

Discussion on ‘4D-Var or EnKF?’

By NILS GUSTAFSSON*, Swedish Meteorological and Hydrological Institute, SE-60176 Norrköping, Sweden

1. Introduction

The development of data assimilation techniques for numerical weather prediction has been very successful ever since the early 1950s until now, starting with simple two-dimensional and univariate spatial interpolation techniques like the successive corrections (SC, Bergthorsson and Döös, 1955) ending up with the four-dimensional variational data assimilation (4D-Var, Rabier et al., 2000) and ensemble Kalman filter (EnKF, Evensen, 1994) techniques of today. Looking a bit closer into the steps of this development, one may see a gradual and continuous development. Already SC schemes were based on the idea of data assimilation, that is, they treated the deviations between observations and a model background field in the spatial interpolation process. SC schemes were generally optimized on statistics of observation minus background data, and included also multivariate relationships. With the introduction of Optimum Interpolation (OI, Eliassen, 1954; Gandin, 1963), both of these aspects of data assimilation were handled more rigorously. An important next step was the generalization of OI to three spatial dimensions (Lorenç, 1981), and after that the step to three-dimensional variational data assimilation (3D-Var, Parrish and Derber, 1992) was not big. Adding the time-development of the assimilation increments over the data assimilation window, we arrive at 4D-Var.

The idea of gradual development should in my opinion be applied in the ongoing discussion on 4D-Var and EnKF. Both methods try to address non-linearities and the errors of the day through an implicit (4D-Var) or an explicit (EnKF) description of flow-dependent forecast error structures. On one hand, 4D-Var in its present strong constraint formulation form is limited to development of flow-dependency over a rather short data assimilation window. On the other hand, EnKF applies a more general flow-dependency from an ensemble of assimilation background states, but is limited due to the small number of ensemble members, which makes great care in the utilization of the derived error covariance structures necessary. In contrast, 4D-Var applies very robust covariance structures, derived as long-term averages, at the start of the assimilation window. Taking these two fundamental and complimentary characteristics of 4D-Var and EnKF

into account, it seems to me more appropriate not to ask the question *4D-Var or EnKF?* but rather *How can ideas from EnKF and 3D-Var or 4D-Var best be combined?*

It is important to note here that 4D-Var will probably not be able to handle strong non-linearities, and we have started to experience this for mesoscale applications of 4D-Var in the handling of, for example, convection and moist physical processes. Following ECMWF staff members we may ask Will non-linearities defeat 4D-Var? As discussed in some depth by Kalnay et al. (2007), EnKF techniques also have limitations in the treatment of non-linearities and the associated non-Gaussian probability distributions.

2. General views on Kalnay et al. (2007)

The Kalnay et al. (2007) paper is well written, it presents some new results on the relative merits of 4D-Var and EnKF, it illustrates the great competence developed by several university groups in EnKF and it provides a valuable discussion needed to support the decisions regarding the development of future data assimilation systems for operational numerical weather prediction. I would like to make the following more general comments on the paper (my statements may be a bit biased in favour of 4D-Var, but this merely reflects the opposite tendency expressed by Kalnay et al., 2007):

(1) In order to answer the question *4D-Var or EnKF?* raised in the title of the Kalnay et al. (2007) paper, we need access to the results of a full-scale test of EnKF in an operational environment utilizing all types of observations, including satellite radiances. Such a full-scale test of EnKF is not available yet, only 4D-Var has reached this advanced state of development. Therefore, when it comes to application of EnKF to advanced primitive equation models, as applied operationally, the paper becomes a bit too speculative with reference to second-hand statements only and with few new results. A problem in this connection is that the competence in EnKF mainly is concentrated to university institutions without the infrastructure to carry out full-scale data assimilation tests (the Canadian Weather Service being an exception). It must be the responsibility of operational weather services to carry out such full-scale tests of EnKF.

(2) The sampling errors associated with the limited (~100) number of ensemble members in EnKF is indeed a crucial problem. Kalnay et al. (2007) discuss this in section 3.2

*Correspondence.
e-mail: nils.gustafsson@smhi.se
DOI: 10.1111/j.1600-0870.2007.00262.x

'*Observation localization*'. In my opinion the discussion is not fully complete. What happens for example with model balances in case of the local ensemble Kalman filter (LEnKF)? Correlations between wind and temperature increments, for example, generally have their maximum values at some distance. What happens with the balances if all correlations are multiplied by a certain distance-dependent factor, and, what happens with the balances over scales larger than the data selection 'boxes' in the LEnKF?

