BACKGROUND ERRORS FOR OBSERVED QUANTITIES AND THEIR PROPAGATION IN TIME

Erik Andersson and Mike Fisher

European Centre for Medium-Range Weather Forecasts

Shinfield Park, RG2 9AX Reading, U.K.

Summary

By applying the observation operators of a variational data assimilation scheme to a set of random vectors, drawn from a population whose p.d.f. is given by the background error covariance matrix, we obtain grid-point fields of background error standard deviations for any observed quantity. These are valuable for diagnosing the data assimilation system's response to data and for tuning the specified observation and background errors in general. The calculated error standard deviations can be compared with those obtained from study of innovation statistics (i.e. observed departures from the background).

The technique has been applied to a range of observed quantites including near-surface wind and temperature, total column water vapour and satellite measured radiances. The latter include TOVS and ATOVS channels of the HIRS, MSU, SSU, MSU-A and MSU-B instruments. We found that operational background errors for humidity are set unrealisticly high in many dry subtropical areas. A case of poor convergence in 4D-Var was found to be due to unrealsiticly high background errors in terms of HIRS channel 12 radiances.

A temporal extension of the technique allows the background errors to be diagnosed at any future time within the 4D-Var assimilation window. This is achieved by applying the tangent-linear of a low-resolution adiabatic version of the forecast model to each random vector. The error reduction due to the use of data in the analysis is also estimated and propagated in time. We thereby obtain an estimate of the analysis error propagated six hours in time, which can be used to define the background error in the next data assimilation cycle. By cycling the scheme for a few days realistic flow-dependent background errors were produced.

1. INTRODUCTION

Modern data assimilation schemes combine the information from a wide variety of observations with prior information in the form of a background atmospheric state (Daley 1991). The background is in operational practice often a short-range (six-hour) forecast from the previous analysis. The resulting analysis is optimal only if accurate error estimates have been assigned to the observational data and to the background (Lorenc 1986). The relative weight given to the observations and the background (and therefore the amplitude of the analysis increments) is fundamentally determined by the specified observation and background error estimates.

When tuning a data assimilation system, one of the most important aspects is to improve the realism of the specified error statistics. The study of 'innovations' (i.e. observed minus background departures) is the most frequently adopted technique (Hollingsworth and Lönnberg 1986; Lönnberg and Hollingsworth 1986; Järvinen 1998 in this volume). It provides estimates, not of observation and background error separately, but of the sum of the two components. The obtained statistics are in terms of the observed quantity.

In earlier schemes, before variational methods were introduced, the observed and the background information were often presented to the analysis in terms of the same physical quantities as the analysis variables, e.g. temperature, wind, humidty and surface pressure, or a linear combination of these. The corresponding error estimates could then easily be compared with those obtained from the innovation statistics. Variational schemes, however, provide much greater flexibility in terms of observation usage (Courtier et al. 1998; Rabier et al. 1998; Andersson et al. 1998). Any observed quantity for which a meaningful model equivalent can be computed can in principle be used directly in the analysis without first being converted (or 'retrieved') into analysed quantities (Lorenc 1986). Satellite-measured radiances (Andersson et al. 1994) and near-surface observations of wind and temperature (Cardinali et al. 1994) are examples of this. For such data it has for this reasons been difficult to carry out tuning based on innovation statistics.

In this paper we present a technique by which the specified background error covariance matrix can be transformed to observation space. The method gives global grid-point values of background error standard deviations in terms of observed quantities. The calculation takes into account the balance constraints built in to the variational background term (Derber and Bouttier 1999) and is therefore consistent with the actual background error covariance matrix. We have computed the background error equivalents for the TOVS-channels of the HIRS, MSU and SSU instruments and for the ATOVS-channels of the AMSU-A and AMSU-B instruments. We have also applied the technique to two-metre temperature, ten-metre wind, total column water vapour, geopotential *et-cetera.*

The ATOVS-channels are not currently used operationally at ECMWF. It is hoped that the current work will help diagnosing the data assimilation response to new data types such as ATOVS and facilitate the development of effective quality control of the new data. As the ECMWF forecast model domain is about to be extended higher into the stratosphere (Untch et al. 1998) the innovation technique will be applied to some of the highest-peaking TOVS/ATOVS channels in an effort to tune the background errors of the extended model in the mid/upper stratosphere. The availability of background errors for observed quantities has also helped diagnosing 4D-Var convergence problems in one occasion.

