Introducing Non-Gaussian Observation Errors into the Local Ensemble Transform Kalman Filters (LETKF)

Chih-Chi Hu

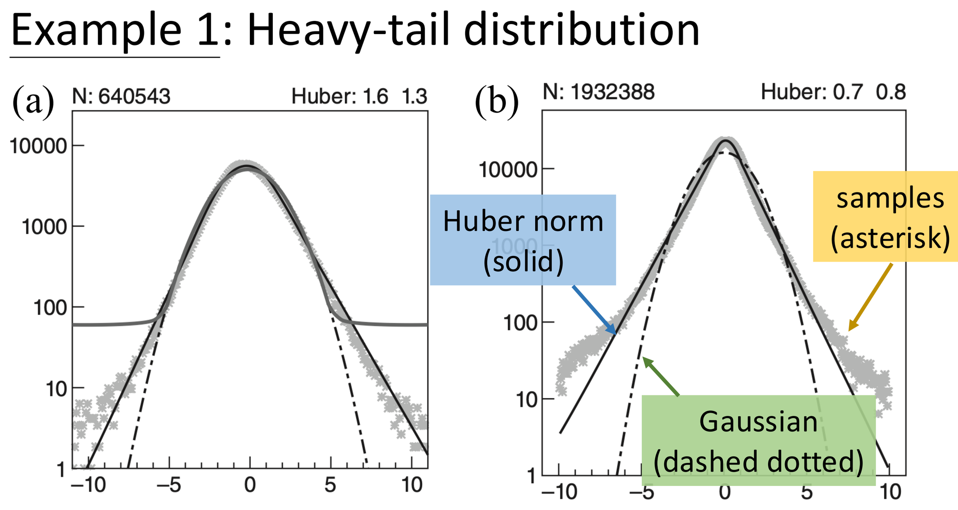
Princeton University, Princeton, NJ, USA

Motivations

- For most DA applications, the observation error is assumed to be Gaussian. However, studies have suggested the observation errors may exhibit non-Gaussian characteristics [a,b].
- The goal of this study:
  1) To evaluate the impact of non-Gaussian observation errors on the LETKF [c], which assumes Gaussian observation errors
  2) Explore methods to account for non-Gaussian errors in the LETKF

Non-Gaussian Observation Errors

- Example 1: Heavy-tail distribution



O-B innovation normalized by the prescribed observation error

(Figure 1 from Tavolato and Isaksen 2015 [a])

(a) Global radiosonde temperature

(b) SH Extratropic surface pressure

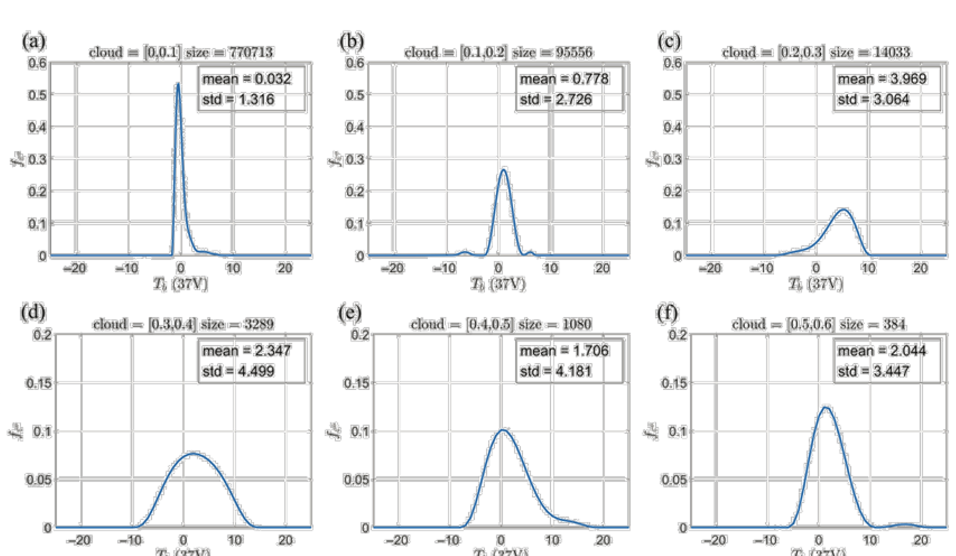
- Example 2: Skewed distribution

Example 2.1: cloudy satellite radiances

(Figure 6 from Hu et al 2024 [b])

