration when the absolute value of the profile increment \( (x_{n+1} - x_n) \) is lower than some fraction, \( \beta \), of the background error (or, more precisely, of the roots of the diagonal of \( C^{-1} \)) for all elements of the profile vector. As \( \beta \) is reduced the solution becomes more accurate but the number of iterations required increases. A value of \( \beta = 0.4 \) has been found to give rapid convergence while not degrading significantly the retrieval accuracy.

(b) Simulated retrievals

The performance of the retrieval scheme has been examined using simulated data for a large number of profiles. For each case, a 'true' profile has been established by taking one of a set of profiles typical of European/N. Atlantic conditions in October (Pescod and Eyre 1983). The profile is interpolated from the special levels of the
radiosonde report to the subset of standard levels which would be provided by the NWP model (see Table 2) and then extrapolated to the 40 levels of the radiative transfer model by regression as described in section 3(a). The microwave surface emissivity has been set to a random value between 0.6 and 0.99 and the cloud-top pressure to a random value between 300 and 900 mb. A random cloud amount has been established such that 14% of cases have a fractional cloud cover of 1, 14% are cloud-free and the remainder have uniformly distributed intermediate values.

Having established a 'true' profile, an appropriate background profile \( x^b \) is calculated by perturbing the 'true' profile \( x^t \) as follows:

\[
x^b = x^t + \sum_i \xi_i \lambda_i^b \cdot \lambda_i
\]  

(17)

where \( \lambda_i \) are the eigenvectors and eigenvalues of the forecast error covariance \( C \), and \( \xi_i \) is a random number drawn from a Gaussian population with zero mean and unit standard deviation. In this way the perturbations have covariance \( C \). Background values of microwave surface emissivity and cloud parameters have been fixed at the values given in Table 3.

Both 'true' profile and background profile are checked and, if necessary, corrected for supersaturation.

'True' radiances and brightness temperatures are calculated from the true profile using the forward model. The simulated measured radiances are found by adding random Gaussian noise of covariance \( E \) (as defined by the data in Table 1) to the 'true' radiances.

4. RESULTS

(a) Studies of theoretical retrieval error covariance

As discussed in section 2, if we find a profile vector \( x \) which satisfies Eq. (3), then its error covariance \( S \) is given by Eq. (9). The roots of the diagonal elements of \( S \) may be interpreted as the standard deviations of retrieval error. By evaluating \( S \) for different values of \( x \), we can study some of the error characteristics of the inversion scheme. One should not place too much emphasis on the absolute values of retrieval error, since they are highly dependent on the values of background error (for which a typical covariance matrix has been taken for illustrative purposes). More attention should be paid to the ratio of retrieval to background error, as this is related to the information which the radiance data supply to the NWP system. Also, this ratio determines the weight which should be given to the retrieved profile (vertical analysis) when including it in the horizontal stage of the NWP analysis (see Lorenc et al. 1986).

Of particular interest here is the behaviour of \( S \) as a function of those variables which make the problem highly nonlinear, namely the cloud-top pressure and cloud amount. \( S \) has been studied for one basic temperature/humidity profile (the mean of 800 profiles representative of the European/N. Atlantic area in October—see Pescod and Eyre (1983)) with a surface emissivity of 0.8, but for a wide range of cloud conditions. The following results highlight the more significant findings.

Figure 1 shows the retrieval errors expected for two cases—cloud-free conditions and full cloud cover at 500 mb—along with the background profile errors. Here we have used 'standard' measurement error (\( E \) matrices calculated from the data given in Table 1) and 'standard' background error (a \( C \) matrix obtained from the data given in Tables 2 and 3). Figure 1 shows that, even in cloud-free conditions, the retrieval errors represent only a moderate reduction from the background errors. This arises because the background errors themselves have been assumed quite small (for the temperature-profile), typical of 12-hour forecast errors for a regional NWP model operating in northern
hemisphere mid-latitudes. Also the inter-level correlations of background error are quite weak and the TOVS weighting functions relatively broad. The combination of these facts limits the information content of the TOVS radiances with respect to the NWP analysis. Nevertheless, the ratio of retrieval to background error represents a decrease in error variance which would be beneficial if obtained in practice.