The method to transform background errors to observation space is described in Chapter 2 and the propagation in time is outlined in Chapter 3. A selection of results is presented in Chapter 5 followed by conlusions in Chapter 6.

2. ANALYSIS AND BACKGROUND ERROR ESTIMATION

The current operational method for analysis and background error estimation was proposed by Fisher and Courtier (1995) with further details described by Fisher (1996). In this section we reproduce the outline of the algorithm.

2.1 Analysis error

The minimisation is performed with respect to a control varable χ which is related to the model variables $\mathbf{x}$ via a change of variable

$$\chi = \mathbf{L}^{-1}(\mathbf{x}-\mathbf{x}_b) \tag{1}$$

where $\mathbf{L}$ is the change of variable operator such that $\mathbf{L}\mathbf{L}^T = \mathbf{B}$. The vector $\mathbf{x}_b$ is the background and $\mathbf{B}$ is the background error covariance matrix. With these definitions the background term is simply $J_b = 0.5\chi^T\chi$ (Courtier et al. 1998; Derber and Bouttier 1999).

As suggested by Fisher and Courtier (1995) the analysis error $\mathbf{A}$ is estimated using the combined Lanczos/conjugate gradient algorithm which finds approximately the leading eigenvectors v_k of the 4D-Var Hessian and the associated eigenvalues λ_k. The leading eigenvectors describe the directions in control-vector space in which the information from observations is most important. By applying the change of variable operator to each eigenvector, an estimate of the analysis error variances (the diagonal of $\mathbf{A}$) in model space is obtained:

$$\mathrm{diag}(\mathbf{A}) \approx \mathrm{diag}(\mathbf{B}) + \sum_{k=1}^{M} (\lambda_k - 1)(\mathbf{L}\mathrm{v}_k)(\mathbf{L}\mathrm{v}_k)^{\mathrm{T}} \tag{2}$$

where M is the number of computed eigenvectors.

2.2 Background error estimation

The *ad-hoc* method. The diagonal of $\mathbf{B}$ is in current 3D/4D-Var obtained by applying a simple error-growth model to the $\mathbf{B}$ of the previous cycle. Vorticity standard deviations at each level are then renormalized to the global amplitude given by the NMC statistics. Temperature, wind, surface pressure and geopotential errors, for later use in the background quality control checks, are obtained by *ad hoc* formulas based on the vorticity standard deviations.

The randomisation method. Alternatively, the randomisation method can be used to calculate a low-rank estimate of $\mathbf{B}$, in terms of model variables (Fisher and Courtier 1995). In particular the diagonal of $\mathbf{B}$ can be estimated as

$$\mathrm{diag}(\mathbf{B}) \approx \sum_{i=1}^{N} (\mathbf{L}\xi_i)(\mathbf{L}\xi_i)^{\mathrm{T}} \tag{3}$$

where ξ_i is a set of N random vectors in control-vector space, drawn from a population with zero mean and unit Gaussian variance. The difference between randomisation and the operational *ad hoc* method is very small for vorticity as the $\mathbf{L}$-operator for vorticity is the identity (see Derber and Bouttier 1999). There is however a significant difference for the mass variables (surface pressure and temperature), which are partly related to vorticity through a balance operator and partly un-balanced. Randomisation takes the actual J_b balance into account, and is therefore more accurate. Variances produced by randomisation are noisier if N is small.

2.3 Transformation to observation space.

The randomiztion method can be extended to compute an approximation to $\mathbf{HBH}^{\mathrm{T}}$ – the background error in terms of *observed* quantities. Let $\mathbf{H}$ be the Tangent Linear observation operators, linearized around the background state, then

$$\mathrm{diag}(\mathbf{HBH}^{\mathrm{T}}) \approx \sum_{i=1}^{N} (\mathbf{HL}\xi_i)(\mathbf{HL}\xi_i)^{\mathrm{T}} \tag{4}$$

This requires that $\mathbf{H}$ can be applied to vertical profiles of model variables at model grid points. The observation operators of 3D/4D-Var (Courtier et al 1998) include vertical interpolation between model levels, radiative transfer and hydrostatic integrations and interpolation in the surface layer. By the application of Eq. (4) maps of background error variances may be obtained in terms of each of the TOVS and ATOVS radiance channels; geopotential height, temperature, wind and humidity; total ozone and total column water; two-metre temperature, two-metre humidity, ten-metre wind and surface pressure. By using the actual $\mathbf{H}$ and $\mathbf{L}$ of the variational analysis, the only approximation in Eq. (4) is in the restricted sample size, $N < \infty$, which inevitably leads to some noise in the estimated variances.

