A STUDY OF THE INNOVATION AND RESIDUAL SEQUENCES IN VARIATIONAL DATA ASSIMILATION

Heikki Järvinen

European Centre for Medium-Range Weather Forecasts

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

Abstract

Innovation and residual sequences from ECMWF 3D- and 4D-Var systems have been studied. First, the Hollingsworth-Lönnberg method is applied to the innovations in order to partition the perceived forecast error variance into contributions from observation and background errors. The estimated forecast error variance of the background trajectory over the 4D-Var assimilation window is characterized by an upwards intensifying growth in the troposphere, as well as with a broadening of its horizontal covariance length scale. The estimation of these statistics for 4D-Var therefore requires innovations which are homogeneously spread in space and time. Second, the residuals are used for checking the fit of the analysis to observations over the assimilation window. This fit should, in theory, reveal the effect of model error in a strong constraint variational problem. In this study a convex curve with a u-shape is found for this fit implying that the perfect model assumption of 4D-Var may be violated with as short an assimilation window as six hours.

1. INTRODUCTION

There is a large number of statistical parameters in an assimilation system which need to be estimated and specified for the system to perform optimally. Among the most important parameters are the observation and background error variances, as the ratio of these statistics determine the amplitude of the analysis increments. In principle, it is not strictly necessary to fix the absolute values of these statistics, as long as their relative magnitude is realistic. With this point of view, the observation error variances describe the relative information content of the observations with respect to the background information in a given assimilation system. The observation error variance can be statistically estimated from the innovation sequence. Application of a best estimate for the observation error variance in one component of the observing network implies that the error statistics of all the other sources of information must be adjusted accordingly and must reflect their true information content. Furthermore, a relevant measure of the magnitude of a model error in a strong constraint formulation of 4D-Var assimilation system can only be obtained if the observation and background error variances are correctly specified. Therefore, even if it is possible in principle to have an assimilation system with the variances specified only in relative sense, a more thorough evaluation of the system requires the absolute values for the statistical parameters.

2. BACKGROUND AND OBSERVATION ERROR VARIANCES

Statistical estimation of observation and background error variances from innovation sequences is well established (*Buell* 1972, *Rutherford* 1972, *Hollingsworth and Lönnberg* 1986, *Lönnberg and Holl-*

ingsworth 1986; for more references, see for *Daley* 1991 and *Bouttier and Courtier* 1998). The emphasis in this paper is laid on the behaviour of these statistics over the assimilation window and on the aspects related to 4D-Var in particular. The innovation and residual sequences are extracted from the ECMWF implementation of 3D-Var (*Courtier et al* 1998, *Andersson et al* 1998, *Rabier et al* 1998) and 4D-Var (*Rabier et al* 1997 and 1999) data assimilation systems which apply an assimilation window of six hours.

The innovation sequence $\boldsymbol{d}_i$ at location i is defined as a difference between the observations $\boldsymbol{y}_i$ and the background $\boldsymbol{x}_b$ as processed by the appropriate observation operator $\boldsymbol{H}_i$

$$\boldsymbol{d}_i = (\boldsymbol{y}_i - \boldsymbol{H}_i \boldsymbol{x}_b)$$

It is assumed here that the observation operator also includes the model integration in time. Replacing the background $\boldsymbol{x}_b$ in this formula with the analysis $\boldsymbol{x}_a$ would define the residual sequence. The covariance between the innovation sequences at locations i and j is given as an average over a sample by

$$cov(i, j) = \overline{\boldsymbol{d}_i \boldsymbol{d}_j^T}$$

In an ideal case, a histogram of covariances behaves as displayed in Fig. 1a. The dots represent the binned covariances as a function of the separation between i and j, in this case for AIREP (aircraft report) temperature over North America at 200hPa. The covariance is zero for large horizontal separations, over 500km, say. With decreasing separation, the covariance builds up and, in case of uncorrelated observation errors, this is entirely due to the horizontal correlation of forecast error. Fitting a curve through the points of the histogram and extrapolating to zero distance, the perceived forecast error variance implied by the sample of innovations can be partitioned into contributions from background error and observation error. In some cases, however, a significant covariance is present even for large station separations. This can be interpreted as a mean difference between the short range forecast and the observations. This is an indication of a bias either in the forecast model, or in the observing system. This estimation method does not provide a clean way to separate this bias from the random part of the covariance. The bias must be deduced using complementary information.

