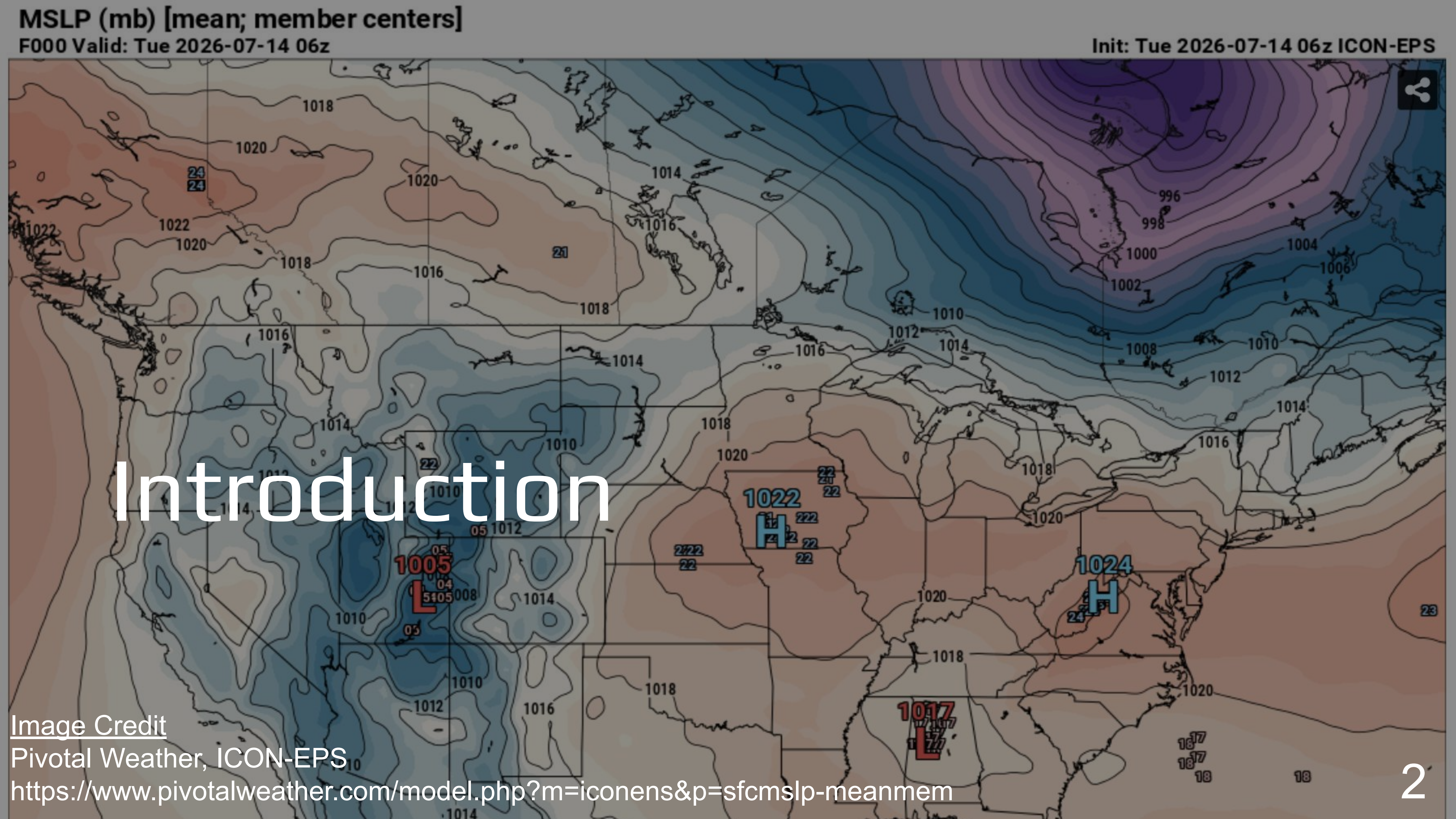
A Novel Data Assimilation Technique to Better Resolve Ensemble Overconfidence

Guan Xin (Max)
Moha Gharamti
Jeffrey Anderson

SIParCS
Summer Internships in Parallel Computational Science
BOULDER · CO
NSF
NCAR
OPERATED BY UCAR



