A Framework for Adaptive Inflation and Covariance Localization for Ensemble Filters

Ahmed Attia [1] Emil Constantinescu [1,2]

[1] Mathematics and Computer Science Division (MCS), Argonne National Laboratory (ANL)

[2] The University of Chicago, Chicago, IL

SIAM Conference on Mathematics of Planet Earth (MPE18)
September 14, 2018

Outline

Outline

Ensemble Kalman Filter (EnKF)

Assimilation cycle over $[t_{k-1}, t_k]$; **Forecast step**

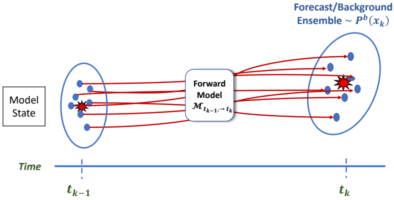
- **Initialize:** an analysis ensemble $\{\mathbf{x}^{\rm a}_{k-1}(e)\}_{e=1,\ldots,\mathrm{N_{ens}}}$ at t_{k-1}
- **Forecast:** use the discretized model $\mathcal{M}_{t_{k-1}\rightarrow t_k}$ to generate a forecast ensemble at t_k:

$$\mathbf{x}^{\rm b}_k(e) = \mathcal{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^{\rm a}_{k-1}(e)) + \eta_k(e), \quad e = 1, \ldots, \mathrm{N_{ens}}$$

- *Forecast/Prior statistics*:

$$\overline{\mathbf{x}}^{\rm b}_k = \frac{1}{\mathrm{N_{ens}}} \sum_{e=1}^{\mathrm{N_{ens}}} \mathbf{x}^{\rm b}_k(e)$$

$$\mathbf{B}_k = \frac{1}{\mathrm{N_{ens}}-1} \mathbf{X}^{\rm b}_k \left(\mathbf{X}^{\rm b}_k\right)^{\mathsf{T}} ; \quad \mathbf{X}^{\rm b}_k = [\mathbf{x}^{\rm b}_k(1) - \overline{\mathbf{x}}^{\rm b}_k, \ldots, \mathbf{x}^{\rm b}_k(\mathrm{N_{ens}}) - \overline{\mathbf{x}}^{\rm b}_k]$$



Ensemble Kalman Filter (EnKF)

Assimilation cycle over $[t_{k-1}, t_k]$; **Analysis step**

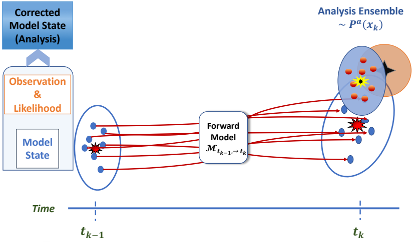
- Given an observation $\mathbf{y}_k$ at time t_k
- **Analysis:** sample the posterior (EnKF update)

$$\mathbf{K}_k = \mathbf{B}_k \mathbf{H}_k^\mathsf{T} \left(\mathbf{H}_k \mathbf{B}_k \mathbf{H}_k^\mathsf{T} + \mathbf{R}_k \right)^{-1}$$

$$\mathbf{x}_k^\mathrm{a}(e) = \mathbf{x}_k^\mathrm{b}(e) + \mathbf{K}_k \left([\mathbf{y}_k + \zeta_k(e)] - \mathcal{H}_k(\mathbf{x}_k^\mathrm{b}(e)) \right)$$

- *The posterior (analysis) error covariance matrix*:

$$\mathbf{A}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{B}_k \equiv \left(\mathbf{B}_k^{-1} + \mathbf{H}_k^\mathsf{T} \mathbf{R}^{-1} \mathbf{H}_k \right)^{-1}$$

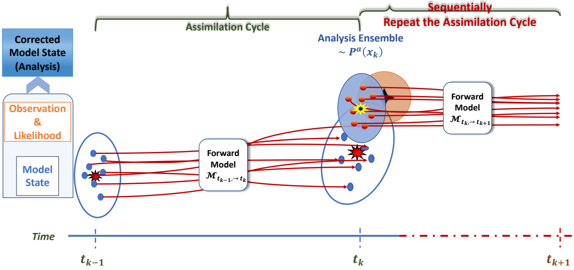


Ensemble Kalman Filter (EnKF)

Sequential EnKF Issues

- Limited-size ensemble results in sampling errors, explained by:
  - **variance underestimation**
  - accumulation of **long-range spurious correlations**
  - filter divergence after a few assimilation cycles

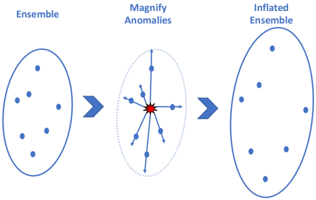
- *EnKF requires* **inflation & localization**



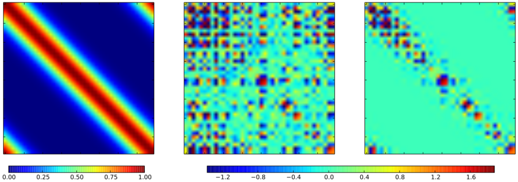
Ensemble Kalman Filter (EnKF)

Inflation & Localization

- Covariance underestimation in EnKF is counteracted, by applying covariance inflation:
  - $\rightarrow$ replace $\mathbf{B}$, with an inflated version $\widetilde{\mathbf{B}}$



