Introduction

The three-day workshop on “Flow-dependent aspects of data assimilation” took place from 11th to 13th of June 2007.

The aim of the workshop was to bring together European and American experts to discuss ways in which the ECMWF analysis could be made more responsive to observations in those particular locations where significant and potentially severe weather is developing. These locations vary from day to day depending on the 'flow', i.e. the atmospheric conditions. It is where the model uncertainty is large that good observations can provide the most beneficial corrections to the model fields through flow-dependent data assimilation.

Ensemble-based methods are well suited to detect and track these areas of higher observation significance as they evolve and develop. The working group participants discussed the techniques for running ensembles of assimilations: how the various sources of error could be represented and accounted for. The workshop recommendations will influence the ECMWF research plans in this area, suggesting the avenues to explore and the priorities for the next couple of years.

The workshop followed the usual format of invited lecturers followed by discussions in working groups and it concluded with a plenary session. Groups were set up to consider the subjects of “Variable transformations”, and the “Principles” and “Implementation issues” of using ensemble-based covariances in 4D-Var. The discussions and recommendations of the three working groups are summarized in the following three reports. The contributions to the workshop have been posted on the ECMWF web site www.ecmwf.int/publications.

ECMWF thank all the participants for contributing to a successful and stimulating workshop.

1. Working Group Report: Variable transformations

Ian Roulston (Chair), Elías Hólm (Secretary), Erik Andersson, Ross Bannister, Luc Fillion, Marta Janisková, Philippe Lopez.

1.1. Coordinate transformations

The working group discussed the possibility of using coordinate transformations for dealing with gradients (e. g. fronts and the tropopause) and the associated asymmetric correlations. Previous work on the use of geostrophic coordinate transforms was mostly dealt with in the horizontal, and difficulties were reported where errors in the background field cause gradients to be misplaced with respect to observations.

> Recommendation: The working group recommended that the fully three-dimensional geostrophic coordinate transform should be investigated at ECMWF. The vertical coordinate in this system is potential temperature, and to accommodate for the boundaries, a hybrid coordinate should be used. The transform can preferably be implemented in terms of a metric instead to avoid numerical difficulties.

1.2. Diabatic balance relationships

The working group noted the effective way of including diabatic terms in the assimilation by normal mode initialization or equivalent approaches, presented by Luc Fillion during the workshop. Here the tendency of the tangent linear model is used, eliminating the need to explicitly program diabatic balances while at the same time increasing the accuracy and consistency of the balance relationships with the model.

> Recommendation: The working group recommended that balance relationships making use of the full tangent linear model, along the line of normal mode initializations applied to the analysis increments, should be tried out.

1.3. Vertical covariances

The working group discussed the importance of correctly specified vertical correlation, in particular in regions of strong gradients like at the boundary layer top and even clouds. The analysis usually spreads increments above the boundary layer or outside clouds due to the averaged nature of the correlations used.

> Recommendation: It is recommended that the possibility of flow dependent vertical correlations be investigated to reduce smoothing of vertical gradients and to change the position, introduce, or eliminate sharp transitions in the vertical.

2. Working Group Report: Ensemble-Based covariances in 4D-Var: Principles

Andrew Lorenc (Chair), Mike Fisher (Secretary), Harald Anlauf, Sue Ballard, Mark Beuhner, Aimé Fournier, Bruce Ingleby, Gert-Jan Marseille, Jim Purser, Istvan Szunyogh, Yannick Trémolet.

2.1. Theoretical basis

The working group noted that variational assimilation attempts to estimate the mean of the analysis pdf, whereas ensemble-based assimilation methods (such as the ensemble Kalman filter) attempt to sample it. There may be fundamental differences in character between the ensemble members and the variational analysis. The mean may not be a likely state of the ensemble.

> Recommendation: The working group recommended that systematic differences between the ensemble members and the control analysis should be investigated. For example, does the use of stochastic physics in the ensemble members make them less balanced than the control?

An ensemble of variational analyses has a connection with the idea of Gaussian mixture models (in which a pdf is approximated as a sum of Gaussians).

> Recommendation: ECMWF should consider this connection, and study the literature on Gaussian mixture models.

2.2. What should be expected, and how should impact be assessed?

The working group discussed the apparent lack of impact of attempts to incorporate more flow-dependence into the ECMWF analysis system. However, it felt that large-scale mean scores may not be the appropriate measure to demonstrate impact. It also noted that comparisons between ensemble Kalman filters and less flow-dependent techniques suggest that the impact of flow dependence may be reduced by the presence of large numbers of satellite-based observations. It may also be the case that flow-dependence of vertical correlations (which has not so far been addressed by ECMWF) are important in allowing better use of satellite-based observations.

