Localization in the EnKF - why does it increase the rank of Pf?*

Ross Bannister, August 2008

1. The raw forecast error covariance matrix provided by the ensemble

Let the ensemble perturbation for member k (of a short forecast) be the vector ε_k. These can be assembled into columns of the vector **E**, ie

$$\mathbf{E} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N). \quad (1)$$

Matrix element E_{ik} is then component i of ε_k. The raw error covariance matrix is then

$$\mathbf{P}_f^{\text{raw}} = \frac{1}{N-1}\mathbf{E}\mathbf{E}^{\text{T}}, \quad (2)$$

which has matrix elements

$$(P_f^{\text{raw}})_{ij} = \frac{1}{N-1}\sum_{k=1}^{N} E_{ik}E_{jk}. \quad (3)$$

The rank of the matrix in (2) will be subject to the inequality

$$\text{rank}\,(\mathbf{P}_f^{\text{raw}}) \leqslant N. \quad (4)$$

2. The localized forecast error covariance matrix

Covariance (2) is likely to contain spurious features if the matrix is undersampled. Multiplying elementwise (Schur product) with a localization matrix **C** gives

$$\mathbf{P}_f^{\text{loc}} = \mathbf{P}_f^{\text{raw}} \odot \mathbf{C}, \quad (5)$$

which has matrix elements

$$(P_f^{\text{loc}})_{ij} = \frac{1}{N-1}\sum_{k=1}^{N} E_{ik}E_{jk}C_{ij}. \quad (6)$$

The **C**-matrix can itself be modelled by a population of N virtual ensemble members, η_k that have the correct covariance. Like the ε_k, the η_k may be assembled into columns of a matrix, **L**. Then **C** is

$$\mathbf{C} \approx \frac{1}{N-1}\mathbf{L}\mathbf{L}^{\text{T}}. \quad (7)$$

Matrix elements of the localized forecast error covariance matrix (6) are then

$$(P_f^{\text{loc}})_{ij} = \frac{1}{(N-1)^2}\sum_{k=1}^{N} E_{ik}E_{jk}\sum_{k'=1}^{N} L_{ik'}L_{jk'}$$
$$= \frac{1}{(N-1)^2}\sum_{k,k'=1}^{N} (E_{ik}L_{ik'})(E_{jk}L_{jk'}). \quad (8)$$

3. The modulated ensemble members

The bracketed terms in (8) can be redefined by the modulated ensemble members

$$\hat{E}_{i(kk')} = E_{ik}L_{ik'},$$
$$\hat{E}_{il} = E_{ik}L_{ik'}, \quad (9)$$

where the index l prescribes a particular k and k'. There are N^2 different combinations (presumably) of modulated ensemble members. Therefore we may expect that the rank of $\mathbf{P}_f^{\text{loc}}$ is

$$\text{rank}\,(\mathbf{P}_f^{\text{loc}}) \leqslant N^2. \quad (10)$$

This is why the rank of the localized forecast error covariance matrix increases.

*Reference - Craig Bishop seminar, "Data assimilation using modulated ensembles (DAMES)", July 2008, Met Office and Univ. of Reading.