Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Localization: Theory and application

Hendrik Reich, África Periáñez, Roland Potthast

DWD, Germany

COSMO General meeting 2012, Lugano

Table of contents

Our problem

- Understand the basic properties of localization in the ensemble Kalman filter scheme.
- Find an adaptive localization scheme depending on the density of data, observation error, ...
- Decomposition of the error sources to determine its effect on the optimal localization length scale.
- We start with a brief description of ensemble Kalman filtering from a mathematical point of view, followed by
- numerical experimental results

Cost function and update formula

The *cost function* to be minimized is

$$J(\varphi) := \|\varphi - \varphi^{(b)}\|^2_{B^{-1}} + \|f - H\varphi^{(b)}\|^2_{R^{-1}}, \quad (1)$$

where $\varphi^{(b)}$ is the *background state*, f are the *data*, H is the *observation operator* and the relation between variables at different points is incorporated by the covariance matrices B and R. Minimizing the cost function gives the *update formula*

$$\varphi^{(a)} = \varphi^{(b)} + BH^*(R + HBH^*)^{-1}(f - H\varphi^{(b)}) \quad (2)$$

Ensemble Kalman Filter

In the EnKF methods the background convariance matrix is represented by $B^{(ens)} := \frac{1}{L-1} Q_k Q_k^*$. The ensemble matrix Q_k is defined as

$$Q_k := \left(\varphi_k^{(1)} - \overline{\varphi}_k^{(b)}, ..., \varphi_k^{(L)} - \overline{\varphi}_k^{(b)} \right), \qquad (3)$$

where $\overline{\varphi}^{(b)}$ denotes the mean $\frac{1}{L} \sum_{l=1}^{L} \varphi^{(l)}$.
Thus, we solve the update in a low-dimensional subspace

$$U^{(L)} := \text{span}\{\varphi_k^{(1)} - \overline{\varphi}_k^{(b)}, ..., \varphi_k^{(L)} - \overline{\varphi}_k^{(b)}\}. \qquad (4)$$

The update formula now is

$$\varphi_k^{(a)} = \varphi_k^{(b)} + Q_k Q_k^* H^* (R + H Q_k Q_k^* H^*)^{-1} (f_k - H\varphi_k^{(b)}) \qquad (5)$$

The updates of the EnKF are a linear combination of the columns of Q_k. We can therefore write

$$\varphi_k - \varphi_k^{(b)} = \sum_{l=1}^{L} \gamma_l \left(\varphi_k^{(l)} - \overline{\varphi_k^{(b)}} \right) = Q_k \gamma \qquad (6)$$

With

$$\widehat{Q}_k := H Q_k, \qquad (7)$$

the resulting the expresion to minimize is

$$J(\gamma) := \| Q_k \gamma \|^2_{B_k^{-1}} + \| f_k - H\varphi_k^{(b)} - \widehat{Q}_k \gamma \|^2_{R^{-1}}, \qquad (8)$$

Error analysis without background contribution

Lemma

Assume that H is injective, that we study true measurement data $f = H\varphi^{(true)}$ and consider the EnKF with data term only

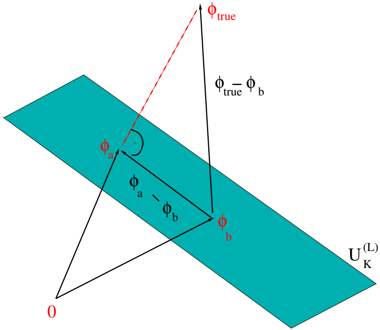
$$J^{(data)}(\gamma) = \|(f - H\varphi^{(b)}) - \hat{Q}_k \gamma\|_{R^{-1}}^2 \quad (9)$$

Then, for the analysis $\varphi^{(a)}$ calculated by the EnKF the difference $\varphi^{(a)} - \varphi^{(b)}$ is the orthogonal projection of $\varphi^{(true)} - \varphi^{(b)}$ onto the ensemble space $U_k^{(L)}$ and the analysis error is given by

$$E_k = d_{H^* R^{-1} H}\Big(U_k^{(L)}, \varphi_k^{(true)} - \varphi^{(b)}\Big), \quad (10)$$

where the right-hand side denotes the distance between a point $\psi = \varphi_k^{(true)} - \varphi^{(b)}$ and the subspace $U^{(L)}$ with respect to the norm induced by the scalar product $< ., . >_{H^* R^{-1} H}$.