The retrieval errors for the cloudy case shown in Fig. 1 demonstrate that, for levels well below the cloud top, retrieval performance degrades to behaviour typical of an MSU-only system. However, at and just above the cloud top, the retrieval error is reduced a little. This effect is demonstrated more clearly in Fig. 2, which shows 500 mb temperature retrieval error as a function of cloud-top pressure and cloud amount. Errors are presented here as fractional unexplained variances, i.e. the variance of retrieval error divided by the variance of the prior (or background or forecast) error. For 500 mb temperature the forecast error variance is assumed to be (1.75 K)^2. These results show that, when treated optimally, the effects of cloud on infrared radiances can improve some
aspects of the retrieval performance. This is not observed in retrieval schemes where cloud-clearing and inversion are separated; in these, cloud can only degrade the retrievals.

The crosses in Fig. 1 indicate the retrieval error in the cloud-free case when the background errors of cloud parameters are set to zero (i.e. when we know a priori that we have a cloud-free case). This shows that, except for surface skin temperature, retrieval performance is only degraded a little by lack of prior knowledge of the cloud conditions.

Figures 3 and 4 show plots for the error in retrieved cloud parameters as a function of these parameters. These plots demonstrate that cloud parameters may be retrieved with reasonable accuracy under most conditions, and particularly well for significant amounts of high cloud. Understandably, cloud-top pressure errors increase at low cloud amount, as do cloud amount errors for low-level cloud. In these cases, cloud uncertainties have little effect on the measured radiances and so do not degrade unduly the retrieval of other parameters.

Experiments have also been conducted to test the sensitivity of retrieval errors to changes in E and C. Figure 5 shows the effect of doubling or halving the assumed forward
model error contribution to \( E \). In most cases this contribution dominates the radiometric component. The figure shows only small changes in retrieval performance except in temperatures near the surface.

Figure 6 shows the effect of changing the inter-level correlations of background error. Assumed standard deviations of error at each level have been kept fixed, but error correlation coefficients \( R_{ij} \) have been changed to \( (r_{ij})^\gamma \) with \( \gamma \) taking values 0.5, 1 and 2. This shows that, for all profile elements except skin temperature, the retrieval performance is very sensitive to the strength of the inter-level correlations.

An additional experiment has been conducted to explore the effects of including the shortwave window channels, HIRS channels 18 and 19. The influence of these additional channels is found to be very small for most variables, with moderate improvements in the retrieval of skin temperature and low-level cloud parameters. However, the atmospheric profile used represents mean mid-latitude conditions, and larger effects may be expected for tropical atmospheres.

\[ (b) \quad \textit{Retrievals from simulated radiances} \]

Using the method described in section 3(c), simulated radiances and realistic background profiles were produced. These were used as inputs to the inversion scheme to obtain retrieved profiles which were then compared with the 'true' profiles to assess the error characteristics of the scheme.

Two hundred profiles with random cloud and surface emissivity conditions have been tested, and successful retrievals (i.e. convergence of Eq. (12) in ten iterations or less) were obtained in 189 cases. In most cases convergence is obtained after three or four iterations. As noted in section 2(c), by testing convergence of the profile increments, the agreement between measured and calculated radiances at convergence may be used as an independent check for anomalous conditions. For each retrieval at convergence, all channels are checked, and the largest value, \( \delta \), of measured minus calculated radiance divided by expected error is noted. For the 'perfect' Gaussian statistics represented by this simulation we find \( \delta > 3 \) in only 10 cases and \( \delta > 4 \) in only 4 cases.

The 11 cases of no convergence have been examined. They are found mainly to be cases in which the iterating profile oscillates about the true profile. However, in all cases, \( \delta \) is found to be less than 4 and usually less than 2. This suggests that such profiles, even though they have not converged, will represent good retrievals provided that they meet

![Figure 4. Theoretical r.m.s. errors in retrieved effective fractional cloud amount as a function of cloud conditions.](image)
some limit on $\delta$. Nevertheless, they have not been included in the statistics reported here.

