Tuning Covariance Localization Using Machine Learning

Azam Moosavi[1], Ahmed Attia[2], and Adrian Sandu[3]

[1] Biomedical Engineering department, Case Western Reserve University
[2] Mathematics and Computer Science Division, Argonne National Laboratory
[3] Computational Science Laboratory, Virginia Tech (CSL)

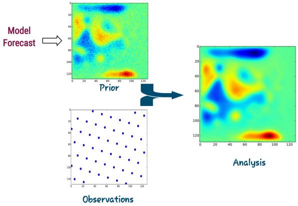
Presented by
Vishwas Rao: ANL

International Conference on Computational Science
ICCS-19
Faro, Portugal
June 12, 2019

Outline

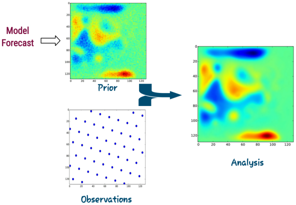
Bayesian Data Assimilation (DA)

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



Bayesian Data Assimilation (DA)

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



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

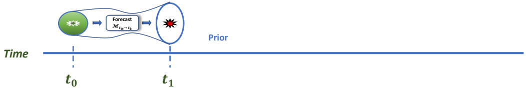
- **The posterior** $\mathbb{P}(\mathbf{x}|\mathbf{y})$: distribution of the parameter $\mathbf{x}$ conditioned on observations

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

Applications: NWP, oil reservoir, ocean, ground water, power flow, etc.

Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

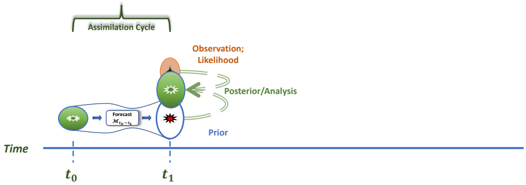


Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

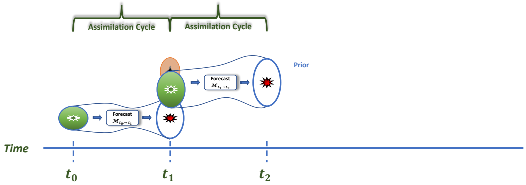
Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



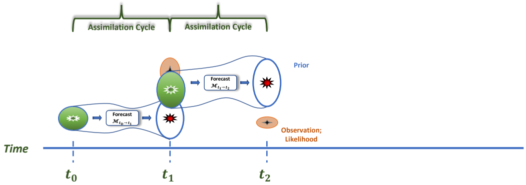
Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



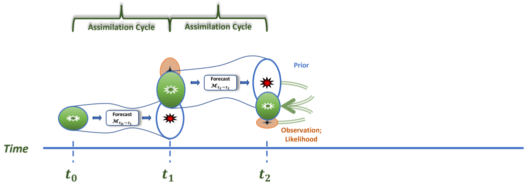
Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



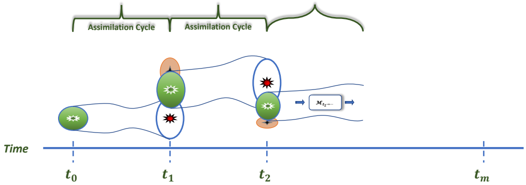
Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



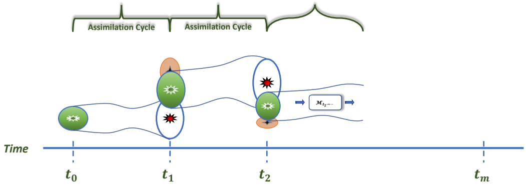
Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$

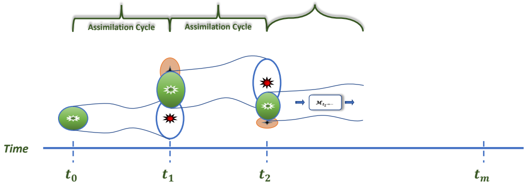


- **The Gaussian framework**:
  - Prior: $\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{true}} \sim \mathcal{N}(0, \mathbf{B})$
  - Likelihood: $\mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathrm{true}}) \sim \mathcal{N}(0, \mathbf{R})$