(3) I find the general statement about adding noise before the model integration in section 4.2 very interesting. It is explained that doing so will allow *the ensemble to explore unstable directions that lie outside the analysis subspace and thus to overcome the tendency of the unperturbed ensemble to collapse towards the dominant unstable directions already included in the ensemble*. I interpret this in the way that the ensemble members in EnKF do contain too limited and restricted information. In 3D-Var and 4D-Var, the B matrix in principle allows variability in all directions, stable and unstable. This could be taken as evidence that we should look for an optimal synthesis of 4D-Var and EnKF.

(4) It is argued that the advantages of 4D-Var with long windows disappear if the model is imperfect or if the adjoint model is not exact. In my view such statements should be made more relativistic. In the first case it must depend on the degree of imperfectness and on how model errors (Tremolet, 2005) can be handled. Also with regard to the non-exact adjoint, it must depend on the degree of approximations applied in the adjoint model. One example is the High Resolution Limited Area Modelling (HIRLAM) 4D-Var (Huang et al., 2002), where the non-linear model is based on a finite difference representation, while the tangent linear and the adjoint models are based on a spectral representation. These seemingly huge model differences do not seem to affect the good results of HIRLAM 4D-Var.

3. Views on the summarizing table 7

Concerning the table at the end of the Kalnay et al. (2007) paper summarizing advantages and disadvantages of 4D-Var and EnKF (table 7), I have the following comments:

(1) I do not agree that EnKF is simple to design and code, in particular if the need for covariance localization, or localization of the filter and the data selection, is taken into account. A global data selection and application of a global covariance matrix was one of the main advantages when 3D-Var and 4D-Var were introduced, in contrast with previous OI schemes with complicated local or regional data selection. In particular, balances were improved with the global data selection (Gustafsson et al., 2001), and I see no reason why this should not apply also to EnKF techniques.

(2) The disadvantage of 4D-Var that tangent-linear and adjoint models have to be developed, and the corresponding advantage of EnKF are often mentioned. Taking the experience

with HIRLAM 4D-Var into account, this argument is certainly valid. It took us 10 yr to develop HIRLAM 4D-Var, and a significant part of these 10 yr was waiting for errors in tangent linear and adjoint models to be detected and corrected. However, there now exist automatic pre-processor techniques that can be used in the derivation of tangent-linear and adjoint models (Giering and Kaminski, 2003). This makes this disadvantage of 4D-Var less obvious.

(3) The potential of rain assimilation has been, at least partially, proven for 4D-Var (Mahfouf et al., 2005). Yet it is not clear, however, whether rain should be assimilated directly or whether one needs a pre-processing to water vapour. The assimilation of rain in EnKF is conceptually straightforward, but the concept still has to be proven in practise.

(4) In 4D-Var, weak digital filters can be included quite easily and without cost (Gauthier and Thépaut, 2001). In EnKF, digital filters probably have to be applied as strong constraints, with all the difficulties associated with the centring of digital filters and the possible need for backward model integrations.

(5) One advantage of 4D-Var is indeed the ability to handle flow-dependencies developing during the time window of the assimilation. In important cases of storm developments, for example, these flow-dependencies develop quite rapidly and should be advantageous also with relatively short assimilation windows (6–12 hr).

4. Example of flow-dependency in 4D-Var

In order to put some substance into the discussion, I will show one example that illustrates the potential of implicit treatment of flow-dependency in 4D-Var, also with short assimilation windows (6 hr). The example is taken from application of the HIRLAM 4D-Var to the mesoscale storm that hit Denmark on the 3 December 1999. A *single simulated observation experiment* was carried out. The question asked was: What would 3D-Var and 4D-Var do in case we had a single surface pressure observation available, telling us that the surface pressure in the centre of the storm ought to be 5 hPa deeper?

A simulated surface pressure observation with a -5 hPa observation increment was thus inserted at 3 December 1999 12 UTC into the data assimilation in the position 57N 3E, in the centre of the storm. Figure 1 shows the surface pressure assimilation increment in case HIRLAM 3D-Var (Gustafsson et al., 2001) is applied. The assimilation increments simply reflect the homogenous and isotropic 3D-Var structure functions on a large spatial scale, reflecting average surface pressure forecast errors.

