TECHNICAL MEMORANDUM

311

Temporal evolution of innovation and residual statistics in the ECMWF variational data assimilation systems

H. Järvinen

Research Department

October 2000

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

ABSTRACT

The temporal evolution of innovation and residual statistics of the ECMWF 3D- and 4D-Var data assimilation systems have been studied. First, the observational method is applied on an hourly basis to the innovation sequences in order to partition the perceived forecast error covariance into contributions from observation and background errors. The 4D-Var background turns out to be significantly more accurate than the background in the 3D-Var. The estimated forecast error variance associated with the 4D-Var background trajectory increases over the assimilation window. There is also a marked broadening of the horizontal error covariance length scale over the assimilation window. Second, the standard deviation of the residuals, i.e. the fit of observations to the analysis is studied on an hourly basis over the assimilation window. This fit should, in theory, reveal the effect of model error in a strong constraint variational problem. A weakly convex curve 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. For improving the optimality of variational data assimilation systems, a sequence of retunes are needed, until the specified and diagnosed error covariances agree.

1 Introduction

Meteorological data assimilation aims at accurately estimating the current state of the atmosphere and thereby providing an initial condition for the prediction model of the atmosphere. There is a large number of statistical parameters in a data 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 covariances, as these determine the distribution of information in space and among the model variables. The characteristic amplitude of these analysis increments is determined by the ratio of the observation and the background error variances. In an assimilation system these error variances describe the relative accuracy of various observation types with respect to the background information. The absolute values of the observation and background error variances can be statistically estimated from the innovation sequence, i.e. from the observation and background differences which is essentially the fit of observations to the *a priori* estimate of the state of the atmosphere. If one wishes to use the best estimate for the error variance in one information source, then the error statistics of all the other sources of information must be adjusted accordingly to reflect their true information content.

Inclusion of the time dimension into the assimilation process, such as 4D-Var, raises the question of the temporal evolution of these statistics. In the implementation of 6-hour 4D-Var at ECMWF, the observations are organized into one hour timeslots, with two half-hour time slots at the beginning and at the end of the assimilation window. The one hour timeslot at

the middle of the assimilation window is thus centered around the nominal analysis time (00, 06, 12 and 18 UTC), coinciding with the main synoptic observing times. In 12-hour 4D-Var at ECMWF, timeslot 10 is centered around the nominal analysis time (00 and 12 UTC), coinciding with the main radiosonde observation coverage. For some observation types, such as aircraft reports, there is enough of statistical material for the estimation of the error covariances separately for each timeslot which is a novel aspect of our results. Comparison of these statistics with those estimated for a static assimilation scheme, such as 3D-Var, reveal important differences in the nature of these two assimilation systems. These differences lend credence to the results of the estimation procedure.

A prerequisite for the hourly error covariance estimation is the availability of observations which are well spread in space and time, such as AIREP (aircraft report), DRIBU (drifting buoy) and SYNOP (synoptic surface station) observations. Polar orbiting satellite radiances and scatterometer winds are also well spread in space and time, but they are affected by the horizontal correlation of observation error, which is hard to separate from the systematic model errors. Radiosonde observations are also very useful for their excellent vertical sampling despite a poor temporal coverage. It should be remembered that using observations for the estimation of error statistics biases the results in favour of the well observed areas.

The hourly calculation of the basic statistics can also be applied to the residual sequences, i.e. to the observation and analysis differences. There are experimental results of idealized systems about what kind of statistical signatures the model errors should bring along to the residual sequences. A measure of the magnitude of model error in a strong constraint formulation of 4D-Var assimilation system can only be obtained if the observation and background error covariances are correctly specified. The first section of this article (Chapter 2) is devoted to the error covariance estimation from the innovations, including a methodological discussion, and the latter one (Chapter 3) is devoted to the residuals and to the question of the model error.

2 Background and observation error covariances

Statistical estimation of observation and background error covariances from innovation sequences is well established (Buell, 1972; Rutherford, 1972; Hollingsworth and Lönnberg, 1986; Lönnberg and Hollingsworth, 1986; for more references, see for instance Daley, 1991, and Bouttier and Courtier, 1998). The emphasis in this paper is laid on the temporal 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; Rabier et al., 1998; Andersson et al., 1998) and 4D-Var (Rabier et al.,

1997, 2000) data assimilation systems. Innovations and the estimation of error covariances apply to the operational 3D-Var system during 1 September 1997 through 15 October 1997, and to the pre-operational 4D-Var system (which was later operationally implemented on the 25th of November 1997 with a 6-hour assimilation window) over the same period. Residuals and the aspects related to the model error apply to an experimental 4D-Var system over a period of 16 to 29 April 1998, as well as to a separate period from an experimental 4D-Var implementation with a 12-hour assimilation window over a period of 16 September through 10 October 1998.

3D-Var data assimilation consists of minimization of a cost function measuring the distance to the background and to the observations. The background is a six hour forecast from the previous analysis and it transports in time the history information of the atmosphere. The background is valid at the middle of the 6-hour assimilation window. The model (or analysis) variables are compared to observed quantities with possibly non-linear observation operators. Observations are collected from the assimilation window and the maximum time difference between an observation and the background is therefore three hours.

4D-Var data assimilation system is a temporal extension of 3D-Var. In 4D-Var the background is valid at the beginning of the assimilation window and the minimization is performed over the data assimilation window to obtain the best least square fit of the six hour model trajectory to the background at the initial time and to the observations spread over the assimilation window. In this implementation of 6-hour 4D-Var, the observations are organized into seven timeslots. Timeslots 1 and 7 cover the first and the last half-hour, respectively, of the 6-hour assimilation window. Timeslots 2 to 6 are one hour timeslots centered around the full hours of the assimilation window. Timeslot 4 is thus centered around the nominal analysis time.

2.1 Methodology

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

$$\boldsymbol{d}_i = \boldsymbol{y}_i - \boldsymbol{H}_i\, \boldsymbol{x}_b \tag{1}$$

In the variational data assimilation the innovation sequence is obtained in the first full nonlinear model integration in the iterative minimization process using $\boldsymbol{x}_b$ as initial condition. It is assumed here that in 4D-Var the observation operator also includes the model integration from the initial time to the observation time. The residual sequence is defined by replacing the background $\boldsymbol{x}_b$ in (1) with the analysis $\boldsymbol{x}_a$. The residual sequence is obtained in the variational

data assimilation in the final nonlinear model integration using the fully converged solution at the minimum of the cost function as initial condition.

The covariance of the innovation sequences at locations i and j as a sample mean is given by

$$cov(i,j) = \overline{\boldsymbol{d}_i \; \boldsymbol{d}_j{}^T} \tag{2}$$

Applying (2) for innovations with different horizontal separation, results in a discretized representation of a probability distribution function (hereafter *dpdf*) of innovation covariances. An example is displayed in Fig. 1a for AIREP (aircraft report) temperature over North America at 200 hPa during 1 Sep 1997 to 15 Oct 1997. The dots represent the sample mean innovation covariances as a function of separation between i and j. The covariance is close to zero for large horizontal separation, over 500 km, say. With decreasing separation, the covariance builds up and, in case of uncorrelated observation errors and in absence of biases, this is entirely due to the horizontal correlation of forecast error. At zero separation, the preceived forecast error variance is due both the forecast error variance and the observation error variance. Fitting a curve through the points of the *dpdf* and extrapolating to zero separation requiring a curve with zero derivative at zero separation, the perceived forecast error covariance implied by the sample of innovations can be partitioned into contributions from the background error and from the observation error.