For each successful inversion, the retrieval error and background (first-guess) error have been calculated, and the means and standard deviations of these data have been computed. The results for the atmospheric profiles are shown in Fig. 7. The retrievals have been partitioned between 'cloudy' (cloud amount $> 0.5$) and 'less cloudy' (cloud amount $\leq 0.5$) cases. It can be seen that these statistics are compatible with the theoretical errors derived from the $S$ matrix and shown in Fig. 1.

Errors in retrieved cloud parameters are expected to vary greatly over their full range, as shown in Figs. 3 and 4. This means that error statistics in terms of overall standard deviations have little meaning. We have therefore excluded from the cloud statistics 'bad cloud' cases, where the retrieved cloud pressure is in error by more than 100 mb. The remainder have the statistics shown in Table 4.
Figure 6. Theoretical r.m.s. errors as a function of inter-level correlations of background error: background errors; other lines are retrieval errors with different values of $\gamma$ (see text).

The 'bad cloud' cases represent about 35% of the total. However, many present no problems since they are cases in which both 'true' and retrieved cloud amounts are very small. Also, some others occur when small amounts of 'true' cloud are retrieved as large amounts of cloud very close to the surface. In summary, the cloud parameters are retrieved with error characteristics consistent with the theoretical values shown in Figs. 3 and 4.

<table>
<thead>
<tr>
<th>TABLE 4. STATISTICS OF CLOUD RETRIEVAL ERRORS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Retrieval error</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cloud-top pressure (mb)</td>
</tr>
<tr>
<td>Cloud amount</td>
</tr>
<tr>
<td>First guess error</td>
</tr>
<tr>
<td>Cloud-top pressure (mb)</td>
</tr>
<tr>
<td>Cloud amount</td>
</tr>
</tbody>
</table>
Figure 7. Root-mean-square errors in simulated retrievals: ––– background errors; ––– retrieval errors for fractional cloud amount <0.5; . . . retrieval errors for cloud amount >0.5.

(c) Anomalous conditions

The simulation study reported above represents 'ideal' conditions: measured radiances and background profile errors are assumed to be Gaussian and unbiased, and the 'true' state conforms to the single-layer cloud model assumed in the forward calculation. Additional experiments have been performed to test the behaviour of the scheme when some of these conditions are not obeyed.

'Rogue' measurements have been simulated by adding a large error (equal to 10 times the expected measurement error, i.e. = 2 K) to one of the tropospheric channels. Under these conditions, convergence is usually obtained, but a check on the radiances at convergence shows a high value of δ in the rogue channel, usually δ > 5. This suggests that measurement errors of this type can be detected.

Biased measurements or calculated radiances have been simulated by adding offsets in brightness temperature to measurements in all channels. An extreme case was
simulated: it was assumed that no bias correction had been applied to the forward model. Appropriate bias errors were obtained from radiosonde co-locations from data in January 1986 (for method, see Eyre (1983)). In this way most channels acquired absolute values of bias in the range 0-5 to 1.5 K. Under these conditions about half the inversion attempts failed to converge. Whether they converged or not, almost all retrievals had values of $\delta$ greater than 3. The successful retrievals had errors with increased standard deviations and, as expected, with significant biases. These results show that we have a mechanism for detecting problems of this type, but that the inversion does not become totally unstable under these conditions. In a routine system, these biases could be monitored and corrected, using biases between measured radiances and those calculated from co-located radiosondes in cloud-free cases.

The limitations of the single-layer cloud model have been examined by simulating mixed cloud. The 'true' radiances were calculated as the mean of two radiance vectors, corresponding to full or partial cloud cover at different levels. Contrary to expectations, the inversion scheme was not very sensitive to these conditions; convergence was usually obtained with low values of $\delta$. This is disappointing in that it suggests we are unable to detect cases of mixed cloud. However, the effects on the temperature and humidity retrieval accuracies were not great; standard deviations of retrieval error were increased only slightly. The worst consequence was the introduction of a bias of about 0.5 K to temperature retrievals in the lower troposphere. Physically, the inversion scheme responds to 2-layer cloud by retrieving a single layer at an intermediate level and by over-estimating slightly the temperature of the atmosphere below the cloud.