- Long-range spurious correlations are reduced by covariance localization (e.g., Schur-product)
  - $\rightarrow$ replace $\mathbf{B}$, with a decorrelated version $\widehat{\mathbf{B}}$



EnKF: Inflation

- *Additive Inflation:*

$$\widetilde{\mathbf{B}} := \mathbf{D} + \mathbf{B}; \quad \text{s.t. } \mathbf{D} = \operatorname{diag}(\boldsymbol{\lambda}) \,,\; \boldsymbol{\lambda} = \left(\lambda_1, \lambda_2, \ldots, \lambda_{\mathrm{N_{state}}}\right)^{\mathsf{T}} ,\; 0 \leq \lambda_i^l \leq \lambda_i \leq \lambda^{\mathrm{u}}$$

- *Multiplicative Inflation:*
  1. **Space-independent inflation:**

$$\widetilde{\mathbf{X}^{\mathrm{b}}} = \left[\sqrt{\lambda}\left(\mathbf{x}^{\mathrm{b}}(1) - \overline{\mathbf{x}}^{\mathrm{b}}\right), \ldots, \sqrt{\lambda}\left(\mathbf{x}^{\mathrm{b}}(\mathrm{N_{ens}}) - \overline{\mathbf{x}}^{\mathrm{b}}\right)\right]; \; 0 < \lambda^l \leq \lambda \leq \lambda^u$$

$$\widetilde{\mathbf{B}} = \frac{1}{\mathrm{N_{ens}} - 1} \widetilde{\mathbf{X}^{\mathrm{b}}} \left(\widetilde{\mathbf{X}^{\mathrm{b}}}\right)^{\mathsf{T}} = \lambda \, \mathbf{B}$$

  2. **Space-dependent inflation:** Let $\mathbf{D} := \operatorname{diag}(\boldsymbol{\lambda}) \equiv \sum_{i=1}^{\mathrm{N_{state}}} \lambda_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$,

$$\widetilde{\mathbf{X}^{\mathrm{b}}} = \mathbf{D}^{\frac{1}{2}} \mathbf{X}^{\mathrm{b}} \,,$$

$$\widetilde{\mathbf{B}} = \frac{1}{\mathrm{N_{ens}} - 1} \widetilde{\mathbf{X}^{\mathrm{b}}} \left(\widetilde{\mathbf{X}^{\mathrm{b}}}\right)^{\mathsf{T}} = \mathbf{D}^{\frac{1}{2}} \mathbf{B} \mathbf{D}^{\frac{1}{2}} \,.$$

- The *inflated* Kalman gain $\widetilde{\mathbf{K}}$, and analysis error covariance matrix $\widetilde{\mathbf{A}}$

$$\widetilde{\mathbf{K}} = \widetilde{\mathbf{B}} \mathbf{H}^{\mathsf{T}} \left(\mathbf{H} \widetilde{\mathbf{B}} \mathbf{H}^{\mathsf{T}} + \mathbf{R}\right)^{-1}; \quad \widetilde{\mathbf{A}} = \left(\mathbf{I} - \widetilde{\mathbf{K}} \mathbf{H}\right) \widetilde{\mathbf{B}} \equiv \left(\widetilde{\mathbf{B}}^{-1} + \mathbf{H}^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{H}\right)^{-1}$$

EnKF: Schur-Product Localization

State-space formulation; $\mathbf{B}-$Localization

- *space-independent covariance localization:*

$$\widehat{\mathbf{B}} := \mathbf{C} \odot \mathbf{B}; \quad \text{s.t. } \mathbf{C} = [\rho_{i,j}]_{i,j=1,2,\ldots,\mathrm{N_{state}}}$$

  - Entries of $\mathbf{C}$ are created using space-dependent localization functions †:

→ Gauss:

$$\rho_{i,j}(L) = \exp\left(\frac{-d(i,j)^2}{2L^2}\right); \quad i,j = 1,2,\ldots,\mathrm{N_{state}}\,,$$

→ 5th-order Gaspari-Cohn:

$$\rho_{i,j}(L) = \begin{cases} -\frac{1}{4}\left(\frac{d(i,j)}{L}\right)^5 + \frac{1}{2}\left(\frac{d(i,j)}{L}\right)^4 + \frac{5}{8}\left(\frac{d(i,j)}{L}\right)^3 - \frac{5}{3}\left(\frac{d(i,j)}{L}\right)^2 + 1\,, & 0 \le d(i,j) \le L \\ \frac{1}{12}\left(\frac{d(i,j)}{L}\right)^5 - \frac{1}{2}\left(\frac{d(i,j)}{L}\right)^4 + \frac{5}{8}\left(\frac{d(i,j)}{L}\right)^3 + \frac{5}{3}\left(\frac{d(i,j)}{L}\right)^2 - 5\left(\frac{d(i,j)}{L}\right) + 4 - \frac{2}{3}\left(\frac{L}{d(i,j)}\right)\,, & L \le d(i,j) \le 2L \\ 0\,. & 2L \le d(i,j) \end{cases}$$

---

†

- $d(i,j)$: distance between ith and jth grid points
- $\mathbf{L}$: radius of influence, i.e. localization radius

