Use of consistent operational ensemble variational assimilation to estimate and diagnose analysis and background error covariances

Loïk Berre, Gérald Desroziers,
Laure Raynaud, Rémi Montroty, and Florian Gibier

Météo-France/CNRS, CNRM/GAME
Toulouse, France

Abstract: The simulation of the error evolution is discussed from a formal point of view, including the representation of non linear effects. The operational ensemble variational assimilation at Météo-France is then briefly summarized and illustrated. It is also shown how such an ensemble assimilation can be used to diagnose analysis effects on the error covariances. A combined use of innovation-based and ensemble-based estimates is finally discussed.

1 Simulation of the error evolution with perturbed assimilations

1.1 Ensemble of perturbed assimilations versus deterministic square-root filters

The principles of ensemble data assimilation (EnDA) are relatively simple and general. In this study, we will focus on the use of perturbed assimilations, based, on one hand, on explicit observation perturbations (representative of observation errors), and on the other hand, on background perturbations which are either fully implicit (from the perturbed previous data assimilation cycles) or partly explicit (to represent model error contributions).

Applying this approach to a Kalman filter algorithm corresponds to an ensemble Kalman filter (EnKF). Similarly, its application to a variational algorithm leads to an ensemble variational assimilation (EnVar), such as the one which is operational at Météo-France since 2008. Such an EnVar system is very easy to implement and run from an existing variational system, as each member is basically quite similar to a usual variational assimilation experiment (albeit less expensive, depending for instance on horizontal resolution and number of outer-loops).

It may be mentioned also that there are other variants called deterministic square-root filters (e.g. Tippett et al 2003). They are based on a linear transformation of background perturbations into analysis perturbations. However, these square-root filters are more restrictive to some extent, as will be briefly discussed in section 1.2 also. Firstly, in order to make the transformation simple, they are based on an assumption that their gain matrix is optimal, which is usually not correct (due to sampling noise and model error approximations for instance). Secondly, because the analysis perturbation update is purely linear, they do not represent non linear effects of analysis schemes such as 4D-Var. Thirdly, such square-root filters are also often (if not always ?) restricted to a low-rank ensemble-based gain matrix, instead of the reference (possibly hybrid) full-rank gain used in 4D-Var for instance.

For these reasons, using an EnVar system can be seen as preferable than building up and using a more restrictive deterministic square-root filter system.

1.2 Analysis error and analysis perturbation equations

From a formal point of view, ensemble data assimilation is a way to simulate the analysis error equation (e.g. Berre et al 2006). For the sake of simplicity, we will start with the usual linear form of the analysis equation, and discuss the non linear case later on.

The (unperturbed) analysis state $\mathbf{x_a}$ can be written as a linear combination of a background $\mathbf{x_b}$ and observations $\mathbf{y}$, with $\mathbf{K}$ the specified gain matrix, and $\mathbf{H}$ the observation operator :

$$\mathbf{x_a} = (\mathbf{I} - \mathbf{KH})\mathbf{x_b} + \mathbf{Ky}$$

The same kind of equation, with the same operators, can be written formally for the true state $\mathbf{x}_\star$, with new specific inputs, corresponding to the true states in model and observation spaces :

$$\mathbf{x}_\star = (\mathbf{I} - \mathbf{KH})\mathbf{x}_\star + \mathbf{Ky}_\star$$

The analysis error equation is given by the difference between these two equations, and again the same kind of equation appears, with new specific inputs, namely background errors $\mathbf{e_b}$ and observation errors $\mathbf{e_o}$:

$$\mathbf{e_a} = (\mathbf{I} - \mathbf{KH})\mathbf{e_b} + \mathbf{Ke_o}$$

So this simple equation indicates that the analysis error evolves in a similar way as the usual analysis state equation : while the inputs change, the basic operators are the same, namely $\mathbf{I} - \mathbf{KH}$ and $\mathbf{K}$.

Another nice feature is that this equation is true even if $\mathbf{K}$ is suboptimal. And the form of the analysis error equation indicates that applying this effective suboptimal $\mathbf{K}$ to observation errors, and the associated $\mathbf{I} - \mathbf{KH}$ to background errors, is a way to take this suboptimality (and its effect on the analysis error) into account.

