Short-range Forecasting Research

Short Range Forecasting Division

Technical Report No. 11

RETRIEVAL AND ASSIMILATION: SYSTEM CONSIDERATIONS

by

Andrew C Lorenc

March 1992

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

RETRIEVAL AND ASSIMILATION: SYSTEM CONSIDERATIONS

Paper presented at

NATO Advanced Research Workshop

High Spectral Resolution Infrared Remote Sensing for Earth's Weather and Climate Studies

Paris, France, 23-26 March 1992

This paper has been submitted for publication by Springer-Verlag in a NATO ASI volume devoted to the above workshop.

S Division Technical Report No 11

March 1992

Andrew C Lorenc

Short-range Forecasting Research Division
Meteorological Office
London Road
Bracknell, RG12 2SZ
England

RETRIEVAL AND ASSIMILATION: SYSTEM CONSIDERATIONS

Andrew Lorenc
Meteorological Office (S)
London Road
Bracknell, RG12 2SZ
England

1 INTRODUCTION

The information content of a piece of information can only be defined with regard to what is already known. An expression, following Shannon, for this is given in section 3. But for practical applications, this is not a sufficient measure. We need also to consider the usefulness of the information.

Data assimilation, like the retrieval of information from remote sensing, can be considered as an inverse problem. The equations derived in section 3 are applicable to either problem. It is shown that the error of the forward process, estimating the observed parameters from the model parameters, must be considered in the inverse process, deducing model parameters from observations. In the infrared remote sensing which is the subject of this workshop, the forward process is the radiative transfer equation. If we cannot calculate that accurately, radiances are less useful. In data assimilation, the forward process is a Numerical Weather Prediction (NWP) model (or a very similar General Circulation Model (GCM) if the assimilation is for climate studies). Since assimilation is an inverse problem, if the model errors in prediction of a parameter are large, observations of that parameter are less useful. Thus, by definition, the model can make reasonably accurate predictions (if it has information from earlier observations) for observations which are useful in the assimilation.

This reasoning has lead some NWP centres to use the forecast fields as background information, when retrieving temperature soundings. Satellite radiances from the current TOVS instrument do not contain information about detailed vertical structure, compared to the model prediction. By using the forecast as prior to fill this lack of information (null space), instead of information from another source, information on larger vertical scales can be extracted from the radiances without introducing less accurate information on smaller scales.

The Met Office was perhaps the first to use this approach operationally; some results are shown in section 2. Experiments at NMC (Baker 1991) and ECMWF (Eyre personal communication, and this workshop), have confirmed that results are better than with separate retrieval and assimilation. Although the next generation of high-spectral resolution sounders will give information on smaller vertical scale, the scales resolved by the model will also have increased, so this qualitative conclusion will probably still hold.

In section 2 the Met Office's research into the use of satellite soundings is reviewed. Development is concentrated on the use of a forecast background, and on accounting for cloud in the radiative transfer (rather than using "cloud-cleared" radiances). Possible strategies for developments to cope with high spectral resolution sounders are discussed. Clouds make the radiative transfer equation very nonlinear. Equations for the inverse problem, are derived in section 3, and nonlinearities are discussed. Final, in section 5, some conclusions and options are discussed.

2 METEOROLOGICAL OFFICE SATELLITE SOUNDING RETRIEVALS

The Met Office has since 1983 run a Local Area Satellite Sounding System (LASS). Initially the system was based on the International TVS Processing Package from the University of Madison, run on a dedicated mini-computer (Turner *et al.* 1985), using direct read-out data from the TIROS satellite. A new cloud clearing scheme was developed (Eyre and Watts, 1987), but it proved difficult to demonstrate a positive impact of the data on the quality of operational forecasts. In 1987, the system was changed to use forecast backgrounds in the retrieval (Lorenc *et al.* 1986, Eyre and Lorenc, 1989). This is the scheme operational in March 1992.