  $\rightarrow$ Posterior: $\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{true}} \sim \mathcal{N}(0, \mathbf{A})$

Data Assimilation: Problem Setup

- **Sequential filtering:** assimilate a single observation at a time $(\mathbf{x}_k|\mathbf{y}_k)$



- **The Gaussian framework**:
  - Prior: $\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{true}} \sim \mathcal{N}(0, \mathbf{B})$
  - Likelihood: $\mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathrm{true}}) \sim \mathcal{N}(0, \mathbf{R})$
  - $\rightarrow$ Posterior: $\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{true}} \sim \mathcal{N}(0, \mathbf{A})$
- **Kalman Filtering (KF):** predict-correct the Gaussian PDFs, i.e., means and covariance matrices

Data Assimilation: Challenges

- **Dimensionality:**
  - Observation space: $\mathrm{N}_{\mathrm{obs}} \ll \mathrm{N}_{\mathrm{state}}$
  - Model state space: $\mathrm{N}_{\mathrm{state}} \sim 10^{8-12}$
  - Covariance matrices $\in \mathbb{R}^{\mathrm{N}_{\mathrm{state}} \times \mathrm{N}_{\mathrm{state}}}$ ($\leftarrow$ **makes KF impractical**)

Data Assimilation

Ensemble Kalman Filter (EnKF): follows a Monte-Carlo approach to approximate the PDFs, i.e. uses an ensemble of model states $\{\mathbf{x}_k(e) \mid e = 1, \ldots, \mathrm{N}_{\mathrm{ens}}\}$

- Forecast step:

$$\mathbf{x}^{\mathrm{f}}_k(e) = \mathbf{M}_{t_{k-1} \to t_k}(\mathbf{x}^{\mathrm{a}}_{k-1}(e)) + \eta_k(e), \quad e = 1, \ldots, \mathrm{N}_{\mathrm{ens}}$$

$$\overline{\mathbf{x}}^{\mathrm{f}} = \frac{1}{\mathrm{N}_{\mathrm{ens}}} \mathbf{x}^{\mathrm{f}}_k(e)$$

$$\mathbf{B}_k = \frac{1}{\mathrm{N}_{\mathrm{ens}} - 1} \mathbf{X}^{\mathrm{f}}_k \left(\mathbf{X}^{\mathrm{f}}_k\right)^T \text{, s.t. } \mathbf{X}^{\mathrm{f}}_k = [\mathbf{x}^{\mathrm{f}}_k(1) - \overline{\mathbf{x}}^{\mathrm{f}}_k, \ldots, \mathbf{x}^{\mathrm{f}}_k(\mathrm{N}_{\mathrm{ens}}) - \overline{\mathbf{x}}^{\mathrm{f}}_k]$$

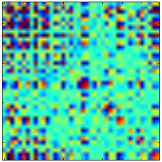
- Analysis step:

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

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

EnKF Issues

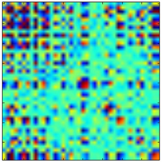
- **EnKF:**
  - → Limited-size ensemble, sampling errors, rank-deficiency
  - → **Spurious long-range correlations:**



  - → Rank-deficiency
  - → Ensemble collapse, and filter divergence

EnKF Issues

- **EnKF:**
  - → Limited-size ensemble, sampling errors, rank-deficiency
  - → **Spurious long-range correlations:**

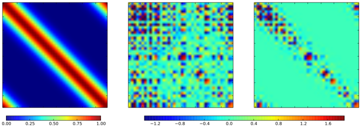


  - → Rank-deficiency
  - → Ensemble collapse, and filter divergence

- **Spurious long-range correlations:** ← **covariance localization**

EnKF: Schur-Product Localization

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



Covariance localization:

