Large-Scale Data Fusion for Improved Model Simulation and Predictability

Ahmed Attia

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

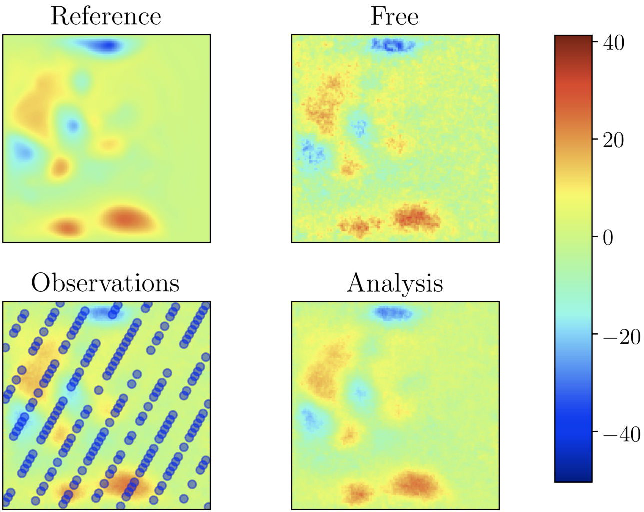
Thanks to
Adrian Sandu: VTech
Emil M. Constantinescu: ANL/UChicago
Alen Alexanderian: SAMSI/NCSU
Arvind K. Saibaba: SAMSI/NCSU
Vishwas Rao: ANL
Azam Moosavi: VTech

Argonne National Laboratory
October 31, 2018

Outline

Motivation: Ocean Simulation

- Consider the sea-surface-height (SSH) in $\Omega \in \mathbb{R}^2$:



Motivation: Advection-Diffusion

- Consider the concentration of a contaminant in $\Omega \in \mathbb{R}^2$:

- **Simulation:** given model parameter $\theta := \mathbf{x}_0$, forward integrator, and model discretization, solve the DEs $\mathbf{x}_0 \to \mathbf{x}_k$

Motivation: Advection-Diffusion

- Consider the concentration of a contaminant in $\Omega \in \mathbb{R}^2$:

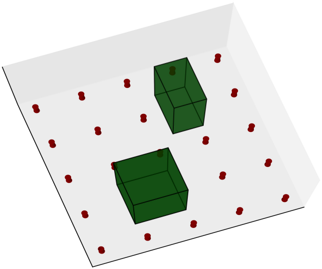
- **Simulation:** given model parameter $\theta := \mathbf{x}_0$, forward integrator, and model discretization, solve the DEs $\mathbf{x}_0 \to \mathbf{x}_k$

- **Forward problem:** given model state/parameter predict model observations $\theta \to \mathbf{y}$

- **Inverse problem:** given noisy observations, and "possibly" uncertain model state/parameter, recover the unknown model state/parameter $\theta \leftarrow \mathbf{y}$

Motivation: Advection-Diffusion

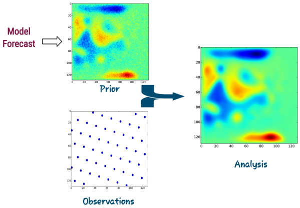
- Consider the concentration of a contaminant in $\Omega \in \mathbb{R}^2$:



- **Simulation:** given model parameter $\theta := \mathbf{x}_0$, forward integrator, and model discretization, solve the DEs $\mathbf{x}_0 \rightarrow \mathbf{x}_k$
- **Forward problem:** given model state/parameter predict model observations $\theta \rightarrow \mathbf{y}$
- **Inverse problem:** given noisy observations, and "possibly" uncertain model state/parameter, recover the unknown model state/parameter $\theta \leftarrow \mathbf{y}$
- **Design of experiments:** e.g., sensor placement for optimal reconstruction of model parameter

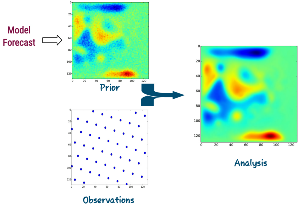
Bayesian Inversion & Data Assimilation

- **The prior** $\mathbb{P}(\theta)$: encapsulates knowledge about θ prior to obtaining new observations
- **The likelihood** $\mathbb{P}(\mathbf{y}|\theta)$: describes the probability distribution of observations conditioned by the model parameter



Bayesian Inversion & Data Assimilation

- **The prior** $\mathbb{P}(\theta)$: encapsulates knowledge about θ prior to obtaining new observations
- **The likelihood** $\mathbb{P}(\mathbf{y}|\theta)$: describes the probability distribution of observations conditioned by the model parameter



Data Assimilation (DA)

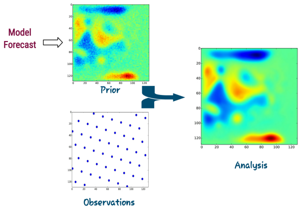
$$\underbrace{\textbf{Model} + \textbf{Prior} + \textbf{Observations}}_{\text{with associated \underline{uncertainties}}} \rightarrow \textbf{Best description of the parameter}$$

- **The posterior** $\mathbb{P}(\theta|\mathbf{y})$: distribution of the parameter θ conditioned on observations

Bayes' theorem: $\textbf{Posterior} \propto \textbf{Likelihood} \times \textbf{Prior}$

Bayesian Inversion & Data Assimilation

- **The prior** $\mathbb{P}(\theta)$: encapsulates knowledge about θ prior to obtaining new observations
- **The likelihood** $\mathbb{P}(\mathbf{y}|\theta)$: describes the probability distribution of observations conditioned by the model parameter



Data Assimilation (DA)

$$\underbrace{\textbf{Model} + \textbf{Prior} + \textbf{Observations}}_{\text{with associated uncertainties}} \rightarrow \textbf{Best description of the parameter}$$

- **The posterior** $\mathbb{P}(\theta|\mathbf{y})$: distribution of the parameter θ conditioned on observations

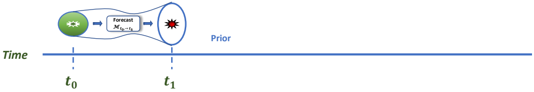
Bayes' theorem: **Posterior** $\propto$ **Likelihood** $\times$ **Prior**

Applications include:

Atmospheric forecasting, power flow, oil reservoir, ocean, ground water, etc.

Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

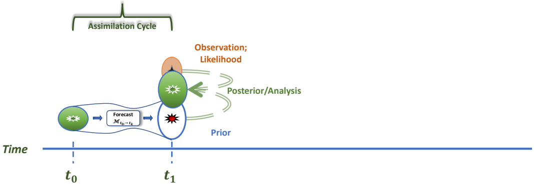


Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

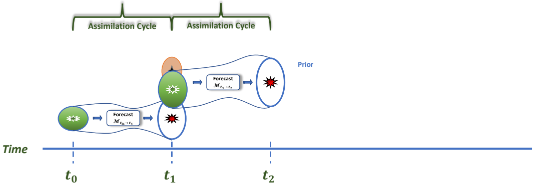
Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



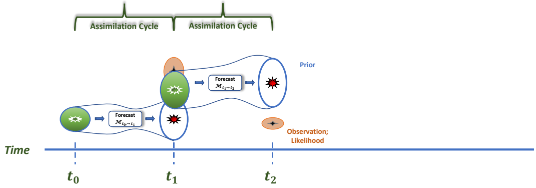
Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



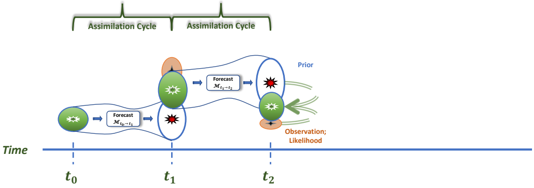
Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



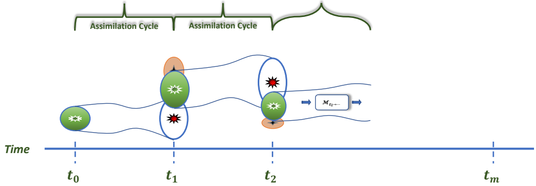
Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



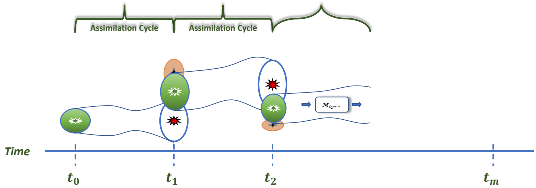
Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

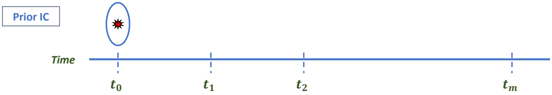


Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

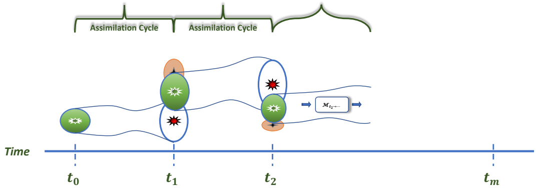


- Smoothing (4D-DA): assimilate multiple observations at once $(\mathbf{x}_0|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$

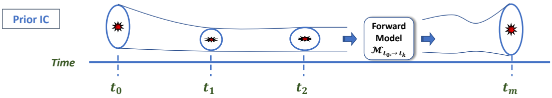


Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

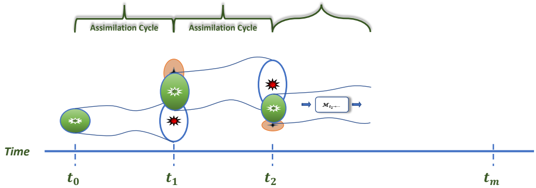


- Smoothing (4D-DA): assimilate multiple observations at once $(\mathbf{x}_0|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$

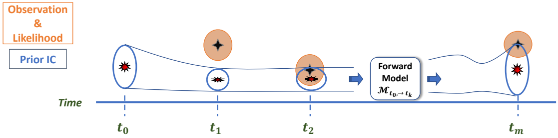


Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

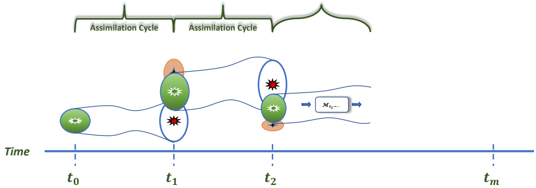


- Smoothing (4D-DA): assimilate multiple observations at once $(\mathbf{x}_0|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$

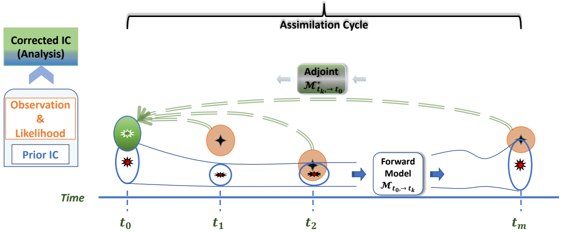


Data Assimilation: Problem Setup

- Filtering (3D-DA): assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



- Smoothing (4D-DA): assimilate multiple observations at once $(\mathbf{x}_0|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$



Data Assimilation: Probabilistic Assumptions & Solvers

Simplifying assumptions are imposed on the error distribution, "*typically*", errors are assumed to be Gaussian (*easy, tractable, ...*).

