Assimilation of IASI and AIRS data: Information Content and Quality Control.

A.D. Collard.
Met Office.
Bracknell, U.K.

Summary.

This paper describes the assimilation of data from the next generation of infrared sounding instruments into numerical weather prediction models. After briefly introducing the instruments in question, the expected retrieval performance of one of them, IASI, is examined in relation to a current infrared sounder, HIRS. An example of their relative performances in characterising a meteorological significant temperature structure is shown. Possible methods for efficient assimilation of these data are explored followed by a study on how correlated radiance errors can affect retrieval performance. Finally cloud detection and quality control methods are discussed.

Introduction.

This paper continues the discussion started in Saunders (2001) of the issues surrounding the assimilation of high spectral resolution infrared sounder data into numerical weather prediction (NWP) models. While Saunders explored the problems involved with calculating radiances from known atmospheric states (the forward problem) this paper explores how information on the state of the atmosphere may be obtained from this new class of observations (the inverse problem) and used to improve the initial atmospheric state for numerical weather prediction.

The advanced sounders to be launched into polar orbit in the next few years are summarised in Table 1. AIRS is a research, rather than an operational, instrument and also differs from the other instruments being considered here in that it is a grating spectrometer rather than an interferometer. However, there is significant interest in the NWP community in using AIRS data as the spectral resolution, information content and data volume will be similar to the operational Fourier transform spectrometers that will follow.

In addition to the polar orbiting instruments listed in Table 1, one may add the Geostationary Imaging Fourier Transform Spectrometer (GIFTS) which will fly as an experimental mission as part of the NASA New Millennium Program.

The advantages of advanced infrared sounders come from the higher spectral resolution and higher number of channels observed. Figure 1 compares a portion of the spectrum as it would be observed by IASI with the instrument spectral response functions (ISRF) of the HIRS[‡] channels in this region. It is clear that the effect of the HIRS ISRF is to smear out

[‡] The High-resolution Infrared Radiation Sounder — part of the ATOVS instrument package on the current operational NOAA polar-orbiting meteorological satellites.

Parameter	Advanced Sounder		
	AIRS	IASI	CrIS
Instrument type	Grating Spectrometer	Interferometer	Interferometer
Platform	Aqua	MetOp-1	NPP & NPOESS
Satellite Agency	NASA/JPL	EUMETSAT/ CNES	NOAA IPO
Spectral range (cm^{-1})	649–1135; 1217–1613; 2169–2674	Contiguous 645–2940	650–1095; 1210–1750; 2155–2550
Number of channels	2378	8461	~1300
Unapodised spectral resolving power	1000–1400	2000–4000	900–1800
Spectral sampling	$\sim \nu/2400$	0.25 cm^{-1}	0.625/1.25/2.5cm^{-1}
Spatial footprint (km)	13.5	12	14
Nominal launch date	May 2001	Dec 2003	2006 & 2009

Table 1.
Summary of Instrument Characteristics (from Saunders, 2001).

information from wavelengths with quite different atmospheric transmissions. The effect of this is that the HIRS weighting functions are broader than the IASI ones and this in turn, together with the effect of the larger number of IASI channels, results in lower vertical resolution for HIRS.

The assimilation of this new type of satellite data into NWP models provides a new set of challenges. This primarily arises from the shear volume of data that the instrument produces (e.g., the AIRS instrument will produce ~35Gbytes/day) and the need to use the observations efficiently. One must also consider how to treat the correlated errors between the thousands of instrumental channels in the most efficient way.

This paper explores some of the problems that are likely to be encountered in assimilating these data. It starts with a summary of the basic theory of data retrieval and assimilation and considers the theoretical performance that we can expect for this class of instrument. Next, possible methods for reducing the volume of data that needs to be assimilated are explored, with particular attention being paid to channel selection. The effect of correlated observation errors on retrieval performance is then examined. Finally possible strategies for cloud detection and quality control are considered.

Basic Retrieval Theory.

The retrieval performance of sounding instruments can be quantified in a variety of ways. In this section, expressions for retrieval accuracy, vertical resolution, information content (entropy reduction) and degrees of freedom for signal (the number of independent pieces of information retrieved) are derived plus equations relating a given atmospheric perturbation to the retrieved profile.

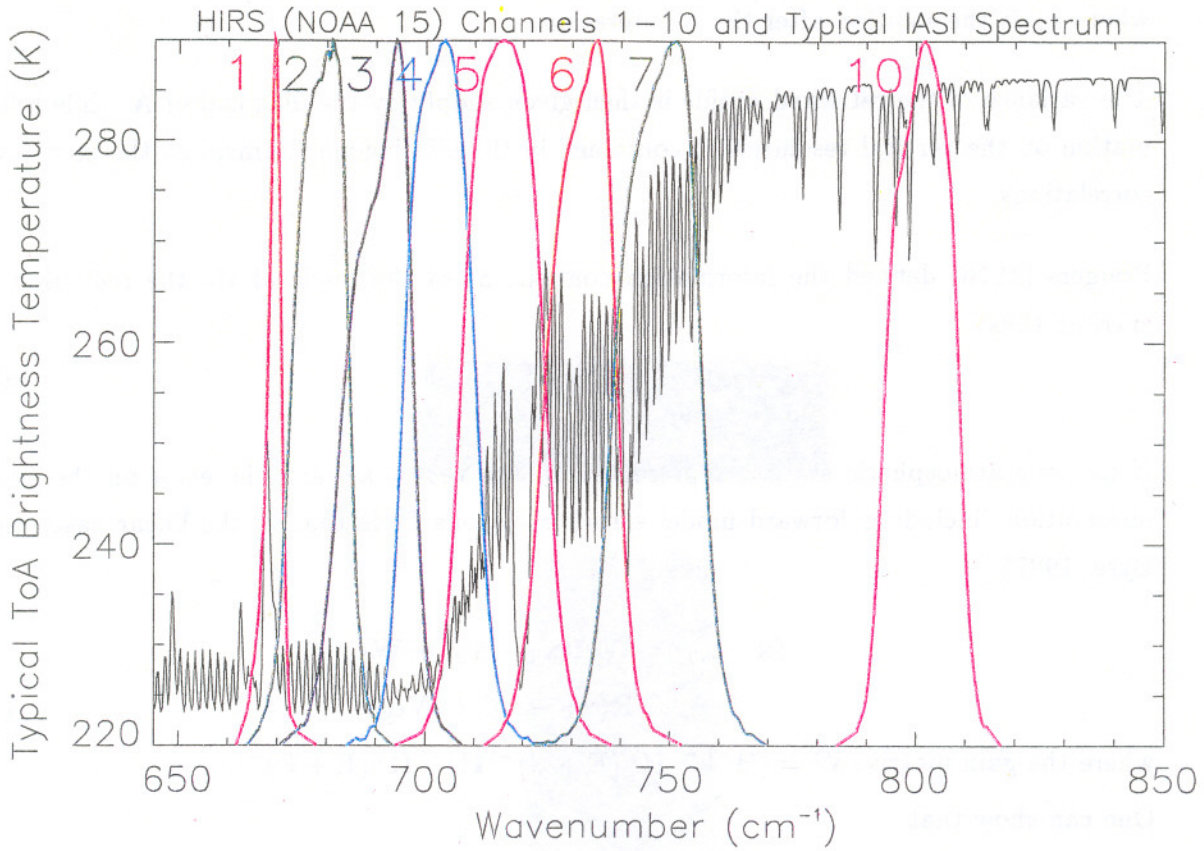

Fig. 1. A comparison of a portion of a typical spectrum as would be observed by IASI with the instrument spectral response functions (ISRF) for HIRS in this region. The smoothing out of spectral features by the HIRS ISRF is clear.

The best estimate, $\hat{\mathbf{x}}$, of the atmospheric state, $\mathbf{x}$, is in general found by minimising the cost function, $J(\mathbf{x})$, where

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_0) + \frac{1}{2}(\mathbf{y}-\mathbf{y}(\mathbf{x}))^T(\mathbf{E}+\mathbf{F})^{-1}(\mathbf{y}-\mathbf{y}(\mathbf{x})) \quad (1)$$

where the observations, $\mathbf{y}$, have observational error covariance $\mathbf{E}$ and forward model error covariance $\mathbf{F}$; $\mathbf{B}$ is the error covariance matrix of the *a priori* measurements $\mathbf{x}_0$; and $\mathbf{y}(\mathbf{x})$ is the observed radiance that would result for a given atmospheric state $\mathbf{x}$.

For strictly linear problems, the solution vector, $\hat{\mathbf{x}}$, and associated error covariance, $\mathbf{A}$, is given by the optimal estimation method of retrieval (Rodgers, 1976) where

$$\hat{\mathbf{x}} = (\mathbf{B}^{-1} + \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H})^{-1}(\mathbf{B}^{-1}\mathbf{x}_0 + \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{y}) \quad (2a)$$

and

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H})^{-1}. \quad (2b)$$

Here, $\mathbf{H} = \nabla_x\mathbf{y}(\mathbf{x})$ is the matrix of instrument weighting functions.

Most realistic problems, however, are not strictly linear in which case one needs alternative methods to seek the minimum of the cost function. One method of doing this, suitable for moderately non-linear problems, is through Newtonian iteration:

$$\mathbf{x}_{n+1} = \mathbf{x}_0 + (\mathbf{B}^{-1} + \mathbf{H}_n{}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H}_n)^{-1}\mathbf{H}_n{}^T(\mathbf{E}+\mathbf{F})^{-1}[\mathbf{y}-\mathbf{y}(\mathbf{x}_n)-\mathbf{H}_n(\mathbf{x}_0-\mathbf{x}_n)] \quad (3)$$

where $\mathbf{x_n}$ is the solution after the n^{th} iteration.

The variance of the retrieved profile is then given simply by the diagonal of $\mathbf{A}$ while information on the vertical resolution is contained in the off-diagonal terms, i.e., the inter-level correlations.

Rodgers (1976) defined the information content, S, of the retrieval via the reduction in entropy thus:

$$S = -\frac{1}{2}\log_2(\mathbf{AB}^{-1}) \tag{4}$$

If the true atmospheric state is represented by the vector $\mathbf{x_T}$ and the error on the given observation (including forward model error) is $\epsilon_\mathbf{y}$, one finds that in the linear case (e.g., Eyre, 1987)

$$\begin{aligned}(\hat{\mathbf{x}} - \mathbf{x_0}) &= \mathbf{WH}(\mathbf{x_T} - \mathbf{x_0}) + \mathbf{W}\epsilon_\mathbf{y} \\ &= \mathbf{R}(\mathbf{x_T} - \mathbf{x_0}) + \mathbf{W}\epsilon_\mathbf{y}\end{aligned} \tag{5}$$

where the gain matrix, $\mathbf{W} = (\mathbf{B}^{-1} + \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}$.
One can show that

$$\mathbf{R} = \mathbf{I} - \mathbf{AB}^{-1} \tag{6}$$

where $\mathbf{I}$ is the identity matrix.