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

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

$\rightarrow$ 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 with Covariance Localization

- Forecast step:

$$
\begin{aligned}
\mathbf{x}^{\rm f}_k(e) &= \mathbf{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^{\rm a}_{k-1}(e)) + \eta_k(e), \quad e = 1, \dots, \mathrm{N}_{\rm ens} \\
\overline{\mathbf{x}}^{\rm f} &= \frac{1}{\mathrm{N}_{\rm ens}} \mathbf{x}^{\rm f}_k(e) \\
\mathbf{B}_k &= \frac{1}{\mathrm{N}_{\rm ens} - 1} \mathbf{X}^{\rm f}_k \left(\mathbf{X}^{\rm f}_k\right)^T \ , \ \text{s.t. } \mathbf{X}^{\rm f}_k = [\mathbf{x}^{\rm f}_k(1) - \overline{\mathbf{x}}^{\rm f}_k, \dots, \mathbf{x}^{\rm f}_k(\mathrm{N}_{\rm ens}) - \overline{\mathbf{x}}^{\rm f}_k]
\end{aligned}
$$

- Analysis step:

$$
\begin{aligned}
\widehat{\mathbf{B}}_k(\mathbf{L}) &= \mathbf{C}(\mathbf{L}) \odot \mathbf{B}_k \\
\widehat{\mathbf{K}}_k &= \widehat{\mathbf{B}}_k \mathbf{H}_k^T \left(\mathbf{H}_k \widehat{\mathbf{B}}_k \mathbf{H}_k^T + \mathbf{R}_k \right)^{-1} \\
\mathbf{x}^{\rm a}_k(e) &= \mathbf{x}^{\rm f}_k(e) + \widehat{\mathbf{K}}_k \left([\mathbf{y}_k + \zeta_k(e)] - \mathcal{H}_k(\mathbf{x}^{\rm f}_k(e)) \right)
\end{aligned}
$$

EnKF with Covariance Localization

- Forecast step:

$$
\begin{aligned}
\mathbf{x}_k^{\mathrm{f}}(e) &= \mathbf{M}_{t_{k-1} \to t_k}(\mathbf{x}_{k-1}^{\mathrm{a}}(e)) + \eta_k(e), \quad e = 1, \dots, \mathrm{N}_{\mathrm{ens}} \\
\overline{\mathbf{x}}^{\mathrm{f}} &= \frac{1}{\mathrm{N}_{\mathrm{ens}}} \mathbf{x}_k^{\mathrm{f}}(e) \\
\mathbf{B}_k &= \frac{1}{\mathrm{N}_{\mathrm{ens}} - 1} \mathbf{X}_k^{\mathrm{f}} \left(\mathbf{X}_k^{\mathrm{f}}\right)^T, \ \text{s.t. } \mathbf{X}_k^{\mathrm{f}} = [\mathbf{x}_k^{\mathrm{f}}(1) - \overline{\mathbf{x}}_k^{\mathrm{f}}, \dots, \mathbf{x}_k^{\mathrm{f}}(\mathrm{N}_{\mathrm{ens}}) - \overline{\mathbf{x}}_k^{\mathrm{f}}]
\end{aligned}
$$

- Analysis step:

$$
\begin{aligned}
\widehat{\mathbf{B}}_k(\mathbf{L}) &= \mathbf{C}(\mathbf{L}) \odot \mathbf{B}_k \\
\widehat{\mathbf{K}}_k &= \widehat{\mathbf{B}}_k \mathbf{H}_k^T \left(\mathbf{H}_k \widehat{\mathbf{B}}_k \mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \\
\mathbf{x}_k^{\mathrm{a}}(e) &= \mathbf{x}_k^{\mathrm{f}}(e) + \widehat{\mathbf{K}}_k \left([\mathbf{y}_k + \zeta_k(e)] - \mathcal{H}_k(\mathbf{x}_k^{\mathrm{f}}(e))\right)
\end{aligned}
$$

How to choose $\mathbf{L}$?