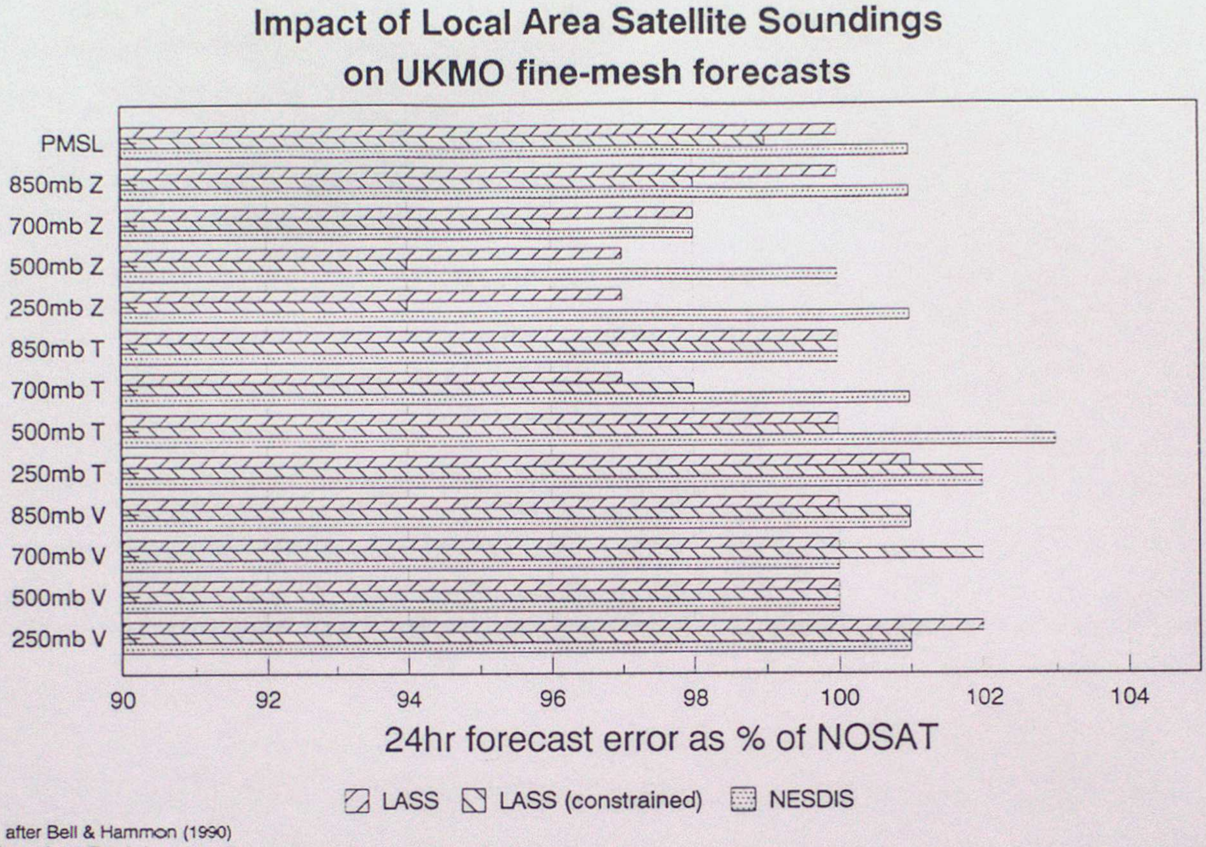

Fig 1. Ratios of rms forecast errors with and without satellite soundings.

Monitoring of the background profiles, and the retrieved profiles, against independent radiosondes does not show a clear impact from the retrieval process; the backgrounds fit the radiosondes about as well as the retrievals do. However impact tests on the quality of forecasts (Bell and Hammon, 1990) do show a slight positive impact. This depends somewhat on how the forecasts are verified. Figure 1 shows the error measured against radiosondes over the whole regional model area, of forecasts made using various forms of the TOVS satellite data, relative to the error in forecasts without the TOVS data. Each set of three bars shows the score for one verification parameter. The top bar in each set is for the operational system, with direct assimilation of the forecast background retrievals. As the retrievals have model information used as background, it is not statistically optimal to use them in this way. Lorenc *et al* (1986) proposed a constraint method to remove the background information (this is discussed further in section 3). The second bar shows results using this method. Results for height verification in particular seem to indicate it is a little better. The third bar shows results from forecasts using the NESDIS retrievals sent out on the Global Telecommunication System (GTS). If anything they have a slight negative impact on forecast quality. In our local area, the forecast backgrounds are usually very good, because even over the Atlantic there is a variety of data. The NESDIS retrievals, made not using these backgrounds, actually damage the model state when we assimilate them. (It might be possible to devise an assimilation method which extracts the useful information while avoiding the damaging bad information from NESDIS's backgrounds, but we have failed to find one).

Since there is little useful information in our local area, we are setting up a scheme (GLOSS) to apply the forecast background retrieval technique to the global cloud-cleared radiances, which are now sent on the GTS in binary (BUFR) code. It is hoped by this means to get useful information in areas where our model backgrounds are less reliable, for instance the Pacific and southern hemisphere. In order to have the computer power for this, and to facilitate closer coupling with the forecast model, GLOSS is being set up on the mainframe Cray YMP computer.

In parallel with the GLOSS development, a scheme for direct use of cloudy radiances is being developed (Eyre, 1989a&b). This technique inverts the radiative transfer equation including clouds. As well as temperature and humidity profiles, cloud parameters can be retrieved (Watts, this workshop). Over the next decade, the accuracy of representation of clouds in forecast models is expected to improve dramatically, so although at present there is no coupling with the model's clouds, it is expected that eventually background cloud parameters will be provided to the retrieval, and the retrieved clouds will be used in the model. The technique also extends naturally to the microwave sounders planned for ATOVS in 1995. The output from the radiative transfer equation is nonlinearly related to the cloud; the effect of low level temperatures depends strongly on the presence or absence of cloud. So an iterative retrieval method is necessary. Experience with the cloudy radiances in LASS is that about four iterations are necessary.