The observation operators can also be applied to the $\mathbf{L}\mathrm{v}_k$-vectors and summed up as in Eq. (2), to form analysis errors in terms of observed quantities.

$$\mathrm{diag}(\mathbf{HAH}^{\mathrm{T}}) \approx \mathrm{diag}(\mathbf{HBH}^{\mathrm{T}}) + \sum_{k=1}^{M} (\lambda_k - 1)(\mathbf{HL}\mathrm{v}_k)(\mathbf{HL}\mathrm{v}_k)^{\mathrm{T}} \tag{5}$$

The first term on the right hand side is obtained from Eq. (4). Using a very simple error-growth model, these analysis errors can be propagated in time by six hours and then be used to define quality control rejection limits in the background check of the next cycle.

3. PROPAGATION IN TIME

The very simple error growth model of Savijärvi (1995) used so far in 3D-4D/Var represents exponential error growth of small errors and the assymptotic behaviour of large errors towards a climatological variance (Fisher 1996). It lacks the dynamical i.e. flow-dependent effects on error growth.

3.1 Prediction error

From Kalman Filter theory we have an expression for the evolution of the prediction error covariance matrix, $\mathbf{P}$:

$$\mathbf{P} = \mathbf{MAM}^{\mathrm{T}} + \mathbf{Q} \tag{6}$$

where $\mathbf{M}$ is the tangent linear of the forecast model and $\mathbf{Q}$ is the model error covariance. Inserting the approximate form for $\mathbf{A}$ from Eq. (2) into Eq. (6), we have:

$$\mathrm{diag}(\mathbf{MAM}^{\mathrm{T}}) \approx \mathrm{diag}(\mathbf{MBM}^{\mathrm{T}}) + \sum_{k=1}^{M} (\lambda_k - 1)(\mathbf{ML}\mathrm{v}_k)(\mathbf{ML}\mathrm{v}_k)^{\mathrm{T}} \tag{7}$$

$$\mathrm{diag}(\mathbf{MBM}^{\mathrm{T}}) \approx \sum_{i=1}^{N} (\mathbf{ML}\xi_i)(\mathbf{ML}\xi_i)^{\mathrm{T}} \tag{8}$$

Eq. (7) and Eq. (8) provide expressions for the evolution of analysis and background errors to any future time within the range of validity of the tangent linear approximation. In the current operational context there are around 90 v_k-vectors. By setting $N = 50$ the additional cost would be 90+50=140 six-hour integrations of the adiabatic tangent linear model $\mathbf{M}$, at low resolution (e.g. T42). It is hoped that this method could replace the current simple error-growth model and introduce the previously lacking flow-dependent effects on error growth.

Flow-dependent BgQC. If, in addition to the above, we apply the observation operators to the evolved random vectors we get a good approximation of the background error in observation space valid at a later time:

$$\mathrm{diag}(\mathbf{HMBM}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}) \approx \sum_{i=1}^{N} (\mathbf{HML}\xi_i)(\mathbf{HML}\xi_i)^{\mathrm{T}}$$

at a small extra cost.

We may also apply $\mathbf{H}$ to the evolved eigenvectors $\mathbf{ML}\mathrm{v}_k$ and sum up as in Eq. (7) to get the analysis errors for observed quantities at the future time:

$$\mathrm{diag}(\mathbf{HMAM}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}) \approx \mathrm{diag}(\mathbf{HMBM}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}) + \sum_{k=1}^{M} (\lambda_k - 1)(\mathbf{HML}\mathrm{v}_k)(\mathbf{HML}\mathrm{v}_k)^{\mathrm{T}} \tag{9}$$

These can be used to provide flow-dependent quality control limits for the background quality control checks of the next analysis cycle.

4.

