STRATEGIES FOR ASSIMILATING SOUNDING INFORMATION

Andrew C Lorenc
Meteorological Office,
Bracknell, England.

1. INTRODUCTION

There is general acceptance of the need to use the best available prior (background) information in the retrieval process of converting radiances to soundings. There is widespread (but not universal) acceptance that a main source of this should be a short-period forecast, carrying the information from earlier observations. This is discussed in section 2. Other potentially useful data are sea-surface temperatures, from satellite observations and ships, temperatures from SYNOP reports, clouds from satellite pictures, and upper atmospheric data from climatology and satellite observations.

These data improve the retrieved soundings in two ways: directly, since combining independent data should give greater accuracy, and indirectly, since a better estimate of the atmospheric environment can give more reliable detection of radiances affected by cloud, and a more accurate linearization of the radiative transfer equation. The direct effect can also be achieved in the NWP analysis process.

Direct effects can be linearized and separated, so that data are added in successive steps, until a total picture is built up. If the relationships between variables are linear, and error distributions Gaussian, then done correctly this gives the optimal result (Lorenc, 1986). Given three estimates $\mathbf{x}$ $\mathbf{y}$ and $\mathbf{z}$ of a profile, represented as a vector of numbers, with Gaussian error distributions with covariance matrices $\mathbf{X}$ $\mathbf{Y}$ $\mathbf{Z}$, then the minimum variance best estimate profile $\mathbf{a}$ is given by:

$$\mathbf{a} = (\mathbf{X}^{-1}+\mathbf{Y}^{-1}+\mathbf{Z}^{-1})^{-1}\ (\mathbf{X}^{-1}\mathbf{x} + \mathbf{Y}^{-1}\mathbf{y} + \mathbf{Z}^{-1}\mathbf{z}) \tag{1}$$

This estimate can also be found in two steps, by first combining $\mathbf{x}$ and $\mathbf{y}$ to give $\mathbf{b}$ and its error covariance $\mathbf{B}$:

$$\mathbf{b} = (\mathbf{X}^{-1}+\mathbf{Y}^{-1})^{-1}\ (\mathbf{X}^{-1}\mathbf{x} + \mathbf{Y}^{-1}\mathbf{y}) \tag{2}$$

$$\mathbf{B} = (\mathbf{X}^{-1}+\mathbf{Y}^{-1})^{-1} \tag{3}$$

and then combining **b** and **z** to give **a** :

$$\mathbf{a} = (B^{-1}+Z^{-1})^{-1}\ (B^{-1}\mathbf{b} + Z^{-1}z) \tag{4}$$

Thus if the relationships between radiances and temperatures were linear, and the error distributions Gaussian, then the inversion process could be done independently, before an analysis which combines them with other data. However the indirect effects are intrinsically nonlinear, and so in principle all data must be processed simultaneously for the optimal result.

For most operational forecast centres this poses practical problems. They cannot always afford to wait until all data are gathered, they cannot afford the scientific effort to monitor and understand the detailed performance characteristics of all the different observing systems, and they cannot afford the computer power to do an integrated nonlinear processing of all the data. In section 3 practical processing options are presented, and in section 4 their merits discussed.

2. USE OF FORECAST

2.1 Advantages

Unless one is using a four-dimensional method (which should anyway be built round a forecast model), then information from earlier observations is best made available through a forecast background field. In many cases such a forecast contains a significant amount of information, compared to the the newly observed information, and should be able to contribute to the nonlinear processes in the retrieval system, such as quality control, and the treatment of clouds. It should also be given a significant (linear) weight in the final product. There is little difference between a system which uses such soundings consistently in an analysis, and an integrated analysis directly using the radiances (Eyre and Lorenc, 1989).