EnKF: Schur-Product Localization

Space-dependent formulation; $\mathbf{B}-$Localization

- *Space-dependent radii, i.e.,* $L \equiv L(i,j)$*:* we need to define localization kernel $\mathbf{C}$
  $\rightarrow$ Examples include [†]:

$$\mathbf{C} := \begin{cases} \mathbf{C}_r = \left[\rho_{i,j}(l_i)\right]_{i,j=1,2,\ldots,\mathrm{N_{state}}} \\ \mathbf{C}_c = \left(\mathbf{C}_r\right)^\mathsf{T} = \left[\rho_{i,j}(l_j)\right]_{i,j=1,2,\ldots,\mathrm{N_{state}}} \\ \frac{1}{2}\left(\mathbf{C}_r + \mathbf{C}_c\right) = \left[\frac{1}{2}\rho_{i,j}(l_i) + \rho_{i,j}(l_j)\right]_{i,j=1,2,\ldots,\mathrm{N_{state}}} \\ \mathbf{C}_d = \left[\rho_{i,j}(l_{\min\,(i,j)})\right]_{i,j=1,2,\ldots,\mathrm{N_{state}}} \\ \mathbf{C}_u = \left[\rho_{i,j}(l_{\max\,(i,j)})\right]_{i,j=1,2,\ldots,\mathrm{N_{state}}} \\ \frac{1}{2}\left(\mathbf{C}_d + \mathbf{C}_u\right) = \left[\frac{1}{2}\rho_{i,j}(l_{\min\,(i,j)}) + \rho_{i,j}(l_{\max\,(i,j)})\right]_{i,j=1,2,\ldots,\mathrm{N_{state}}} \\ \mathbf{C}_G = \left[\rho_{i,j}\left(\sqrt{l_i\, l_j}\right)\right]_{i,j=1,2,\ldots,\mathrm{N_{state}}} \end{cases}$$

- **We focus here on the symmetric kernel:**

$$\mathbf{C} := \frac{1}{2}\left(\mathbf{C}_r + \mathbf{C}_c\right) = \frac{1}{2}\left[\rho_{i,j}(l_i) + \rho_{i,j}(l_j)\right]_{i,j=1,2,\ldots,\mathrm{N_{state}}}$$

†Ahmed Attia, and Emil Constantinescu. "An Optimal Experimental Design Framework for Adaptive Inflation and Covariance Localization for Ensemble Filters." arXiv preprint arXiv:1806.10655 (2018).

EnKF: Schur-Product Localization

Space-dependent formulation; $\mathbf{R}-$Localization

- **Localization in observation space ($\mathbf{R}-$localization):**
  - $\mathbf{HB}$ is replaced with $\widehat{\mathbf{HB}} = \mathbf{C}^{\mathrm{loc},1} \odot \mathbf{HB}$, where

    $$\mathbf{C}^{\mathrm{loc},1} = \left[\rho_{i,j}^{o|m}\right] \; ; \, i = 1, 2, \ldots \mathrm{N}_{\mathrm{obs}} \, ; \, j = 1, 2, \ldots \mathrm{N}_{\mathrm{state}}$$

  - $\mathbf{HBH}^{\mathsf{T}}$ can be replaced with $\widehat{\mathbf{HBH}^{\mathsf{T}}} = \mathbf{C}^{\mathrm{loc},2} \odot \mathbf{HBH}^{\mathsf{T}}$, where

    $$\mathbf{C}^{\mathrm{loc},2} \equiv \mathbf{C}^{o|o} = \left[\rho_{i,j}^{o|o}\right] \; ; \, i, j = 1, 2, \ldots \mathrm{N}_{\mathrm{obs}}$$

    - $\rho_{i,j}^{o|m}$ is calculated between the ith observation grid point and the jth model grid point.
    - $\rho_{i,j}^{o|o}$ is calculated between the ith and jth observation grid points.

- Assign radii to state grid points vs. observation grid points:
  - Let $\mathbf{L} \in \mathbb{R}^{\mathrm{N}_{\mathrm{obs}}}$ to model grid points, and project to observations for $\mathbf{C}^{\mathrm{loc},2}$ [hard/unknown]
  - Let $\mathbf{L} \in \mathbb{R}^{\mathrm{N}_{\mathrm{obs}}}$ to observation grid points; [efficient; followed here]

Inflation & Localization

Tuning the parameters

- Tuning the inflation parameter/factors $\boldsymbol{\lambda}$
  - Bayesian approach for adaptive inflation exists, and still requires improvements
  - mostly for uncorrelated observation errors
- Tuning the localization radii of influence $\mathbf{L}$
  - adaptive localization approaches are limited, especially in the vertical
  - mostly for uncorrelated observation errors
  - expert knowledge, especially with observation system, is required
  - theory is lacking

The parameters $\boldsymbol{\lambda}$, $\mathbf{L}$ are generally tuned empirically!

Outline

Optimal Experimental Design (OED)

An OED problem seeks an optimal design $\mathbf{w}$ that solves

$$\min_{\boldsymbol{\lambda}\in\mathbb{R}^{\mathrm{N_{state}}}} \Psi^{\mathrm{OED}}(\mathbf{w}) + \alpha\,\Phi(\mathbf{w})$$

$$\text{subject to} \quad \mathbf{w}^{l} \leq \mathbf{w} \leq \mathbf{w}^{u}$$

- $\Psi^{\mathrm{OED}(\mathbf{w})}$ is the specific design criterion
  - For sensor placement, the design decides which sensors to activate
  - The optimal design minimizes the **uncertainty in the posterior state**
  - OED famous criteria:
    1. A-optimality: Trace of posterior covariance
    2. D-optimality: Determinant of the posterior covariance
    3. etc.
- $\Phi(\boldsymbol{\lambda}) : \mathbb{R}_{+}^{\mathrm{N_s}} \mapsto [0, \infty)$ is a regularization function (e.g.,ℓ_1, ℓ_0, etc.)
- $\alpha > 0$ is a user-defined penalty parameter that controls the sparsity of the design