5. RESULTS

The results presented in this section have been obtained by cycling 4D-Var for five days at resolution T319 outer loop (i.e. resolution of background fields), T63 inner loop (i.e. resolution of analysis increments) and L50 (the number of model levels), with the top level at 0.1 hPa (Untch et al 1998). The assimilation started 19981201-00 UTC and the results presented refer to 19981206-00. Background errors have been propagated using Eq. (7) and Eq. (8) and transformed to observation space using Eq. (9).

The background 500 hPa height field is shown contoured in Fig. 1 with the associated background error shaded. Larger background errors are shown darker. We can see that the troughs over the oceans generally are associated with higher background errors - with the largest errors in the frontal regions to the south of the cyclone centres. Data dense continental areas (Europe, America and China) generally show smaller background errors, except along the American West coast were in this case larger errors have drifted in with a trough from the Pacific.

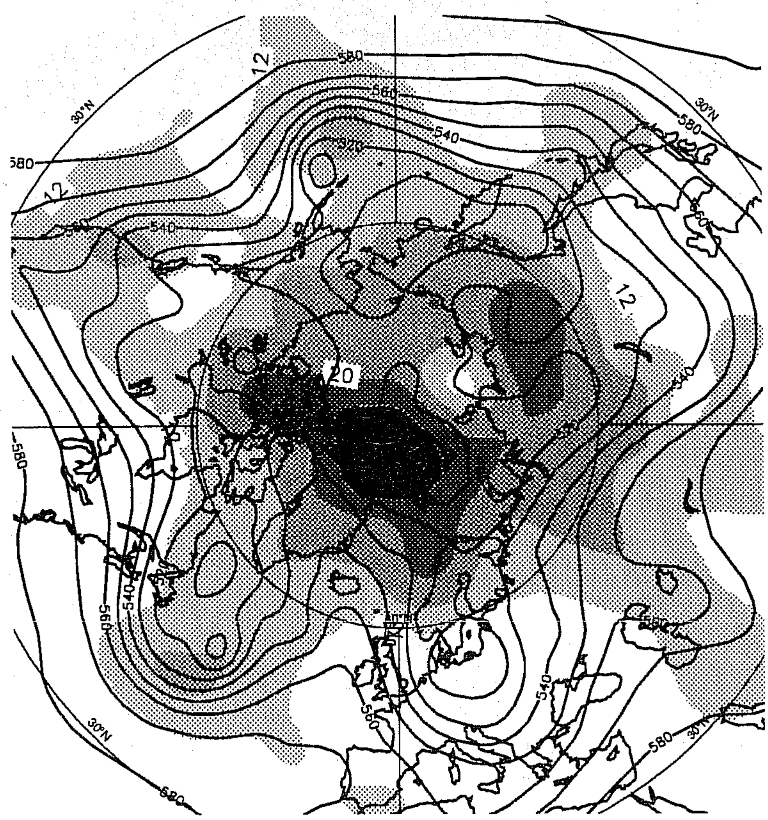

Figure 1: Background field of 500 hPa geopotential (contoured) and estimated flow-dependent background error standard deviation (shaded), 19981206-00 UTC. Shading starts at 12 m with an interval of 4 m.

The remaining figures show examples of background errors in terms of TOVS radiance channels. The observation operators (radiative transfer calculations in the case of TOVS) have been linearized around the background state. The estimated background errors are therefore a reflection of the model state itself as well as the specified background error covariance matrix. This may be important particularly for some humidity sensing channels for which the weighting function varies strongly with the amount and the distribution of humidity in the atmosphere. Fig. 2 shows MSU-2 background errors (shaded) in the Southern Hemisphere together with the background 500 hPa height field (contoured). We can see that the background errors used in this experimental version of 4D-Var translate to MSU-2 errors of less than 0.4 degrees (brightness temperature) in most oceanic areas. Errors in excess of 0.4 degrees (shaded) occur mostly in the dynamically most active troughs. Errors are also large over sea ice and over land were this channel is sensitive to errors in surface skin temperature