Moreover, this equation can be shown to be also valid for a weakly non linear analysis system such as 4D-Var (Desroziers et al 2009). In practice, using perturbed non linear analyses is in fact a way to represent non linear effects in this analysis part of the error evolution. This is similar to the representation of non linear effects in the forecast part of the error evolution, through the use of perturbed non linear forecasts (instead of using a linear model for the forecast evolution of initial perturbations).

These last two features (about suboptimality and non linearity) suggest that using observation and background perturbations, to produce analysis perturbations (in a non linear way), is likely to be more general and appropriate than using a direct linear transformation of background perturbations into analysis perturbations (as done in deterministic square-root filters).

The equation of the analysis perturbation can be derived in a similar way as for the analysis error. The perturbed analysis $\mathbf{x'_a}$ is a linear combination of perturbed inputs $\mathbf{x'_b}$ and $\mathbf{y'}$:

$$\mathbf{x'_a} = (\mathbf{I} - \mathbf{KH})\mathbf{x'_b} + \mathbf{Ky'}$$

The equation of the analysis perturbation $\varepsilon_\mathbf{a}$ corresponds to the difference between this perturbed analysis and the unperturbed analysis:

$$\varepsilon_\mathbf{a} = (\mathbf{I} - \mathbf{KH})\varepsilon_\mathbf{b} + \mathbf{K}\varepsilon_\mathbf{o}$$

It appears that the perturbation equation is again very close to the analysis error equation, with new specific inputs, namely background perturbations ε_b and observation perturbations ε_o.

So this illustrates the fact that an ensemble of perturbed assimilations mimics in a relevant way the manner in which errors evolve in the data assimilation cycle.

1.3 Formal comparison with the NMC method

To emphasize further this point, the analysis error and perturbation equations can be compared with the equation of the analysis increment $\delta\mathbf{x} = \mathbf{x_a} - \mathbf{x_b}$, which is the basic ingredient in the NMC method (e.g. Berre et al 2006):

$$\delta\mathbf{x} = -\mathbf{KHe_b} + \mathbf{Ke_o}$$

In the analysis increment equation, the inputs are the same as in the analysis error equation, but one of the operators is different. In fact, $\mathbf{I} - \mathbf{KH}$, which can be seen as a high-pass filter (Daley 1991), is replaced by $-\mathbf{KH}$, which is a low-pass filter. This is consistent with the fact that estimated correlation functions are too broad in the NMC method, compared to ensemble assimilation (e.g. Belo Pereira and Berre 2006).

More generally, this kind of formal comparison indicates that the analysis error equation is better simulated in EnDA than with the NMC method.

1.4 Open issues in the error simulation technique

There are of course open issues in the way of simulating the error evolution with ensemble assimilation. Suppose for instance that we want to simulate errors in a 4D-Var cycle with a high resolution model. Due to numerical cost, one may be interested by two possible approximations in the ensemble simulation of errors.

Firstly, one may consider to reduce the horizontal resolution of the model, which is relatively classical. Secondly, one may approximate the reference gain matrix of 4D-Var, either with 3D-Fgat, or with 4D-Var and fewer outer loops. Another possibility is to use either EnKF or ETKF, in the error simulation part, but in this case the gain $\mathbf{K}$ in this part will be derived from the ensemble information essentially. In other words, the error simulation will be based on a low-rank ensemble-based gain matrix, which potentially can be a rather coarse approximation of the reference (possibly hybrid) full-rank gain used in 4D-Var.

For reasons evoked in section 1.1, using a "consistent ensemble variational assimilation" (i.e. a variational approach in both perturbed and unperturbed components) seems preferable than using either EnKF or ETKF in the error simulation part (together with a variational system in the deterministic part).

2 The operational Météo-France ensemble variational assimilation

2.1 Summary of the EnVar system

The main features of this system are briefly evoked here, as a summary of their description in Berre et al (2007).

This ensemble is made of 6 perturbed global members, with truncation T359, 60 vertical levels, and 3D-Fgat for the Arpege model. This system will be upgraded in 2009, by using 4D-Var with one outer loop, 70 vertical model levels, and truncation T399.