The curve fitting through the points of the *dpdf* is performed as follows. The horizontal range is doubled and all the points are duplicated symmetrically across the true maximum range (obtaining a u-shaped group of points). Mean covariance is then calculated and a series of cosine functions is fitted in the least squares sense using N functions in the series and using the mean covariance as a constant term. The fitted curve has thus zero derivative at zero separation (and at maximum separation) and resolution can be regulated by varying N.

The curve fitting aims at filtering out variations in the sample which are not believed to be forecast error covariances. Different choices of base functions have been presented in the literature, such as gaussian or Bessel functions (Julian and Thiebaux, 1975). The cosine functions are chosen here for methodological simplicity and because it is believed that other sources of uncertainty are more important for the results than those arising from different choices of the base functions. Other sources of uncertainty are, for instance, biases in the observing and forecasting systems and the general bias of the estimation procedure in favour of the areas covered by dense observing networks. This method has nevertheless proven robust and has

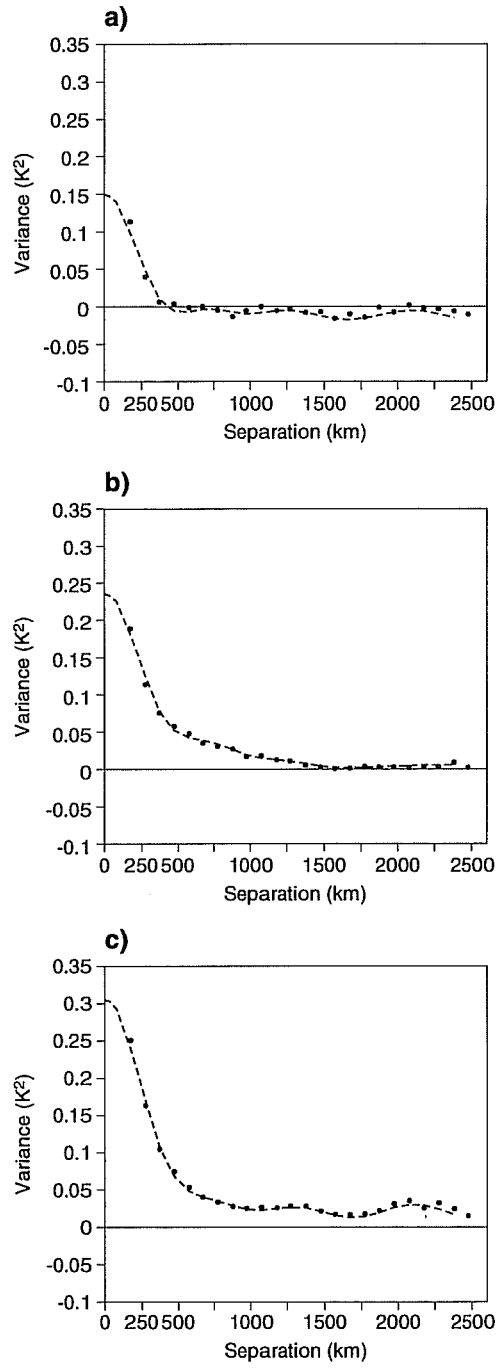

Figure 1: A discretized representation of a probability distribution function of innovation covariances for a half-hour timeslot 1 (a), for one hour timeslot 4 (b) and for one hour timeslot 6 (c) for AIREP temperature over North America at 200 hPa. The 4D-Var innovation sequence is from 1 Sep 1997 to 15 Oct 1997.

been used for innovations from various types of observations. N has been chosen subjectively to be the largest value which still produces a smooth appearance for the fitted curve without large variations for small horizontal scales. The associated estimation error for the partitioning remains small provided the sample of innovations is large and the mean innovation in the sample is small. It is not proven however whether this estimation procedure produces a correct positive definite covariance function.

The following parameter values have been used. The number of cosine functions N equals 8. Pairs of innovations with horizontal separation less than 125 km has been disregarded. The bin interval is 100 km for all points of the *dpdf*. The first point of the *dpdf* is therefore at 175 km separation in Fig. 1. Vertical interval is ± 25 hPa. These parameter choices are subjective containing a certain degree of arbitrariness, but the results are not too sensitive to variations in these parameter values. One anonymous reviewer suspects that the estimated standard deviation for the background error may be an underestimate with this method as opposed to a method using Bessel functions, and in particular because the shortest sampling separation for the covariance calculation is as large as 175 km. The intention is to avoid with the chosen 125 km cut-off separation for pairs of innovations the possible covariaces due to the sub-grid distance interpolations in the forward observation operators to appear and to be mixed with the horizontally correlated background errors.

In some cases 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, i.e. a mean innovation in the sample, which 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 the bias from the random part of the covariance. The bias must be deduced using complementary information. In some studies, for instance in Hollingsworth and Lönnberg (1986), the mean innovation is subtracted separately for each station. This removes the effect of both the instrumental bias and the mean forecast errors. In this study the biases are not removed, but care has been exercised in interpreting the results by excluding cases where significant biases seem to be present.

2.2 Growth of background error covariance

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 *dpdf* for innovation covariance of Fig. 1a corresponds to the first half-hour of the 6-hour assimilation window of 4D-Var. The background error variance is about $0.15K^2$ in this case. At the middle of the 6-hour assimilation window (Fig. 1b; timeslot 4) the forecast

error variance of the background trajectory has increased to about $0.23K^2$ in three hours. 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 the initial time. This is because of transfer of energy across the scales which also happens with the tangent linear model (eg. Tanguay et al., 1995).

By the end of the assimilation window (Fig. 1c; timeslot 6), the covariance has increased further to about $0.31K^2$. Part of the increased covariance is due to a small bias which is present in this particular sample of innovations. It is not trivial to attribute the bias in this case either to the observing system or to the forecasting system, as the usually reliable reference network of radiosondes operates only at the main synoptic observing hours, and it is not completely free of biases either. In this case, the bias seem to develop during the six hour background integration, supporting the idea that it originates from the forecast system rather than from the observing system.

By definition, the background error covariance of 4D-Var corresponds to the covariance at the initial time of the assimilation window. In contrast with this definition, the forecast error covariance associated with the background trajectory is used in a specific meaning to distinguish it from the background error covariance. Similarly, the distinction should be clear for the evaluation of the covariance length scale implied by the background penalty term of the 4D-Var data assimilation system. The comparison is relevant only with the covariance length scale deduced from the innovations at the initial time of the assimilation window. Such a comparison is performed with the present data. The length scales of the background error covariace implied by the background constraint term (Derber and Bouttier, 1999) represent hemispheric averages over both well observed and poorly observed areas, and are in general much broader than the covariance length scales estimated with the observational method for a limited and well observed area, such as North America (Fig. 1a).

