Beyond optimal estimation:
sensitivity of analysis error to the specification of background error

John Eyre and Fiona Hilton
Met Office, Exeter, UK

1.   Introduction

Optimal estimation theory, on which most retrieval and data assimilation methods are based, assumes that the error covariances of the observations and of the *a priori* (background) information are known, and that they are used (at least approximately) in the retrieval/analysis procedure. In data assimilation for numerical weather prediction (NWP), background errors may be known reasonably well in the sense of global statistics. However, this can disguise substantial spatial and temporal variability. Moreover, it is possible for the overall error variance to be correct whilst the partitioning of errors between different spatial scales is significantly incorrect. The specification of the magnitude and vertical structure of the background error covariance is crucial to the appropriate interpretation of radiance information from satellite sounders within a NWP data assimilation system. Uncertainties in the specification of error covariances are inevitable, but an improved understanding of the acceptable range of misspecification is likely to lead to improved impact of these data in NWP. This problem is likely to be important for advanced infra-red sounders, as we attempt to extract information on smaller vertical scales, whilst retaining the information that the NWP model usually contains on sharp vertical structures, such as the tropopause and the boundary layer top.

In this paper, we present initial investigations of the sensitivity of analysis error to the misspecification of background error. We first discuss further the motivation for the work and summarise previous work on this topic. We then present the theory for analysis/retrieval error, and for the differences between the optimal case and the general, sub-optimal case. We illustrate this theory for a simple scalar case, where we identify a "danger zone", i.e. a level of misspecification of background error that will degrade the analysis, which we quantify as a function of the ratio of observation error to background error. We then present preliminary results for a more complex case: a one-dimensional variational retrieval (1D-Var) analogue of the assimilation of radiances from the Infra-red Atmospheric Sounding Interferometer (IASI) on the MetOp satellite (Simeoni et al., 1997) into the Met Office global NWP system. We examine the vertical eigen-modes of a forecast error covariance matrix and calculate the "information content" of a IASI brightness temperature spectrum, i.e. how much the analysis error is reduced compared to the background error, in each eigen-mode. Finally we present conclusions and suggestions for further work.

2.   Background

Compared with most types of new data from satellites, initial exploitation of IASI data by the NWP community has been very rapid and successful; radiance data of high quality were made available to the community by EUMETSAT soon after launch in 2006, and several centres proceeded to demonstrate positive impact on NWP performance (e.g. Hilton et al., 2009; Collard and McNally, 2009). Nevertheless, it is clear that the full information content potentially available from IASI and other advanced IR sounder data is far from being fully exploited in NWP. Active research is in progress to seek improvements in several areas including: efficient processing of the full spectrum, observation errors including inter-channel correlations, residual biases, surface properties over land and ice, treatment of cloud, and background error statistics. In this paper we examine only the last of these.

Most data assimilation and satellite retrieval methods are based on optimal estimation (OE) theory (e.g. see Rodgers 1976, Lorenc, 1988). With this theory, the best estimate of the state of the system is derived by combining observed estimates with prior estimates, weighting each according to its error covariance. Such retrieval/analysis methods implicitly assume that the error covariances are known

whereas, in practice, they are usually known only approximately. There are two ways forward to address this problem. The first is to strive to improve our estimates of these quantities, and this is a field of continuing work. The second is to try to make assimilation/retrieval robust against our **inevitable** lack of knowledge of the error covariances. This general problem applies equally to both background and observation error covariances. In this paper, we only consider the former.

Why are background error covariances inevitably in error? Global averages of their statistics can be estimated quite accurately from the diagnostics of NWP systems, but such averages mask large spatial and temporal variability. We need therefore to understand our sensitivity to background error covariance and its inevitable mis-specification. It has also been suggested that the sensitivity may be particularly important for the assimilation of satellite radiances or other non-local observations, and this issue is explored in a companion paper (Hilton and Eyre, 2010).