The retrieval methods described above are one-dimensional. Although background information from a model is horizontally consistent, and most cloud clearing algorithms use information from adjacent fields of view, the actual inversion from radiances to a profile is done for each sounding separately. The nonlinear scheme does not even use any horizontal consistence prior "information" about clouds. If the retrieval were made part of the model assimilation, the derived profiles would be constrained to be consistent with the model's physical parametrisation of cloud, and with other observations, both from the same instrument, and others. Since the assimilation method used at the Met Office (Lorenc Bell and Macpherson 1991) is iterative, it would be feasible to combine the iterative retrieval with the assimilation, by making the iteration module of GLOSS a subroutine of the assimilation. We hope to perform some simple tests along these lines. ECMWF are developing a combined retrieval (of clear-column radiances) and analysis (3DVAR), using the variational method set out in section 3. However it will be hard to extend this to use cloud radiances (Eyre, personal communication).

It is necessary when planning future operational systems to consider the computer requirements. In setting up GLOSS we estimate that for: ~75000 soundings per satellite per day of ~20 cloud-cleared channels, at ~120km spacing, GLOSS will take ~3500 CPUsec/day, on the CRAY YMP.

To extrapolate this we need the following factors:

2 satellites	2	
40km spacing	9	
cloudy radiances	4	(extra iterations)
ATOVS (1995?)	2	(~40 channels instead of 20)
IASI or AIRS (1998?)	100	(~2000 channels instead of 20)

Thus for ATOVS we need 100 Cray YMP single CPU hours per day. Although the computer is a multi processor, and is to be upgraded, this is still a very significant part of the expected capacity. We will have to justify this with a significant benefit, and probably cut back on the ideal system (e.g. by not processing every sounding; 40km resolution may not be needed globally).

For IASI or AIRS we might need 50 times the computer power of needed for ATOVS. By the end of the decade, we are likely to have available massively parallel computers that can provide the required raw CPU power. It will however be difficult to apply these computers to the very large and complex forecasting systems used today in NWP. Reverting to a system dedicated to the retrieval problem may be the most economical way of providing the power. This however rules out our ideas about a closer coupling of the retrieval and assimilation.

3 BAYESIAN DERIVATION OF INVERSION, AND NONLINEARITIES

3.1 **Bayes Theorem**

Notation	
$\mathbf{x}$	atmosphere as represented in model
$\mathbf{x}_t$	model representation of the true state of the atmosphere
$\mathbf{x}_b$	prior estimate of $\mathbf{x}_t$ (e.g. from forecast)
$\mathbf{y}$	observations
$\mathbf{y}_t$	observations that would be given by error-free instruments
$K(\mathbf{x})$	forward operator for calculating $\mathbf{y}$ from $\mathbf{x}$
$\mathbf{K}$	tangent linear operator of K, such that $K(\mathbf{x}+\delta\mathbf{x})=K(\mathbf{x})+\mathbf{K}\delta\mathbf{x}+O(\delta\mathbf{x}^2)$.
P	probability
p	probability distribution function
$P(\mathbf{x})$	= probability that $\mathbf{x}\le\mathbf{x}_t<\mathbf{x}+\mathbf{dx}$ = $p(\mathbf{x})\mathbf{dx}$
$P(A\|B)$	the conditional probability of A, given B.

For satellite retrievals, $\mathbf{x}$ is an atmospheric profile, $\mathbf{y}$ is the set of radiances, and K is the radiative transfer equation. For assimilation, $\mathbf{x}$ is the model state, $\mathbf{y}$ is the set of all observations, and k is an interpolation or extrapolation (including an extrapolation in time).

This derivation follows Lorenc (1986). Probabilities are used in a Bayesian way to describe the state of information. We have some prior information[1]

---

[1]Note that the whole retrieval philosophy is based on the idea that there is some useful prior information about $\mathbf{x}$. Hence the name "prior", or

about $\mathbf{x}$. We add to this information from observations $\mathbf{y}$. We need to know the posterior knowledge about $\mathbf{x}$. Operator K does not have a normal inverse. Since we start with our prior knowledge, all probabilities are conditional on knowing $\mathbf{x}_b$. To simplify notation we write $P(\cdot)$ instead of $P(\cdot|\mathbf{x}_b)$. Basic relationships for conditional probabilities give:

$$P(\mathbf{x}\cap\mathbf{y}) = P(\mathbf{x}|\mathbf{y}) \quad P(\mathbf{y}) \quad = \quad P(\mathbf{y}|\mathbf{x}) \quad P(\mathbf{x})$$
$$= p_a(\mathbf{x}|\mathbf{y})\mathbf{dx}\; p_{ofb}(\mathbf{y})\mathbf{dy} = p_{of}(\mathbf{y}|\mathbf{x})\mathbf{dy}\; p_b(\mathbf{x})\mathbf{dx}$$

where

$P(\mathbf{x}|\mathbf{y})$ $=p_a(\mathbf{x}|\mathbf{y})\mathbf{dx}$, is the analysis probability, i.e. the probability that $\mathbf{x}\le\mathbf{x}_t<\mathbf{x}+\mathbf{dx}$, given the background $\mathbf{x}_b$ and the observations.