In 4D-Var, the background is specified at the beginning of the assimilation window and there is a forecast error associated with the background trajectory which extends over the assimilation window. The background error covariance of Fig. 1a corresponds to the innovations of the first half an hour of the six hour assimilation window of 4D-Var. At the middle of the six hour assimilation window (Fig. 1b) the forecast error variance of the background trajectory is increased compared to the variance at the initial time. The horizontal length scale of the forecast error covariance at the middle of the assimilation window is also larger than the length scale of the background error covariance at initial time. The estimation of the background error variance must therefore correspond the value at the initial time of the assimilation window. The same is also true if one wishes to evaluate the covariance length scale implied the background penalty term of the variational assimilation system with the covariance length scale as deduced from the innovations.

a)

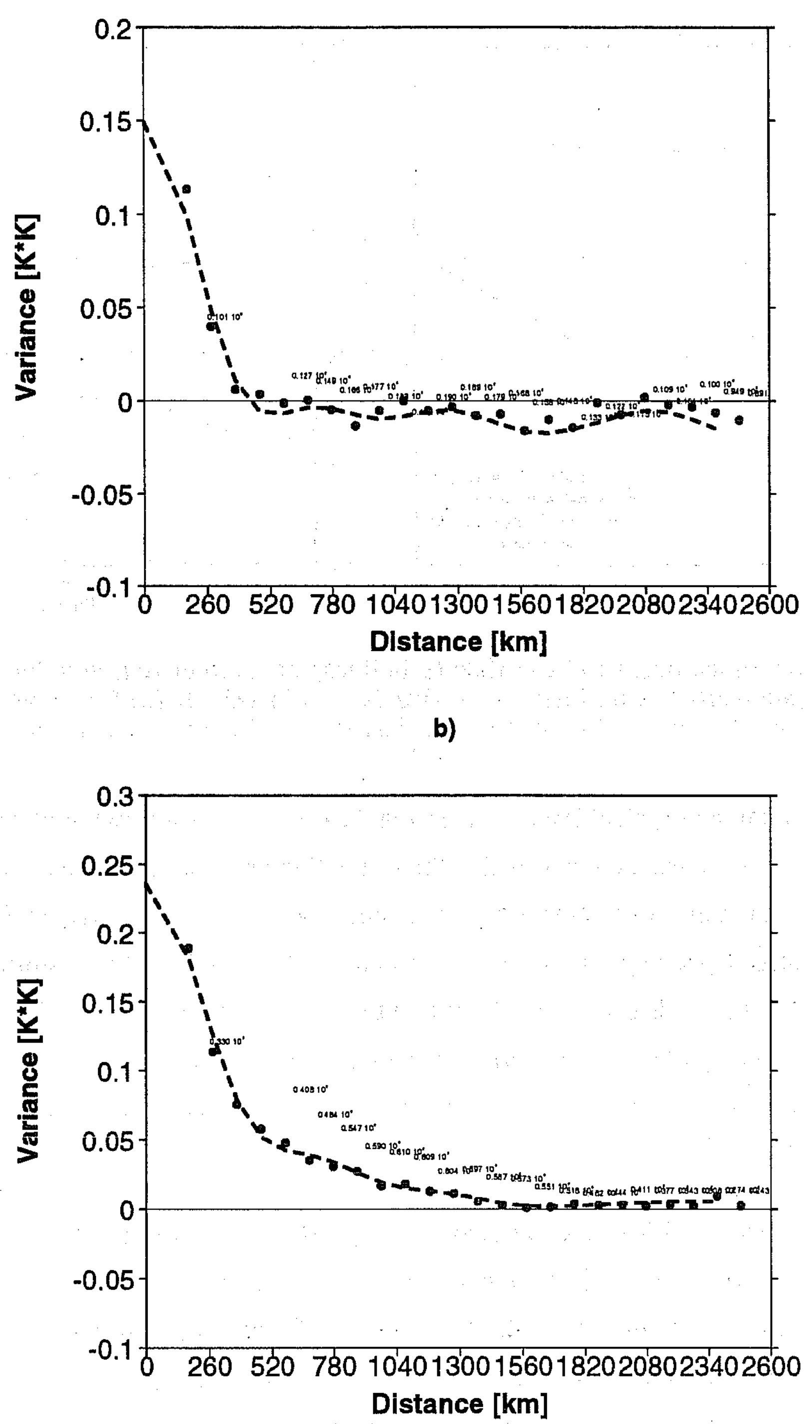

Figure 1: Histogram of innovation covariances at the beginning (a) and at the middle (b) of the six hour assimilation window for AIREP temperature over North America at 200hPa. The 4D-Var innovation sequence is from 1 Sep 1997 to 15 Oct 1997. Note the different scale of ordinate in (a) and (b).

