An observing system simulation experiment (OSSE) using the Particle Flow Filter (PFF) in a high-dimensional geophysical system in the Data Assimilation Research Testbed (DART)

Chih-Chi Hu[1,2], Peter Jan van Leeuwen[2], Jeffrey Anderson[3]

[1]Princeton University, Princeton, NJ, USA

[2]Colorado State University, Fort Collins, CO, USA

[3]National Center for Atmospheric Research, Boulder, CO, USA

Introduction – Particle Flow Filter (PFF)

- Particle filter (PF) is a fully nonlinear data assimilation (DA) method, while it is known to suffer from the weight degeneracy issue in a high-dimensional system.
- The Particle Flow Filter (PFF), avoids the weight by construction, has shown the potential to solve a high-dimensional nonlinear DA problem [a].

$$\frac{d}{ds}\mathbf{x}_s = \mathbf{f}_s(\mathbf{x}_s),\ s \in [0, \infty)$$

(model state at pseudo time s; the "particle flow" is)

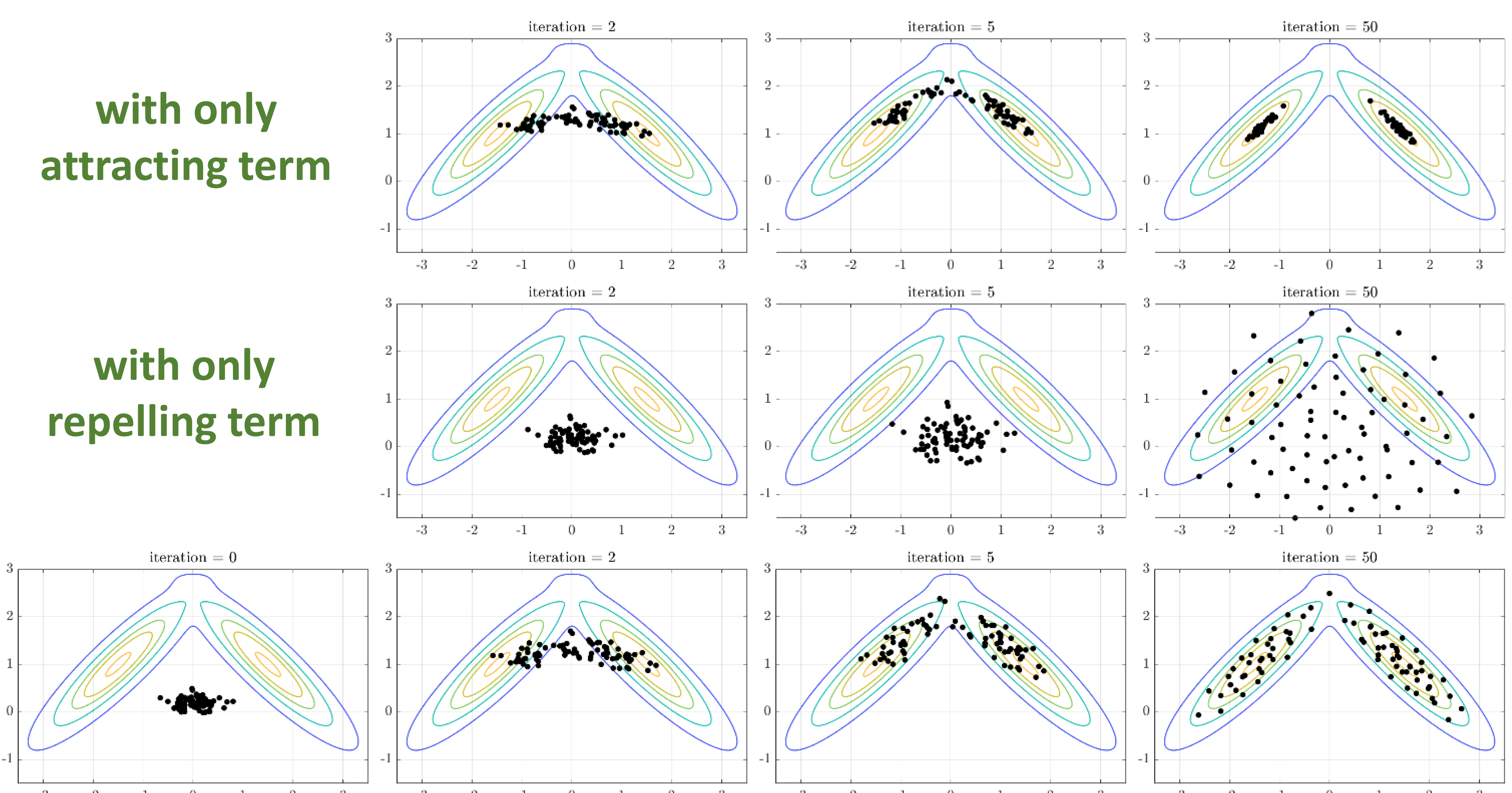
$$q_0(\mathbf{x}) = p(\mathbf{x})$$ (the pdf at initial pseudo time)

