THE EFFECTS OF NONLINEARITY ON RETRIEVAL ERRORS: IMPLICATIONS FOR
THE INTERPRETATION OF ADVANCED INFRA-RED SOUNDER DATA

J R Eyre and A D Collard
Meteorological Office
Bracknell, UK

1. INTRODUCTION

When analyzing meteorological fields, it is important to consider sources of significant error, both in the observations available to the analysis and in the analysis procedure itself. For mathematically optimal results, it is necessary to account correctly for the error characteristics of the observations and of any other information used in the analysis. For practical applications, where sub-optimal results are usually acceptable, it is nevertheless important to characterize significant error sources approximately; serious deficiencies in the analysis usually result if an important source of error is overlooked or greatly under-estimated.

Amato et al. (1996) have drawn attention to a particular problem caused by nonlinear relationships between some observed quantities and those to be analyzed. They have demonstrated detrimental effects in the case of retrieving atmospheric temperature profiles from spectra of infra-red radiances. Similar effects are potential sources of significant error for general problems of this type: when retrieving/analyzing a set of geophysical variables from a set of observations related to them in a nonlinear way (see Joiner and da Silva 1998; Stoffelen 1998, section VI.1.2).

This paper examines the errors introduced by these nonlinearities within analysis/retrieval schemes that are formulated to be optimal in the absence of such effects. In particular, it explores the theoretical error characteristics of nonlinear variational methods now in use for analysis of the 3D atmospheric state from global observations (e.g. Andersson et al. 1994) or the retrieval of atmospheric temperature and humidity profiles from satellite sounding radiances (e.g. Eyre et al. 1993). It discusses how these errors will tend to manifest themselves and suggests some approaches to estimating their effects, and to mitigating them.

As an illustration, the effects of nonlinearity errors have been quantified for the Infra-red Atmospheric Sounding Interferometer (IASI) (Phulpin et al. 1997). Previous error analyses based on linear theory (e.g. Amato et al. 1995, Collard 1998) have suggested that the information on temperature in IASI water vapour channels is surprisingly high. However, linear error analysis is equivalent to assuming that the weighting functions are known exactly. In practice, because of uncertainties in humidity (and, to a lesser extent, temperature), they are only known to finite accuracy. Does this lead to an over-optimistic assessment of the temperature information available from these channels? We attempt to answer this question.
2. THEORY

2.1 Solution of the variational problem

The problem of variational analysis is usually posed as one of minimising a penalty function containing a number of quadratic terms, each representing a source of information on the variables to be analyzed (e.g., see Lorenc 1986, Le Dimet and Talagrand 1986). Using the notation of Ide et al. (1997), the penalty function for a combination of "observed" information with "background" information is:

\[
J(x) = \frac{1}{2}(x-x^b)^T B^{-1}(x-x^b) + \frac{1}{2}(y^o-H(x))^T (E+F)^{-1}(y^o-H(x))
\]

(1)

where \( x \) a vector representing the state to be analyzed, \( x^b \) is an estimate of \( x \) obtained from "prior" or "background" information, \( B \) is the error covariance of \( x^b \), \( y^o \) a vector of observations, \( H(x) \) is a vector containing equivalent values in the observation space corresponding to the state \( x \) and is calculated through a "forward model" (or "observation operator"), \( H(\ldots) \), \( E \) is the error covariance of the observations, and \( F \) is the error covariance of the forward model. \( ^T \) and \( ^{-1} \) denote matrix transpose and inverse respectively.

In general, the retrieval/analysis problem may be nonlinear in several ways. For example, \( B, E \) or \( F \) could be a function of the state \( x \) or the measurements \( y^o \). However, the only potential nonlinearity represented in eq.(1) is in the observation operator \( H(x) \), and only this aspect of nonlinearity is considered in this paper.

Eq.(1) is minimised by solving its gradient equation:

\[
\nabla_x J(x) = B^{-1}(x-x^b) - H'(x)^T (E+F)^{-1}(y^o-H(x)) = 0,
\]

(2)

where \( H'(x) = \nabla_x H(x) \). We define the solution of this equation as the analysis, \( x = x^a \). Then,

\[
B^{-1}(x^a-x^b) = H'(x^a)^T (E+F)^{-1}(y^o-H(x^a)).
\]

(3)

Lorenc (1986) shows that this solution is the most probable solution for \( x \) (i.e. the mode of its probability density function), if the error characteristics of observations and background are Gaussian, and observation errors are uncorrelated with background errors. There is no assumption of a linear observation operator in the derivation of the maximum probability solution; \( H(x) \) can be a nonlinear function of \( x \). However, this solution of the problem will only be truly optimal if the error characteristics of the real problem correspond to those we assume in the analysis of the data.

