A Randomized Dormant Ensemble Kalman Filter

Sampling Errors: An Alternative Look

Moha Gharamti, NCAR

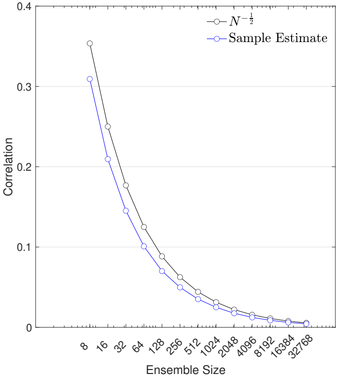
ISDA, Fort Collins

June 8, 2022

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

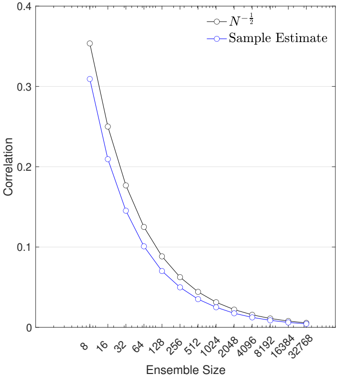
Sampling Errors in the EnKF

- *Bluntly:* Basic statistics tells us small ensemble sizes just won't cut it



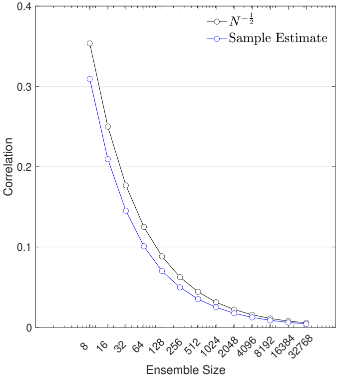
Sampling Errors in the EnKF

- *Bluntly:* Basic statistics tells us small ensemble sizes just won't cut it
    - Variance underestimation
    - Rank deficient covariances; $N << N_x$
    - Noisy and spurious correlations



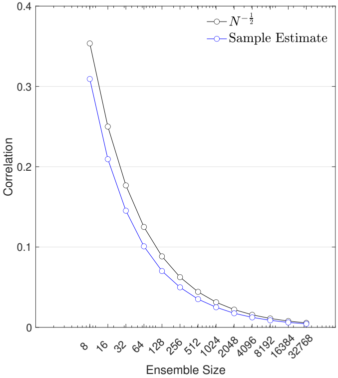
Sampling Errors in the EnKF

- *Bluntly:* Basic statistics tells us small ensemble sizes just won't cut it
  - Variance underestimation
  - Rank deficient covariances; $N << N_x$
  - Noisy and spurious correlations
- Too expensive to run with large ensemble sizes; $N \leq 100$ is typical



Sampling Errors in the EnKF

- *Bluntly:* Basic statistics tells us small ensemble sizes just won't cut it
  - Variance underestimation
  - Rank deficient covariances; $N << N_x$
  - Noisy and spurious correlations
- Too expensive to run with large ensemble sizes; $N \leq 100$ is typical
- Can become even more complex in non-Gaussian and non-linear regimes



Ways to Reduce Sampling Errors

Two major remedies:

1. **Localization:** Localize the impact of the observations to nearby state variables only (Houtekamer and Mitchell, 2001)

Ways to Reduce Sampling Errors

Two major remedies:

1. **Localization:** Localize the impact of the observations to nearby state variables only (Houtekamer and Mitchell, 2001)
2. **Inflation:** Ensemble state covariance is increased by linearly inflating each scalar component of the state while preserving the mean (Pham et al. 1998) :

$$
\begin{aligned}
\widetilde{x}_i &\leftarrow \sqrt{\lambda}\,(x_i - \overline{x}) + \overline{x}, \\
\widehat{\sigma}^2 &= \lambda \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2 ,
\end{aligned}
$$

where $\lambda > 1$ is an inflation factor and $i = 1, 2, \ldots, N$.

- Adaptive forms (Anderson 2007, 2009; El Gharamti 2018)
- Posterior (El Gharamti 2019), RTPS (Whitaker and Hamill, 2012)

What We Know About Inflation?

- <u>Usually very effective:</u>
  i. avoids filter divergence
  ii. improves assimilation quality
  iii. enhances performance even in biased scenarios

What We Know About Inflation?

- Usually very effective:
  i. avoids filter divergence
  ii. improves assimilation quality
  iii. enhances performance even in biased scenarios
- Frequently used: low-order models and large earth systems e.g., atmosphere (Raeder et al., 2021), land (Raczka et al., 2021)

What We Know About Inflation?

- Usually very effective:
  - i. avoids filter divergence
  - ii. improves assimilation quality
  - iii. enhances performance even in biased scenarios
- Frequently used: low-order models and large earth systems e.g., atmosphere (Raeder et al., 2021), land (Raczka et al., 2021)
- Large inflation can cause issues for ocean models

Received: 24 June 2017 | Revised: 28 March 2018 | Accepted: 30 March 2018

DOI: 10.1002/qj.3308

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

A global coupled ensemble data assimilation system using the Community Earth System Model and the Data Assimilation Research Testbed

Alicia R. Karspeck[1,4] | Gokhan Danabasoglu[1] | Jeffrey Anderson[1] | Svetlana Karol[2] | Nancy Collins[1] | Mariana Vertenstein[1] | Kevin Raeder[1] | Tim Hoar[1] | Richard Neale[1] | Jim Edwards[1] | Anthony Craig[3]

Randomized Dormant EnKF

RD-EnKF Motivation:

As a way to avoid numerical instabilities in certain models, is it possible to retain sufficient ensemble spread without the need for excessive inflation?

Randomized Dormant EnKF

RD-EnKF Motivation:

As a way to avoid numerical instabilities in certain models, is it possible to retain sufficient ensemble spread without the need for excessive inflation?

- **RD-EnKF** update step inherently leads to less spread reduction

Randomized Dormant EnKF

RD-EnKF Motivation:

As a way to avoid numerical instabilities in certain models, is it possible to retain sufficient ensemble spread without the need for excessive inflation?

- **RD-EnKF** update step inherently leads to less spread reduction
- The idea is to randomly break down the ensemble each assimilation step into 2 subsets:
  I. **Active:** Members within this subset go through a regular EnKF update
  II. **Dormant:** Members within this subset just sit and wait

$$N = N_a + N_d, \quad N_d = \lfloor \alpha N \rceil \quad \alpha \in [0, 1]$$

Randomized Dormant EnKF

RD-EnKF Motivation:

As a way to avoid numerical instabilities in certain models, is it possible to retain sufficient ensemble spread without the need for excessive inflation?