$P(\mathbf{y})$ $=p_{ofb}(\mathbf{y})\mathbf{dy}$, is the probability of getting observations $\mathbf{y}$.

$P(\mathbf{y}|\mathbf{x})$ $=p_{of}(\mathbf{y}|\mathbf{x})\mathbf{dy}$ is the probability of getting observations $\mathbf{y}$ given that $\mathbf{x}=\mathbf{x}_t$.

$P(\mathbf{x})$ $=p_b(\mathbf{x})\mathbf{dx}$, is the probability that $\mathbf{x}\le\mathbf{x}_t<\mathbf{x}+\mathbf{dx}$, given only the prior knowledge of $\mathbf{x}_b$.

$$p_{of}(\mathbf{y}|\mathbf{x}) = \int p_o(\mathbf{y}|\mathbf{y}_t\cap\mathbf{x})\; p_f(\mathbf{y}_t|\mathbf{x})\; \mathbf{dy}_t$$

$p_o(\mathbf{y}|\mathbf{y}_t\cap\mathbf{x})$ is the instrumental error distribution.

$p_f(\mathbf{y}_t|\mathbf{x})$ is the forward operator error distribution.

$$p_{ofb}(\mathbf{y}) = \int p_{of}(\mathbf{y}|\mathbf{x})\; p_b(\mathbf{x})\; \mathbf{dx}$$
$$= \int\int p_o(\mathbf{y}|\mathbf{y}_t\cap\mathbf{x})\; p_f(\mathbf{y}_t|\mathbf{x})\; \mathbf{dy}_t\; p_b(\mathbf{x})\; \mathbf{dx}$$

Bayes' Theorem can be derived from the above:

$$p_a(\mathbf{x}|\mathbf{y}) = p_{of}(\mathbf{y}|\mathbf{x})\; p_b(\mathbf{x}) \;/\; p_{ofb}(\mathbf{y})$$

"background", is to be preferred to "first-guess", which is often used. An ideal iterative method should be independent of the first-guess, it should not be independent of the background.

$$= \frac{\int p_o(\mathbf{y}|\mathbf{y}_t \cap \mathbf{x})\ p_f(\mathbf{y}_t|\mathbf{x})\ d\mathbf{y}_t\ p_b(\mathbf{x})}{\int\int p_o(\mathbf{y}|\mathbf{y}_t \cap \mathbf{x})\ p_f(\mathbf{y}_t|\mathbf{x})\ d\mathbf{y}_t\ p_b(\mathbf{x})\ d\mathbf{x}}$$

This p.d.f. describes our total posterior information about $\mathbf{x}$, given $\mathbf{x}_b$ and $\mathbf{y}$. It is exact; no assumptions have been made. The information content, following Shannon's ideas of information entropy (e.g. see Tarantola 1987 p28) is given by

$$\mathcal{I} = \int p_a(\mathbf{x}|\mathbf{y})\ \ln\left(\frac{p_a(\mathbf{x}|\mathbf{y})}{p_b(\mathbf{x})}\right)\ d\mathbf{x}.$$

Note that this definition makes no allowance for the relative importance or usefulness of the different components of $\mathbf{x}$.

3.2 **Linear retrieval for Gaussian p.d.f.s**

If we assume all the p.d.f.s are Gaussian, and K can be linearized in the region of $\mathbf{x}_b$ and $\mathbf{x}_a$ such that

$$K(\mathbf{x}_a) = K(\mathbf{x}_b) + \mathbf{K}\ (\mathbf{x}_a - \mathbf{x}_b),$$

then, using the properties of Gaussians described in the appendix, we have:
the background error distribution:

$$p_b(\mathbf{x}) = \mathrm{N}(\mathbf{x}_b - \mathbf{x}, \mathbf{B}),$$

the instrumental error distribution:

$$p_o(\mathbf{y}|\mathbf{y}_t \cap \mathbf{x}) = \mathrm{N}(\mathbf{y} - \mathbf{y}_t, \mathbf{O}),$$

the forward operator error distribution:

$$p_f(\mathbf{y}_t|\mathbf{x}) = \mathrm{N}(\mathbf{y}_t - K(\mathbf{x}), \mathbf{F}),$$