Further inspection of Eqn. 5 reveals that the expected retrieved profile for given *a priori* information is

$$E[(\hat{\mathbf{x}} - \mathbf{x_0})] = \mathbf{R}(\mathbf{x_T} - \mathbf{x_0}) \tag{7a}$$

with the noise on this solution being

$$E[(\hat{\mathbf{x}} - \mathbf{x_0})(\hat{\mathbf{x}} - \mathbf{x_0})^T] = \mathbf{W}(\mathbf{E}+\mathbf{F})\mathbf{W}^T \tag{7b}$$

$\mathbf{R}$ is known as either the *Averaging Kernel* (Backus and Gilbert, 1970) or the *Model Resolution Matrix* (Menke, 1984)‡. The rows of $\mathbf{R}$ give the contributions from each level of the $\mathbf{x_T} - \mathbf{x_0}$ profile (i.e., the difference between the truth and *a priori* profiles) to a given level in the $\hat{\mathbf{x}} - \mathbf{x_0}$ profile (the difference between the retrieved and *a priori* profiles). Conversely, the columns tell one how a perturbation in a single level of $\mathbf{x_T} - \mathbf{x_0}$ is distributed over $\hat{\mathbf{x}} - \mathbf{x_0}$.

Rodgers (2000) has shown that the number of degrees of freedom for signal (i.e., the number of separate information eigenvectors provided by the measurements) in the retrieval is given by the trace of the resolution matrix, $Tr(\mathbf{R})$.

‡ $\mathbf{R}$ here should not be confused with the common usage of $\mathbf{R}$ to indicate the total observational error $(\mathbf{E}+\mathbf{F})$.

Purser and Huang (1993) use this property of the resolution matrix to define vertical resolution in terms of effective data density, ρ, where the data density, ρ_i, for layer i of thickness Δz_i is given by

$$\rho_i = R_{ii}/\Delta z_i \tag{8}$$

The vertical resolution by this definition is thus $1/\rho_i$ for level i. Purser and Huang go on to define a smoothing operator based on the correlation structure of $\mathbf{R}$ to eliminate the worst of the level-to-level fluctuations in ρ, but for the examples in this document this does not appear to be a problem and the simpler definition is preferred.

Rodgers (1990) shows that the analysis error covariance, $\mathbf{A}$, can be thought of as the sum of two error covariances, $\mathbf{A}_N$ and $\mathbf{A}_M$ corresponding to the contributions from the null-space and measurement errors respectively ($\mathbf{A}_M$ is hereinafter referred to as the "propagated measurement error"). For the optimal estimation method,

$$\mathbf{A}_N = (\mathbf{B}^{-1} + \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H})^{-1}\mathbf{B}^{-1}(\mathbf{B}^{-1} + \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H})^{-1} \tag{9a}$$

and

$$\mathbf{A}_M = (\mathbf{B}^{-1} + \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H}(\mathbf{B}^{-1} + \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H})^{-1}. \tag{9b}$$

The null-space error arises from the fact that the retrieved profile is a smoothed version of the truth with the *a priori* profile providing those components where the observations do not add information. Thus the analysis error covariance, $\mathbf{A} = \mathrm{E}\left[(\hat{\mathbf{x}} - \mathbf{x}_t)(\hat{\mathbf{x}} - \mathbf{x}_t)^T\right]$, has a component which accounts for the smoothing error.

Comparing Eqns 7b and 9b, one can see that the deviation of the retrieved profile from the expected value is given by the propagated measurement noise.

Calculation of Retrieval Properties for IASI.

In this section we shall compare the expected performance of a currently operational infrared sounding instrument, HIRS, with that expected of IASI.

These investigations require the analysis covariance matrix, $\mathbf{A}$, to be calculated from the *a priori* data error covariance and the measurements' error covariance using appropriate IASI weighting functions, $\mathbf{H}$. The weighting functions used were calculated for a mid-latitude summer case using the IASI fastmodel, RTIASI (Matricardi and Saunders, 1999).

For apodised (Level 1c) IASI data, the instrumental errors will be correlated between nearby channels and so the observational error covariance matrix is approximated by a penta-diagonal matrix. The noise levels used are the current (Oct. 2000) best estimate of the instrument noise on launch and were provided by F. Cayla of CNES. The interchannel error correlations are calculated (using the method of Amato *et al.*, 1998) assuming a Gaussian apodisation function which is truncated at an optical path difference of 2cm and which

produces an instrument spectral response function with a full width at half maximum of 0.5cm^{-1}.

In addition to the instrumental noise, a greatly simplified approximation to the forward model noise is included in the form of an extra $(0.2\text{K})^2$ being added to the diagonal of the matrix. As will be seen later (see also Saunders (2001) and Sherlock(2000b)), this representation of forward model noise as uncorrelated between channels is not realistic.

The HIRS noise values come from Table 3.2.1-2 of the NOAA KLM Users Guide which can be found on-line at `http://perigee.ncdc.noaa.gov/docs/klm/html/c3/sec3-2.htm#t321-2` to which a forward modelling error of 0.2K is added in quadrature (i.e., the same forward model error is assumed as for IASI).

The *a priori* covariance matrix used is an ECMWF background error covariance on the 43 levels used by RTIASI.

Expected Retrieval Performance.

The expected retrieval errors (i.e., the square root of the diagonals of the matrix **A** defined in Eqn. 2b.) for IASI and HIRS are compared in Figure 2 for temperature and humidity. In both cases, it can be seen that the retrieval accuracy that we can expect from IASI is a significant improvement over that from HIRS.

Similarly, Figures 3a and 3b show that the vertical resolution (defined in Eqn. 8) is significantly improved for the case of IASI. Indeed for humidity retrievals the vertical resolution is so high that it is of the order of the spacing in the levels in the radiative transfer/retrieval scheme and thus motivates the production of future fast radiative transfer models with higher vertical resolution.

The number of degrees of freedom for signal for the IASI retrievals in this case is 26.8 which is a five-fold improvement on the HIRS value of 5.5. Similarly the information content as defined in Eqn. 4 is 84.2 bits for IASI compared to 12.4 for HIRS.

One further test of instrument performance is to investigate how a given structure in the atmosphere will be interpreted in the retrieval scheme. To do this one defines a given deviation from the *a priori* profile and then uses Eqn. 7a to determine the expected retrieved profile. Eqn. 7b is then used to determine the expected deviation about this expected retrieved profile.

Figures 4 shows this for a temperature perturbation that was used in an earlier study by Prunet *et al.* (1998). This profile was one of those that Prunet *et al.* extracted from a study by Rabier *et al.* (1996) which identified error structures in the initial atmospheric state of an NWP model run that would grow rapidly with time and would result in erroneous forecasts.

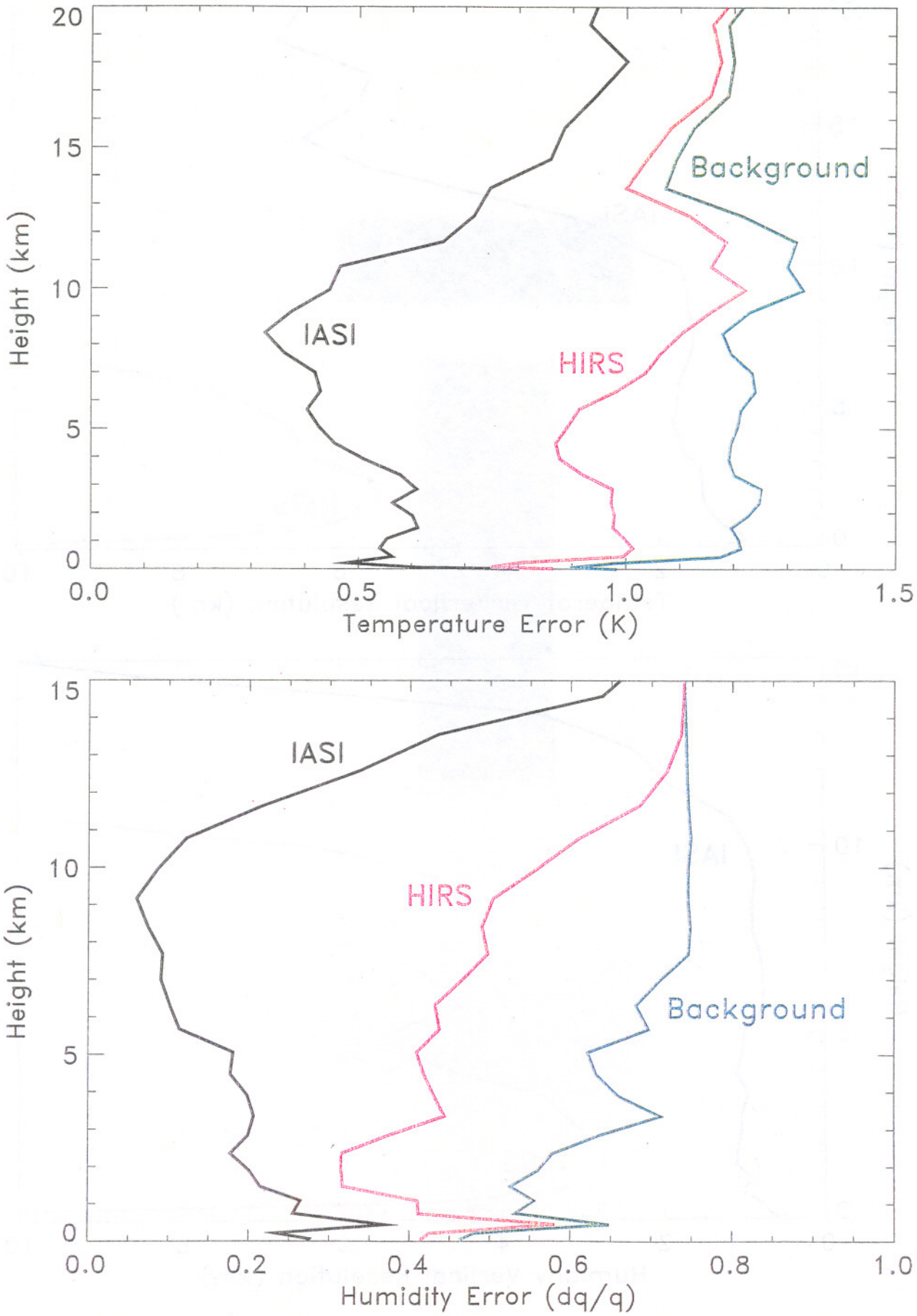

Fig. 2. IASI and HIRS temperature and humidity retrieval accuracies.