- **RD-EnKF** update step inherently leads to less spread reduction
- The idea is to randomly break down the ensemble each assimilation step into 2 subsets:
  I. **Active:** Members within this subset go through a regular EnKF update
  II. **Dormant:** Members within this subset just sit and wait

$$N = N_a + N_d, \quad N_d = \lfloor \alpha N \rceil \quad \alpha \in [0, 1]$$

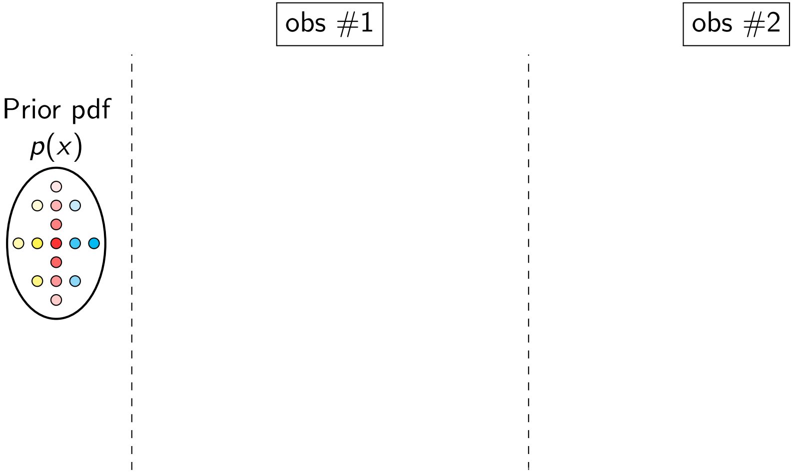
- When the observations are assimilated serially (e.g., DART, GSI), the dormant subset can change for each observation

Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$

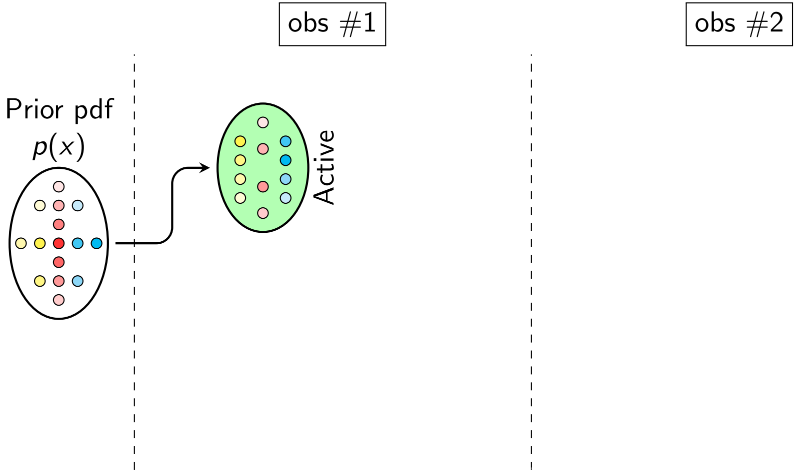
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



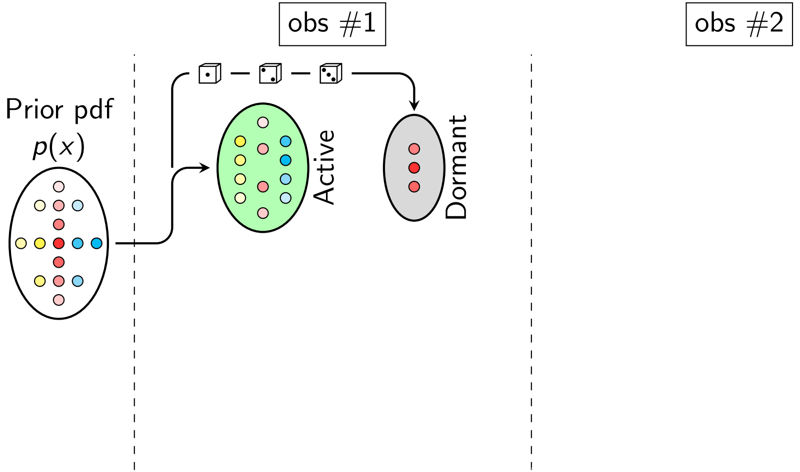
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



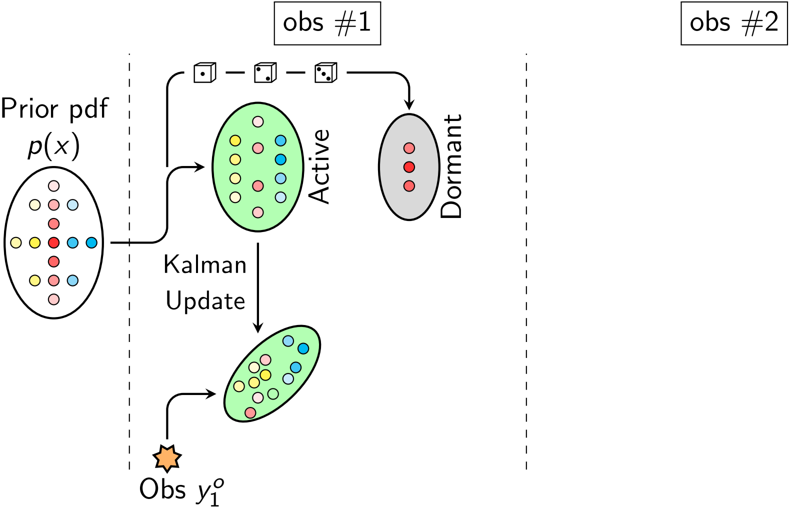
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



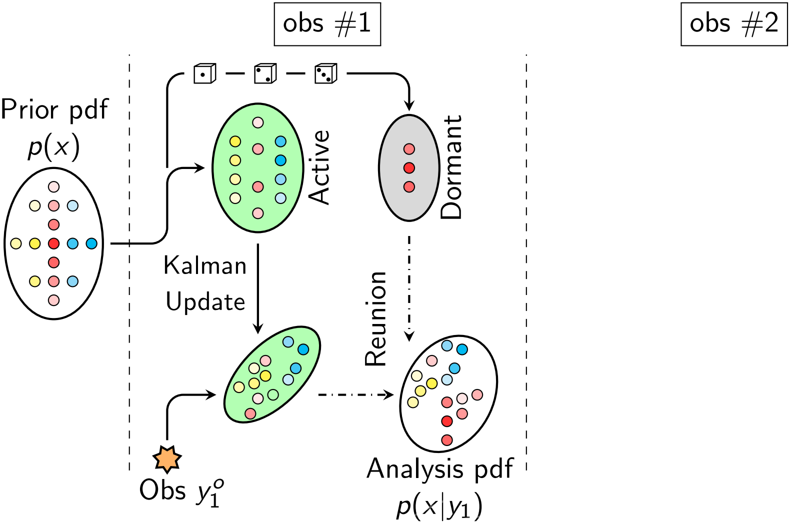
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



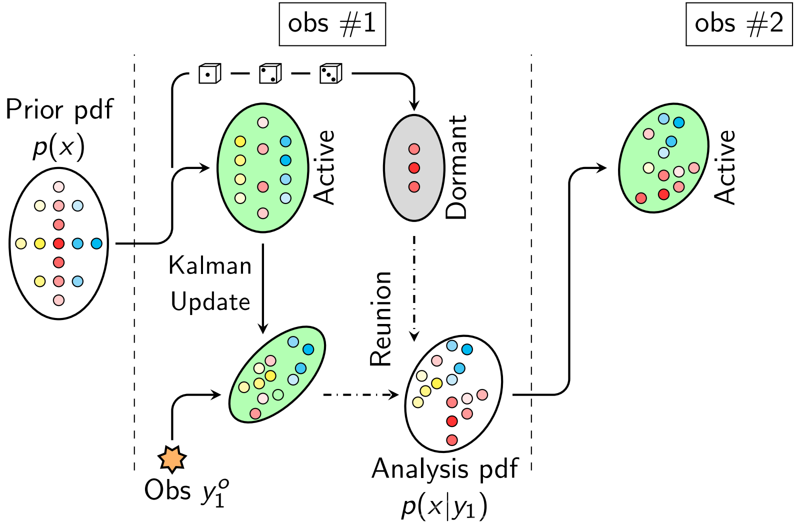
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