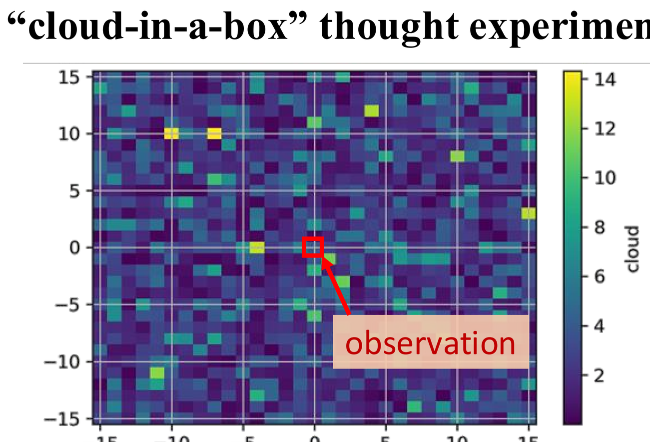
Estimated observation error PDFs for SSMIS satellite radiances at 37 GHz (V) for different cloud amount.

The non-parametric estimation is performed using the Deconvolution-based Observation Error Estimation (DOEE)[b] method.



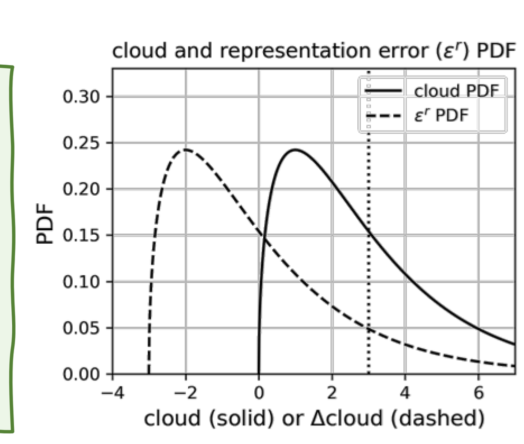
Example 2.2: representation error due to scale mismatch

"cloud-in-a-box" thought experiment



X – large-scale (model grid)
Y – small-scale (observation)

Assume identity mapping, and
$Y \sim Gamma(\alpha, \theta)$
$X^{truth} = mean(Y)$
$\varepsilon^r = Y - X^{truth}$



A Modification of the Analysis Mean in LETKF

- The ensemble mean update equations for the Local Ensemble Transform Kalman Filter (LETKF) [c] considering one scalar observation are as follows:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b \bar{\mathbf{w}}^a \qquad \text{Eqn (1)}$$

$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T R^{-1} \boxed{(y^o - \bar{y}^b)} \qquad \text{Eqn (2)}$$

$$\tilde{\mathbf{P}}^a = \left[(k-1)\mathbf{I} + (\mathbf{Y}^b)^T R^{-1} \mathbf{Y}^b\right]^{-1} \qquad \text{Eqn (3)}$$

- Idea: In 1-D space, *the analytical posterior PDF* can be identified. We can adjust the term $d = (y^o - \bar{y}^b)$ so that the analysis mean of LETKF fits to the mean or the mode of the *analytical posterior PDF*.

$$\Delta y^{target} = y^{target} - \bar{y}^b \quad \text{(analysis increment in the observation space)}$$

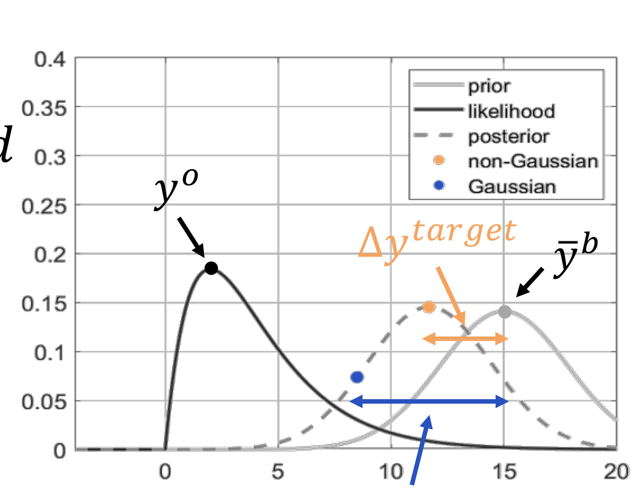
can be the mode or mean of the analytical posterior PDF

$$\Delta y^{Gaussian} = \frac{\sigma_b^{\,2}}{\sigma_o^{\,2} + \sigma_b^{\,2}} (y^o - \bar{y}^b) \equiv \frac{\sigma_b^{\,2}}{\sigma_o^{\,2} + \sigma_b^{\,2}} d$$

$$\Delta y^{target} = \frac{\sigma_b^{\,2}}{\sigma_o^{\,2} + \sigma_b^{\,2}} d^{target}$$

$$\Rightarrow d^{target} = \frac{\sigma_o^{\,2} + \sigma_b^{\,2}}{\sigma_b^{\,2}} \Delta y^{target}$$