An optimized spatial filter is applied to error variances, in order to increase the robustness of variance estimates. Noise to signal studies (Raynaud et al 2009) indicate that with such an optimized spatial filter applied to the 6-member variance estimate, the relative error estimation variance (of the background error variance field) is around 10%, which is similar to the quality of raw variances estimated from a 21-member ensemble (according

to $\frac{2}{N-1} = \frac{2}{20} \approx 10\%$). Moreover, background error standard deviations are inflated by a factor 1.3, in order to represent model error contributions.

The Arpege 4D-Var uses these sigmab's of the day for vorticity (and thus, implicitly, also for the associated balanced parts of temperature, surface pressure and divergence), and this is operational since July 2008. The upgrading in 2009 will include an extension of this, by using the ensemble to specify flow-dependent variances also for specific humidity and for the unbalanced parts of temperature, surface pressure and divergence. This extension contributes to additional positive impacts, in particular due to flow-dependent humidity variances.

An experimental coupling with regional models has been carried out also, both at 10 km resolution with the Aladin model (Desroziers et al 2007), and at 2.5 km resolution with the Arome model (Brousseau et al 2007). This is applied in order to estimate the static part of regional error covariances, and an extension to specify flow-dependent covariances is also investigated.

2.2 Flow-dependent error variances and their impact

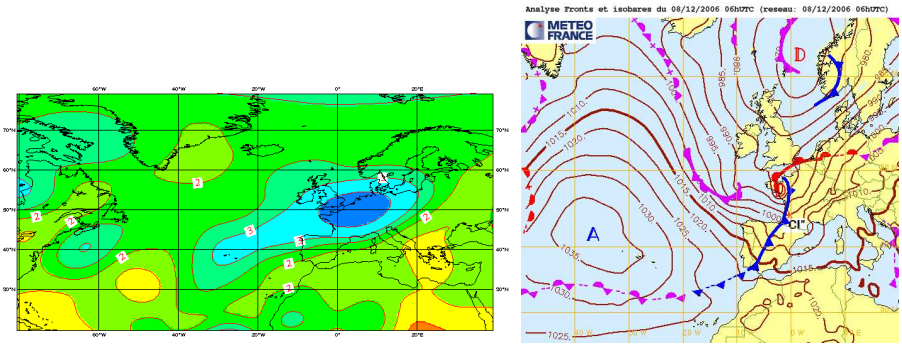

Figure 1: Left panel: ensemble-based background error standard deviations for vorticity near 500 hPa (isoline interval: $10^{-4}s^{-1}$). Right panel: mean sea level pressure field. Both fields are valid on 8 December 2006.

The left panel of Figure 1 is an example of field of sigmab's of the day, for vorticity near 500 hPa, on the 8th of December 2006. Large values of sigmab's are in blue, so they are located over Europe for this situation.

The right panel shows the associated weather situation, in terms of mean sea level pressure. This indicates the occurence of a severe winter storm over France, which is connected to large values of sigmab's, in accordance with expected uncertainties in this kind of intense situation.

As illustrated by Figure 7 in Berre et al (2007), it should be mentioned also that the largest flow-dependent changes of the variance field (compared to climatological variance fields) are relatively localized spatially. This suggests that one should not expect a huge and systematic improvement when using flow-dependent sigmab's instead of static sigmab's, but rather positive impacts which can be relatively localized in space and in time, and which are connected to intense weather situations. This general expectation will be illustrated experimentally a bit below in this section.

The impact of this kind of flow-dependent sigmab's is illustrated in Figure 2 for a severe winter storm which occured over France on 10 February 2009. The red isolines correspond to 36 hour forecasts of mean sea level pressure, based on static sigmab's in the left panel, and based on sigmab's of the day in the right panel. The blue isolines correspond to the verifying analysis. It appears that using flow-dependent sigmab's has a positive impact on the forecast of this severe storm, in terms of depth of the low and gradient intensity.

