The operational implementation of 4D-Var

Reprinted from ECMWF Newsletter Number 78

On 25 November 1997, the ECMWF operational system was switched to use a 4D-Var assimilation algorithm. 4D-Var is a four-dimensional variational data assimilation technique that performs a statistical interpolation between a distribution of meteorological observations in time and space (Figure 1) and an *a priori* estimate of the model state (called the background). A special property of 4D-Var is that it takes into account the dynamics and the physics of the forecast model in order to ensure that the observations are used in a meteorologically consistent way. As will be illustrated below, this implies that asynoptic data can be used at their appropriate observation times, the structure functions of the assimilation are flow-dependent, and the physical imbalances in the model are reduced.

This is the first ever operational application of the 4D-Var technique successfully applied to a high-resolution assimilation and forecast system. After the initial proposal in 1985 (Lewis and Derber 1985, Courtier and Talagrand 1987, Talagrand and Courtier 1987), this was made possible by more than ten years of scientific and technical developments in and around ECMWF's Integrated Forecast System (IFS) (Pailleux, 1997) as well as the availability of a powerful new computer system organized around a Fujitsu VPP700 with 116 processors. The pre-operational and early operational experience with 4D-Var has shown a very clear improvement in the performance of the forecasting system. This is an impressive yet young assimilation system that offers an exceptional scope for future improvements.

From 3D-Var to 4D-Var

The variational analysis system was designed to allow a smooth transition from 3D-Var to 4D Var; technically speaking, the software changes only slightly with 4D-Var (although it is computationally more expensive). Both systems work with a T213L31 forecast system, and compute analysis increment fields at a lower resolution of T63L31, as well as analysis and forecast error estimates at T42L31. The observations used are almost the same: conventional synoptic stations (including Australian pseudo-observations), buoys, radiosondes, aircraft reports, cloud-motion winds, cloud-cleared TOVS radiances and ambiguous scatterometer winds. In both systems, the analysis is produced every 6 hours, e.g. for the 12 UTC analysis (the initial condition for the medium-range forecast and the Ensemble Prediction System (EPS)), the observations between 9 UTC and 15 UTC are used. The formulation of the background term is identical in 3D- and in 4D Var, as is the penalization of the tendency of the gravity wave modes. The variational analysis problem is solved at a low resolution using the incremental technique in order to reduce the computer costs: first, the observations are compared to the high-resolution T213 background model state, but the minimization of the cost function of the analysis is then carried out at a lower T63 resolution using the observation operators linearized in the vicinity of the high resolution background. Finally, the T63 increments are converted back into T213 increments using normal mode initialization and added to the T213 background. This is roughly equivalent to replacing the minimization of a T213 cost-function by the minimization of its T63 quadratic approximation in the vicinity of the background. As explained below, the technique is slightly modified in 4D-Var.

