A RANDOMIZED DORMANT ENSEMBLE KALMAN FILTER

"An Alternative Look at Sampling Errors"

MOHA GHARAMTI, NCAR
gharamti@ucar.edu – Boulder, CO

BACKGROUND

Ensemble Kalman filters suffer from sampling errors mainly due to small ensemble sizes. Sampling errors often cause: [1] Variance underestimation, [2] Rank deficient covariances ($N << N_x$) and [3] Noisy and spurious correlations. Strategies to reduce/mitigate sampling errors include *localization* and *inflation*:

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

where $\lambda > 1$ is an inflation factor and $i = 1, 2, \ldots, N$. Adaptive prior [Anderson 2009, El Gharamti 2018] and posterior inflation [Zhang et al. 2004, Whitaker and Hamill 2012, El Gharamti 2019] forms exist. Inflation is quite an effective tool for atmospheric and land applications, however, studies have argued that it could cause numerical instabilities for certain applications such as the ocean.

Objective

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

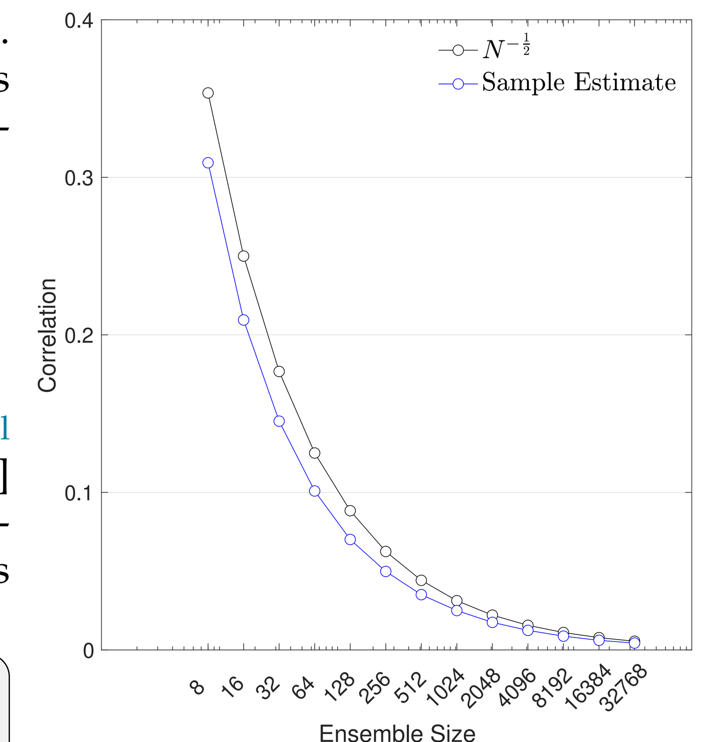


Figure 1: Sample correlation of two independent random variables using random draws with different ensemble sizes. $1/\sqrt{N}$ curve depicts the convergence rate.

RANDOMIZED DORMANT EnKF

https://dart.ucar.edu/

The Randomized Dormant EnKF (**RD-EnKF**) update step inherently leads to *less spread reduction* than a regular EnKF. The idea is to break down the ensemble each assimilation step into 2 subsets:

1. **Active:** Members within this subset go through a regular EnKF update,
2. **Dormant:** Members (chosen randomly) within this subset just sit and wait, such that:

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

where $\lfloor \cdot \rceil$ denotes the rounding or the nearest integer function and α is the dormancy rate. When the observations are assimilated serially (as in DART), the dormant subset can change for each observation. **RD-EnKF** cycling procedure is illustrated in Figure 2.