Advanced infra-red sounders sounder such as IASI are often characterised as having a vertical resolution of ~1 km (Prunet et al., 1998; Collard, 1998). This is equivalent to stating that they are sensitive with low error to scales significantly greater than 1 km, and not sensitive to scales significantly smaller than1 km, but sensitive to scales of approximately1 km with errors comparable to background errors. This illustrates why it is important to understand the vertical structure of the background error covariance and its magnitude on different scales, as this will determine how measurements and prior information are weighted **on each scale**, and the effect of mis-specifying the background error on each scale.

There appears to be surprisingly little in the assimilation/retrieval literature on effects of misspecifying background errors. Some mathematical aspects are considered by Strand (1974). Daley (1991) references three papers on the topic: Seaman (1977), Seaman (1983) and Franke (1985), which are most relevant to the horizontal analysis of relatively sparse observations. Watts and McNally (1988) and McNally (2000) show the effects of misspecifying background error for some specific examples of the vertical analysis problem for passive sounder data.

In this paper, we explore some general properties of the problem by considering the scalar (zero-dimensional) analysis problem. We also present a framework for applying these ideas to a specific one-dimensional problem: the analysis of the vertical profiles of temperature and humidity from IASI data.

3.   Theory

The general analysis/retrieval equation can be expressed in its linearised form as:

$$x^a \;=\; x^b \;+\; K \cdot (y^o - H[x^b]), \tag{1}$$

where x^a is the analysed state vector,
  x^b is the background state vector,
  y^o is the observation vector,
  $H[\ldots]$ is the (generally nonlinear) operator for calculating the equivalent of the state vector in the observation space,
and   K is the general linearised analysis/retrieval operator.

This equation gives rise to equivalent equations connecting the errors in these quantities:

$$\varepsilon^a \;=\; \varepsilon^b \;+\; K \cdot (\varepsilon^o - H \cdot \varepsilon^b), \tag{2}$$

or $$\varepsilon^a \;=\; (I - K.H) \cdot \varepsilon^b \;+\; K \cdot \varepsilon^o, \tag{3}$$

where ε^a and ε^b are the errors in x^a and x^b respectively,
  ε^o is the combined error in y^o and operator $H[\ldots]$,
  H is the Jacobian matrix of $H[\ldots]$ with respect to x,
and   I is a unit matrix.

An expression for the analysis error covariance, A, can then be derived from eq.(3):

$$A = (I–K.H).B.(I–K.H)^T + K.R.K^T, \tag{4}$$

where B is the background error covariance,
and   R is the combined error covariance of the observation and the observation operator.

Here we have had to assume only that the background and observation errors are uncorrelated with each other. Equation (4) is therefore a very general expression.

In the optimal case, in which the diagonal values of A are minimised with respect to the elements of K, we can derive usual expression for the optimal linear estimator:

$$K(B) = B.H^T.(H.B.H^T+R)^{-1}. \tag{5}$$

Here it is important to note that the optimal estimate will only be obtained if the background error covariance **assumed** by the analysis system (which we subsequently denote as B_A) is equal to the true value of the background error covariance, B, such that the assumptions of OE are valid.

By substituting eq.(5) into eq.(4), we can derive expressions for the analysis error covariance in the optimal case:

$$A_{OPT} = (I–K.H).B, \tag{6}$$

or $$A_{OPT}^{-1} = B^{-1} + H^T.R^{-1}.H . \tag{7}$$

Because K is a function of B, it can be seen from eq.(6) that A will be a nonlinear function of B.

Consider now the general, sub-optimal case in which A is given by eq.(4). In this case, although the true background error covariance is B, K has been derived using the assumed value B_A, where $B \neq B_A$:

$$K(B_A) = B_A.H^T.(H.B_A.H^T+R)^{-1}. \tag{8}$$

And the analysis error covariance, from eq.(4), is given by:

$$A(B) = (I–K(B_A).H).B.(I–K(B_A).H)^T + K(B_A).R.K(B_A)^T \tag{9}$$

or $$A(B) = A_{OPT}(B_A) + (I–K(B_A).H).(B-B_A).(I–K(B_A).H)^T \tag{10}$$

Note that, unlike eq.(6), equations (9) and (10) are linear in B.