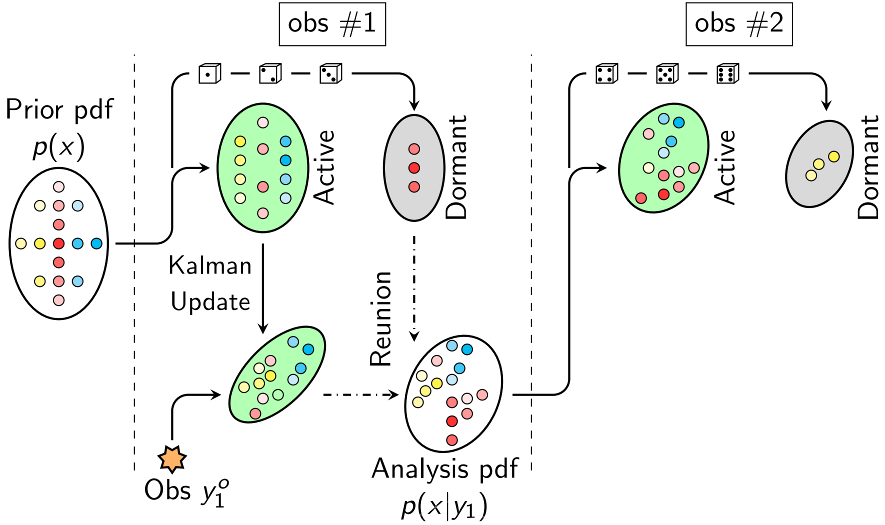
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



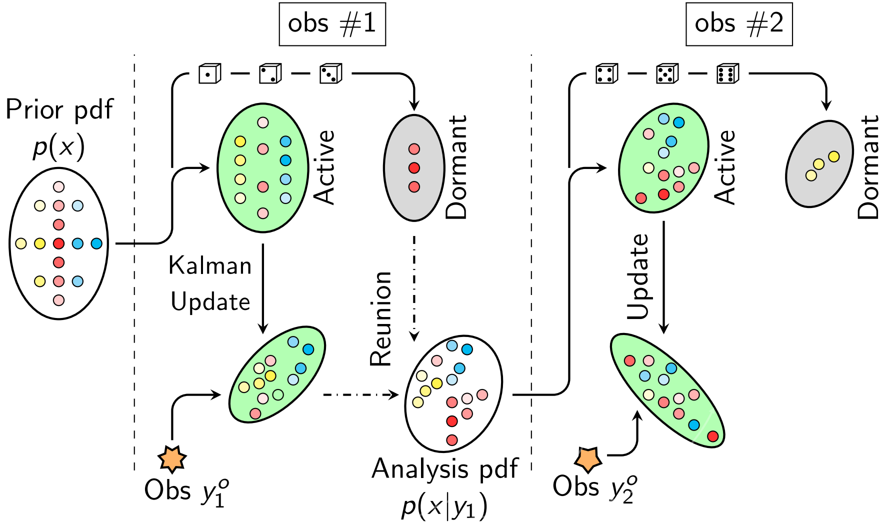
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



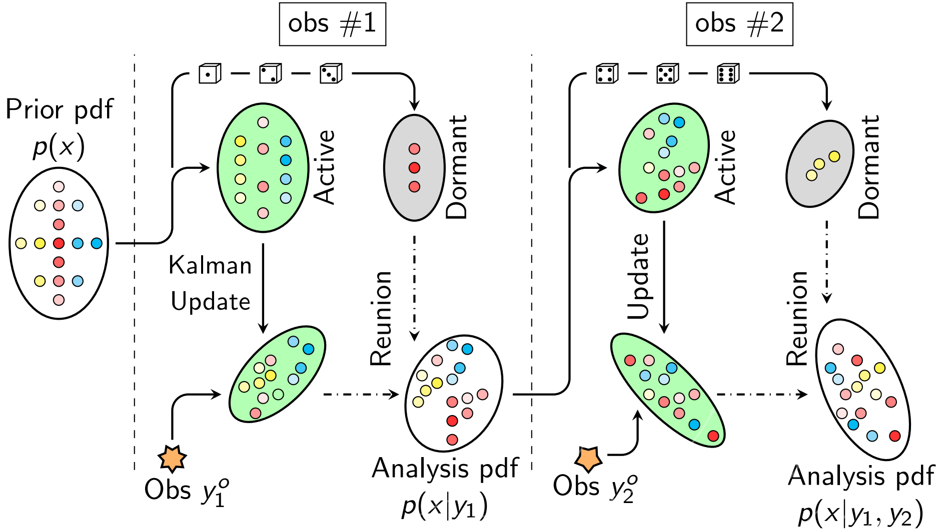
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



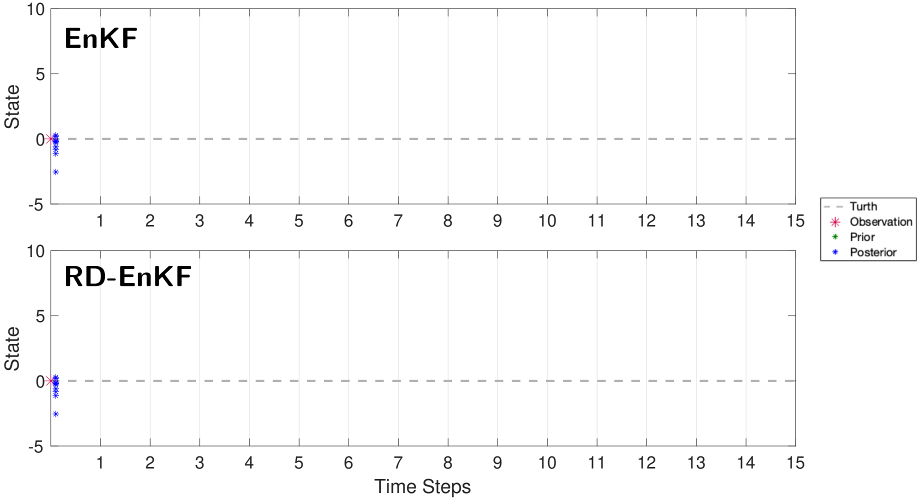
Method Illustration

RD-EnKF: 15 members, 2 available observations, $\alpha = 20\%$



A Scalar Example

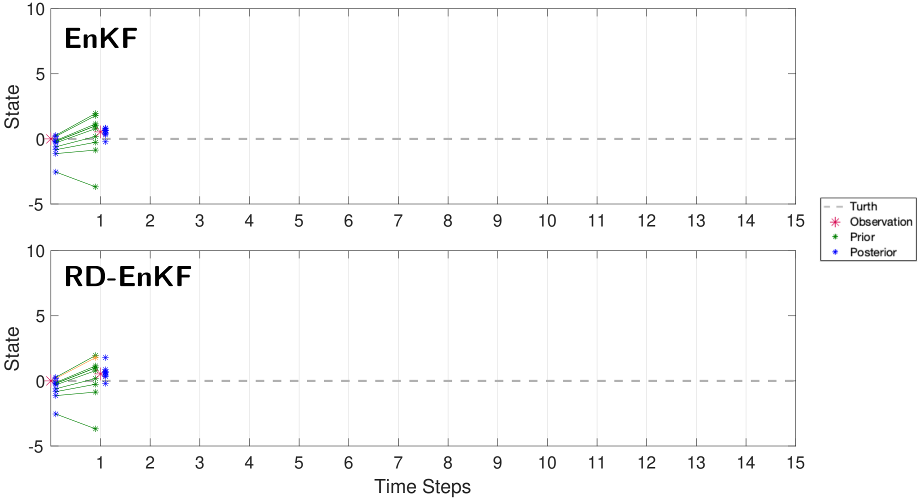
10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