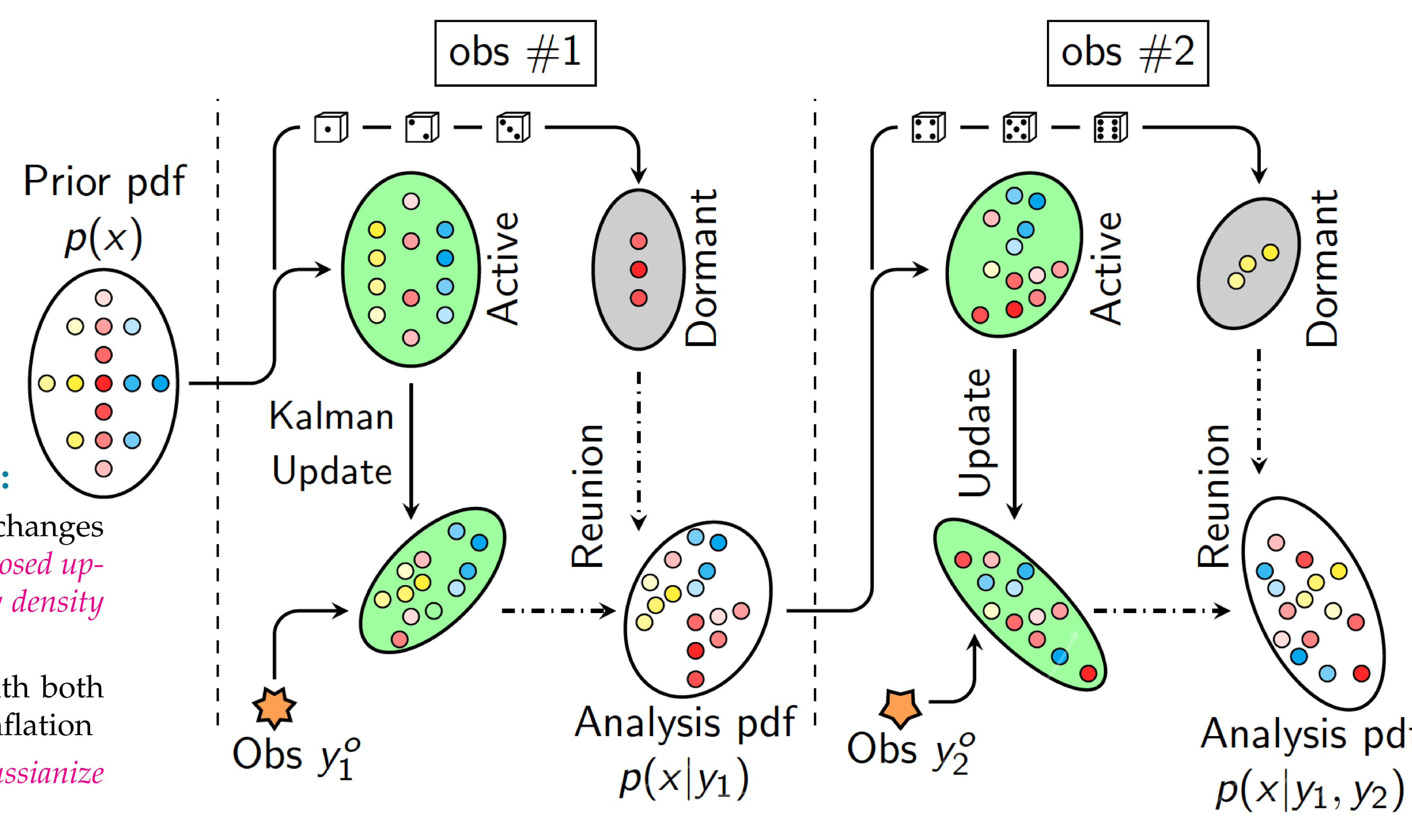


Figure 2: **RD-EnKF** illustration using an ensemble of 15 members. Each colored circle represents a single member. Two available observations y_1^o and y_2^o are assimilated one after the other. The goal is to go from the prior pdf $p(x)$ to the full analysis pdf $p(x|y_1, y_2)$. The dice on the top arrows indicate random sampling.