Adaptive Covariance Localization

Adaptive tuning of the localization radius/radii

- **Idea: Machine learning for adaptive localization** Train a ML model, and use it to predict a proper localization radius at the analysis step

Adaptive Covariance Localization

Adaptive tuning of the localization radius/radii

- **Idea: Machine learning for adaptive localization** Train a ML model, and use it to predict a proper localization radius at the analysis step
- **ML model:**
  1. Lasso
  2. Random Forest
- **space and/or time adaptive localization**
  1. Adaptive localization in time: space-independent localization
  2. Adaptive localization in time and space: space-dependent localization

Adaptive Covariance Localization

Features from forecast information

- Statistical descriptive summaries of the forecast ensemble
  1. First-order moment
  2. Second-order moment; e.g., blocks of the correlation matrix
- Forecast-observation root-mean-squared error (RMSE):

$$RMSE^{\mathbf{x}^{\rm f}|\mathbf{y}} = \frac{1}{\rm N_{obs}} \left\| \mathcal{H}\left(\mathbf{x}^{\rm f}\right) - \mathbf{y} \right\|_2$$

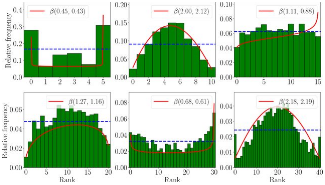
Adaptive Covariance Localization

Decision criteria

- Accuracy: analysis-observation RMSE:

$$RMSE^{\mathbf{x}^{\mathrm{a}}|\mathbf{y}} = \frac{1}{\mathrm{N_{obs}}} \left\| \mathcal{H}\left(\mathbf{x}^{\mathrm{a}}\right) - \mathbf{y} \right\|_2$$

- Dispersion: uniformity of the rank histogram.



$$D_{KL} Beta(\alpha, \beta) \| \mathcal{U} = D_{KL} Beta(\alpha, \beta) \| Beta(1.0, 1.0)$$

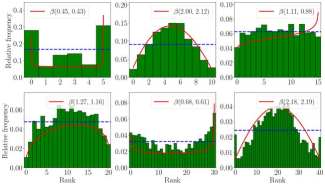
Adaptive Covariance Localization

Decision criteria

- Accuracy: analysis-observation RMSE:

$$RMSE^{\mathbf{x}^{\mathrm{a}}|\mathbf{y}} = \frac{1}{\mathrm{N_{obs}}} \left\| \mathcal{H}\left(\mathbf{x}^{\mathrm{a}}\right) - \mathbf{y} \right\|_2$$

- Dispersion: uniformity of the rank histogram.



$$D_{KL} Beta(\alpha, \beta) \| \mathcal{U} = D_{KL} Beta(\alpha, \beta) \| Beta(1.0, 1.0)$$

The decision criterion:

$$\mathcal{C}_{\mathbf{r}} = w_1\, RMSE^{\mathbf{x}^{\mathrm{a}}|\mathbf{y}} + w_2\, D_{KL} Beta(\alpha, \beta) \| Beta(1.0, 1.0)$$

with $0 \leq w_1, w2 \leq 1, \quad w_1 + w_2 = 1$

EnKF with Adaptive Localization

- Forecast step: $\mathbf{x}_k^{\mathrm{f}}(e) = \mathbf{M}_{t_{k-1}\to t_k}(\mathbf{x}_{k-1}^{\mathrm{a}}(e)) + \eta_k(e), \quad e = 1, \ldots, \mathrm{N}_{\mathrm{ens}}$
- ML training/testing:
  - Training phase:
    1. extract the features from the forecast ensemble
    2. sample a set of localization radius/radii and use each to calculate the analysis, and the decision criterion to train the model
    3. pick the winner (minimum value of the decision criterion) for the analysis step
  - Testing phase: use the fitted model to yield proper localization radius/radii
- Analysis step: use the winner/predicted radius for localization, and calculate the analysis ensemble