from prior to posterior

$$q_\infty(\mathbf{x}) = p(\mathbf{x}|\mathbf{y}_o)$$ (the pdf at final pseudo time (targeted pdf))

$$\mathbf{f}_s(\mathbf{x}) = \mathbf{D}\sum_{i=1}^{N_p}\left\{\mathbf{K}(\mathbf{x}_s^i,\mathbf{x})\nabla_{\mathbf{x}_s^i}\log p(\mathbf{x}_s^i|\mathbf{y}_o) + \nabla_{\mathbf{x}_s^i}\cdot\mathbf{K}(\mathbf{x}_s^i,\mathbf{x})\right\}$$

"attracting term" "repelling term"



Motivations

- Geophysical DA problems are high-dimensional, and can be quite nonlinear.
- Challenges:
  - Extremely high-dimensional problem ⇒ need a ***parallelizable*** algorithm
  - Can't afford 'too many' particles (typically 100 is a lot), and therefore 'sampling error' issue ⇒ require localization (how?)
- **Goal: develop a scalable algorithm for PFF and conduct an OSSE to test this PFF algorithm in an simplified atmospheric model**

Data Assimilation Research Testbed (DART)

- DART is a open-source community software for ensemble DA, which includes interfaces to a variety of geophysical models and observation operators.
- The core DA algorithm is based on the two-step ensemble filtering [b,c] framework:

$$p(\mathbf{x}|y) = \int p(\mathbf{x},\mathbf{z}|y)d\mathbf{z} = \int \frac{p(y|\mathbf{x},\mathbf{z})p(\mathbf{x},\mathbf{z})}{p(y)}d\mathbf{z} = \int \frac{p(y|\mathbf{z})p(\mathbf{x},\mathbf{z})}{p(y)}d\mathbf{z} \cdots = \int p(\mathbf{z}|y)p(\mathbf{x}|\mathbf{z})d\mathbf{z}$$

a new variable "z" is introduced that satisfies

$$p(y|\mathbf{x},\mathbf{z}) = p(y|\mathbf{z})$$

Examples of "z":

1) $z = H(\mathbf{x})$
2) $\mathbf{z}$ = the "nontrivial" input variables to $H(\mathbf{z}_y)$ ("inner domain")

- A new algorithm for PFF (called PFF-DART) is proposed as follows (note $\mathbf{z} = \mathbf{z}_y$):
  - **1st step**: use the PFF to draw $\mathbf{z}^i$ from $p(\mathbf{z}|y)$
  - **2nd step**: draw from $p(\mathbf{x}|\mathbf{z}^i)$ *(can run in parallel)*

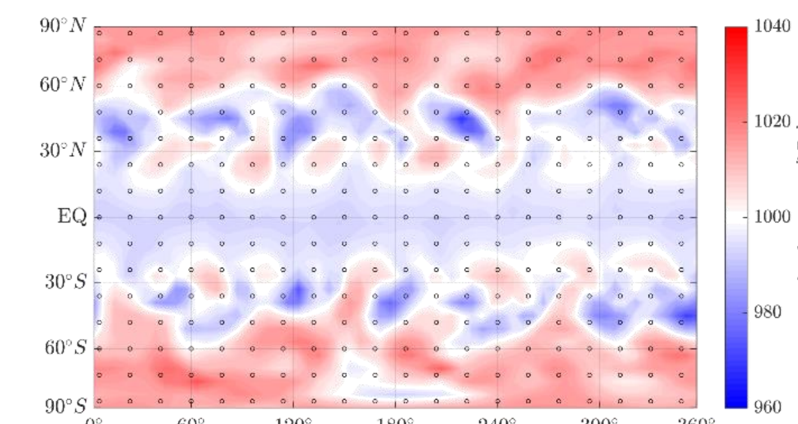
So $p(\mathbf{z}|y) = \frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{z}-\mathbf{z}^i)$

and $p(\mathbf{x}|y) = \frac{1}{N}\sum_{i=1}^{N}p(\mathbf{x}|\mathbf{z}^i)$

We assume Gaussian $p(\mathbf{x}|\mathbf{z}^i)$ so the linear regression update is used to draw **x**, and the "increment localization" (the default in DART) can be naturally applied here.