Later in this paper, we will also consider error covariances projected on to the eigenvectors of B_A. In the optimal case, where $B = B_A$, eq.(7) becomes:

$$V^T.A_{OPT}^{-1}.V = V^T.B_A^{-1}.V + V^T.H^T.R^{-1}.H.V \tag{11}$$

$$V^T.A_{OPT}^{-1}.V = \Lambda^{-1} + V^T.H^T.R^{-1}.H.V \tag{12}$$

where V contains the eigenvectors of B_A and Λ is a diagonal matrix containing the eigenvalues of B_A.

Why project on to the eigenvectors of B_A, rather than B, or any other matrix? Firstly, because it is known; whether it is right or wrong, it is the matrix used in the retrieval/analysis system, and thus it is the effective "filter" applied to the observations, affecting how much information is extracted from the observations on each scale.

It should be noted that the eigenvectors and eigenvalues of the background error covariance have no intrinsic physical meaning; they are sensitive to purely mathematical aspects of their representation,

such as whether humidity is expressed as specific humidity or its logarithm, and the temperature eigenvectors will be sensitive to the units of humidity if the two variables have correlated errors. Nevertheless, they are one way of representing the background errors on different scales (see Figs. 4 and 5) and have the convenient property of orthogonality. Moreover, in the Met Office operational 4D-Var system (Rawlins et al., 2007), the vertical part of the B-matrix is actually represented by an operator in this eigen-space.

4.   The scalar case

Let us consider the results of applying the theory developed in section 3 to a simple scalar case, in which there is one state variable and one observation of this variable, such that the Jacobian of the observation operator, $H = 1$. Figure 1 illustrates these results.
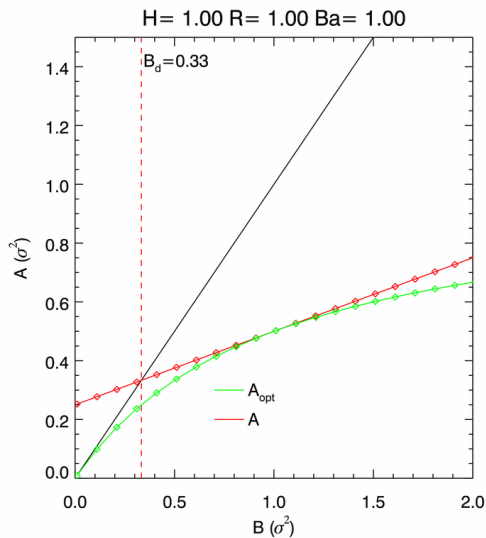

Figure 1. Analysis error variance, A, as a function of the true background error variance, B, where the assumed background error variance, $B_A = 1$, and the assumed (and true) observation error variance, $R = 1$.

The red line represents eq.(9) or (10), and it shows the analysis error for this sub-optimal analysis problem. Note that A is linear in B. The green line shows the optimal solution, i.e. the analysis error variance that results when the correct value of B is assumed for each actual value of B. The region below the green line, i.e. "better than optimal", cannot be reached. As expected, the two lines on coincide at $B = B_A = 1$. When B is greater than B_A, the analysis is sub-optimal because the process assumes that the background is more accurate than it actually is, and hence extracts too little information from the observations. When B is less than B_A, the analysis is sub-optimal because the process assumes that the background is less accurate than it actually is, and hence attempts to extract too much information from the observations. In this case, it is possible to make the analysis error variance greater than the background error variance. This occurs when the red line crosses the line A = B. We call the region beyond this point the "danger zone", i.e. the portion of the red line to the left of the line A = B, in which the analysis is degraded relative to the background. We should be aware of the presence of this zone and strive to avoid it. On the other hand, we can see that if we attempt to avoid it by decreasing B_A, we are more likely to be at the right-hand end of the red line, where the analysis fails to extract available information from the observations.

Figure 2 shows the effects of varying the ratio between background and observation error variance. At higher values of observation error, there is an increase in the value of B at which the "danger zone" begins. However, the increased slope of the red line means that the negative consequences of crossing this line are reduced. At lower values of observations error, the danger zone is more remote but the negative consequences of entering it are greater. This is illustrated further in Figure 3, which shows the value of B at which the danger zone is entered as a function of the R, all for $B_A = 1$. It can be seen that, for a wide range of values of R, i.e. R > 0.5, the danger zone lies within a factor of 0.25-

0.45 of the assumed value of B. In other words, in this range of R/B_A, if the assumed background error variance is over-estimated by more than a factor of 2.2-4.0 then the analysis is likely to be degraded, and the problem will only be negligible if $R \ll B_A$.

