DEVELOPMENTS IN ECMWF'S 4D-VAR SYSTEM

E. Andersson[*], C. Cardinali, M. Fisher, E. Hólm, L. Isaksen, Y. Trémolet and A. Hollingsworth
ECMWF, Shinfield Park, Reading, U.K.

1. INTRODUCTION

The operational 4D-Var data assimilation system (Rabier et al. 2000) at the European Centre for Medium-Range Weather Forecasts (ECMWF) is gradually being enhanced for the use of frequent and high-density data and for assimilation of cloud and rain information at relatively high analysis resolution. A more accurate and efficient 4D-Var solution algorithm was implemented in January 2003, and a new humidity analysis (Hólm et al. 2002) was implemented in October 2003.

An important trend in the development of the 4D-Var data assimilation system is the rapid increase in the use of satellite data (Thépaut and Andersson 2003). The additional data contribute significantly towards improved analysis accuracy (Simmons and Hollingsworth 2002). However, very dense or very frequent observations can reduce the rate of convergence of the iterative solution algorithm (Andersson et al. 2000), adding substantial computational load to 4D-Var, or requiring limiting the accuracy of the solution at fixed computational cost. In 4D-Var the main computational expense is related to minimizing an objective cost function. This is currently done at T159 resolution (~120 km) having first compared the observations against short-range forecast fields produced by the operational T511 (40 km) model.

2. THE REVISED 4D-VAR ALGORITHM

In January 2003, with the increasing data volumes in mind, the 4D-Var solution algorithm was comprehensively revised to improve its accuracy and efficiency. Future increases in analysis resolution (to T255 or higher) should be possible with the new algorithm. A schematic is shown in Figure 1. Following an approach implemented at Météo-France (Veersé and Thépaut 1998), the new solution algorithm first solves the 4D-Var problem approximately at low resolution (T95). This provides an accurate preliminary analysis for subsequent calculations at full analysis resolution (T159), thereby reducing the number of computationally expensive iterations that must be performed at full resolution. The low-resolution (T95 and T159) model fields and their evolution over the 12-hour assimilation period are now generated through interpolation of a high-resolution (T511) forecast for improved consistency (Figure 1), (Trémolet 2003).

In addition, information gleaned during the T95 step about the "shape" of the analysis cost function is used to accelerate ("precondition") the T159 minimization (Appendix B of Fisher and Andersson 2001). The iterations are carried out until an objective stopping-criterion is reached, ensuring that the solution is reached to within similar error bounds from day to day in the operational suite. The scheme was further modified to take advantage of the conjugate gradient algorithm, which is an efficient minimizer for this type of application. The revised 4D-Var furthermore generates its estimate of analysis error simultaneously with the actual analysis (rather than in a separate calculation), which also contributes to the efficiency gains.

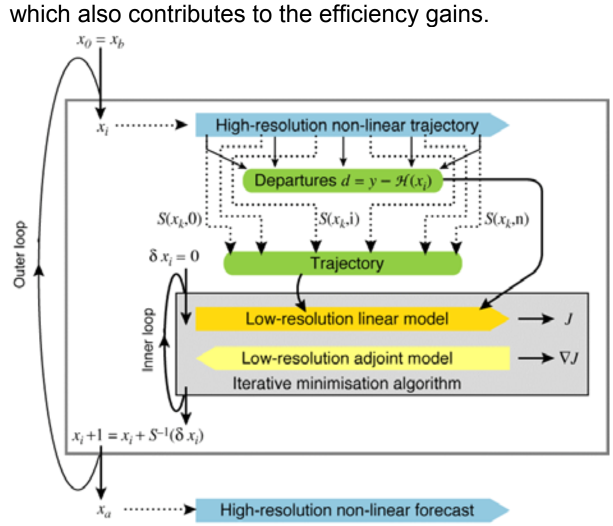

Figure 1 Schematic of the revised 4D-Var solution algorithm implemented in January 2003. Outer loops are performed at high resolution (T511) using the full non-linear model (blue). Inner iterations are performed at lower resolution (first T95, then T159) using the tangent-linear forecast model (yellow), linearized around a 12-hour succession of model states ('the trajectory') obtained through interpolation from high resolution (S denotes the truncation operator, J the cost function and x the atmospheric state vector). From Trémolet (2003).