NCAR
UCAR

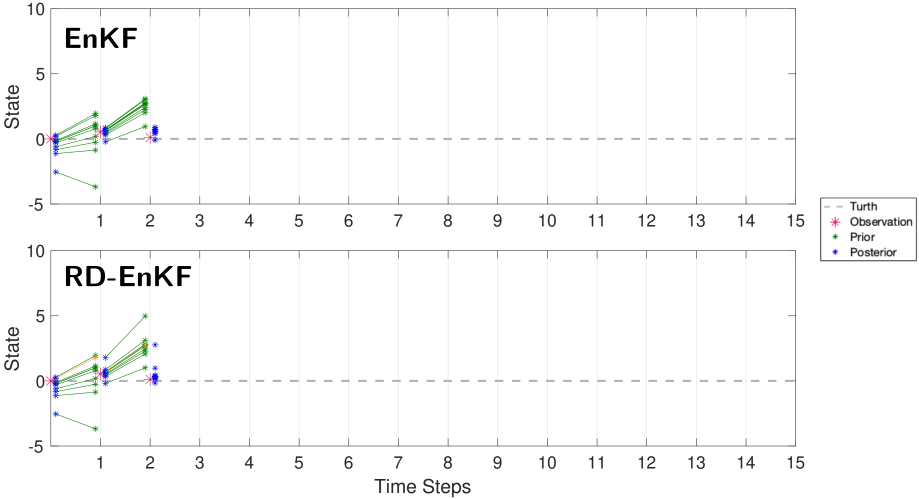
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



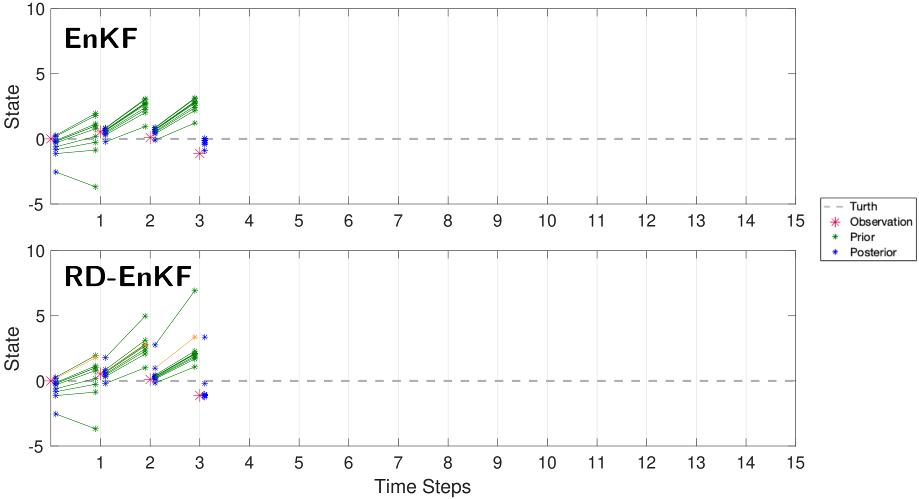
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



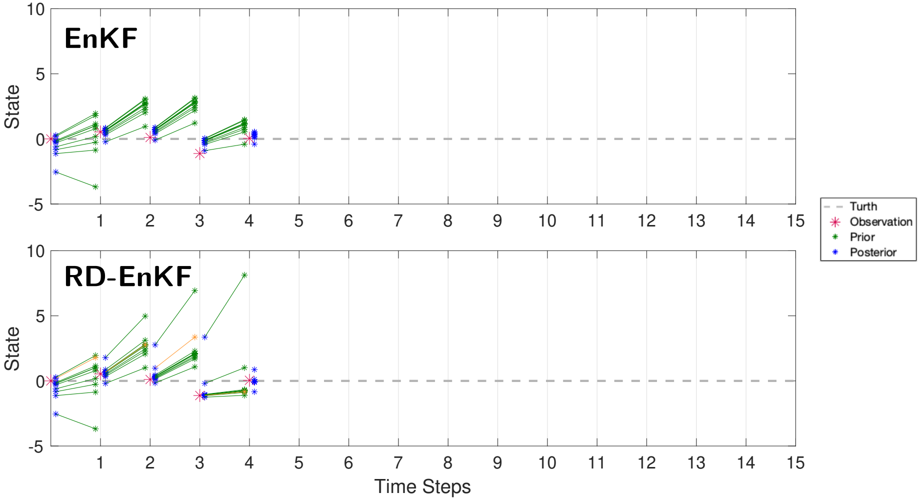
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



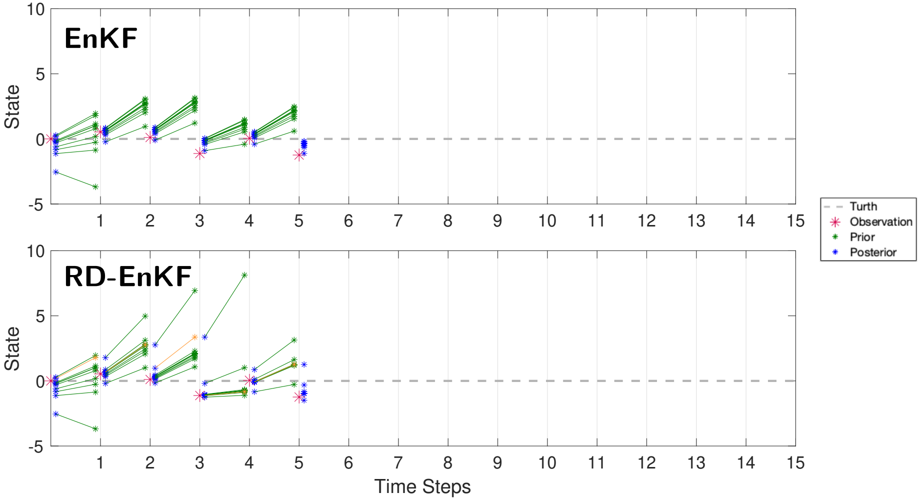
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



