A theory for why even simple localization is so useful in ensemble data assimilation

Matthias Morzfeld, Scripps Institution of Oceanography, UC San Diego
Daniel Hodyss, Naval Research Laboratory

UC San Diego | SCRIPPS INSTITUTION OF OCEANOGRAPHY | U.S. NAVAL RESEARCH LABORATORY

MM is supported by the US Office of Naval Research (ONR) grant N00014-21-1-2309. **DH** is supported by the US Office of Naval Research (ONR) grant N0001422WX00451.

SCAN ME

Summary

Motivation. Covariance localization has been the key to the success of ensemble data assimilation in high dimensional problems, especially in global numerical weather prediction.

Idea. We review and synthesize optimal and adaptive localization methods that are rooted in sampling error theory and that are defined by optimality criteria, e.g., minimizing errors in forecast covariances or in the Kalman gain.

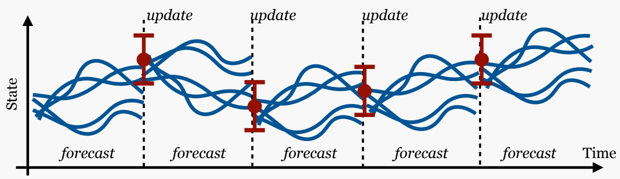


Illustration of a cycling EnKF. The ensemble forecast (blue) is updated in view of observations (red) and associated errors (error bars).

Ensemble Kalman filter (EnKF)

The ensemble Kalman filter (EnKF) refers to a class of algorithms that update the forecast of a numerical model in view of an observation

- Use an ensemble to describe the state of the atmosphere and its associated uncertainty
- Update each ensemble member in view of the observation $\mathbf{y}$

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}(\mathbf{y} - (\mathbf{H}\mathbf{x}_i^f + \boldsymbol{\eta})),$$
$$\mathbf{K} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1},$$

where $\mathbf{x}_i^f$ are the forecast ensemble members, $\mathbf{H}$ is a linear or linearized observation operator, $\mathbf{R}$ is the observation error covariance, $\boldsymbol{\eta}$ is a draw from a normal distribution with mean zero and covariance $\mathbf{R}$, $\mathbf{P}$ is the forecast covariance, and $\mathbf{K}$ is the Kalman gain

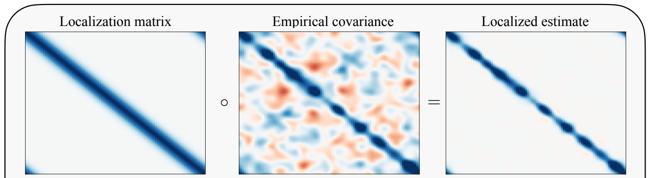


Illustration of Schur-product-localization. The Schur product of the localization matrix with the empirical covariance matrix is the localized estimate.

Localization

- The ensemble size n_e is much smaller than the number of state variables, n_x, or the number of observations, n_y ($n_e \ll n_y < n_x$)
- The small ensemble size renders the empirical estimate of the forecast covariance $\mathbf{P}$ inaccurate and, hence, the Kalman gain and EnKF are also inaccurate
- Localization is a suite of methods that boosts the accuracy of the forecast covariance by enforcing on the empirical covariance the assumption that correlations decay with distance
- The localized estimate after Schur-product-localization is

$$\mathbf{P}_{\text{loc}} = \mathbf{L} \circ \mathbf{P},$$

where $\mathbf{L}$ is a symmetric positive semi-definite localization matrix (Gaspari-Cohn) that dampens spurious, far-field covariances

- Within an ensemble adjustment Kalman filter (EAKF), we localize the Kalman gain

$$\mathbf{K}_{j,\text{loc}} = \mathbf{l}_j \circ \mathbf{K}_j,$$

where $\mathbf{l}_j$ is a vector that reduces the effects of far-away observations, $\mathbf{K}_j = \mathbf{P}\mathbf{h}_j^T(\mathbf{h}_j\mathbf{P}\mathbf{h}_j^T + r_j)^{-1}$ is the Kalman gain associated with the jth observation, and where r_j is the associated observation error variance

Optimal localization

- How do we construct localization schemes? We search for localization schemes that minimize errors

Prior optimal localization

- Find the best Schur product estimator that minimizes

$$\min_{\mathbf{L}} \langle \,||\mathbf{L} \circ \mathbf{P} - \mathbf{P}_{\text{true}}||_{\text{Fro}}^2 \rangle \rightarrow \mathbf{L}_{ij}^{\text{opt}} = \frac{\rho_{ij}^2(n_e - 1)}{1 + \rho_{ij}^2 n_e},$$

where ρ_{ij} is the correlation between components i and j, and where $\langle \cdot \rangle$ is an average (Morzfeld & Hodyss, 2023; Ménétrier et al., 2015)