An advantage of using a forecast in retrievals is that it facilitates a simple but stringent test that the satellite data processing and retrieval has been done sufficiently well for the soundings to be useful in numerical weather prediction (NWP). For forecast-independent retrievals it is not clear how to do this. The soundings can be verified against independent "collocated" radiosondes, but imprecise collocation, and observational and representativeness errors in the sondes, mean that a residual is expected even for perfect retrievals. The agreements achieved in practice are similar to those from a short-period forecast; thus it is not clear how good they must be before the satellite soundings are a useful addition to the NWP process. Results of NWP impact studies are dependent on the details of the NWP system used, and cannot necessarily be generalized. If on the other hand both the radiance information and the forecast information are made available to the retrieval process, then it is easy to show that the resulting sounding should be better than both the forecast alone, and the forecast-independent retrievals. This follows for minimum variance retrieval algorithms, from equations similar to (3). Since the error covariance matrices are all positive definite, then any linear combination of elements of **b** has an expected error variance which is less than or equal to the same combination of **x** or **y**. Thus temperatures, thicknesses, and lapse-rates should all be more accurate. The relative accuracy of the forecast, forecast-background retrievals, and forecast-independent retrievals can be measure using radiosonde collocation statistics, without needing to know the absolute accuracy of the radiosondes or the collocation method.

2.2 Example

For the Meteorological Office Local Area Satellite Sounding (LASS) system, it was fairly simple to show that forecast-background retrievals were better than forecast independent ones. Fig.1 shows results from an early trial of the forecast-background system, before it had been fully tuned to remove biases in the forecast, and with a very simple forecast→sounding interpolation. Nevertheless the SDs (ie excluding mean biases) and most of the RMS differences for temperature are less for the forecast-background retrievals. Subsequent tuning of this system has significantly improved results (Fig.2). However it has proved to be much harder to show that forecast-background retrievals are better than the forecast alone. An example from a recent month is shown in Fig.2. Apart from in the stratosphere, where the forecast is extrapolated using climatological data, and near the surface, there is not a significant consistent improvement. Three contributary causes for this disappointing result have been suggested:

(i) The nonlinear effects such as quality control and cloud clearing are not being handled sufficiently well for the linear minimum variance theory to be applicable.

(ii) There really is only a little information that a single set of satellite radiances can add to a typical forecast.