- **The Gaussian framework**:
  - Prior: $\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{true}} \sim \mathcal{N}(0, \mathbf{B})$
  - Likelihood: $\mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathrm{true}}) \sim \mathcal{N}(0, \mathbf{R})$
  - → **Posterior:** $\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{true}} \sim \mathcal{N}(0, \mathbf{A})$

Data Assimilation: Probabilistic Assumptions & Solvers

Simplifying assumptions are imposed on the error distribution, "*typically*", errors are assumed to be Gaussian (*easy, tractable, ...*).

- **The Gaussian framework**:
  - Prior: $\mathbf{x}^{\rm b} - \mathbf{x}^{\rm true} \sim \mathcal{N}(0, \mathbf{B})$
  - Likelihood: $\mathbf{y} - \mathcal{H}(\mathbf{x}^{\rm true}) \sim \mathcal{N}(0, \mathbf{R})$
  - → **Posterior:** $\mathbf{x}^{\rm a} - \mathbf{x}^{\rm true} \sim \mathcal{N}(0, \mathbf{A})$
- **Approaches & solvers:**

		Approach	
		Variational: *solve an optimization problem, e.g., minimize the negative-log of the posterior, to get an analysis state*	**Ensemble:** *use Bayes' theorem, with Monte-Carlo representation of errors and states/parameters*
Problem Setup	**Filtering (3D)**	• **3DVar:** three-dimensional Variational DA	• **EnKF:** Ensemble Kalman filter • **MLEF:** Maximum-likelihood ensemble filter • **IEnKF:** Iterative Ensemble Kalman filter • **PF:** Particle filters • **MCMC:** Markov Chain Monte-Carlo sampling • ...
	Smoothing (4D)	• **4DVar:** four-dimensional Variational DA	• **EnKS:** Ensemble Kalman Smoother • **MCMC:** Markov Chain Monte-Carlo sampling • ...

Data Assimilation: Challenges

- **Dimensionality:**
  - Model state space: $\mathrm{N_{state}} \sim 10^{8-12}$
  - Observation space: $\mathrm{N_{obs}} \ll \mathrm{N_{state}}$
  - Ensemble size: $\mathrm{N_{ens}} \sim 100$
- **Gaussian framework:**
  - Strong assumption that holds for linear dynamics and linear observation operator $\mathcal{H}$
  - **EnKF is the most popular filter for "linear-Gaussian" settings:**
    - $\rightarrow$ Sampling errors
    - $\rightarrow$ Spurious long-range correlations,
    - $\rightarrow$ Rank-deficiency
    - $\rightarrow$ Ensemble collapse, and filter divergence
  - $\mathcal{H}$ is becoming more *nonlinear*, leading to *non-Gaussian* posterior
- **Non-Gaussian DA**
  - PF: Degeneracy
  - MCMC: Gold standard, yet computationally unaffordable

Data Assimilation: Challenges

- **Dimensionality:**
  - Model state space: $\mathrm{N_{state}} \sim 10^{8-12}$
  - Observation space: $\mathrm{N_{obs}} \ll \mathrm{N_{state}}$
  - Ensemble size: $\mathrm{N_{ens}} \sim 100$

- **Gaussian framework:**
  - Strong assumption that holds for linear dynamics and linear observation operator $\mathcal{H}$
  - **EnKF is the most popular filter for "linear-Gaussian" settings:**
    - $\rightarrow$ Sampling errors
    - $\rightarrow$ Spurious long-range correlations,
    - $\rightarrow$ Rank-deficiency
    - $\rightarrow$ Ensemble collapse, and filter divergence
  - $\mathcal{H}$ is becoming more *nonlinear*, leading to *non-Gaussian* posterior

- **Non-Gaussian DA**
  - PF: Degeneracy
  - MCMC: Gold standard, yet computationally unaffordable

- **Resampling family:** Gradient-based MCMC (e.g. HMC) filtering and smoothing algorithms for non-Gaussian DA

Hybrid Monte-Carlo (HMC) sampling

To draw samples $\{\mathbf{x}(e)\}_{e=1,2,\ldots}$ from $\propto \pi(\mathbf{x}) = e^{-\mathcal{J}(\mathbf{x})}$:

- $\mathbf{x}$: viewed as a position variable,
- Add synthetic "*momentum*" $\mathbf{p} \sim \mathcal{N}(0, \mathbf{M})$ and sample the joint PDF, then discard $\mathbf{p}$.
- Generate a **MC** with invariant distribution $\propto \exp\left(-H(\mathbf{p}, \mathbf{x})\right)$;
- **HMC proposal**: symplectic integrator plays the role of a proposal density.

- The Hamiltonian:

$$H(\mathbf{p}, \mathbf{x}) = \underbrace{\frac{1}{2}\,\mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}}_{\text{kinetic energy}} + \underbrace{\mathcal{J}(\mathbf{x})}_{\text{potential energy}} = \frac{1}{2}\,\mathbf{p}^T \mathbf{M}^{-1} \mathbf{p} - \log(\pi(\mathbf{x}))$$

- The Hamiltonian dynamics (a symplectic integrator used):

$$\frac{d\mathbf{x}}{dt} = \nabla_{\mathbf{p}} H = \mathbf{M}^{-1}\mathbf{p}\,, \qquad \frac{d\mathbf{p}}{dt} = -\nabla_{\mathbf{x}} H = -\nabla_{\mathbf{x}} \mathcal{J}(\mathbf{x})$$

- The canonical PDF of $(\mathbf{p}, \mathbf{x})$:

$$\propto \exp\left(-H(\mathbf{p}, \mathbf{x})\right) = \exp\left(-\frac{1}{2}\mathbf{p}^T \mathbf{M}^{-1} \mathbf{p} - \mathcal{J}(\mathbf{x})\right) = \exp\left(-\frac{1}{2}\mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}\right) \pi(\mathbf{x})$$

Hybrid Monte-Carlo (HMC) sampling

Symplectic integrators

- To integrate the solution of the Hamiltonian equations from pseudo time t_k to time $t_{k+1} = t_k + h$:
  1. **Position Verlet integrator**

$$\begin{aligned} \mathbf{x}_{k+1/2} &= \mathbf{x}_k + \frac{h}{2}\,\mathbf{M}^{-1}\,\mathbf{p}_k\,, \\ \mathbf{p}_{k+1} &= \mathbf{p}_k - h\,\nabla_{\mathbf{x}}\mathcal{J}(\mathbf{x}_{k+1/2})\,, \\ \mathbf{x}_{k+1} &= \mathbf{x}_{k+1/2} + \frac{h}{2}\,\mathbf{M}^{-1}\,\mathbf{p}_{k+1}\,. \end{aligned}$$

  2. **Two-stage integrator**

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x}_k + (a_1 h)\mathbf{M}^{-1}\mathbf{p}_k\,, \\ \mathbf{p}_1 &= \mathbf{p}_k - (b_1 h)\nabla_{\mathbf{x}}\mathcal{J}(\mathbf{x}_1)\,, \\ \mathbf{x}_2 &= \mathbf{x}_1 + (a_2 h)\mathbf{M}^{-1}\mathbf{p}_1\,, \\ \mathbf{p}_{k+1} &= \mathbf{p}_1 - (b_1 h)\nabla_{\mathbf{x}}\mathcal{J}(\mathbf{x}_2)\,, \\ \mathbf{x}_{k+1} &= \mathbf{x}_2 + (a_2 h)\mathbf{M}^{-1}\mathbf{p}_{k+1}\,, \end{aligned}$$

$$a_1 = 0.21132\,, \quad a_2 = 1 - 2a_1\,, \quad b_1 = 0.5\,.$$

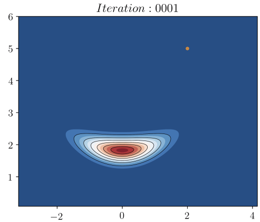
- MH: Acceptance Probability: $a^{(k)} = 1 \wedge e^{-\Delta H}, \quad \Delta H = H(\mathbf{p}^*, \mathbf{x}^*) - H(\mathbf{p}_k, \mathbf{x}_k)$

$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{x}^* & \text{with probability } a^{(k)} \\ \mathbf{x}_k & \text{with probability } 1 - a^{(k)} \end{cases}$$

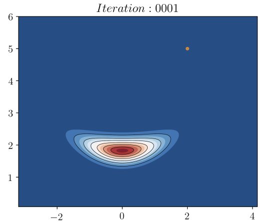
Hybrid Monte-Carlo (HMC) sampling

Examples; code is available from: `https://www.mcs.anl.gov/~attia/software.html`

- MH Sampling



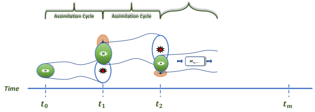
- HMC Sampling



1- HMC sampling filter † (for sequential DA)

The analysis step

Assimilate given information (e.g. background and observations) at a single time instance t_k.



- **Gaussian framework**:

$$\mathbb{P}^{\rm b}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|\mathbf{x}-\mathbf{x}^{\rm b}\|_{\mathbf{B}^{-1}}\right) ; \quad \mathbb{P}(\mathbf{y}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|\mathcal{H}(\mathbf{x})-\mathbf{y}\|_{\mathbf{R}^{-1}}\right) .$$

$$\mathbb{P}^{\rm a}(\mathbf{x}) \propto \overbrace{\exp\left(-\mathcal{J}(\mathbf{x})\right)}^{\pi(\mathbf{x})} ,$$

- **Potential energy and gradient**:

$$\begin{aligned} \mathcal{J}(\mathbf{x}) &= \frac{1}{2}\|\mathbf{x}-\mathbf{x}^{\rm b}\|_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathcal{H}(\mathbf{x})-\mathbf{y}\|_{\mathbf{R}^{-1}} , \\ \nabla_{\mathbf{x}}\mathcal{J}(\mathbf{x}) &= \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^{\rm b}) + \mathbf{H}^T\mathbf{R}^{-1}(\mathcal{H}(\mathbf{x})-\mathbf{y}) . \end{aligned}$$

- **The Hamiltonian**:

$$H(\mathbf{p}, \mathbf{x}) = \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + \mathcal{J}(\mathbf{x}).$$

† Attia, Ahmed, and Adrian Sandu. "A hybrid Monte Carlo sampling filter for non-Gaussian data assimilation." AIMS Geosciences 1, no. geosci-01-00041 (2015): 41-78.

