Filtering of ensemble-based forecast error variance maps, a toy-models approach

Benjamin Ménétrier, Thibaut Montmerle, Loik Berre and Yann Michel
CNRM-GAME (Météo-France/CNRS)

Necessity of new development tools

The flow-dependency of forecast error variance maps requires daily computations of ensemble of forecasts. Due to the high computational cost of NWP models, only small ensembles can be run, generating a strong sampling noise. This issue has been successfully addressed for the global model ARPEGE with two filtering methods (Berre, MWR, 2010): local spatial averaging and objective spectral filtering, which are based on strong assumptions about the sampling noise statistical properties (gaussian, homogeneous, isotropic), that are reasonable at global scale. However, these assumptions are less valid for non-hydrostatic models with explicit moist convection, as in the operational cloud-resolving model AROME, where the assimilation of cloud and precipitations related observations (radars, cloudy radiances) requires flow-dependent and precise forecast error variance maps. To test new filtering techniques, low-cost experiments are conducted with two approaches: randomization and toy-models.

Randomization approach

Methodology:
1. A modeled $B_{\text{ref}}$ matrix is explicitly specified.
2. An ensemble is generated by randomization: $\epsilon_k = B_{\text{ref}}^{1/2} \zeta$, where $\zeta \sim \mathcal{N}(0,1)$.
3. The variance map computed from $\langle \epsilon_k \rangle$ is compared to those of $B_{\text{ref}}$.

Advantages:
- Arbitrary maps can be specified for the true forecast error variance,
- Scores are computed with respect to an exact reference,
- Very large ensembles are available at a very low computational cost,

Drawbacks:
- True background error statistics are modeled statistics (e.g. homogeneous and isotropic correlations),
- Forecast error estimates have a specific PDF (gaussian here),
- Extrapolation of new methods and results to NWP model ensembles is very hazardous.

New toy-models approach

Methodology:
1. An appropriate toy-model is built,
   - Typical 2D models in geophysics: vorticity advection or shallow-water,
   - NWP models-mimicked dynamical core (spectral representation, truncature, LAM specificities),
2. An ensemble assimilation is performed with a full variational system,
   - Heterogeneous observation network, perturbed observations, L-BFGS method minimizer.
3. Maps of forecast error variance are computed from the whole ensemble (to get a reference) and from smaller sub-samples, and then compared.

Advantages:
- Background error statistics can be implicitly adjusted (through model forcing, observations location, ...), to mimic structures of NWP models ones,
- Background error statistics have non-idealized properties (correlations heterogeneity here),
- Forecast error estimates do not always have a specific PDF (non-gaussianity here),
- Large ensembles are still available at a reasonable computational cost,
- Extrapolation of new methods and results to NWP model ensembles is possible, with care.

Variance map filtering techniques

Local spatial averaging (Raynaud et al., QJRMS, 2008):
- The variance map is convoluted in grid-point space with a uniform compactly supported kernel,
- The averaging length-scale requires a manual tuning,
- The computational cost is higher (convolution + tuning).

Objective spectral filtering (Raynaud et al., QJRMS, 2009):
- The power spectrum of the variance map is filtered,
- Equivalent to a convolution in grid-point space with an homogeneous and isotropic kernel,
- Under restrictive assumptions on sampling noise statistics (gaussian noise, homogeneous covariances), an objective filter can be determined.

Preliminary conclusions and perspectives

Both filtering methods presented here are equivalent to a convolution with a smoothing kernel. Therefore, they are efficient provided that the typical variations length-scale $L_p$ of the true variance map is almost uniform, and is larger than the length-scale of sampling noise, closely related to the forecast error correlation length-scale (Raynaud et al., QJRMS, 2009). Moreover, as displayed on the right panel, objective filtering is no longer optimal if the assumptions on sampling noise statistics are lacking, which is the case in the toy-model experiment (heterogeneous correlations).

If the fields to filter are strongly heterogeneous and anisotropic, as it is the case in the toy-model presented here (or in CRMs), and since the true variance map structure is not well known, the new denoising methods should be adaptive, and they should not rely on restrictive assumptions about the sampling noise statistics. Wavelet filtering seems to provide promising algorithms (e.g. non-linear thresholding), and is currently under experiment. Limited-Area Models domain specificities still need to be taken in account.