Using \( ^t \) to denote true values, we can re-write eq.(3) as
$$B^{-1}(x^a-x^b;x^b+y^b) = H'(x^a)^T.(E+F)^{-1}.(y^b-H(x^a)+y^b).$$  \hspace{1cm} (4)

Introducing $e^a$, $e^b$ and $e^f$ for the errors in $x^a$, $x^b$ and $y^o$, and defining forward model error, $e^f = H(x^f)-y^f$,

$$B^{-1}(e^a-e^b) = H'(x^a)^T.(E+F)^{-1}.(e^a-e^b).H(x^a)+H(x^f).$$ \hspace{1cm} (5)

2.2 The linear case

For the linear problem, i.e. where $H(x)$ is a linear function of $x$, and so $H'(x) = H = a$ constant, we can solve eq.(3) analytically to give

$$x^a = x^b + K.(y^o-H(x^b))$$ \hspace{1cm} (6)

where $K = B.H'^T.(H'.B.H'^T+E+F)^{-1}$. \hspace{1cm} (7)

$x^a$ is "optimal" in the following respects: if $x^b$ and $y^o$ are unbiased, then $x^a$ is unbiased and has minimum error variance; additionally, if the errors of $x^b$ and $y^o$ are Gaussian, $x^a$ also corresponds to the maximum likelihood solution.

It can be shown (see Rodgers 1976) that the covarion of $e^a$ is given by

$$A^{-1} = B^{-1} + H'^T.(E+F)^{-1}.H'$$ \hspace{1cm} (8)

or

$$A = B - K.H'.B = (I-K.H').B.$$ \hspace{1cm} (9)

It can also be shown (from eq.(5), or see Eyre 1987) that

$$e^a = A.B^{-1}.e^b + A.H'^T.(E+F)^{-1}.(e^f-e^f)$$ \hspace{1cm} (10)

or

$$e^a = (I-K.H').e^b + K.(e^f-e^f).$$ \hspace{1cm} (11)

From this equation for the error in a single retrieval/analysis, an equivalent equation for the mean error of an ensemble of retrievals/analyses can be obtained in terms of the means of the various components of error.

$$<e^b> = (I-K.H').<e^b> + K.<e^f>-<e^f>,$$ \hspace{1cm} (12)

where $<...>$ denotes a mean value.

Clearly, if $e^b$, $e^f$ and $e^f$ are all unbiased, then $e^a$ will also be unbiased; otherwise eq.(12) gives us a method for assessing the propagation of biases from different sources into the solution.
Let us denote the error, the mean error and the covariance of error for the linear case, given by equations (11), (12) and (8), as \( \varepsilon^a, \langle \varepsilon^a \rangle \) and \( A_L \) respectively.

### 2.3 The case of the nonlinear observation operator

When \( y(x) \) is nonlinear, two types of problem arise. The first concerns the practical problems of finding the solution \( x^a \). Minimisation methods for solution of eq.(2) or (3), or their equivalents, tend to become slower to converge as nonlinearity increases. Also the nonlinearity may introduce multiple minima, making it more difficult to find the global minimum. These may indeed be real problems in practice, but they are not the primary concern here. We shall assume that \( x^a \), corresponding to a suitable minimum of \( J(x) \), can be found and that we are concerned with the second type of problem, namely the error characteristics of \( x^a \).

With the introduction of nonlinearity, the modal solution of eq.(2) or (3) is no longer the minimum variance solution. However, it is not our intention here to re-examine the various possible definitions of the "optimal" solution. We accept that the solution of eq.(2) is our "best" estimate, but we are interested to understand its error characteristics.