Next, the observation and background error variances are estimated over the six hour assimilation window to reveal their temporal behaviour in 3D-Var as well as in 4D-Var. The variance estimation presented here uses AIREP component wind over North America at 200hPa. In 3D-Var the background is valid at the middle of the assimilation window, and the estimated standard deviation of background

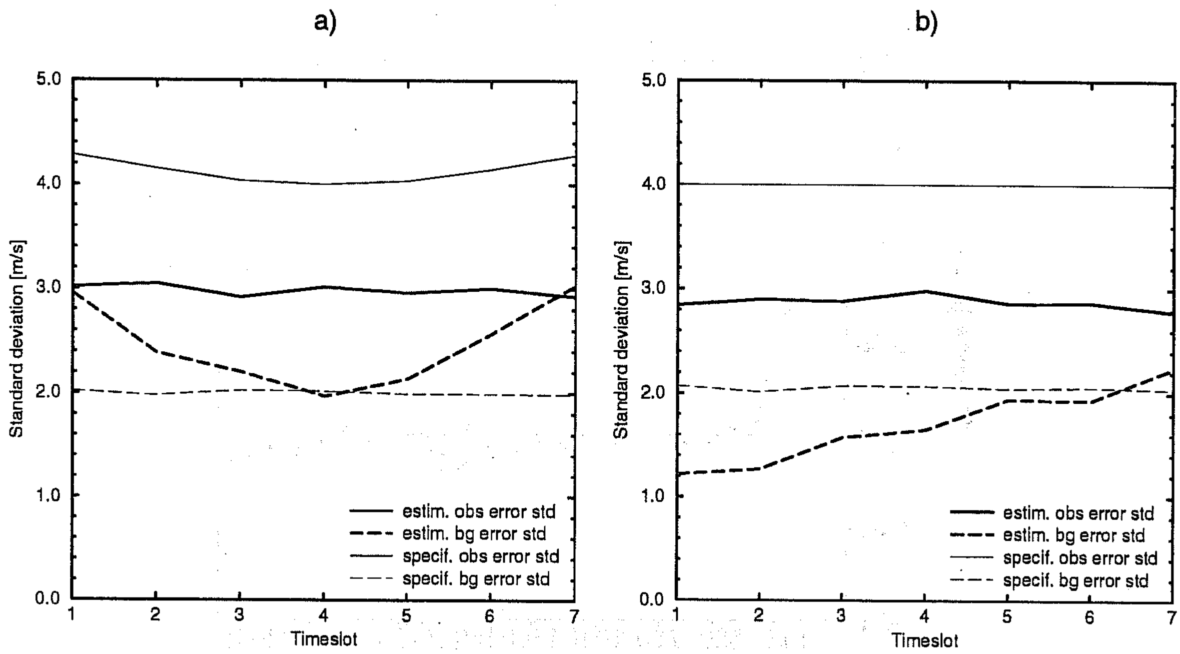

Figure 2: Estimated (thick lines) and specified (thin lines) standard deviation of background (dashed lines) and observation error (solid lines) in 3D-Var (a) and in 4D-Var (b) for AIREP component wind over North America at 200hPa. The estimation is based on an innovation sequence from 1 Sep 1997 to 15 Oct 1997.

error has its minimum there (thick dashed line in Fig. 2a). The estimated standard deviation of observation error remains at a constant level, within the estimation uncertainty (thick solid line). The specified standard deviation of observation error (thin solid line) is in this case larger than the estimated one and it is inflated depending on the time difference to the middle of the assimilation window in order to account for the persistence error. Note that the persistence error in 3D-Var appears as a horizontally correlated error which in the variance estimation contributes to the background error variance. The correction in the specified observation error standard deviation is qualitatively right but not quite large enough in this case.