5. SUMMARY AND CONCLUDING REMARKS

This study has demonstrated a new inversion method for TOVS data which could be adopted for assimilating cloudy radiances directly into the analysis of a NWP model. In this scheme, the conventional pre-processing and cloud-clearing stages of the retrieval procedure are bypassed and the errors associated with the data treated in a more nearly optimal manner.

The method is a nonlinear, maximum probability technique which can accommodate most aspects of the real radiative transfer problem. It is an iterative method but is found to be stable and to converge rapidly in most cases. The scheme also incorporates quality control features which allow certain types of anomalous conditions and data to be recognized, thus preventing gross retrieval errors. As part of the new scheme, a fast model has been developed for calculating the derivatives of the radiative transfer equation with respect to all the profile, surface and cloud variables which appear in it.

In the presence of clouds, as in other schemes, retrieval accuracies are degraded below the cloud top. However, at and just above the cloud top, the retrieval performance can be improved as we take advantage of the sharper infrared weighting functions which the cloud produces. Also, retrieval of the cloud parameters themselves shows considerable skill.

In these studies, a particular matrix has been used for the forecast error covariance. However, it has been shown that the retrieval accuracies, and hence the weight which should be given to the data in the subsequent NWP analysis, are strongly dependent on the inter-level correlations of error assumed. The values used here are the best available at present, but there is considerable scope for further study of the vertical error correlation structures of the NWP models with which such an inversion scheme might be applied. It would be useful to have more detailed information on the vertical covariances of forecast
errors as a function of forecast range, synoptic conditions, thermal gradients, land/sea, etc., for the NWP models of a number of forecasting centres.

The scheme described here has been applied successfully to real TOVS data; the results are reported in part II (Eyre 1989). In principle, the technique as applied to TOVS could be improved by including other parameters which lead to a better description of the forward problem. The effect of ozone on the HIRS longwave channels might be treated by the addition of total ozone amount to the list of simultaneous variables and HIRS channel 9 to the measurement vector. In the daytime, the effect of reflected solar radiation might be included by adding parameters representing surface and cloud reflectances to the variables and HIRS channels 18 and 19 to the measurement vector. Information from simultaneous measurements of the Advanced Very High Resolution Radiometer (AVHRR) are potentially very useful. For example, the AVHRR pixels coincident with each HIRS field-of-view might be analysed to extract the clear-column radiances in the 11 μm window region. This could be used, in clear and partly cloudy conditions, to construct an additional ‘cloud-free HIRS’ channel and might be expected to improve retrievals close to the surface in partly cloudy conditions.

Finally, it is noted that the scheme described here has potential for wide application; it could easily be adapted to other sounding systems. In particular, it is expected to be appropriate for the combination of microwave and infrared instruments of the Advanced TOVS system to be flown on the satellites NOAA-K, -L and -M in the mid-1990s. Its adaptation for this purpose will be considered in future studies.

**Acknowledgements**

This work was conducted while I was a visiting scientist at the Co-operative Institute for Meteorological Satellite Studies (CIMSS), University of Wisconsin-Madison. I wish to thank Dr W. L. Smith and colleagues at CIMSS for their support. I am indebted to Mr R. J. Purser (lately CIMSS, now Meteorological Office) for invaluable suggestions in the early stages of this work. I am also grateful to Mr A. C. Lorenc, Dr R. W. Saunders and Mr P. D. Watts (Meteorological Office) and Dr J. J. Barnett and Mr A. P. McNally (Oxford University) for their comments on the paper.

**Appendix A: Calculation of the K Matrix**

The radiative transfer equation for a cloud-free TOVS channel $j$ may be written as

$$R_j = -\varepsilon_s B_j(T_s) \tau_j(p_s) + \int_0^{p_s} B_j(T(p)) \frac{d\tau_j(p)}{dp} dp +$$

$$+ (1 - \varepsilon_s) \int_0^{p_s} B_j(T(p)) \frac{d\tau_j^* (p)}{dp} dp$$

(A1)

where $\varepsilon_s$ is the surface emissivity; $B_j(T)$ is the mean Planck function at temperature $T$ in channel $j$; $T_s$ is the surface skin temperature; $p_s$ is the surface pressure; $\tau_j(p)$ is the mean transmittance appropriate to channel $j$ from pressure $p$ to space along the satellite viewing direction; and $\tau_j^*(p)$ is the 2-path transmittance from pressure $p$ down to the surface and then to space.