To illustrate the effect of the implicit treatment of flow-dependency in 4D-Var, the same simulated surface pressure observation was inserted into a 4D-Var assimilation, with the assimilation window starting at 3 December 1999 06 UTC and ending at 12 UTC. Figure 2 shows the 4D-Var surface pressure assimilation increments valid at 12 UTC. We can notice that the spatial scale of the surface pressure increments has shrunk

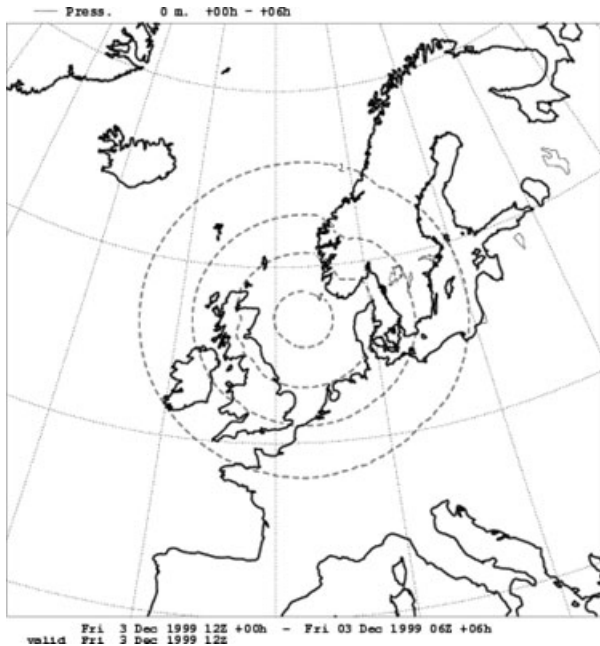


Fig. 1. 3D-Var mean sea level pressure assimilation increments from a single simulated surface pressure observation in the position 57N 3E at 3 December 1999 12 UTC.

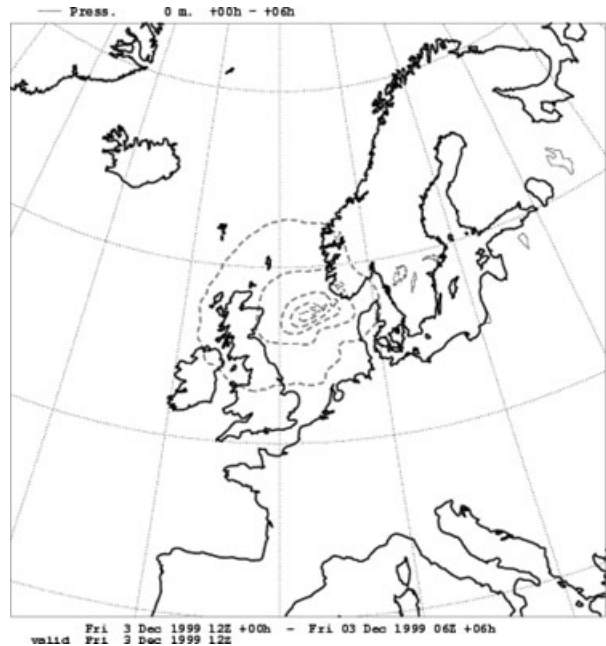


Fig. 2. 4D-Var mean sea level pressure assimilation increments from a single simulated surface pressure observation in the position 57N 3E at 3 December 1999 12 UTC. 4D-Var was applied over the assimilation window 3 December 1999 06–12 UTC.

significantly, roughly corresponding to the scale of the storm itself, as given by observations. Furthermore, the increments no longer have the isotropic horizontal structure of typical 3D-Var increments.

How are these flow-dependent assimilation increments achieved in 4D-Var? First of all, it can be shown that 4D-Var is equivalent to a full rank extended Kalman filter (EKF) over the data assimilation window, with the important limitation that the covariance structures at the start of the assimilation window are just the static (and robust) ones applied in 3D-Var. From this we can conclude that it is the application of the tangent linear model, linearized around a trajectory calculated by the full non-linear model, that provides the flow-dependency 6 hr later into the data assimilation window. The assimilation increments at the start of the assimilation window were also investigated. The 4D-Var surface pressure increments at 3 December 06 UTC turned out to be very small (not shown). Figure 3 shows a NW–SE cross-section of upper-air temperature and wind increments, centred at 55N 0EW in an area upstream of the storm development 6 hr later. We may notice an increase of the vertical wind shear as well as a slight vertical tilt in the assimilation increments at the start of assimilation window. Thus we can simply conclude that 4D-Var manages to intensify the storm development by increasing the degree of baroclinicity in the model state at the start of the assimilation window, and this provides a faster growth of the storm during the tangent linear propagation of the assimilation