In 4D-Var, in contrast to 3D-Var, the background is specified at the initial time, where the standard deviation of forecast error of the background trajectory (thick dashed line in Fig. 2b) has its minimum. There is a monotonic growth of the standard deviation of forecast error throughout the assimilation window. The background in this implementation of 3D- and 4D-Var is a six hour forecast from the previous analysis, with the remark that the forecast producing the 4D-Var background is constrained by the observations in the previous assimilation window. Thus, comparing the level of background/ forecast error standard deviation in Fig. 2a and Fig. 2b, one can note, first, that the initial background error standard deviation is lower in the 4D-Var system and, second, that the 4D-Var forecast error standard deviation associated with the background trajectory reaches the level of 3D-Var background error standard deviation in about four to five hours into the assimilation window. In other words, the 4D-Var assimilation is ahead of 3D-Var assimilation by four to five hours at the very earliest forecast range.

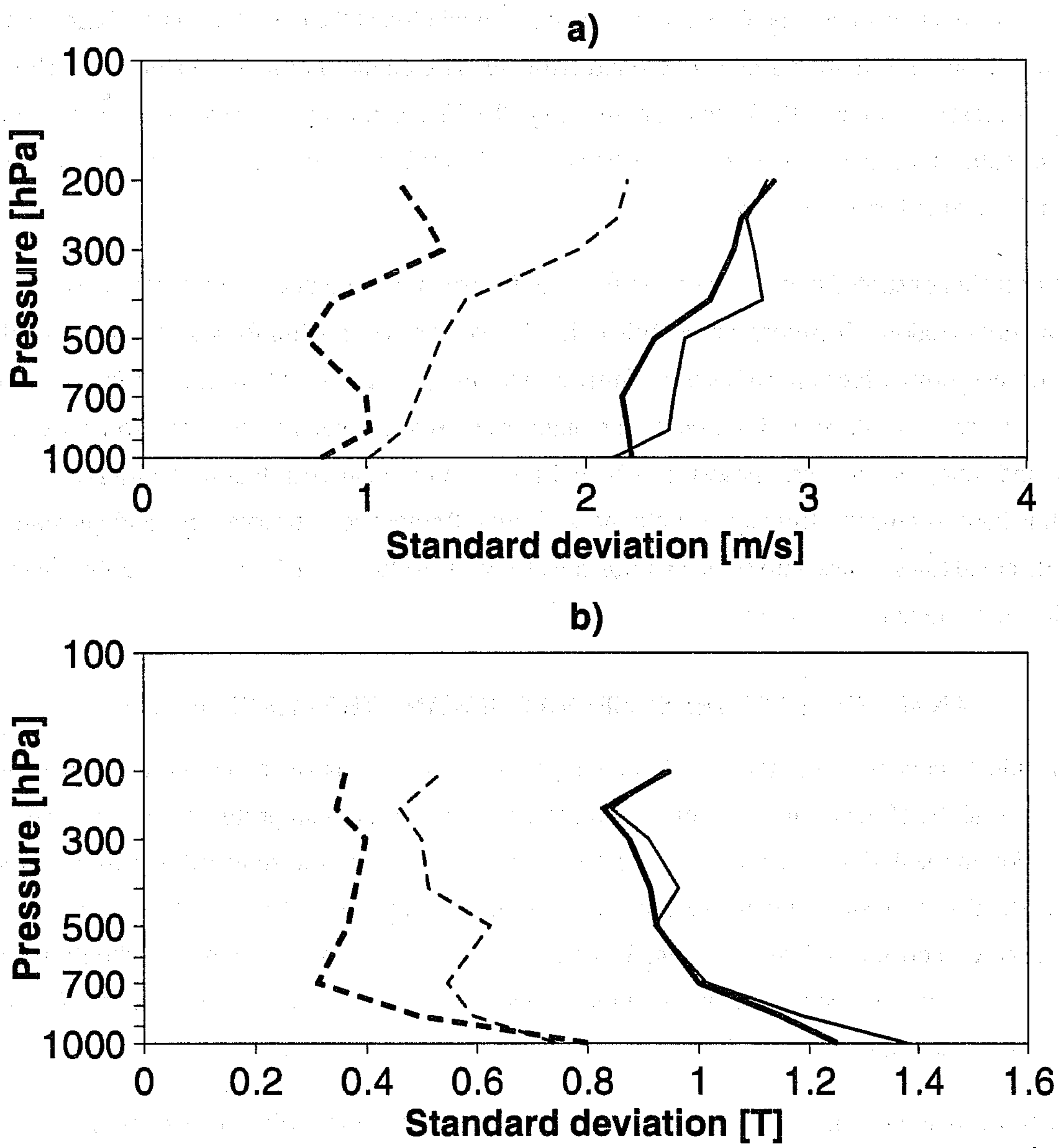

Figure 3: Vertical profiles of the background (dashed line) and observation (solid line) error standard deviation at the beginning (thick line) and at the end (thin line) of the six hour assimilation window for aircraft component wind (a) and temperature (b) over North America. The 4D-Var innovation sequence is from 1 Sep 1997 to 15 Oct 1997.

Reassuringly for the variance estimation procedure, the estimated standard deviation of observation error remains at a constant level throughout the assimilation window, and at the same level in both 3D- and 4D-Var systems (thick solid lines in Fig. 2). This is logical as there is no real reason for the quality of observations to be function of time.

The growth rate of the forecast error standard deviation associated with the background trajectory over the 4D-Var assimilation window increases with height in the troposphere. Over well observed areas, like North America, the forecast error standard deviation of the background trajectory in the lower troposphere at the end of the assimilation window (thin dashed line in Fig. 3a) is only marginally larger than the initial background error standard deviation (thick dashed line in Fig. 3a), in this case for AIREP component wind. This is also the case for AIREP temperature errors over the same area (Fig. 3b).

Near the jet level, however, the forecast error standard deviation at the end of the assimilation window is almost doubled from the initial background error standard deviation for component wind (Fig. 3a), and is increased by about 50% for temperature (Fig. 3b). The estimated observation error standard deviation (solid lines in Fig. 3) remains unchanged at all levels throughout the assimilation window, within the estimation uncertainty.