where **B O** and **F** are covariances.
The observational error distribution, (knowing $\mathbf{x}_t$) is given by:

$$p_{of}(\mathbf{y}|\mathbf{x}) = \int p_o(\mathbf{y}|\mathbf{y}_t \cap \mathbf{x})\ p_f(\mathbf{y}_t|\mathbf{x})\ d\mathbf{y}_t$$

$$= N(\mathbf{y}-K(\mathbf{x}_t),\mathbf{O}+\mathbf{F})$$

where **O+F** (=**E**) is the observational error covariance.
The observation distribution, (only knowing $\mathbf{x}_b$) is given by:

$$p_{ofb}(\mathbf{y}) = N(\mathbf{y}-K(\mathbf{x}_b),\mathbf{O}+\mathbf{F}+\mathbf{KBK}^T).$$

Substituting in Bayes' Theorem gives:

$$\begin{aligned} p_a(\mathbf{x}|\mathbf{y}) &= p_{of}(\mathbf{y}|\mathbf{x})\ p_b(\mathbf{x})\ /\ p_{ofb}(\mathbf{y}) \\ &= N(\mathbf{y}-K(\mathbf{x}_t),\mathbf{O}+\mathbf{F})\ N(\mathbf{x}_b-\mathbf{x},\mathbf{B})\ /\ N(\mathbf{y}-K(\mathbf{x}_b),\mathbf{O}+\mathbf{F}+\mathbf{KBK}^T) \\ &= N(\mathbf{x}_a-\mathbf{x},\mathbf{A}). \end{aligned}$$

where $\mathbf{x}_a$ and **A** are defined by

$$\mathbf{A} = \mathbf{B} - \mathbf{BK}^T(\mathbf{KBK}^T+\mathbf{E})^{-1}\mathbf{KB}$$

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{BK}^T(\mathbf{KBK}^T+\mathbf{E})^{-1}(\mathbf{y}-K(\mathbf{x}_b)).$$

3.3 **Variational retrievals**

If K is more nonlinear, or the p.d.f.s are non-Gaussian, then the direct solution derived above cannot be used, although the Bayes' Theorem for the analysis p.d.f. is still valid:

$$p_a(\mathbf{x}|\mathbf{y}) = p_{of}(\mathbf{y}|\mathbf{x})\ p_b(\mathbf{x})\ /\ p_{ofb}(\mathbf{y}).$$

The expression for p_a which results is usually too complicated to be very useful in describing our knowledge about **x**; we want an estimate of the "best", most probable, **x**. In order to avoid a definition of "best" which depends of the choice of model basis for **x**, we normalize the probabilities with respect to a null prior, $p_{null}(\mathbf{x})$, a probability distribution which describes our knowledge before we knew even $\mathbf{x}_b$. We further assume, for simplicity, that the basis for **x** has been chosen so that $p_{null}(\mathbf{x})$ is a constant over the range of values of interest[2]. Then the most probable **x** is

---

[2]Note that if we change variables from **x** to **u** by a nonlinear mapping, then $p_{null}(\mathbf{u})$ is not a constant.

that which maximizes $p_a(\mathbf{x}|\mathbf{y})/p_{null}(\mathbf{x})$. Since $p_{null}(\mathbf{x})$ and $p_{ofb}(\mathbf{y})$ are independent of $\mathbf{x}$, this is the same as the $\mathbf{x}$ which minimizes a penalty function $\mathcal{J}$ given by

$$\mathcal{J} = -\ln(p_{of}(\mathbf{y}|\mathbf{x})) - \ln(p_b(\mathbf{x})).$$

If we substitute the Gaussian p.d.f.s of the last section into this, we get:

$$\mathcal{J} = \tfrac{1}{2}(\mathbf{y}-K(\mathbf{x}))^T(\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y}-K(\mathbf{x})) + \tfrac{1}{2}(\mathbf{x}_b-\mathbf{x})^T\mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x}) + \text{constant}.$$

If furthermore we make K linearizable, we see why the linear problem with Gaussians is easier to solve: $\mathcal{J}$ becomes a quadratic in $\mathbf{x}$. Using the same algebraic manipulations as are needed to establish the properties of Gaussians used in the last section, gives:

$$\mathcal{J} = \tfrac{1}{2}(\mathbf{x}_a-\mathbf{x})^T\mathbf{A}^{-1}(\mathbf{x}_a-\mathbf{x}) + \text{constant}.$$

where $\mathbf{x}_a$ and $\mathbf{A}$ are defined by

$$\mathbf{A} = \mathbf{B} - \mathbf{BK}^T(\mathbf{KBK}^T+\mathbf{O}+\mathbf{F})^{-1}\mathbf{KB}$$

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{BK}^T(\mathbf{KBK}^T+\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y}-K(\mathbf{x}_b)).$$