Flowerdew's optimal localization

- Flowerdew (2015) derives the optimal localization (similar assumptions as above)

$$\mathbf{L}_{ij} = \frac{\rho_{ij}^2}{\rho_{ij}^2 + \sigma_s},$$

where σ_s is computed from ρ_{ij} and n_e

Posterior optimal localization

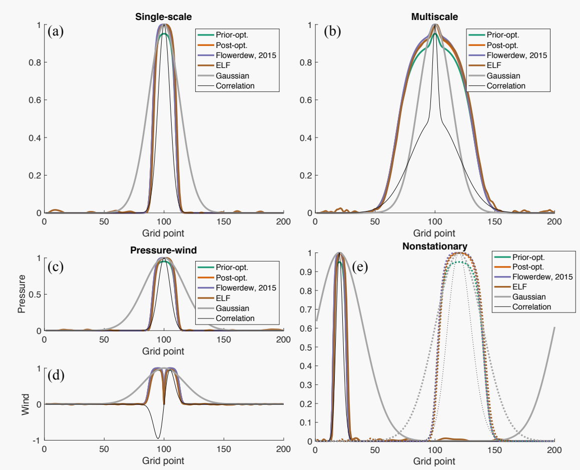
- We find the best approximation to the Kalman gain within an EAKF

$$\min_{\alpha} \langle (\alpha K - K_{\text{true}})^2 \rangle$$

- Using a Taylor expansion, we find (Morzfeld & Hodyss, 2023)

$$\alpha_{\text{opt}} = \frac{\rho_{ij}^2(n_e - 1 + 2\sigma_y\beta(\sigma_y^2\beta - 1))}{1 + \rho_{ij}^2(n_e + 2\sigma_y^2\beta(3\sigma_y^2\beta - 4))},$$

where r is the observation error, $\sigma_y^2 = \mathbf{h}\mathbf{P}\mathbf{h}^T$ and $\beta = 1/(\sigma_y^2 + r)$



Comparison of several optimal/adaptive localization schemes. All methods damp far-field correlations and all methods are surprisingly similar.

Universal localization law

- All optimal localizations follow a universal law

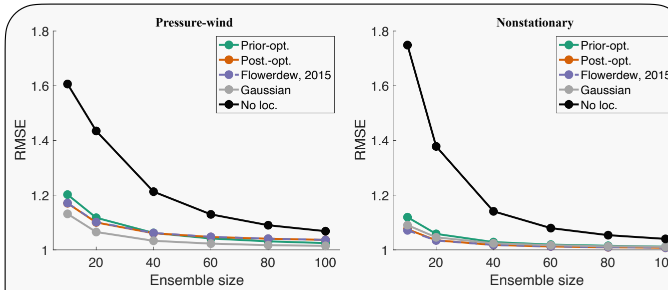
$$\alpha_{\text{opt}} = \frac{c_1\rho_{ij}^2}{1 + c_2\rho_{ij}^2},$$

where the constants c_1 and c_2 define the "*type of optimality*"

- Localization schemes based on this "*universal*" law may be ***broadly applicable and useful***

Numerical tests

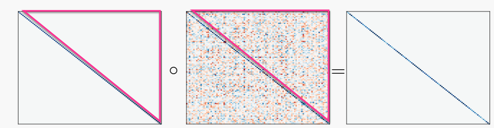
- We applied optimal and sub-optimal localization methods to four covariance matrices
- All schemes perform very similarly in terms of analysis RMSE of an EnKF update



Summary of numerical tests. All localization methods lead to a very similar analysis RMSE (scaled by infinite ensemble size RMSE) on four different covariance types (two are not shown).

Analysis

- We analyze errors of localization schemes with the help of the canonical example of a high-dimensional isotropic Gaussian (due to Peter Bickel and others)
- The analysis reveals that deleting far-field correlations drastically reduces errors, but correlations in the near-field are a second-order effect (theory in a 12 page Appendix)



Optimal localization *vs.* ***No localization***

$$||\hat{\mathbf{P}}_{\text{polo}} - \mathbf{P}||_{\text{Fro}}^2 = \underbrace{\frac{2n_x}{n_e + 1}}_{\text{diagonal}} \propto \frac{n_x}{n_e}$$

$$||\hat{\mathbf{P}} - \mathbf{P}||_{\text{Fro}}^2 = \underbrace{\frac{2n_x}{n_e - 1}}_{\text{diagonal}} + \underbrace{\frac{n_x(n_x - 1)}{n_e - 1}}_{\text{off-diagonal}} = \frac{n_x(n_x + 1)}{n_e - 1} \propto \frac{n_x^2}{n_e}$$

$$\langle \text{MSE}_{\text{post-opt}} \rangle - \langle \text{MSE}_{\text{KF}} \rangle = \frac{1}{n_x}\frac{1}{4(n_e - 1)}$$

$$\langle \text{MSE}_{\text{unloc}} \rangle - \langle \text{MSE}_{\text{KF}} \rangle = \frac{1}{4(n_e - 1)}$$

Results

1. All optimal localization methods follow a universal law and are quite similar.
2. We confirm the similarity of the various schemes in idealized numerical experiments, where we observe that all localization schemes we test – optimal and non-adaptive schemes – perform quite similarly in a wide-array of problems.
3. We explain this surprising finding with mathematical rigor on an idealized class of problems, first put forward by Bickel and others to study the collapse of particle filters.
4. The numerical experiments and the theory show that the most important attribute of a localization scheme is the *well-known property that one should dampen spurious long-range correlations. The details of the correlation structure have a second order effect.*

References

1. Anderson and Lei (2013). *Empirical localization of observation impact in ensemble Kalman filters*. Monthly Weather Review 141:4140–4153.
2. Flowerdew (2015). *Towards a theory of optimal localisation*. Tellus A 67(1):25257.
3. Hodyss and Morzfeld (2023), *How sampling errors in covariance estimates cause bias in the Kalman gain and impact ensemble data assimilation*, Monthly Weather Review 151(9):2413- 2426.
4. Ménétrier, Montmerle, Michel, and Berre (2015). *Linear filtering of sample co- variances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization.* Monthly Weather Review 143:1622–1643.
5. Morzfeld and Hodyss (2023), *A theory for why even simple covariance localization is so useful in ensemble data assimilation*, Monthly Weather Review 151(3):717-736.