It has been demonstrated that the error variances and the error covariance length scale evolve over the assimilation window. Therefore a prerequisite for the variance estimation for 4D-Var is the availability of observations which are well spread in space and time, such as AIREP, DRIBU (drifting buoy) and SYNOP observations. Polar orbiting satellite radiances and scatterometer winds are also well spread in space and time, but they are, however, affected by the horizontal correlation of observation error, which is hard to separate from systematic model errors. Radiosonde observations, and geostationary satellite cloud track winds with some caution, are also very useful for their excellent vertical sampling despite a poor temporal coverage.

3. ANALYSIS FIT TO THE OBSERVATIONS AND THE MODEL ERROR

The residual sequences, i.e. the fit of the analysis to observations over the assimilation window, should reveal the effect of model error in a strong constraint variational problem (*Ménard and Daley* 1996). For an ideal 4D-Var with a valid perfect model assumption, the standard deviation of the residuals should be constant or concave over the assimilation window, and at a level slightly lower than the observation error standard deviation. Violation of the perfect model assumption should manifest itself in a convex u-shaped curve, with a more pronounced curvature and a higher level of misfit to observations the higher the model error.

Standard deviation of innovations and residuals over the assimilation window is displayed in Fig. 4 for AIREP temperature (panel a) and for component wind (panel b) for North Atlantic at 250hPa. The main features of the observation fit to the background (solid lines in Fig. 4) and to the analysis (dashed lines) are very similar for these two different observing systems. In both cases, the background fit degrades monotonically over the assimilation window, and the analysis fit has a minimum close to the middle of the assimilation window. Figure 5 shows similar plots for DRIBU surface pressure observations over North and South Atlantic. The fit of the observations to the analysis (dashed lines) is fairly similar in both areas with a minimum close to the middle of the assimilation window. The quality of the background (solid lines) at the initial time in the South Atlantic is only slightly worse than in the North Atlantic. The growth of forecast error standard deviation of surface pressure is however much more rapid on the South Atlantic.