Numerical Experiments

- **Lorenz96 model:**

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + 8, \quad k = 1, 2, \cdots, 40$$

  All state vector components here are observed, i.e. $\mathcal{H} = \mathbf{I}$

- **Quasi Geostrophic (QG) model:**

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

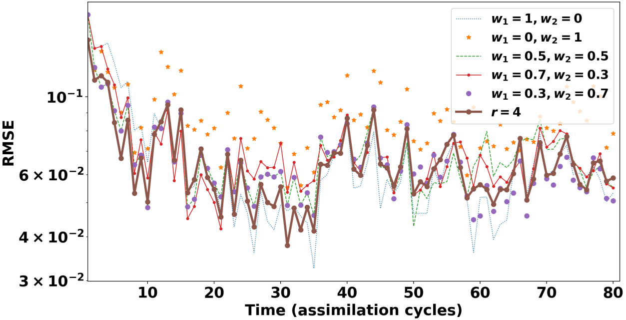
  where ψ is surface elevation or the stream function, and $F = 1600$, $\varepsilon = 10^{-5}$, $A = 2 \times 10^{-12}$

  Model discretized by a grid of size 129×129.

  A standard linear operator to observe 300 components of ψ is used.

Numerical experiments: adaptive-in-time localization

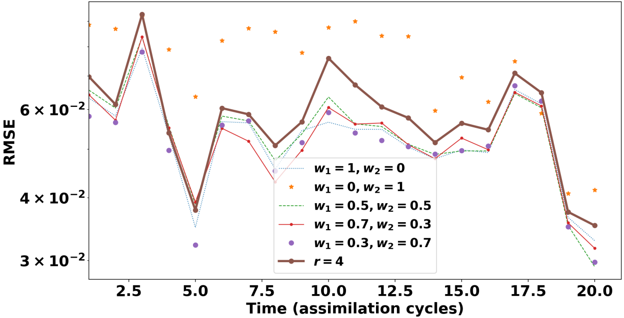
Lorenz model



Training-phase results. RMSE of adaptive localization in time vs fixed localization in Lorenz model for training (first 80 assimilation cycles)

Numerical experiments: adaptive-in-time localization

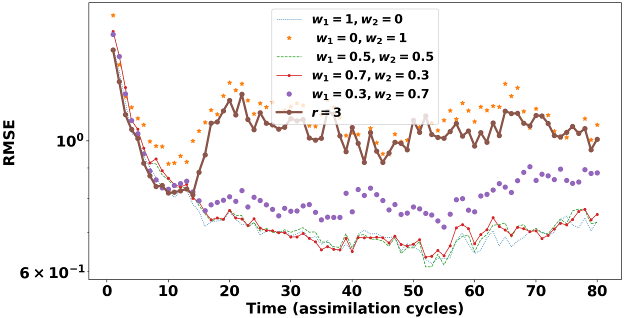
Lorenz model



Testing-phase results. RMSE over the next 20 assimilation cycles.

Numerical experiments: adaptive-in-time localization

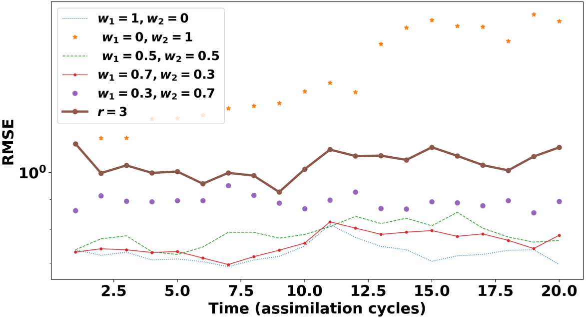
QG model



Training-phase results. RMSE of adaptive localization in time vs fixed localization in QG model for training (first 80 assimilation cycles)