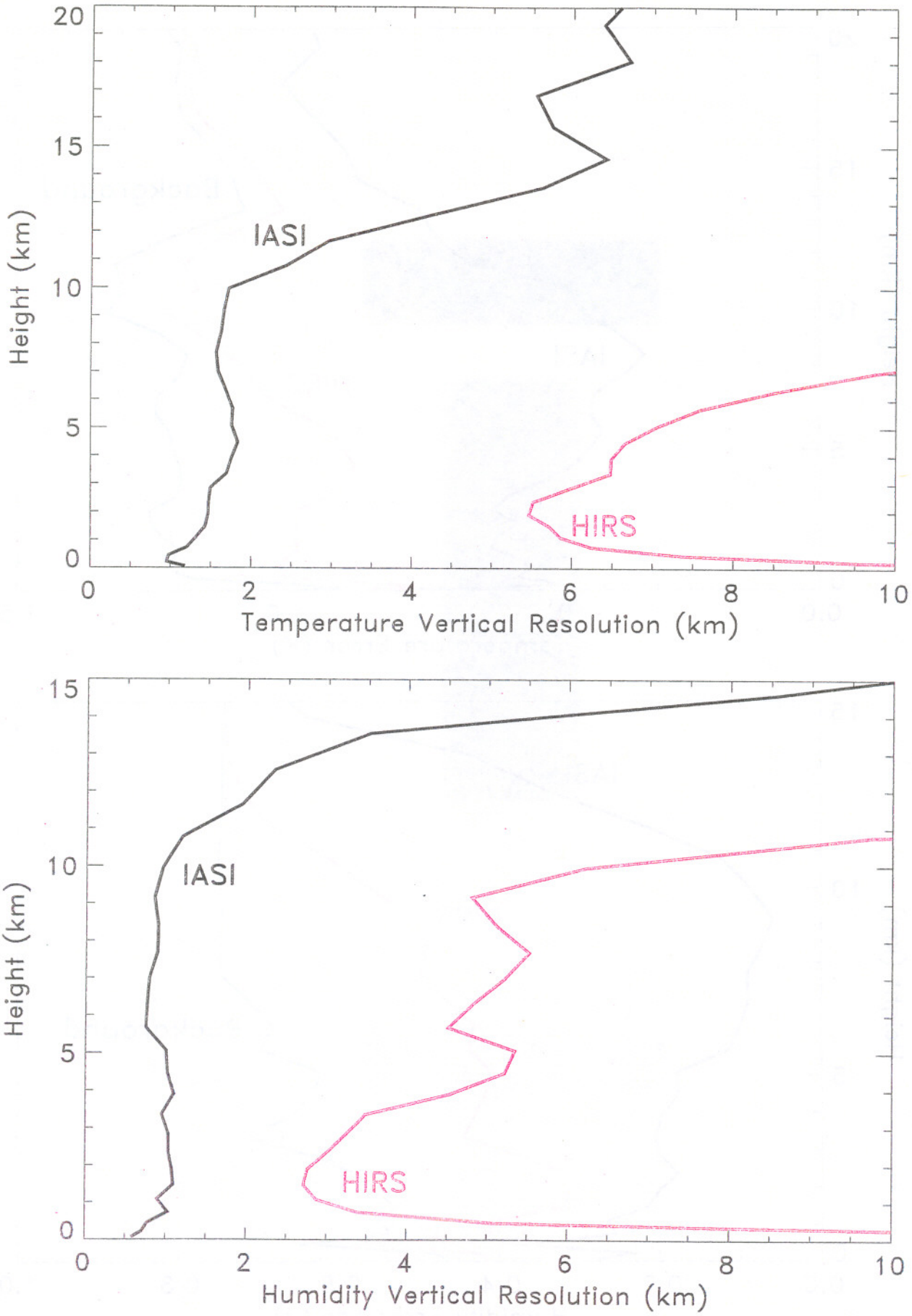

Fig. 3. IASI and HIRS temperature and humidity retrieval vertical resolution.

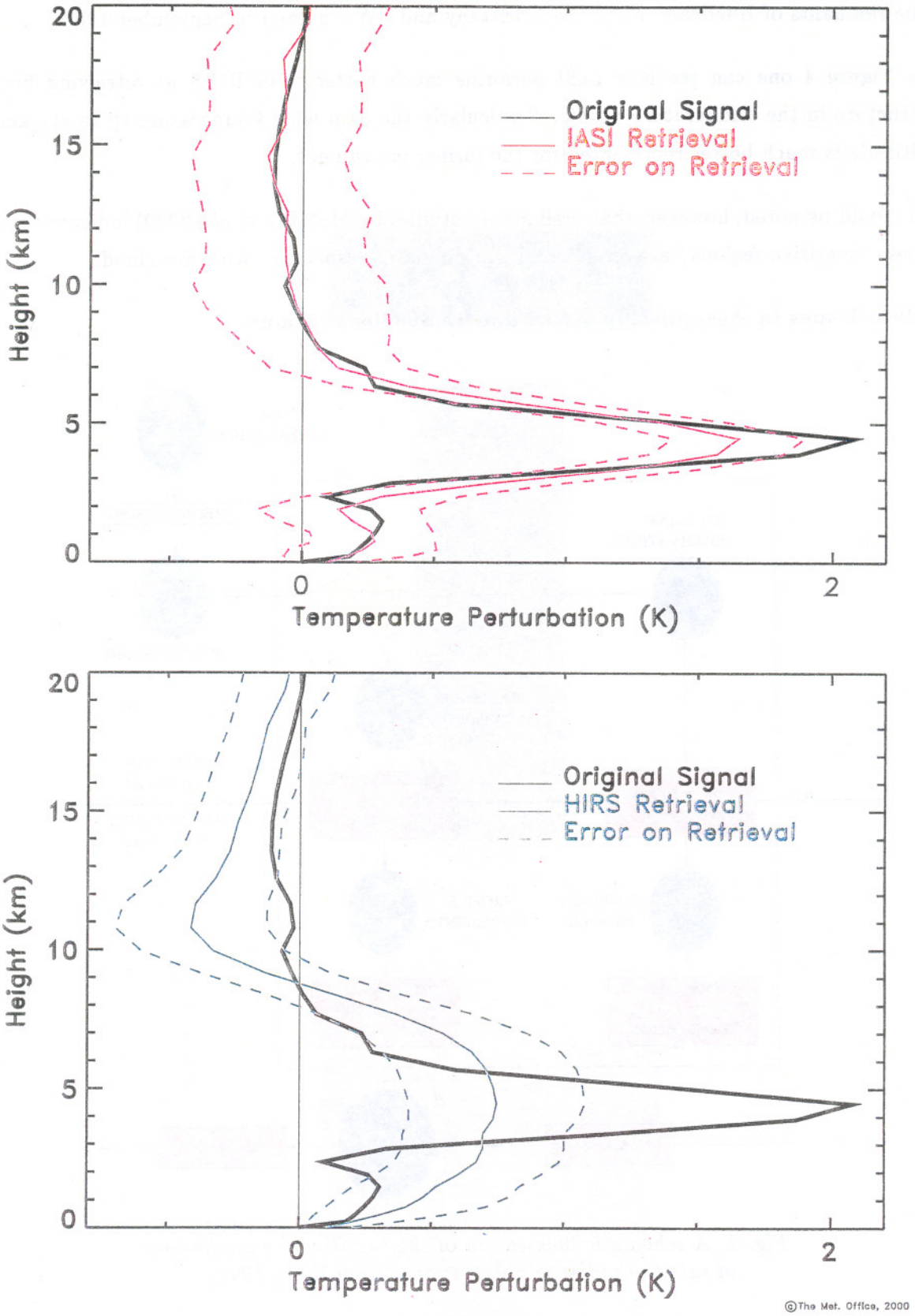

Fig. 4. One of the "Rabier curves" used by Prunet et al. and the corresponding expected profiles that would be retrieved by the HIRS and IASI instruments as calculated via the resolution matrix plus the expected errors on those profiles (see text for discussion of these errors).

These error structures were important in the failed forecasting of the reintensification of the remnants of Hurricane Floyd over Brittany and SW England in September 1993.

In Figure 4 one can see how IASI performs much better than HIRS at retrieving fine structure in the temperature profile. Particularly the 3km wide temperature spike at 4km altitude is much better represented for the former instrument.

It should be noted, however, that preliminary studies by McNally *et al.* (2000) indicate that these "sensitive regions" are often cloud contaminated, especially with low cloud.

Main Issues in Assimilating AIRS and IASI Observations.