Figure 2 shows an example of analysis increments calculated at T95 by the first outer-loop iteration (Figure 2a) and the update calculated at T159 (Figure 2b) by the second (final) outer-loop iteration. The example is for temperature at model level 50. It shows that most of the increment is formed at T95, with finer-scale and smaller additions or corrections obtained from the T159 iterations.

[*] *Corresponding author address*: Erik Andersson, ECMWF, Shinfield Park, Reading, RG2 9AX, U.K.; e-mail: erik.Andersson@ecmwf.int.

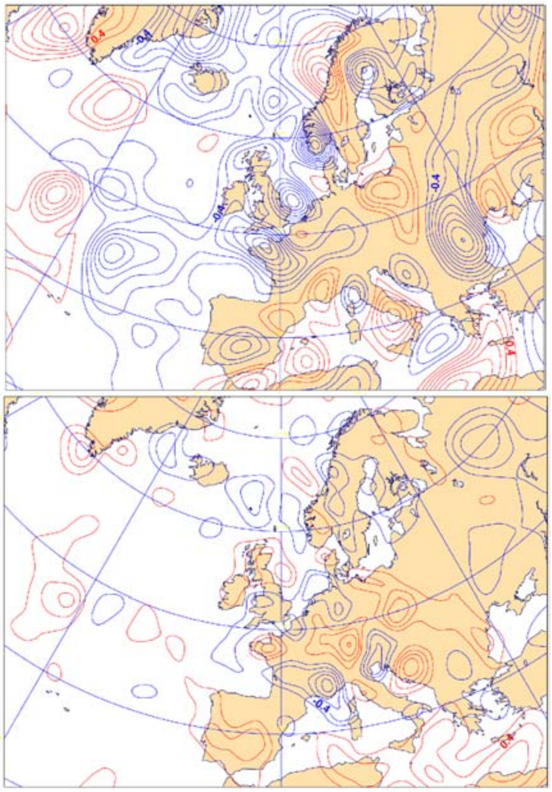

Figure 2 Analysis increments for temperature (0.2 K contours, positive in red and negative in blue) at model level 50 (~880 hPa), 20020925 at 00 UTC, produced by the (a) T95 and (b) T159 iterations of the revised 4D-Var solution algorithm. The total analysis increment is the sum of the two. Most of the increment is formed at the lower resolution with smaller additions and corrections obtained at the higher resolution.

3. THE NEW HUMIDITY ANALYSIS

A new formulation of the humidity analysis has been developed (Hólm et al. 2002). Humidity is in many respects a harder quantity to analyze than wind and temperature, for example. Humidity errors show large variability over short distances and can vary with several orders of magnitude in the vertical. The analysis needs to respect the physical limits due to condensation effects near saturation and the strict limit at zero humidity. A large set of 3-hour forecast differences, generated by an ensemble of data assimilations with randomly perturbed observations (Fisher 2003a), provided ample data for study of short-range forecast error distributions for humidity, and various transformed humidity variables. It was found that when error distributions were stratified according to the value of the relative humidity they were easier to approximate with a Gaussian function. Based on these findings, following the approach suggested by Dee and DaSilva (2003), a normalized relative humidity variable was chosen as the new analysis variable for the revised humidity analysis. The asymmetries in error distributions for conditions near zero humidity and near saturation are also accounted for. The resulting error standard deviations in terms of either specific or relative humidity are thereby strongly dependent on the atmospheric state (Figure 3). The humidity variable transform has been implemented in the background constraint of 4D-Var. It has been verified that the new humidity analysis in a broad sense gives the correct weight to all humidity sensitive radiance data (HIRS, SSMI, Meteosat, GOES, AMSU-B) and also to SYNOP 2m-relative humidity and radiosonde specific humidity data. Data impact studies (OSEs) for these main types of humidity data are now being carried out.