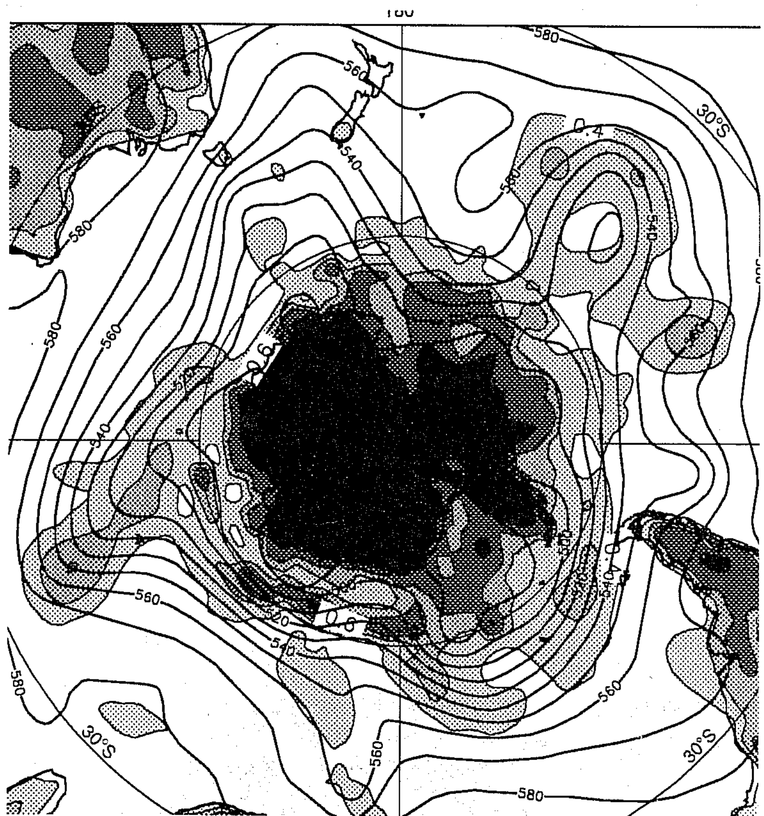

Figure 2: Background field of 500 hPa geopotential (contoured) and estimated flow-dependent background error standard deviation (shaded) for TOVS channel MSU-2, 19981206-00 UTC. Shading starts at 0.4 degrees (brightness temperature) with an interval of 0.1 degrees.

The specification of background errors is currently the most uncertain in the upper-most part of the model where there are few conventional data to verify short-range forecasts against. The NMC-method is also less reliable in data sparse parts of the atmosphere. Comparison with innovation statistics for some high peaking AMSU channels may provide some guidance in the near future. The transformation of the current **B**-matrix into equivalents of AMSU-12, which peaks around 10 hPa, yields standard deviations between 0.4 and 0.6 degrees in most parts of the mid-latitudes, Fig. 3. The specified background error for temperature, however, is between 0.8 and 2.0 degrees in these parts of the stratosphere (not shown). The much lower errors in terms of vertically broad TOVS channels are a consequence of important negative correlations of error.

The final example shows HIRS channel 12. This channel is sensitive to humidity in the mid and upper troposphere. Where there is a lot of humidity in the atmosphere the channel tends to peak higher up and where the atmosphere is dry it peaks lower down. The signal in HIRS-12 may also be saturated if the humidity is very high in the mid troposphere. These factors combine to create a very complicated field of background error for HIRS-12. An example is shown in Fig. 4. It shows the background total column water vapour (contoured) and HIRS-12 background errors (shaded). Light shading indicates low errors (less than 2 degrees) and darker shading indicates higher errors (more than 4 degrees). We can see that the lowest errors tend to occur where the humidity is high, and vice versa. It was found that in extreme cases of dry background even small additions of humidity will result in a large response in terms of HIRS-12 brightness temperature. This characteristic of HIRS-12 in combination with crudly specified background errors for specific humidity in dry regions can in some cases result in very large estimated background errors for HIRS-12 - in one case exceeding 30 degrees (not shown). With observations errors in the order of 2 degrees poor conditioning of the 4D-Var minimisation was observed. Statistics of HIRS-12 innovations indicate that such very large backkground er-