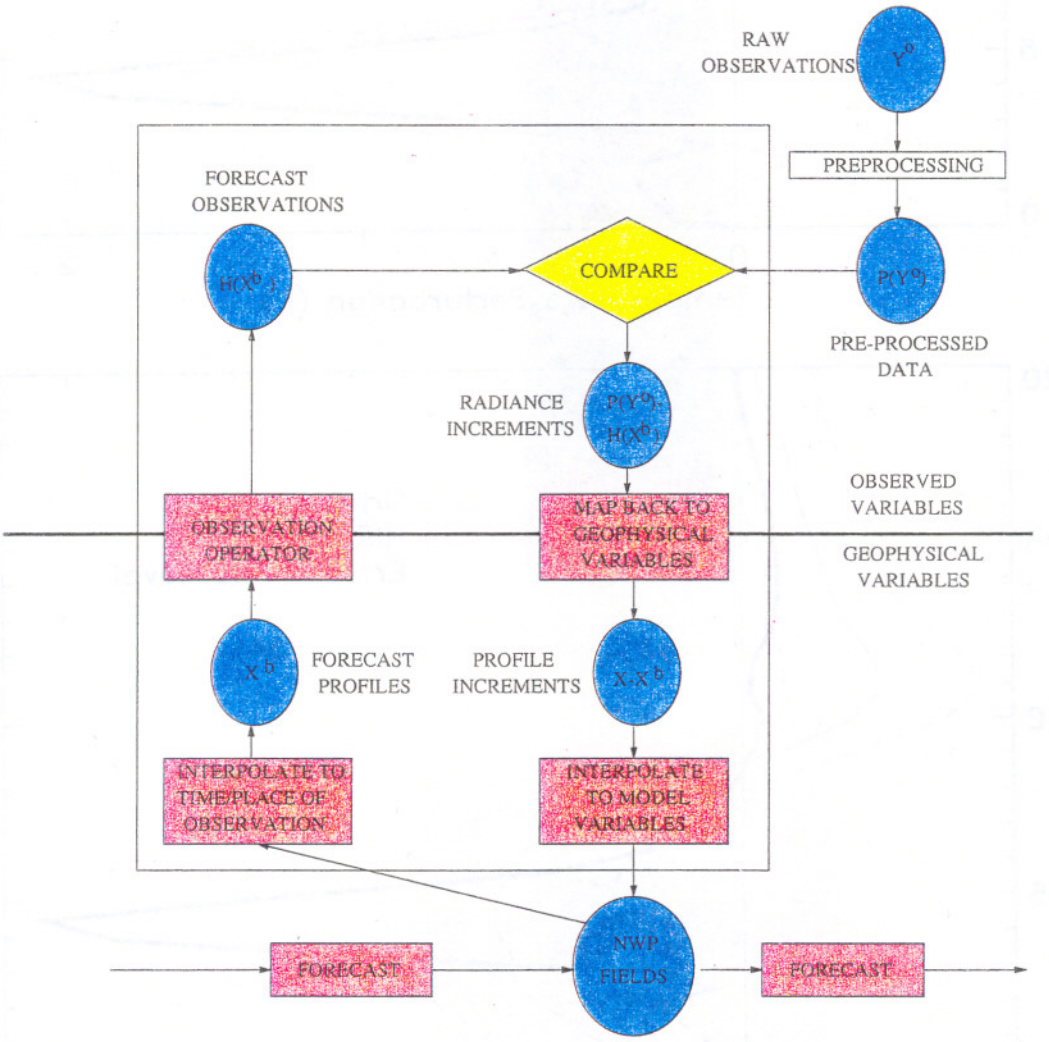

Fig. 5. A schematic illustration of the variational assimilation of satellite radiance observations (from Eyre, 1997).

Figure 5 illustrates how observations are assimilated into the NWP fields in a 3-Dimensional Variational Analysis (3DVar) framework (Eyre, 1997). *A priori* data from the model forecast fields is optimally combined (Eqn. 3) with preprocessed radiance data to produce a new forecast field that is an improved estimate of the true atmospheric state. The preprocessing

of the observations includes cloud-detection and/or cloud-clearing plus other quality control processes. This stage may also include a one-dimensional retrieval stage which allows for the determination of parameters (e.g., skin temperature) which may be required by the radiative transfer code but which are not part of the 3DVar control variable. When considering the issues specific to the assimilation of advanced IR sounder data, one must consider both the 1DVar (where the state vector has ~ 100 elements) and 3DVar stages (where the equivalent vector is $\sim 10^7$ elements long).

The most obvious property of the new advanced IR sounder data is the very large number of channels — over 1000 for all the instruments being considered here. This is compared with the 19 infrared channels of HIRS (or the total of 40 for ATOVS).

One can theoretically, given very large computational resources, try to assimilate all the radiance data available. This is not only very computationally intensive (especially if one has to deal with non-diagonal observational error covariance matrices — see later) but is also very inefficient as the number of independent pieces of information contained in the data is much fewer than the number of channels.

The simplest way of reducing the size of the problem is to use only a subset of channels. One method to do this is to be explored in the next section. Typically 100–1000 channels would be selected.

A more sophisticated version of the channel selection method is to select a subset of channels and then find nearby channels with similar properties and add them together. Known as super-channelling, this method improves over the simple channel selection method through the reduction in the instrumental noise. However, it is probable that in co-adding similar but not identical channels the combined channel's Jacobian will not be as sharp as each individual channel. One must also consider that the fast-modelling of super-channels (at least as it is currently done) requires that a single Planck function is assumed for the whole range of channels in the super-channel and this may also introduce errors if the spectral range is too large.

One may extend the idea of super-channels further and consider a more general set of linear combinations of channels, hereinafter referred to as pseudo-channels. This may be desirable for one of two reasons:

Firstly, one may attempt to find a co-ordinate system in measurement space so that the new observational error covariance matrix is diagonal. One does this by determining the eigenvectors, $\mathbf{X}$, of the $(\mathbf{E}+\mathbf{F})$ matrix (the eigenvalues of which lie along the diagonal of the diagonal matrix $\mathbf{\Lambda}$) and then using pseudo-channels represented by the vector $\mathbf{z} = \mathbf{X}^T\mathbf{y}$. Eqn. 3 can then be rewritten

$$\mathbf{x_{n+1}} = \mathbf{x_0} + (\mathbf{B^{-1}} + \mathbf{\Theta_n}^T\mathbf{\Lambda^{-1}}\mathbf{\Theta_n})^{-1}\mathbf{\Theta_n}^T\mathbf{\Lambda^{-1}}[\mathbf{z} - \mathbf{z}(\mathbf{x_n}) - \mathbf{\Theta_n}(\mathbf{x_0} - \mathbf{x_n})] \qquad (10)$$

where $\mathbf{\Theta_n} = \mathbf{X}^T\mathbf{H_n}$. One may then select only those pseudo-channels with small variances (given by the elements of $\mathbf{\Lambda}$). If one cannot make a large reduction in the number of pseudo-channels in this way, this method is rather limited as one still has to calculate and store $\mathbf{X}$ which has the same dimensions as $(\mathbf{E}+\mathbf{F})$.

Alternatively, one may use *a priori* data (either in the form of a climatological dataset or a previous set of observations) to produce a set of empirical orthogonal functions (EOFs) of the covariance of the instrument spectra. Those EOFs with small amplitudes may be then discarded as not being representative of atmospheric signals and thus reduce the data volume. A side effect of this is that in the act of discarding the small amplitude EOFs one is also greatly decreasing the noise in the spectrum (if one were to reform the spectrum from the remaining EOFs). This noise reduction is simply a result of the regularisation imposed by the *a priori* data and a similar regularisation will occur during the assimilation process with the NWP forecast fields being the *a priori* data. It should also be noted that this noise will tend to be highly correlated between channels.

All of the above methods of reducing the observational data volume may be used in the 1DVar preprocessing stage, the 3DVar stage or both. However, one may also assimilate the retrieved profiles obtained at the 1DVar stage into the model at the 3DVar stage. This method has been used at most NWP centres for TOVS observations, but many are now adopting the direct assimilation of radiances explained above.

The two main difficulties with the assimilation of profile data are that the *a priori* information used in the retrieval of the profile will also be assimilated in the 3DVar stage and that the retrieval errors are highly correlated between levels and this should be allowed for. In many cases retrieved profiles have been treated like degraded radiosonde profiles with errors uncorrelated in the vertical which results in a sub-optimal assimilation. Theoretical work by Joiner and da Silva (1998) and Joiner and Dee (2000) has further investigated the assimilation of retrieved profiles. The latter paper tentatively concludes that assimilation of retrieved radiances can be close to optimal providing that the *a priori* estimate used in the retrieval step is the same as that used in assimilation and that the correct error covariance for the retrieval is used in the assimilation step.

It is anticipated that — at the Met Office at least — radiances rather than retrievals will be assimilated and that the data volume will be reduced at both the 1DVar and 3DVar stages by channel selection. In the next section channel selection will be described.

Channel Selection.

Many different methods have been suggested for channel selection. The method described here was suggested by Rodgers (1996). In experiments comparing various channel selection techniques in the context of variational assimilation for NWP, Rabier (2000) found this to

be significantly better than the others.

Rodgers's method involves choosing a figure of merit — either information content (Eqn. 4) or Degrees of Freedom for Signal (DFS = $Tr(\mathbf{R})$) — and determine which channels have the biggest impact on this.

This is done by starting with the background error covariance matrix, $\mathbf{B}$, and then evaluating $\mathbf{A}$ through Eqn. 2b for each individual channel. After all the channels have been tried the one that results in the highest figure of merit is kept for the next iteration. For the next iteration two channels are used, one of them being the best one from the previous iteration. This process is repeated until the required number of channels have been chosen.

Rodgers noted that if one can assume that the observational and forward model errors are uncorrelated between channels, one can greatly speed up this calculation by determining the new solution error covariance, $\mathbf{A}_i$, from the previous one, $\mathbf{A}_{i-1}$ via the relation:

$$\mathbf{A}_i = \mathbf{A}_{i-1}\left\{\mathbf{I} - \mathbf{h}_i(\mathbf{A}_{i-1}\mathbf{h}_i)^T/[1 + (\mathbf{A}_{i-1}\mathbf{h}_i)^T\mathbf{h}_i]\right\} \tag{11}$$

Here $\mathbf{h}_i$, is the Jacobian of the channel being tested divided by the standard deviation of the observation plus forward model error.

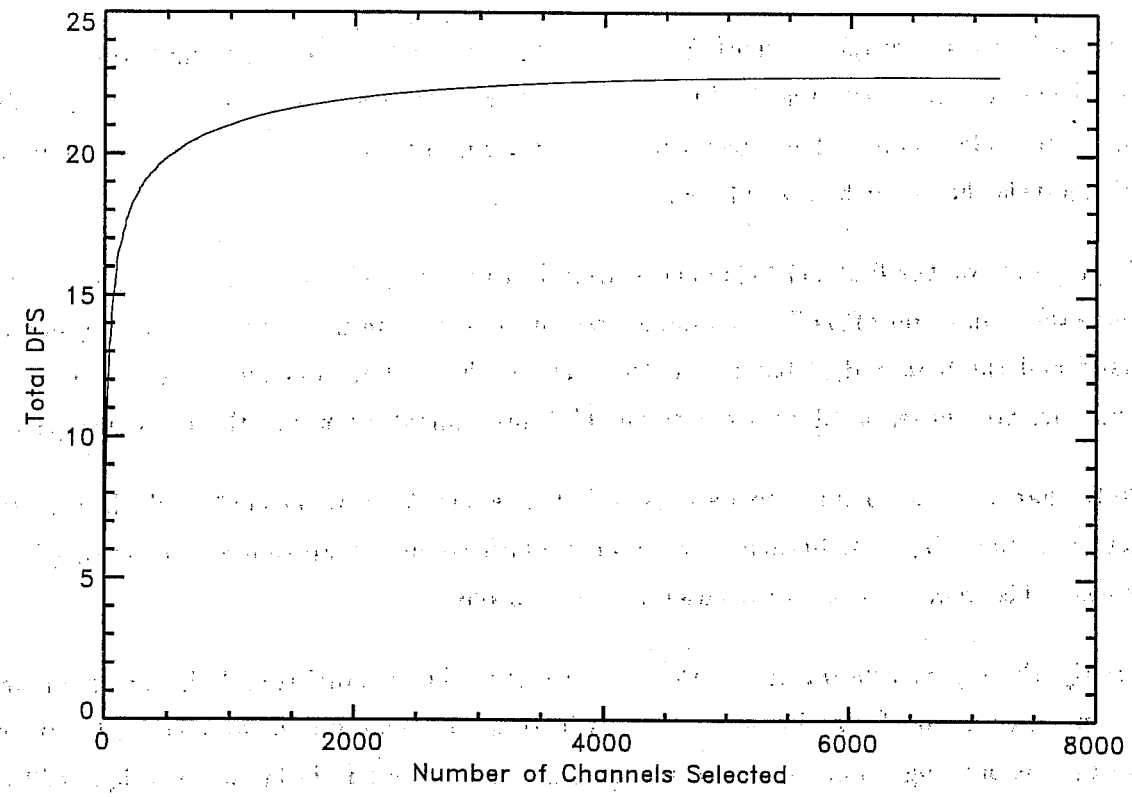

Fig. 6. Degrees of freedom for signal versus number of channels used for a typical IASI case. Much of the impact of a full retrieval can be obtained with less than 1000 channels.