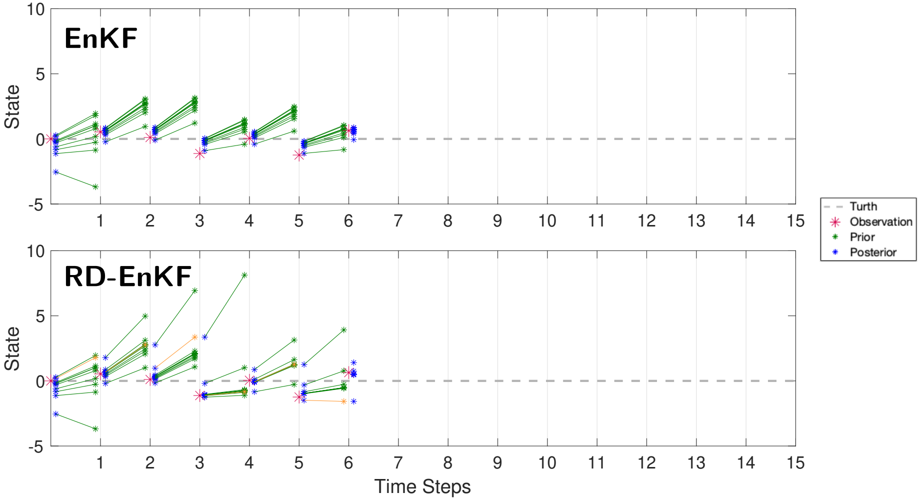
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



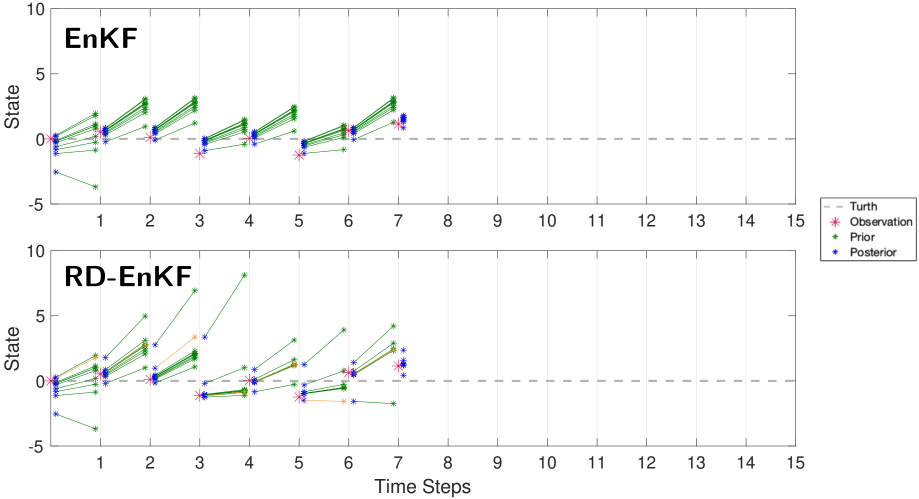
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



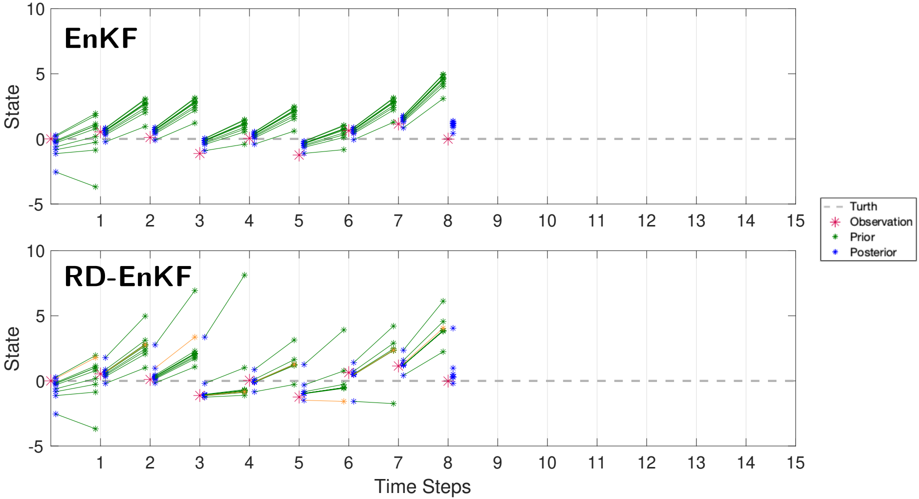
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



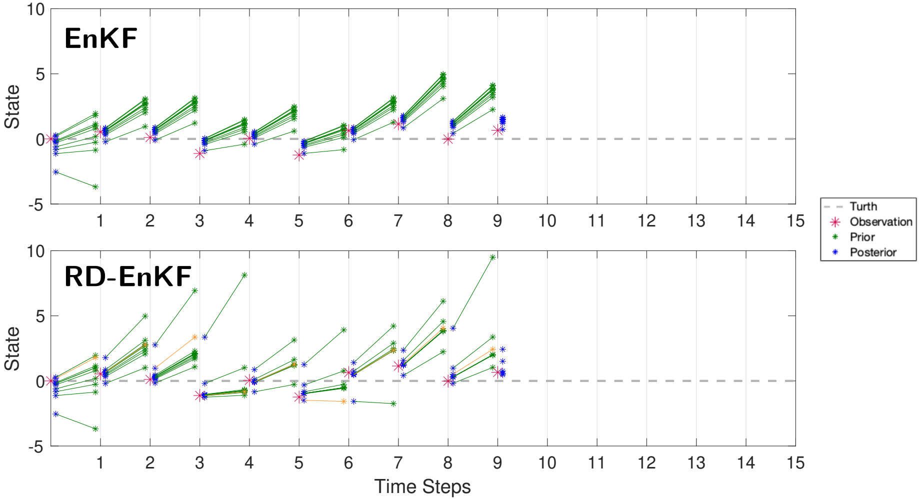
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



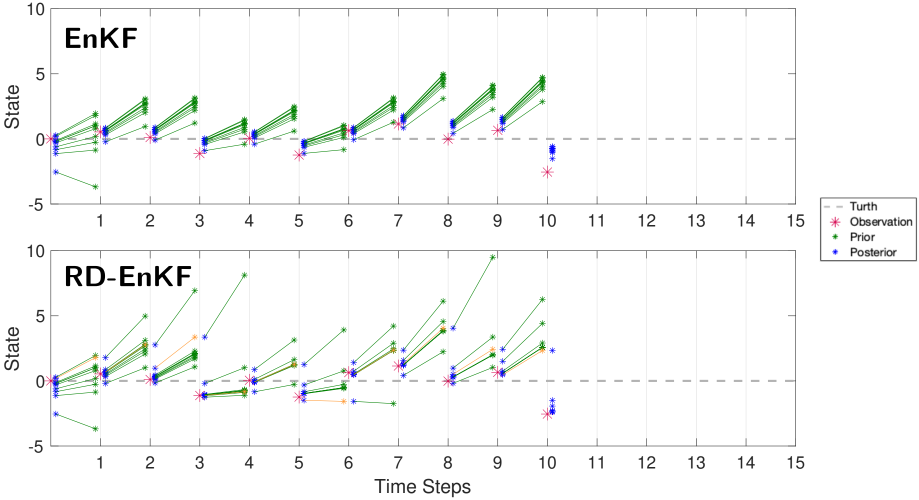
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