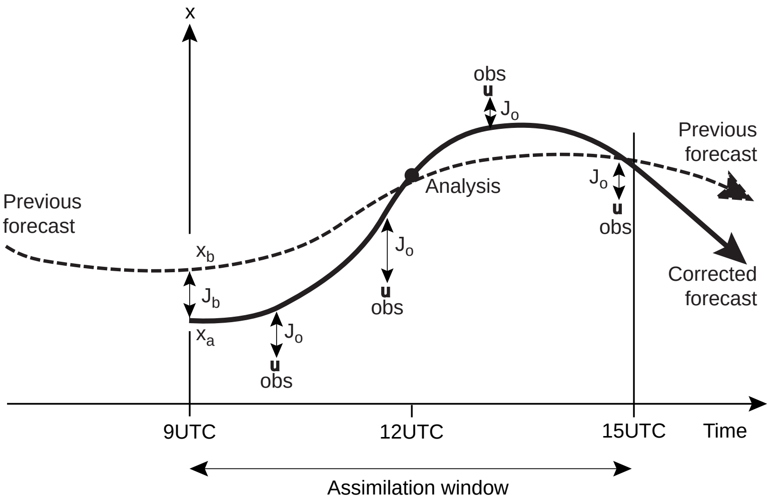
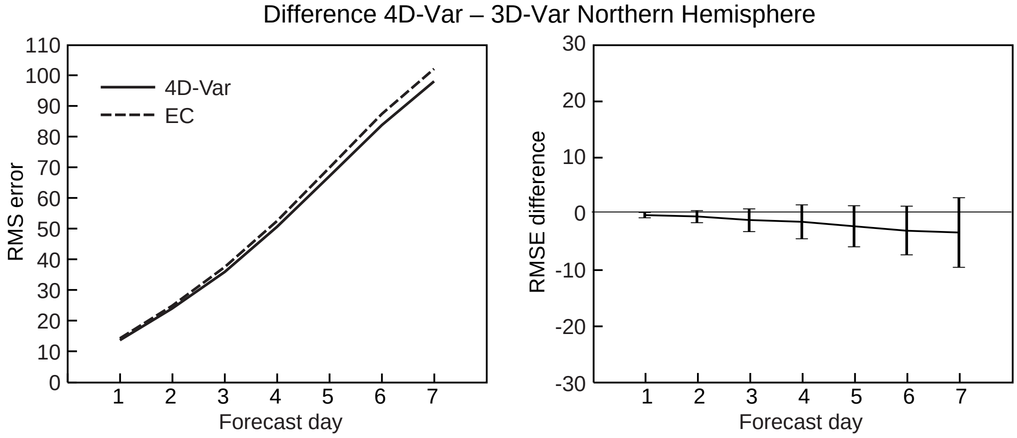
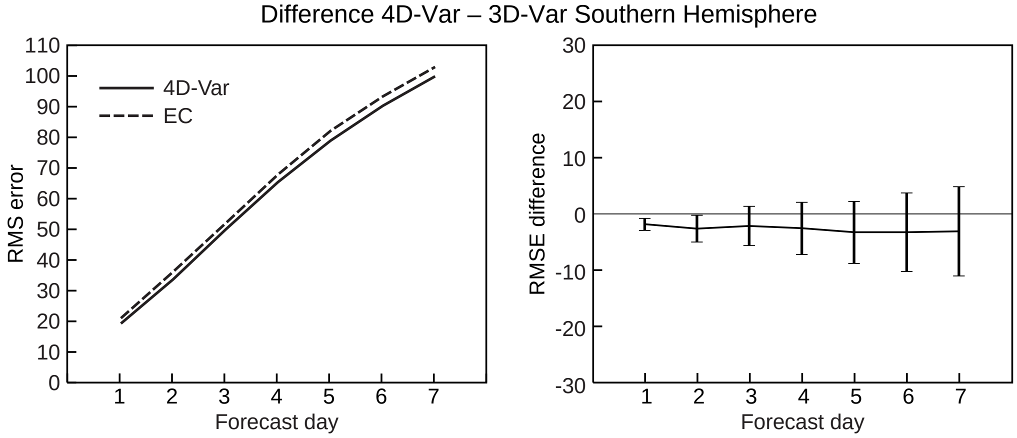


Fig. 1 Simplified view of the 4D-Var assimilation technique for a single parameter x. Over a given time window (six hours here), the observations are compared at their appropriate time with a short- range forecast issued from the previous analysis. The model state at the initial time of the window is then adjusted to achieve a statistically good compromise x_a between the fit J_b to the previous forecast x_b, and the fit J_o to the observations.



Difference 4D-Var – 3D-Var Northern Hemisphere



Difference 4D-Var – 3D-Var Southern Hemisphere

Fig. 2 Root-mean square of the differences between analyses and forecasts of the 500hPa geopotential height from 4D-Var (solid) and from 3D-Var (dashed), averaged over 40 cases of the preoperational suite in the Northern (top) and the Southern (bottom) Hemispheres. The left panels show the scores, the right panels show the differences between the scores with vertical bars depicting the 95% confidence level of the Student t-test.

There are, however, some important new features. Due to the nature of the 4D-Var algorithm, the model evolution during the 6-hour interval, or assimilation ‘window’, is explicitly taken into account. While in 3D-Var the observations were assumed to have been produced at the central time of the interval (except for a tendency correction of the surface pressure observations), in 4D-Var the observations are compared to a model trajectory at 1-hour intervals. This means that the asynoptic observations are compared more precisely with the model forecast, thus improving the timing of meteorological events. There is some potential for using more data than in 3D-Var by allowing the inclusion of more than one report from the same observing station; however, experimentation has shown that temporal observation error correlation must be explicitly taken into account to use the added information correctly.

Because the observations are compared with a sequence of model states over the 6-hour window, the concept of analysis time is blurred in 4D-Var. At the beginning of the minimization, the observations are compared with the previous forecast; at the end, they are compared with a corrected forecast that starts from a modified model state at the beginning at the assimilation window, as shown in Figure 1. Since the corrected forecast is the result of an adjustment to observations distributed over the whole window, the forecast errors do not grow with time as they would with a forecast issued from a static analysis. It can be shown that the forecast errors are at a minimum near the middle of the window, at 12 UTC in our example. For this reason, the 3-hour forecast state is labelled as the ‘official’ atmospheric analysis at 12 UTC. Together with the analyzed surface fields, this is the initial state for the medium-range forecast. This procedure has the added advantage that the change of the assimilation technique is transparent for the dissemination of products to the Member States. When the 4D-Var window is extended to 12 or even 24 hours, the choice of cut-off technique will need to be reassessed.