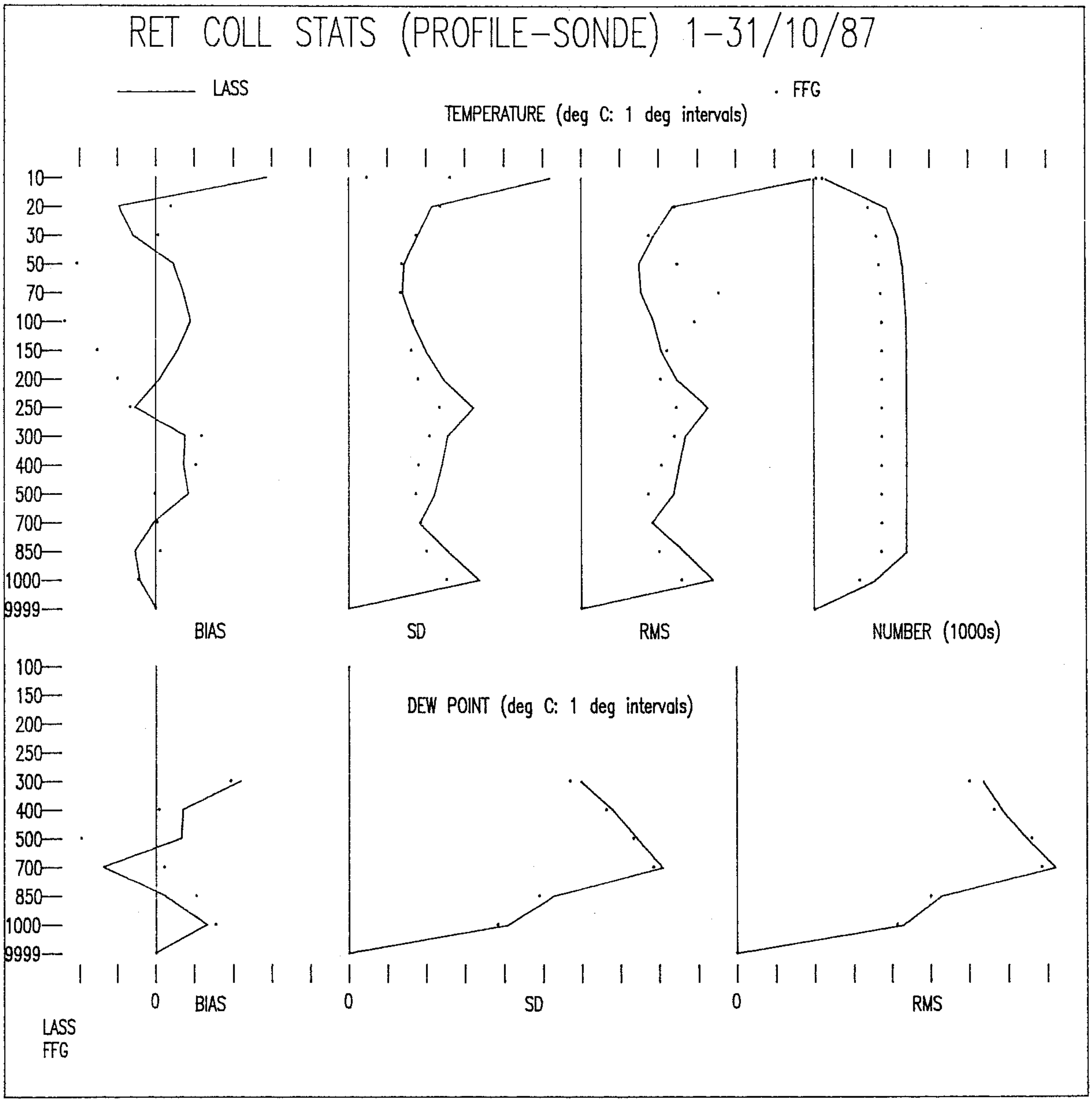

Fig.1 Mean, SD of, and RMS differences between LASS retrievals and independent radiosondes, for forecast-independent retrievals from the then operational system (solid lines),
and for forecast-background retrievals from a test system (dots),
for temperature (upper curves) and dew-points (lower curves) at pressure levels.
(from P.Dibben, personal communication)