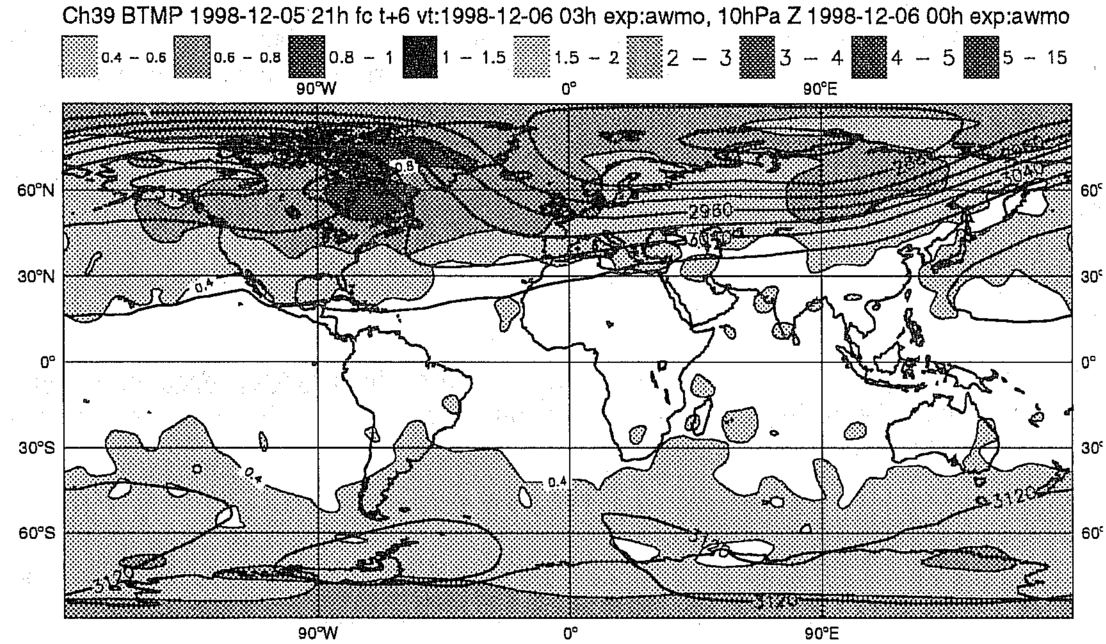

Figure 3: Background field of 10 hPa geopotential (contoured) and estimated flow-dependent background error standard deviation (shaded) for TOVS channel AMSU-12, 19981206-00 UTC. Shading starts at 0.4 degrees (brightness temperature) with an interval of 0.2 degrees.