If K cannot be linearized over the whole range containing $\mathbf{x}_b$ and possible $\mathbf{x}_a$s, then an explicit solution is not possible. If K is still differentiable, so that

$$K(\mathbf{x}+\delta\mathbf{x}) = K(\mathbf{x})+\mathbf{K}_\mathbf{x}\delta\mathbf{x}, \quad \text{as } \delta\mathbf{x}\to 0$$

then we can look for the minimum of $\mathcal{J}$ using a descent algorithm. At the minimum, the gradient of $\mathcal{J}$ with respect to the components of $\mathbf{x}$ is zero:

$$\mathcal{J}' = -\mathbf{K}_\mathbf{x}{}^T(\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y}-K(\mathbf{x})) - \mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x}) = 0.$$

This formula is exact; we can find the most probable $\mathbf{x}$. The next stage of generalization is to allow the p.d.f.s to be weakly non-Gaussian. That is, we use the Gaussian formulae with $\mathbf{O}_\mathbf{x}$ $\mathbf{F}_\mathbf{x}$ and $\mathbf{B}_\mathbf{x}$ being slowly varying functions of $\mathbf{x}$, whose derivatives we can neglect. We also neglect derivatives of $\mathbf{K}_\mathbf{x}$. Then if we define $\mathbf{x}_a$ as the $\mathbf{x}$ which minimizes $\mathcal{J}$, i.e.

$$\mathcal{J}' = -\mathbf{K}_{\mathbf{x}a}{}^T(\mathbf{O}_{\mathbf{x}a}+\mathbf{F}_{\mathbf{x}a})^{-1}(\mathbf{y}-K(\mathbf{x}_a)) - \mathbf{B}_{\mathbf{x}a}{}^{-1}(\mathbf{x}_b-\mathbf{x}_a) = 0.$$

Then

$$\mathcal{J}'' \cong \mathbf{K}_{\mathbf{x}a}{}^{T}(\mathbf{O}_{\mathbf{x}a}+\mathbf{F}_{\mathbf{x}a})^{-1}\mathbf{K}_{\mathbf{x}a} + \mathbf{B}_{\mathbf{x}a}{}^{-1} = \mathbf{A}^{-1}.$$

Then, in the neighbourhood of $\mathbf{x}_a$,

$$p_a(\mathbf{x}|\mathbf{y}) \propto N(\mathbf{x}_a-\mathbf{x}, \mathcal{J}''^{-1}).$$

If K is sufficiently nonlinear, or the p.d.f.s are sufficiently non-Gaussian, $p_a(\mathbf{x}|\mathbf{y})$ may have multiple maxima. We have then to consider how to decide which is best. It may well be that the "best" $\mathbf{x}$ is that which is most likely to be in a finite, rather than an infinitesimal, region around $\mathbf{x}_t$. This likelihood may then be more related to the integral under the above approximate Gaussian, rather than its peak value. To find the $\mathbf{x}$ which maximizes this integral, we should not search for the maximum of $p_a(\mathbf{x}|\mathbf{y})$, but rather $|\mathbf{A}|^{1/2}p_a(\mathbf{x}|\mathbf{y})$. Note that between local maxima, $|\mathcal{J}''|$ may become zero. In these regions our assumptions about slowly varying $\mathbf{K}$ $\mathbf{B}$ $\mathbf{O}$ and $\mathbf{F}$ are not good. However our approximation $|\mathbf{A}^{-1}|$ remains positive.

3.4 Weight given to Background

If we perturb $\mathbf{x}_a$ to $\mathbf{x}_a+\delta\mathbf{x}_a$, and $\mathbf{x}_b$ to $\mathbf{x}_b+\delta\mathbf{x}_b$, then

$$\mathcal{J}' = \mathcal{J}''\delta\mathbf{x}_a - \mathbf{B}^{-1}\delta\mathbf{x}_b.$$

So to remain a solution of $\mathcal{J}'=0$

$$\delta\mathbf{x}_a = \mathcal{J}''^{-1}\ \mathbf{B}^{-1}\delta\mathbf{x}_b = \mathbf{A}\ \mathbf{B}^{-1}\delta\mathbf{x}_b.$$

This can be used to find the $\tilde{\mathbf{x}}_a$ which would have been found using a slightly different background $\tilde{\mathbf{x}}_b$, without repeating the full nonlinear solution.
If we wish to repeat the analysis, using a slightly different background and perhaps additional observations, then we can use

$$\tfrac{1}{2}(\mathbf{y}-K(\mathbf{x}))^T(\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y}-K(\mathbf{x})) = \tfrac{1}{2}(\mathbf{x}_a-\mathbf{x})^T\mathbf{A}^{-1}(\mathbf{x}_a-\mathbf{x}) - \tfrac{1}{2}(\mathbf{x}_b-\mathbf{x})^T\mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x}) + \text{const.}$$

Then we have the new penalty given by

$$\mathcal{J} = \tfrac{1}{2}(\mathbf{x}_a-\mathbf{x})^T\mathbf{A}^{-1}(\mathbf{x}_a-\mathbf{x}) - \tfrac{1}{2}(\mathbf{x}_b-\mathbf{x})^T\mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x}) + \tfrac{1}{2}(\tilde{\mathbf{x}}_b-\mathbf{x})^T\mathbf{B}^{-1}(\tilde{\mathbf{x}}_b-\mathbf{x})$$

+ penalty term for additional observations + constant.