For specular reflection at the surface, both paths in $\tau_j^*(p)$ have the same surface zenith angle, $\theta$, as does $\tau_j(p)$. (If the reflection is other than specular but the radiation may be treated as monochromatic, the downward part of the path is calculated at some appropriate mean angle $\theta^*$, and transmittances may be related through the expression
\[ \ln \tau(p, \theta) = (\sec \theta^* / \sec \theta) \ln \tau(p, \theta) \]. We have neglected here any atmospheric scattering and any contributions from reflected solar radiation.

Dropping the subscript \( j \) and simplifying the notation (i.e. \( \tau_i(p_i) \) becomes \( \tau_i \), etc.), Eq. (A1) may be approximated for numerical evaluation as follows:

\[ R^c = \varepsilon_s B_s \tau_s + \sum_{i=1}^{I} \left\{ (B_i + B_{i-1}) \{ \tau_{i-1} - \tau_i + (1 - \varepsilon_s)(\tau_i^* - \tau_{i-1}^*) \} \right\} \]  \hspace{1cm} (A2)

where the atmospheric profile has been expressed in discrete form at \( I \) levels. At the top of the atmosphere, \( p_0 = 0, T_o = T_1 \), and \( \tau_o = 1 \). At the surface \( p_I = p_s \) (but \( T_I \neq T_s \)).

The elements of \( \mathbf{K} \) are found by differentiating Eq. (A2) with respect to all the elements of the profile vector:

\[ K_{mj} = dR_j/dx_m. \]  \hspace{1cm} (A3)

Ideally, the algorithm representing the forward problem should be differentiated exactly. Here, some approximations have been made in order to speed the computation.

(a) **Temperature profile, \( x_m = T_i \)**

If we make the approximation that the transmittance is independent of temperature then

\[
\frac{dR^c}{dT_i} = \begin{cases} 
\frac{1}{2} \frac{dB_i}{dT_i} \{ \tau_{i-1} - \tau_{i+1} + (1 - \varepsilon_s)(\tau_{i+1}^* - \tau_{i-1}^*) \} & i < I \\
\frac{1}{2} \frac{dB_i}{dT_i} \{ \tau_{i-1} - \tau_i + (1 - \varepsilon_s)(\tau_i^* - \tau_{i-1}^*) \} & i = I.
\end{cases} \]  \hspace{1cm} (A4a)

If the surface is black (i.e. \( \varepsilon_s = 1 \)), Eq. (A4a) becomes

\[ dR^c/dT_i = \frac{1}{2}(dB_i/dT_i)(\tau_{i-1} - \tau_{i+1}) \]  \hspace{1cm} (A5a)

and similarly for Eq. (A4b).

If \( \varepsilon_s < 1 \), but the absorption may be treated as monochromatic (as in the microwave region) and the surface reflection is specular, then

\[ \tau_i^* = \tau_i^2/\tau_i. \]  \hspace{1cm} (A6)

In this case, Eq. (A4a) becomes

\[
\frac{dR^c}{dT_i} = \frac{1}{2} \frac{dB_i}{dT_i} \left\{ 1 + (1 - \varepsilon_s) \frac{\tau_i^2}{\tau_{i-1} \tau_{i+1}} \right\} (\tau_{i-1} - \tau_{i+1}) \]  \hspace{1cm} (A7a)

and similarly for Eq. (A4b).