Illustration of Lemma



Error analysis with background term

Theorem

Assume that H is injective, that we study true measurement data $f = H\varphi^{(true)}$ and consider an assimilation step using the EnKF. Then, for the analysis error in the step k we have the analysis error estimate

$$\|\varphi_k^{(true)} - \varphi^{(b)}\|_{H^*R^{-1}H} \geq E_k \geq d_{H^*R^{-1}H}\Big(U_k^{(L)}, \varphi_k^{(true)} - \varphi_k^{(b)}\Big). \tag{11}$$

Localization

LETKF basic idea: Localization to D, leading to

$$Q_{k,loc} := \left(\chi_D(\varphi_k^{(1)} - \overline{\varphi}_k^{(b)}), ..., \chi_D(\varphi_k^{(L)} - \overline{\varphi}_k^{(b)}) \right). \qquad (12)$$

We now have

$$B = \frac{1}{L-1} Q_{k,loc} Q_{k,loc}^T \qquad (13)$$

and

$$f_{k,loc} = \chi_D f_k \qquad (14)$$

We now solve the equations in the locally low-dimensional subspace

$$U_k^{(L,D)} := \text{span}\{\chi_D(\varphi_k^{(1)} - \overline{\varphi}_k^{(b)}), ..., \chi_D(\varphi_k^{(L)} - \overline{\varphi}_k^{(b)})\}. \qquad (15)$$

Localization

Thus, in the above error estimates we just have to replace

$$E_k \to E_{k,loc}$$
$$(\varphi_k^{(true)} - \varphi^{(b)}) \to \chi_D(\varphi_k^{(true)} - \varphi^{(b)}) \tag{16}$$
$$U_k^{(L)} \to U_{k,loc}^{(L)}.$$

to get the local error estimates.

Localization

Theorem

Assume that there is $c,\ C > 0$ *such that for all* $x \in D$ *there is* $l \in 1, ..., L$ *such that*

$$| \varphi^{(l)}(x) | \geq c, \tag{17}$$

and that

$$\left\| \nabla \varphi^{(l)}(x) \right\|_{\infty} \leq c, \quad x \in D. \tag{18}$$

Then with sufficiently rich data and the true solution in $H(D)$ *we have*

$$\sup_{x \in D} E_{k,loc}(x, \rho) \to 0, \quad \rho \to 0. \tag{19}$$

1d toy model

- 1d model without cycling, uses least-square estimate to obtain an analysis (LSA). The truth is given by a (higher order) function.
- Either “pure” least square estimate without background (free LSA), or correction of a background state (bg LSA)
- Observations are generated from the truth with a specified observation error σ_{obs}.
- here, the analysis is given by straight lines $a + bx$ where a, b are estimated from the observations.
- do this *globally* using all available observations or *step by step* in several intervals using a local subset of observations.
- The use of straight lines in some sense mimics the behaviour of an ensemble method, which also tries to approximate a high order state within a (lower order) subspace spanned by the background ensemble members (detailed later).

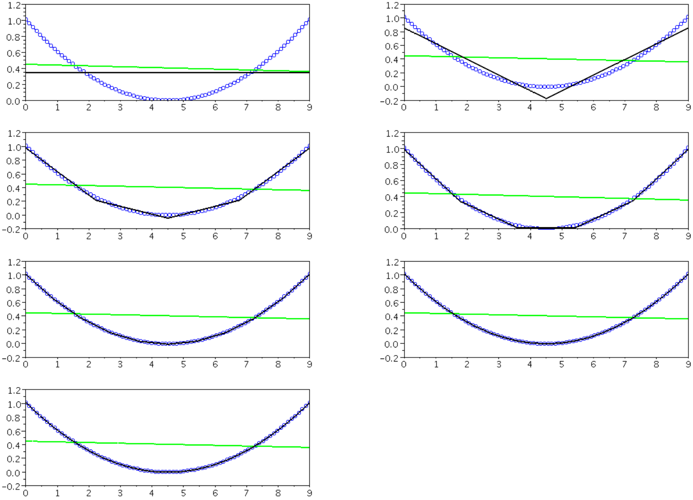


Fig.1: truth (blue line), observations (blue circles), background (green), free LSA (red) and bg LSA (black) for the set of localization radii, $\sigma_{obs} = 0.0005$