> Recommendations:

- A wider range of measures should be studied to determine the impact of flow-dependent modifications to the analysis system. Feature-based verification should also be used.
- The flow-dependence of vertical covariance should be studied.
- The possibility of introducing flow-dependent covariances through the use of vertical coordinate transforms (or equivalent metric-based methods) should be investigated.
- Overall, the working group felt that more use of observation-based verification statistics would be welcome

2.3. Ensemble size

Estimates of the standard deviation of background error (and other statistics) derived from a small ensemble are noisy.

> Recommendation: The possibility of smoothing ensemble-derived statistics should be investigated, as a means to reduce sampling noise.

2.4. Model error

Forecasts initialized from current ensembles of analyses have insufficient spread, and perturbations grow too slowly.

Recommendations:

- OSSE-like perfect-model studies should be conducted to investigate the lack of spread of the analysis ensemble, and the slow growth of spread in forecasts initialized from the ensemble
- Current attempts to represent the covariance matrix of model error (Q) lack any flow-dependence.
- The possibility of introducing a degree of flow dependence into the Q matrix should be investigated. For example, the formulation of the stochastic backscatter scheme may be helpful in deciding where model error variances are large.
- Statistics of analysis increments should be studied to try to diagnose model bias. The random component of model error might be estimated from the statistics of short (e.g. 6h) forecasts from identical initial states using a model that incorporates stochastic backscatter.
- The impact of stochastic physics on the rate of growth of ensemble spread should be studied by comparing the observation-minus-forecast statistics for the control forecast with those of the ensemble members.

2.5. Miscellaneous

The working group discussed the possibility of ECMWF developing an ensemble Kalman filter. Although there are some advantages in this idea, the group recognised that this would entail significant effort, and that there is some merit in ECMWF taking a somewhat different approach to that of other centres.

Recommendations:

- ECMWF should continue to develop its existing ensemble of 4D-Var analyses.
- The working group felt it important to retain a static component to the background error statistics.

The working group felt that correlations between ensemble spread and diagnostics of the background field (e.g. Eady index) could be exploited to allow a considerable degree of flow-dependence to be incorporated into the analysis, independently of an analysis ensemble.

Recommendation: The correlation between ensemble spread and the diagnostics of the background field should be studied.

3. Working Group Report
Ensemble-Based covariances in 4D-Var: Implementation issues

Jeff Whitaker (Chair), Lars Isaksen (Secretary), Martin Leutbecher, Loïk Berre, Andreas Rhodin, Peter Steinle, Gerald Desroziers, Sue Ballard, Magnus Lindskog, Marek Wlasak

3.1. The current implementation

The current ensemble data assimilation (DA) system implemented at ECMWF consists of a 10-member ensemble of 4D-Var DA cycles with perturbed observations, SST and soil moisture. The experimental stochastic backscatter model error parameterization scheme is used in the trajectory job and the first-guess forecasts. The 10-member ensemble of first-guess forecasts is used to construct a background error variance (sigma-b) estimate. This variance estimate is multiplied by a factor of 2.25 - 4 (square of the so-called 'REDNMC' factor) and is then used in a separate 4D-Var analysis without perturbed observations, SST or soil moisture. The sigma-b estimate is not used within the 10 member ensemble of 4D-Var analyses. The sigma-b estimate is also used in the quality control background check for the parallel, unperturbed 4D-Var system. The impact of flow-dependent background error variances is neutral for hemispheric averaged 500 hPa forecast skill metrics, averaged over 20 - 30 cases. This result is consistent with ensemble Kalman filter (EnKF) results presented at the workshop. The use of the ensemble perturbations to initialize the EPS system instead of the evolved singular vectors results in a small improvement in EPS performance in mid-latitudes, and a significant improvement in the tropics.

It is important to remember that 4D-Var already has flow-dependent background error covariances (B) towards the end of the 12 hour assimilation window (but not at the beginning of the window, except from effects of non-linear balances implemented in the background constraint). The ensemble-based estimate of sigma-b introduces flow dependence at the beginning of the assimilation window.

3.2. The fundamental question

Why is the apparent impact of flow-dependent background errors so small in the operational 4D-Var?

3.3. Hypothesis and workshop recommendations

3.3.1. Time and space averaged 500 hPa forecast skill is not a sensitive metric to detect the impact of flow-dependence in the 4D-Var background error covariance (B).