(b) **Mixing ratio profile, \( x_m = \ln (c_i) \)**

If \( c_i \) is the water vapour mixing ratio at \( p_i \), then

\[ dR/d\ln(c_i) = c_i \frac{dR}{dc_i}. \]  \hspace{1cm} (A8)

Differentiation of Eq. (A2) with respect to \( c_i \) gives

\[ \frac{dR^c}{dc_i} = \varepsilon_s B_s \frac{d\tau_s}{dc_i} + \sum_{k=1}^{I} \frac{1}{2} (B_k + B_{k-1}) \frac{d}{dc_i} \{ \tau_{k-1} - \tau_k + (1 - \varepsilon_s)(\tau_k^* - \tau_{k-1}^*) \}. \]  \hspace{1cm} (A9)

For a black surface, this reduces to

\[ dR^c/dc_i = \{ B_s - \frac{1}{2} (B_I + B_{I-1}) \} d\tau_s/dc_i + \sum_{k=1}^{I-1} \frac{1}{2} (B_{k+1} + B_{k-1}) \frac{d\tau_k}{dc_i}. \]  \hspace{1cm} (A10)
and for a specularly-reflecting surface to

\[
\frac{dR_c}{dc_i} = \left[ \varepsilon_s B_s - \frac{1}{2} (B_f + B_{f-1})(2 - \varepsilon_s) + 2(1 - \varepsilon_s) \tau_s \sum_{k=1}^{l} (B_k + B_{k-1}) \left( \frac{1}{\tau_i} - \frac{1}{\tau_{i-1}} \right) \right] \frac{d\tau_s}{dc_i} + \\
+ \sum_{k=1}^{l-1} \frac{1}{2} (B_{k+1} - B_{k-1}) \left[ 1 + (1 - \varepsilon_s) \frac{\tau^2_k}{\tau^4_k} \right] \frac{d\tau_k}{dc_i}.
\]  
(A11)

\[d\tau_k/\,dc_i\] has been evaluated in different ways for HIRS and MSU. For HIRS we proceed as follows:

\[
\frac{d\tau_k}{dc_i} = \tau_k \frac{d\ln \tau_k}{dc_i} = \frac{d\ln \tau_k}{du_k} \frac{du_k}{dc_i}
\]  
(A12)

where \(u_k\) is the integrated water vapour content from \(p_k\) to space at angle \(\theta\):

\[u_k = (\sec \theta/2g) \sum_{n=1}^{k} (c_n + c_{n-1})(p_n - p_{n-1})
\]  
(A13)

where \(g\) is the acceleration due to gravity. Then

\[
\frac{du_k}{dc_i} = \begin{cases} 
(\sec \theta/2g)(p_{i+1} - p_{i-1}); & i < k \\
(\sec \theta/2g)(p_i - p_{i-1}); & i = k \\
0; & i > k.
\end{cases}
\]  
(A14)

Also, we approximate:

\[\tau_k = \tau_k^w \tau_k^t
\]  
(A15)

where \(\tau_k^w\) is the transmittance of water vapour (including both spectral line and continuum contributions) and \(\tau_k^t\) is the transmittance of all other absorbers. Hence,

\[
d\ln \tau_k/du_k = d\ln \tau_k^w/du_k.
\]  
(A16)

If we now make the approximation that \(d\ln \tau_k^w/du_k\) is independent of the distribution of water vapour along the path, then \(d\ln \tau_k^w/du_k\) may be evaluated numerically from finite differences of the profiles of \(\tau_k^w\) and \(u_k\) at standard pressure levels. These are calculated during the evaluation of Eq. (A2). \(d\tau_k/\,dc_i\) is then found from Eqs. (A12), (A14) and (A16).

For MSU, fewer approximations are required, as the transmittance due to water vapour is calculated as follows:

\[
\tau_k^w = \exp \left\{-\sec \theta \sum_{n=1}^{k} \frac{1}{2} (\alpha_n + \alpha_{n-1})(p_n - p_{n-1})\right\}
\]  
(A17)

where

\[
\alpha_n = w_{1n} T_n + w_{2n} c_n + w_{3n} T_n c_n + w_{4n},
\]  
(A18)

and \(w_{1n}, w_{2n}, w_{3n}, w_{4n}\) are fixed coefficients (see Weinreb et al. 1981). Therefore,