Algorithmic Features:

- Unlike inflation that changes the spread only, *the proposed update affects the probability density function as a whole*
- Remains compatible with both localization and prior inflation
- Update tends to *de-gaussianize* the analysis pdf

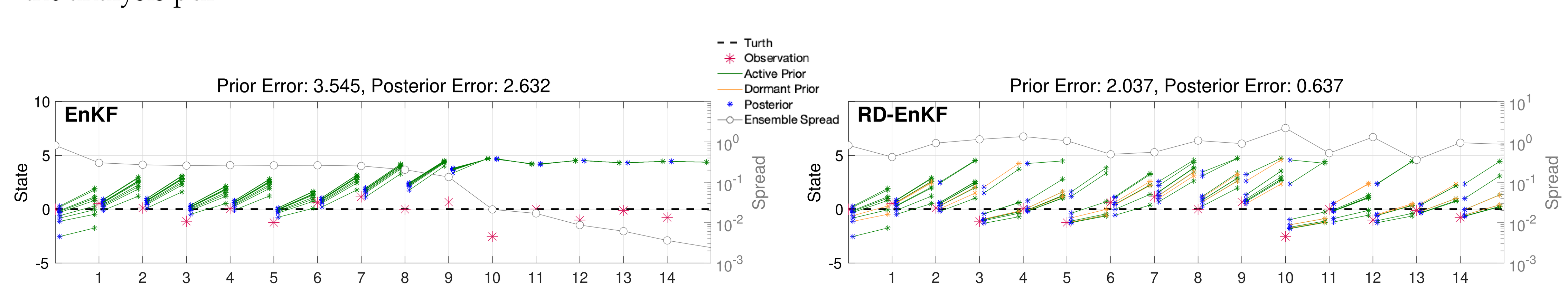
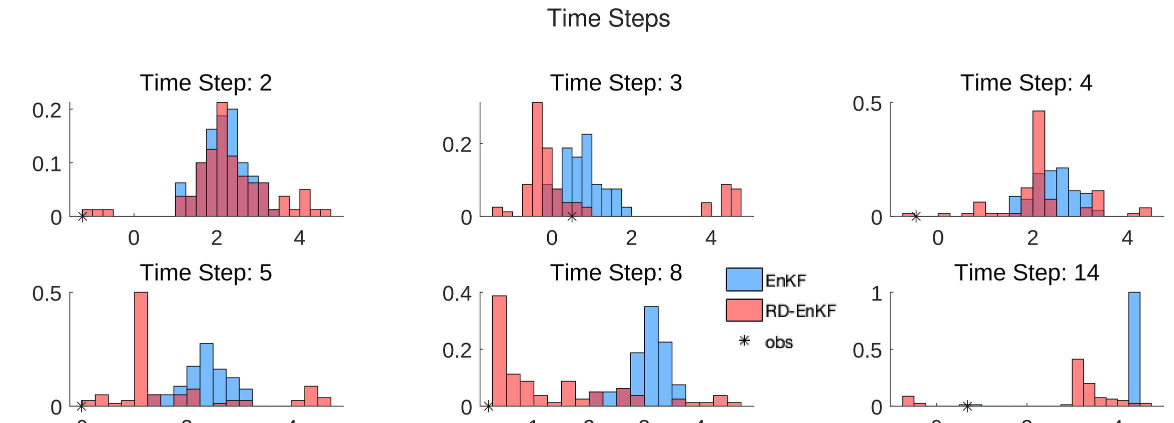


Figure 3: A 1D example from DART's "DART LAB" tutorial illustrating the behavior of the **RD-EnKF** in a controlled cycling DA system. 10 members are integrated forward in time using a biased nonlinear model. Scalar observations are sampled from the *truth* which is set to 0. The dormancy rate is set to 20% i.e., 2 dormant members (DMs; shown in orange). Overall, the **RD-EnKF** yields better prior and posterior accuracy than the **EnKF** while maintaining sufficient spread. The **EnKF** diverges after ~10 cycles. The evolution of the associated pdfs, using 80 members, in time for each scheme are displayed to the right.