2.3 Temporal behaviour of the error variances

So far we have seen qualitatively the increase of forecast error covariance over the assimilation window. Next, the observation and background error variances are estimated over the six hour assimilation window to quantitatively reveal their temporal behaviour in 3D-Var as well as in 4D-Var. The error variances are estimated using innovations from AIREP component winds over North America at 200 hPa over a period of 1 Sep 1997 to 15 Oct 1997.

In 3D-Var the background is valid at the middle of the assimilation window, and the estimated standard deviation of background error has its minimum of about 2 m/s there (thick

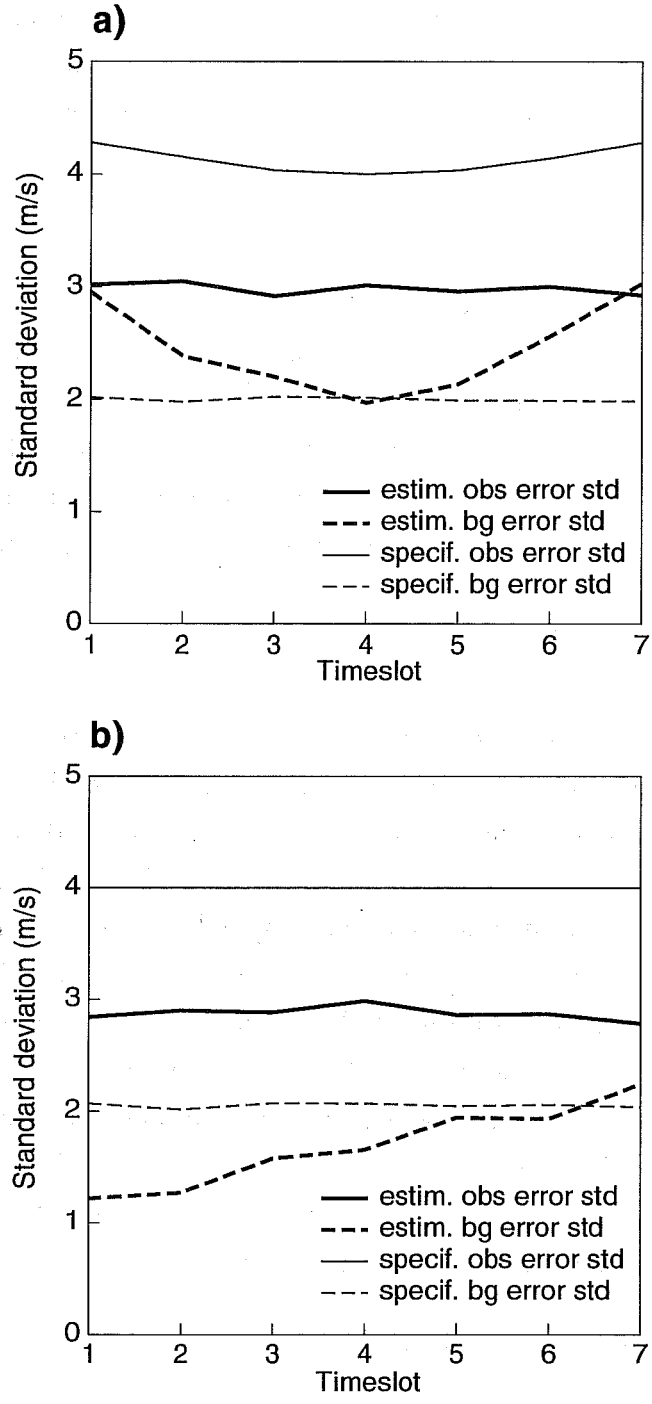

Figure 2: Estimated (thick lines) and specified (thin lines) standard deviation of background error (dashed lines) and observation error (solid lines) in 3D-Var (a) and in 4D-Var (b), respectively, for AIREP component wind over North America at 200 hPa. The estimation is based on an innovation sequence from 1 Sep 1997 to 15 Oct 1997. The numbering of timeslots is as follows. Timeslot 1 (7) cover the first (last) half-hour of the 6-hour assimilation window. One hour timeslots 2 ... 6 are centered around full hours of the window. Timeslot around 12UTC, for instance, is number 4.

dashed line in Fig. 2a) in timeslot 4. The estimated standard deviation of observation error remains constant at about 3 m/s, within the estimation uncertainty (thick solid line). The

specified standard deviation of observation error (thin solid line) of 4 m/s 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, i.e. the error due to the time difference between the observation and the analysis. Note that in the error variance estimation from the innovations, the persistence error in 3D-Var appears as a horizontally correlated error which contributes to the background error variance. The correction of persistence error in the specified observation error standard deviation is qualitatively right but not quite large enough, i.e. the u-shape in the curve for specified observation error standard deviation is not quite as steep as in the curve for estimated background error standard deviation. The background error variance specified in this implementation of 3D-Var (thin dashed line) agrees well with the estimated background error variance. In this curve there is a tiny variation with the timeslot. This is because this curve represents the geographically varying specified background error variance interpolated to observation locations, and there is a different set of observations in each timeslot. In summary, it seems that there is too little weight on these observations for this 3D-Var assimilation system to perform optimally, i.e. too large specified observation error variance for AIREP winds. Tuning of a global assimilation system based on limited area statistics is not however straightforward, and a sequence of retunes is usually needed.

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 at a level of about 1.2 m/s. 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 two features. First, at the middle of the assimilation window (timeslot 4), the 4D-Var background error standard deviation is about 20% smaller than the corresponding value for 3D-Var, i.e. about 1.6 m/s for 4D-Var and about 2.0 m/s for 3D-Var. Second, the 4D-Var forecast error standard deviation associated with the background trajectory reaches the (minimum) level of 3D-Var background error standard deviation of about 2 m/s in timeslot 5 or 6. In other words, forecasts from the 4D-Var data assimilation are one or two hours better than forecasts from the 3D-Var data assimilation at the very earliest forecast range.

The specified background error standard deviation (thin dashed line) is the same as in 3D-Var, with no increase with time. (For cycling of the background error variances, a local growth law of forecast errors (Savijärvi, 1995) is applied in this implementation of 3D/4D-Var.)

Reassuringly for the variance estimation procedure, the estimated standard deviation of observation error (thick solid line) remains almost constant throughout the assimilation window, and is the same both in 3D-Var and in 4D-Var systems. This is reasonable as there is no particular reason for the quality of observations to depend on time or on the assimilation system. Persistence error is non-existent in 4D-Var, and therefore the specified observation error standard deviation (thin solid line) is kept constant.

There is a variety of observed parameters and geographical areas for which the innovations behave just as illustrated in Fig. 2. This particular parameter and area is selected because of an abundance of observations there, thus providing very stable estimates for these statistics.