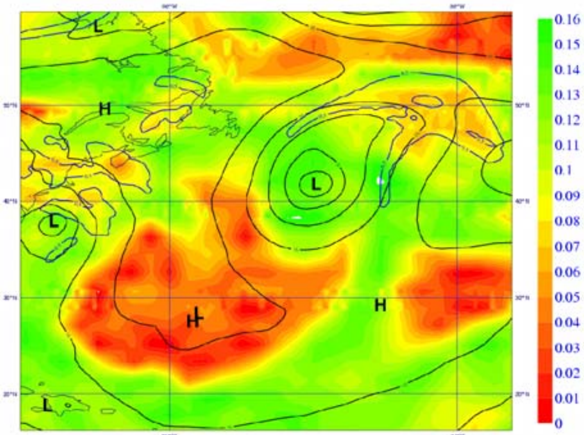

Figure 3 Humidity background error at level 44 (~700 hPa) in terms of relative humidity (shaded, see legend), 20030524-15 UTC. Geopotential at 1000 hPa is contoured in black (4 dm interval) and the blue contour shows cloud fraction = 0.5. The new statistical model for humidity background errors assigns low (orange) relative-humidity error standard-deviations within the cold-air outbreak and within frontal clouds, and high background errors (green) e.g. within the depression.

4. DATA USAGE

Radiance data from several additional spacecraft have been introduced in recent years (Thépaut and Andersson 2003). Figure 4 shows the continuing increase in the number of data used per 3D or 4D-Var assimilation cycle (at 12 UTC) over the last seven years. Current numbers are in excess of 1,600,000 data per cycle – a 20-fold increase compared to the operational 3D-Var system in 1997. With the introduction of AIRS data in October 2003 the total has risen further, to ~3,500,000 data per assimilation cycle.

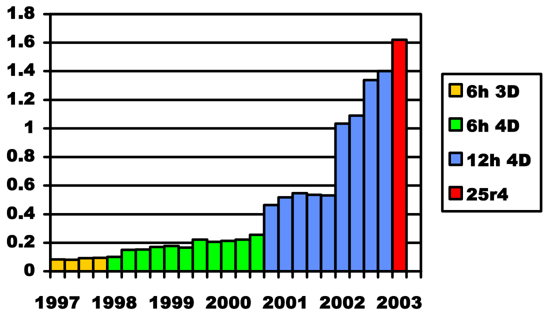

Figure 4 Number of observational data used in ECMWF's operational system per assimilation cycle (millions): 6-hourly 3D-Var (yellow), 6-hourly 4D-Var (green), 12-hourly 4D-Var (blue) and since the operational change (red) implemented 14 January 2003. A further substantial increase (to about 3.5 million) occurred when AIRS data were activated in October 2003.

5. THE BACKGROUND TERM

A new set of background-error statistics has been derived based on an ensemble of 4D-Var assimilations (Fisher and Andersson 2001; Fisher 2003a). The new statistics have smaller amplitude reflecting the improved accuracy of the assimilation system, especially in the stratosphere. The error structures described are generally sharper in both the horizontal and the vertical, which is partly due to the increased availability of observations.

The dynamical balance between wind and mass analysis increments was improved to account for the effects described by the non-linear balance equation and the quasi-geostrophic omega equation (Fisher 2003a). This is likely to be of particular importance in ageostrophic flows near jet-streams (Figure 5) and in the tropics.

6. THE REDUCED-RANK KALMAN FILTER

A possible shortcoming of earlier attempts to formulate a reduced rank Kalman filter (RRKF) was that the subspace used to define the flow-dependent covariance matrix for one cycle of analysis evolved to be nearly orthogonal to the subspace used at the next cycle (Fisher and Andersson 2001). As a consequence, much of the flow-dependent covariance information was thrown away each cycle.

To overcome this problem, a new formulation of the RRKF has been developed, based closely on the "reduced order Kalman filter" (ROKF) described by Farrell and Ioannou (2001a, 2001b). The ROKF applies a technique from robust control theory, which defines an oblique projection onto a subspace that, in a well-defined sense, optimally captures the preferred responses and forcings of the model dynamics. The technique is called balanced truncation, and in recent years has become a well-established element of modern control theory. By using a subspace that represents both the optimal forcings and the optimal responses of the dynamics, the ROKF ensures that the evolved covariance information is retained in the subspace, and is propagated to the next cycle of analysis.