In 4D-Var, the model state at the beginning of the assimilation window is optimized in order to minimize the distance to the background (defined at the same time) and to the observations; whenever it is modified, the new fit to the observations is calculated by running the model forecast until the time of each observation, interpolating the forecast state to the observation location and variable, weighting the (obs - model) mismatch by the assumed observation errors, and computing the adjoint of this process (i.e. the transpose of its derivatives, with a suitable inner product) in order to obtain the sensitivity of the mismatch with respect to the model state at the beginning of the window. This introduction of the forecast model and its adjoint into the assimilation algorithm introduces flow-dependent structure functions for all observations, except the ones made right at the beginning of the assimilation window, which are used exactly as they would be in 3D-Var. When the flow is unstable, the observations are given more weight; when a wave develops, the analysis increments caused by the observations tend to follow the shape of the wave; when the advection of a meteorological feature (e.g. the humidity field) does not agree with the sequences of observations made in the area, the wind is suitably corrected, even if it is not directly observed. This flow dependence is stronger for observations made near the end of the assimilation window; over a 6-hour window as used now, the design of the background term Jb is still quite important in determining the structure functions in most parts of the analysis.

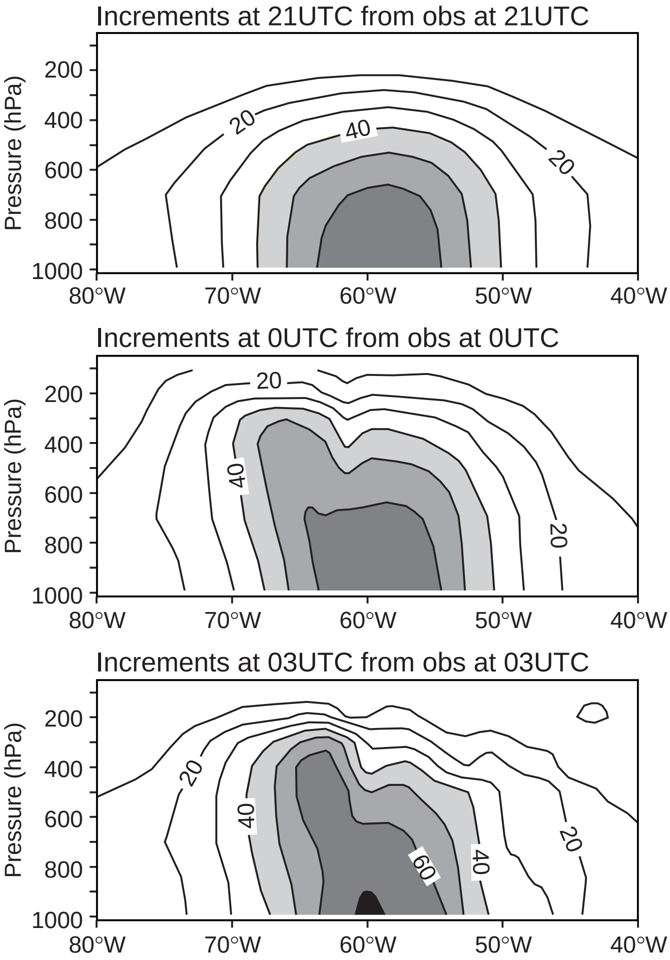


Fig. 3 Corrections to the previous forecast caused by a height observation at 850hPa, (40N,60W). Isolines show the resulting increment, in geopotential units, for an observation at three different times, as explained in the panel captions.

The introduction of the model into the assimilation procedure brings flow-dependence of the structure functions and use of observations at the appropriate time, but it comes at a price. Firstly, since typically 80 evaluations of the variational cost-function and its gradient have to be done to minimize the 4D-Var cost function, the computer cost of about 160 6-hour T63 model forecasts has to be paid on top of the cost of the other software components which already existed in 3D-Var. In practice, the model itself vectorizes and distributes rather well on the Fujitsu computer, and most of the cost actually comes from the management of observations and from the fact that 4D-Var uses the tangent-linear and adjoint versions of the model (adiabatic core, plus simplified physics for part of the minimization). Secondly, the realism of the model used inside 4D-Var is important because it determines how perturbations to the initial state of the model are translated into perturbations in terms of observed parameters. It is limited by the low resolution (T63L31) and by the simplification of the physics; consequently, problems related to those model limitations are likely to affect the use of observations near the end of the assimilation window and near the ground, as well as wherever there are non-linearities, threshold processes or interactions with the water cycle. This could limit the usefulness of 4D Var. For the time being, the 4D-Var incremental procedure is carried out in two low-resolution minimization steps (instead of one in 3D-Var):

- a high-resolution T213 comparison with observations to linearize the problem in the vicinity of the background forecast,
- a long minimization at T63 with minimal linearized physics,
- an update of the high-resolution T213 model with this first set of increments in order to correct the T63 linearization around a state closer to the analysis,
- another short minimization at T63 with a much more complete version of the linearized physics,
- finally, this second set of increments is added to the current high-resolution initial estimate of the initial conditions to produce the final T213 analysis.