1- HMC sampling filter (for sequential DA)

Numerical experiments: 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 $x_0 \equiv x_{40}$, and $F = 8$
- Initial background ensemble & uncertainty:
  - reference IC: $\mathbf{x}_0^{\rm True} = \mathcal{M}_{t=0 \rightarrow t=5}(-2, \ldots, 2)^{\sf T}$
  - $\mathbf{B}_0 = \sigma_0 \mathbf{I} \in \mathbb{R}^{\rm N_{state} \times N_{state}}$, 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

$$\mathcal{H}(\mathbf{x}) = \begin{cases} (x_1,\ x_4,\ x_7,\ \ldots,\ x_{37},\ x_{40})^T \in \mathbb{R}^{14} \\ (x'_1,\ x'_4,\ x'_7,\ \ldots,\ x'_{37},\ x'_{40})^T \in \mathbb{R}^{14} \end{cases} \quad \text{with} \quad x'_i = \begin{cases} x_i^2 & : x_i \geq 0.5 \\ -x_i^2 & : x_i < 0.5 \end{cases}$$

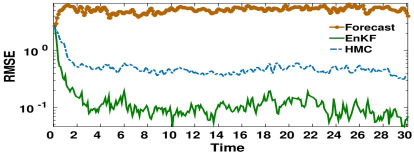
- Benchmark EnKF flavor: DEnKF with Gaspari-Cohn (GC) localization

Experiments are carried out using DATeS

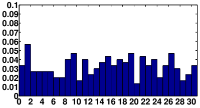
- Ahmed Attia and Adrian Sandu, DATeS: A Highly-Extensible Data Assimilation Testing Suite, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-30, in review, 2018.
- http://people.cs.vt.edu/~attia/DATeS/

1- HMC sampling filter (for sequential DA)

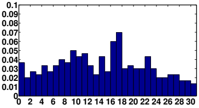
Numerical experiments: results with linear $\mathcal{H}$



(a) RMSE

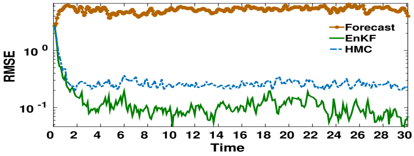


(b) R. Hist: $\mathbf{x}_1$

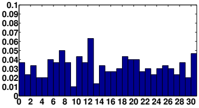


(c) R. Hist: $\mathbf{x}_2$

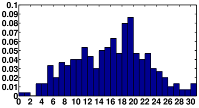
Position Verlet symplectic integrator is used with time step $T = 0.1$ with $h = 0.01$, $\ell = 10$, and 10 mixing steps. The $(\log)$ RMSE reported for the HMC filter is the average taken over the 100 realizations of the filter.



(a) RMSE



(b) R. Hist: $\mathbf{x}_1$

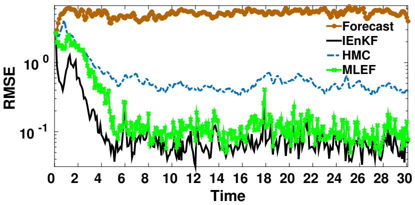


(c) R. Hist: $\mathbf{x}_2$

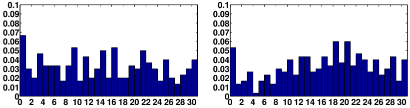
Two-stage symplectic integrator is used with time step $T = 0.1$ with $h = 0.01$, $\ell = 10$, and 10 mixing steps. The $(\log)$ RMSE reported for the HMC filter is the average taken over the 100 realizations of the filter.

1- HMC sampling filter (for sequential DA)

Numerical experiments: results with discontinuous quadratic $\mathcal{H}$

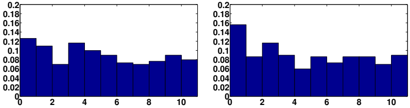


Three-stage symplectic integrator is used with time step $T = 0.1$ with $h = 0.01$, $\ell = 10$, and 10 mixing steps. The $(\log)$ RMSE reported for the HMC filter is the average taken over the 100 realizations of the filter.



(a) $\mathbf{x}_1$ (b) $\mathbf{x}_2$

Rank histograms for observed and unobserved components of the state vector with $\mathrm{N_{ens}} = 30$.

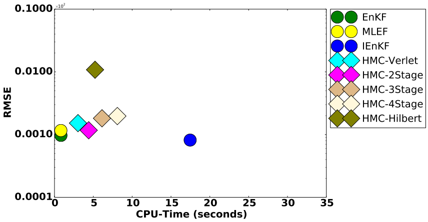


(a) $\mathbf{x}_1$ (b) $\mathbf{x}_2$

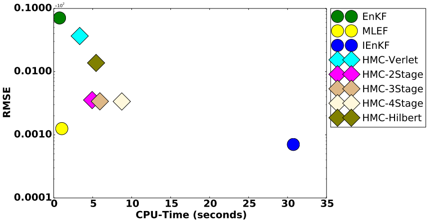
Rank histograms for observed and unobserved components of the state vector with $\mathrm{N_{ens}} = 10$.

1- HMC sampling filter (for sequential DA)

Numerical experiments: accuracy vs cost



(a) Linear observation operator



(b) Quadratic observation operator with a threshold

RMSE Vs CPU-time per assimilation cycle of DA with the Lorenz-96 model. The time reported is the average CPU-time taken over 100 identical runs of each experiment. The ensemble size is fixed to 30 members for all experiments here.

Relaxing the Gaussian-Prior Assumption

- To this point, we have assumed that "the prior can be well-approximated by a Gaussian".
- In practice, the prior is generally expected to be non-Gaussian.
- The prior PDF can hardly be formulated explicitly or even upto a scaling factor.

Relaxing the Gaussian-Prior Assumption

- To this point, we have assumed that "the prior can be well-approximated by a Gaussian".
- In practice, the prior is generally expected to be non-Gaussian.
- The prior PDF can hardly be formulated explicitly or even upto a scaling factor.
- **Can we efficiently approximate the prior distribution given the ensemble of forecasts?**

Relaxing the Gaussian-Prior Assumption

- To this point, we have assumed that "the prior can be well-approximated by a Gaussian".
- In practice, the prior is generally expected to be non-Gaussian.
- The prior PDF can hardly be formulated explicitly or even upto a scaling factor.
- **Can we efficiently approximate the prior distribution given the ensemble of forecasts?**
- **Idea: approximate the prior density by fitting a Gaussian mixture model (GMM) to the forecast ensemble† (e.g. using EM).**

† Attia, Ahmed, Azam Moosavi, and Adrian Sandu. "Cluster sampling filters for non-Gaussian data assimilation." Atmosphere 9, no. 6 (2018): 213.

2- Sampling filters with GMM prior; cluster sampling filters

- Prior (GMM):

$$\mathbb{P}^{\rm b}(\mathbf{x}_k) = \sum_{i=1}^{\rm N_c} \tau_{k,i}\, \mathcal{N}(\mu_{k,i},\, \Sigma_{k,i})$$
$$= \sum_{i=1}^{\rm N_c} \tau_{k,i} \frac{(2\pi)^{-\frac{\rm N_{state}}{2}}}{\sqrt{|\Sigma_{k,i}|}} \exp\left(-\frac{1}{2}\|\mathbf{x}_k - \mu_{k,i}\|^2_{\Sigma_{k,i}^{-1}}\right),$$

where $\tau_{k,i} = \mathbb{P}(\mathbf{x}_k(e) \in i^{th}$ component, and $(\mu_{k,i},\, \Sigma_{k,i})$ are the mean and the covariance matrix associated with the i^{th} component.

- Likelihood:

$$\mathbb{P}(\mathbf{y}_k|\mathbf{x}_k) = \frac{(2\pi)^{-\frac{\rm N_{obs}}{2}}}{\sqrt{|\mathbf{R}_k|}} \exp\left(-\frac{1}{2}\|\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k\|^2_{\mathbf{R}_k^{-1}}\right)$$

- Posterior:

$$\mathbb{P}^{\rm a}(\mathbf{x}_k) \propto \sum_{i=1}^{\rm N_c} \frac{\tau_{k,i}}{\sqrt{|\Sigma_{k,i}|}} \exp\left(-\frac{1}{2}\|\mathbf{x}_k - \mu_{k,i}\|^2_{\Sigma_{\mathbf{k},\mathbf{i}}^{-1}} - \frac{1}{2}\|\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k\|^2_{\mathbf{R}_k^{-1}}\right)$$

2- Sampling filters with GMM prior; cluster sampling filters

HMC sampling filter with GMM prior ($\mathcal{C}\ell$HMC)

- Potential energy and gradient:

$$
\begin{aligned}
\mathcal{J}(\mathbf{x}_k) &= \frac{1}{2}\|\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k\|^2_{\mathbf{R}_k^{-1}} - \log\left(\sum_{i=1}^{\mathrm{N_c}} \frac{\tau_{k,i}}{\sqrt{|\Sigma_{k,i}|}} \exp\left(-\frac{1}{2}\|\mathbf{x}_k - \mu_{k,i}\|^2_{\Sigma_{k,i}^{-1}}\right)\right) \\
&= \frac{1}{2}\|\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k\|^2_{\mathbf{R}_k^{-1}} + \mathcal{J}_{k,1}(\mathbf{x}_k) - \log\left(\frac{\tau_{k,1}}{\sqrt{|\Sigma_{k,1}|}}\right) \\
&\quad - \log\left(1 + \sum_{i=2}^{\mathrm{N_c}} \frac{\tau_{k,i}\sqrt{|\Sigma_{k,1}|}}{\tau_{k,1}\sqrt{|\Sigma_{k,i}|}} \exp\left(\mathcal{J}_{k,1}(\mathbf{x}_k) - \mathcal{J}_{k,i}(\mathbf{x}_k)\right)\right).
\end{aligned}
$$

$$
\begin{aligned}
\nabla_{\mathbf{x}_k}\mathcal{J}(\mathbf{x}_k) &= \mathbf{H}_k^T\mathbf{R}_k^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k) + \nabla_{\mathbf{x}_k}\mathcal{J}_{k,1}(\mathbf{x}_k) \\
&\quad - \frac{1}{\left(1 + \sum_{j=2}^{\mathrm{N_c}} \frac{\tau_{k,j}\sqrt{|\Sigma_{k,1}|}}{\tau_{k,1}\sqrt{|\Sigma_{k,j}|}} \exp\left(\mathcal{J}_{k,1}(\mathbf{x}_k) - \mathcal{J}_{k,j}(\mathbf{x}_k)\right)\right)} \\
&\quad \sum_{i=2}^{\mathrm{N_c}} \left(\frac{\tau_{k,i}\sqrt{|\Sigma_{k,1}|}}{\tau_{k,1}\sqrt{|\Sigma_{k,i}|}} \exp\left(\mathcal{J}_{k,1}(\mathbf{x}_k) - \mathcal{J}_{k,i}(\mathbf{x}_k)\right)\right)\left[\nabla_{\mathbf{x}_k}\mathcal{J}_{k,1} - \nabla_{\mathbf{x}_k}\mathcal{J}_{k,i}\right], \\
\nabla_{\mathbf{x}_k}\mathcal{J}_{k,i} &= \nabla_{\mathbf{x}_k}\mathcal{J}_{k,i}(\mathbf{x}_k) = \Sigma_{k,i}^{-1}(\mathbf{x}_k - \mu_{k,i}); \quad \forall\, i = 1, 2, \ldots, \mathrm{N_c}.
\end{aligned}
$$