If we define $\mathbf{C}$ and $\mathbf{x}_c$ by

$$\mathbf{C}^{-1} = \mathbf{A}^{-1} - \mathbf{B}^{-1}$$

$$\mathbf{C}^{-1}\mathbf{x}_c = \mathbf{A}^{-1}\mathbf{x}_a - \mathbf{B}^{-1}\mathbf{x}_b.$$

Or equivalently

$$\mathbf{C} = (\mathbf{I}-\mathbf{A}\mathbf{B}^{-1})^{-1}\mathbf{A}$$

$$\mathbf{x}_c = \mathbf{x}_b + (\mathbf{I}-\mathbf{A}\mathbf{B}^{-1})^{-1}(\mathbf{x}_a-\mathbf{x}_b).$$

Then

$$\mathcal{J} = \tfrac{1}{2}(\mathbf{x}_c-\mathbf{x})^T\mathbf{C}^{-1}(\mathbf{x}_c-\mathbf{x}) + \tfrac{1}{2}(\tilde{\mathbf{x}}_b-\mathbf{x})^T\mathbf{B}^{-1}(\tilde{\mathbf{x}}_b-\mathbf{x})$$

+ penalty term for additional observations + constant.

$\mathbf{x}_c$ can be thought of as a linearized representation of the information in the original observations, without any information from $\mathbf{x}_b$. It can be reused in a subsequent analysis (with the same or a different background) as a pseudo-observation.

In practice, things are not this simple. Because of the null-space not properly defined by the observations, $(\mathbf{I}-\mathbf{A}\mathbf{B}^{-1})$ is not invertible. A pseudo-inverse can however be defined by excluding the null-space (Purser 1990).

The method used in assimilating the LASS retrievals for the second experiment shown in figure 1, was a simplified version of this (Lorenc *et al.* 1986), appropriate for the horizontal and vertical splitting of the assimilation method (Lorenc *et al.* 1991).

4 DISCUSSION OF THE FUTURE

Experience with the current generation of sounders, in particular the failure despite years of effort to get consistent positive impact in the northern hemisphere, until the forecast background methods were developed, indicates that for data assimilation uses we should stay with the forecast background retrieval method. In this, information from the model is available during the retrieval, and only the new information from the sounding is used in the model. (Note that this conclusion may not apply to other uses, such as atmospheric studies, and climate change detection.)

Improved cloud and moisture information in the soundings from future instruments, and in the NWP models and GCMs currently being developed, means that this two-way interaction will be extended to cloud and moisture parameters.

There are theoretical advantages in combining the retrieval and assimilation even more closely, so that the interaction takes place within a single (iterative) solution of the combined inverse problem.

However this will require a large computer system, capable of simultaneously handling both the retrieval and assimilation steps. It may be more cost effective to perform a forecast background retrieval on one system, and use it on another. A linearized correction to the retrieval, to take account of the background information used, is possible during the assimilation process. This requires that background profile used for the retrieval, and some covariance statistics, to be passed to the assimilation along with the retrieval. (Unfortunately the effect of cloud are nonlinear, so an accurate correction for changes in prior cloud information is not possible in this method).

Until such time as decisions have to be made about computer and system architecture, it would be wise to develop retrieval and assimilation as compatible systems, to keep both options open. Future decisions must take account of the operational benefits achieved, relative to the effort expended.

REFERENCES

Baker, Wayman. 1991. "Joint NESDIS/NASA/NMC effort to develop an advanced satellite retrieval system". A.M.S. Preprints of 9th Conf. on NWP. 113-114.

Bell,R.S., and Olive M Hammon, 1990. "The impact of satellite sounding data in the Meteorological Office fine-mesh model". Preprints, WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, Cleremont-Ferand, France. July 1990.

Eyre,J.R. 1989. "Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation: theory and simulation for TOVS." Quart. J. Roy. Met. Soc., 115, 1001-1026.

Eyre,J.R. 1989. "Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation: application to TOVS data." Quart. J. Roy. Met. Soc., 115, 1027-1038.

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

Eyre,J.R., and Watts,P.D. 1987. "A sequential estimation approach to cloud clearing for satellite temperature sounding" Quart.J.Roy.Met.Soc., 113, 1349-1376.

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

Lorenc,A.C., Bell,R.S., and Macpherson,B. 1991. "The Meteorological Office "Analysis Correction" data assimilation scheme". Quart. J. Roy. Met. Soc. 117, 59-89

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.