LORENZ '96 EXPERIMENTS

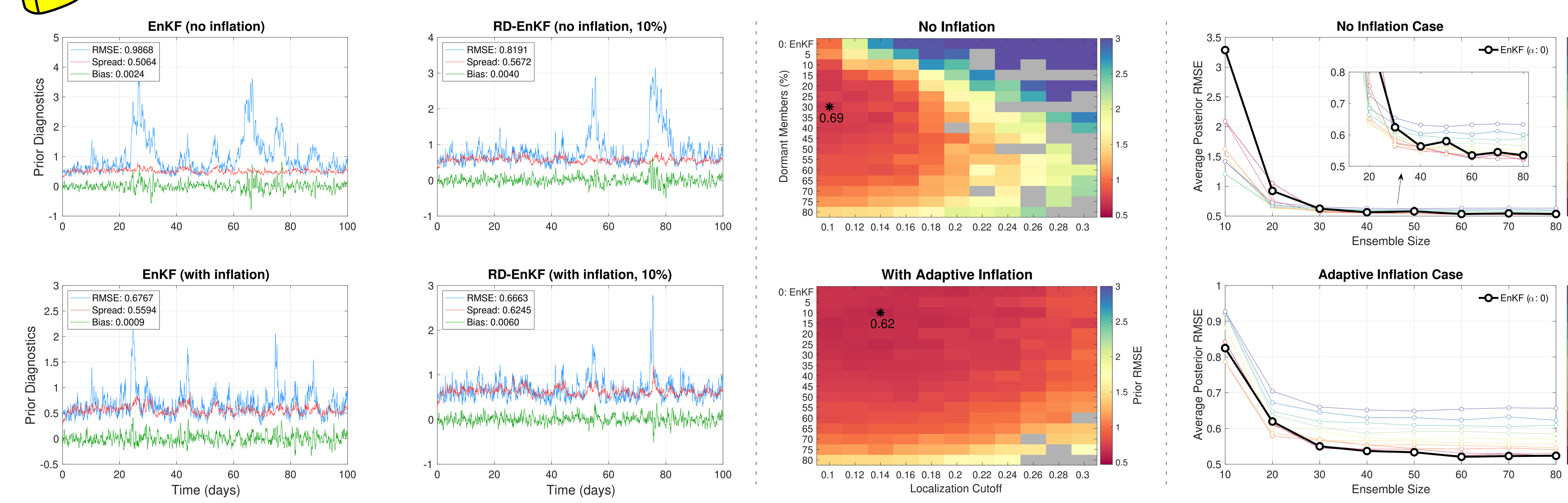


Figure 4: Observing System Simulation Experiments (perfect model) are performed using the L96 model and DART. Cycling is performed every hour using 20 observations distributed equally throughout the domain. The left panels show time-series plots of RMSE, Spread and Bias of the stochastic **EnKF** and the **RD-EnKF** using 20 members. Localization sensitivity runs are shown for both filters in the middle panels. The right panels show the overall RMSE of each scheme using a fixed localization (cutoff is 0.1) and different ensemble sizes.