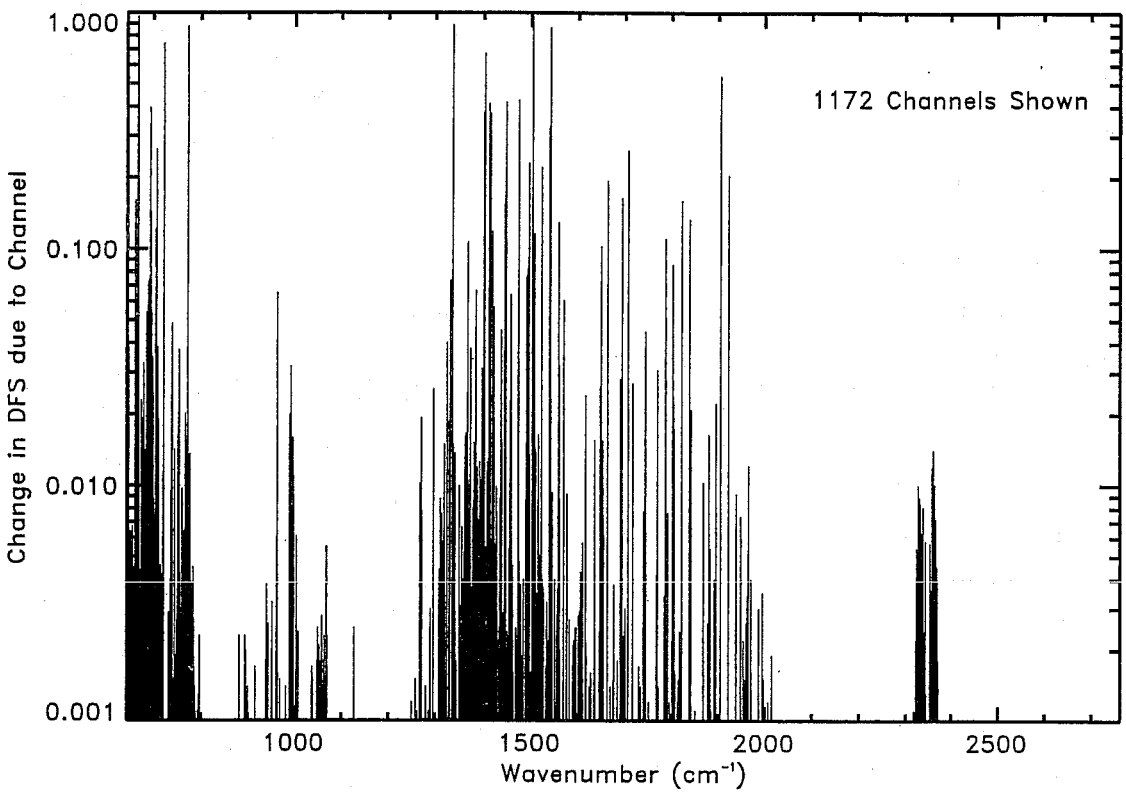

Fig. 7. The positions and impact, in terms of change in degrees of freedom for signal, of the first 1172 channels chosen for a typical IASI situation.

Figure 6 shows how the figure of merit — in this case DFS — typically improves with the number of channels selected for IASI. The number of DFS is close to the maximum value after 500–1000 channels have been chosen resulting in a 8–16-fold reduction in the number of channels that need be assimilated.

Figure 7 shows the first 1172 channels chosen in this case together with the impact on the retrieval — in terms of DFS — of adding each of these channels. Channels in the 15μm CO_2 band and the 6.3μm H_2O band tend to be chosen first while channels in the 4.3μm CO_2 band are not chosen until later due to the high instrumental noise at these wavelengths.

Note that the first channel to be chosen in this example is at 667cm^{-1}, which is in the centre of the CO_2 ν_2 Q-branch — a channel which sounds stratospheric and mesospheric levels. This channel is the first chosen for two reasons:

Firstly, the *a priori* knowledge at the altitudes that this channel sounds is very poor and so any information from the observations will result in a large improvement. This is important as one might find that even though some levels are poorly known, it might still be more important to get more information on other, more accurately known, regions of the atmosphere (e.g., levels where errors in the knowledge of the stage of the atmosphere can result in large forecast errors as described by Rabier (1996)). Therefore, one might wish to modify the **B** matrix used in the channel selection calculations to reflect not only where

knowledge is worst but also where knowledge is most desired.

Secondly, the assumed forward model error for all of the channels in this example is the same (0.2K), which is clearly unrealistic. One might expect the Q-branch to be less well modelled in which case the assumed forward model error might be far too low. In this situation, it is less likely that this channel would be chosen so soon. In other words, one must endeavour to make sure that the observational error covariances used are realistic.

This methodology works on a case by case basis, i.e., one can calculate the optimal channels to choose for a given atmospheric state. One can also estimate the optimal set of channels for all possible atmospheric states by extending the calculations described above so that the state vector assumed is a concatenation of state vectors from a range of representative atmospheric situations (Rodgers, *priv. comm.*). One may tune this process further, so that more important atmospheric situations (for NWP) are given more weight. Of course, it is impossible to include all possible atmospheric states in such a calculation and more work needs to be done to determine how to allow for all situations that may be important.

The Effect of Correlated Observation Errors.

A complication in the use of high-spectral resolution infrared sounder observations is that the $(\mathbf{E} + \mathbf{F})$ matrix is thousands of elements square and thus has millions of elements in total. This makes its manipulation unwieldy. If one can assume that the $(\mathbf{E} + \mathbf{F})$ matrix is diagonal, however, the problems associated with inverting large matrices would disappear.

In reality the $(\mathbf{E} + \mathbf{F})$ matrix is not diagonal. The forward model errors, $\mathbf{F}$, are certainly correlated to some degree from channel to channel (Sherlock, 2000a), while for apodised interferometer observations the instrumental error covariance matrix, $\mathbf{E}$, will always have some non-zero near-diagonal elements.

The effect of another type of correlated error — that of the inclusion of undetected cloud in the observations — is explored in this paper for illustrative purposes. In addition, the effect on retrieval accuracy of assuming that the highly correlated cloud error covariance is diagonal is explored. This follows work by Collard (1999), while work examining the effect of correlated forward model error has been done by Sherlock (2000b)

Watts and McNally (1988) showed that if a minimal variance retrieval is attempted where the true error covariances, $\mathbf{E}'$, $\mathbf{F}'$, and $\mathbf{B}'$, are approximated by $\mathbf{E}$, $\mathbf{F}$, and $\mathbf{B}$ respectively, the analysis error covariance can be derived as follows:

The linear minimum variance solution to the retrieval problem (Eqn. 2a) may be re-written as

$$\hat{\mathbf{x}} = \mathbf{x_0} + \mathbf{W}(\mathbf{y} - \mathbf{y_0}) \tag{12}$$

Here, $\mathbf{W} = (\mathbf{B^{-1}} + \mathbf{H^T}(\mathbf{E}+\mathbf{F})^{-1}\mathbf{H})^{-1}\mathbf{H^T}(\mathbf{E}+\mathbf{F})^{-1}$ is the gain matrix being assumed by the retrieval scheme. If the true Jacobian is $\mathbf{H'}$ then $\mathbf{y} - \mathbf{y_0} = \mathbf{H'}(\mathbf{x_T} - \mathbf{x_0})$[‡] and

$$\hat{\mathbf{x}} - \mathbf{x_0} = \mathbf{W}\mathbf{H'}(\mathbf{x_T} - \mathbf{x_0}) + \mathbf{W}\epsilon_\mathbf{y} \tag{13}$$

where $\epsilon_\mathbf{y}$ is the measurement and forward model error.

From this it can be seen that the resolution matrix , $\mathbf{WH'}$, depends only on the *assumed* observational and background error covariances rather than the true values. The changes in the characteristics of the retrieval on using the wrong input error covariances therefore only manifest themselves in the noise levels and bias, not the resolution.

Rearranging Eqn. 13 gives

$$\hat{\mathbf{x}} - \mathbf{x_T} = (\mathbf{I} - \mathbf{W}\mathbf{H'})(\mathbf{x_0} - \mathbf{x_T}) + \mathbf{W}\epsilon_\mathbf{y} \tag{14}$$

As $(\mathbf{E'} + \mathbf{F'}) = E[\epsilon_\mathbf{y}\epsilon_\mathbf{y}^T]$, $\mathbf{B'} = E[(\mathbf{x_0} - \mathbf{x_T})(\mathbf{x_0} - \mathbf{x_T})^T]$, and the true analysis error covariance matrix, $\mathbf{A'} = E[(\hat{\mathbf{x}} - \mathbf{x_T})(\hat{\mathbf{x}} - \mathbf{x_T})^T]$, and assuming that observational and background errors are uncorrelated,

$$\begin{aligned}
\mathbf{A'} &= (\mathbf{I} - \mathbf{W}\mathbf{H'})\mathbf{B'}(\mathbf{I} - \mathbf{W}\mathbf{H'})^T + \mathbf{W}\mathbf{O'}\mathbf{W}^T \\
&= \mathbf{B'} - \mathbf{W}\mathbf{H'}\mathbf{B'} - \mathbf{B'}(\mathbf{W}\mathbf{H'})^T + \mathbf{W}\mathbf{H'}\mathbf{B'}(\mathbf{W}\mathbf{H'})^T + \mathbf{W}\mathbf{O'}\mathbf{W}^T \\
&= \mathbf{B'} - \mathbf{W}\mathbf{H'}\mathbf{B'} - (\mathbf{W}\mathbf{H'}\mathbf{B'})^T + \mathbf{W}(\mathbf{H'}\mathbf{B'}\mathbf{H'}^T + (\mathbf{E'} + \mathbf{F'}))\mathbf{W}^T
\end{aligned} \tag{15}$$

When $(\mathbf{E'} + \mathbf{F'}) = (\mathbf{E} + \mathbf{F})$, $\mathbf{B'} = \mathbf{B}$ and $\mathbf{H'} = \mathbf{H}$ this reduces to Eqn. 2b.