2.4 Vertical structure of error variances

Temporal behaviour of forecast and observation error standard deviations is illustrated in Fig. 2 at jet level where there is a remarkable growth of the forecast error associated with the background trajectory over the 4D-Var assimilation window. The vertical structure of forecast error standard deviation for wind and temperature is illustrated next at the beginning and at the end of the 6-hour assimilation window of 4D-Var. This also reveals the vertical structure of the error growth over the assimilation window. The innovations are for AIREP component wind and temperature observations over North America. These observation are rarely available much above 200 hPa which is therefore the vertical limit for this study.

In Fig. 3, there are two pairs of curves in both panels. Dashed (solid) curves are for the background (observation) error standard deviation. Thick (thin) lines denote the values at the beginning (end) of the assimilation window. It is known *a priori* that the estimated observation error standard deviation should be unchanged over the assimilation window. Therefore the differences in the two curves for the observation error standard deviation give some indication of the estimation uncertainty. For instance, the observation error standard deviation for wind (Fig. 3a) has its maximum difference of about 0.2 m/s at 700 hPa. All the wind curves in Fig. 3a can thus be thought to have such error bars due to random differences in the samples of innovations. Temperature curves in Fig. 3b should have the respective error bars of about 0.1 K.

Observation error standard deviation for AIREP component wind (Fig. 3a) increases steadily with height from about 2.2 m/s at 1000 hPa to about 2.8 m/s at 200 hPa. Radiosonde data reveal similar vertical structure of observation error standard deviation with somewhat larger values around the jet level (not shown). For the AIREP temperature (Fig. 3b), the observation error standard deviation decreases from the near surface value of about 1.3 K to its minimum

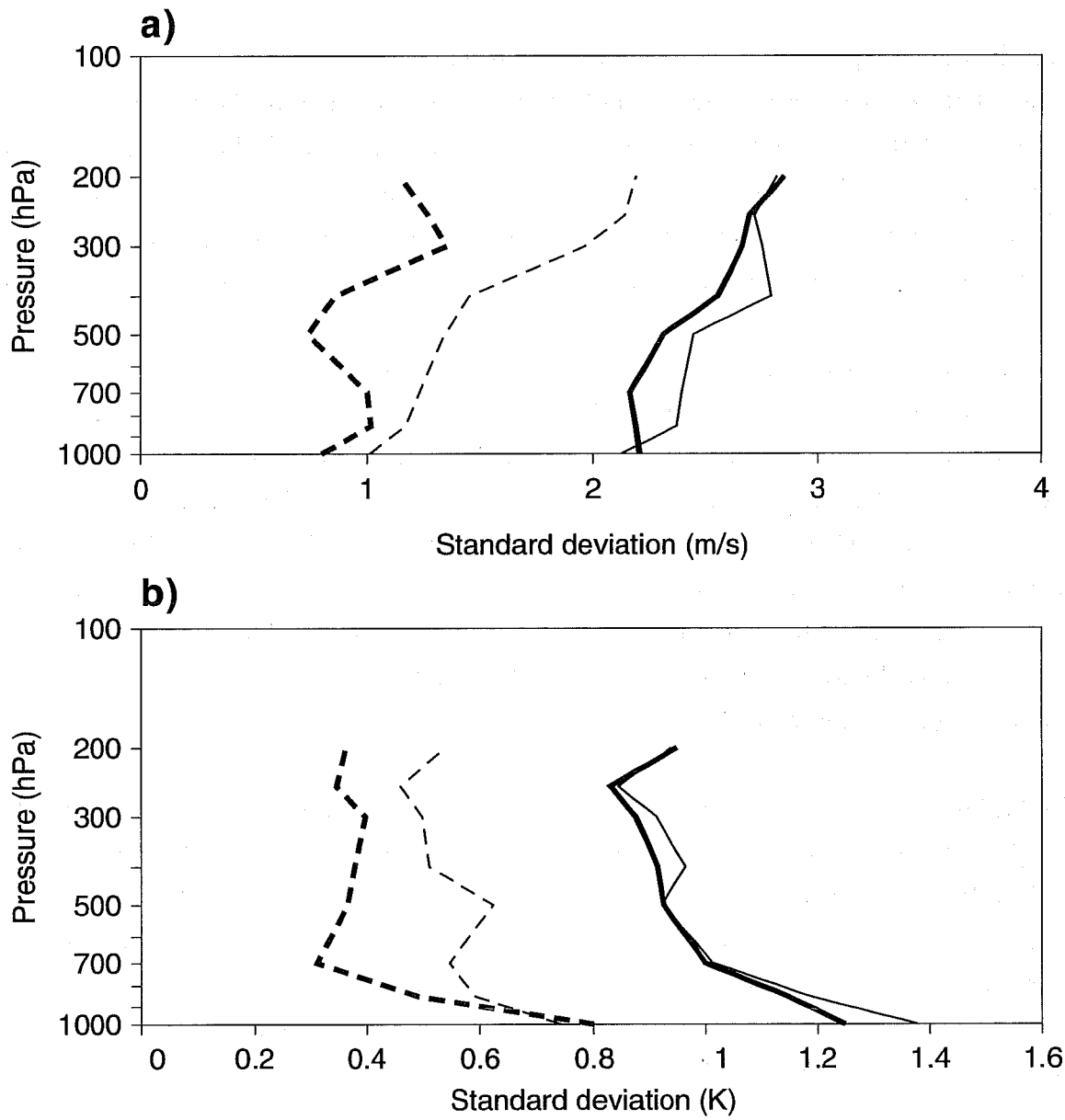

Figure 3: Vertical profiles of the background (dashed line) and observation (solid line) error standard deviation at the beginning (thick line) and at the end (thin line) of the 6-hour assimilation window for AIREP component wind (a) and temperature (b) over North America. The 4D-Var innovation sequence is from 1 Sep 1997 to 15 Oct 1997.

of about 0.83 K at 250 hPa level, followed by an increase above 250 hPa. The maximum near the surface is likely to be related to a larger error of representativity there. The vertical structure of the estimated temperature observation error standard deviation is very similar as for radiosonde data (not shown).

Overall, the background error standard deviation for wind and for temperature (Fig. 3) is smaller that the observation error standard deviation throughout the troposphere both at the beginning of the assimilation window and at the end of the assimilation window. This is remarkable as the background at the initial time of the assimilation window is a three hour forecast as measured from the middle of the previous assimilation window, and the background at the final time of the assimilation window is a nine hour forecast.

The background error standard deviation for component wind (Fig. 3a) is initially about 0.9 m/s below 400 hPa level and about 1.2 m/s above 400 hPa level. There are two local maxima at 850 hPa and at 300 hPa. The largest increase of the background error standard deviation occurs at 200 hPa level, where the increase is from the initial value of 1.2 m/s to 2.2 m/s six hours later. There is also a large increase at around 500 hPa level. Near the surface the increase of the error standard deviation is very small; for instance, at 850 hPa from about 1.0 m/s to 1.1 m/s six hours later. Radiosonde data also reveal similar features in the vertical structure, such as the local minimum around 500 hPa level, as estimated at the middle of the assimilation window (not shown).