The specified observation error variance in this assimilation system at the levels corresponding to Figs. 4 and 5 for AIREP temperature is 1.3K, for AIREP wind 4.0m/s and for DRIBU surface pressure 140hPa, respectively. Comparing these values to the actual analysis fit in Figs. 4 and 5, one can note that the values are over-estimates and therefore there is no way to define the magnitude of the associ-

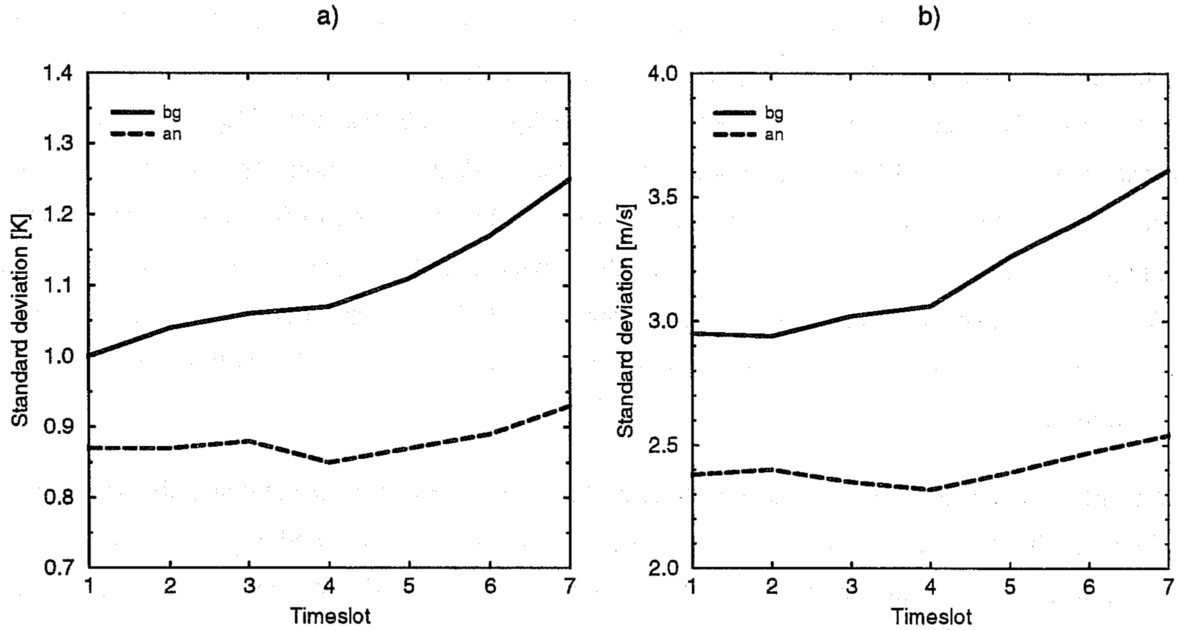

Figure 4: Standard deviation of innovations (solid lines) and residuals (dashed lines) in 4D-Var for AIREP acar temperature (a) and AIREP amdar component wind (b) at 250hPa over North Atlantic. The curves are based on innovation and residual sequences from 16 to 29 April1998.