Figure 3 is another illustration for a case of tropical cyclone near Madagascar, taken from a study by Montroty

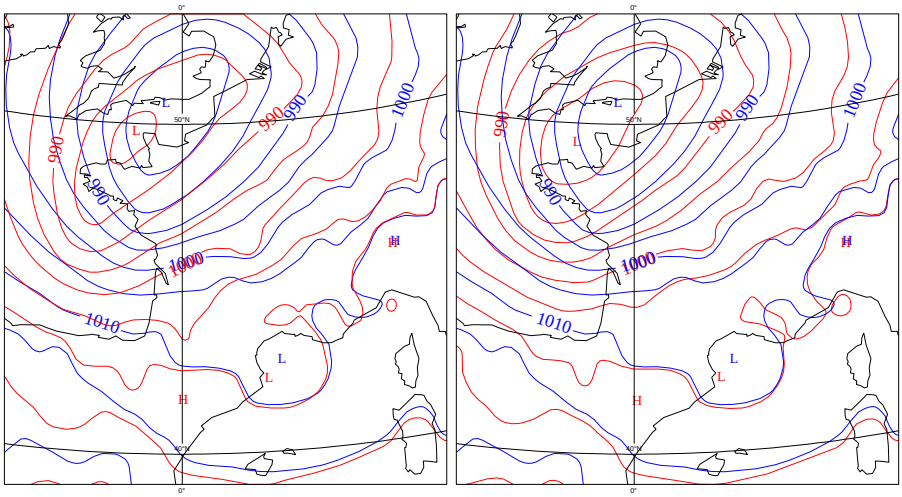

Figure 2: Severe winter storm over France on 10 February 2009. The red isolines correspond to 36h forecasts of mean sea level pressure, while the blue isolines are from the verifying analysis. Left panel: 36h forecast based on static sigmab's. Right panel: 36h forecast based on flow-dependent sigmab's.

(2008). The left panel shows trajectory forecasts, and the left panel corresponds to forecasts of intensity. The black line is the verifying observation, the yellow curves are based on static sigmab's, and the purple curves are based on flow-dependent sigmab's. It can be seen that using the flow-dependent sigmab has a positive impact on the trajectory and on the intensity of this tropical cyclone. A detailed case study has shown that this positive impact arises from a beneficial amplification of analysis increments, in some important sensitive areas.

Extended impact runs have been carried out over several monthly periods to evaluate the average impact of flow-dependent sigmab's.

This is illustrated by Figure 4, where the three panels correspond to the decrease of the geopotential forecast RMS over Northern America, as a function of height and forecast range, when using flow-dependent sigmab's instead of static sigmab's. The blue isolines correspond to a positive decrease of the RMS, i.e. to an improvement due to the use of flow-dependent sigmab's.

It appears that using flow-dependent sigmab's has a positive impact on the average, which tends to be more pronounced during the two winter seasons than in autumn. This is likely to be related to the more intense cyclogeneses in winter.

Moreover, examination of time series of RMS indicates that these average improvements correspond to a tendency to reduce local RMS peaks, as shown by Figure 8 in Berre et al (2007). This is consistent with the aforementioned idea that using flow-dependent sigmab's is likely to be particularly beneficial for local intense weather situations.

3 Diagnostics of analysis and background errors

3.1 Expressions of the analysis error covariance

When the analysis is optimal, a classical estimate of the analysis error covariance is : $\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}$, which indicates that analysis errors are expected to be smaller than background errors.

The analysis error estimate which is provided by ensemble assimilation (based on an ensemble of perturbed