Following Amato et al. (1996) we expand the forward model \( H(x) \) as a Taylor series (except that here we do so around the analyzed value \( x^a \)):

\[
H(x') = H(x^a) + H'(x^a)(x' - x^a) + \text{higher order terms in } (x' - x^a).
\]  

(13)

The higher order terms give rise to a "nonlinearity error" \( \varepsilon^a \) which, from eq.(13), is given by:

\[
\varepsilon^a = H(x^a) - H(x') - H'(x^a)\varepsilon^a.
\]  

(14)

\( \varepsilon^a \) arises because \( H'(x^a) \) differs from \( H'(x') \). Figure 1 illustrates a scalar example.

Substituting eq.(14) into eq.(5) gives

\[
B^{-1}(\varepsilon^a - \varepsilon^b) = H'(x^a)^T(E+F)^{-1}(\varepsilon^a - \varepsilon^b) - H'(x^a).\varepsilon^a.
\]  

(15)

or

\[
(B^{-1} + H'(x^a)^T(E+F)^{-1}H'(x^a))\varepsilon^a = B^{-1}\varepsilon^a + H'(x^a)^T(E+F)^{-1}(\varepsilon^a - \varepsilon^b),
\]  

(16)

from which it can be shown that

\[
\varepsilon^a = \varepsilon^a_{L} - K(x^a)\varepsilon^a.
\]  

(17)
Figure 1. Illustrating the source of nonlinearity error for a scalar case.

Figure 2. Illustrating, for a scalar case, the iterative estimation of nonlinearity error $N$ and analysis error $A$, starting either from the analysis error in the linear limit $A_L$ or from the background error $B$. 
At this point it is possible simply to absorb the nonlinearity error $\epsilon^n$ as a component of the forward model error $\epsilon$. Then equations (8) to (12) can be applied to the nonlinear case. However, because the characteristics of the nonlinearity error are different from those of other sources of forward model error (as will be shown below in section 3.1), it is instructive to keep $\epsilon^n$ separate.

The mean error and covariance of error for the nonlinear case can now be expressed in terms of their linear counterparts as follows:

$$<\epsilon^b> = <\epsilon^l> - K(x^b),<\epsilon^n>, \tag{18}$$

and

$$A = A_L + K(x^b)N.K(x^b)^T - K(x^b)Q - Q^T.K(x^b)^T, \tag{19}$$

where $N$ is the covariance of nonlinearity error and $Q$ is the cross-covariance of nonlinearity error with analysis error in the linear limit (as given by eq.(11)).

Equations (18) and (19) show how the nonlinearity error (in the measurement space) is mapped by the retrieval/analysis operator $K(x^b)$ into addition terms contributing to the mean error in the retrieval/analysis and its error covariance respectively. Note that presence of this new source of error does not (in general) change the expression representing the modal retrieval/analysis, eq.(3). However, it does change the value of the error in this analysis.

Alternatively, as suggested above, nonlinearity error may be treated as a new component of forward model error and simply absorbed into $F$, i.e. $F$ is given a value in eq.(3) which includes a contribution from nonlinearity error. However, in doing so the question of "optimality" must be re-examined for the reasons discussed in section 3.2.

3. DISCUSSION

3.1 Error distributions and the introduction of bias

Equations (13) and (14) show that the nonlinearity error $\epsilon^n$ represents quadratic and higher order terms in $\epsilon^a$ in the expansion of $H(x^l)$. If the quadratic term is dominant, then $\epsilon^n$ will be quadratic in $\epsilon^a$. If all other error sources are Gaussian then, in the linear case, $\epsilon^a$ will also be Gaussian. Therefore, in the nonlinear case, $\epsilon^n$ will not have a Gaussian distribution. In fact, the quadratic relation between $\epsilon^a$ and $\epsilon^n$ will give the same sign to values of $\epsilon^n$ for all values of $\epsilon^a$, and thus the mean value of $\epsilon^n$ will be non-zero.

The effect of non-zero bias in $\epsilon^n$, if uncorrected, will cause a bias in the analysis/retrieval through eq.(18),
even if there are no other sources of bias. (i.e. $\langle e_i^2 \rangle = 0$). Such biases have been observed in practice for the problem of retrieving wind speed and direction from scatterometer data (see Stoffelen and Anderson 1997).

For the scatterometer retrieval problem, there is an additional complication: although the forward model errors are close to Gaussian in the retrieval space, they become highly non-Gaussian when mapped into the observation space, and so the solution of eq.(2) or (3) no longer represents the most probable solution.