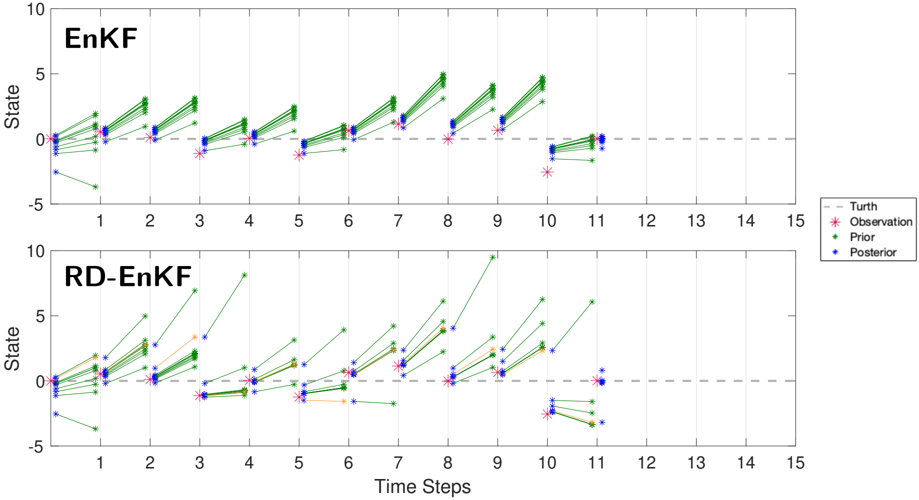
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



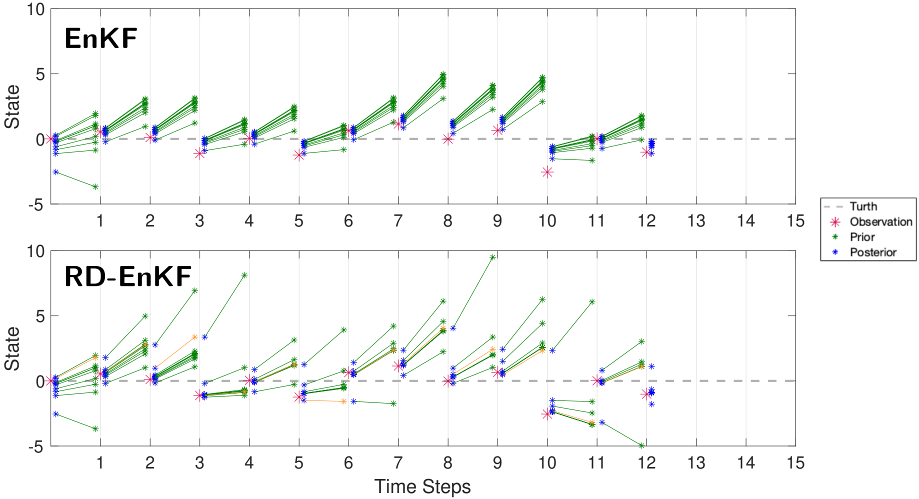
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



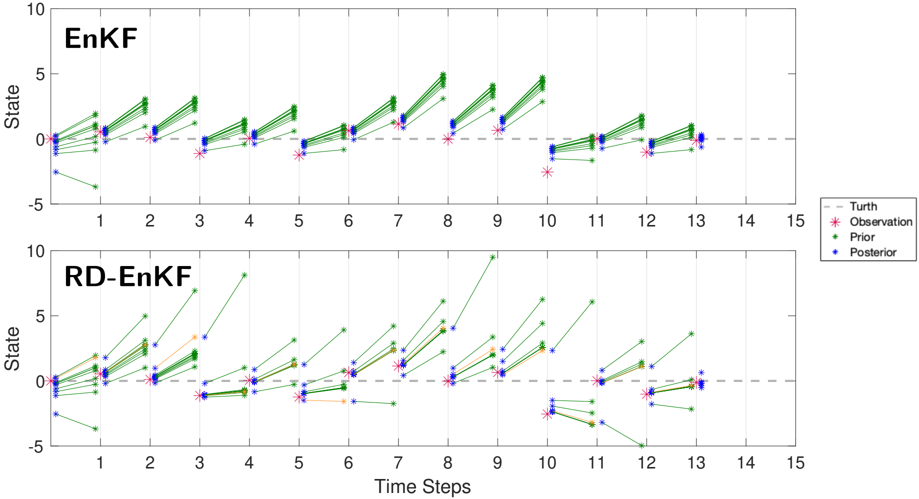
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



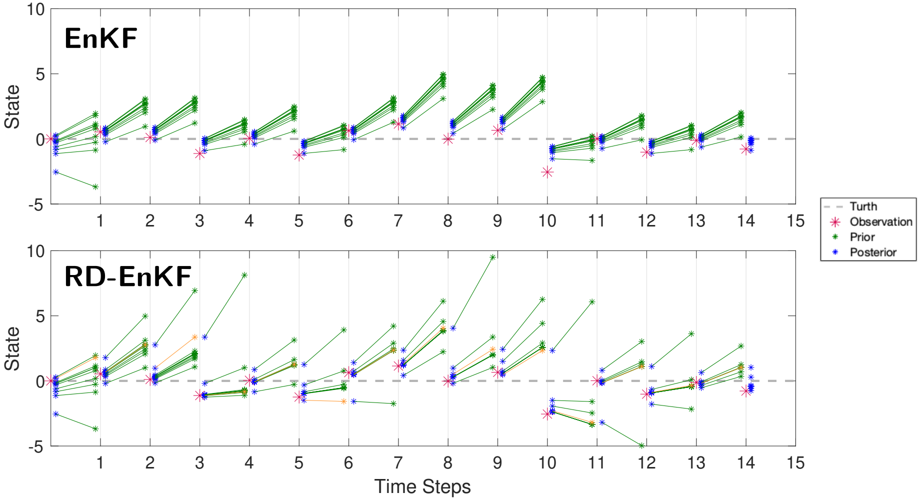
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



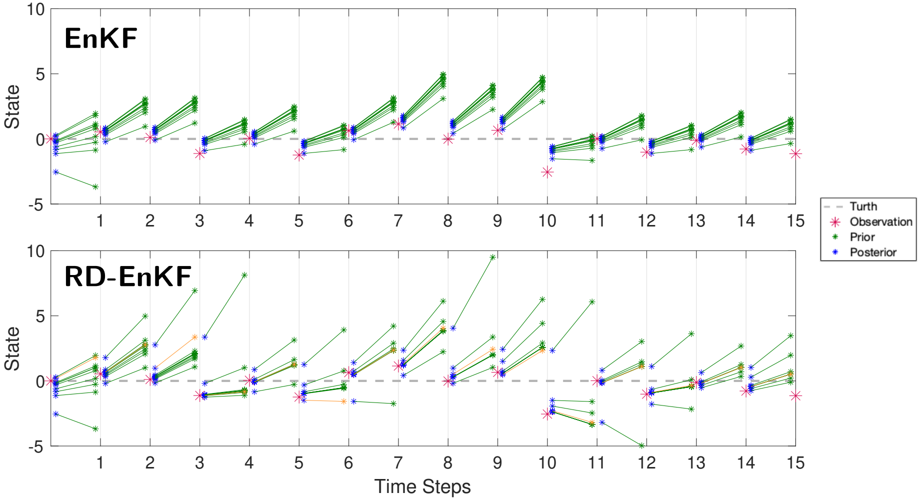
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