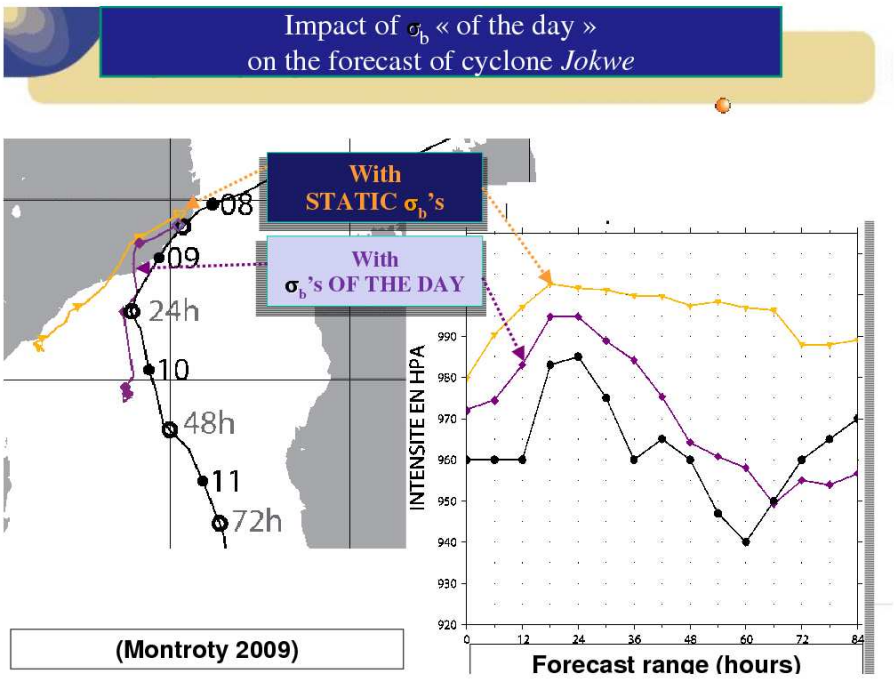

Figure 3: Forecasts of the tropical cyclone Jokwe, based on static sigmab's (in orange) and on flow-dependent sigmab's (in purple), as a function of forecast range. The verifying observation is in black. Left panel: trajectory forecast. Right panel: intensity forecast.

assimilations) is more general, as it is valid also for a suboptimal analysis system :

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}(\mathbf{I} - \mathbf{KH})^{\mathbf{T}} + \mathbf{KRK}^{\mathbf{T}}$$

As shown by this equation, the analysis error equation is a bit more complex in the general case, and ensemble assimilation can be an efficient way to handle this complexity. In particular, ensemble-based estimates of **A** and **B** can be compared, in order to diagnose analysis effects.

3.2 Local covariance estimates

One interesting thing to look at is the horizontal distribution of error standard deviations. This is shown for instance in Figure 7.a in Belo Pereira and Berre (2006) for background errors of vorticity near 500 hPa, averaged over a one-month period. It can be seen that sigmab values tend to be larger in data-sparse oceanic areas such as the Pacific and the Atlantic, and that relatively small sigmab's are found in data-dense areas such as Northern America and Europe.

It is also possible to calculate the difference between estimates of sigmab's and sigmaa's, calculated with the energy norm in this case. This is shown in Figure 5, for statistics averaged over a one month period. One can notice that the error reduction is stronger in data-dense areas, and also over isolated islands where radiosondes are available.

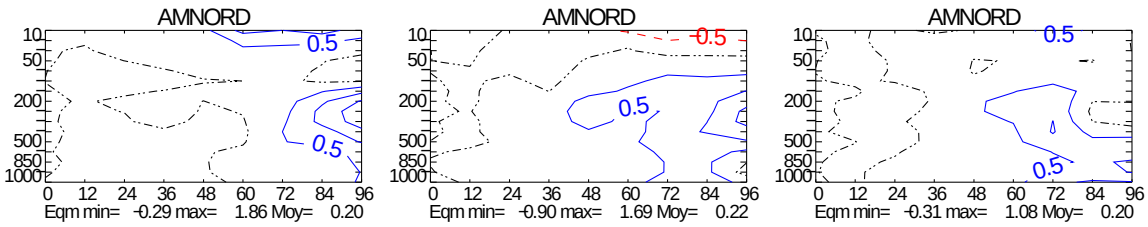

Figure 4: Reduction of average geopotential RMSE over Northern America, when using flow-dependent sigmab's, instead of static sigmab's, as a function of height (y axis) and forecast range (x axis). Blue isolines correspond to a positive impact of flow-dependent sigmab's, while the black isoline corresponds to a neutral impact. Isoline spacing: 0.5m. Left: November 2006 - January 2007 (3 months). Middle: February - March 2008 (1 month). Right: September - October 2007 (1 month).