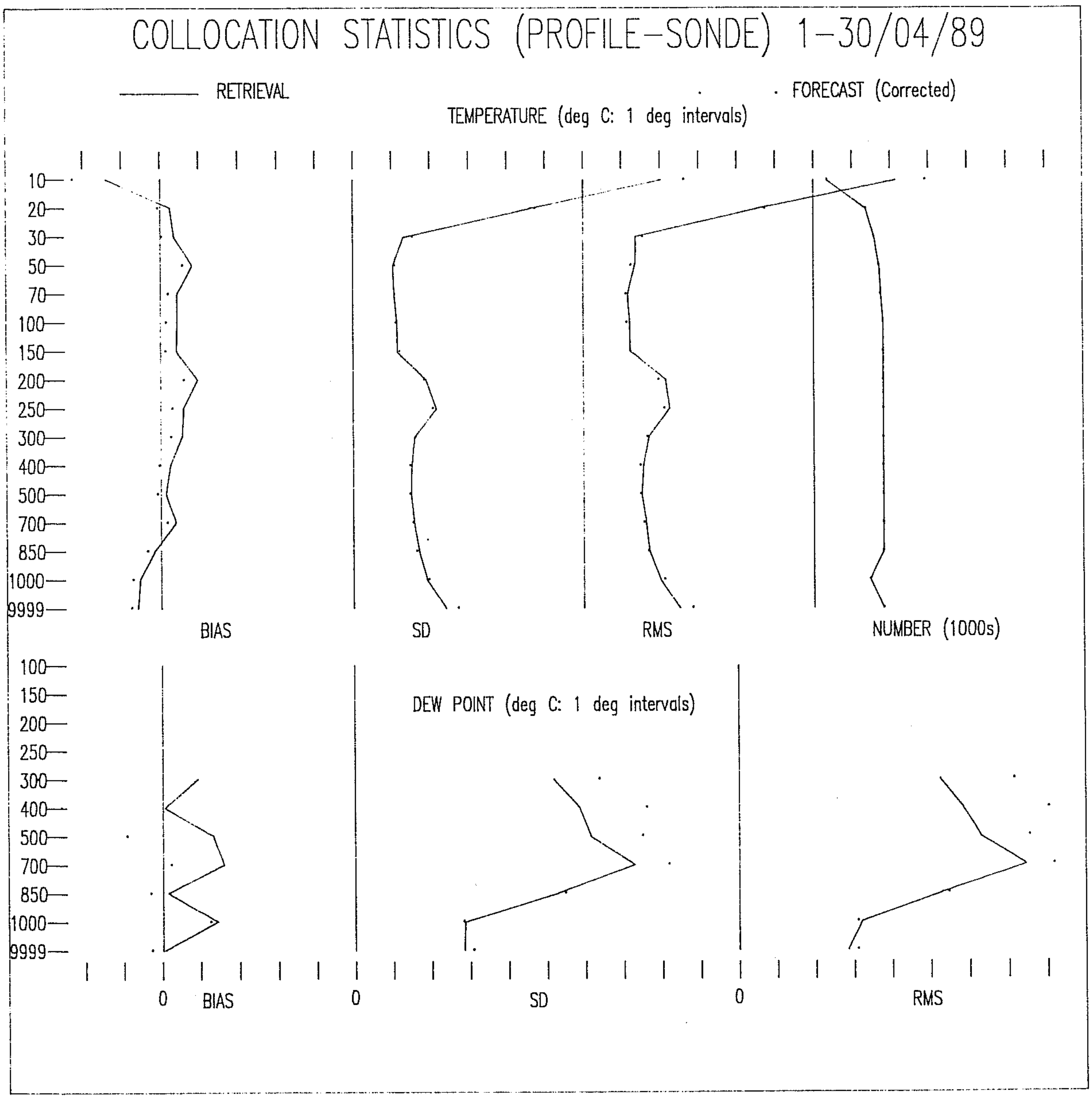

Fig.2 Mean, SD of, and RMS differences between LASS retrievals and independent radiosondes (solid lines),
Mean, SD of, and RMS differences between forecast profiles used as backgrounds in the retrievals and the same radiosondes (dots),
for temperature (upper curves) and dew-points (lower curves) at pressure levels.
(from P.Dibben, LASS Report April 1989, Met O 24)

(iii) This verification against independent radiosondes is not a good measure of useful information.

2.3 Disadvantages

The main argument against forecast-background retrieval schemes is that errors in the forecast are carried though to the retrievals. This of course is true for any background in a minimum variance scheme; the contribution being dependent on the relative weighting of radiances and background. A linearized minimum variance retrieval is given by:

$$\mathbf{x}_{a} = \mathbf{x}_{b} + \mathbf{W}(\mathbf{y}_{o} - K(\mathbf{x}_{b})) \tag{5}$$

$$\mathbf{W} = \mathbf{B}\,\mathbf{K}^{T}(\mathbf{O} + \mathbf{F} + \mathbf{KBK}^{T})^{-1} \tag{6}$$