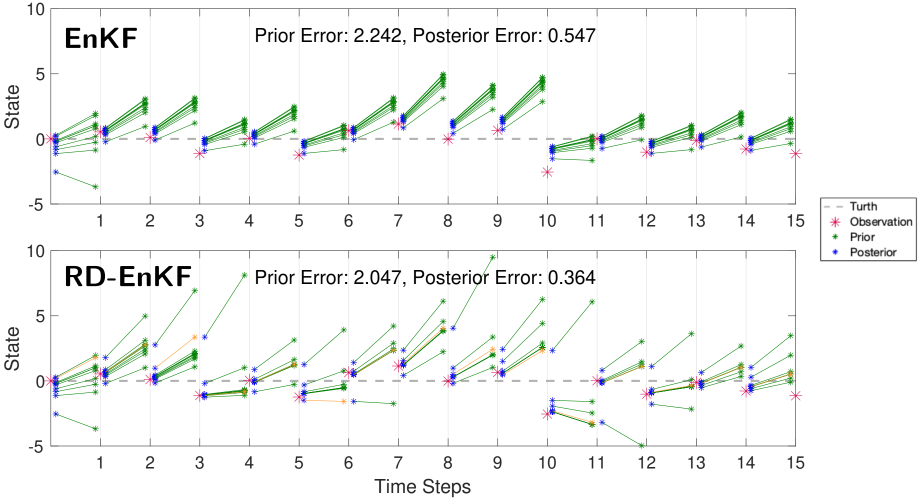
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



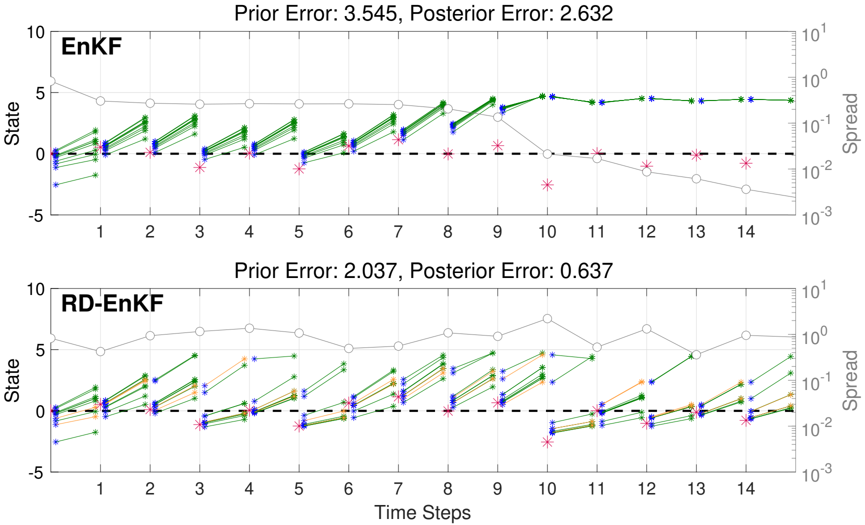
A Scalar Example

10 members, *biased* linear model, **EnKF** vs **RD-EnKF** ($\alpha = 10\%$)



A Scalar Example

10 members, *biased & nonlinear* model, **EnKF** vs **RD-EnKF** ($\alpha = 20\%$)

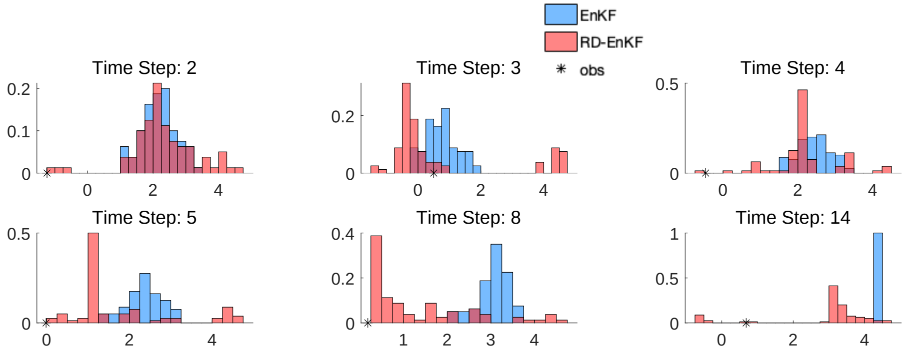


RD-EnKF: Algorithmic Features

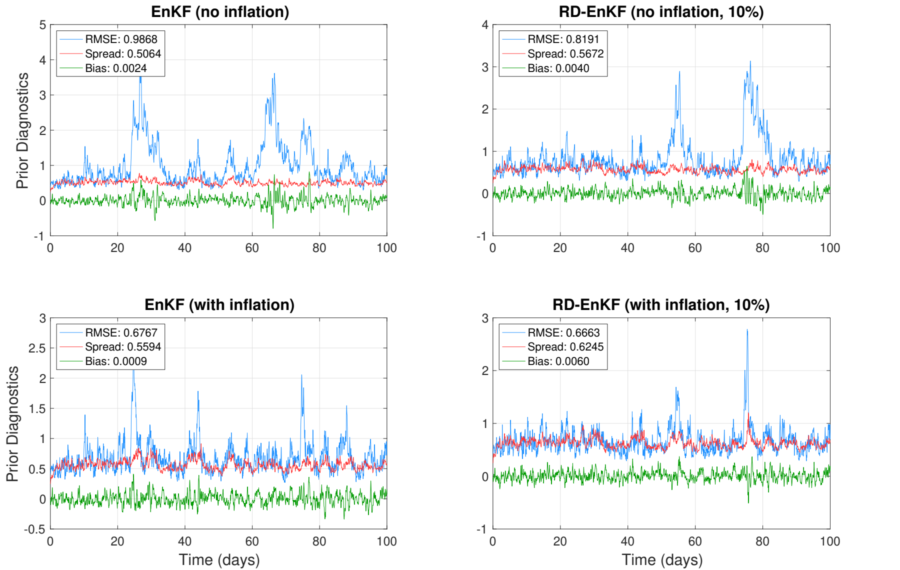
- Can be viewed as some form of posterior inflation
- Complexity: it requires random permutations of the ensemble
- Compared to the EnKF, the analysis pdf is not only inflated *but rather it's fully modified*

RD-EnKF: Algorithmic Features

- Can be viewed as some form of posterior inflation
- Complexity: it requires random permutations of the ensemble
- Compared to the EnKF, the analysis pdf is not only inflated *but rather it's fully modified*