Outline

OED Approach for Adaptive Inflation

The A-optimal design (inflation parameter, $\boldsymbol{\lambda}^{\rm A-opt}$) minimizes:

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}^{\rm N_{state}}} \operatorname{Tr}\left(\widetilde{\mathbf{A}}(\boldsymbol{\lambda})\right) - \alpha \, \|\boldsymbol{\lambda} - \mathbf{1}\|_1$$

$$\text{subject to} \quad 1 = \lambda_i^l \leq \lambda_i \leq \lambda_i^u, \quad i = 1, \ldots, \rm N_{state}$$

Remark: we choose the sign of the regularization term to be negative, unlike the traditional formulation

- Let $\mathcal{H} = \mathbf{H} = \mathbf{I}$ with uncorrelated observation noise, the design criterion becomes:

$$\Psi^{\rm Infl}(\boldsymbol{\lambda}) := \operatorname{Tr}\left(\widetilde{\mathbf{A}}\right) = \sum_{i=1}^{\rm N_{state}} \left(\lambda_i^{-1}\sigma_i^{-2} + r_i^{-2}\right)^{-1}$$

- Decreasing λ_i reduces $\Psi^{\rm Infl}$, i.e. the optimizer will always move toward $\boldsymbol{\lambda}^l$

OED Approach for Adaptive Inflation

Solving the A-OED problem, requires evaluating the objective, and the gradient:

- **The design criterion:**

$$\Psi^{\mathrm{Infl}}(\boldsymbol{\lambda}) := \mathrm{Tr}\left(\widetilde{\mathbf{A}}\right) = \mathrm{Tr}\left(\widetilde{\mathbf{B}}\right) - \mathrm{Tr}\left(\left(\mathbf{R} + \mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\right)^{-1}\mathbf{H}\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\right)$$

- **The gradient:**

$$\nabla_{\boldsymbol{\lambda}}\Psi^{\mathrm{Infl}}(\boldsymbol{\lambda}) = \sum_{i=1}^{\mathrm{N_{state}}} \lambda_i^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\left(z_1 - z2 - z3 + z4\right)$$

$$z_1 = \widetilde{\mathbf{B}}\mathbf{e}_i$$

$$z_2 = \mathbf{H}^{\mathsf{T}}\left(\mathbf{R} + \mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\right)^{-1}\mathbf{H}\widetilde{\mathbf{B}}z1$$

$$z_3 = \widetilde{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\left(\mathbf{R} + \mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\right)^{-1}\mathbf{H}z_1$$

$$z_4 = \mathbf{H}^{\mathsf{T}}\left(\mathbf{R} + \mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\right)^{-1}\mathbf{H}\widetilde{\mathbf{B}}z_3$$

$\mathbf{e}_i \in \mathbb{R}^{\mathrm{N_{state}}}$ is the ith cardinality vector

OED Adaptive $\mathbf{B}-$Localization (State-Space)

$$\min_{\mathbf{L}\in\mathbb{R}^{\mathrm{N_{state}}}} \Psi^{\mathrm{B-Loc}}(\mathbf{L}) + \gamma\,\Phi(\mathbf{L}) := \mathrm{Tr}\left(\widehat{\mathbf{A}}(\mathbf{L})\right) + \gamma\,\|\mathbf{L}\|_2$$

subject to $l_i^l \leq l_i \leq l_i^u, \quad i = 1,\ldots,\mathrm{N_{state}}$

- **The design criterion:**

$$\Psi^{B-Loc}(\mathbf{L}) = \mathrm{Tr}\left(\widehat{\mathbf{B}}\right) - \mathrm{Tr}\left(\left(\mathbf{R} + \mathbf{H}\widehat{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\right)^{-1} \mathbf{H}\widehat{\mathbf{B}}\widehat{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\right)$$

- **The gradient**:

$$\nabla_{\mathbf{L}}\Psi^{B-Loc} = \sum_{i=1}^{\mathrm{N_{state}}} \mathbf{e_i}\,\mathbf{l}_{B,i}\left(\mathbf{I} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H}\widehat{\mathbf{B}}\right)^{-1}\left(\mathbf{I} + \widehat{\mathbf{B}}\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{e}_i$$

$$\mathbf{l}_{B,i} = \mathbf{l}_i^{\mathsf{T}} \odot \left(\mathbf{e}_i^{\mathsf{T}}\mathbf{B}\right)$$

$$\mathbf{l}_i = \left(\frac{\partial\rho_{i,1}(l_i)}{\partial l_i}, \frac{\partial\rho_{i,2}(l_i)}{\partial l_i}, \ldots, \frac{\partial\rho_{i,\mathrm{N_{state}}}(l_i)}{\partial l_i}\right)^{\mathsf{T}}$$

$\mathbf{e}_i \in \mathbb{R}^{\mathrm{N_{state}}}$ is the ith cardinality vector

OED Adaptive: Observation-Space Localization

- So far, we assumed full state-space formulation, i.e. $\mathbf{L} \in \mathbb{R}^{\mathrm{N_{state}}}$
  1. the OED problem is solved to find $\mathbf{L}^{\mathrm{A-opt}}$ in the model state space
  2. $\mathbf{L}^{\mathrm{A-opt}}$ is projected, in the analysis step, into observation space to localize $\mathbf{HB}$, and $\mathbf{HBH}^{\mathsf{T}}$
  - **Pros:**
    - reduces the cost of calculating the analysis
  - **Cons:**
    - same cost for the optimization problem
    - projecting of $\mathbf{L}^{\mathrm{A-opt}}$ might be challenging or unknown