Here, following the notation of Lorenc (1986), $\mathbf{x}$ is a temperature and humidity profile, and $\mathbf{y}$ is a set of radiances. Subscripts $_{a}$ $_{t}$ $_{o}$ and $_{b}$ denote retrieved, true, observed, and background. $K(\cdot)$ denotes the radiative transfer calculation of radiances from a profile and $\mathbf{K}$ is its linearization. $\mathbf{O}$ $\mathbf{F}$ $\mathbf{B}$ are the error covariances of observations, radiative transfer calculation, and background respectively. $\mathbf{W}$, given by (6), is the optimal set of weights given to the radiances in the retrieval.
Then the errors in $\mathbf{x}_{a}$ are given by:

$$\mathbf{x}_{a} - \mathbf{x}_{t} = \mathbf{W}(\mathbf{y}_{o} - \mathbf{y}_{t}) + \mathbf{W}(\mathbf{y}_{t} - K(\mathbf{x}_{t})) + (\mathbf{I} - \mathbf{WK})(\mathbf{x}_{b} - \mathbf{x}_{t}) \tag{7}$$

It can be seen from (7) that the retrieved profile's error contains contributions from the observational error, the error in $K(\cdot)$ and the background error. The contribution from a forecast background might be more important than that from another type of background because:

(i) A forecast is more accurate than other backgrounds, hence its weighting is greater and $\mathbf{W}$ is smaller.

(ii) It is when the forecast errors are large that we most need accurate soundings.

(iii) Soundings with errors which are correlated with the forecast background are slightly more complicated to use correctly in an analysis which uses the same forecast background.

(iv) In an analysis which uses a different forecast background, we might not want the soundings to be "contaminated" with somebody else's forecast.

Methods for alleviating these problems are discussed in the next section.

3. ANALYSIS OPTIONS

In this section we concentrate on the NWP analysis end of the processing chain, discussing options for including satellite data, and ways of minimizing the effect of the various types of retrieval error. The total integrated retrieval and analysis process is shown schematically in Fig.3; we are thus considering options for the upper boundary of the bottom module in this figure.

3.1 Retrieved profiles

The traditional way of using satellite soundings is as profiles of temperature and humidity, often known by their telecommunication code name, SATEMs. For the past decade the straightforward use of these data has had a beneficial impact on NWP in the southern hemisphere, where other information sources are sparse. However the impact on the better observed northern hemisphere has been more marginal. It has long been recognized that these soundings have systematic and correlated errors, which make their effective use in combination with other data difficult. Attempts are made to allow for the correlations in OI analysis methods (e.g. Lorenc, 1981), and correlations have been measured empirically (e.g. Schlatter and Branstator, 1979). However this is difficult, because of the lack of a reliable $\mathbf{x}_t$. Now that a more physical basis for retrieval processes is in use, it is worthwhile considering the errors in terms of (7).

(i) The first term, from the observational errors, leads to large vertical correlations in retrieval error, because of the large vertical scale of the so-called *weighting functions* which form $\mathbf{K}$ and hence affect $\mathbf{W}$. Horizontal correlations will only come through instrument calibration drift, and horizontally correlated errors in $\mathbf{y}_o$.

(ii) The second term, representing inaccuracies in our knowledge and calculation of the radiative transfer equation, can cause systematic biases with large horizontal scales, or synoptic scale correlations if $K(\cdot)$ contains approximations whose validity is a function of the atmospheric profile.

(iii) The third term consists of background errors carried through to the retrieval; its magnitude will depend on the weighting given to the background, and the scales will depend largely on the scales of the highly nonlinear processes _ _ _ _ _ _ _ _ _ _ _ linearizable processes

OTHER DATA ______________________________ RAW RADIANCES

pre-processing

⇒ de-clouding

→ field of view correction

→ zenith angle correction

⇒ quality control

______________________________ CLEAR-COLUMN RADIANCES

retrieval

⇒ radiative transfer calculation

→ inversion

______________________________ RETRIEVED PROFILES

NWP analysis

⇒ quality control

⇒ analysis

⇒ 4-D data assimilation

FORECAST BACKGROUND ______________________________ NWP ANALYSIS

Fig.3 **processes** and **DATA** in an integrated retrieval and NWP analysis suite.
Possible modularization is marked by ____________

background error.