In this work, only the effect of using the incorrect observational error covariance is being explored, therefore Eqn. 15 becomes.

$$\begin{aligned}
\mathbf{A'} &= \mathbf{B} - \mathbf{W}\mathbf{H}\mathbf{B} - (\mathbf{W}\mathbf{H}\mathbf{B})^T + \mathbf{W}(\mathbf{H}\mathbf{B}\mathbf{H}^T + (\mathbf{E'} + \mathbf{F'}))\mathbf{W}^T \\
&= \mathbf{B} - \mathbf{W}\mathbf{H}\mathbf{B} - (\mathbf{W}\mathbf{H}\mathbf{B})^T + \mathbf{W}(\mathbf{H}\mathbf{B}\mathbf{H}^T + (\mathbf{E} + \mathbf{F}))\mathbf{W}^T - \mathbf{W}(\mathbf{E} + \mathbf{F})\mathbf{W}^T + \mathbf{W}(\mathbf{E'} + \mathbf{F'})\mathbf{W}^T \\
&= \mathbf{A} + \mathbf{W}((\mathbf{E'} + \mathbf{F'}) - (\mathbf{E} + \mathbf{F}))\mathbf{W}^T
\end{aligned} \tag{16}$$

[‡] In fact (assuming $\mathbf{x_0}$ is the linearisation point), $\mathbf{y} - \mathbf{y_0} = \mathbf{H'}(\mathbf{x_T} - \mathbf{x_0}) + \mathbf{y'}(\mathbf{x_0}) - \mathbf{y}(\mathbf{x_0})$ where $\mathbf{y'}(\mathbf{x_0})$ and $\mathbf{y}(\mathbf{x_0})$ are the true and calculated observations vectors at $\mathbf{x_0}$ respectively. In the purely linear case, the extra terms will produce only a bias in $\hat{\mathbf{x}}$ but when the problem is non-linear this will result in extra terms in the covariance of $\hat{\mathbf{x}}$ also. This should be allowed for as part of the forward model error covariance matrix.

Eqn. 16 is used in the discussions to follow.

In the following, the gross effect of correlated cloud contamination errors is studied by calculating a set of spectra for a range of cloud heights, phases, optical depths and particle sizes. Only clouds that result in a maximum brightness temperature difference of less than 1K relative to the clear sky case are included in the next step. This is because it is expected that clouds with relatively high radiometric signals are likely to be detected and allowed for in the retrieval process (either by ignoring the observation altogether, attempting some form of cloud-clearing or simultaneously retrieving temperature, humidity and those cloud properties that would affect the retrieval of the clear sky parameters) and only the clouds with relatively small signals will directly affect the clear-sky retrieval statistics.

The error covariance matrix for undetected cloud is then calculated via the relation

$$\mathbf{S}_{cloud} = \frac{1}{n}\sum(\Delta\mathbf{I}_i - \Delta\bar{\mathbf{I}})(\Delta\mathbf{I}_i - \Delta\bar{\mathbf{I}})^T \tag{17}$$

where $\Delta\mathbf{I}_i$ is the array of brightness temperature differences (cloudy minus clear) for the i^{th} spectrum and $\Delta\bar{\mathbf{I}}$ is the mean brightness temperature spectrum for all n spectra considered. The mean deviation of the cloudy brightness temperature spectrum from that for clear sky, $\Delta\bar{\mathbf{I}}$, would in practice manifest itself as an extra term in the bias correction. The error covariance matrix that results from this and its correlation matrix are shown in Figure 8, while the diagonal of this matrix (compared to IASI noise) is shown in Figure 9.

In Figure 10a, the effect of attempting to retrieve the temperature profile[‡] with correlated forward model errors due to cloud is shown, both where the retrieval assumes the true observational error and where the observational error covariance matrix is assumed to be diagonal. Figure 10b shows the same thing but with the cloud error multiplied by a factor of 10 (i.e., the cloudy error covariance is multiplied by 100).

In the case where the original cloud error covariance is used (and where the error associated with undetected cloud is a fraction of the instrumental noise), the degradation of the solution accuracy when this error covariance matrix is approximated by a diagonal is small. However, when the residual cloud error covariance is multiplied one hundred-fold (standard deviation multiplied by ten), the diagonal approximation is seen to cause large degradation to the solution standard deviation. This large reduction in retrieval accuracy can be avoided if the full error covariance matrix is used.

Using a full observational error covariance matrix in the solution of the retrieval problem is problematic for two reasons. The first one is, of course, the size of the matrix, but it may be possible to mitigate these problems if linear combinations of channels may be used as

‡ Note that for ease of calculation only every eighth IASI channel is used.

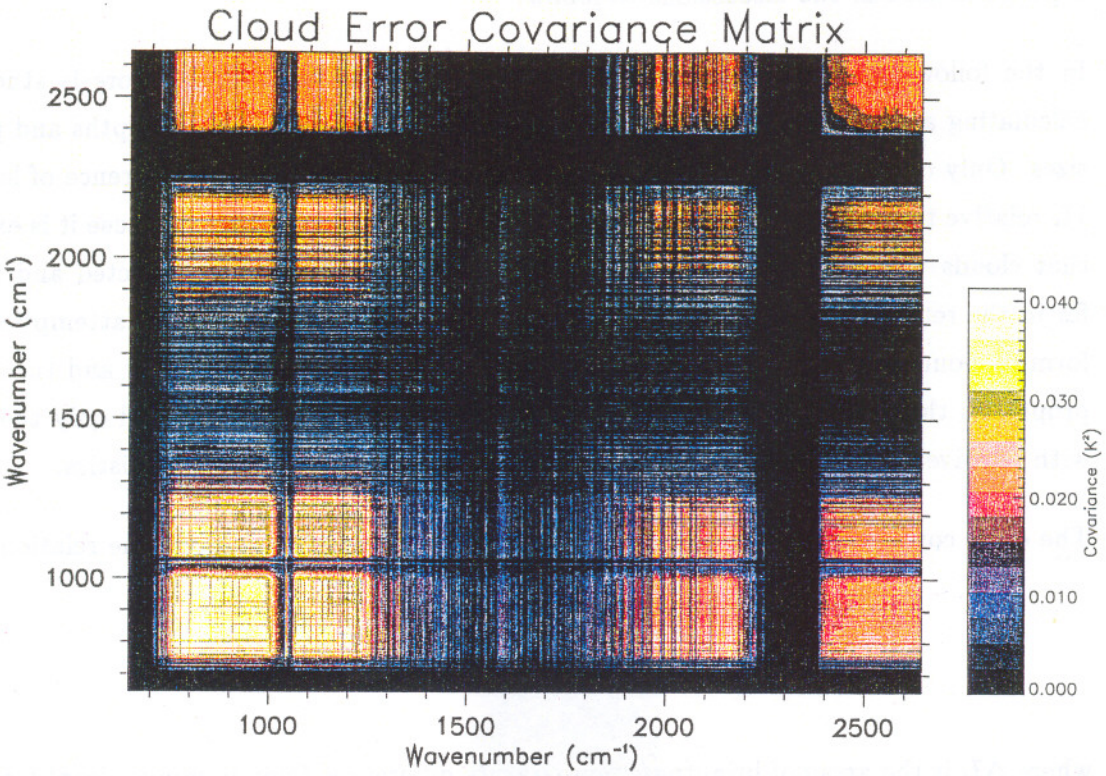

Fig. 8a. A simulated observational error covariance matrix due to the presence of undetected cloud. The relatively high error covariances in the window regions and the low covariances in the absorption bands due to CO_2, H_2O and O_3 can be seen clearly.

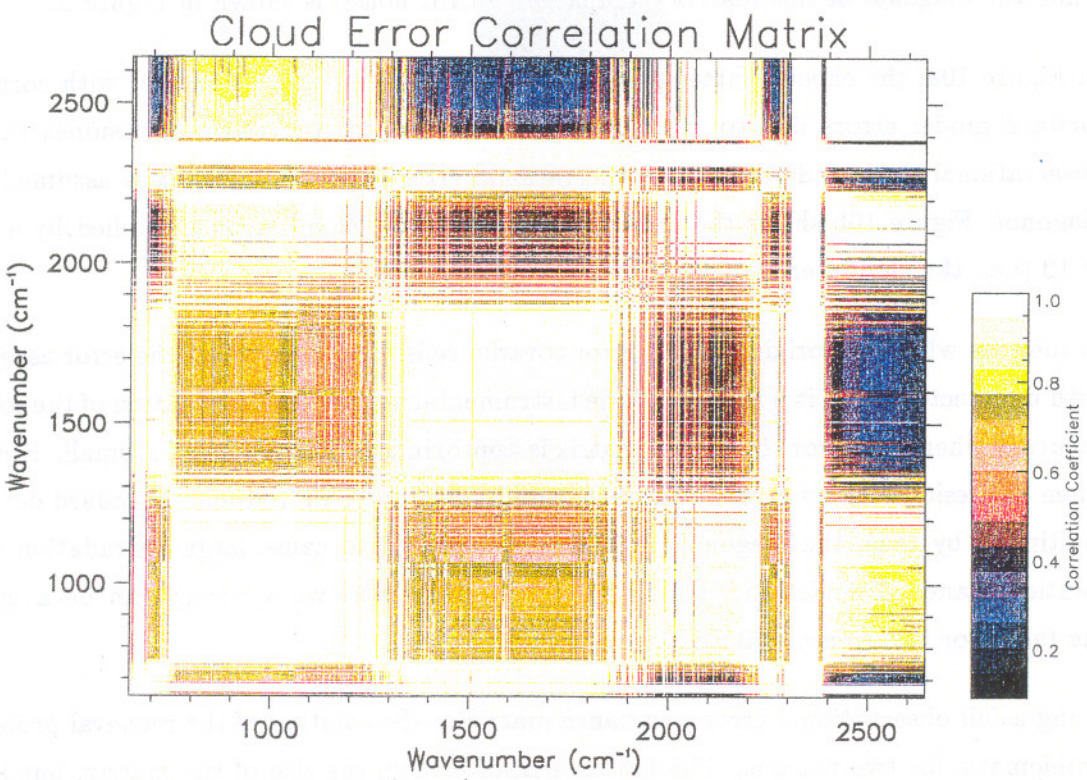

Fig. 8b. The correlation matrix corresponding to the error covariance matrix shown in Fig. 8a.