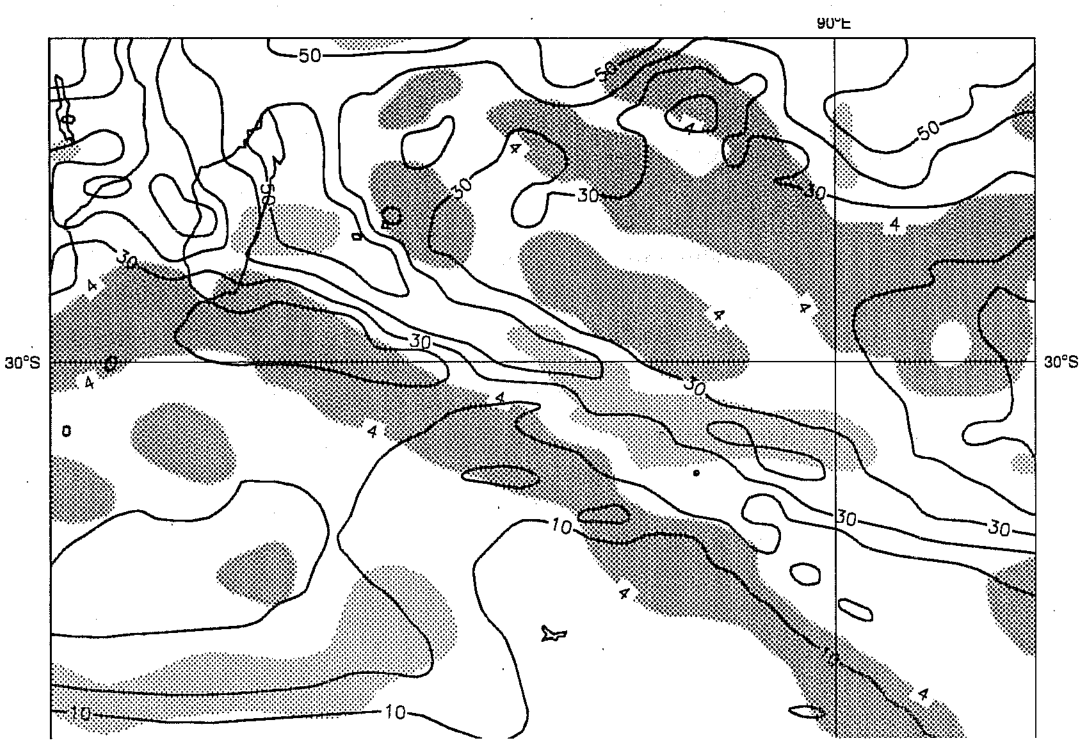

Figure 4: Background field of total column water vapour (contoured) and estimated flow-dependent background error standard deviation (shaded) for TOVS channel HIRS-12, 19981206-00 UTC, South East of Madagascar. Light shading below 2.0 degrees (brightness temperature) and darker shading above 4.0 degrees.

rors are spurious and should be corrected by improved specification of humidity background errors.

6. CONCLUSIONS

A method to dignose background and errors for observed quantities has been developed. The observation operators (linearized around the background model state) are applied to a set of random vectors drawn from a population with the p.d.f. of the background error covariance matrix. The method can be used to tune the setting of observation and/or background errors, by comparison with innovation statistics (observation minus background). This may be especially useful as: 1) The model domain is about to be extended higher into the stratosphere; 2) New observations are introduced (AMSU-A and B) and 3) New analysis variables are introduced, e.g. ozone. Result of such studies are not available at the current time.

The technique has been used to diagnose a real case of 4D-Var convergence problem, with respect to humidity sensitive radiance observations.

Randomisation also provides a practical and cheap way of getting flow-dependent background error standard deviations, for cycling of 4D-Var and for improvement of quality control rejection criteria. Forecast and quality control impact is yet to be evaluated.

References.

Andersson, E., Pailleux, J., Thépaut, J-N., Eyre, J., McNally, A. P., Kelly, G. and Courtier, P., 1994: Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **120**, 627-653.

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljevic, D., Brankovic, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J-N. and Viterbo, P., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part III: Experimental results. *Q. J. R. Meteorol. Soc.*, **124**, 1831-1860.

Andersson, E. and Järvinen, H., 1999: Variational quality control. To appear in *Q. J. R. Meteorol. Soc.* Also available as ECMWF Tech Memo 250.

Cardinali, C., Andersson, E., Viterbo, P., Thépaut, J-N., and Vasiljevic, D., 1994: Use of conventional surface observations in three-dimensional variational data assimilation. ECMWF Tech. Memo. 207. Available from ECMWF.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. *Q. J. R. Meteorol. Soc.*, **124**, 1783-1808.

Derber, J. and Bouttier, F., 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. Submitted to *Tellus*.

Daley, R., 1992: The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 164-177.

Fisher, M. and Courtier, P., 1995: Estimating the covariance matrix of analysis and forecast error in variational data assimilation. ECMWF Tech. Memo. 220.

Fisher, M., 1996: The specification of background error variances in the ECMWF variational analysis system. Proc. ECMWF Workshop on "Non-linear aspects of data assimilation", Reading, 9-11 September 1996, 645-652.

Hollingsworth, A. and Lönnberg, P. 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136.

Järvinen, H. and Undén, P., 1997: Observation screening and first guess quality control in the ECMWF 3D-Var data assimilation system. ECMWF Tech. Memo. 236. Available from ECMWF.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177-1194.

Lönnberg, P. and Hollingsworth, A., 1986: The Statistical Structure of Short-Range Forecast Errors as determined from Radiosonde Data. Part II: The Covariance of Height and Wind Errors. *Tellus*, **38A**, 137–161.

Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A., and Bouttier, F., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part II: Structure functions. *Q. J. R. Meteorol. Soc.*, **124**, 1809-1829.

Savijärvi, H., 1995: Error growth in a large numerical forecast system. *Mon. Wea. Rev.*, Vol **123,** 212–221.

Untch, A., Simmons, A., Hortal, M., Jakob, C. and colleagues, 1998: Increased stratospheric resolution in the ECMWF forecasting system. Proc. Workshop on “Chemical data assimilation”, deBilt 9-10 December 1998.