\[
\frac{d\tau_k}{dc_i} = \tau_k \frac{d\ln \tau_k^w}{dc_i} = \tau_k \frac{d\ln \tau_k^w}{d\alpha_i} \frac{d\alpha_i}{dc_i}.
\]  
(A19)
From Eq. (A17)
\[
\frac{d \ln \tau_i}{d \alpha_i} = \begin{cases} 
- \frac{1}{2} \sec \theta (p_{i+1} - p_{i-1}); & i < k \\
- \frac{1}{2} \sec \theta (p_i - p_{i-1}); & i = k \\
0; & i > k
\end{cases}
\tag{A20}
\]
and from Eq. (A18),
\[
d\alpha_i/dc_i = w_{zi} + w_{zi}T_i, \tag{A21}
\]
(c) Surface temperature, \(x_m = T_s\)
\[
dR^c/dT_s = \varepsilon_s \tau_s dB_s/dT_s. \tag{A22}
\]
(d) Surface emissivity, \(x_m = \varepsilon_s\)
\[
dR^c/d\varepsilon_s = B_s \tau_s - \sum_{i=1}^{l} \frac{1}{2}(B_i + B_{i-1})(\tau_i - \tau_{i-1}) \tau_i^2/\tau_i/\tau_{i-1}. \tag{A23}
\]
For specular reflection at the surface, this becomes
\[
dR^c/d\varepsilon_s = B_s \tau_s - \sum_{i=1}^{l} \frac{1}{2}(B_i + B_{i-1})(\tau_i - \tau_{i-1}) \tau_i^2/\tau_i/\tau_{i-1}. \tag{A24}
\]
(e) Surface pressure, \(x_m = p_s\)
\[
\frac{dR^c}{dp_s} = (\varepsilon_s B_s - \frac{1}{2}(B_1 + B_{l-1})) \frac{d\tau_s}{dp_s} + \sum_{i=1}^{l} \frac{1}{2}(B_i + B_{i-1})(1 - \varepsilon_s) \frac{d(\tau_i^2 - \tau_{i-1}^2)}{dp_s}. \tag{A25}
\]
For a black surface,
\[
dR^c/dp_s = (B_s - \frac{1}{2}(B_1 + B_{l-1})) \frac{d\tau_s}{dp_s} \tag{A26}
\]
and for a specularly-reflecting surface,
\[
\frac{dR^c}{dp_s} = (\varepsilon_s B_s - \frac{1}{2}(B_1 + B_{l-1})(2 - \varepsilon_s) + 2(1 - \varepsilon_s) \tau_s \sum_{i=1}^{l} (B_i + B_{i-1}) \frac{\tau_{i-1} - \tau_i}{\tau_i/\tau_{i-1}}) \frac{d\tau_s}{dp_s}. \tag{A27}
\]
(f) The effects of clouds

The radiance from an atmosphere containing complete cover of black cloud at pressure \(p_c\) is calculated by
\[
R^c = B_c \tau_c + \sum_{i=1}^{I_c} \frac{1}{2}(B_i + B_{i-1})(\tau_{i-1} - \tau_i) \tag{A28}
\]
where \(i = I_c\) at \(p = p_c\), \(\tau_c = \tau_f(p_c)\) and \(B_c = B_f(T(p_c))\). Differentiation with respect to \(T_i\) gives
\[
\frac{dR^c}{dT_i} = \begin{cases} 
\frac{1}{2}(dB_i/dT_i)(\tau_{i-1} - \tau_{i+1}); & i < I_c \\
\frac{1}{2}(dB_i/dT_i)(\tau_{i-1} - \tau_i); & i = I_c \\
0; & i > I_c
\end{cases} \tag{A29}
\]
Differentiation with respect to \( c_i \) gives

\[
\frac{dR^0}{dc_i} = \begin{cases} 
\frac{1}{2}(B_i - B_{i-1}) \frac{d \tau_c}{dc_i} + \sum_{k=1}^{i-1} \frac{1}{2}(B_{k+1} - B_{k-1}) \frac{d \tau_k}{dc_i}, & i < I_c \\
\frac{1}{2}(B_i - B_{i-1}) \frac{d \tau_c}{dc_i}, & i = I_c \\
0, & i > I_c.
\end{cases}
\]  
(A30)

For infrared channels, derivatives with respect to all surface variables are zero.