The background error standard deviation for temperature (Fig. 3b) is initially about 0.37 K throughout the troposphere, except much higher values at the lowest levels with a maximum of 0.8 K at 1000 hPa. Six hours later, the values have increased to about 0.5 K, except near 500 hPa level with a local maximum of over 0.6 K, and near the surface with no increase at all. There is actually a peculiar decrease of forecast error at 1000 hPa level, but this is within the estimation error bar of 0.1 K. The vertical structure with the large values nearest to the surface can also be seen in the radiosonde data (not shown).

The vertical structure of background error for 3D-Var is qualitatively similar as for 4D-Var, except that the 3D-Var minimum occurs at all levels at the middle of the assimilation window (not shown), and increases toward the beginning and the end of the assimilation window. The vertical structure of observation error is practically the same in 3D-Var as in 4D-Var.

It would be interesting to compare the accuracy of the ECMWF data assimilation system of today with earlier operational systems of ECMWF. The statistical properties of one such

Level	Observation error std (m/s)		Background error std (m/s)	
	OI 1983	4D-Var 1997	OI 1983	4D-Var 1997
200hPa	3.3	4.0	5.2	2.4
500hPa	4.0	3.3	4.2	1.5
850hPa	3.4	3.2	2.6	1.6

Table 1: Estimates of the vector wind observation and background error standard deviations for the ECMWF OI system of 1983 and for the ECMWF 4D-Var system of 1997, respectively.

system are documented in details in the literature, against which we will briefly compare our results. Hollingsworth and Lönnberg (1986) studied the innovations from the ECMWF global grid point model and Optimum Interpolation analysis system in a winter period in 1983. Their Figure 3 displays the vertical profiles of observation and prediction error standard deviation for vector wind at North American radiosonde stations. Their Figure 3 differs qualitatively from our Fig. 3a with respect to the relative magnitude of observation and forecast error. In their work, the forecast error standard deviation exceeds the observation error standard deviation above 500 hPa level, whereas in the results of this stydy, the background/forecast error standard deviation is clearly smaller than the observation error standard deviation throughout the troposphere, and throughout the 4D-Var assimilation window of six hours. The estimated error standard deviations for 200/500/850hPa level (Table 1) for "OI 1983" are extracted from Figure 3 in Hollingsworth and Lönnberg (1986), and the values "4D-Var 1997" are extracted from Fig. 3a of this study which values are multiplied by $\sqrt{2}$ to convert the error in component wind to the error in vector wind.

It should be remembered that the error estimates are obtained in these studies with different methods of estimation and for different observing systems, and they should be compared with caution. It is justified however to say that a remarkable increase in the accuracy of ECMWF data assimilation system has taken place over the last 15 years. Assuming the accuracy of the observing systems of this comparison to be similar, the background error standard deviation is now about half the value it used to be 15 years ago. There are many contributing factors behind these gains which come from improved forecast model formulation, from increased model resolution, from better data assimilation technique, from better exploitation of remote sensing data and from vastly improved AIREP availability over the area of interest, for instance.

3 The fit of observations to the analysis and the model error

The residual sequence represents the fit of observations to the analysis, or the fit of observations to the *a posteriori* estimate of the state of the atmosphere. The residuals are a very useful source of information for diagnosing the properties of data assimilation systems, as the pioneering

research of Hollingsworth and Lönnberg (1989) shows. Here, the diagnosis of residual sequences of 4D-Var is restricted to the question of the model error, as the ECMWF implementation of 4D-Var makes an implicit perfect model assumption.

3.1 The expectation

The model error in a strong constraint variational problem, if not properly dealt with, leaves a statistical signature to the residual sequence. Ménard and Daley (1996) have explained how the fit of observations to the analysis should behave over the assimilation window. The following considerations hold in a statistical sense when a large sample of residuals is studied.

In an idealized case of a perfect model of a dynamically neutral system, the standard deviation of the residuals should be constant in time over the assimilation window, and lower than the observation error standard deviation. The presence of growing modes makes it easier to fit observations at the end of the assimilation window and so the fit of observations to the analysis should improve towards the end of the window. The presence of decaying modes in the model makes it easier to fit observations at the beginning of the assimilation window and so the fit of observations to the analysis should improve towards the beginning of the window. Therefore a perfect model with mixed modes, i.e. growing, neutral and decaying modes present, should produce a concave[1] fit of observations to the analysis.

Model error should manifest itself in the following way. With neutral dynamics, the fit should be convex and the misfit of analysis to observations should be higher than with perfect model. Growing modes diverge from reality in the presence of model error, making it harder to fit observations at the end of the assimilation window. Erroneous decaying modes, on the other hand, deteriorate the fit at the beginning of the assimilation window. Therefore the fit of observations to the analysis for an imperfect model with mixed modes should be gradually shifted from a concave to a convex with increasing model error, associated with an increasing level of misfit to observations.

3.2 Results with the 6-hour 4D-Var

The innovation and residual sequences studied here cover a period of 16 to 29 April 1998. A separate period from an experimental 4D-Var implementation with a 12-hour assimilation window cover a period of 16 September to 10 October 1998. The interest here is on the spread of

[1] Concave and convex curves are defined here by the sign of second time derivative of the curve for the fit of observations to the analysis over the assimilation window: a negative second derivative for a concave curve and a positive second derivative for a convex curve.

the sample of innovations and residuals.

Figure 4 displays the evolution of the standard deviation of innovations (solid curves) and residuals (dashed curves) over the assimilation window for AIREP (amdar) temperature (Fig. 4a) and for AIREP (acar) component wind (Fig. 4b) for the North Atlantic at 250 hPa. Amdar and acar are two slightly different aircraft observation reporting systems. The standard deviation of innovations monotonically increases over the assimilation window for these two observing systems. The standard deviation of the residuals is lower than of the innovations. The standard deviation of the residuals has a minimum exactly at the middle of the assimilation window, both for component wind and temperature.

The specified standard deviation of observation error at 250 hPa is 1.3 K for AIREP temperature, and 4.0 m/s for AIREP component wind. These are larger than the standard deviation of innovations, and obviously misspecified to be too large. The estimated standard deviations are 0.83 K for temperature (see for Fig. 3b) and 2.7 m/s for component wind (see for Fig. 3a), respectively. The AIREP wind observations are fitted within the estimated observation error standard deviation, but the temperature observations are not fitted so closely. There is clearly some sub-optimality in the assimilation system, due to misspecified observation error standard deviations.

Figure 5 displays the standard deviation of the innovations (solid curves) and of the residuals (dashed curves) for DRIBU surface pressure observations over the North Atlantic (Fig. 5a) and over the South Atlantic (Fig. 5b). Note that the curves for the South Atlantic are more noisy than for the North Atlantic because of a much smaller amount of DRIBU observations there. Over the North Atlantic, the general shape of the curves at sea level is fairly similar to those over the same area at upper levels in Fig. 4. At sea level, there is an increase of the standard deviation of the innovations over the assimilation window, and the minimum of the standard deviation of the residuls occurs in timeslot 5. Over the South Atlantic the standard deviation of innovations increases rapidly over the assmilation window, and the minimum in the standard deviation of the residuals is in timeslot 3, i.e. one hour before the middle of the assimilation window. By ignoring some of the noise in this curve, one can note a decreasing trend of standard deviation for the first three timeslots of the assimilation window, and an increasing trend of standard deviation over the following four timeslots of the assimilation window.

