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 \( \epsilon_k \). These can be assembled into columns of the vector \( \mathbf{E} \), ie

\[
\mathbf{E} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_N).
\]

Matrix element \( E_{ik} \) is then component \( i \) of \( \epsilon_k \).

The raw error covariance matrix is then

\[
P_{\text{raw}}^f = \frac{1}{N-1}\mathbf{E}\mathbf{E}^T,
\]

which has matrix elements

\[
(P_{\text{raw}}^f)_{ij} = \frac{1}{N-1} \sum_{k=1}^{N} E_{ik} E_{jk}.
\]

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

\[
\text{rank}(P_{\text{raw}}^f) \leq N.
\]

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 \( \mathbf{C} \) gives

\[
P_{\text{loc}}^f = P_{\text{raw}}^f \odot \mathbf{C},
\]

which has matrix elements

\[
(P_{\text{loc}}^f)_{ij} = \frac{1}{N-1} \sum_{k=1}^{N} E_{ik} E_{jk} C_{ij}.
\]

The \( \mathbf{C} \)-matrix can itself be modelled by a population of \( N \) virtual ensemble members, \( \eta_k \) that have the correct covariance. Like the \( \epsilon_k \), the \( \eta_k \) may be assembled into columns of a matrix, \( \mathbf{L} \). Then \( \mathbf{C} \) is

\[
\mathbf{C} \approx \frac{1}{N-1} \mathbf{LL}^T.
\]

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

\[
(P_{\text{loc}}^f)_{ij} = \frac{1}{(N-1)^2} \sum_{k=1}^{N} E_{ik} E_{jk} \sum_{k'=1}^{N} L_{ik'} L_{jk'}.
\]

\[
= \frac{1}{(N-1)^2} \sum_{k,k'=1}^{N} (E_{ik} L_{ik'}) (E_{jk} L_{jk'}). \quad (8)
\]

3. The modulated ensemble members

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

\[
\bar{E}_{ik} = E_{ik} L_{ik'},
\]

\[
\bar{E}_{il} = E_{ik} L_{jl},
\]

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 \( P_{\text{loc}}^f \) is

\[
\text{rank}(P_{\text{loc}}^f) \leq N^2.
\]

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.