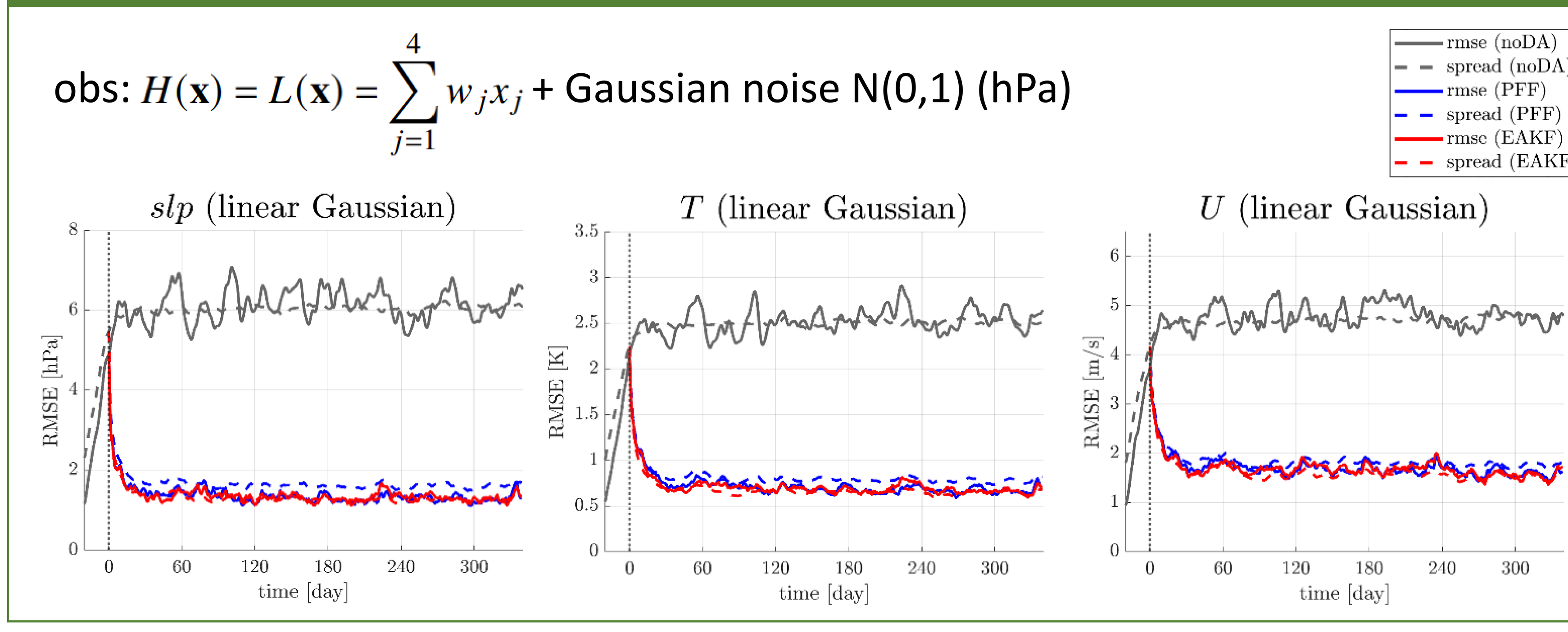
OSSE setup

- The "Bgrid" model (var = U, V, T, ps)
- 1 year cycling DA, with 300 obs every day (black circles, not located on grid points, **only ps is observed**)
- Compare: noDA, EAKF, PFF-DART
- 25 ensemble members, no inflation (default), the same localization