The curves for standard deviation of the residuals for AIREP and DRIBU observations (dashed curves in Figs. 4 and 5) are convex rather than concave. No assessment of the statistical significance of the shape will be presented. Therefore two alternative conclusions regarding the

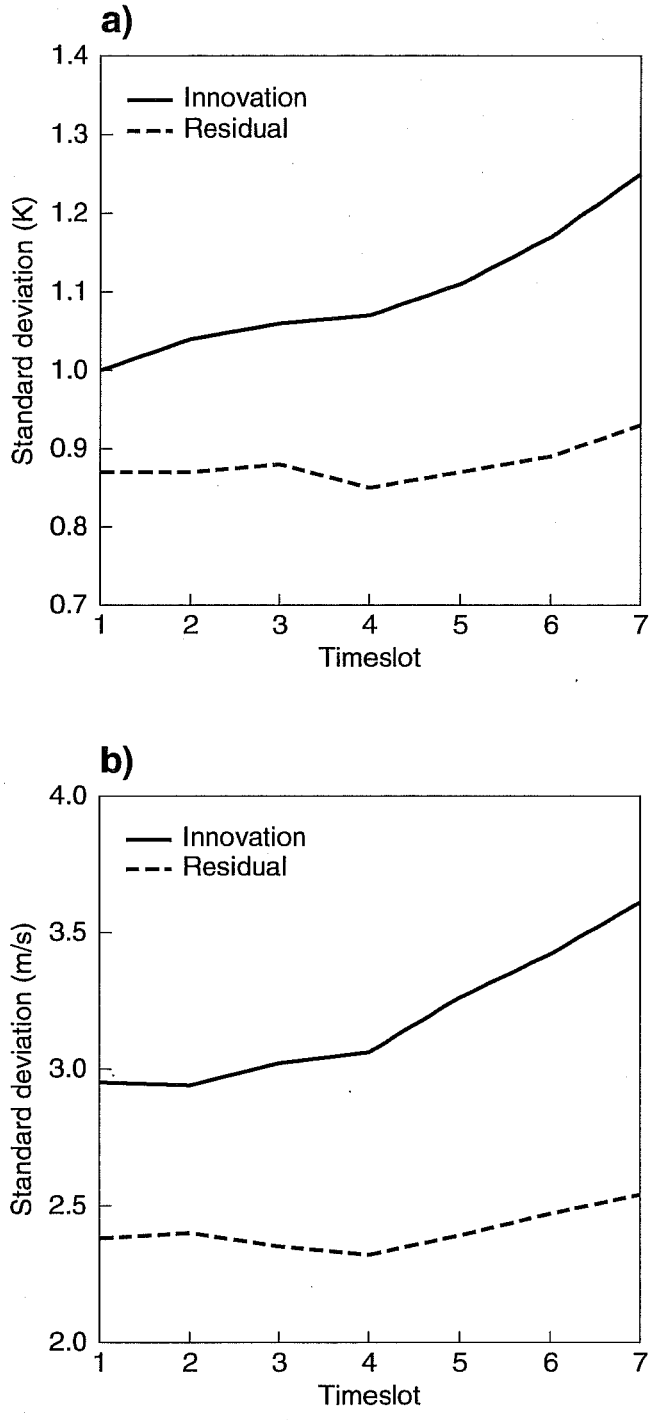

Figure 4: Standard deviation of innovations (solid lines) and residuals (dashed lines) in 4D-Var for AIREP (amdar) temperature (a) and AIREP (acar) component wind (b) at 250 hPa over the North Atlantic. The curves are based on innovation and residual sequences from 16 to 29 April 1998. Amdar and acar are two slightly different aircraft observation reporting systems.

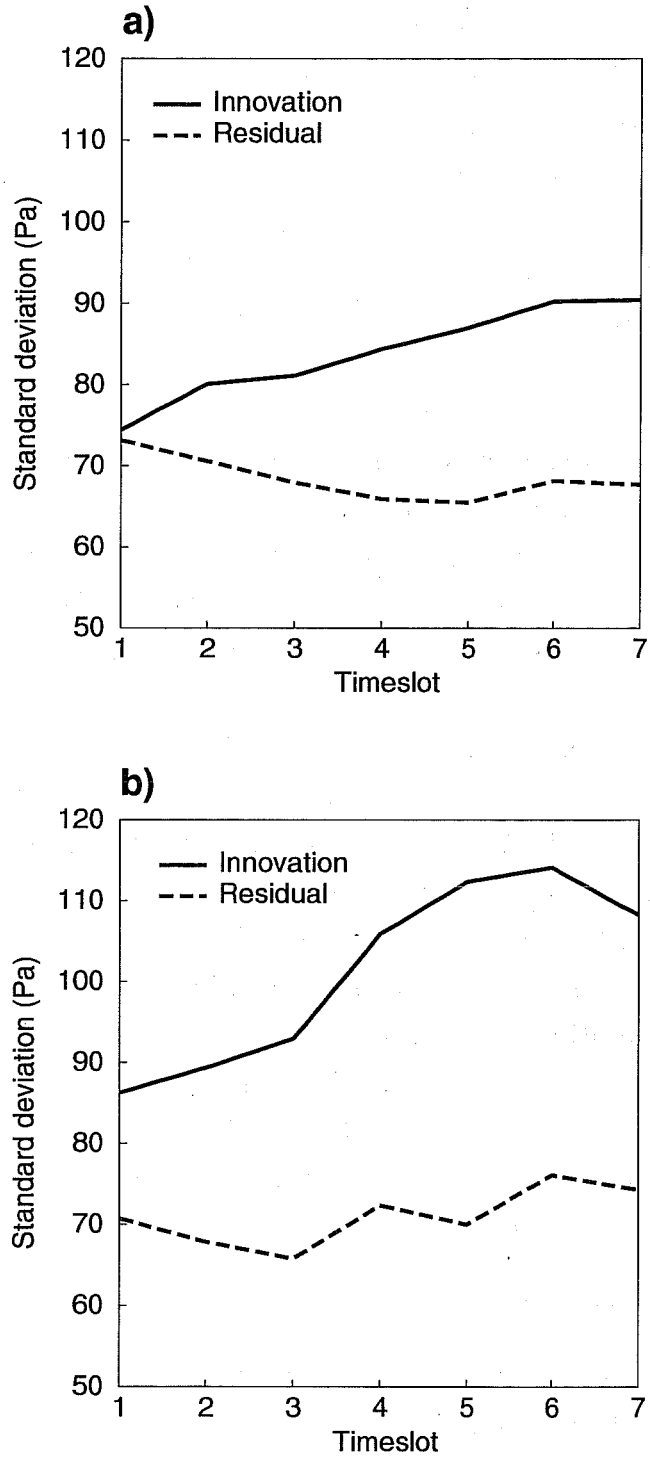

Figure 5: Standard deviation of innovations (solid lines) and residuals (dashed lines) in 4D-Var for DRIBU surface pressure in the North Atlantic (a) and in the South Atlantic (b). The curves are based on innovation and residual sequences from 16 to 29 April 1998.