---

- **Alternative:** observation-space formulation:

  $\rightarrow$ formulate OED optimization problem in the observation space; i.e., $\mathbf{L} \in \mathbb{R}^{\mathrm{N_{obs}}}$

OED Adaptive: Observation-Space Localization

- Assume $\mathbf{L} \in \mathbb{R}^{\mathrm{N_{obs}}}$ is attached to observation grid points
- $\mathbf{HB}$ is replaced with $\widehat{\mathbf{HB}} = \mathbf{C}^{\mathrm{loc},1} \odot \mathbf{HB}$, with

$$\mathbf{C}^{\mathrm{loc},1} = \left[\rho_{i,j}^{o|m}(l_i) \right] \; ; \; i = 1, 2, \ldots \mathrm{N_{obs}} \; ; \; j = 1, 2, \ldots \mathrm{N_{state}}$$

- $\mathbf{HBH}^\mathsf{T}$ can be replaced with $\widehat{\mathbf{HBH}^\mathsf{T}} = \mathbf{C}^{\mathrm{loc},2} \odot \mathbf{HBH}^\mathsf{T}$, with

$$\mathbf{C}^{o|o} := \frac{1}{2}\left(\mathbf{C}_r^o + \mathbf{C}_c^o \right) = \frac{1}{2}\left[\rho_{i,j}^{o|o}(l_i) + \rho_{i,j}^{o|o}(l_j) \right]_{i,j=1,2,\ldots,\mathrm{N_{state}}}$$

- **Localized posterior covariances:**
  - Localize $\mathbf{HB}$:

$$\widehat{\mathbf{A}} = \mathbf{B} - \widehat{\mathbf{HB}}^\mathsf{T} \left(\mathbf{R} + \mathbf{HBH}^\mathsf{T} \right)^{-1} \widehat{\mathbf{HB}}$$

  - Localize both $\mathbf{HB}$ and $\mathbf{HBH}^\mathsf{T}$:

$$\widehat{\mathbf{A}} = \mathbf{B} - \widehat{\mathbf{HB}}^\mathsf{T} \left(\mathbf{R} + \widehat{\mathbf{HBH}^\mathsf{T}} \right)^{-1} \widehat{\mathbf{HB}}$$

OED Adaptive $\mathbf{R}-$Localization

Decorrelate $\mathbf{HB}$

- **The design criterion:**

$$\Psi^{R-Loc}(\mathbf{L}) = \mathrm{Tr}\left(\mathbf{B}\right) - \mathrm{Tr}\left(\widehat{\mathbf{HB}}\, \widehat{\mathbf{HB}}^{\mathsf{T}} \left(\mathbf{R} + \mathbf{HBH}^{\mathsf{T}} \right)^{-1} \right)$$

- **The gradient**:

$$\begin{aligned}
\nabla_{\mathbf{L}} \Psi^{R-Loc} &= -2 \sum_{i=1}^{\mathrm{N_{obs}}} \mathbf{e_i}\, \mathbf{l}_{\mathrm{HB,i}}^{\mathsf{T}}\, \psi_i \\
\psi_i &= \widehat{\mathbf{HB}}^{\mathsf{T}} \left(\mathbf{R} + \mathbf{HBH}^{\mathsf{T}} \right)^{-1} \mathbf{e}_i \\
\mathbf{l}_{\mathrm{HB,i}} &= \left(\mathbf{l}_i^s \right)^{\mathsf{T}} \odot \left(\mathbf{e}_i^{\mathsf{T}} \mathbf{HB} \right) \\
\mathbf{l}_i^s &= \left(\frac{\partial \rho_{i,1}(l_i)}{\partial l_i}, \frac{\partial \rho_{i,2}(l_i)}{\partial l_i}, \ldots, \frac{\partial \rho_{i,\mathrm{N_{state}}}(l_i)}{\partial l_i} \right)^{\mathsf{T}}
\end{aligned}$$

$\mathbf{e}_i \in \mathbb{R}^{\mathrm{N_{obs}}}$ is the ith cardinality vector

OED Adaptive $\mathbf{R}$−Localization

Decorrelate $\mathbf{HB}$ and $\mathbf{HBH}^\mathsf{T}$

- **The design criterion**:

$$\Psi^{R-Loc}(\mathbf{L}) = \mathrm{Tr}\left(\mathbf{B}\right) - \mathrm{Tr}\left(\widehat{\mathbf{HB}}\, \widehat{\mathbf{HB}}^\mathsf{T} \left(\mathbf{R} + \widehat{\mathbf{HBH}^\mathsf{T}} \right)^{-1} \right)$$