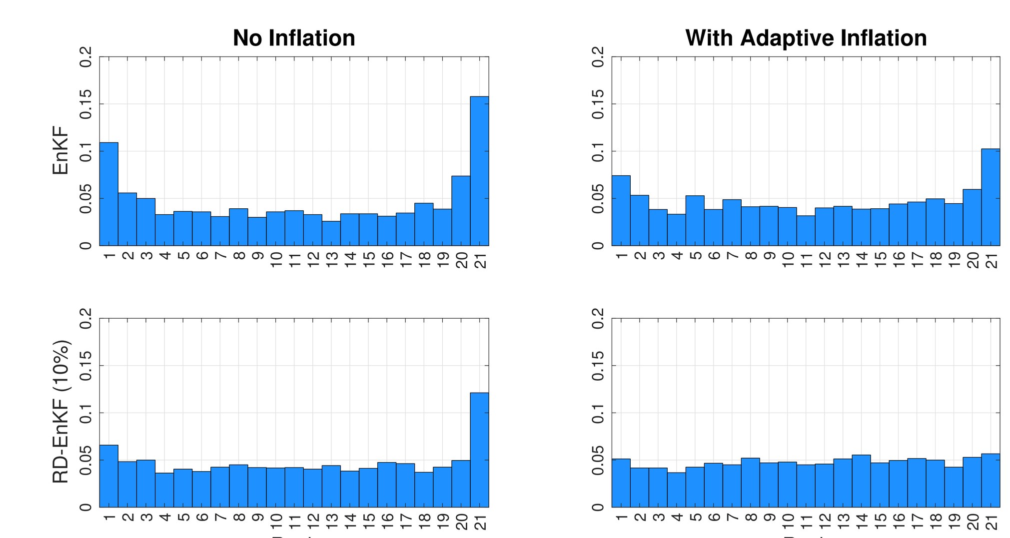
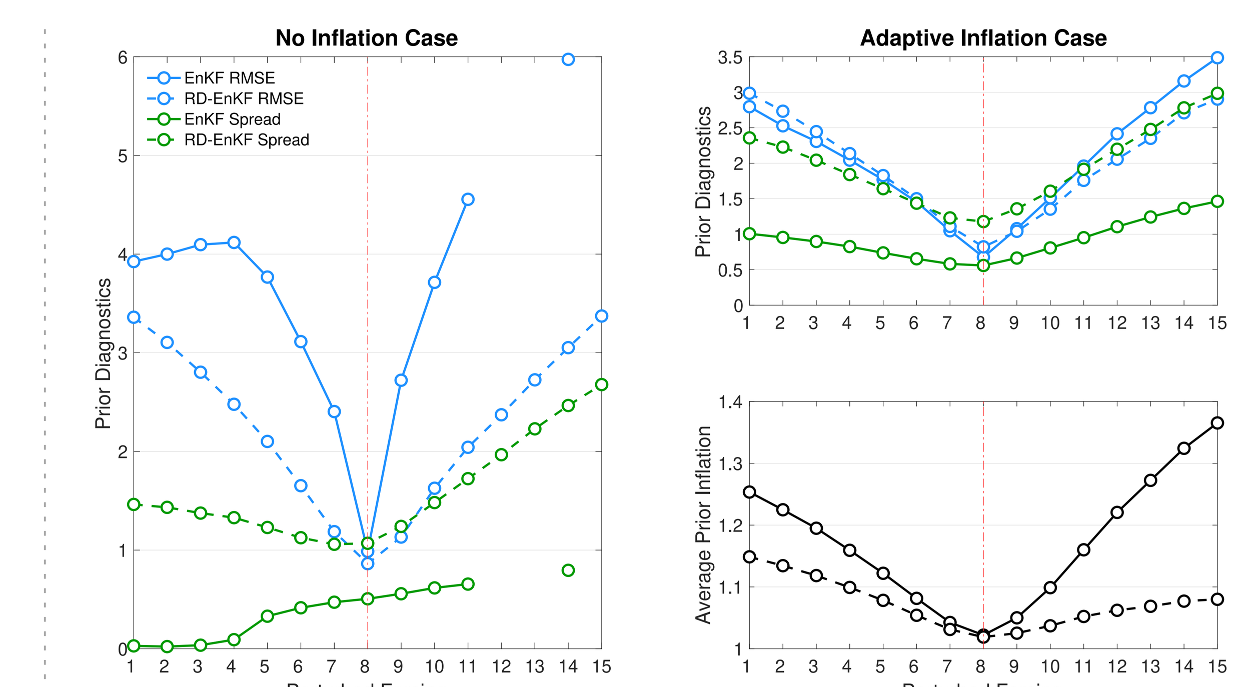