presence of model error to be drawn from the shape of the curves for standard deviation of the residuals in Figs. 4 and 5. In the first alternative, the convexity in the curves is regarded as non-significant which is the same as considering the curves as flat over the assimilation window. In this case, if the dynamical system is considered neutral over the 6-hour assimilation window, the conclusion would be that the model error is negligible. But if the dynamical system is considered to include mixed modes, the conclusion would be that the model error is non-negligible. In the second alternative, the convexity in the curves is regarded as significant. The conclusion would be that the model error is non-negligible, regardless of the nature of the model dynamics. Regarding the first alternative, there are several examples in the literature (Rabier and Courtier, 1992; Thépaut et al., 1996; Järvinen et al., 1999) of the presence of growing modes in 4D-Var, even with an assimilation window of six hours. Therefore both a flat and a convex curve for standard deviation of the residuals would indicate the presence of a model error in this assimilation system. Thus proving the statistical significance whether the curve is convex rather than flat, may not be absolutely necessary. One can therefore say that the perfect model assumption may not be correct in the ECMWF assimilation system, and it may affect the accuracy of the system, even with as short an assimilation window as six hours.

There is another interesting feature in Fig. 5 which is related to the quality of the background and the analysis of surface pressure over the Atlantic Ocean. The standard deviation of the residuals (dashed lines in Fig. 5) is very similar both over the North and South Atlantic. It is perhaps slightly larger towards the end of the assimilation window over the South Atlantic. Therefore, not suprisingly, the quality of background both over the North and South Atlantic at the beginning of the assimilation window (solid lines in Fig. 5) is very similar, i.e. the standard deviation of the innovations is equal initially. The standard deviation of the innovations increases however much more rapidly over the South Atlantic, i.e. the forecast error growth for surface pressure is much quicker over the South Atlantic than over the North Atlantic. This is most likely because of a more inaccurate initial state of the atmosphere upsteam of the South Atlantic than upstream of the North Atlantic.

3.3 Results with the 12-hour 4D-Var

Further insight into the question of presence of model error is given by considering an assimilation experiment with a longer assimilation window of 12 hours, instead of the standard six hours. Longer assimilation window increases the non-linearity of the system and makes the growth and decay of unstable modes more pronounced. The innovation and residual sequences cover a period of 16 September to 10 October 1998, and are extracted for the area of North America.

The thirteen timeslots of the 12-hour 4D-Var of ECMWF are organized for 12 UTC nominal analysis time such that the first half-hour timeslot extends from 0300 to 0330 UTC, followed by a one hour time slot extending from 0330 to 0430 UTC, and so on. Timeslot 10 is thus centered around the nominal analysis time of 12 UTC. The assimilation window for the 00 UTC nominal analysis time extends from 1500 to 0300 UTC. In the 6-hour 4D-Var the main synoptic observing network operates at the middle of the assimilation window. In the 12-hour 4D-Var the main synoptic observation coverage is closer to the end of the assimilation window at timeslot 10 and the observation distribution in time is not as symmetric as with the 6-hour 4D-Var.

Figure 6 displays the standard deviation of the innovations (solid curves) and of the residuals (dashed curves) for AIREP temperature at 200 hPa level (thin lines in Fig. 6a) and at 250 hPa level (thick lines in Fig. 6a). The statistics are also displayed for the component wind at 250 hPa level for two different kind of AIREP reports: for manual reports (thin lines in Fig. 6b) and for acar reports (thick lines in Fig. 6b). The standard deviation of the innovations (solid curves) increases throughout the 12-hour assimilation window for both observed variables and for different levels and observing systems. The standard deviation of the residuals (dashed curves) have a minimum in all cases within one hour from the middle of the assimilation window, i.e. in timeslot 6 or 7. The minimum is followed by an enhanced increase of standard deviation towards the end of the assimilation window. The standard deviation of the residuals is higher at the end than at the beginning of the assimilation window. This behaviour could be anticipated for an imperfect model in the presence of growing modes. There is a local maximum at timeslot 10, and to a lesser extent also in timeslot 4, coinciding with large amounts of observations from the main synoptic network. This indicates that AIREP data is less closely fitted in those timeslots as there is an abundance of other observations to be fitted.

Another statistical signature of the model error is that with an increasing length of the assimilation window, the standard deviation of the residuals should also increase. The standard deviation of the residuals for AIREP temperature with a 6-hour assimilation window (Fig. 4a) and with a 12-hour assimilation window (Fig. 6a) is at the same level of about 0.87 K. The standard deviation of the residuals for AIREP component wind with a 6-hour assimilation window (Fig. 6a) is about 2.4 m/s and it is with a 12 hour assimilation window about 2.8 m/s, i.e. a larger standard deviation of the residuals for a longer assimilation window. This could be explained with the assumption of the presence of model error.

There are some statistical signatures of the presence of the model error in this implementation of 4D-Var. For defining the *magnitude* of the model error, the observation and background