3.2 Error covariances and optimality assumptions

Values of forward model error covariance that are appropriate for the linear case will not be appropriate to achieve a solution of minimum error variance for the nonlinear case. Even if the mean bias is corrected (e.g. empirically), the analysis will still have an error covariance given by eq.(19). This will not only represent error variances higher than for the linear case $A_L$, but higher than if the forward model error covariance is increased to account for nonlinearity error (i.e. where the error covariance is given by eq.(8) in which $F$ is replaced by $F + N$). If $F$ is adjusted in this way and the mean bias corrected, then solution accuracy will be improved. However, because analysis error and nonlinearity error are mutually dependent (through equations (14) and (17)), the covariance of the latter cannot now be estimated analytically (but see section 3.3).

Also, because the distribution of the nonlinearity error is inherently non-Gaussian, if $N$ is absorbed into $F$ then eq.(3) will no longer represent the exact maximum likelihood solution. To achieve this, we would have to construct a penalty function containing a term appropriate to the particular distribution of the nonlinearity error (i.e. proportional to the log of its probability density function - see, e.g., Lorenc 1986, Ingleby and Lorenc 1993, Andersson and Jarvinen 1999). By absorbing $N$ into $F$ we may reduce the variance of the solution error, but at the expense of departing from the maximum likelihood solution.

3.3 Assessment and mitigation of nonlinear effects

The significance of nonlinearity error will vary greatly from problem to problem. It is essentially related to the change in the value of $H'$ for typical values of analysis error. Note that there is a significant difference here from the problem presented by Amato et al. (1996): in their linearized one-step retrieval, the nonlinearity error is related to the background error, whereas in the fully nonlinear variational framework it is related to the analysis/retrieval error, which can be much smaller. (For example, for a retrieval using climatological background information, as illustrated by Amato et al., the retrieval error will be much smaller than the background error.)

When an observation is related nonlinearly to the analysis variables, the magnitude of the nonlinearity error should be investigated. Significant degradation of performance is expected when the nonlinearity error exceeds the combined measurement + forward model error but is not accounted for in the assumed statistics.

197
In practice, if forward model error is deduced empirically (e.g. from observation-background statistics), it is likely to include contributions from nonlinearity error. Similarly, biases derived from observation-background statistics will contain contributions from nonlinearity error, in addition to any caused by measurement, forward model or background biases. Empirical bias tuning procedures based on observation-background statistics will tend to compensate for these effects. However, they will not do so exactly, since the magnitudes of nonlinearity effects associated with background errors will be different from those associated with analysis errors. It is therefore important to consider in which space (analysis or measurement) and for which variable (background or analysis) it is most desirable that results be unbiased.

It will be particularly important to assess and allow for the effects of nonlinearity error when other sources of noise are low, and particularly (as shown by Amato et al.) when background constraints are non-existent or very weak.

The effects of nonlinearity error may be quantified as follows:

Method 1.

(a) For a typical analyzed profile $x^*$, calculate $H(x^*)$ and $H'(x^*)$, and also the analysis error covariance in the linear limit, $A = A_L$, using eq.(8).

(b) Calculate an ensemble of typical analysis errors, with error covariance $A$, and hence corresponding "true" profiles, $x'$, and "true" measurements, $H(x')$.

(c) Using these values in eq.(14), calculate the corresponding ensemble of nonlinearity errors.

(d) Calculate from this ensemble the mean nonlinearity error and its covariance, $N$.

(e) Add the estimate of $N$ to $F$ and re-evaluate the analysis error covariance, $A$, using eq.(8).

(f) If the nonlinearity error has changed the analysis error significantly, repeat (c) to (f).

This iterative procedure is illustrated in Fig.2. For the well-behaved functions shown here, it appears to be a convergent process. Fig.2 also suggests that the process could be initiated at (b), with nonlinearity error estimated from the background error (in place of analysis error in the linear limit).

Method 2.

Perform steps (a)-(d) as above, then at step (e) use eq.(19) to calculate analysis error, and ignore step (f). This leads to a larger analysis error variance as explained in 3.2; it represents the error covariance of the maximum probability solution, but not the solution of minimum error variance.

4. RESULTS FOR IASI

To illustrate the concepts discussed in this paper, simulations were made exploring the theoretical
performance of the IASI instrument. IASI is an infra-red interferometer due to fly on the METOP series of polar orbiting satellites. Its data will be processed on-board to provide radiance spectra of 8461 channels with apodised spectral resolution of 0.5 cm\(^{-1}\) covering the frequency range 645-2760 cm\(^{-1}\).