In partly cloudy conditions with fractional cloud cover \( N_c \), the radiance is given by

\[
R = (1 - N_c)R^c + N_c R^o.
\]  
(A31)

Therefore, for all elements \( x_m \) discussed above,

\[
dR/dx_m = (1 - N_c) dR^c/dx_m + N_c dR^o/dx_m.
\]  
(A32)

Also, from Eq. (A31)

\[
dR/dN_c = R^o - R^c
\]  
(A33)

and

\[
dR/dp_c = N_c dR^o/dp_c.
\]  
(A34)

\( dR^o/dp_c \) is evaluated numerically from the overcast radiances calculated for cloud at standard pressure levels through Eq. (A28).

For MSU channels, clouds are assumed to have no effect on the radiances or their derivatives.

**APPENDIX B. CALCULATION OF HUMIDITY ERROR COVARIANCES**

The inversion scheme uses \( \ln(\text{mixing ratio}) \) as its humidity variable. The forecast error statistics currently available are in the form of relative humidity covariances. It is therefore necessary to convert covariances between relative humidity errors at different levels and between relative humidity and temperature errors to respective covariances in terms of \( \ln(\text{mixing ratio}) \). The two variables are related as follows:

\[
c = c_s RH
\]
or

\[
\ln c = \ln c_s + \ln RH
\]  
(B1)

where \( c \) is mixing ratio, \( c_s \) is saturated mixing ratio, and \( RH \) is relative humidity. At a given pressure level, \( c_s \) is a function of temperature only, and so

\[
\delta \ln c = \frac{d \ln c}{d RH} \delta RH + \frac{d \ln c_s}{dT} \delta T = \frac{\delta RH}{RH} + \frac{d \ln c_s}{dT} \delta T.
\]  
(B2)

Using Eq. (B2), we calculate the required covariances between variables at pressure levels \( i \) and \( j \):

\[
\delta \ln c_i, \delta T_j = \frac{1}{RH_i} \delta RH_i \delta T_j + \left( \frac{d \ln c_s}{dT} \right)_i \delta T_i \delta T_j
\]  
(B3)
\[
\delta \ln c_i, \delta \ln c_j = \frac{1}{RH_iRH_j} \delta RH_i, \delta RH_j + \frac{1}{RH_i} \left( \frac{d \ln c_k}{dT} \right)_i \delta RH_j, \delta T_j + \frac{1}{RH_j} \left( \frac{d \ln c_k}{dT} \right)_j \delta RH_i, \delta T_i \]
\[
+ \left( \frac{d \ln c_i}{dT} \right)_i \delta RH_j, \delta T_j - \left( \frac{d \ln c_j}{dT} \right)_j \delta RH_i, \delta T_i.
\]  

(B4)

In this work we have evaluated Eqs. (B3) and (B4) using typical mean values of \(d \ln c_a/dT = 0.08 \text{ K}^{-1}\) and of \(RH = 0.70\).

REFERENCES

Bengtsson, L.  
1985  Medium-range forecasting at the ECMWF. *Adv. Geophys.*, 28B, 3-54

Chedin, A., Scott, N. A., Wahiche C. and Moulinier, P.  

Durand, Y. and Juvanon du Vachat, R.  

Eyre, J. R.  

1983  ‘Comparisons between measured and simulated TOVS brightness temperatures’. Met. O.19 Branch Memorandum 68, Meteorological Office, Bracknell

1984  ‘User guide to TOVSRAD. A program for calculating synthetic HIRS-2 and MSU equivalent blackbody temperatures’. Met. O.19 Branch Memorandum 75, Meteorological Office, Bracknell


1989  Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation: application to TOVS data. *ibid.*, 115, 1027–1037

Eyre, J. R. and Menzel, W. P.  

Eyre, J. R. and Watts, P. D.  

Eyre, J. R., Watts, P. D., Turner, J. and Lorenc, A. C.  


Halem, J., Kalnay, E., Baker, W. E. and Atlas, R.  
1982  Hoffman, R. N.  


Huang, H.-L. A. and Smith, W. L.  


Purser, R. J. 1983 ‘Accumulation of a set of radiosonde profiles for use with satellite sounding data from the European and N. Atlantic areas’. Met. O.19 Branch Memorandum 71, Meteorological Office, Bracknell