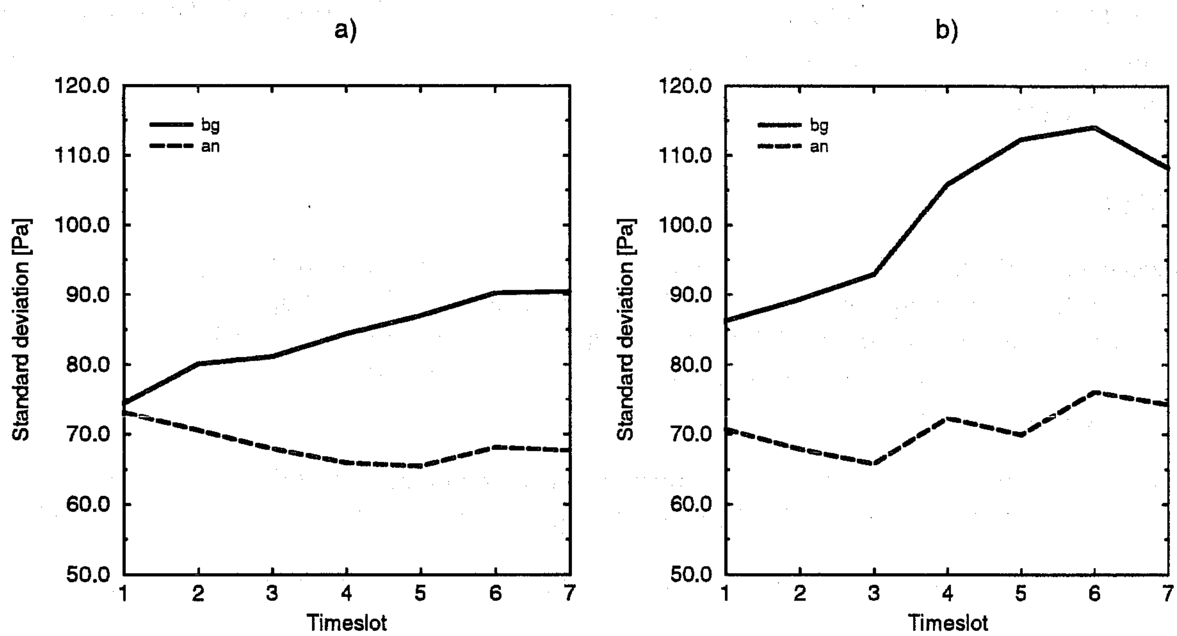

Figure 5: Standard deviation of innovations (solid lines) and residuals (dashed lines) in 4D-Var for DRIBU surface pressure in North (a) and South (b) Atlantic. The curves are based on an innovation and residual sequences from 16 to 29 April 1998.

ated model error. Figure 2 also indicates that the observation error variances are not accurately specified.

4. CONCLUSIONS

It is instructive to study the innovation and residual sequences over the assimilation window. For 4D-Var the innovations reveal that the forecast error variance associated with the background trajectory grows monotonically over the assimilation window from the initial value of the background error variance. There is an associated broadening of the horizontal length scale of the forecast error covariance. The growth of forecast error variance intensifies with height in the troposphere. The background error variance is lower in the 4D-Var assimilation than in the 3D-Var assimilation and the shortest range forecasts of 4D-Var system are ahead of 3D-Var forecasts by about four to five hours. The variance estimation for 4D-Var requires observations which are well spread in space and time, such as aircraft, drifting buoy and synop observations, and polar orbiting satellite measurements as well. Radiosonde observations, and geostationary satellite cloud track wind are also very useful for their excellent vertical sampling.

The residual sequences may be used for studying the model error in a strong constraint variational problem. In this study a convex curve with a u-shape is found for surface and upper air observations. This implies that the perfect model assumption may not be correct and it may affect a 4D-Var with as short an assimilation window as six hours.

ACKNOWLEDGEMENTS

I would like to thank my colleagues Erik Andersson, François Bouttier, Michael Fisher and Anthony Hollingsworth for their support and intrest on this work, as well as Olivier Talagrand for suggesting to diagnose the model error.

REFERENCES

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 P Viterbo, 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part III: Experimental results. *Q. J. R. Meteorol. Soc.*, **124**, 1831-1860.

Bouttier F and P Courtier, 1998: Data assimilation concepts and methods. ECMWF Training Course Lecture Series. Available from ECMWF, Shinfield Park, RG2 9AX Reading, Berkshire, U.K.

Buell, C E, 1972: Correlation functions for wind and geopotential on isobaric surfaces. *J. Appl. Meteorol.*, **11**, 51-59.

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

Daley, R, 1991: Atmospheric Data Analysis. Cambridge Atmospheric and Space Science Series. Cambridge University Press. ISBN 0-521-38215-7. 457 pages.

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

Lönnberg, P and A Hollingsworth, 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.

Ménard, R and R Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A**, 221-237.

Rabier, F, Mahfouf, J-F, Fisher, M, Järvinen, H, Simmons, A, Andersson, E, Bouttier, F, Courtier, P, Hamrud, M, Haseler, J, Hollingsworth, A, Isaksen, L, Klinker, E, Saarinen, S, Temperton, C, Thépaut, J-N, Undén, P and D Vasiljevic, 1997: Recent experimentation on 4D-Var and first results from a Simplified Kalman Filter. ECMWF Tech Memo 240. Available from ECMWF.

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.

Rabier, F, Järvinen, H, Klinker, E, Mahfouf, J-F and A Simmons, 1999: The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics. 28pp. Submitted to *Q. J. R. Meteorol. Soc.*

Rutherford, I O, 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29**, 809-815.