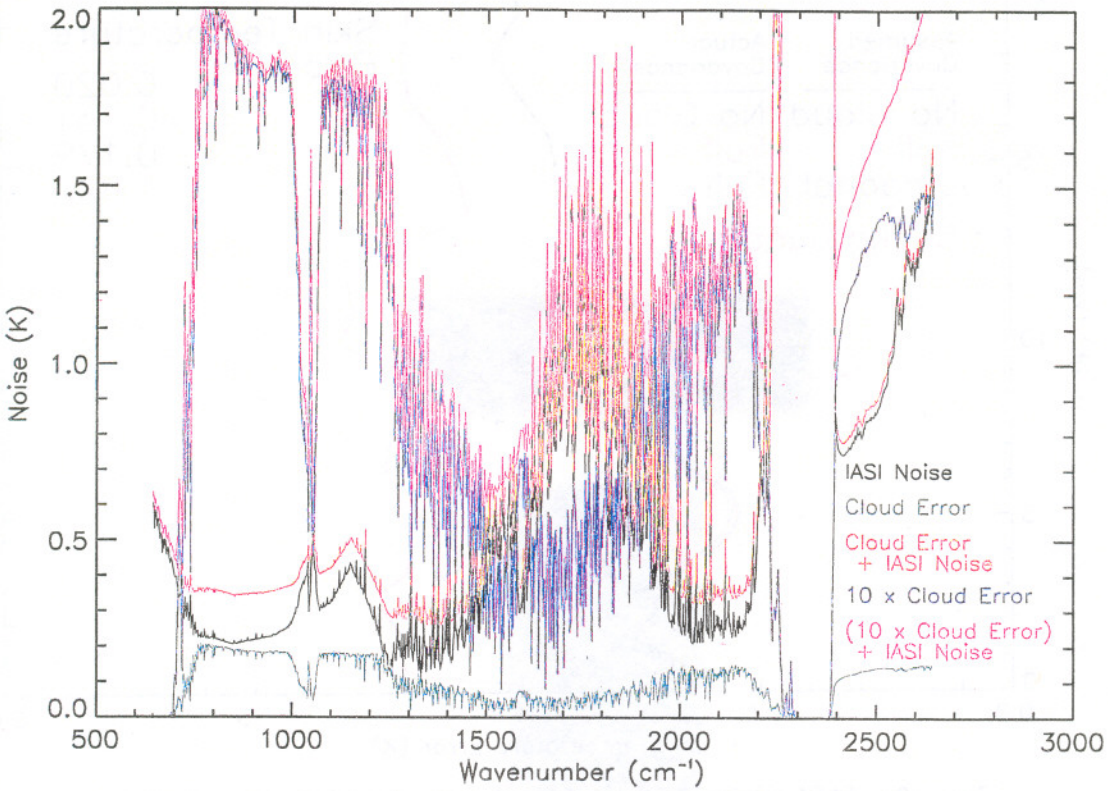

Fig. 9. The diagonal elements of the cloud error covariance of Figure 8a compared to the IASI instrument noise. Five curves are shown: The noise due to undetected cloud is shown in green while the blue curve is this noise multiplied by ten. The black curve is the IASI instrument noise while the red and magenta curves result from the IASI noise being added in quadrature to the noise represented by the green and blue curves respectively.

shown in Eqn. 10. The second difficulty will be with the determination of the true $(\mathbf{E}+\mathbf{F})$ matrix which will require much effort both in observation monitoring and theoretical work on the expected form of $(\mathbf{E}+\mathbf{F})$ (particularly the correlation structure).

Cloud Detection and the Use of Cloudy Radiances.

There are three approaches to dealing with observations which may contain cloudy data: detect and ignore all observations with clouds in the field of view (cloud detection), attempt to remove the effect of clouds from the observation (cloud clearing), or attempt to retrieve profile information simultaneously with cloud properties.

Detecting and rejecting observations affected by cloud is the most conservative approach to using the observations. An elegant scheme developed for numerical weather prediction is that of English *et al.* (1999) which uses the observations and *a priori* data to determine the probability of cloud being in the field of view using a variational approach. This has been demonstrated successfully for TOVS and ATOVS.

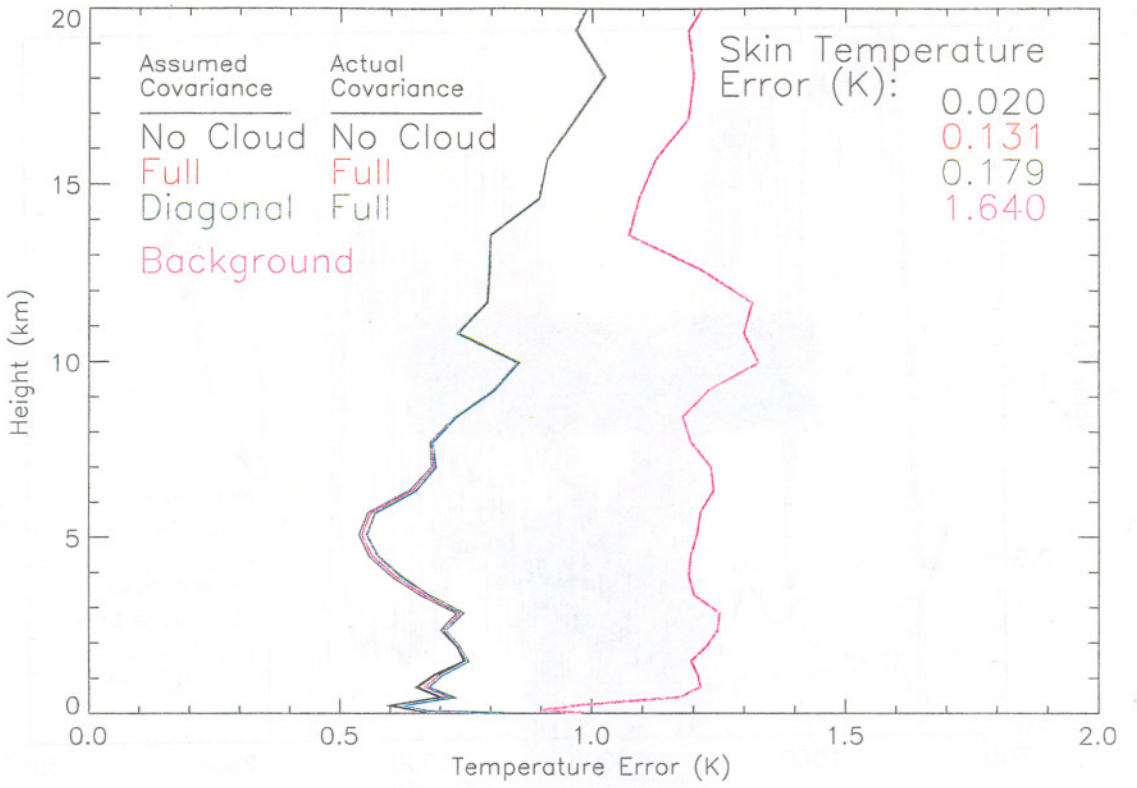

Fig. 10a. IASI retrieval accuracies using every eighth channel where the true observational error covariance is described by the full error covariance matrix shown in Figure 8a. Retrieval accuracies when the assumed observational error covariance is the same as the truth, is just the diagonal elements of the truth and for the clear case are shown.

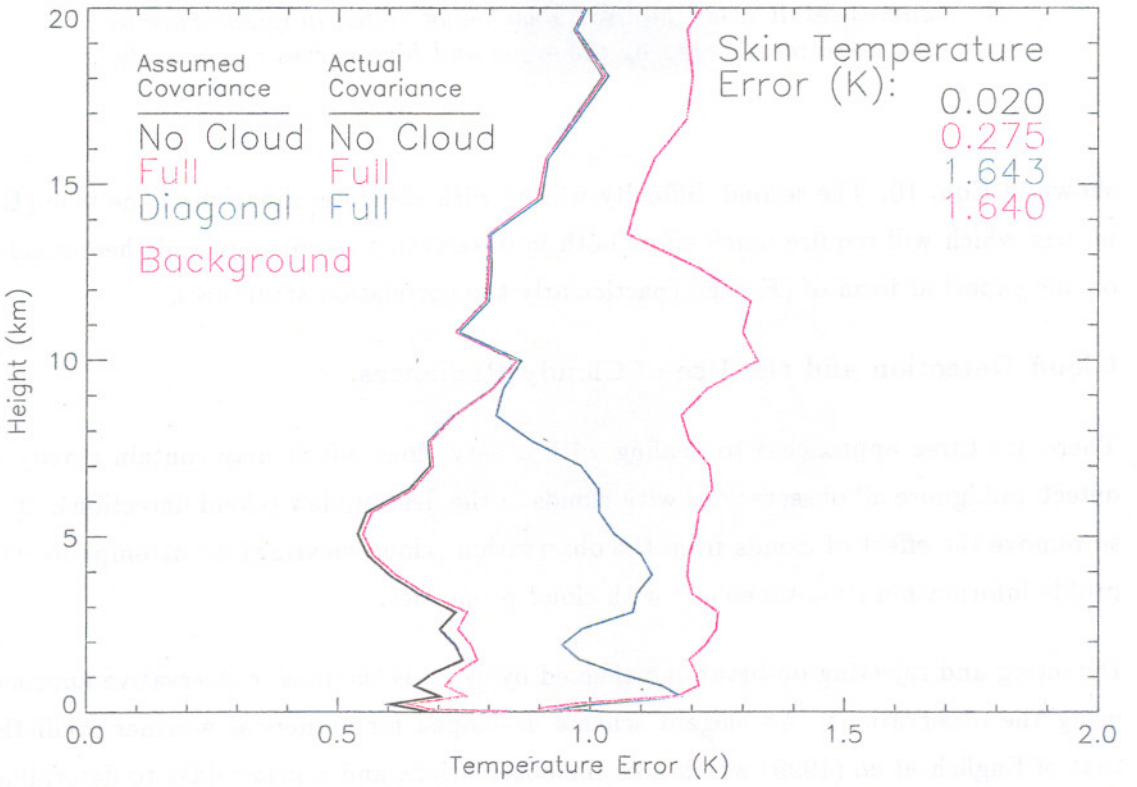

Fig. 10b. As figure 10a except the cloudy error covariance is multiplied by a factor of one hundred.