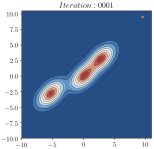
- The Hamiltonian:

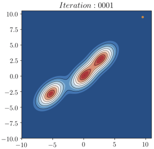
$$
H(\mathbf{p}_k,\, \mathbf{x}_k) = \frac{1}{2}\,\mathbf{p}_k{}^T\,\mathbf{M}^{-1}\,\mathbf{p}_k + \mathcal{J}(\mathbf{x}_k).
$$

2- Sampling filters with GMM prior; cluster sampling filters

limitations & multi-chain samplers (MC-$\mathcal{C}\ell$HMC , MC-$\mathcal{C}\ell$MCMC)

- **If GMM has too many components, $\mathcal{C}\ell$HMC may collapse (i.e. filter degeneracy)**

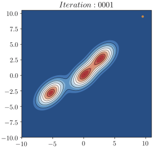


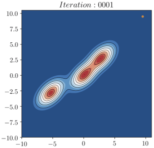


2- Sampling filters with GMM prior; cluster sampling filters

limitations & multi-chain samplers (MC-$\mathcal{C}\ell$HMC , MC-$\mathcal{C}\ell$MCMC)

- **If GMM has too many components, $\mathcal{C}\ell$HMC may collapse (i.e. filter degeneracy)**





- This could be avoided if we force the sampler to collect ensemble members from all the probability modes

---

Idea: construct a Markov chain to sample each of the components in the posterior
$\rightarrow$ Multi-chain cluster sampling filter (MC-$\mathcal{C}\ell$MCMC , MC-$\mathcal{C}\ell$HMC)

- The parameters of each chain can be **tuned locally**
  - Chains are initialized to the components' means in the prior mixture
  - The **local ensemble size** (sample size per chain) can be specified for example based on the *prior weight of the corresponding component, multiplied by the likelihood of the mean of that component*

2- HMC sampling filter with GMM prior; cluster sampling filters

Numerical experiments: setup

- **The model (QG-1.5):**

$$
\begin{aligned}
q_t &= \psi_x - \varepsilon J(\psi, q) - A\Delta^3\psi + 2\pi \sin(2\pi y)\,, \\
q &= \Delta\psi - F\psi\,, \\
J(\psi, q) &\equiv \psi_x q_x - \psi_y q_y\,,
\end{aligned}
$$

where $\Delta \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

  - State vector: $\psi \in \mathbb{R}^{16641}$.
  - Model subspace dimension of the order of $10^2 - 10^3$.
  - ψ is interpreted as either a *stream function* or *surface elevation*.
  - Here $F = 1600$, $\varepsilon = 10^{-5}$, and $A = 2 \times 10^{-12}$.
  - Boundary conditions: $\psi = \Delta\psi = \Delta^2\psi = 0$.

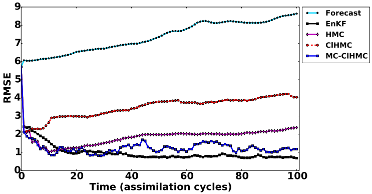
- **The observations:**

  1. A linear operator with random offset,
  2. A flow-velocity magnitude operator:

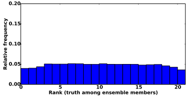
$$
\mathcal{H} : \mathbb{R}^{16641} \rightarrow \mathbb{R}^{300}\,; \quad \mathcal{H} : \psi \rightarrow \sqrt{u^2 + v^2}\,; \quad u = +\frac{\partial \psi}{\partial y}\,, \quad v = -\frac{\partial \psi}{\partial x}\,.
$$

2- HMC sampling filter with GMM prior; cluster sampling filters

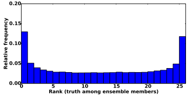
Numerical experiments: QG-1.5 results with linear $\mathcal{H}$



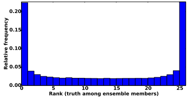
Data assimilation results with the linear observation operator. RMSE of the analyses obtained by EnKF, HMC, $\mathcal{C}\ell$HMC , and MC-$\mathcal{C}\ell$HMC filters. The ensemble size is 25. The symplectic integrator used is 3-stage, with $h = 0.0075$, $\ell = 25$, for HMC and $\mathcal{C}\ell$HMC , and $h = 0.05/\mathrm{N_c}, \ell = 15$ for MC-$\mathcal{C}\ell$HMC .



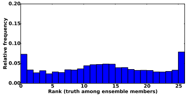
(a) DEnKF



(b) HMC



(c) $\mathcal{C}\ell$HMC +AIC

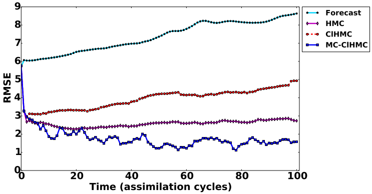


(d) MC-$\mathcal{C}\ell$HMC +AIC

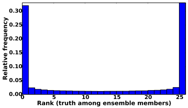
Data assimilation results with the linear observation operator. The rank histograms of where the truth ranks among posterior ensemble members. The ranks are evaluated for every 16^{th} variable in the state vector (past the correlation bound) at 100 assimilation times.

2- HMC sampling filter with GMM prior; cluster sampling filters

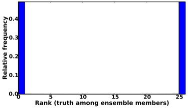
Numerical experiments: QG-1.5 results with flow magnitude $\mathcal{H}$



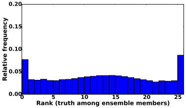
RMSE of the analyses obtained by HMC, $\mathcal{C}\ell$HMC , and MC-$\mathcal{C}\ell$HMC filtering schemes. In this experiment, EnKF analysis diverged after the third cycle, and it's RMSE results have been omitted for clarity.



(a) HMC



(b) $\mathcal{C}\ell$HMC +AIC

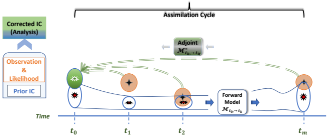


(c) MC-$\mathcal{C}\ell$HMC +AIC

The rank histograms of where the truth ranks among posterior ensemble members. The ranks are evaluated for every 16^{th} variable in the state vector (past the correlation bound) at 100 assimilation times. The filtering scheme used is indicated under each panel.

[3, 4] HMC sampling smoothers

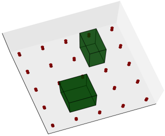
Assimilate a set of observations $\mathbf{y}_0, \mathbf{y}_1, \ldots \mathbf{y}_m$ at once, to a background $\mathbf{x}_0$.



1. Attia, Ahmed, Vishwas Rao, and Adrian Sandu. "A sampling approach for four dimensional data assimilation." In Dynamic Data-Driven Environmental Systems Science, pp. 215-226. Springer, Cham, 2015.
2. Attia, Ahmed, Vishwas Rao, and Adrian Sandu. "A hybrid Monte-Carlo sampling smoother for four-dimensional data assimilation." International Journal for Numerical Methods in Fluids 83, no. 1 (2017): 90-112.
3. Attia, Ahmed, Razvan Stefanescu, and Adrian Sandu. "The reduced-order hybrid Monte-Carlo sampling smoother." International Journal for Numerical Methods in Fluids 83, no. 1 (2017): 28-51.

Optimal Experimental Design

- Sensor placement for *optimal parameter recovery*



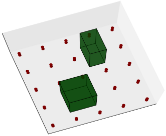
Experimental design: $\xi := \left\{ \begin{array}{c} \mathbf{y}_1, \ldots, \mathbf{y}_{\mathrm{N_s}} \\ w_1, \ldots, w_{\mathrm{N_s}} \end{array} \right\}$

$\mathbf{y}_1, \ldots, \mathbf{y}_{\mathrm{N_s}}$: candidate sensor locations; we can vary weights $w_i = \left\{ \begin{array}{ll} 0 & \text{sensor inactive} \\ 1: & \text{sensor active} \end{array} \right\}$

- Find the best r sensor location such as to maximize some utility function (e.g. identification accuracy, information gain, etc.)

Optimal Experimental Design

- Sensor placement for *optimal parameter recovery*



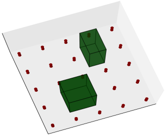
Experimental design: $\xi := \left\{ \begin{array}{c} \mathbf{y}_1, \ldots, \mathbf{y}_{\mathrm{N_s}} \\ w_1, \ldots, w_{\mathrm{N_s}} \end{array} \right\}$

$\mathbf{y}_1, \ldots, \mathbf{y}_{\mathrm{N_s}}$: candidate sensor locations; we can vary weights $w_i = \left\{ \begin{array}{cc} 0 & \text{sensor inactive} \\ 1: & \text{sensor active} \end{array} \right\}$

- Find the best r sensor location such as to maximize some utility function (e.g. identification accuracy, information gain, etc.)
- **Challenges:**
  1. Brute force search for an optimal design is combinatorially prohibitive. It requires $\binom{\mathrm{N_s}}{r}$ function evaluations; e.g., for $\mathrm{N_s} = 35$, and $r = 10$, then $\sim 2 \times 10^8$ function evaluations
  2. Each function evaluations is prohibitively expensive
     * The covariance matrix can have over 10^{12} entries ~ 8 TB
     * Need to evaluate the determinant or the trace repeatedly

Optimal Experimental Design

- Sensor placement for *optimal parameter recovery*



Experimental design: $\xi := \left\{ \begin{array}{c} \mathbf{y}_1, \ldots, \mathbf{y}_{\mathrm{N_s}} \\ w_1, \ldots, w_{\mathrm{N_s}} \end{array} \right\}$

$\mathbf{y}_1, \ldots, \mathbf{y}_{\mathrm{N_s}}$: candidate sensor locations; we can vary weights $w_i = \left\{ \begin{array}{ll} 0 & \text{sensor inactive} \\ 1 : & \text{sensor active} \end{array} \right\}$