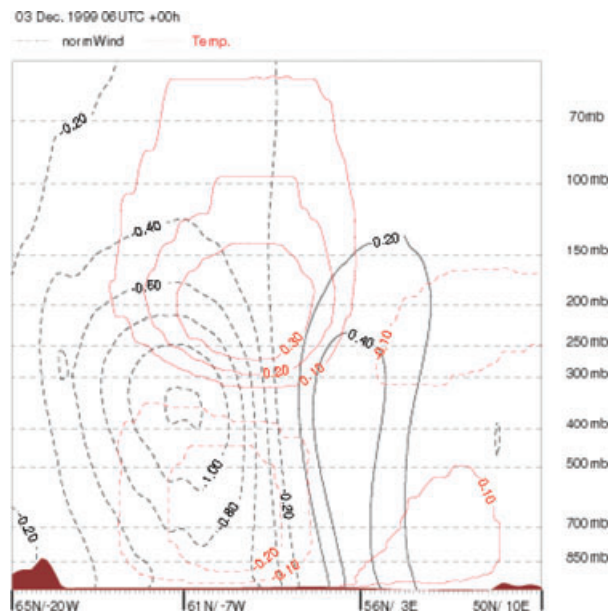


Fig. 3. Vertical cross-section in the NW–SE direction, centred at 55N 0EW, showing 4D-Var upper-air temperature and wind assimilation increments valid at 3 December 1999 06 UTC. A single simulated surface pressure observation was inserted in the position 57N 3E at 3 December 1999 12 UTC. 4D-Var was applied over assimilation window 3 December 1999 06–12 UTC.

increments up to the 12 UTC, the time of the inserted simulated observation.

Considering the assimilation increments at the start of the assimilation window, for example those illustrated in Fig. 3, it needs to be mentioned again that these are heavily constrained by the applied isotropic and homogeneous structure functions.

5. Concluding remarks

It has been a great pleasure for me to participate in this discussion on the future development of data assimilation for numerical weather prediction. Great thanks to Eugenia Kalnay et al. and to the editor of *Tellus*.

To summarize my opinion, I am in favour of optimally combining the ideas of four-dimensional variational data assimilation and Ensemble Kalman Filtering. I have not said much about how to make this optimized combination. I believe that one promising approach is the one taken at NCEP and ECMWF, that is, to gradually introduce inhomogeneity, anisotropy and flow-dependency into the background error covariance matrix applied in 3D-Var and at the start of the assimilation window in 4D-Var. In order to model the flow-dependent part of the background error covariance, an ensemble of background states certainly is a natural source of information.

Finally, to avoid possible misunderstanding, I am a strong believer in probabilistic weather forecasting. Ensemble Prediction System (EPS) is one approach taken, and for EPS ensemble assimilation techniques are naturally applied. Ensemble assimilation techniques can also be applied within the framework of 3D-Var or 4D-Var, for example through perturbation of observations. Several groups (ECMWF, Météo-France and HIRLAM) have tried this approach, in the first instance for derivation of

background error structure functions, but also with EPS applications in perspective.

References

- Bergthorsson, P. and Döös, B. 1955. Numerical weather map analysis. *Tellus* **7**, 329–340.
- Eliassen, A. 1954. Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. *Inst. Weather and Climate Research*, Acad. Sci., Oslo, Report No. 5
- Gandin, L. S. 1963. *Objective Analysis of Meteorological Fields*. Hydromet Press, Leningrad.
- English translation by Israel Program for Scientific Translation Ltd. IDST Cat. No., 1373, 1965
- Gauthier, P. and Thépaut, J.-N. 2001. Impact of the Digital Filter as a Weak Constraint in the Preoperational 4DVAR Assimilation System of Météo-France. *Mon. Wea. Rev.* **129**, 2089–2102.
- Giering, R. and Kaminski, T. 2003. Applying TAF to generate efficient derivative code of Fortran 77–95 programs. *PAMM* **2**(1), 54–57.
- Gustafsson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M., and co-authors. 2001. Three-dimensional variational data assimilation for a limited area model. Part I: general formulation and the background error constraint. *Tellus* **53A**, 425–446.
- Huang, X.-Y., Yang, X., Gustafsson, N., Mogensen, K. and Lindskog, M. 2002. Four-dimensional variational data assimilation for a limited area. HIRLAM Technical Report, No. 57, pp 44, Available from SMHI.
- Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. and Ballabrera-Poy, J. 2007. 4D-Var or Ensemble Kalman Filter? *Tellus* **59A**, this issue.
- Lorenc, A. 1981. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.* **109**, 701–721.
- Mahfouf, J.-F., Bauer, P. and Maréchal, V. 2005. The assimilation of SSM/I and TMI rainfall rates in the ECMWF 4D-Var system. *Q.J.R. Meteorol. Soc.* **131**, 437–458.