The observation quality control procedure has been adapted from 3D- to 4D-Var in order to take into account the use of observation at the appropriate time. This has meant only slight changes to the assumed observation errors. The screening and variational quality control procedures have not been changed, although some improvements are planned.

Impact on operational performance

Extensive pre-operational validation of 4D-Var was carried out before deciding on the operational implementation. The impact of going from 3D-Var to 4D-Var (everything else being kept identical) has been assessed with several configurations in the selection of observations, the use of physics, and the number of high-resolution updates in the incremental formulation. It has led to the choice of the system described above. The impact on some forecast scores is presented in figure 2 (cf. Rabier et al, 1998, for more details). The fit of the background fields to the observations is improved as well, and the short-range spin-up of precipitation is reduced.

The impact might look small, but it is an essential step toward further improvements of the system. Since the implementation, the operational scores have compared very well with other forecast centres, although this is certainly a combination of effects between 4D-Var and previous changes like the revision to the background term Jb (cf. article by F. Bouttier in this newsletter), the extension of the use of TOVS radiances, and the slightly later changes in the physics package, among others.

A principal feature of the 4D-Var is the flow-dependence of the structure functions, already documented by Thépaut et al (1993). Although a 6-hour time window is rather short for it to take its full development, it is already benefitting the assimilation in dynamically active areas, as shown in figure 3 (taken from Rabier et al, 1997).

Future evolution

4D-Var is viewed as an essential tool for future improvements of the assimilation system. First of all, it is a very new assimilation system which needs to be retuned: because the observations are used differently and the forecasts are improved with respect to 3D-Var, the assumed observation and background errors and the rejection thresholds need to be readjusted using objective methods. More data than in 3D-Var can be used now because several reports from the same station can be used, assuming that serial correlation of observation error is properly taken into account. The dynamical initialization problem is different from the one in 3D-Var because some model properties are included in the 4D-Var increments. More efficient incremental strategies can be devised to accelerate the convergence of the 4D-Var variational assimilation problem. All these could be implemented soon at almost no computational cost, but with potentially significant impacts on the forecast quality.

In 4D-Var, the model is naturally integrated within the assimilation. As the forecast model is improved, the 4D-Var configuration will evolve too. It is planned to improve the vertical model resolution in the stratosphere and to include ozone as a prognostic variable. The same changes will be brought into 4D-Var, paving the way for a coupled ozone/wind assimilation in the stratosphere. The realism of the low-resolution model used in the minimization will be improved by the inclusion of better physics and a better horizontal resolution, which are expected to improve the use of data over land as well as observations related to clouds and precipitation. In return this should improve the assimilation and forecast of actual weather parameters.

In the longer term, the full promise of 4D-Var will only be realized if we are able to handle sophisticated flow-dependent structure functions. In other words, the future system should be able to decide automatically to which observations the quality of the ensuing forecast will be sensitive, and to use them in a meteorologically optimal way. The natural way of doing this will be to extend the time window of 4D-Var to 12 and perhaps even 24 hours. In order for the flow dependence to be managed better in the assimilation system, an extension to 4D-Var called Simplified Kalman Filter is being developed, which will introduce flow-dependence into the Jb term. The improved estimates of analysis error it provides will also improve the generation of ensemble members in the EPS.

F. Bouttier and F. Rabier

References

Bouttier, F., J. Derber and M. Fisher, 1997: the 1997 revision of the Jb term in 3D/4D-Var. ECMWF Research Dept. Technical Memorandum no. 238.

Lewis, J. and J. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus,* **37A,** 309-322

Courtier, P. and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Quart. J. Roy. Meteor. Soc.,* **113,** 1329 1347.

Pailleux, J., 1997: 1987-1997, Ten years of research and operational activities with the integrated forecasting system (IFS). ECMWF Newsletter no. 75, Spring 1997.

Rabier, F., J.-F. Mahfouf, M Fisher, H. Järvinen, A. Simmons, E. Andersson, F. Bouttier, P. Courtier, M. Hamrud, J. Haseler, A. Hollingsworth, L. Isaksen, E. Klinker, S. Saarinen, C. Temperton, J.-N. Thépaut, P. Undén and D. Vasiljevic, 1997: The ECMWF operational implementation of four-dimensional variational assimilation. ECMWF Research Dept. Technical Memorandum no. 240.

Talagrand, O. and P. Courtier, 1987 Variational assimilation of meterological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Met. Soc.,* **113,** 1311-1328

Thépaut, J.-N., R. Hoffman and P. Courtier, 1993: Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Wea.Rev.,* **121,** 3393-3414.