- Find the best r sensor location such as to maximize some utility function (e.g. identification accuracy, information gain, etc.)
- **Challenges:**
  1. Brute force search for an optimal design is combinatorially prohibitive. It requires $\binom{\mathrm{N_s}}{r}$ function evaluations; e.g., for $\mathrm{N_s} = 35$, and $r = 10$, then $\sim 2 \times 10^8$ function evaluations
  2. Each function evaluations is prohibitively expensive
     * The covariance matrix can have over 10^{12} entries ~ 8 TB
     * Need to evaluate the determinant or the trace repeatedly

- **Solution strategy:**
  - Gradient based optimization with relaxation $w_i \in [0, 1]$, and
  - use sparsifying penalty functions

Inverse Problem & Sensor Placement

Bayesian inverse problem: Gaussian framework

- **Forward operator:**

$$\mathbf{y} = \mathbf{F}(\theta) + \eta\,; \quad \eta \sim \mathcal{N}(0, \mathbf{\Gamma}_{\rm noise})$$

- **The prior and the likelihood:**

$$\mathbb{P}(\theta) = \mathcal{N}(\theta_{\rm pr}, \mathbf{\Gamma}_{\rm pr})\ , \quad \mathbb{P}(\mathbf{y}|\theta) = \mathcal{N}(\mathbf{F}(\theta), \mathbf{\Gamma}_{\rm noise})\ ,$$

For time-dependent model, with temporally-uncorrelated observational noise: $\mathbf{\Gamma}_{\rm noise}$ is a block diagonal with k^{th} equal to $\mathbf{R}_k$, observation error covariances at time instance $t_{\rm k}$

- **The posterior:** $\mathcal{N}(\theta^{\mathbf{y}}_{\rm post}, \mathbf{\Gamma}_{\rm post})$**:**

$$\mathbf{\Gamma}_{\rm post} = \left(\mathbf{F}^*\mathbf{\Gamma}^{-1}_{\rm noise}\mathbf{F} + \mathbf{\Gamma}^{-1}_{\rm pr}\right)^{-1} \equiv \left(\mathbf{H}_{\rm misfit} + \mathbf{\Gamma}^{-1}_{\rm pr}\right)^{-1} = \mathbf{H}^{-1}$$

$$\theta^{\mathbf{y}}_{\rm post} = \mathbf{\Gamma}_{\rm post}\left(\mathbf{\Gamma}^{-1}_{\rm pr}\theta_{\rm pr} + \mathbf{F}^*\mathbf{\Gamma}^{-1}_{\rm noise}\,\mathbf{y}\right)\ , \text{where}$$

  - $\mathbf{F}^*$ is the adjoint of the forward operator $\mathbf{F}$
  - $\mathbf{H}$ is the Hessian of the negative posterior-log
  - $\mathbf{H}_{\rm misfit}$ is the data misfit term of the Hessian (i.e. Hessian-misfit)

Experimental Design

Standard formulation

- The design $\mathbf{w}$ enters the Bayesian inverse problem through the data likelihood:

$$\pi_{\mathsf{like}}(\mathbf{y}|\theta;\mathbf{w}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{F}(\theta)-\mathbf{y}\right)^{\mathsf{T}}\mathbf{W}_{\Gamma}\left(\mathbf{F}(\theta)-\mathbf{y}\right)\right); \quad \mathbf{W}_{\Gamma} = \mathbf{\Gamma}_{\rm noise}^{-1/2}\mathbf{W}\mathbf{\Gamma}_{\rm noise}^{-1/2}$$

  where $\mathbf{W} = \mathbf{I}_{\mathsf{m}} \otimes \mathbf{W}_s$, and $\mathbf{W}_s = \mathrm{diag}\left(w_1, \ldots, w_{\rm N_s}\right)$

- Given the weighted likelihood, the posterior covariance of θ reads:

$$\mathbf{\Gamma}_{\rm post}(\mathbf{w}) = \left[\mathbf{H}(\mathbf{w})\right]^{-1} = \left(\mathbf{F}^*\mathbf{W}_{\Gamma}\mathbf{F} + \mathbf{\Gamma}_{\rm pr}^{-1}\right)^{-1} = \left(\mathbf{H}_{\rm misfit}(\mathbf{w}) + \mathbf{\Gamma}_{\rm pr}^{-1}\right)^{-1}$$

Here, $\otimes$ is the Kronecker product.

Experimental Design

Standard formulation

- The design $\mathbf{w}$ enters the Bayesian inverse problem through the data likelihood:

$$\pi_{\text{like}}(\mathbf{y}|\theta;\mathbf{w}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{F}(\theta)-\mathbf{y}\right)^{\mathsf{T}}\mathbf{W}_{\Gamma}\left(\mathbf{F}(\theta)-\mathbf{y}\right)\right); \quad \mathbf{W}_{\Gamma} = \mathbf{\Gamma}_{\text{noise}}^{-1/2}\mathbf{W}\mathbf{\Gamma}_{\text{noise}}^{-1/2}$$

  where $\mathbf{W} = \mathbf{I}_{\mathsf{m}} \otimes \mathbf{W}_s,$ and $\mathbf{W}_s = \operatorname{diag}\left(w_1, \ldots, w_{\mathrm{N_s}}\right)$

- Given the weighted likelihood, the posterior covariance of θ reads:

$$\mathbf{\Gamma}_{\text{post}}(\mathbf{w}) = \left[\mathbf{H}(\mathbf{w})\right]^{-1} = \left(\mathbf{F}^{*}\mathbf{W}_{\Gamma}\mathbf{F} + \mathbf{\Gamma}_{\text{pr}}^{-1}\right)^{-1} = \left(\mathbf{H}_{\text{misfit}}(\mathbf{w}) + \mathbf{\Gamma}_{\text{pr}}^{-1}\right)^{-1}$$

- **Standard Approach for ODE:** find $\mathbf{w}$ that minimizes posterior uncertainty, e.g.:
  - A-optimality: $\operatorname{Tr}\left(\mathbf{\Gamma}_{\text{post}}\right)$
  - D-optimality: $\det\left(\mathbf{\Gamma}_{\text{post}}\right)$
  - etc.

Here, $\otimes$ is the Kronecker product.

Experimental Design

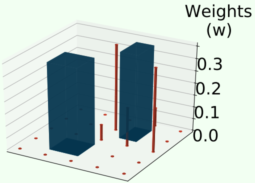
Goal-oriented formulation

what if we are interested in a prediction quantity

$$\rho = \mathcal{P}(\theta)\,,$$

rather than the parameter itself?

e.g. the average contaminant concentration within a specific distance from the buildings' walls;



Goal-Oriented ODE (GOODE)

GOODE

- Consider a linear prediction:

$$\rho = \mathbf{P}\theta \,,$$

where $\mathbf{P}$ is a linear prediction operator

- In the linear-Gaussian settings: ρ follows a Gaussian prior $\mathcal{N}(\rho_{\rm pr}, \mathbf{\Sigma}_{\rm pr})$

$$\rho_{\rm pr} = \mathbf{P}\theta_{\rm pr} \qquad \mathbf{\Sigma}_{\rm pr} = \mathbf{P}\mathbf{\Gamma}_{\rm pr}\mathbf{P}^*$$

GOODE

- Consider a linear prediction:

$$\rho = \mathbf{P}\theta \,,$$

where $\mathbf{P}$ is a linear prediction operator

- In the linear-Gaussian settings: ρ follows a Gaussian prior $\mathcal{N}(\rho_{\rm pr}, \mathbf{\Sigma}_{\rm pr})$

$$\rho_{\rm pr} = \mathbf{P}\theta_{\rm pr} \qquad \mathbf{\Sigma}_{\rm pr} = \mathbf{P}\mathbf{\Gamma}_{\rm pr}\mathbf{P}^*$$

- Given the observation $\mathbf{y}$, and the design $\mathbf{w}$, the posterior distribution of ρ is $\mathcal{N}(\rho_{\rm post}, \mathbf{\Sigma}_{\rm post})$, with

$$\rho_{\rm post} = \mathbf{P}\theta^{\mathbf{y}}_{\rm post}$$

$$\mathbf{\Sigma}_{\rm post} = \mathbf{P}\mathbf{\Gamma}_{\rm post}\mathbf{P}^* = \mathbf{P}\mathbf{H}^{-1}\mathbf{P}^* = \mathbf{P}\left(\mathbf{H}_{\rm misfit} + \mathbf{\Gamma}_{\rm pr}^{-1}\right)^{-1}\mathbf{P}^*$$

GOODE

- Consider a linear prediction:

$$\rho = \mathbf{P}\theta \,,$$

where $\mathbf{P}$ is a linear prediction operator

- In the linear-Gaussian settings: ρ follows a Gaussian prior $\mathcal{N}(\rho_{\rm pr}, \mathbf{\Sigma}_{\rm pr})$

$$\rho_{\rm pr} = \mathbf{P}\theta_{\rm pr} \qquad \mathbf{\Sigma}_{\rm pr} = \mathbf{P}\mathbf{\Gamma}_{\rm pr}\mathbf{P}^*$$

- Given the observation $\mathbf{y}$, and the design $\mathbf{w}$, the posterior distribution of ρ is $\mathcal{N}(\rho_{\rm post}, \mathbf{\Sigma}_{\rm post})$, with

$$\rho_{\rm post} = \mathbf{P}\theta_{\rm post}^{\mathbf{y}}$$

$$\mathbf{\Sigma}_{\rm post} = \mathbf{P}\mathbf{\Gamma}_{\rm post}\mathbf{P}^* = \mathbf{P}\mathbf{H}^{-1}\mathbf{P}^* = \mathbf{P}\left(\mathbf{H}_{\rm misfit} + \mathbf{\Gamma}_{\rm pr}^{-1}\right)^{-1}\mathbf{P}^*$$

GOODE Objective:

Find the design $\mathbf{w}$ that minimizes the uncertainty in ρ

GOODE: A-Optimality

space-time formulation

The G-O A-optimal design ($\mathbf{w}_{\rm opt}^{\rm GA}$)

$$
\mathbf{w}_{\rm opt}^{\rm GA} = \operatorname*{arg\,min}_{\mathbf{w} \in \mathbb{R}^{\rm N_s}} \operatorname{Tr}(\mathbf{\Sigma}_{\rm post}(\mathbf{w})) + \alpha \, \|\mathbf{w}\|
$$

$$
\text{s.t. } 0 \leq w_i \leq 1, \quad i = 1, \ldots, \rm N_s
$$

GOODE: A-Optimality

space-time formulation

The G-O A-optimal design ($\mathbf{w}_{\rm opt}^{\rm GA}$)

$$\mathbf{w}_{\rm opt}^{\rm GA} = \underset{\mathbf{w}\in\mathbb{R}^{\rm N_s}}{\arg\min}\, \mathrm{Tr}(\mathbf{\Sigma}_{\rm post}(\mathbf{w})) + \alpha\,\|\mathbf{w}\|$$
$$\text{s.t. } 0 \leq w_i \leq 1, \quad i = 1, \ldots, \mathrm{N_s}$$