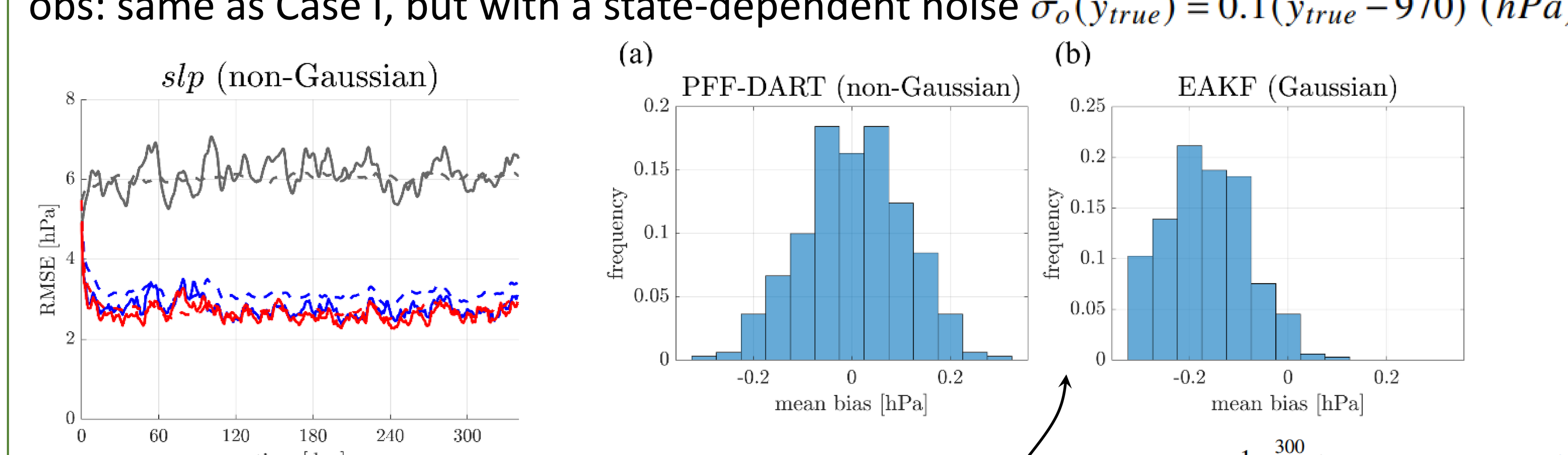
Case I – linear and Gaussian obs

obs: $H(\mathbf{x}) = L(\mathbf{x}) = \sum_{j=1}^{4} w_j x_j$ + Gaussian noise N(0,1) (hPa)



Case II – linear and non-Gaussian obs

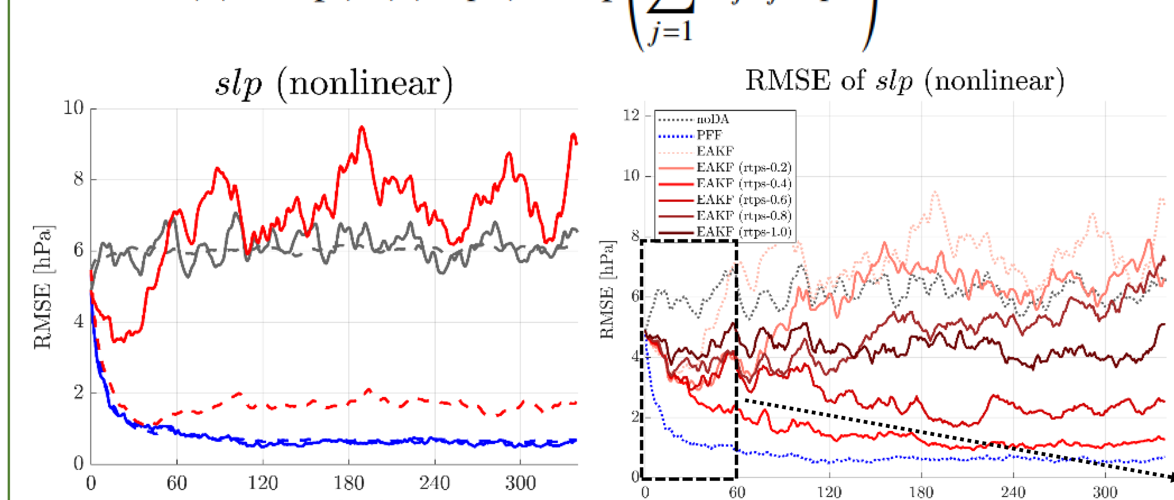
obs: same as Case I, but with a state-dependent noise $\sigma_o(y_{true}) = 0.1(y_{true} - 970)\ (hPa)$



the distribution of the posterior mean bias $b(t) = \frac{1}{300}\sum_{k=1}^{300}\left(\bar{y}^{(k)}_{posterior}(t) - y^{(k)}_{true}(t)\right)$

Case III – nonlinear and non-Gaussian obs

obs: $H(\mathbf{x}) = \exp(L(\mathbf{x}) - p_0) = \exp\left(\sum_{j=1}^{4} w_j x_j - p_0\right)$ with noise $\sigma_o(y_{true}) = \beta y_{true}\ (\beta = 0.25)$



EAKF with RTPS

$$\sigma_a \leftarrow (1-\alpha)\sigma_{posterior} + \alpha\sigma_{prior}$$

EAKF with the best tuned inflation is worse than the PFF without any inflation.

Inflation does not address the issues from the nonlinearities.

Conclusions

- We develop an algorithm for the PFF in DART that can
  1) run in parallel for high-dimensional, spatially extended geophysical models
  2) efficiently apply the localization (via the existing DART modules) to reduce the sampling errors due to the small number of particles
- The PFF performs comparably to EAKF for linear and Gaussian observations, while the PFF outperforms EAKF for nonlinear and non-Gaussian observations.

References

a. Hu, C.-C., and P. J. van Leeuwen, 2021: A particle flow filter for high-dimensional system applications. Quarterly Journal of the Royal Meteorological Society, 147 (737), 2352–2374.

b. Anderson, J. L., 2003: A local least squares framework for ensemble filtering. Monthly Weather Review, 131 (4), 634–642.

c. Grooms, I., 2022: A comparison of nonlinear extensions to the ensemble kalman filter: Gaussian anamorphosis and two-step ensemble filters. Computational Geosciences, 26 (3), 633–650.