- **The gradient**:

$$\nabla_{\mathbf{L}} \Psi^{R-Loc} = \sum_{i=1}^{\mathrm{N_{obs}}} \mathbf{e_i} \left(\eta_i^o - 2\, \mathbf{l}_{\mathrm{HB,i}}^\mathsf{T} \right) \psi_i^o$$

$$\psi_i^o = \widehat{\mathbf{HB}}^\mathsf{T} \left(\mathbf{R} + \widehat{\mathbf{HBH}^\mathsf{T}} \right)^{-1} \mathbf{e}_i$$

$$\eta_i^o = \mathbf{l}_{B,i}^o \left(\mathbf{R} + \widehat{\mathbf{HBH}^\mathsf{T}} \right)^{-1} \widehat{\mathbf{HB}}$$

$$\mathbf{l}_{B,i}^o = \left(\mathbf{l}_i^o \right)^\mathsf{T} \odot \left(\mathbf{e}_i^\mathsf{T} \mathbf{HBH}^\mathsf{T} \right)$$

$$\mathbf{l}_i^o = \left(\frac{\partial \rho_{i,1}(l_i)}{\partial l_i}, \frac{\partial \rho_{i,2}(l_i)}{\partial l_i}, \ldots, \frac{\partial \rho_{i,\mathrm{N_{obs}}}(l_i)}{\partial l_i} \right)^\mathsf{T}$$

Outline

Experimental Setup

- The model (Lorenz-96):

$$\frac{dx_i}{dt} = x_{i-1}\left(x_{i+1} - x_{i-2}\right) - x_i + F\,; \quad i = 1, 2, \ldots, 40\,,$$

  - $\mathbf{x} \in \mathbb{R}^{40}$ is the state vector, with $x_0 \equiv x_{40}$
  - $F = 8$
- Initial background ensemble & uncertainty:
  - reference IC: $\mathbf{x}_0^{\rm True} = \mathcal{M}_{t=0\rightarrow t=5}(-2, \ldots, 2)^{\mathsf{T}}$
  - $\mathbf{B}_0 = \sigma_0 \mathbf{I} \in \mathbb{R}^{\rm N_{state} \times N_{state}}$, with $\sigma_0 = 0.08 \left\| \mathbf{x}_0^{\rm True} \right\|_2$
- Observations:
  - $\sigma_{\rm obs} = 5\%$ of the average magnitude of the observed reference trajectory
  - $\mathbf{R} = \sigma_{\rm obs} \mathbf{I} \in \mathbb{R}^{\rm N_{obs} \times N_{obs}}$
  - Synthetic observations are generated every 20 time steps, with

$$\mathcal{H}(\mathbf{x}) = \mathbf{H}\mathbf{x} = \left(x_1,\ x_3,\ x_5,\ \ldots,\ x_{37},\ x_{39}\right)^T \in \mathbb{R}^{20}\,.$$

- EnKF flavor used here: DEnKF with Gaspari-Cohn (GC) localization

All experiments are carried out using **DATeS**

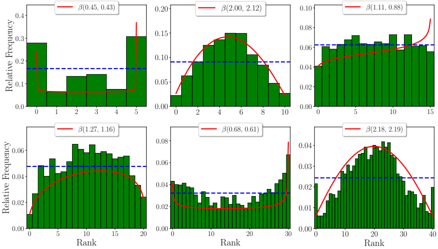
- `http://people.cs.vt.edu/~attia/DATeS/`
- `https://doi.org/10.5281/zenodo.1247464`
- Ahmed Attia and Adrian Sandu, DATeS: A Highly-Extensible Data Assimilation Testing Suite, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-30, in review, 2018.

Numerical Results: Performance Metrics

- RMSE:

$$\mathbf{RMSE} = \sqrt{\frac{1}{\mathrm{N_{state}}} \sum_{i=1}^{\mathrm{N_{state}}} (x_i - x_i^{\mathrm{True}})^2}\,,$$

- KL-distance to uniform Rank histogram

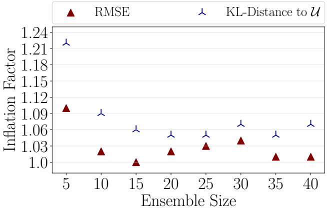


→ The KL divergence between two Beta distributions $\mathrm{Beta}(\alpha, \beta)$, and $\mathrm{Beta}(\alpha', \beta')$:

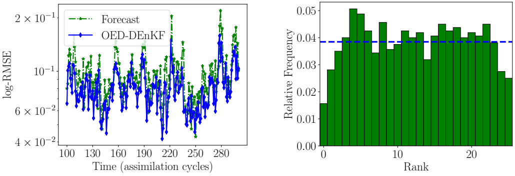
$$D_{\mathrm{KL}}(\mathrm{Beta}(\alpha,\beta) \mid \mathrm{Beta}(\alpha' \beta')) = \ln\Gamma(\alpha+\beta) - \ln(\alpha\,\beta) - \ln\Gamma(\alpha'+\beta') + \ln(\alpha'\,\beta') + (\alpha-\alpha')\Big(\psi(\alpha)-\psi(\alpha')\Big) + (\beta-\beta')\Big(\psi(\beta)-\psi(\beta')\Big)$$

- $\psi(\cdot) = \frac{\Gamma'(\cdot)}{\Gamma(\cdot)}$ is the digamma function, i.e. the logarithmic derivative of the gamma function
- $\mathcal{U}(0,1) \equiv Beta(\alpha' = 1,\ \beta' = 1)$

Numerical Results: Benchmark



The minimum average RMSE over the interval $[10, 30]$, for every choice of $\mathrm{N_{ens}}$, is indicated by red a triangle. Blue tripods indicate the minimum KL distance between the analysis rank histogram and a uniformly distributed rank histogram. Space-independent radius of influence $\mathbf{L} = 4$ is used.