A thorough assessment of these error sources can lead to a more effective use of these soundings. However this cancels one of the main benefits of using the data in this form; simplicity. The user has to understand the satellite instrument and the processing algorithm. This better understanding can lead to improved observational error covariance statistics to be used in an optimal statistical analysis scheme. However most schemes are forced to restrict the number of data selected, and are thus sub-optimal. For these the understanding should for instance suggest that sounding data should be processed as layer means, or thicknesses, consistent with the scale of the vertical correlations discussed in (i) above.

3.2 Retrieved profiles + background used

It was proposed by Lorenc et al (1985) that a simple improvement of the use of soundings would be possible if the background $\mathbf{x_b}$ were made available as well as the sounding. Users would be able to make their own assessment of the magnitude and correlation of its errors. Given a reasonably accurate estimate of **W** and **K**, they can correct for the effect of a sub-optimal background $\mathbf{x_b}$ by calculating what the retrieval would have been using a slightly different background $\mathbf{x_b}'$:

$$\mathbf{x_a}' = \mathbf{x_a} + (\mathbf{I} - \mathbf{WK})\ (\mathbf{x_b}'-\mathbf{x_b}) \qquad (8)$$

Code to do this for LASS soundings, replacing an older forecast-background by an assimilation background incorporating more recent data, has been included in the Met Office's recently implemented "AC" assimilation scheme; it will be evaluated in impact tests currently being planned.

Alternatively, a modified retrieval can be calculated whose errors, for diagonal elements of the covariance matrix, are uncorrelated with those of the background (Lorenc et al 1985, eqn.5). This is the minimum variance retrieval subject to the additional constraint that the background value at a particular level is given zero weight in the retrieval at that level. It is one way of getting round the problem stated in 2.3 (iii). The effect of the modification is to amplify the difference $\mathbf{x_a}'-\mathbf{x_b}$, but to increase the estimated uncertainty in $\mathbf{x_a}'$, so that its assigned weight in the analysis is decreased.

3.3 Pre-processed *clear-column* radiances

Most operational processing of satellite soundings has been split into two stages:

(i) pre-processing, which includes quality control, and the removal of effects due to clouds, fields of view, and viewing angles. This produces equivalent clear-column radiances

(ii) inversion of these radiances into a sounding.

One way of better matching the form of the data to their information content is to use these clear-column radiances directly in the analysis, integrating the bottom two modules in fig.3. To do this the analysis scheme needs an algorithm for evaluating $\boldsymbol{K}(\cdot)$. Errors in this can still lead to complicated and poorly understood systematic and correlated errors, but at least the errors due to the background term in (7) are avoided. Note that the use of the instrument's channels is still not necessarily the optimal representation, since it is based on $\mathbf{K}$ rather than $\mathbf{W}$. Future satellite instruments, using interferometry, are being developed which may provide many thousands of radiance channels. For these some linear combinations of channels may be used in order to reduce data amounts; $\boldsymbol{K}(\cdot)$ would have to be altered accordingly.

3.4 Raw radiances

In principle the algorithms used to incorporate radiances directly in an NWP analysis can be extended to use the raw radiances, integrating all three modules of fig.3. $\mathbf{x}$ would have to contain cloud information, and $\boldsymbol{K}(\cdot)$ would need to include the effects of clouds, as well as the other effects dealt with in the pre-processing of clear-column radiances. As explained in the introduction, these effects are best dealt with in an integrated processing of all available information, since many are nonlinear.

NWP models are increasingly being used to predict cloud, for instance the Met Office mesoscale model carries cloud liquid water content as a predictive variable, and it is planned that by the end of 1990 our global and regional fine-mesh models should do likewise. These models require cloud data to analyse, and should be able to provide a useful forecast cloud field for use as background. Since the largest signal in satellite radiances comes from the clouds, it is logical to develop methods which use this, rather than simply attempting to eliminate it as at present, to produce cloud-cleared radiances.