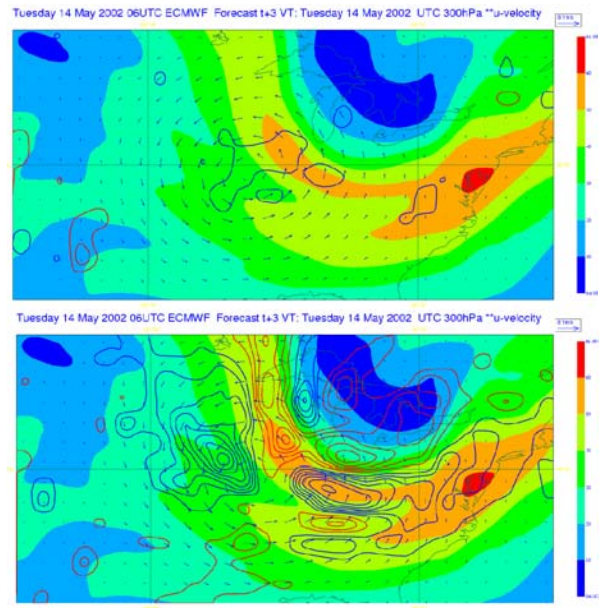

Figure 5 Analysis increments due to a single height observation at 300 hPa near a jet entrance, 20020514-09 UTC. The wind speed is colour-shaded in 10 m/s bands according to the legend (orange is 50-60 m/s). The wind increments are shown as arrows and the contours indicate their divergence (red) and convergence (blue). In this case the single height observation lowers the geopotential (not shown) at the jet entrance resulting in cyclonic wind increments. The revised 4D-Var (lower panel) produces a convergence/divergence pattern consistent with the secondary circulation expected in the region of a jet-entrance, in contrast to the previous version (top). The analysis response to observations can thus be strongly dependent on the atmospheric flow.

Several months of analysis experimentation have been conducted with ROKF in a range of configurations (Fisher et al. 2003). However, the impact on forecast scores has remained stubbornly neutral. It seems likely that the explanation for the lack of impact of the ROKF (and probably for the RRKF too) is that too small a subspace was used. Two pieces of evidence support this explanation. First, the method of balanced truncation provides bounds on the truncation error of the model dynamics as a function of the Hankel singular values corresponding to the neglected directions in phase space. By extrapolating the rate of decay of the leading Hankel singular values, it is possible to give a rough estimate of the truncation error. The calculation suggests that at least a few thousand vectors would be required to give an accurate truncation of the model dynamics.

The second piece of evidence comes from analysis of a large sample of forecast differences taken from the ERA-40 reanalysis. Differences between 36h and 12h forecasts verifying at the same time were used. The differences were truncated to a spectral resolution of T42. It was found that a projection of one forecast difference onto the preceding 2500 forecast differences was only around 50%. This shows that the dimension of subspace required to produce a good approximation to the static covariance matrix of forecast differences (a popular surrogate for background error) is at least few thousand.

Subspace dimensions of thousands of vectors are not computationally feasible with current computers, since they require thousands of integrations of the tangent linear model at every analysis cycle. A glimmer of hope is provided by the fact that when the analysis of forecast differences was repeated using differences that were restricted to a limited horizontal domain of 1000 km x 1000 km, the projection of a given forecast difference onto the preceding 200 vectors was around 90%. It appears from this finding, that some form of localization of the background co-variances is necessary. It is not clear how to do this for a deterministic subspace based on singular vectors. However, we note that such localization techniques are an important component of a number of implementations of the ensemble Kalman filter (EnKF).

7. DIAGNOSTICS OF THE 4D-VAR SYSTEM

Different types of diagnostic can be obtained from the 'influence matrix' (Cardinali et al. 2003), which in statistical regression models is used to measure the leverage of the observations. The diagonal elements of the influence matrix represent the analysis sensitivity to the observations, and are called 'self-sensitivities'. Self-sensitivities for SYNOP surface pressure observations are shown in Figure 6. Each box indicates the observation influence at the observation location. Low-influence data points have large background influence, which is the case in data-rich areas such as North America and Europe (observation influence ~ 0.2). In data-sparse areas observations have larger influence: in the Polar regions, where there are only few isolated observations, self-sensitivity is close to 1 and the background has small influence in the analysis.