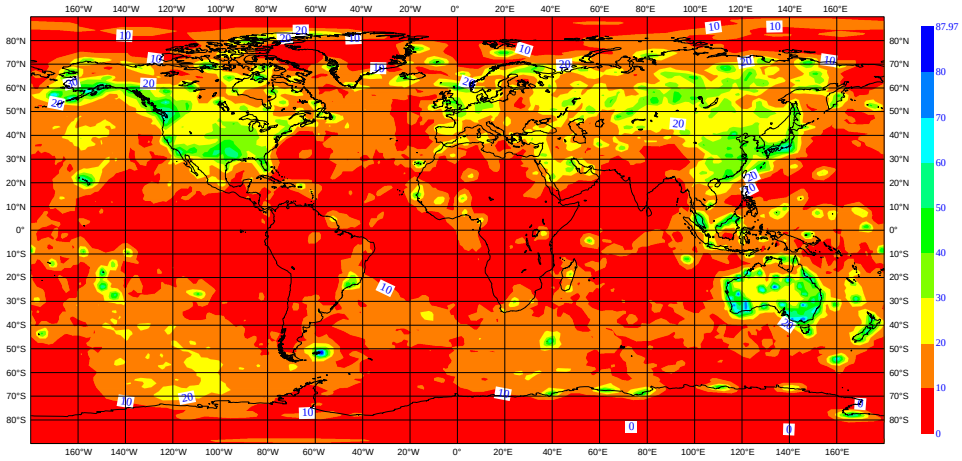

Figure 5: Difference between local estimates of sigmab and sigmaa, averaged over a one-month period, with a total energy norm.

3.3 Global covariance estimates

Another typical diagnostic corresponds to error variance spectra. This is shown in Figure 2.a of Ştefănescu et al (2006) for surface pressure, with a full curve for the background and a dashed curve for the analysis. These two spectra have their maximum in the large scales, which is typical for a variable like surface pressure.

Moreover, the analysis error variance tends to be smaller than the background error variance. This is consistent with the expectation that errors tend to be reduced by the analysis. It can be noticed also that the error reduction tends to be larger in the large scales than in the small scales. As shown in Daley (1991) for instance, this is consistent with the expectation that the error reduction is maximum for components at which the amplitude of background error is maximum, compared to the amplitude of observation error.

It is also interesting to look at vertical profiles of error standard deviations. Figure 2.c in Ştefănescu et al (2006) is an example for temperature. The full curve is for the background, whereas the dashed line is for the analysis. It can be seen that the two profiles are relatively similar, and that the analysis error tends to be smaller than the background error. This is again consistent with the expectation that errors tend to be reduced by the analysis.

It can be noticed also that the error reduction tends to be larger in the mid-troposphere than near the surface. This is likely to be related to the fact that background errors are larger scale in the mid-troposphere. This allows their reduction to be relatively strong through the analysis step.

It is also interesting to diagnose how analysis errors evolve into forecast errors, in the ensemble assimilation. This is illustrated in Figure 16 in Belo Pereira and Berre (2006), by comparing vertical profiles of wind error standard deviations. The full line is for the analysis, and the dotted line is for the associated 6h forecast. It can be seen that there is an increase of spread during the 6h forecast, which is likely to correspond to effects of baroclinic instabilities for instance (knowing that there was no model error simulation in this ensemble).

It may be mentioned that data impact studies can be carried out also by using ensemble assimilation. The idea is to compare the analysis spread when using different observation systems. This has been done by Tan et al (2007) for instance, in order to examine the impact of wind lidars and radiosondes.

4 Combined use of innovation-based and EnDA estimates

One way to validate ensemble sigmab estimates is to calculate innovation-based sigmab estimates, using techniques proposed for instance by Hollingsworth and Lönnberg (1986) and Desroziers et al (2005). Following Desroziers et al (2005), it can be shown in fact that the covariance between the analysis increment and the innovation is an estimate of the background error covariance, in observation space.

In principle, this can be calculated for a specific date, but then the local sigmab is calculated from a single error realization, as is if we had only ONE member in an ensemble. Conversely, if we calculate local spatial averages of these sigmab's, the sample size will be increased, and comparison with ensemble estimates can be considered.