- **The gradient** (discarding the regularization term) :

$$\nabla_{\mathbf{w}} \mathrm{Tr}(\mathbf{\Sigma}_{\rm post}(\mathbf{w})) = -\sum_{i=1}^{\rm N_{pred}} \zeta_i \odot \zeta_i$$

where $\zeta_i = \left(\mathbf{\Gamma}_{\rm noise}^{-\frac{1}{2}}\mathbf{F}\, [\mathbf{H}(\mathbf{w})]^{-1}\, \mathbf{P}^*\, \mathbf{e}_i\right)$, and $\mathbf{e}_i$ is the i^{th} coordinate vector in $\mathbb{R}^{\rm N_{pred}}$

Here, $\odot$ is the pointwise Hadamard product

GOODE: A-Optimality

4D-Var formulation

- **Efficient computation of the gradient:** for temporally-uncorrelated observational noise, the gradient:

$$\nabla_{\mathbf{w}} \mathrm{Tr}(\mathbf{\Sigma}_{\mathrm{post}}(\mathbf{w})) = -\sum_{k=1}^{\mathrm{m}} \sum_{j=1}^{\mathrm{N_{pred}}} \zeta_{k,j} \odot \zeta_{k,j} \, ,$$

  where

$$\zeta_{k,j} = \mathbf{R}_k^{-\frac{1}{2}} \mathbf{F}_{0,k} \, [\mathbf{H}(\mathbf{w})]^{-1} \, \mathbf{P}^* \, \mathbf{e}_i$$

  and

  * $\mathbf{e}_i$ is the i^{th} coordinate vector in $\mathbb{R}^{\mathrm{N_{pred}}}$
  * $\mathbf{F}_{0,k}$ is the forward operator that maps the parameter to the equivalent observation at time instance $t_{\mathrm{k}} \, ; k = 1, 2, \ldots, \mathrm{m}$

The G-O D-optimal design ($\mathbf{w}_{\rm opt}^{\rm GD}$)

$$\mathbf{w}_{\rm opt}^{\rm GD} = \underset{\mathbf{w} \in \mathbb{R}^{\rm N_s}}{\arg\min} \log\det\left(\mathbf{\Sigma}_{\rm post}(\mathbf{w})\right) + \alpha \left\|\mathbf{w}\right\|$$

$$\text{s.t. } 0 \leq w_i \leq 1, \quad i = 1, \ldots, \rm N_s$$

- **The gradient** (discarding the regularization term):

$$\nabla_{\mathbf{w}} \left(\log\det\left(\mathbf{\Sigma}_{\rm post}(\mathbf{w})\right)\right) = -\sum_{k=1}^{\rm m} \sum_{j=1}^{\rm N_{pred}} \xi_{k,j} \odot \xi_{k,j}$$

  where

$$\xi_{k,j} = \mathbf{R}_k^{-1/2} \mathbf{F}_{0,k} \left[\mathbf{H}(\mathbf{w})\right]^{-1} \mathbf{P}^* \mathbf{\Sigma}_{\rm post}^{-1/2}(\mathbf{w}) \mathbf{e}_j$$

  and

  1. $\mathbf{e}_i$ is the i^{th} coordinate vector in $\mathbb{R}^{\rm N_{pred}}$
  2. $\mathbf{\Sigma}_{\rm post}^{-1}(\mathbf{w}) = \mathbf{\Sigma}_{\rm post}^{-1/2}(\mathbf{w})\, \mathbf{\Sigma}_{\rm post}^{-1/2}(\mathbf{w})$

GOODE: D-Optimality

Alternative 4D-Var formulation

- **Efficient computation of the gradient:** for temporally-uncorrelated observational noise, the gradient is equivalent to:

$$\nabla_{\mathbf{w}} \left(\log \det \left(\mathbf{\Sigma}_{\rm post}(\mathbf{w}) \right) \right) = - \sum_{k=1}^{\rm m} \sum_{i=1}^{\rm N_s} \mathbf{e}_i \left(\eta_{k,i}^{\mathsf{T}} \mathbf{\Sigma}_{\rm post}^{-1} \eta_{k,i} \right)$$

with

$$\eta_{k,i} = \mathbf{P} \, \left[\mathbf{H}(\mathbf{w}) \right]^{-1} \mathbf{F}_{k,0}^{*} \, \mathbf{R}_k^{-1/2} \mathbf{e}_i$$

where $\mathbf{e}_i$ is the i^{th} coordinate vector in $\mathbb{R}^{\rm N_s}$, i.e. in the observation space

GOODE

Experiments using Advection-Diffusion Model: Setup I

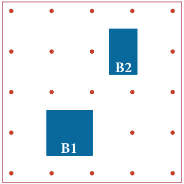
- **Numerical model (A-D):** u solves:

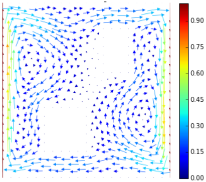
$$
\begin{aligned}
u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u &= 0 \quad \text{in } \Omega \times [0, T] \\
u(0,\, x) &= u_0 \quad \text{in } \Omega \\
\kappa \nabla u \cdot \mathbf{n} &= 0 \quad \text{on } \partial\Omega \times [0, T]
\end{aligned}
$$

  * $\Omega \in \mathbb{R}^2$ is an open and bounded domain
  * u the concentration of a contaminant in the domain Ω
  * κ is the diffusivity, and $\mathbf{v}$ is the velocity field

- **Observations:** $\mathrm{N_s} = 22$ candidate sensor locations, with
  * $t_0 = 0$, and $T = 0.8$
  * and observations are taken at time instances $\{t_k\} = \{0.4,\ 0.6,\ 0.8\}$ respectively

Domain, observational grid, and velocity field





GOODE Experiments are implemented in **hIPPYlib**

- `https://hippylib.github.io/`

Predictions: $\mathbf{P}$ predicts u at the degrees of freedom of the FE discretization withing distance ϵ from **one or both** buildings at t_{pred}.

	Vector-valued prediction	Scalar-valued prediction
	u within distance ϵ from the internal boundaries at time t_{pred}	the *"average"* u within distance ϵ from the internal boundaries at time t_{pred}
B1	$\mathbf{P}_{v0}$	$\mathbf{P}_{s0} \equiv \mathbf{v}^{\mathsf{T}} \mathbf{P}_{v0}$
B2	$\mathbf{P}_{v1}$	$\mathbf{P}_{s1} \equiv \mathbf{v}^{\mathsf{T}} \mathbf{P}_{v1}$
B1 & B2	$\mathbf{P}_{v2}$	$\mathbf{P}_{s2} \equiv \mathbf{v}^{\mathsf{T}} \mathbf{P}_{v2}$

The vector-valued operators, predict the value of u at the prediction grid-points, at prediction time. The scalar-valued operators average the vector-valued prediction QoI, i.e. $\mathbf{v} = \left(\frac{1}{\mathrm{N}_{\text{pred}}}, \ldots, \frac{1}{\mathrm{N}_{\text{pred}}} \right)^{\mathsf{T}} \in \mathbb{R}^{\mathrm{N}_{\text{pred}}}$

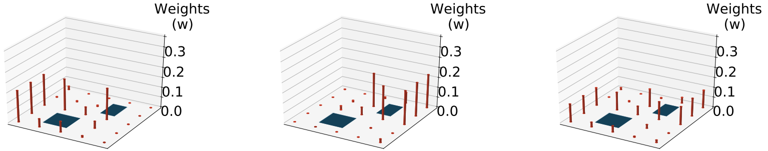
Here, we show A-GOODE results for:

Prediction operator	t_{pred}	ϵ	N_{pred}
$\mathbf{P}_{v0}$	1.0	0.02	164
$\mathbf{P}_{v1}$	1.0	0.02	138
$\mathbf{P}_{v2}$	1.0	0.02	302

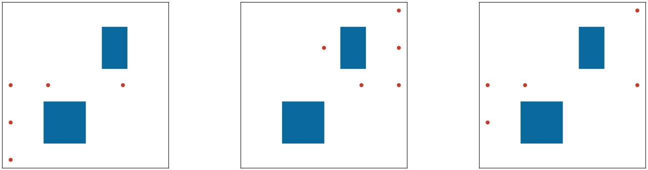
Regularization: ℓ_1 norm is used

Attia, Ahmed, Alen Alexanderian, and Arvind K. Saibaba. "Goal-Oriented Optimal Design of Experiments for Large-Scale Bayesian Linear Inverse Problems." Inverse Problems, Vol . 34, Number 9, Pages 095009 (2018).

Numerical Results: $\mathbf{P} = \mathbf{P}_{\upsilon *}$; A-GOODE



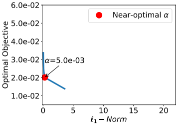
(c) $\mathbf{P}_{\upsilon 0}$; $\alpha = 10^{-4}$ (d) $\mathbf{P}_{\upsilon 1}$; $\alpha = 10^{-4}$ (e) $\mathbf{P}_{\upsilon 2}$; $\alpha = 10^{-4}$



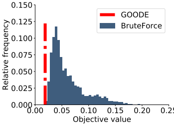
(f) $\mathbf{P}_{\upsilon 0}$; $\alpha = 10^{-4}$ (g) $\mathbf{P}_{\upsilon 1}$; $\alpha = 10^{-4}$ (h) $\mathbf{P}_{\upsilon 2}$; $\alpha = 10^{-4}$

The optimal weights $\{w_i\}_{i=1,\ldots,\mathrm{N_s}}$ are plotted on the z-axis, where the weights are normalized to add up to 1 (top row); the corresponding active sensors are plotted on the bottom row.

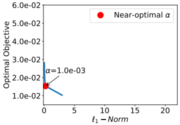
Choosing the penalty parameter: $\mathbf{P} = \mathbf{P}_{v*}$; A-GOODE



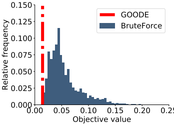
(a) $\mathbf{P}_{v0}$



(b) $\mathbf{P}_{v0}$

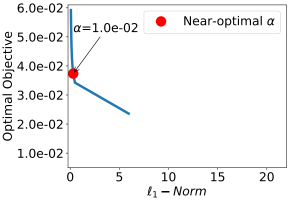


(c) $\mathbf{P}_{v1}$

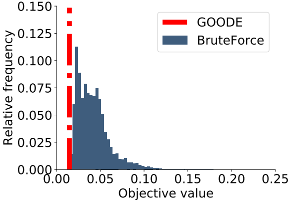


(d) $\mathbf{P}_{v1}$

A-GOODE results with a sequence of 75 penalty parameter values spaced between $[10^{-7},\ 0.2]$.



