At Météo-France, the convective-scale model AROME draws its analysis from a 3D-Var scheme. The background error covariances provided to the assimilation system have to be modelled in an appropriate way, since they have a deep impact on the analysis. In the current operational configuration, a spectral covariance model is used with homogeneous climatological variances over the domain whereas they should be heterogeneous and flow-dependent. Consequently, our current system assimilates observations in a sub-optimal way, especially during intense weather events.

Two strategies are investigated to get flow-dependent variances:
- interpolating variances from an ensemble assimilation at global scale (AEARP, operational at Météo France since July 2008),
- computing variances from a small ensemble assimilation at convective scale (6 members), and removing the sampling noise with a spatial filter designed especially for convective-scale variances.

Whereas the first option has been evaluated with mixed results, the second is still under development and several open questions remain.

Diagnostic study of flow-dependent variance maps over the AROME domain

Unbalanced component of specific humidity background error variances at level 50 (~ 945 hPa), for six configurations. This is one of the cases with the highest correlations.

Sample noise in estimated covariances: the Wishart theory

Given an ensemble of \( N \) model states \( \{x^k\}_k \), the sample covariance can be computed by:

\[
B = \frac{1}{N-1} S
\]

where

\[
S = \sum_{k=1}^{N} (x_k^2 - \bar{x}_x)(x_k^2 - \bar{x}_x)^T
\]

with \( \bar{x}_x = \frac{1}{N} \sum_{k=1}^{N} x_k^2 \)

If the states are following a Gaussian distribution \( x_k^2 \sim N(x^0, B) \), then \( S \) is following a Wishart distribution [3], giving the distribution of \( B \).

In NWP application, \( B \) is singular and does not have an explicit PDF, but all moments can yet be computed. The sampling noise \( B^* = B - B^\dagger \) two first moments are thus:

\[
E[B^k] = 0
\]

\[
E[(B^1 - E[B^1])(B^2 - E[B^2])] = \frac{1}{N-1}(B_{12}B_{12} + B_{21}B_{21})
\]

Methodologies and open questions for the linear filtering of variances

Several approaches can be used to compute the filter gain \( H \):

- Expressing approximations of \( \text{Cov}(\tilde{v}) \) and \( \text{Cov}(v^0) \) in a base where they are diagonal, so that \( H \) is diagonal too (e.g. spectral filters in [1] and [2]).
  - A spectral base assumes that signal and noise have homogeneous statistics over the domain, which is not the case according to the Wishart theory. Is it efficient even so ?
  - What base would be the best (wavelets, curvelets, ...) ?
- Setting a parametric model of the filter gain \( H \), whose parameters can be computed from:
  - climatological values,
  - external data (e.g. background correlation length-scales),
  - the raw estimation \( v^\dagger \) itself, making the filter non-linear.
  - What model would be the best (wavelets thresholding, recursive filters, ...) ?
  - By what mean can we specify the filter parameters from other data (external calibration, internal optimization, ...) ?

Linear filtering theory

The raw estimation \( \tilde{v} \) of the noise-free signal \( v^\dagger \) can be linearly filtered:

\[
\tilde{v} = Hv + h
\]

Minimizing the Euclidean norm expectation between filtered and noise-free signal, \( E[|v - \tilde{v}|^2] \), leads to:

\[
\begin{align*}
H &= \text{Cov}(v)\text{Cov}(v^\dagger)^{-1} \\
h &= E[v^\dagger] - HS(v)
\end{align*}
\]

Assuming that the signal \( v^\dagger \) and the noise \( v^0 = v - v^\dagger \) are independent, the linear filter gain can be rewritten:

\[
H = (\text{Cov}(v) - \text{Cov}(v^0))\text{Cov}(v^\dagger)^{-1}
\]

Since we want to filter variances, the noise covariance can be found from the Wishart theory:

\[
\text{Cov}(v^0) = \frac{2}{N-1} B - B^\dagger
\]

References