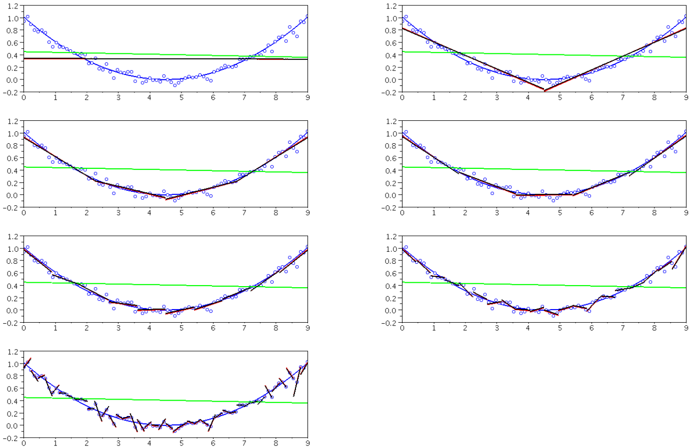


Fig.2: same as Fig.1, but for $\sigma_{obs} = 0.05$

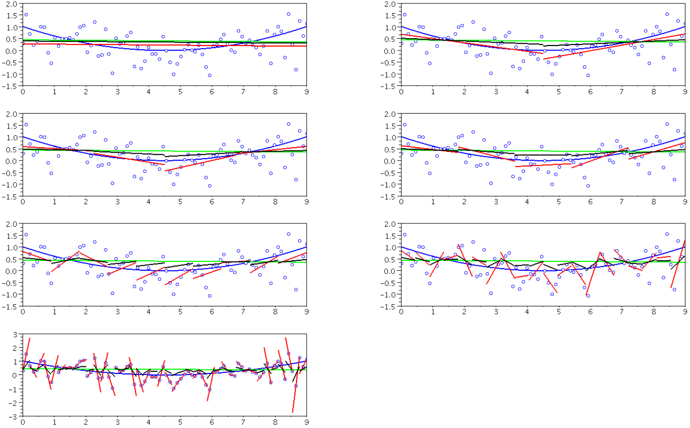


Fig.3: same as Fig. 1, but for $\sigma_{obs} = 0.5$

Results

- for all values of σ_{obs} the bg LSA is better than the first guess.
- For large σ_{obs} the free LSA is worse than the bg analysis.
- for small σ_{obs} the results of the free and the bg LSA become very similar.
- the optimal value of ρ_{loc} moves to smaller values with decreasing σ_{obs}.

Optimal localization radius

- estimate the optimal localization radius ρ_{loc} as a function of σ_{obs} and observation density d for the free analysis.
- two error sources: *approximation error* and *sampling error*.
- *approximation error* should decrease with smaller localization radii as a higher order function can be better approximated by a large number of straight lines. $\sim \rho_{loc}^2$ (theorems on numerical interpolation)
- *sampling error* should decrease with larger localization radii as a larger number of observations gives a statistical better estimate. $\sim 1/\sqrt{N_{obs}}$, where N_{obs} is the number of observations.
- N_{obs} can be expressed as $N_{obs} = \int\limits_V d(x)dV = 2d\rho_{loc}$.

$$\hat{e} \sim \rho_{loc}^2 \quad + \quad \frac{\sigma_{obs}}{\sqrt{2d\rho_{loc}}}, \tag{20}$$

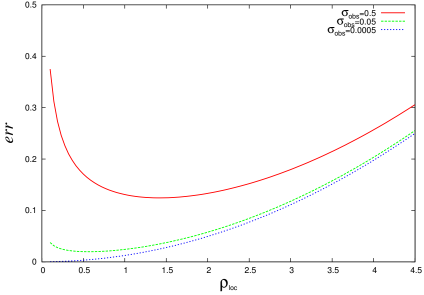
Optimal localization radius

The minimum of this error (as a function of the localization radius ρ_{loc}) can be obtained, leading to

$$\rho_{loc}^{opt} \sim \left(\frac{\alpha}{4}\right)^{2/5}, \tag{21}$$

where $\alpha = (\sigma_{obs}/\sqrt{2d})$.
Thus, ρ_{loc}^{opt} as a function of σ_{obs} can be described by

$$\rho_{loc}^{opt} \sim \sigma_{obs}^{2/5}, \tag{22}$$



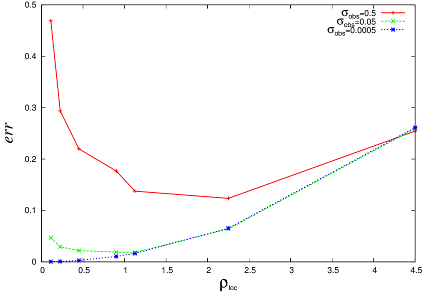
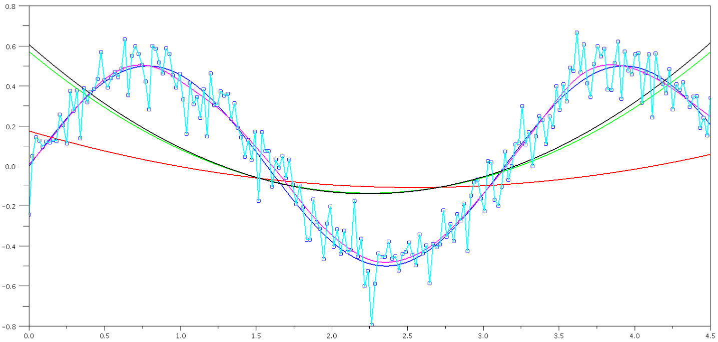


Fig.4: theoretical and numerical results for error as a function of ρ_{loc}, $\sigma_{obs} = [0.0005\, 0.05\, 0.5]$.

the optimal value of ρ_{loc} moves to smaller values with decreasing σ_{obs}.