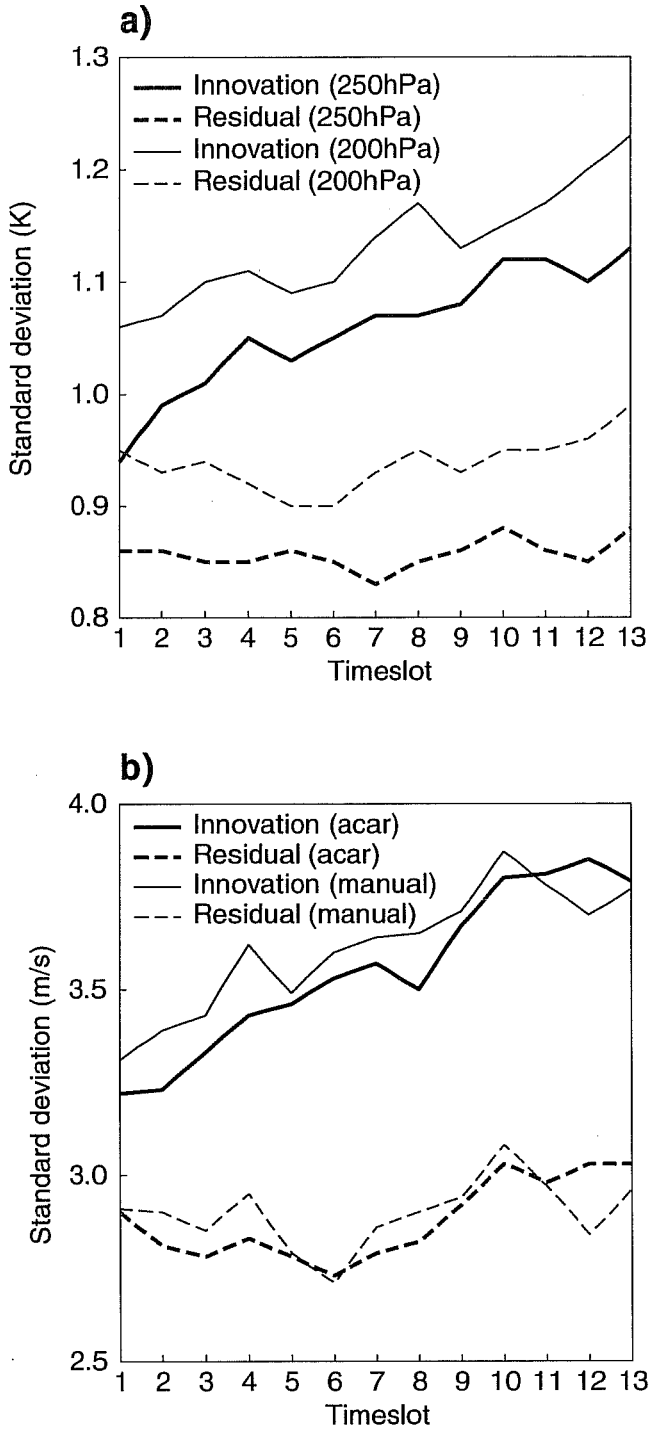

Figure 6: Standard deviation of innovations (solid lines) and residuals (dashed lines) in 4D-Var for AIREP temperature at 200 hPa and 250 hPa (a), and for AIREP component wind separately for two different kind of aircraft reports (manual and acar) at 250 hPa (b) over North America in an assimilation experiment with a 12 hour assimilation window. The curves are based on innovation and residual sequences from 16 September to 10 October 1998.

error standard deviations should be correctly specified. This can be checked from the statistics collected so far. For instance, the specified observation error standard deviation for AIREP temperature at 250 hPa level is 1.3 K. The maximum standard deviation of the innovations in Fig. 4a is 1.25 K. The corresponding values for AIREP wind is 4.0 m/s and 3.6 m/s (Fig. 4b), and for DRIBU surface pressure 140 Pa and 120 Pa (Fig. 5b), respectively. This means that the specified observation error standard deviations are over-estimates. Figure 2 also indicates that the observation error standard deviations are specified to be too large. Therefore there is no reliable way at present to define the *magnitude* of the associated model error. A renewed attempt can be made after retuning the specified error standard deviations.

4 Summary and conclusions

It is instructive to study the temporal aspects related to the innovation and residual sequences over the assimilation window. The statistical parameter estimation for 3D-Var requires observations which are well spread in space. Radiosonde observations are very useful for this purpose because of their excellent vertical sampling. The estimation for 4D-Var benefits of a good spread of observations also in time, such as aircraft reports, drifting buoy and synoptic observations, and polar orbiting satellite measurements as well.

The innovations reveal that in the ECMWF 6-hour 4D-Var data assimilation system the background is significantly more accurate than in the 3D-Var. At the middle of the 6-hour assimilation window the background error standard deviation is about 20% smaller in 4D-Var than in 3D-Var. Forecasts (or background trajectories) of the 4D-Var system are better than 3D-Var forecasts by about one to two hours at the very earliest forecast range. The 4D-Var forecast error variance associated with the background trajectory grows monotonically over the assimilation window from the initial minimum value of the background error variance. There is an associated broadening of the horizontal length scale of the forecast error covariance. The growth rate of forecast error variance increases with height in the troposphere.

A significant improvement in the accuracy of the data assimilation and forecasting system at ECMWF has taken place over the last 15 years. This improvement is revealed by comparing the estimated observation and background error covariances of this paper with those of Hollingsworth and Lönnberg (1986). There are many factors contributing to this achievement, such as improvements in forecast model formulation, increased model resolution, better data assimilation technique and better exploitation of the remote sensing data. It is also important to note the vastly improved AIREP availability, in particular over the North America, which enables the hourly estimation of the relevant statistical parameters presented in this paper.

The residual sequences can be used for diagnosing data assimilation systems. Here the effect of model error is studied in a strong constraint variational problem. A weakly convex curve is found for the standard deviation of residuals over the assimilation window both for surface pressure and for component wind and temperature at the jet level. Also the standard deviation of the residuals of the component wind at the jet level increases if the assimilation window is increased from 6 to 12 hours. These findings indicate that the perfect model assumption of the ECMWF 4D-Var assimilation system may not be correct and, it may affect the system with as short an assimilation window as six hours, and is more likely to affect the system when the length of assimilation window is increased up to 12 or 24 hours.

For improving the optimality of variational data assimilation systems, it is necessary to perform a sequence of retunes, until the specified and diagnosed error covariances agree. Every major change in the data assimilation system, including the forecast model, or in the observing systems enforce yet another retune. It is therefore desirable to aim at a more automated process of estimation of statistical parameters as a part of the operational production and monitoring, rather than making the estimation and retuning only occasionally, perhaps as a part of research activities.

5 Acknowledgements

I would like to thank my colleagues at ECMWF. François Bouttier supervised the work providing with valuable advice and support. Anthony Hollingsworth brought the work into perspective with earlier work in this field, and commented the manuscript most usefully. Thanks are also due to Erik Andersson and Michael Fisher for many discussions. Rob Hine finalized the figures most professionally. I would like to thank Olivier Talagrand (LMD, École Normale Supérieure) for his guidance on the diagnosis of the model error and for the time spent with the manuscript. Comments on the manuscript by Juhani Rinne at Finnish Meteorological Institute are also warmly acknowledged. The author is grateful for the two anonymous reviewers for greatly improving the manuscript with their constructive criticism.

REFERENCES

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljević, D., Branković, 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.

Bouttier, F. and Courtier, P. 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., Vasiljević, 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.

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

Derber, J. and Bouttier, F. 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**, 195-221.

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.

Hollingsworth, A. and Lönnberg, P. 1989. The verification of objective analyses: diagnostics of analysis system performance. *Meteorol. Atmos. Phys.*, **40**, 3-27.

Julian, P. R. and Thiebaux, H. J. 1975. On some properties of correlation functions used in optimum interpolation schemes. *Mon. Wea. Rev.*, **103**, 605-616.

Järvinen, H., Andersson, E. and Bouttier, F. 1999. Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus*, **51A**, 469-488.

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.

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

Rabier, F. and Courtier, P. 1992. Four-dimensional assimilation in the presence of baroclinic instability. *Q. J. R. Meteorol. Soc.*, **118**, 649-672.

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 Vasiljevic, D. 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 Simmons, A. 2000. The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126**, 1143-1170.

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

Savijärvi, H. 1995. Error growth in large numerical forecast systems. *Mon. Wea. Rev.*, **123**, 212-221.

Tanguay, M., Bartello, P. and P. Gauthier, 1995: Four-dimensional data assimilation with wide range of scales. *Tellus*, **47A**, 974-997.

Thépaut, J.-N., Courtier, P., Belaud, G. and Lemaître, G. 1996. Dynamical structure functions in a four-dimensional variational assimilation: a case study. *Q. J. R. Meteorol. Soc.*, **122**, 535-561.