(a) $\mathbf{P}_{v2}$



(b) $\mathbf{P}_{v2}$

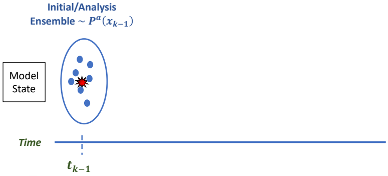
Test with a prediction operator $\mathbf{P}_{\mathrm{v}2}$.

Outline

Ensemble Kalman Filter (EnKF)

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

- **Initialize:** an analysis ensemble $\{\mathbf{x}^{\mathrm{a}}_{k-1}(e)\}_{e=1,\ldots,\mathrm{N_{ens}}}$ at t_{k-1}

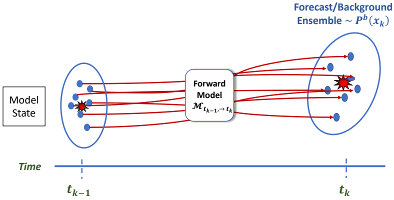


Ensemble Kalman Filter (EnKF)

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:

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



Ensemble Kalman Filter (EnKF)

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

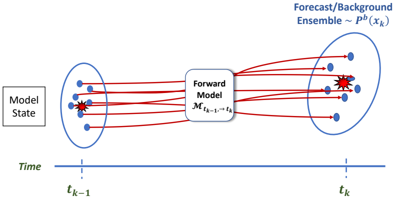
- **Initialize:** an analysis ensemble $\{\mathbf{x}^{\mathrm{a}}_{k-1}(e)\}_{e=1,\ldots,\mathrm{N_{ens}}}$ at t_{k-1}
- **Forecast:** use the discretized model $\mathcal{M}_{t_{k-1}\rightarrow t_k}$ to generate a forecast ensemble at t_k:

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

- *Forecast/Prior statistics*:

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

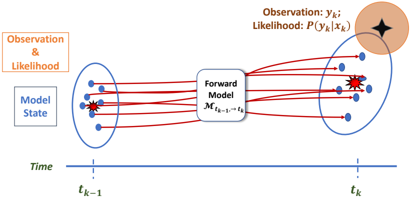
$$\mathbf{B}_k = \frac{1}{\mathrm{N_{ens}} - 1} \mathbf{X}^{\mathrm{b}}_k \left(\mathbf{X}^{\mathrm{b}}_k\right)^{\mathsf{T}} ; \quad \mathbf{X}^{\mathrm{b}}_k = [\mathbf{x}^{\mathrm{b}}_k(1) - \overline{\mathbf{x}}^{\mathrm{b}}_k, \ldots, \mathbf{x}^{\mathrm{b}}_k(\mathrm{N_{ens}}) - \overline{\mathbf{x}}^{\mathrm{b}}_k]$$



Ensemble Kalman Filter (EnKF)

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

- Given an observation $\mathbf{y}_k$ at time t_k



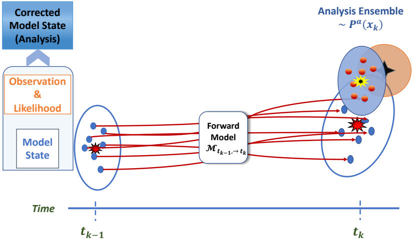
Ensemble Kalman Filter (EnKF)

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)

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

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



Ensemble Kalman Filter (EnKF)

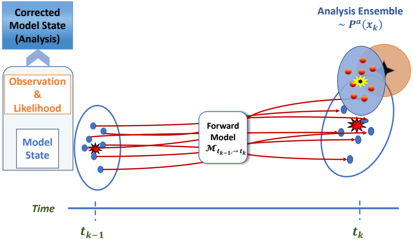
Assimilation cycle over $[t_{k-1}, t_k]$; **Analysis step**

- 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*:

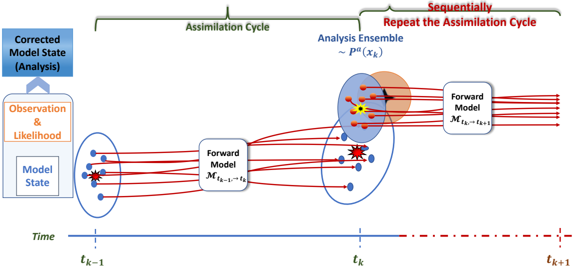
$$\mathbf{A}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{B}_k \equiv \left(\mathbf{B}_k^{-1} + \mathbf{H}_k^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{H}_k \right)^{-1}$$



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

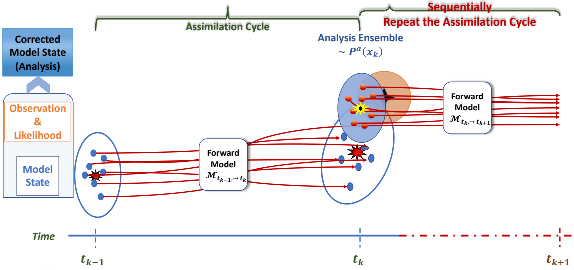


Ensemble Kalman Filter (EnKF)

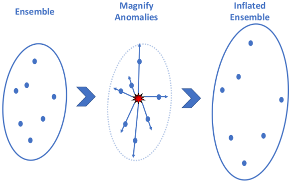
Sequential EnKF Issues

- 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**



EnKF: Inflation



Space-independent inflation:

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

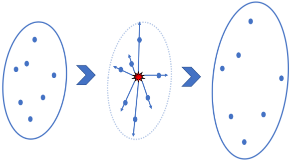
$$\widetilde{\mathbf{B}} = \frac{1}{\mathrm{N}_{\mathrm{ens}} - 1} \widetilde{\mathbf{X}^{\mathrm{b}}} \left(\widetilde{\mathbf{X}^{\mathrm{b}}} \right)^{\mathsf{T}} = \lambda \, \mathbf{B}$$

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

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

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

EnKF: Inflation



Space-independent inflation:

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

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

Space-dependent inflation: Let $\mathbf{D} := \mathrm{diag}\left(\boldsymbol{\lambda}\right) \equiv \sum_{i=1}^{\mathrm{N}_{\mathrm{state}}} \lambda_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$,

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

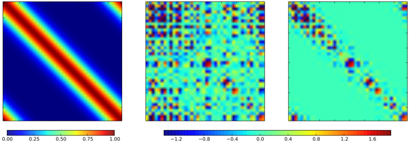
$$\widetilde{\mathbf{B}} = \frac{1}{\mathrm{N}_{\mathrm{ens}} - 1} \widetilde{\mathbf{X}^{\mathrm{b}}} \left(\widetilde{\mathbf{X}^{\mathrm{b}}}\right)^{\mathsf{T}} = \mathbf{D}^{\frac{1}{2}} \mathbf{B} \mathbf{D}^{\frac{1}{2}} .$$

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



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:

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

†

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

EnKF: Schur-Product Localization

Observation-space formulation; $\mathbf{R}-$Localization

- **Localization in observation space ($\mathrm{R}-$localization):**
  - $\mathbf{HB}$ is replaced with $\widehat{\mathbf{HB}} = \mathbf{C}^{\mathrm{loc},1} \odot \mathbf{HB}$, where
  $$\mathbf{C}^{\mathrm{loc},1} = \left[\rho_{i,j}^{o|m}\right] \; ; \, i = 1, 2, \ldots \mathrm{N_{obs}} \, ; \, j = 1, 2, \ldots \mathrm{N_{state}}$$
  - $\mathbf{HBH}^\mathsf{T}$ can be replaced with $\widehat{\mathbf{HBH}^\mathsf{T}} = \mathbf{C}^{\mathrm{loc},2} \odot \mathbf{HBH}^\mathsf{T}$, where
  $$\mathbf{C}^{\mathrm{loc},2} \equiv \mathbf{C}^{o|o} = \left[\rho_{i,j}^{o|o}\right] \; ; \, i, j = 1, 2, \ldots \mathrm{N_{obs}}$$

    - $\rho_{i,j}^{o|m}$ 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.

EnKF: Schur-Product Localization

Observation-space formulation; $\mathbf{R}-$Localization

- **Localization in observation space ($\mathbf{R}-$localization):**
  - $\mathbf{HB}$ is replaced with $\widehat{\mathbf{HB}} = \mathbf{C}^{\rm loc,1} \odot \mathbf{HB}$, where
  $$\mathbf{C}^{\rm loc,1} = \left[\rho_{i,j}^{o|m} \right] \; ; \; i = 1, 2, \ldots \mathrm{N_{obs}} \; ; \; j = 1, 2, \ldots \mathrm{N_{state}}$$
  - $\mathbf{HBH}^\mathsf{T}$ can be replaced with $\widehat{\mathbf{HBH}^\mathsf{T}} = \mathbf{C}^{\rm loc,2} \odot \mathbf{HBH}^\mathsf{T}$, where
  $$\mathbf{C}^{\rm loc,2} \equiv \mathbf{C}^{o|o} = \left[\rho_{i,j}^{o|o} \right] \; ; \; i, j = 1, 2, \ldots \mathrm{N_{obs}}$$
    - $\rho_{i,j}^{o|m}$ 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 $\mathbf{L} \in \mathbb{R}^{\mathrm{N_{obs}}}$ to model grid points, and project to observations for $\mathbf{C}^{\rm loc,2}$ [hard/unknown]
  - Let $\mathbf{L} \in \mathbb{R}^{\mathrm{N_{obs}}}$ to observation grid points; [efficient; followed here]

EnKF: Schur-Product Localization

Observation-space 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]

The parameters $\boldsymbol{\lambda} \in \mathbb{R}^{\mathrm{N}_{\mathrm{state}}}$, $\mathbf{L} \in (\mathbb{R}^{Nstate}$ or $\mathbb{R}^{Nobs})$, are generally tuned empirically!

EnKF: Schur-Product Localization

Observation-space 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

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

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

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

    - $\rho_{i,j}^{o|m}$ 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 $\mathbf{L} \in \mathbb{R}^{\mathrm{N_{obs}}}$ to model grid points, and project to observations for $\mathbf{C}^{\mathrm{loc},2}$ [hard/unknown]
  - Let $\mathbf{L} \in \mathbb{R}^{\mathrm{N_{obs}}}$ to observation grid points; [efficient; followed here]

The parameters $\boldsymbol{\lambda} \in \mathbb{R}^{\mathrm{N_{state}}}$, $\mathbf{L} \in (\mathbb{R}^{Nstate}$ or $\mathbb{R}^{Nobs})$, are generally tuned empirically!

We proposed an OED approach to automatically tune/ these parameters.

