Localization (Houtekamer, Mitchell 2001)

In EnKF, the ensemble size N is often much smaller than the state dimension n. For atmospheric models, $n \sim 10^7$, but $N \sim 10^2$ is affordable.

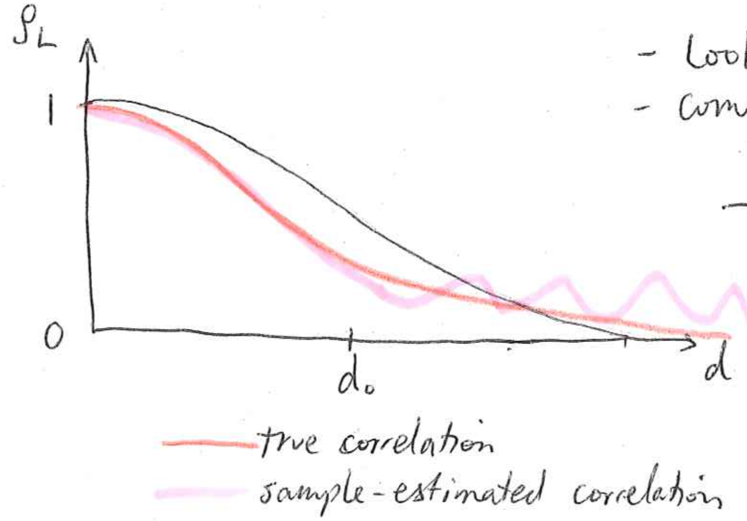
Estimated P^b using N ensemble members can only span a N-dimensional subspace; Spurious long-distance correlations will occur if the analysis domain dimension is much larger than N $\rightarrow$ the "rank problem" (Anderson 2001) (Hamill et al 2001) (Lorenc 2003)

Use local analysis to reduce the dimension of the problem.
$\rightarrow$ Localization.

- Localization function:
  (Gaspari Cohn 1999) fifth-order Polynomial
  Let d be the distance between observation and the analysis variable, d_0 be a cutoff distance for the Localization fcn.

$$\rho_L = \begin{cases} -\frac{r^5}{4} + \frac{r^4}{2} + \frac{5}{8}r^3 - \frac{5}{3}r^2 + 1, & 0 \leq r < 1, \\ \frac{r^5}{12} - \frac{r^4}{2} + \frac{5}{8}r^3 + \frac{5}{3}r^2 - 5r + 4 - \frac{2}{3r}, & 1 \leq r < 2, \\ 0, & r \geq 2 \end{cases} \quad (1)$$

where $r \equiv \frac{|d|}{d_0}$



- Looks like a Gaussian
- compact support $(0, 2d_0)$

$\rightarrow$ tune d_0 so that spurious correlations are removed, yet useful correlations are kept untouched.

① Model-space localization:

$$x^a = x^b + (\rho_L \circ P^b) H^T \left(H(\rho_L \circ P^b) H^T + R \right)^{-1} (y^o - h(x^b)) \quad (2)$$

Schur (element-wise) product

② Observation-space Localization:

$$x^a = x^b + \rho_L \circ (P^b H^T) \left(\rho_L \circ (H P^b H^T) + R \right)^{-1} (y^o - h(x^b)) \quad (3)$$

or in serial EnKF:

$$x_{(j+1)} = x_{(j)} + \rho_{L_j} \circ K_j (y_j^o - \bar{y}_{(j)})$$

$\Rightarrow$ for nonlinear $(H X^b)_{jk} \approx h_j(x_k^b) - \overline{h_j(x^b)}$, ① and ② are very different!

③ B localization: $K = \rho_L \circ (P^b H^T) (H P^b H^T + R)^{-1}$, $\rho_L = e^{-d^2/2L^2}$, (4)

④ R localization: $K = P^b H^T (H P^b H^T + \rho_L' \circ R)^{-1}$, $\rho_L' = e^{d^2/2L^2}$, (5)

$$= \rho_L \circ (P^b H^T) (\rho_L \circ (H P^b H^T) + R)^{-1}$$

Adaptive Localization

Anderson 2007 - hierarchical filters (group filter)

- use a "group" of ensembles to estimate the amount of sampling errors.
- when assimilating one observation, increment δy, is regressed to increments in each state variable, δx

$$\delta \vec{x} = \vec{\beta}\, \delta y \quad (6)$$

- $\vec{\beta}$ is a "correlation map" telling each state variable how to adjust according to observed information δy.

- most sampling errors are from the correlation, not standard deviation; for a given observation and state variable, the $\hat{\beta}$ regression coefficient can be written as $\hat{\beta} = \hat{r}\hat{\sigma}_x / \hat{\sigma}_y$, hat denotes sample estimated

Idea: use m groups of ensembles of size N to obtain not just one but m estimates $\hat{\beta}_i$, $i = 1, 2, \cdots, m$, of the true regression coefficient β.

- Assume the true β is a random draw from the same distribution from which the $\hat{\beta}_i$ are drawn. The optimal localization factor ρ can be found by minimizing "sampling error"

$$\sqrt{\sum_{j=1}^{m} \sum_{i=1,\, i\neq j}^{m} \| \rho \hat{\beta}_i - \hat{\beta}_j \|^2} \qquad (7)$$

- Here $\hat{\beta}_j$ is used in place of the true β, which is unknown. Similar to Houtekamer and Mitchell (1998), who used one ensemble's statistics to update another ensemble, avoiding "inbreeding".
  $\Rightarrow$ bootstrap sampling, similar to Zhang and Oliver (2010)

Anderson 2012: Sampling Error Correction (SEC) algorithm.
Given N, determine the distribution $\mathcal{N}(\bar{r}_N, \sigma_{r,N})$ from which $\hat{r}$ is drawn. Use offline Monte Carlo sampling method: Draw m samples of size N from bivariate normal distribution with covariance $\begin{pmatrix} 1 & r_k \\ r_k & 1 \end{pmatrix}$, and calculate sample-estimated $\hat{r}_{k,m}$, for $k = 1, 2, \cdots K$ each with chosen r_k value from $[-1, 1]$.
Find $\bar{r}_N$, $\sigma_{r,N}$ from these $\hat{r}_{k,m}$ and r_k.