Recommendation: Since flow-dependent background errors should be most helpful when the error in the background forecast is extreme (very different than that implied by the operational B), metrics which target these situations should be used. If the ensemble estimate of B correctly captures these changes, the observations should be used more effectively in the DA system and the analyses (and subsequent forecasts) should be more accurate on average. The tails of the short-term forecast pdf should be examined to see if the probability of extreme forecast 'busts' is reduced. EPS diagnostics which measure the consistency between ensemble spread and ensemble mean error should also be used. Since the largest impact is likely to be on short range forecasts, observation-based verification metrics should be employed.

3.3.2. There are so many observations now being assimilated (seven million per day) that the initial B used in the 4D-Var just doesn't matter much.

Recommendations: Perform experiments with a reduced observing network (only the 'conventional' data), where flow-dependent B estimates have been shown to have a larger impact in EnKF systems. Compare the resulting B estimates with parallel runs of EnKF systems at other centres to see if the estimates are model and/or system dependent. Such experiments should demonstrate whether the ECMWF ensemble DA implementation performs consistently with other EnKF-based systems, in situations when flow-dependent B estimates are known to be important.

3.3.3. The degree of optimality of the flow-dependent variance estimates may be further studied.

Recommendations:

a. Develop validation tools utilizing innovation statistics to evaluate B. Is the ensemble system producing the expected B? Are the sigma-b estimates correctly identifying situations where background forecast error is large or small?

b. Apply filtering (smoothing) of variances to reduce sampling error.

c. Use flow-dependent variance estimate within the evolving perturbed 4D-Var DA ensemble. This may result in a different, and possibly better (closer to optimal in the Kalman filter sense) B estimate (see section 1.3.4 also).

3.3.4. Flow-dependent changes in the structure of background correlations are as important as changing variances.

Recommendations (ordered roughly in terms of difficulty of implementation)

a. Use the existing wavelet analysis tools to derive flow-dependent horizontal and vertical covariance structures from ensemble (promising results from Météo-France, DWD, and the ALADIN/HIRLAM group using this approach were presented at the workshop).

b. Also examine use of the ‘alpha control vector’ approach to integrate ensemble information more directly in 4D-Var. Environment Canada, the NCAR/WRF DA group, and the Met Office are testing this approach. However, setting the parameters of B directly (as is currently being done with the sigma-b estimate) is more straightforward, should be examined first (as described in item (a)).

c. Use flow-dependent B estimate within the evolving perturbed 4D-Var DA ensemble. This may result in a different, and possibly better (closer to optimal in the Kalman filter sense) B estimate. However, more ensemble members than 10 may be needed, and missing error sources must be treated carefully to prevent filter divergence (see section 3.5).

d. Since step (d) brings the system closer (in spirit) to an EnKF, it would help to bring a visiting scientist to ECMWF to implement a parallel EnKF system. This would enable direct comparison of background error estimates, and would aid in understanding the relative strengths and weaknesses of the two approaches. The resulting EnKF system may also be useful for future reanalysis efforts (since EnKF systems have certain advantages over 4D-Var for reanalysis, including the ability to adapt to changing observing networks, and their ability to provide direct estimates of analysis uncertainty). The working group estimated

that, due to the simplicity of EnKF algorithms, a one-year visit should be long enough to implement an EnKF system and perform preliminary experiments.

3.3.5. Missing error sources, not sampled by the DA ensemble, are limiting the accuracy of the B estimates. The need to heavily inflate the sigma-b variance estimates may be a symptom of this problem.

Recommendations:

a. Investigate the use of adaptive covariance/variance inflation using innovation statistics (instead of applying a fixed global constant).

b. Use correlated observation error perturbations for satellite radiances (but not necessarily in the DA itself).

c. Consider adding forcing singular vectors to DA ensemble perturbations (since singular vectors appear to be doing a good job in accounting for growth of spread associated with model errors in the EPS system).

d. Continue to develop stochastic schemes to account for flow-dependent errors. Possible candidates include

- Utilizing the stochastic (Monte-Carlo) cloud properties in radiation scheme to increase spread by using different random numbers in each ensemble member.
- Further development of the Stochastic Backscatter scheme
- Develop a stochastic radiative transfer algorithm to account for errors in forward operator used for satellite radiances.

3.4. Concluding remarks

Despite the fact that initial experiments have not shown a large impact on forecast skill, the working group strongly encourages further development of ensemble DA systems at the ECMWF.