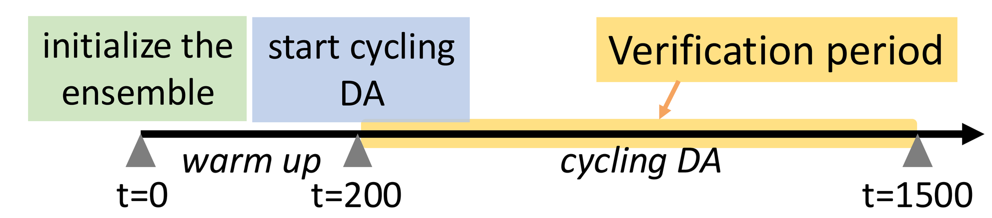
$$\Rightarrow \bar{\mathbf{w}}^{a,modified} = \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T R^{-1} \boxed{d^{target}} \qquad \text{Eqn (4)}$$



- Equation (4) is used to replace Equation (2) in the ensemble mean update equations. The ensemble perturbation update equations remain the same.

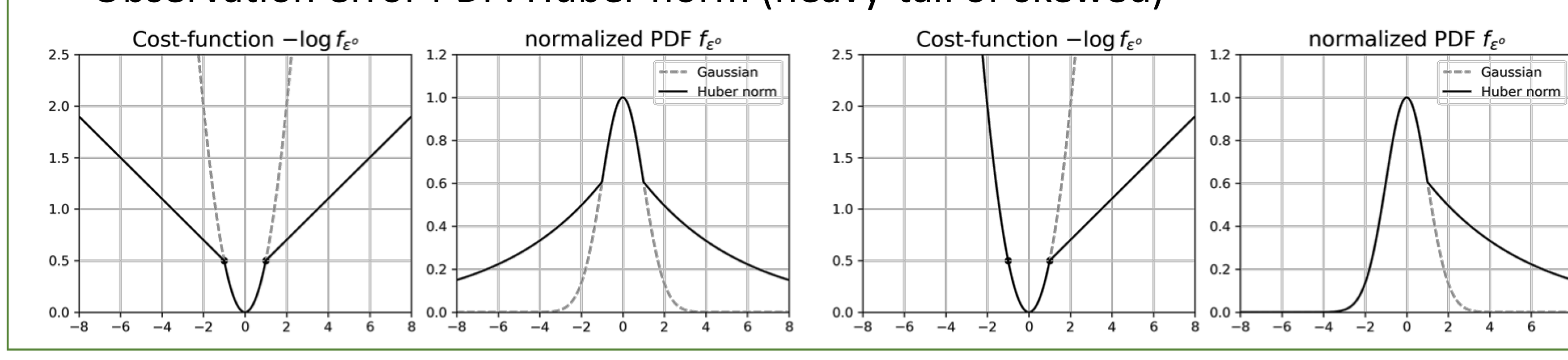
Experimental Design

- Lorenz 96; Perfect Model Observing System Simulation Experiment (OSSE)