OED Approach for Adaptive Inflation

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

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

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

OED Approach for Adaptive Inflation

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

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

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

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:

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

- Decreasing λ_i reduces Ψ^{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)

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

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

- **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**:

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

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

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

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

OED Adaptive: Observation-Space Localization

- 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

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

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:**

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

- **The gradient**:

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

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

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

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

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

OED Adaptive $\mathbf{R}$−Localization

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

- **The design criterion**:

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

- **The gradient**:

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

Experimental Setup

- The model (Lorenz-96):

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

  - $\mathbf{x} \in \mathbb{R}^{40}$ is the state vector, with $x_0 \equiv x_{40}$
  - $F = 8$
- Initial background ensemble & uncertainty:
  - reference IC: $\mathbf{x}_0^{\rm True} = \mathcal{M}_{t=0 \rightarrow t=5}(-2, \ldots, 2)^{\sf T}$
  - $\mathbf{B}_0 = \sigma_0 \mathbf{I} \in \mathbb{R}^{\rm N_{state} \times N_{state}}$, 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

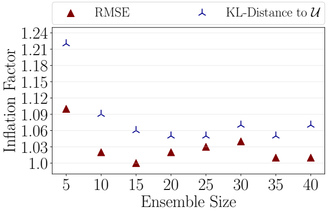
$$\mathcal{H}(\mathbf{x}) = \mathbf{H}\mathbf{x} = \left(x_1,\ x_3,\ x_5,\ \ldots,\ x_{37},\ x_{39}\right)^T \in \mathbb{R}^{20}\,.$$

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

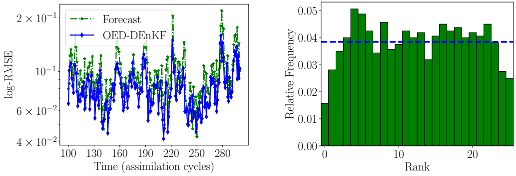
Experiments are implemented in **DATeS**

- http://people.cs.vt.edu/~attia/DATeS/
- 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: 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.



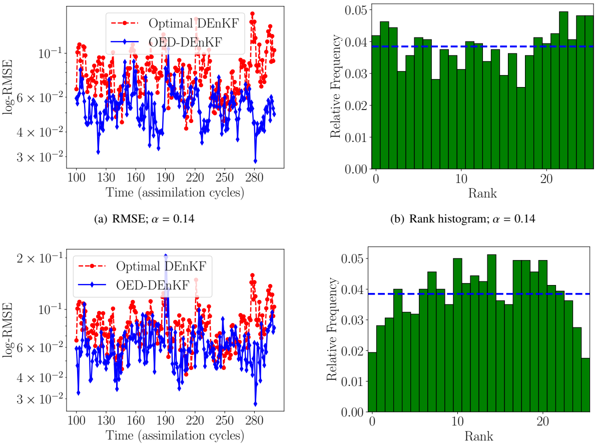
(a) RMSE

(b) Rank histogram

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

Benchmark EnKF Results

Numerical Results: A-OED Adaptive Space-Time Inflation I



(a) RMSE; $\alpha = 0.14$

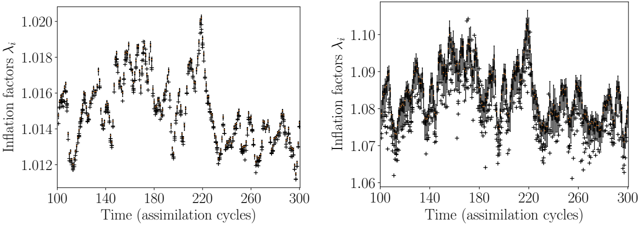
(b) Rank histogram; $\alpha = 0.14$

(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: A-OED Adaptive Space-Time Inflation II

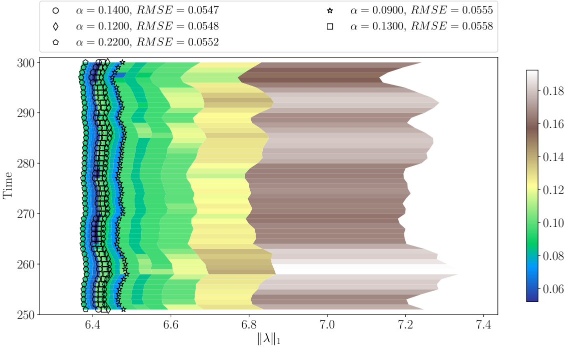


(a) $\alpha = 0.14$ (b) $\alpha = 0.04$

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

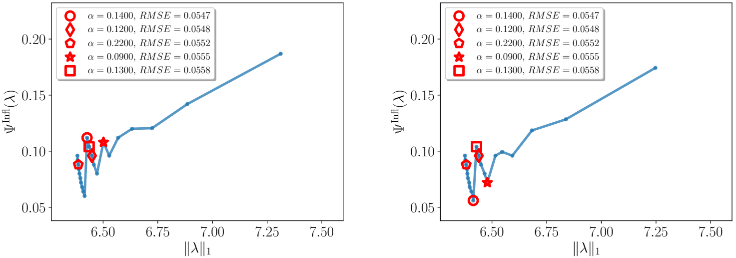
Choosing α



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.

Numerical Results; A-OED Inflation Regularization II

Choosing α

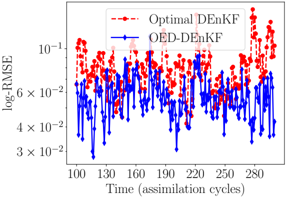


(a) Cycle 100   (b) Cycle 150

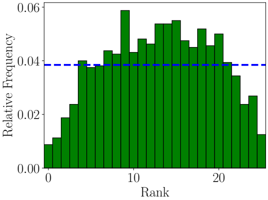
L-curve plots are are plotted for 25 equidistant values of the penalty parameter at assimilation cycles 100 and 150, respectively.

Numerical Results: A-OED Adaptive Space-Time Localization I

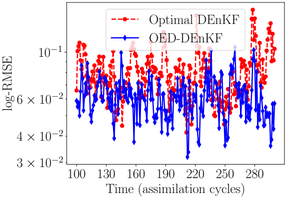
State-space formulation



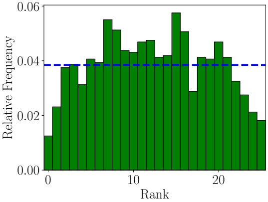
(a) RMSE; $\gamma = 0$



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



(c) RMSE; $\gamma = 0.001$

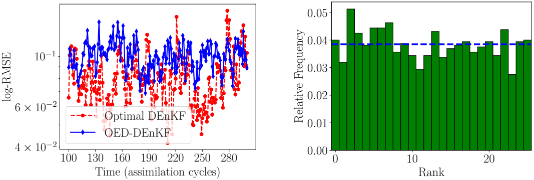


(d) Rank histogram; $\gamma = 0.001$

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

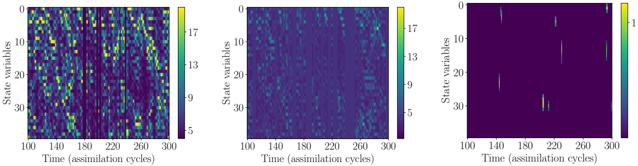
Numerical Results: A-OED Adaptive Space-Time Localization II

State-space formulation



(a) RMSE (b) Rank histogram

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

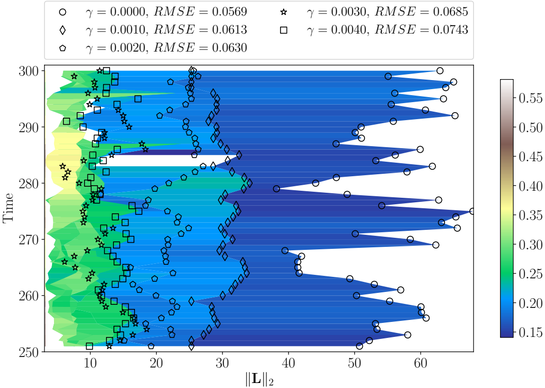


(a) $\gamma = 0.0$ (b) $\gamma = 0.001$ (c) $\gamma = 0.04$

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

Numerical Results: A-OED Adaptive Space-Time Localization

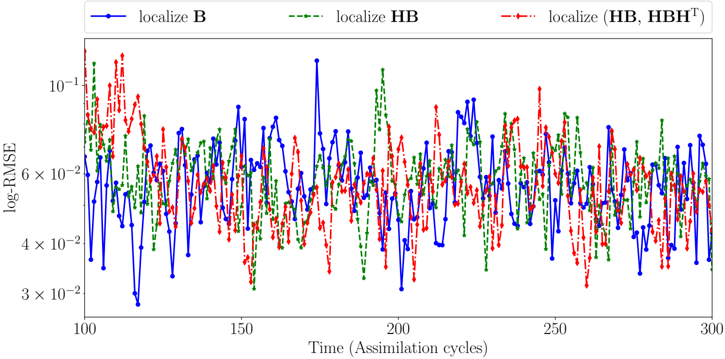
Choosing γ



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

Numerical Results: A-OED Adaptive Space-Time Localization I

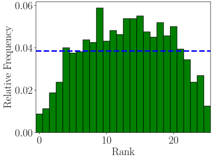
Observation-space formulation



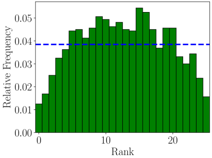
A-OED optimal localization radii $\mathbf{L}$ found by solving the OED localization problems in model state-space, and observation space respectively. No regularization is applied, i.e., $\gamma = 0$

Numerical Results: A-OED Adaptive Space-Time Localization II

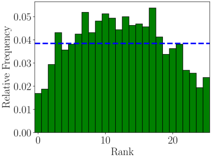
Observation-space formulation



(a) Localize $\mathbf{B}$

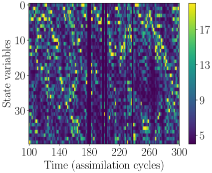


(b) Localize $\mathbf{HB}$

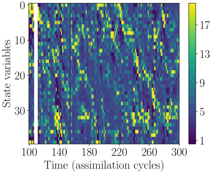


(c) Localize $\left(\mathbf{HB},\ \mathbf{HBH}^{\mathsf{T}}\right)$

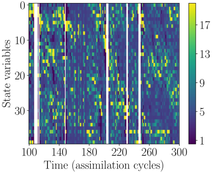
Rank histogram for A-OED localization solved in model state-space, and observation space respectively.



(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.

Concluding Remarks

- A family of sampling algorithms for Non-Gaussian DA
  - HMC sampling filter, and Cluster sampling filters
  - HMC smoother, and Reduced HMC smoothers
- Goal oriented Optimal Design of Experiments (GOODE)
  - Mathematical and algorithmic foundations for goal-oriented optimal design of experiments, for PDE-based Bayesian linear inverse problems
- OED framework for adaptive localization and inflation
  - Either A-OED inflation or localization is carried out each cycle
  - Can create a weighted objective to account for both inflation and localization
  - Unlike localization, regularization is a must for adaptive inflation