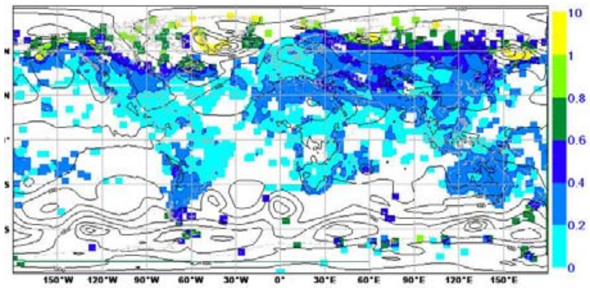

Figure 6 Observation Influence, (or self-sensitivity, Cardinali et al. 2003) of surface pressure SYNOP data valid for 13 February 200312 UTC, in ECMWF's 4D-Var. The influence of individual data is large in data-sparse regions, and for isolated stations.

Recently, the diagonal elements of the Influence matrix has been used to highlight areas that may contribute to poor convergence in the analysis, known also as the ill-conditioning problem. It has also been shown that the trace of the influence matrix is a measure of the information content extracted from the observations by the analysis scheme (Figure 7). This trace is sometimes referred to as degrees of freedom for signal (DFS). It should be noted that this measure of information content does not necessarily correspond to forecast impact. A large part of the information in HIRS and SSMI (Figure 7) is with respect to humidity, and parts of the AMSU-A information is in stratospheric temperature.

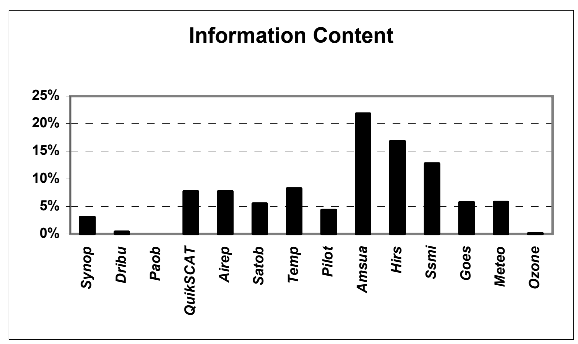

Figure 7 Information content i.e. degree of freedom for signal for the main data types in the assimilation. From Cardinali et al. (2003).

Another method that allows accurate global calculation of DFS for the complete 4D-Var system has been developed by Fisher (2003b). For example the diagram shown in Figure 8 indicates that more information is extracted from observations with increasing analysis resolution. In particular, an increase of inner-loop resolution from T159 (current) to T255 (planned for 2004/05) would significantly increase DFS of 4D-Var. The calculations are based on current background error statistics and current used observations (January 2003).

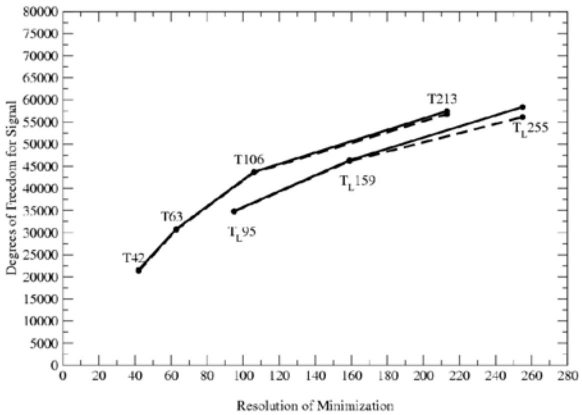

Figure 8 Degrees of freedom for signal (DFS) is a measure of information content in the analysis. A newly developed method (Fisher 2003b) allows efficient calculation of DFS for the global 4D-Var system. The diagram shows that the information extracted from available data increases as the inner-loop resolution of the analysis increases. An increase beyond the current operational resolution (T159) to T255 or higher, would allow additional information to be extracted from the available observations.

8. EVOLUTION OF FORECAST ERRORS

Figure 9 provides indication of the link between reducing forecast error in the very short range (and by implication reducing analysis error) and reducing error in the medium range. It shows smoothed time series of the ranges at which ECMWF's operational 500 hPa height forecasts have reached certain levels of anomaly correlation over the past decades (Simmons 2003; Simmons and Hollingsworth 2002). In recent years the improvement of forecasts as measured by the increase in these forecast ranges has been by an amount that varies little beyond a day or so ahead. Improvements in medium-range 500 hPa height forecasts thus appear to have stemmed directly from model, analysis and observing-system improvements that have reduced analysis and short-range forecast error.