$\Rightarrow$ $\rho\hat{\beta}$ is considered optimal in terms of minimum sampling error,

$$\rho = \frac{Q^2}{1+Q^2}\frac{\bar{r}_N}{\hat{r}}, \quad Q = \frac{\bar{\beta}_N}{\sigma_{\beta,N}}, \quad \bar{\beta}_N = \bar{r}_N \frac{\hat{\sigma}_x}{\hat{\sigma}_y}, \qquad (8)$$

$$\sigma_{\beta,N} = \sigma_{r,N}\frac{\hat{\sigma}_x}{\hat{\sigma}_y}.$$

Anderson and Lei 2013
Lei et al. 2014. 2015 - Empirical Localization Function (ELF)

Idea: find optimal localization by conducting an observing system simulation experiment (OSSE) and minimize the analysis error variance:

Let x be a set of state variable instances archived in an OSSE and y be observations.

For a given distance d between x and y, the update eqn is

$$\delta x = \rho(d)\,\hat{\beta}\,\delta y$$

There are K instances found in OSSE where x and y are separated at distance d, indexed by $k=1, 2, \cdots, K$

Not members here!

$\Rightarrow$ find best $\rho(d)$ by minimizing

$$\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\bar{x}_k + \rho(d)\hat{\beta}_k\,\delta y_k - x_k^{tr}\right)^2} \qquad (9)$$

$$\rho(d) = \sum_{k=1}^{K}(x_k^{tr}-\bar{x}_k)\hat{\beta}_k\,\delta y_k \Big/ \sum_{k=1}^{K}(\hat{\beta}_k\,\delta y_k)^2 \qquad (10)$$

Do the same calculation for a range of d values, then an ELF is obtained. ELF can be calculated for a training period and applied to a system afterwards.

Inflation

When ensemble spread is too small, uncertainties in background state are under-represented. This will cause "filter divergence".

To prevent this, the spread of prior / posterior ensemble can be inflated.

① Multiplicative Inflation:

$$(x_k')_{inf} = x_k' \cdot \lambda \quad , \text{ for } k=1,2,\cdots,N \qquad (1)$$

$\rightarrow P_{inf} = \lambda^2 P$, not changing the covariance structure
→ Does not introduce new directions for analysis increments to take place.

② Additive Inflation:

$$(x_k')_{inf} = x_k' + \varepsilon_k \quad , \text{ for } k=1,2,\cdots,N. \qquad (2)$$

$\varepsilon_k \sim N(0, Q)$ or other distribution.

$\rightarrow P_{inf} = P + Q$ → covariance structure changed.
Introduces new directions for analysis incr.

③ Covariance Relaxation

Zhang et al. 2004: relax-to-prior-perturbation (RTPP).

$$(x_k^{a\prime})_{new} = x_k^{a\prime}(1-\alpha) + x_k^{b\prime}\alpha \quad , \text{ for } k=1,2,\cdots,N \qquad (3)$$

$$(x_k^a)'_{new} = x_k^{a\prime}\left(\alpha\frac{\sigma_b - \sigma_a}{\sigma_a} + 1\right)$$

since ensemble spread reduce after assimilation.

→ relax-to-prior-spread: (RTPS)
Whitaker and Hamill 2012

Adaptive Inflation

Use innovation statistics, one can detect the deficiency in ensemble spread $\rightarrow$ base of adaptive inflation methods.

Amount of inflation needed for prior:

$$\lambda^o = \sqrt{\frac{tr(\mathbb{E}(d^{o-b}(d^{o-b})^T) - R)}{tr(HP^bH^T)}} \quad (4)$$

Problem: sample estimates of the expectation can be noisy when sample size is small.

Anderson 2007, 2009:

Consider d^{o-b} a random draw from $N(0, \lambda^o HP^bH^T + R)$ where λ^o is the expected inflation suggested by d^{o-b}. use this as likelihood function and update λ field with a Bayesian filter

$$p(\lambda | d^{o-b}) \propto p(d^{o-b} | \lambda)\, p(\lambda) \quad (5)$$

$\rightarrow$ assume Gaussian distribution for λ: $p(\lambda) = N(\bar{\lambda}, \sigma_\lambda^2)$

tunable

Miyoshi 2011: Gaussian approximation to Anderson 2009, use (4) to find λ^o and update $\bar{\lambda}^b$:

$$\bar{\lambda}^a = \frac{\bar{\lambda}^b \sigma_\lambda^{o2} + \lambda^o \sigma_\lambda^2}{\sigma_\lambda^{o2} + \sigma_\lambda^2}, \quad \sigma_\lambda^{o2} = \frac{2}{p}\left(\frac{\bar{\lambda}^b tr(HP^bH^T + R)}{tr(HP^bH^T)}\right)^2$$

Simplification: use smoothing time scale τ: $\bar{\lambda}^a = \bar{\lambda}^b + \frac{\lambda^o - \bar{\lambda}^b}{\tau}$.

Ying and Zhang 2015: calculate $\lambda^o = (\alpha \frac{\bar{\sigma}_b - \bar{\sigma}_a}{\bar{\sigma}_a} + 1)$ and find adaptive α for RTPS.