Figure 6 in Berre et al (2007) is an example of comparison for HIRS 7 for a specific single date. The top panel is ensemble estimates of sigmab, with local spatial averages over a 500 km radius. The bottom panel corresponds to innovation-based sigmab estimates, using Desroziers' formula, with a similar 500 km spatial average. It is striking to notice that similar patterns can be seen in these two independent estimates, such as large sigmab values over Central Pacific, and small sigmab values over Southern Atlantic. So this kind of comparison is a way to validate ensemble estimates, and this may be also a way to estimate model error covariances in particular, as proposed for instance by Daley (1992).

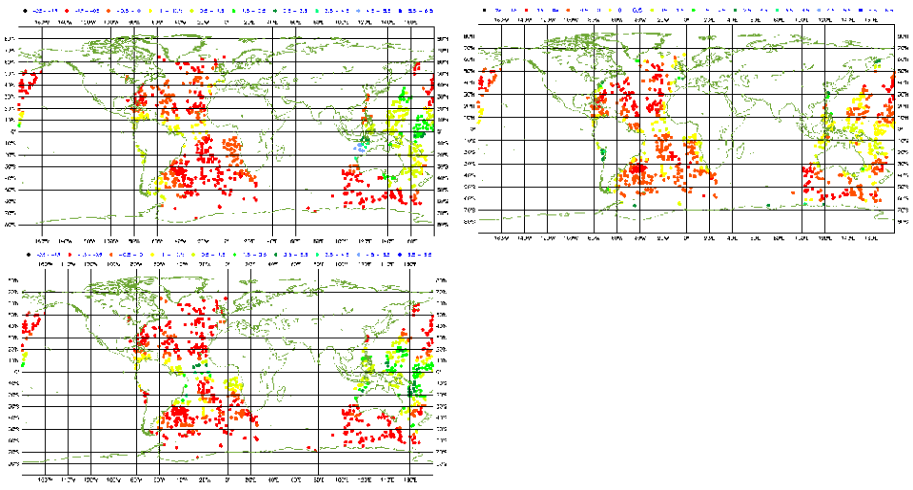

Figure 6: Estimates of sigmab for the channel HIRS-7, on 24 January 2009 at 00 UTC, centered and normalized. Top left: innovation-based estimates. Top right: ensemble 3D-Fgat estimates. Bottom left: ensemble 4D-Var estimates.

More recently, these diagnostics have been used to evaluate the possibility to use a 4D-Var ensemble at Météo-France, instead of the current 3D-Fgat ensemble. The top left panel of Figure 6 is the innovation-based estimate

of sigmab for the channel HIRS-7, from the reference unperturbed 4D-Var system. The top right panel is the 3D-Fgat ensemble estimate, and the bottom left panel is the 4D-Var ensemble estimate. It appears that the ensemble 4D-Var estimate is much closer to the innovation-based estimate (see the green areas over Central Pacific for instance ; the global correlation with the innovation-based map is increased from 0.2 to 0.6 when replacing ensemble 3D-Fgat by ensemble 4D-Var).

This is also consistent with the expectation that ensemble 4D-Var should simulate the error evolution of the reference deterministic 4D-Var system in a more accurate way (thanks to the use of the 4D-Var gain matrix in the analysis perturbation update). Impact experiments indicate also that using flow-dependent sigmab's provided by ensemble 4D-Var contributes to an additional positive impact, compared to the use of flow-dependent sigmab's from ensemble 3D-Fgat. So these results support the idea to use 4D-Var in the ensemble assimilation system, instead of the current ensemble 3D-Fgat.

Another important aspect to evoke is the possibility to estimate model error covariances when combining innovations and ensemble assimilation. Typically, **B** can be written as the sum of analysis errors evolved by the model **M**, and model error covariances contained in **Q** :

$$\mathbf{B} = \mathbf{M}\mathbf{A}\mathbf{M}^{\mathbf{T}} + \mathbf{Q}$$

As (more or less) proposed by Daley (1992), ensemble assimilation can be used to estimate the evolved analysis error component. Moreover, as shown in the previous slides, innovations can be used to estimate background errors in observation space. This indicates that a natural technique to estimate **Q** is to use differences between ensemble- and innovation-based estimates of covariances.