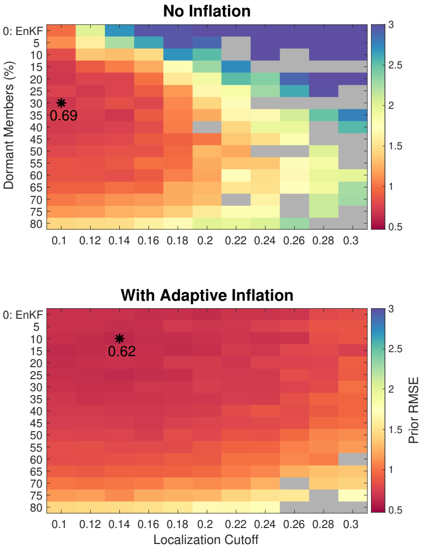


Experiments using Lorenz'96



NCAR
UCAR

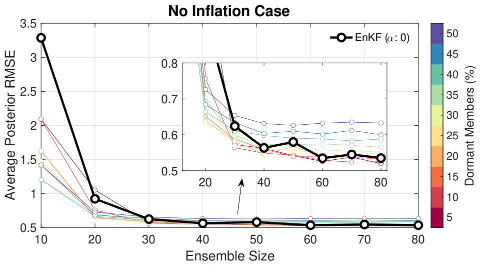
L96: Localization Sensitivity



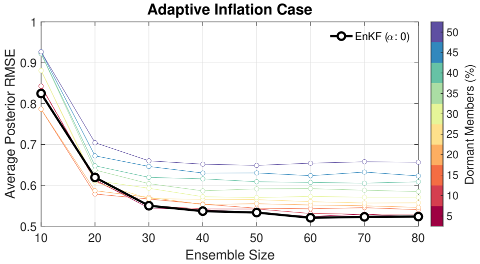
- Without inflation, the most accurate prior estimates are obtained with a `cutoff` of 0.1 radians and α : 30%
- **RD-EnKF** performs well even in poorly localized regimes
- Adaptive prior inflation stabilizes the performance and improves the accuracy
- **RD-EnKF** performance *degrades* as α increases

L96: Ensemble Size Sensitivty

- Without inflation and $N = 10$, **RD-EnKF** outperforms the standard **EnKF** for all tested dormancy rates
- For $N = 80$, **RD-EnKF** still outperforms the **EnKF** using $\alpha \in [5, 10]\%$



- With adaptive inflation, the **RD-EnKF** is only more accurate than the **EnKF** for small ensemble sizes; $N \leq 20$



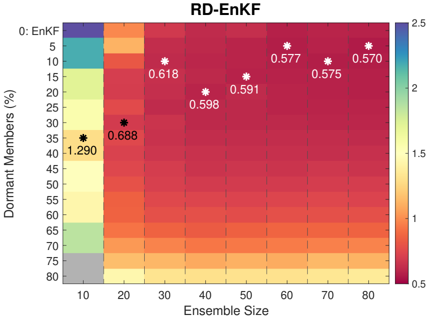
NCAR
UCAR

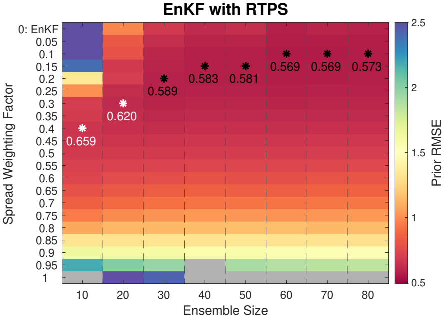
L96: Comparison to RTPS

- Unlike the **RD-EnKF**, **RTPS** update is performed with the entire ensemble and then posterior spread is partially relaxed back to the prior: $\sigma^a \leftarrow \beta\left(\sigma^f - \sigma^a\right) + \sigma^a$

L96: Comparison to RTPS

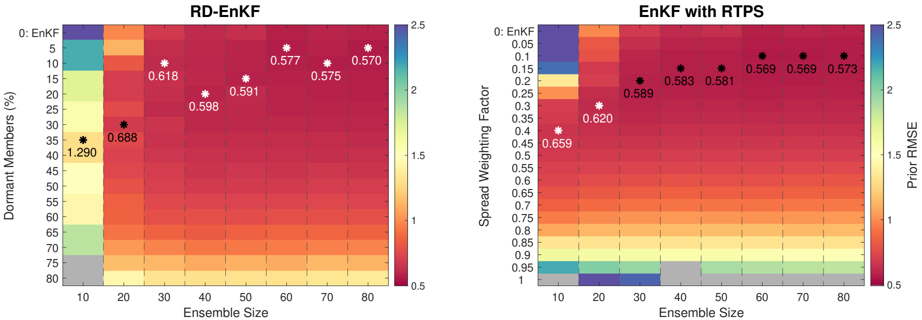
- Unlike the **RD-EnKF**, **RTPS** update is performed with the entire ensemble and then posterior spread is partially relaxed back to the prior: $\sigma^a \leftarrow \beta\left(\sigma^f - \sigma^a\right) + \sigma^a$





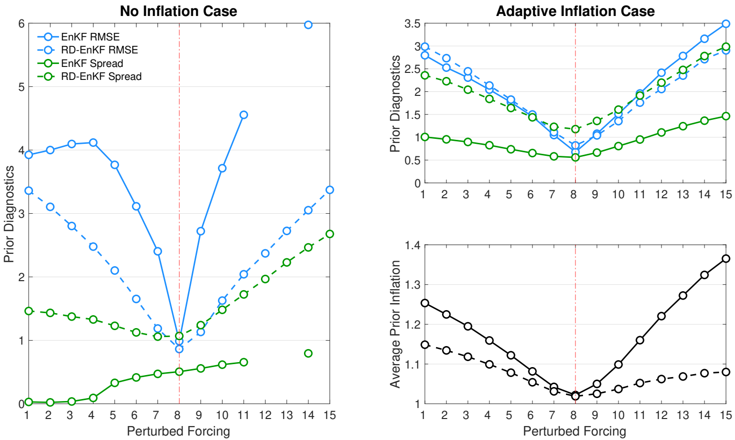
L96: Comparison to RTPS

- Unlike the **RD-EnKF**, **RTPS** update is performed with the entire ensemble and then posterior spread is partially relaxed back to the prior: $\sigma^a \leftarrow \beta\left(\sigma^f - \sigma^a\right) + \sigma^a$



- For small ensemble sizes (i.e., $N \leq 20$), **RD-EnKF** estimates are less accurate than those obtained with **RTPS**
- Ensemble spread retained by the **RD-EnKF** is consistently larger than the **RTPS**

L96: Model Errors



- Without inflation: **RD-EnKF** is *significantly* better than the **EnKF**
- With inflation: Both schemes perform equally well. The **RD-EnKF** uses less inflation but still yields larger ensemble variability

Summary

- A new method for tackling sampling errors is proposed
- The randomized dormant EnKF (**RD-EnKF**) is specifically designed for models that are less tolerant to inflation and for regimes with limited uncertainty growth

Summary

- A new method for tackling sampling errors is proposed
- The randomized dormant EnKF (**RD-EnKF**) is specifically designed for models that are less tolerant to inflation and for regimes with limited uncertainty growth

- Preliminary results in toy models show several promising aspects about the **RD-EnKF** scheme, particularly:
  (1) Ability to maintain sufficient ensemble spread after the update
  (2) Robust performance even in poorly localized domains
  (3) The need for less inflation given the inherent spread retention by the dormant members

Summary

- A new method for tackling sampling errors is proposed
- The randomized dormant EnKF (**RD-EnKF**) is specifically designed for models that are less tolerant to inflation and for regimes with limited uncertainty growth

- Preliminary results in toy models show several promising aspects about the **RD-EnKF** scheme, particularly:
  (1) Ability to maintain sufficient ensemble spread after the update
  (2) Robust performance even in poorly localized domains
  (3) The need for less inflation given the inherent spread retention by the dormant members

- Extensive testing: Ocean DA application
- Different ways to randomly (or not) select the dormant members