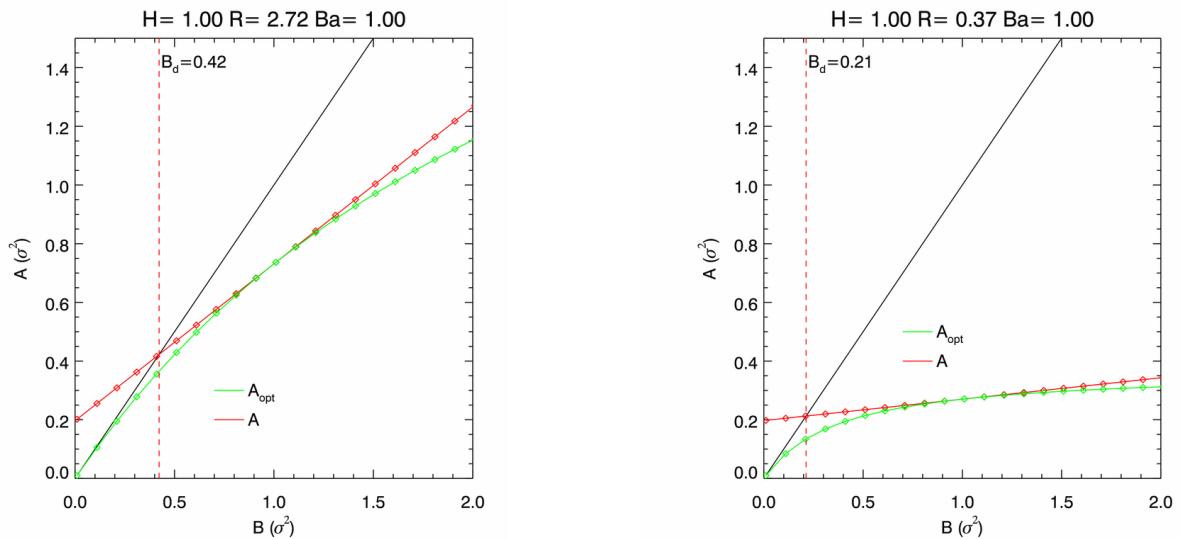This demonstrates the dangers of over-estimating the value of B in the analysis/retrieval process.

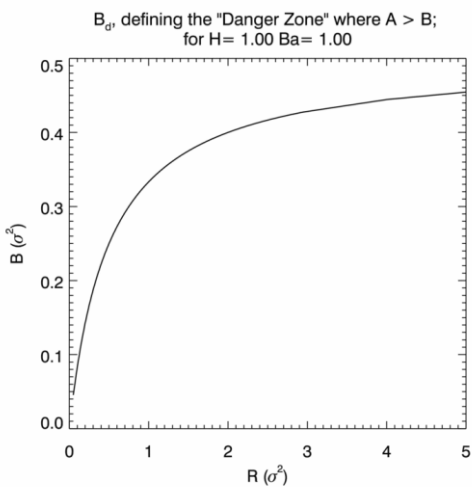Figure 2. As Figure 1, but for R = 4 (left) and R = 0.25 (right).

Figure 3. The value of B at which the danger zone begins as a function of R, for $B_A = 1$.