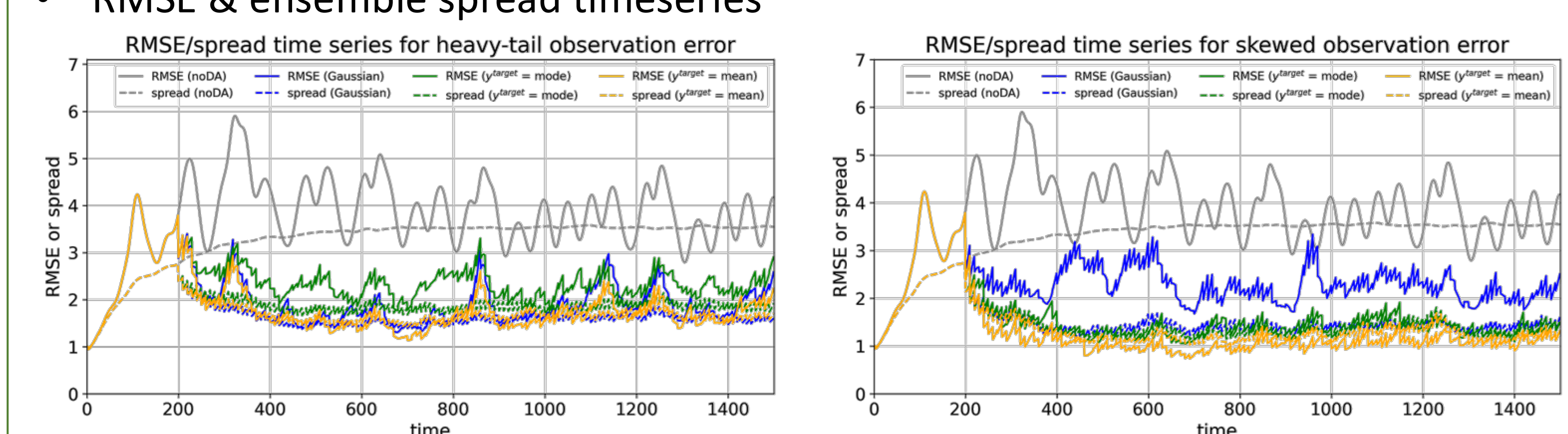
Parameters:
nx = 200
ne = 25
inflation: None
localization: R (Gaspari-Cohn)

- Observation error PDF: Huber norm (heavy-tail or skewed)