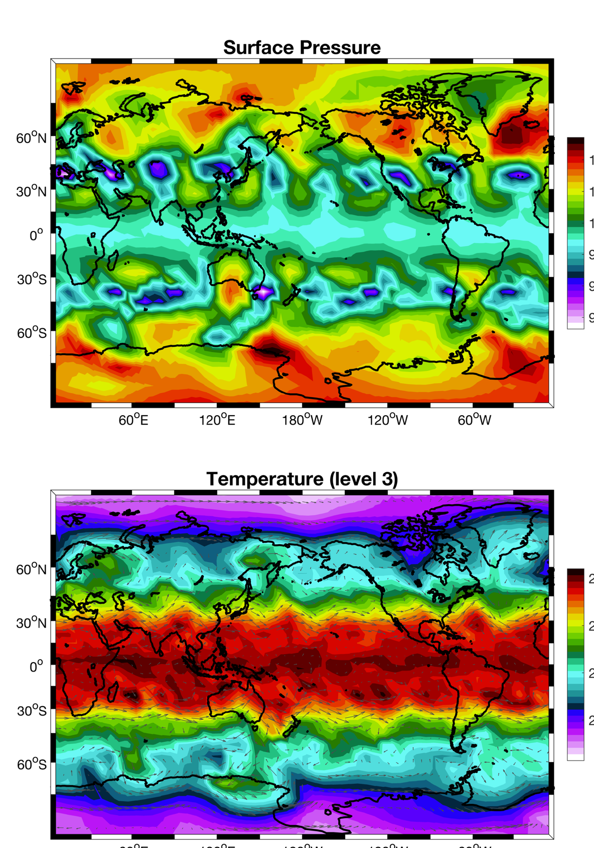