Purser,R.J. 1990. "Vertical aspects of the assimilation of sounding data". Preprints, WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, Cleremont-Ferand, France. July 1990.

Tarantola,A. 1987. "Inverse Problem Theory - Methods for data fitting and model parameter estimation". publ. Elsevier. 613pp. ISBN 0-444-42765-1

Turner,J., Eyre,J.R., Jerrett,D., and McCallum,E. 1985. "The HERMES System". Meteorol.Mag., 114, 161-173.

Appendix PROPERTIES OF GAUSSIAN DISTRIBUTIONS

A multi-dimensional Gaussian ("normal") distribution is described by the function:

$$N(\mathbf{x},\mathbf{B}) = ((2\pi)^N|\mathbf{B}|)^{-1/2}\exp(-\tfrac{1}{2}\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x})$$

where $\mathbf{B}$ is an $N\times N$ positive definite matrix, and $|\mathbf{B}|$ is its determinant.

Mean of $\mathbf{x}$ $= \int \mathbf{x}\ N(\mathbf{x},\mathbf{B})\ d\mathbf{x} = 0.$

Covariance of $\mathbf{x}$ $= \int \mathbf{x}\mathbf{x}^T N(\mathbf{x},\mathbf{B})\ d\mathbf{x} = \mathbf{B}.$

Product of two Gaussians:

$$N(\mathbf{x}_b-\mathbf{x},\mathbf{B})\ N(\mathbf{x}_c-\mathbf{x},\mathbf{C}) = N(\mathbf{x}_b-\mathbf{x}_c,\mathbf{B}+\mathbf{C})\ N(\mathbf{x}_a-\mathbf{x},\mathbf{A})$$

where $\mathbf{x}_a$ and $\mathbf{A}$ are defined by

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{C}^{-1}$$

$$\mathbf{A}^{-1}\mathbf{x}_a = \mathbf{B}^{-1}\mathbf{x}_b + \mathbf{C}^{-1}\mathbf{x}_c.$$

Hence the convolution of two Gaussians:

$$\int N(\mathbf{x}_b-\mathbf{x},\mathbf{B})\ N(\mathbf{x}_c-\mathbf{x},\mathbf{C})\ d\mathbf{x} = N(\mathbf{x}_b-\mathbf{x}_c,\mathbf{B}+\mathbf{C}) \int N(\mathbf{x}_a-\mathbf{x},\mathbf{A})\ d\mathbf{x}$$

$$= N(\mathbf{x}_b-\mathbf{x}_c,\mathbf{B}+\mathbf{C}).$$

Product of two Gaussians in variables related by an operator K:

As long as K can be linearized in the region of $\mathbf{x}_b$ and $\mathbf{x}_a$ such that

$$K(\mathbf{x}_a) = K(\mathbf{x}_b) + \mathbf{K}\ (\mathbf{x}_a-\mathbf{x}_b),$$

then

$$N(\mathbf{x}_b-\mathbf{x},\mathbf{B})\ N(\mathbf{y}-K(\mathbf{x}),\mathbf{E}) = N(\mathbf{y}-K(\mathbf{x}_b),\mathbf{E}+\mathbf{KBK}^T)\ N(\mathbf{x}_a-\mathbf{x},\mathbf{A})$$

where $\mathbf{x}_a$ and $\mathbf{A}$ are defined by

$$\mathbf{A} = \mathbf{B} - \mathbf{BK}^T(\mathbf{KBK}^T+\mathbf{E})^{-1}\mathbf{KB}$$

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{BK}^T(\mathbf{KBK}^T+\mathbf{E})^{-1}(\mathbf{y}-K(\mathbf{x}_b)).$$

If $K(\mathbf{x})=\mathbf{Kx}$, then equivalent definitions for $\mathbf{x}_a$ and $\mathbf{A}$ are more like those earlier:

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{K}^T\mathbf{E}^{-1}\mathbf{K}$$

$$\mathbf{A}^{-1}\mathbf{x}_a = \mathbf{B}^{-1}\mathbf{x}_b + \mathbf{K}^T\mathbf{E}^{-1}\mathbf{y}.$$

Hence the convolution of two Gaussians:

$$\int N(\mathbf{x}_b-\mathbf{x},\mathbf{B})\ N(\mathbf{y}-\mathbf{Kx},\mathbf{E})\ d\mathbf{x} = N(\mathbf{y}-\mathbf{Kx}_b,\mathbf{E}+\mathbf{KBK}^T) \int N(\mathbf{x}_a-\mathbf{x},\mathbf{A})\ d\mathbf{x}$$

$$= N(\mathbf{y}-\mathbf{K}\mathbf{x}_b, \mathbf{E}+\mathbf{K}\mathbf{B}\mathbf{K}^T).$$