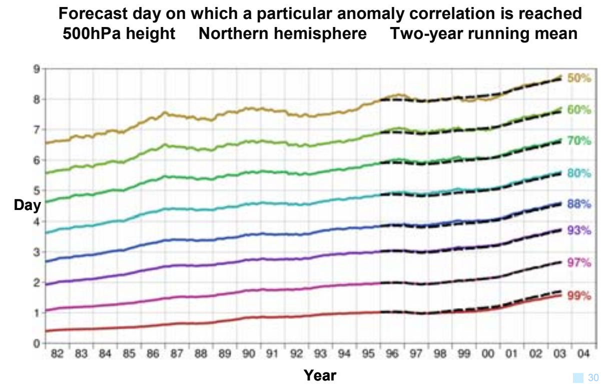

Figure 9 The forecast ranges at which the monthly-mean anomaly correlations of 500 hPa height forecasts for the extratropical northern hemisphere reach a set of values form 99% to 50%, based on ECMWF operational forecasts from January 1980 to July 2003. The colored curves denote actual values plotted as two-year running means, each point on the graph representing the average over the preceding two years. The black dashed curves show displacements of the curve for the 97% level from January 1996 onwards. The curves are displaced so as to match actual values for each percentage level for January 1996. (From Simmons 2003).

REFERENCES

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

Cardinali, C., S. Pezzulli and E. Andersson, 2003: Influence matrix diagnostic of a data assimilation system. Proc. ECMWF Seminar on "Recent Developments in Data Assimilation for Atmosphere and Ocean", 8-12 Sept 2003, Reading, U.K. *In press.* http://www.ecmwf.int/newsevents/meetings/annual_seminar/index.html

Dee, D. P. and A. M. Da Silva, 2003: The choice of variable for atmospheric moisture analysis. *Mon. Wea. Rev.*, **131,** 155-171.

Farrell, B. F. and P. J. Ioannu, 2001a: Accurate low dimensional approximation of the linear dynamics of fluid flow. *J. Atmos. Sci.*, **58,** 2771—2789.

Farrell, B. F. and P. J. Ioannu, 2001b: State estimation using a reduced order Kalman filter. *J. Atmos. Sci.,* **58,** 3666—3680.

Fisher, M. and E. Andersson, 2001: Developments in 4D-Var and Kalman Filtering. *ECMWF Tech Memo*, **347,** pp 36. http://www.ecmwf.int/publications/library

Fisher, M., 2003a: Background error covariance modeling. Proc. ECMWF Seminar on "Recent Developments in Data Assimilation for Atmosphere and Ocean", 8-12 Sept 2003, Reading, U.K. *In press.*

http://www.ecmwf.int/newsevents/meetings/annual_seminar/index.html

Fisher, M., 2003b: Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *ECMWF Tech Memo*, **397,** pp 18. http://www.ecmwf.int/publications/library

Fisher, M., L. Isaksen, M. Ehrendorfer, A. Beck and E. Andersson, 2003: A critical evaluation of the reduced-rank Kalman filter (RRKF) approach to flow-dependent cycling of background error covariances. *ECMWF Tech. Memo. In Press.* http://www.ecmwf.int/publications/library

Hólm, E., E. Andersson, A. Beljaars, P. Lopez, J-F. Mahfouf, A. J. Simmons and J.-N. Thépaut, 2002: Assimilation and modeling of the hydrological cycle: ECMWF's status and plans. *ECMWF Tech. Memo*., **383,** pp 55. http://www.ecmwf.int/publications/library

Rabier, F., H. Järvinen, E. Klinker, J.F. Mahfouf and A. Simmons, 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.

Simmons, A. J. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.,* **128,** 647—678.

Simmons, A., 2003: Observations, assimilation and the improvement of global weather prediction – Some results from operational forecasting and ERA-40. Proc. ECMWF Seminar on "Recent Developments in Data Assimilation for Atmosphere and Ocean", 8-12 Sept 2003, Reading, U.K. *In press.* http://www.ecmwf.int/newsevents/meetings/annual_seminar/index.html

Thépaut, J.N. and E. Andersson, 2003: Assimilation of high-resolution satellite data. *ECMWF Newsletter*, **97,** 6—12. http://www.ecmwf.int/publications/newsletters

Trémolet, Y., 2003: Diagnostics of linear and incremental approximations in 4D-Var. *Q. J. R. Meteorol. Soc.*, Accepted.

Veersé, F. and J.-N. Thépaut, 1998: Multiple-truncation incremental approach for four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.* **124,** 1889—1908.