5.   Illustration: assimilation of IASI radiances into a global NWP system

In this section, we assess the error characteristics of a 1D-Var problem that simulates the assimilation of IASI radiances using 4D-Var. We start by examining the eigenvectors of B matrices for vertical profiles of temperature and humidity derived to be consistent with the B matrix used within the Met Office operational global NWP system as at December 2009. The B matrices were derived by taking the covariance statistics of a set of column profiles of temperature and humidity representing random perturbations consistent with the 4D-Var statistics (Andersson et al., 2000).
Figure 4 shows the 10 leading eigenvectors, ranked by order of decreasing eigenvalue, for humidity expressed as the natural logarithm of specific humidity, ln(q). They form a "well-behaved" family of eigenvectors, with increasing vertical structure for humidity in the troposphere.
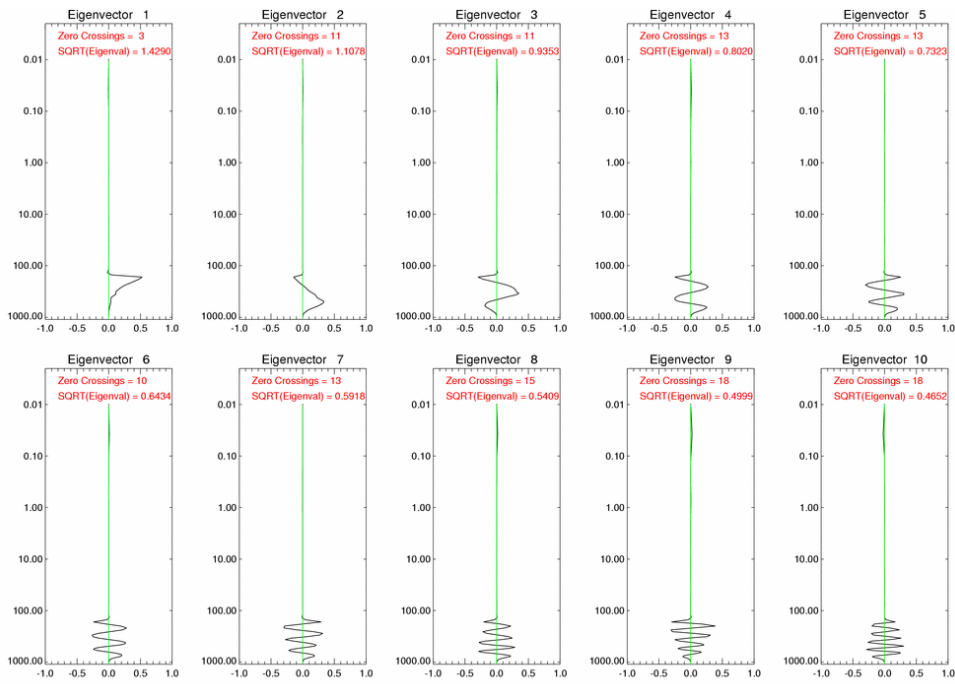
Figure 4. The leading 10 eigenvectors of B_A for the vertical background error covariance used in 4D-Var for the Met Office 70-level global NWP model, for humidity expressed as $\ln(q)$.
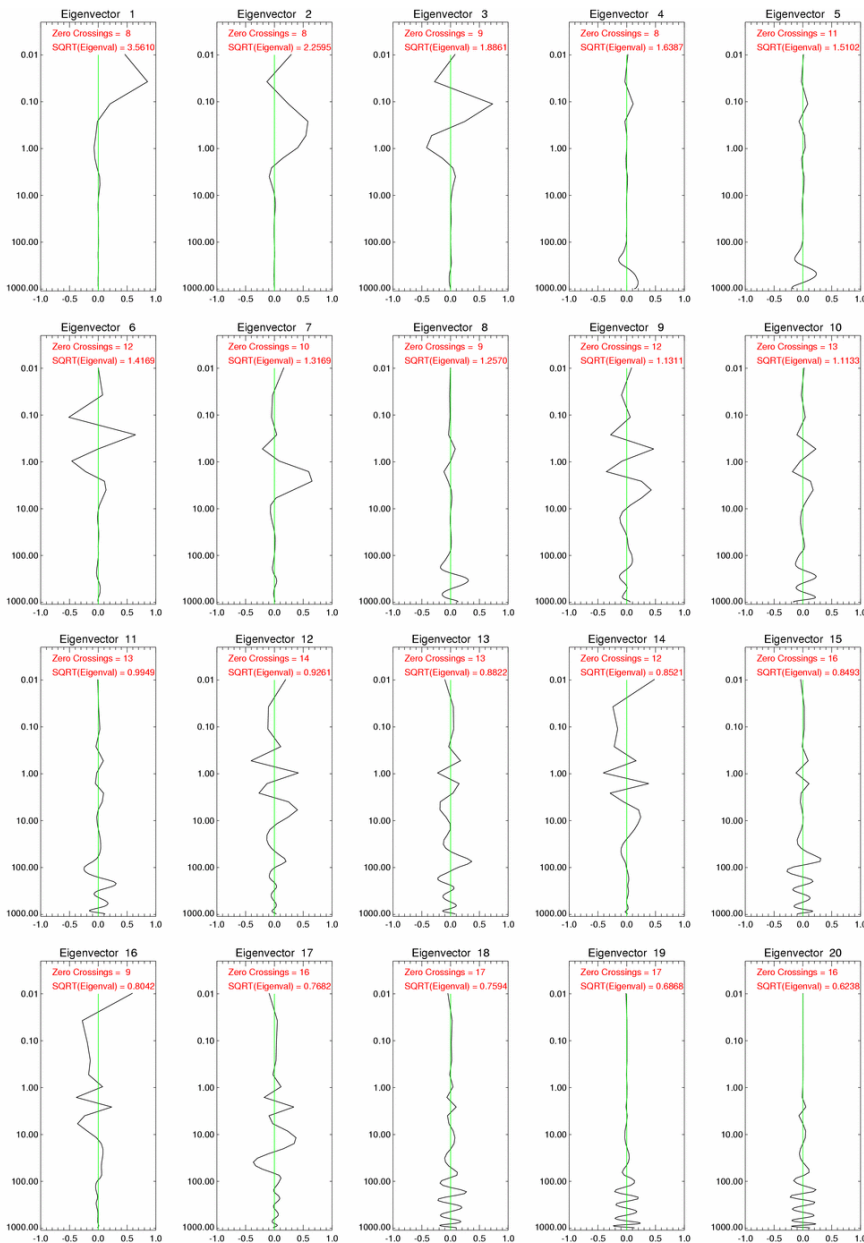
Figure 5. The leading 20 eigenvectors of B_A for the vertical background error covariance used in 4D-Var for the Met Office 70-level global NWP model, for temperature.