However there are already large computational problems in performing a four-dimensional data assimilation for NWP, properly to use past data. The addition complications of processing raw satellite data would make a fully integrated system a large undertaking. One option is to integrate the top two modules of fig.3, with an interface of retrievals and background used to the third, as discussed in 3.2 above. Eyre (1989a & b) has proposed a nonlinear inversion method which integrates the preprocessing and retrieval steps. A proper use of these retrievals in the NWP analysis is almost equivalent to use of the raw radiances in the analysis.

4. DISCUSSION

All of the methods discussed in section 3 have advantages and disadvantages; each has NWP systems for which it is most suited. The Met Office intend to develop and maintain some capability for each approach.

Use of SATEMs (section 3.1) is the simplest and cheapest, and is currently the most common. It will need to be maintained as fall-back for some time. Effort in developing more sophisticated methods are concentrating on the TIROS operational sounders; facilities for processing other satellites need to be maintained, in case the TIROS system should fail. Apart from its theoretical limitations, the use of centrally processed SATEMs has practical drawbacks for time-critical regional NWP. There are inevitable time delays in collecting and processing global data, which can be avoided by a local direct readout and processing. This was the main reason for setting up the Met Office LASS system.

Use of soundings with knowledge of their background (section 3.2) is one way of partitioning the ideal, fully integrated, system into modules that are more managable, letting a forecast contribute to the prior knowledge used in the retrieval, while allowing for the background's contribution to the retrieval error. It avoids the expense of evaluation $\boldsymbol{K}(\cdot)$ for each sounding as part of the NWP analysis system. It retains the option of integrating the preprocessing and de-clouding in a nonlinear retrieval.

Use of clear-column radiances (section 3.3) is another way of partitioning the system. It avoids integating many of the satellite specific details into the NWP analysis, while also avoiding the problem of allowing for errors fed through from the background. A radiative transfer calculation $\boldsymbol{K}(\cdot)$ must be coded, tuned and calibrated. Ideally, if this method of processing is to spread, the instrument operators who do the preprocessing should provide routines for $K(\cdot)$, and estimates of its errors. $K(\cdot)$ will be nonlinearly dependent on prior temperature and humidity profiles. But because of the omission of clouds, it will be sufficiently close to linear to avoid most mathematical difficulties in the analysis, whether it is done using nonlinear variational methods, or a linearized method such as optimal interpolation (OI).

5. CONCLUSIONS

Partitioning of the retrieval-analysis process is valid for linear processes, and Gaussian errors. The best method depends on the needs and resources of the user.

For nonlinear processes, and non-Gaussian errors, the best available prior information should be used at every stage. An NWP system can help provide this.

References

Eyre, J.R. 1989: "Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation: theory and simulation for TOVS." Quart. J. Roy. Met. Soc. *to appear*

Eyre, J.R. 1989: "Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation: application to TOVS data." Quart. J. Roy. Met. Soc. *to appear*

Eyre, J.R. and Lorenc, A.C. 1989: 'Direct use of satellite sounding radiances in numerical weather prediction'. Met. Mag. **118**, 13-16

Lorenc, A.C. 1981: 'A global three-dimensional multivariate statistical analysis scheme.' Mon. Wea. Rev. 109 , 701-721.

Lorenc, A.C. 1986: 'Analysis methods for numerical weather prediction.' Quart. J. Roy. Met. Soc. 112, 1177-1194.

Lorenc, A.C., Adams, W., and Eyre, J.R. 1985: "The analysis of high-resolution satellite data in the Met Office." Met O 11 Tech. Note 212.
Proceedings of ECMWF workshop on high-resolution analysis, June 1985. 69-88.

Schlatter, T.W., and Branstator, G.W. 1979: "Errors in Nimbus 6 temperature profiles and their spatial correlation". Mon. Wea. Rev. 107, 1402-1413.