Figure 5: Rank histograms for unobserved variable #32 obtained using the **EnKF** and the **RD-EnKF** (perfect model). The right panels show the RMSE and Spread for both schemes under imperfect modeling conditions. Spatially and temporally averaged prior inflation values for each run are also shown.



IDEALIZED ATMOSPHERIC GCM EXPERIMENTS

- Dynamical core of the GFDL AM2 Bgrid model
- Minimum resolution that generates baroclinic instabilities: $60 \times 30 \times 5$
- Prognostic variables: PS, T, U, V
- 1-hr time step (total of 200 days)
- DART-callable routine: `bgrid_solo` [Anderson et al. 2005]



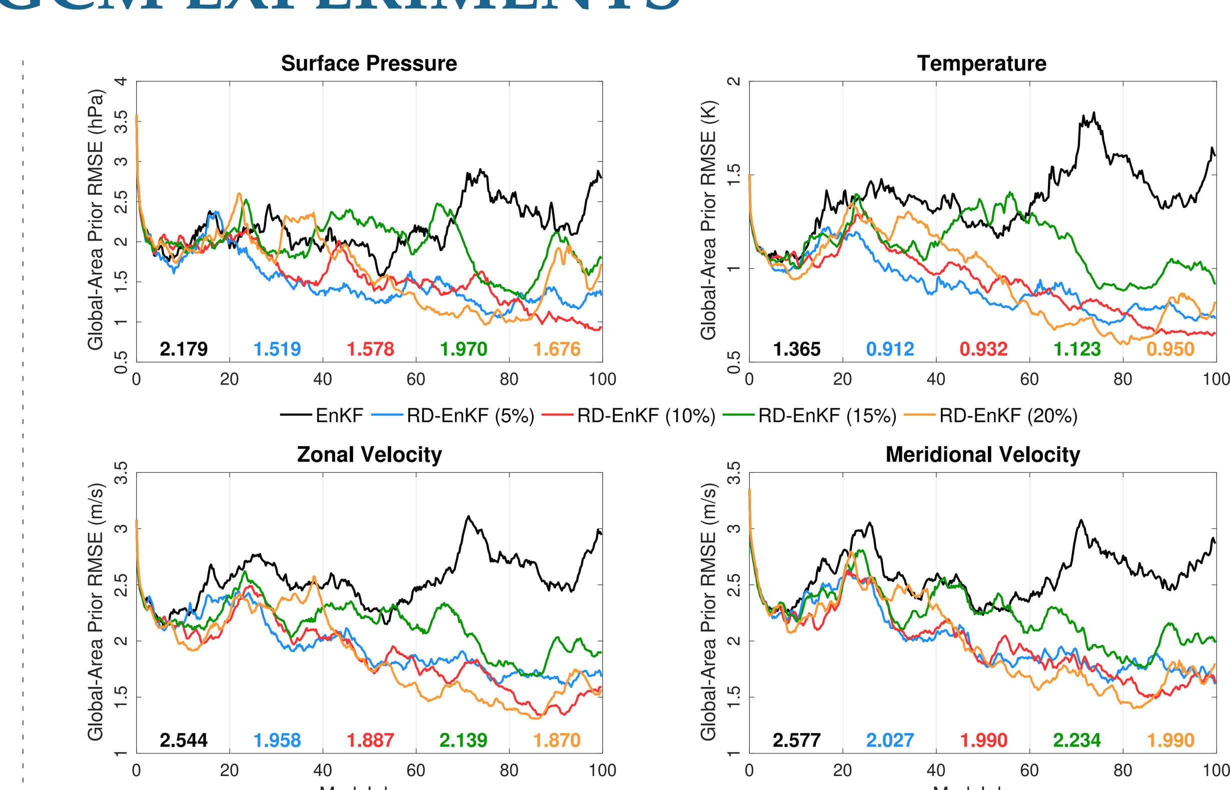


Figure 6: Time evolution of global-area prior RMSE for PS, T, U and V using the **EnKF** and the **RD-EnKF**; $\alpha \in [0.5, 0.2]$. The ensemble size is set to 20 with a localization cutoff of 0.2 radians. PS observations are sampled from 300 different sites (randomly selected) with an observation error variance of 1 hPa.

CONCLUSIONS

- The randomized dormant EnKF is specifically designed for models that are less tolerant to inflation or those with limited uncertainty growth
- Preliminary results show several promising aspects about the **RD-EnKF**:
  (1) Ability to maintain sufficient ensemble spread after the update
  (2) Robust performance even in poorly localized domains
  (3) Need for less inflation given the inherent spread retention by DMs
- More extensive testing with other models (e.g., ocean) will be conducted
- Can select the dormant members differently

REFERENCES

- Anderson, J. L. "Spatially and temporally varying adaptive covariance inflation for ensemble filters." Tellus A: Dynamic meteorology and oceanography 61.1 (2009): 72-83.
- Anderson, J. L., Wyman, B., Zhang, S., and Hoar, T. (2005). Assimilation of surface pressure observations using an ensemble filter in an idealized global atmospheric prediction system. Journal of the atmospheric sciences, 62(8), 2925-2938.
- El Gharamti, M. "Enhanced adaptive inflation algorithm for ensemble filters." Monthly Weather Review 146.2 (2018): 623-640.
- El Gharamti, M. et al. "Comparing adaptive prior and posterior inflation for ensemble filters using an atmospheric general circulation model." Monthly Weather Review 147.7 (2019): 2535-2553.
- The Data Assimilation Research Testbed (Version 9.16.3) [Software]. (2019). Boulder, Colorado: UCAR/NCAR/CISL/DAReS. http://doi.org/10.5065/D6WQ0202
- Whitaker, J. S. and Hamill T. M. "Evaluating methods to account for system errors in ensemble data assimilation." Monthly Weather Review 140.9 (2012): 3078-3089.
- Zhang, F. et al. "Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter." Monthly Weather Review 132.5 (2004): 1238-1253.