5 Conclusions and perspectives

Ensemble assimilation allows analysis/background error cycling to be simulated and diagnosed. Using an ensemble of perturbed 4D-Var assimilations is relatively easy to implement, and it allows non linear analysis effects to be represented in the error simulation.

Flow-dependent covariances can be estimated, with positive impacts on intense/severe weather events such as mid-latitude storms and tropical cyclones. Diagnostic comparison between analysis and background spread provides information about analysis effects. Moreover, comparisons with innovation-based estimates can be carried out, for validation, and also for estimation of model error covariances.

Some examples of open issues are the optimization of the error simulation (i.e. choice of a good compromise between approximations of the resolution and analysis scheme used in the deterministic part), and the covariance filtering technique (e.g. spectral/wavelet optimized filters versus Schur filter for instance).

References

Belo Pereira, M. and L. Berre, 2006: The Use of an Ensemble Approach to Study the Background Error Covariances in a Global NWP Model. *Mon. Wea. Rev.*, **134**, 2466-2489.

Berre, L., 2000: Estimation of synoptic and meso scale forecast error covariances in a limited area model. *Mon. Wea. Rev.*, **128**, 644-667.

Berre, L., O. Pannekoucke, G. Desroziers, S.E. Ştefănescu, B. Chapnik and L. Raynaud, 2007: A variational assimilation ensemble and the spatial filtering of its error covariances: increase of sample size by local spatial averaging. Proceedings of the ECMWF workshop on flow-dependent aspects of data assimilation, 11-13 June 2007, 151-168.

(http://www.ecmwf.int/publications/library/do/references/list/14092007)

Brousseau, P., L. Berre, et al, 2008: A prototype convective-scale data assimilation system for operations: the Arome-RUC. Hirlam technical report N.68 on the SRNWP workshop on data assimilation, 21-23 March 2007, 23-30.

Daley, R., 1991: Atmospheric data analysis. Cambridge Univerisity Press, 460 pp.

Daley, R., 1992: Estimating Model-Error Covariances for Application to Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **120**, 17351746.

Desroziers, G., L. Berre, B. Chapnik, et P. Poli, 2005: Diagnosis of observation, background and analysis error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385-3396.

Desroziers, G., L. Berre, O. Pannekoucke, S.E. Ştefănescu, P. Brousseau, L. Auger, B. Chapnik and L. Raynaud, 2008b: Flow-dependent error covariances from variational assimilation ensembles on global and regional domains. Hirlam technical report N.68 on the SRNWP workshop on data assimilation, 21-23 March 2007, 3-22.

Desroziers, G., L. Berre, V. Chabot and B. Chapnik, 2009: A posteriori diagnostics in an ensemble of perturbed analyses. To appear in *Mon. Wea. Rev.*

Hollingsworth, A. and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111-136.

Montroty, R., 2008: Impact d'une assimilation de données à mésoéchelle. Ph.D. dissertation, Université Paul Sabatier, 221 pp. [Available from Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France.]

Raynaud, L., L. Berre , and G. Desroziers, 2008: Spatial averaging of ensemble-based background-error variances. *Quart. J. Roy. Meteor. Soc.*, **134**, 1003-1014.

Raynaud, L., L. Berre , and G. Desroziers, 2009: Objective filtering of ensemble-based background-error variances. To appear in *Quart. J. Roy. Meteor. Soc*.

Ştefănescu, S.E., L. Berre and M. Belo Pereira, 2006: The Evolution of Dispersion Spectra and the Evaluation of Model Differences in an Ensemble Estimation of Error Statistics for a Limited Area Analysis. *Mon. Wea. Rev.*, **134**, 3454-3476.

Tan, D.G.H., E. Andersson, M. Fisher and L. Isaksen, 2007: Observing system impact assessment using a data assimilation ensemble technique: application to the ADM-Aeolus wind profiling mission. ECMWF Technical Memorandum, **510**, 18pp.

Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M. and J.S. Whitaker, 2003: Ensemble Square Root Filters *Mon. Wea. Rev.*, **131**, 1485-1490.