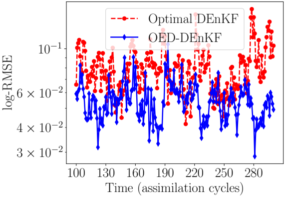


(a) RMSE (b) Rank histogram

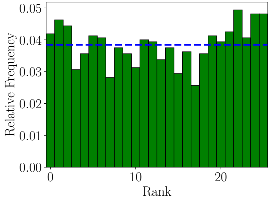
Analysis RMSE and rank histogram of DEnKF with $\mathbf{L} = 4$, and $\lambda = 1.05$.

Benchmark EnKF Results

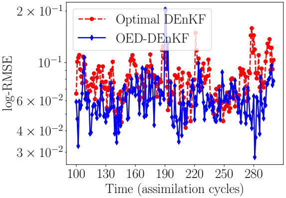
Numerical Results: OED Adaptive Space-Time Inflation I



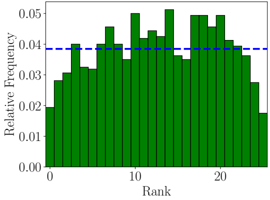
(a) RMSE; $\alpha = 0.14$



(b) Rank histogram; $\alpha = 0.14$



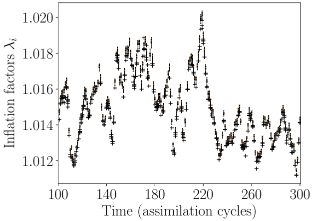
(c) RMSE; $\alpha = 0.04$



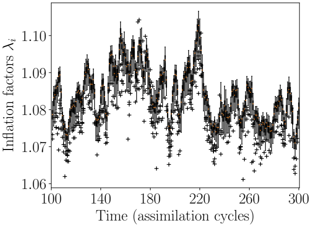
(d) Rank histogram; $\alpha = 0.04$

The localization radius is fixed to $\mathbf{L} = 4$. The optimization penalty parameter α is indicated under each panel.

Numerical Results: OED Adaptive Space-Time Inflation II



(a) $\alpha = 0.14$

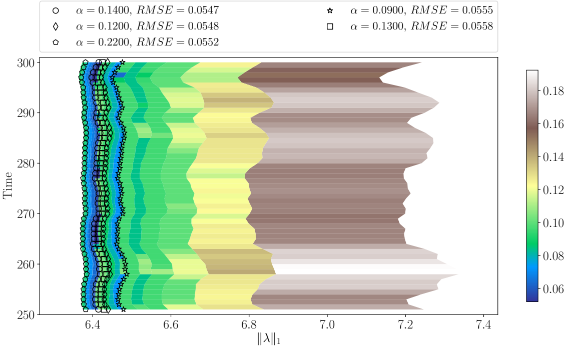


(b) $\alpha = 0.04$

Box plots expressing the range of values of the inflation coefficients at each time instant, over the testing timespan $[10, 30]$.

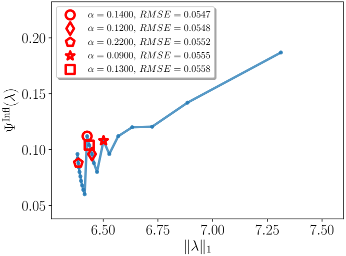
Numerical Results; A-OED Inflation Regularization I

Choosing α

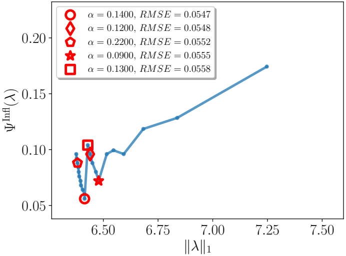


L-curve plots are are plotted for 25 equidistant values of the penalty parameter, at every assimilation time instant over the testing timespan $[0.03, 0.24]$. The values of the penalty parameter α that resulted in the 5 smallest average RMSEs, over all experiments carried out with different penalties, are highlighted on the plot and indicated in the legend along with the corresponding average RMSE.

Choosing α



(a) Cycle 100

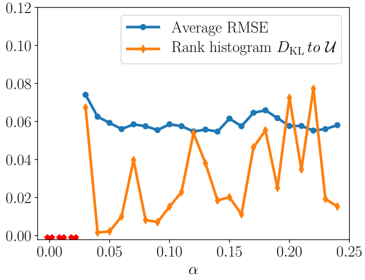


(b) Cycle 150

L-curve plots are are plotted for 25 equidistant values of the penalty parameter at assimilation cycles 100 and 150, respectively.

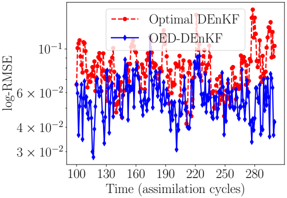
Numerical Results; A-OED Inflation Regularization III

Choosing α

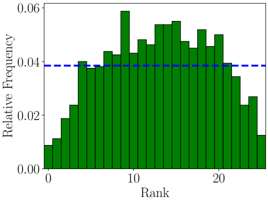


Average RMSE and KL-divergence from a uniform rank histogram resulted for 22 equidistant values of the penalty parameter in the interval $[0.03, 0.24]$. Values of the penalty parameter α that led to filter or optimizer divergence are indicated by red **x** marks.