Figure 5 shows the 20 leading eigenvectors for temperature. By contrast, these show a more complex structure, with several of the modes which represent tropospheric structure also showing substantial contributions in the upper stratosphere and mesosphere. Whilst such structures may not be "wrong", they may have problematic consequences through a tendency to project observed tropospheric information into analysis increments at much higher levels, and vice versa. At least some of the features appearing in the stratosphere and mesosphere are thought to result from a forecast model problem which is under investigation.

Using these B-matrices, the IASI 1D-Var problem is simulated as follows: As in the Met Office operational 1D-Var for IASI (Hilton et al., 2009), 183 channels are selected, of which 31 are in the water vapour band. Two levels of observation error have been simulated: the instrument noise alone, and instrument noise plus forward model error (0.2K standard deviation) with an additional contribution for the effects of unmodelled trace gases (Collard, 2007; Hilton et al., 2009). In contrast to the 70-level representation of the NWP profile shown in Figs 4 and 5 above, the analysis is performed on the 43 pressure levels RTTOV radiative transfer model (Saunders et al, 1999; Saunders, 2002) using Jacobians for a US standard atmosphere.

Using these inputs, the analysis error covariance in the eigen-space of B is calculated using eq.(12) and the standard deviations of analysis error (i.e. square roots of the diagonal elements) are compared with the equivalent values for the background error covariances. Results are shown in Figures 6 and 7, for both cases of observation error described above.
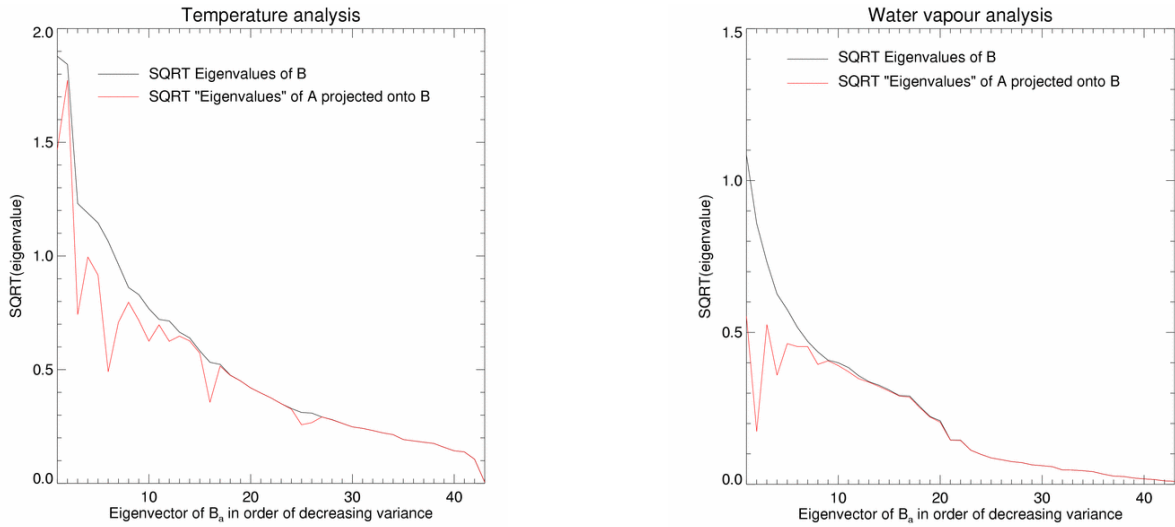

Figure 6. Standard deviations of analysis error and background error in the eigen-space of the background error covariance, for observation error variance = instrument noise variance + forward model error variance.
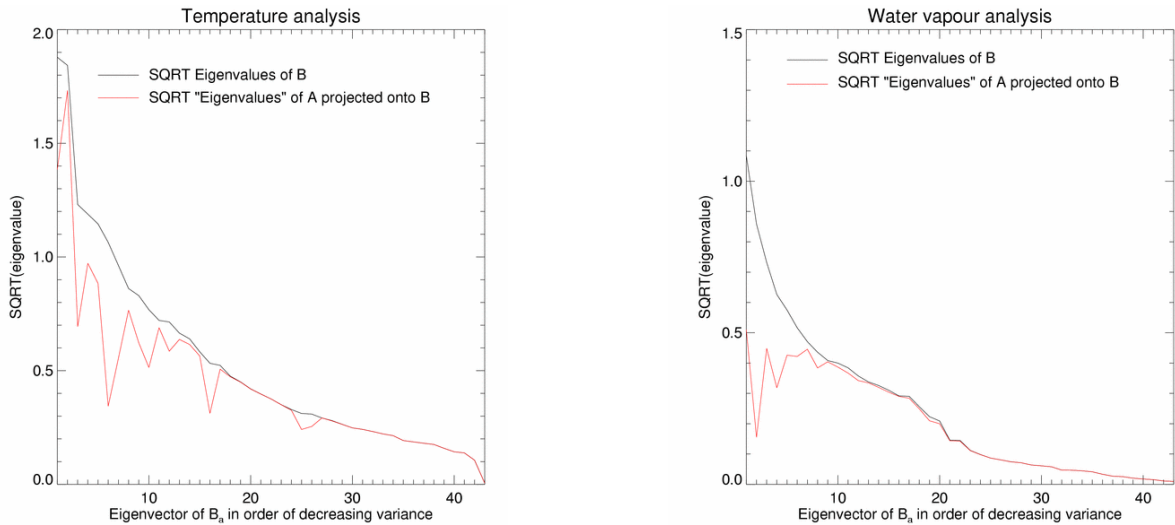

Figure 7. As Figure 6, but for observation error = instrument noise only.

Figure 6 shows that the variance in ~12 eigen-modes for temperature and ~8 eigen-modes for humidity are significantly reduced by assimilation of the IASI radiances, and Figure 7 shows that the impact is enhanced when the observation error is reduced to the level of the instrument noise only. However, these are only preliminary results requiring further investigation and analysis.

6.   Conclusions and further work

The long-term goal of this work is to make retrievals/analyses robust against the inevitable errors in the background error covariance, which may be particularly important for effective assimilation of satellite sounder data. We argue that, for NWP, it is the structure of the B matrix **assumed** by the data assimilation, B_A, that is likely to be crucial to achieve the desired results, and hence it is the mis-

specification of this matrix that merits study.