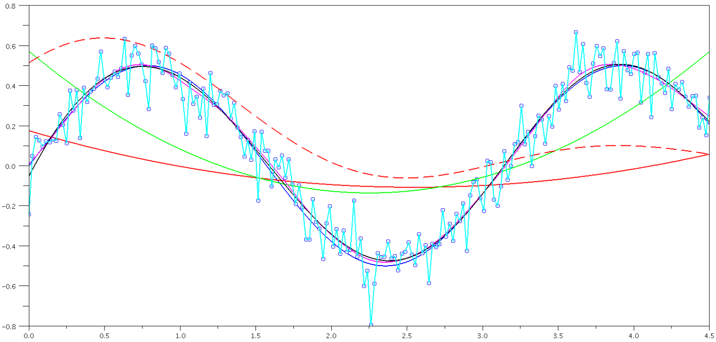
- when is LETKF similar to (bg) LSA?
  - “polynomial order” of fg-ens members and LSA background base functions ($:= N_P$) should be the same; additionally we need $(N_{ens} - 1) \geq N_P$
- LETKF cannot “fit” more than N_{ens} observations, but we have to distinguish two cases:
  - if N_P and N_{ens} are comparable to “order” of the truth (ensemble subspace is good approximation to truth $\rightarrow$ *approximation error* small), the error will decrease $\sim \frac{1}{\sqrt{N_{obs}}}$ even if $N_{obs} > (N_{ens} - 1)$ (*sampling error*)
  - if LETKF subspace is too small/not appropiate (model error?); *approximation error* dominates, additional obs don't have positive impact for $N_{obs} > (N_{ens} - 1)$



o o o observations —— truth —— LSA/LETKF/3dVar background
—— LSA analysis —— LETKF —— 3dVar —— spline

Fig.5 $\sigma_{obs} = 0.1,\ N_{ens} = 10,\ N_P = 3$ in LETKF bg ens

LETKF similar to bg LSA; 3dVar ana is best



o o o observations —— truth —— LSA/3dVar background
- - - LETKF background —— LSA analysis —— LETKF —— 3dVar —— spline

Fig.6: $\sigma_{obs} = 0.1$, $N_{ens} = 10$, sin/cos base functions as LETKF bg ens

LETKF similar to 3dVar

adaptive horizontal localization

- localization length scales depend on weather situation, observation density ...
- simple adaptive method: keep number of *effective observations* fixed, vary localization radius (*effective observations*: sum of **observation weights**)
- up to now only implemented in horizontal direction
- one has to define minimum / maximum radius, number of *effective observations* $N_{obs}^{eff} = \alpha(N_{ens} - 1)$, $\alpha \geq 1$
- ideal number of effective observations depends on ensemble size, ...
- Christoph already showed first results

Outlook / Conclusion

- 1d model: optimal localization length ρ_{loc} depends on σ_{obs}; this (first results) also seems to be the case for the L95-LETKF.
- 1d model: 2-step ana gives better results if two obs types and $\sigma^1_{obs} >> \sigma^2_{obs}$.
- 1d model: for fixed ρ_{loc} in LETKF: $N_{obs} > (N_{ens} - 1)$ gives better results only if ensemble-subspace is appropiate
- 2d model LETKF: similar results found
- "classical" view on localization in EnKF: up to which distance can we trust the correlations in the ensenmble?
- (How) are both approaches connected? Do they lead to similar optimal localization radii? Should be tested (L95-LETKF).

Outlook / Conclusion

- COSMO: conventional data with large ρ_{loc} to get large scale analysis increments, radar data in second analysis step with small ρ_{loc} to get small scale variations. Maybe 3rd step to get *nonlocal* radiance observations (without vertical localization).
- with one analysis step only, different kinds of obs with different observation density d_{obs}:
  - d_{obs} "high" $\rightarrow \rho_{loc}$ "small", d_{obs} low $\rightarrow \rho_{loc}$ large.
- in order to save time: reduced grid (weights) can be different in the analysis steps. Problem: 4d-aspect, observation operators in COSMO-model. Linear approximation (as in obs impact studies): $Y_a = Y_b W$.
- next steps: investigate influence of observation density, nonlocal observations within 2d-model / L95-LETKF