The effect of nonlinearity error on the retrievals was examined following the two recipes detailed in Section 3.3. For these preliminary studies, a typical mid-latitude winter profile was chosen and interpolated on to 44 temperature and 28 humidity levels. This profile was used as the analyzed profile \(x^a\), and corresponding brightness temperatures \(H(x^a)\) and their Jacobians \(H'(x^a)\) were calculated using the fast radiative transfer model RTIASI developed at ECMWF. An ensemble of "true" profiles was generated using random analysis error vectors of appropriate covariance. The matrices \(B\), \(O\) and \(F\) were chosen to be consistent with those used in other information content studies within the IASI Sounder Science Working Group (e.g. see Collard 1998). \(B\) represents the error covariance of a 6-hour forecast and was obtained from ECMWF; \(O\) was taken as the specified instrument noise for IASI; \(F\) was set as a diagonal matrix with values of \((0.2K)^2\).

The mean and standard deviation of the resultant nonlinearity error (in observation space) are shown in Fig.3. The nonlinearity error is found to be generally much smaller than the assumed instrument and forward model errors. As expected, the nonlinearity error is greatest in the region of the 6.3 \(\mu\)m water vapour band, because these elements of the Jacobians are most sensitive to the water vapour profile. However even in this region other sources of error are dominant.

For these calculations, the nonlinearity error covariance was approximated as diagonal before adding it to the forward model error. (The calculation with the full 8461x8461 nonlinearity error covariance is computationally difficult. However, in an experiment using 1400 randomly-selected channels, analysis errors obtained using the full nonlinearity error covariance matrix were not significantly different from those using the diagonal approximation.)

The effects on the expected analysis error covariance of the nonlinearity error are presented in Fig.4. When the analysis error covariance is recalculated using Method 1, only a very small increase in the analysis error variance is seen (such that further iteration of the method is unnecessary). When the revised analysis error covariance is calculated using Method 2 (eq.19), a somewhat larger error is seen, but the increase relative to the linear case is still at most only 0.1K in temperature or 0.1 in log of specific humidity.

5. CONCLUSIONS

The presence of nonlinearity in the relationship between observed and analyzed variables is a potential source of bias in the analysis and of increased error variance. The importance of these effects will need to be
studied case by case, but significant effects are to be expected when the nonlinearity error is comparable to or greater than other sources of measurement and forward model error. Moreover the magnitude of the nonlinearity error is related to the magnitude of the analysis/retrieval error itself. Section 3.3 suggests a recipe for assessing quantitatively the effects of nonlinearity error.

We have applied this recipe to simulation of IASI performance. Given typical values of retrieval error corresponding to reasonable values of measurement, forward model and background error, the nonlinearity error is found to be small - much less than other assumed sources of forward model error - even in the most nonlinear channels (in the water vapour absorption band). As a consequence, the effects on information content (compared with linear analysis) are also small. These results support previous studies which have suggested that information content on temperature in water vapour channels is high. They also suggest that, in practice, nonlinearity error will not be a significant source of additional error for variational analysis/retrieval methods applied to IASI data.

REFERENCES


Le Dimet, F. X. and Talagrand, O., 1986: Variational algorithms for analysis and assimilation of
meteorological observations: theoretical aspects. Tellus, 38A, 97-100.


Figure 3. The spectra of nonlinearity error (bias and standard deviation) for IASI, compared with the spectrum of the standard deviation of combined instrument and forward model error.
Figure 4. Standard deviation of retrieval error for IASI, both ignoring nonlinearity error and including it (methods 1 and 2), compared with standard deviation of background error, for temperature (above) and humidity (below). Also shown is retrieval bias induced by nonlinearity error.
TECHNICAL PROCEEDINGS OF THE TENTH
INTERNATIONAL ATOVS STUDY CONFERENCE

Boulder, Colorado
27 January - 2 February 1999

Edited by
J. Le Marshall and J.D. Jasper
Bureau of Meteorology Research Centre, Melbourne, Australia

Published by
Bureau of Meteorology Research Centre
PO Box 1289K, GPO Melbourne, Vic., 3001, Australia

December 1999