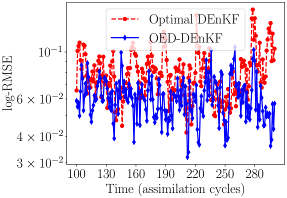
Numerical Results: OED Adaptive Space-Time Localization I



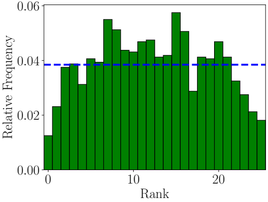
(a) RMSE; $\gamma = 0$



(b) Rank histogram; $\gamma = 0$



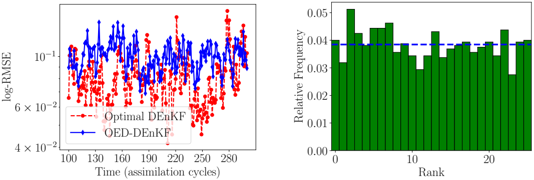
(c) RMSE; $\gamma = 0.001$



(d) Rank histogram; $\gamma = 0.001$

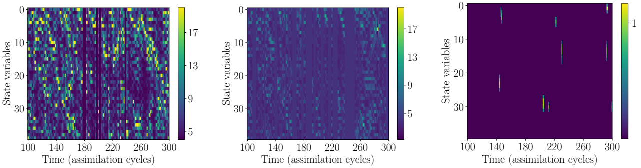
The inflation factor is fixed to $\lambda = 1.05$. The optimization penalty parameter γ is shown under each panel.

Numerical Results: OED Adaptive Space-Time Localization II



(a) RMSE (b) Rank histogram

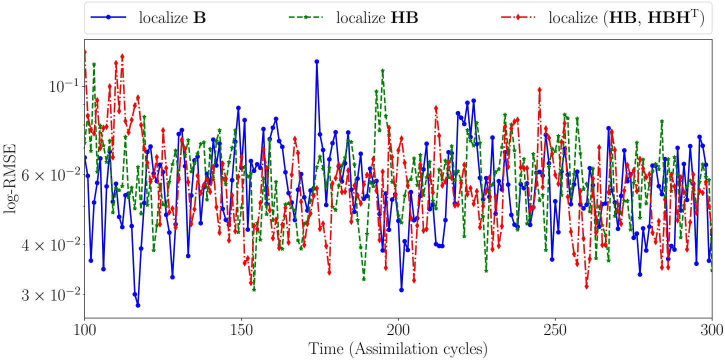
Results for $\lambda = 1.05$, and $\gamma = 0.04$.



(a) $\gamma = 0.0$ (b) $\gamma = 0.001$ (c) $\gamma = 0.04$

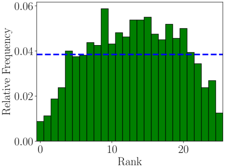
Localization radii at each time points, over the testing timespan $[10, 30]$. The optimization penalty parameter γ is shown under each panel.

Numerical Results: OED Adaptive Space-Time Localization III

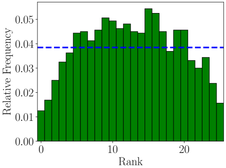


A-OED optimal localization radii $\mathbf{L}$ found by solving the OED localization problems in model state-space, and observation space respectively. No regularization is applied, i.e., $\gamma = 0$

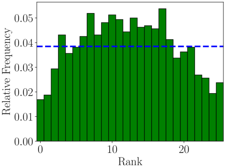
Numerical Results: OED Adaptive Space-Time Localization IV



(a) Localize $\mathbf{B}$

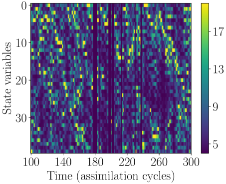


(b) Localize $\mathbf{HB}$

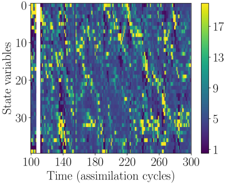


(c) Localize $\left(\mathbf{HB}, \mathbf{HBH}^{\mathsf{T}}\right)$

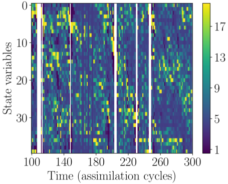
Rank histogram for A-OED localization solved in model state-space, and observation space respectively.



(d) Localize $\mathbf{B}$



(e) Localize $\mathbf{HB}$

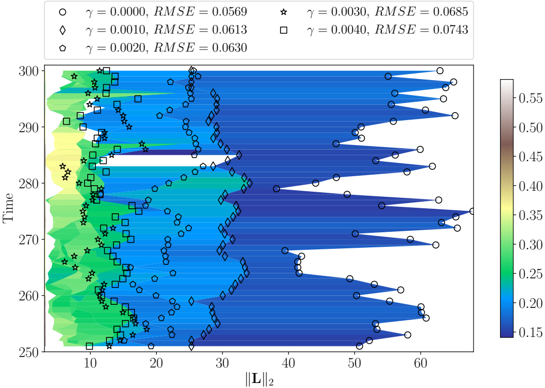


(f) Localize $\left(\mathbf{HB}, \mathbf{HBH}^{\mathsf{T}}\right)$

Space-time optimal localization radii over the testing timespan.

Numerical Results; A-OED Localization Regularization I

Choosing γ



L-curve plots are shown for values of the penalty parameter $\gamma = 0,\ 0.001,\ \ldots,\ 0.34.$

Concluding Remarks

- Introduced an OED approach for adaptive inflation and localization
- Either A-OED inflation or localization is carried out each cycle
- Can create a weighted objective to account for both inflation and localization
- Regularization is a must for adaptive inflation
- Regularization may not be needed, in general, for adaptive localization

---

- Definiteness of the localization Kernel $\mathbf{D}$
- Regularization norm
- Other OED criteria; e.g., D-optimality
- Adaptive Bayesian A-OED!

Thank You