We have drawn attention to the existence of a "danger zone": if the background error variance is over-estimated by too large a factor, then the analysis errors are expected to be higher than background errors. For a large range in the ratio of observation error to assumed background error, this factor lies in the range 2.2-4.0.

We have examined the B-matrices for temperature and humidity implied by the background error statistics currently used in operational Met Office 4D-Var B-matrix. For temperature, these structures reveal some potential problems for the assimilation of radiance investigation that merit further investigation. We have also presented a preliminary analysis, in the eigen-space of this B-matrix, for the information extracted from IASI data by an assimilation system which uses this B-matrix. This illustrates how the B-matrix effectively acts as a filter for different scales of information in the observations.

We plan to extend this investigation and perform a more complete error analysis for IASI. In this way, we aim to improve our understanding of the behaviour of the assumed B-matrix within the assimilation process on each vertical scale and, if necessary, to improve its representation. Thus we aim to make our assumptions about background error statistics more robust against inevitable errors in these assumptions.

References

Andersson E., M. Fisher, R. Munro and A. McNally, 2000. Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Q. J. R. Meteorol. Soc.*, 126, 1455-1472.

Collard, A. 1998. Notes on IASI performance. *Met Office Forecasting Research Technical Report No. 256.* Available from the Met Office.

Collard, A.D., 2007. Selection of IASI channels for use in numerical weather prediction. *Q.J.R. Meteorol. Soc.*, **133**, 1977-1991.

Collard A. D. and A. P. McNally, 2009. The assimilation of Infrared Atmospheric Sounding Interferometer radiances at ECMWF. *Q. J. R. Meteorol. Soc.*, 135, 1044-1058.

Daley, R., 1991. *Atmospheric Data Assimilation*. Cambridge University Press, Cambridge, UK.

Franke, R., 1985. Sources of error in objective analysis. *Mon. Wea. Rev.*, 113, 260-270.

Hilton, F., N. Atkinson, N., S. English and J. Eyre, 2009. Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments. *Q. J. R. Meteorol. Soc.*, 135, 495-505.

Lorenc, A. C., 1988. Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc.,* **114**, 205-240.

McNally, A. P., 2000. Estimates of short-range forecast temperature error correlations and the implications for radiance data assimilation. *Q. J. R. Meteorol. Soc.,* **126**, 371-373.

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.

Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc and T. J. Payne, 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.,* **133**, 347-362.

Rodgers, C. D., 1976. Retrieval of atmospheric temperature and composition from remote

measurements of thermal radiation. *Rev. Geophys. Space Phys.,* **14**, 609-624.

Saunders, R. W., 2002, RTTOV-7 Users Guide, EUMETSAT NWP-SAF Report,
http://research.metoffice.gov.uk/research/interproj/nwpsaf/rtm/rttov7_ug.pdf

Saunders R. W., M. Matricardi and P. Brunel, 1999. An improved fast radiative transfer model for assimilation of satellite radiance observations. *Q. J. R. Meteorol. Soc.*, **125**, 1407-1425.

Seaman, R. S., 1977. Absolute and differential accuracy of analyses achievable with specified observational network characteristics. *Mon. Wea. Rev.,* 105, 1211-1222.

Seaman, R. S., 1983. Objective analysis accuracies of statistical interpolation and succesive correction schemes. *Aus. Meteorol. Mag.,* 31, 225-240.

Siméoni, D., C. Singer and G. Chalon, G. 1997. Infrared atmospheric sounding interferometer. *Acta Astronautica*, **40**, 113-118.

Strand, O. N., 1974. Coefficient errors caused by using the wrong covariance matrix in the general linear model. *The Annals of Statistics*, 2, 935-949.

Watts, P. D. and A. P. McNally, 1988. The sensitivity of a minimum variance retrieval scheme to the values of its principal components. Tech. Proc. 4th Int. TOVS Study Conf., ITSC-IV; Igls, Austria; 16-22 March 1988; Ed.: W. P. Menzel; Report of CIMSS, University of Wisconsin-Madison; p.399-412.