Numerical experiments: adaptive-in-time localization

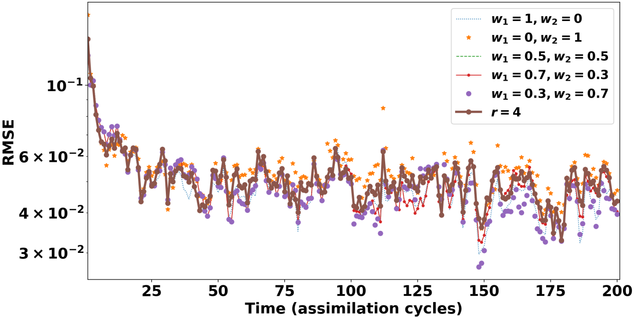
QG model



Testing-phase results, over the following 20 assimilation cycles.

Numerical experiments: space-time adaptive localization

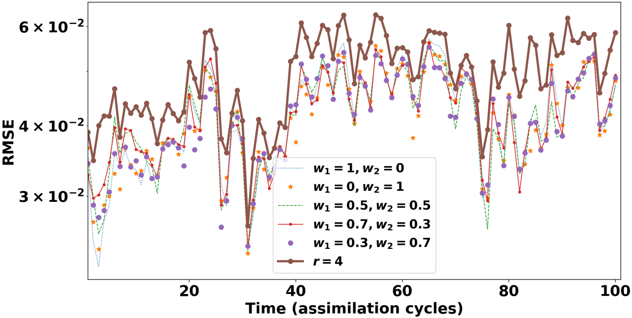
Lorenz model



Training-phase over 200 assimilation cycles.

Numerical experiments: space-time adaptive localization

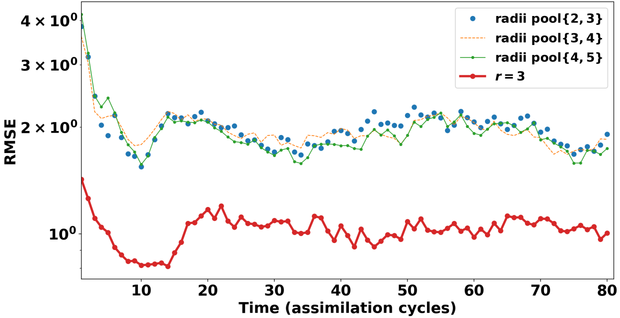
Lorenz model



Testing-phase over 100 assimilation cycles.

Numerical experiments: space-time adaptive localization

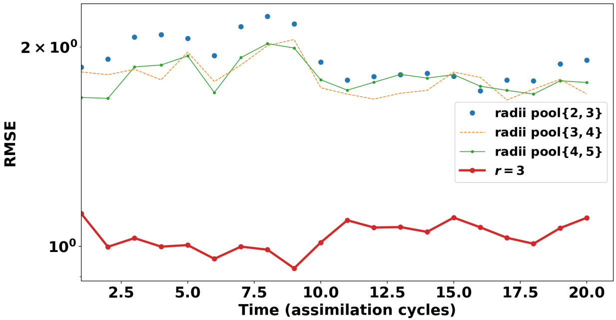
QG model



Training-phase

Numerical experiments: space-time adaptive localization

QG model



Testing-phase

Remarks & References

Numerical experiments are carried out using **DATeS**, available from:
`http://people.cs.vt.edu/~attia/dates/` or
`https://bitbucket.org/a_attia/dates/` or
`https://github.com/a-attia/dates/`

References:

1. Azam Moosavi, Ahmed Attia, and Adrian Sandu, A machine learning approach to adaptive covariance localization. arXiv preprint arXiv:1801.00548, 2018.
2. Ahmed Attia, and Adrian Sandu, DATeS: a highly extensible data assimilation testing suite v1.0, Geosci. Model Dev., 12, 629-649, https://doi.org/10.5194/gmd-12-629-2019, 2019.
3. Ahmed Attia, and Emil Constantinescu. An Optimal Experimental Design framework for adaptive inflation and covariance localization for ensemble filters. arXiv preprint arXiv:1806.10655, Submitted (2019).