One may also use other instruments to aid in cloud detection. This can be through the use of microwave channels (which are less affected by cloud) to predict what the radiance for a chosen infrared channel should be (e.g., NOAA/NESDIS plan to use AMSU channels 4, 5 and 6 to predict the 2390.91cm^{-1} brightness temperature for AIRS (Goldberg, 2000)). One can also use a co-located imaging instrument to detect the presence of cloud without using the IR sounder observations at all, AVHRR cloud detection being a good example of this (Saunders and Kriebel, 1988).

One can go a step further than cloud detection and attempt to remove the effect of clouds on the observed radiances. This is known as cloud-clearing and the most common method for doing this is the N^* technique (Smith, 1968) which is used by NOAA in its operational TOVS retrievals (McMillin and Dean, 1982). The technique uses observations in adjacent fields of view and assumes that only the cloud amount varies between those fields of view. Joiner and Rokke (2000) have re-examined this technique in a variational context.

Eventually, one would wish to include cloud parameters in the state vector (Joiner and Rokke include the "N^*" in their state vector) and to do a full simultaneous retrieval of temperature, constituents and clouds. This has been examined by Eyre (1989) in the context of TOVS observations. The main issues that need to be addressed is that the problem is very non-linear, the *a priori* data is usually poor and the calculation of the radiative properties of clouds is less accurate than for clear air (especially transmissive cirrus clouds).

There is great incentive to solve the problem of assimilation of cloudy radiances given the preliminary result from McNally *et al.* (2000) that the regions that most affect forecast uncertainty tend to be cloudy. McNally noted that the field often just contained low clouds in which case one might at least be able to obtain information on the atmosphere above the cloud tops (low stratiform cloud decks are the simplest to model in the thermal infrared).

Quality Control.

Quality control is the detection and rejection of observations that would have a detrimental effect on the accuracy of the model field if they were to be assimilated.

One important part of quality control is the effective detection and characterisation of cloud as described above. It has already been shown how poor knowledge of undetected cloud in the field of view can have a highly detrimental effect on retrieval accuracy.

There can be many other reasons why the observed radiances are in error ranging from instrumental noise to errors in the modelling of the observed radiances.

Theoretically, the most powerful way to determine whether the observations are as expected is to evaluate the cost function, J, (Eqn. 1). The value of the cost function should be on

average equal to half the number of observations, $n_{obs}/2$, and the standard deviation about this mean should be $\sqrt{2n_{obs}}/2$. Therefore, if the value of the cost function for a given observation is significantly outside of this range then the observation may be erroneous.

A value of J that is too high indicates that the fit is poorer than expected. This can be caused through one of three reasons:

i) The observation is truly erroneous.

ii) The assumed errors are too small. In this case, monitoring of the statistics for J can provide insight into how to refine the $\mathbf{E}$, $\mathbf{F}$ and, potentially, $\mathbf{B}$ matrices.

iii) The measurement errors are not normally distributed. In this case one would have to determine what would be a reasonable range for J given the actual distribution of error. One source of observational error where this might be the case is that due to undetected cloud as was explored earlier.

If J is too small, by far the most likely explanation is that the assumed errors are too large. Again, this will provide information to help improve the error covariance matrices used.

In addition to cost function tests, one may also simply determine whether the difference between the observed and modelled radiances (based on either the *a priori* or retrieved atmospheric state) is reasonable.

One may also attempt to detect climatologically unlikely radiance spectra by projecting them along the eigenvectors of the observed covariance matrix for the observations and comparing the amplitude of the signal in these directions with the range of values that would be expected given the eigenvalues of the matrix.

Conclusions.

This paper has attempted to outline the major issues that concern the future assimilation of data from high spectral resolution infrared sounding instruments.

It has been shown that this next generation of instruments can potentially greatly increase the amount of information that may be assimilated into NWP models. However, care must be taken to ensure that the data is used most efficiently.

All of the areas explored in this paper — the efficient use of observations and channel selection, the effect of correlated error, cloud detection and quality control — will require more research in the future, both in preparation for the satellites' launches and when real data becomes available. Initial assimilation schemes will tend to be conservative, relying primarily on methods that are in use for the current generation of instruments. For instance, a subset of channels will initially be processed rather than using some of the potentially

more advantageous channel combination methods. Similarly, the initial assumption will be that the observational error covariances are diagonal until it can be shown that the correct error covariance can be accurately estimated and that its use will justify the increase in computational resources that may be required.

For the foreseeable future the greatest computational burden (in terms of time) during the assimilation of infrared sounder radiances will be the forward modelling of radiances (and the associated gradients). The greatest gains in the speed of assimilation can therefore be made here. For example, if radiances (and gradients) for a linear combinations of channels can be modelled as quickly as for a single channel the efficiency of advanced sounder radiance assimilation will be greatly increased.

Acknowledgements.

Material for this paper and many useful discussions were provided by John Eyre, Roger Saunders, Vanessa Sherlock, Sean Healy, Peter Rayer and Richard Renshaw (all of the Met Office).

References.

AMATO, U., D. DE CANDITIIS, AND C. SERIO (1998). Effect of apodisation on the retrieval of geophysical parameters from Fourier-transform spectrometers. *Appl. Optics*, **37**, 6537–6543.

BACKUS, G.E. AND J.F. GILBERT (1970). Uniqueness in the inversion of inaccurate gross Earth data. *Philos. Trans. R. Soc. London, Ser. A*, **266**, 123–192.

COLLARD, A.D. (1999). The effect of undetected cloud on IASI retrievals. *Numerical Weather Prediction Branch Technical Note, No. 261, The Met Office, Bracknell, U.K.*

ENGLISH, S.J., J.R. EYRE, AND J.A. SMITH (1999). A cloud-detection scheme for use with satellite sounding radiances in the context of data assimilation for numerical weather prediction. *Q.J.R. Meteorol. Soc.*, **125**, 2359–2378.

EYRE, J.R. (1987). On systematic errors in satellite sounding products and their climatological mean values. *Q.J.R. Meteorol. Soc.*, **113**, 279–292.

EYRE, J.R. (1989). Inversion of cloudy satellite radiances by non-linear optimal estimation. *Q.J.R. Meteorol. Soc.*, **113**, 279–292.

EYRE, J.R. (1997). Variational assimilation of remotely-sensed observations of the atmosphere. *J. Meteorol. Soc. Jpn.*, **75**, 221–228.

GOLDBERG, M. (2000). Operational processing and distribution of AIRS. To appear in *The Technical Proceedings of the Eleventh International TOVS Study Conference, Budapest, Hungary. 20^{th}–26^{th} September 2000.* Ed. J. Le Marshall, Bureau of Meteorology, Melbourne, Australia.

JOINER, J., AND A.M. DA SILVA (1998). Efficient methods to assimilate remotely sensed data based on information content. *Q.J.R. Meteorol. Soc.*, **124**, 1669–1694.

JOINER, J., AND D.P. DEE (2000). An error analysis of radiance and suboptimal retrieval assimilation. *Q.J.R. Meteorol. Soc.*, **126**, 1495–1514.

JOINER, J., AND L. ROKKE (2000). Variational cloud-clearing with TOVS data. *Q.J.R. Meteorol. Soc.*, **126**, 725–748.

MCMILLIN, L.M., AND C. DEAN (1982). EVALUATION OF A NEW OPERATIONAL TECHNIQUE FOR PRODUCING CLEAR RADIANCES. *J. Appl. Meteorol.*, **21** , 1005–1014.

McNally, A.P. (2000). The occurrence of cloud in meteorologically sensitive areas and the implications for advanced IR sounders. To appear in *The Technical Proceedings of the Eleventh International TOVS Study Conference, Budapest, Hungary. 20^{th}–26^{th} September 2000.* Ed. J. Le Marshall, Bureau of Meteorology, Melbourne, Australia.

Matricardi, M., and R.W. Saunders (1999). A fast radiative transfer model for simulation of IASI radiances. *Applied Optics.*, **38**, 5679–5691.

Menke, W. (1984). *Geophysical data analysis: discrete inverse theory.* Academic Press, San Diego CA, USA.

Prunet, P., J.-N. Thépaut and V. Cassé (1998). The information content of clear sky IASI radiances and their potential for numerical weather prediction. *Q.J.R. Meteorol. Soc.*, **124**, 211–241.

Purser, R.J. and H.-L. Huang (1993). Estimating effective data density in a satellite retrieval or an objective analysis. *J. Appl. Meteor.*, 1092–1107.

Rabier, F., E. Klinker, P. Courtier and A. Hollingsworth (1996). Sensitivity of forecast errors to initial conditions. *Q.J.R. Meteorol. Soc.*, **122**, 121–150.

Rabier, F., D. Chafai, and R. Abboubi (2000). Comparison of channel selection methods for IASI. To appear in *The Technical Proceedings of the Eleventh International TOVS Study Conference, Budapest, Hungary. 20^{th}–26^{th} September 2000.* Ed. J. Le Marshall, Bureau of Meteorology, Melbourne, Australia.

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

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

Rodgers, C.D. (1996). Information content and optimisation of high spectral resolution measurements. *SPIE*, **2830**, *Optical spectroscopic techniques and instrumentation for atmospheric and space research II,* Paul B. Hays and Jinxue Wang, eds., pp136–147.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheres: Theory and Practice.* World Scientific Publishing.

Saunders, R.W., and K.T. Kriebel (1988). An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sensing*, **9**, 123–150.

Saunders, R.W. (2001). Assimilation of IASI and AIRS data: Forward Modelling. In *Proceedings of the ECMWF Seminar on 'Exploitation of the new generation of satellite instruments for numerical weather prediction' 4–8 Sept. 2000.*ECMWF Report, Reading, U.K.

Smith, W.L. (1968). An improved method for calculating tropospheric temperature and moisture from satellite radiometer measurements. *Mon. Weather Rev.*, **96**, 387–396.

Sherlock, V.J. (2000a). Results from the first UKMO IASI radiative transfer model intercomparison. *Numerical Weather Prediction Branch Technical Note, No. 287, The Met Office, Bracknell, U.K.*

Sherlock, V.J. (2000b). Impact of IASI fast radiative transfer model error on IASI retrieval accuracy. *Numerical Weather Prediction Branch Technical Note, No. 319, The Met Office, Bracknell, U.K.*

Watts, P.D., and A.P. McNally (1988). The sensitivity of a minimum variance retrieval scheme to the values of its principal parameters. *The Technical Proceedings of the Fourth International TOVS Study Conference.*, 399–407. Ed. W.P. Menzel, Univ. of Wisconsin, Madison WI, U.S.A.