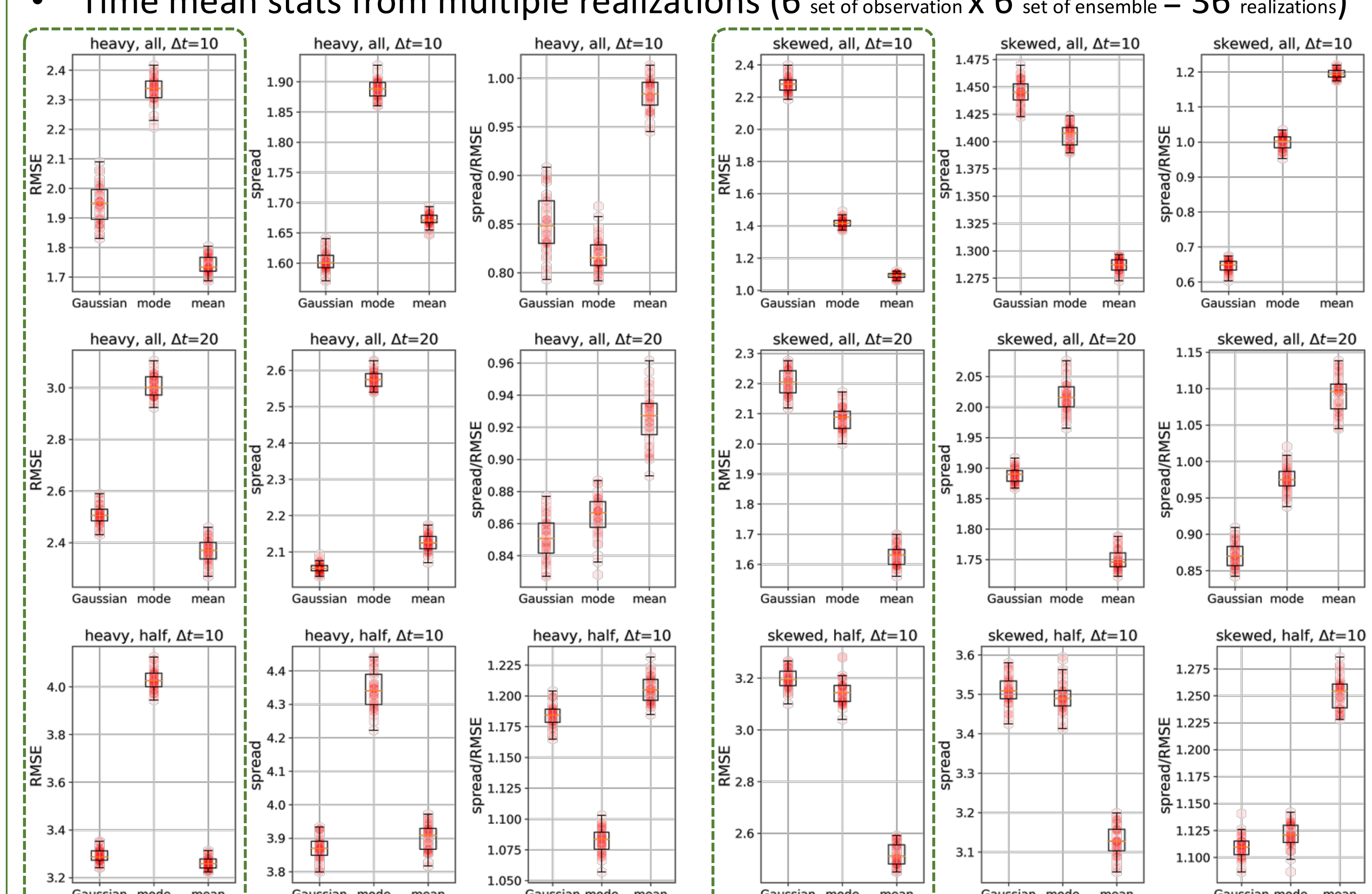


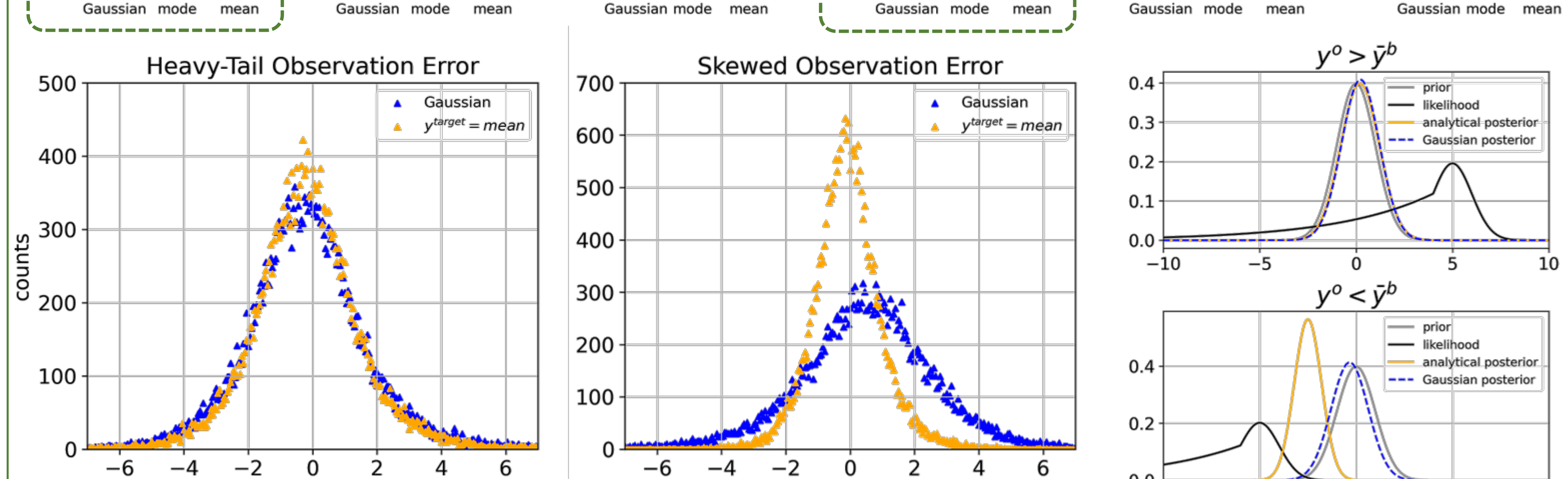
Results

- RMSE & ensemble spread timeseries



- Time mean stats from multiple realizations (6 set of observation x 6 set of ensemble = 36 realizations)





Conclusions

- Non-Gaussian representation error can arise from scale mismatch when the small-scale features follow non-Gaussian distribution.
- Assimilating observations with non-Gaussian observation error can degrade the performance of LETKF or introduce biases.
- Modifying the LETKF ensemble mean update equation to fit to the mean of the analytical posterior PDF improves the performance and reduces biases.

References

a. Tavolato, C., & Isaksen, L. (2015). On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, *141*(690), 1514-1527.

b. Hu, C. C., Van Leeuwen, P. J., & Geer, A. J. (2024). A non-parametric way to estimate observation errors based on ensemble innovations. *Quarterly Journal of the Royal Meteorological Society*.

c. Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1-2), 112-126.