Introduction

Image Credit
Pivotal Weather, ICON-EPS
https://www.pivotalweather.com/model.php?m=iconens&p=sfcmslp-meanmem

Numerical Weather Prediction (NWP)

- Cornerstone of modern weather forecasting

- Computationally efficient, but approximate

Image Credit
WRF Logo: https://www2.mmm.ucar.edu/wrf/users/
ECMWF Logo:
https://commons.wikimedia.org/wiki/File:ECMWF_logo.svg

Ensemble NWP

- Ensemble to represent probability distribution

- Constrained via Ensemble Data Assimilation (EnsDA)

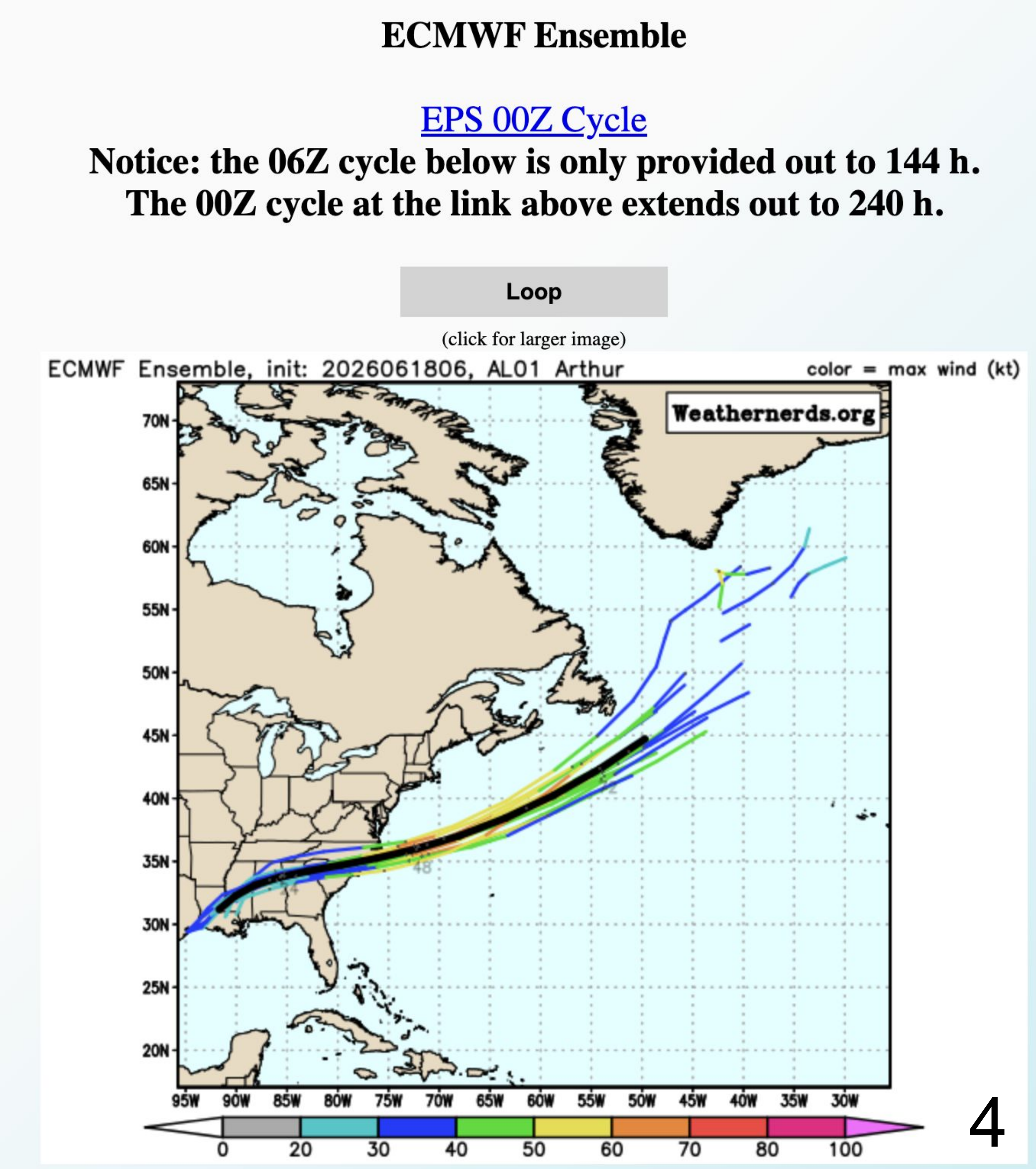
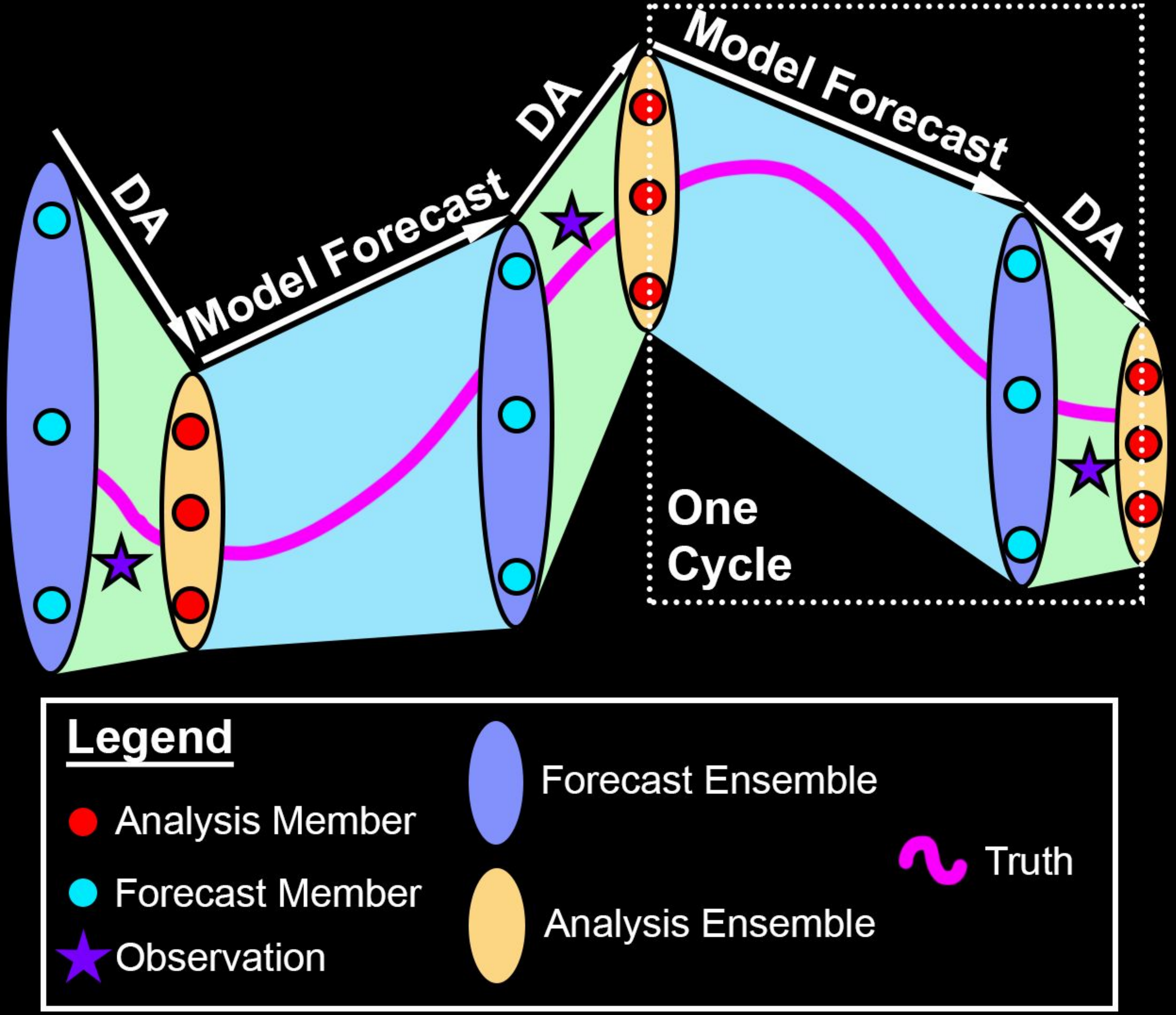


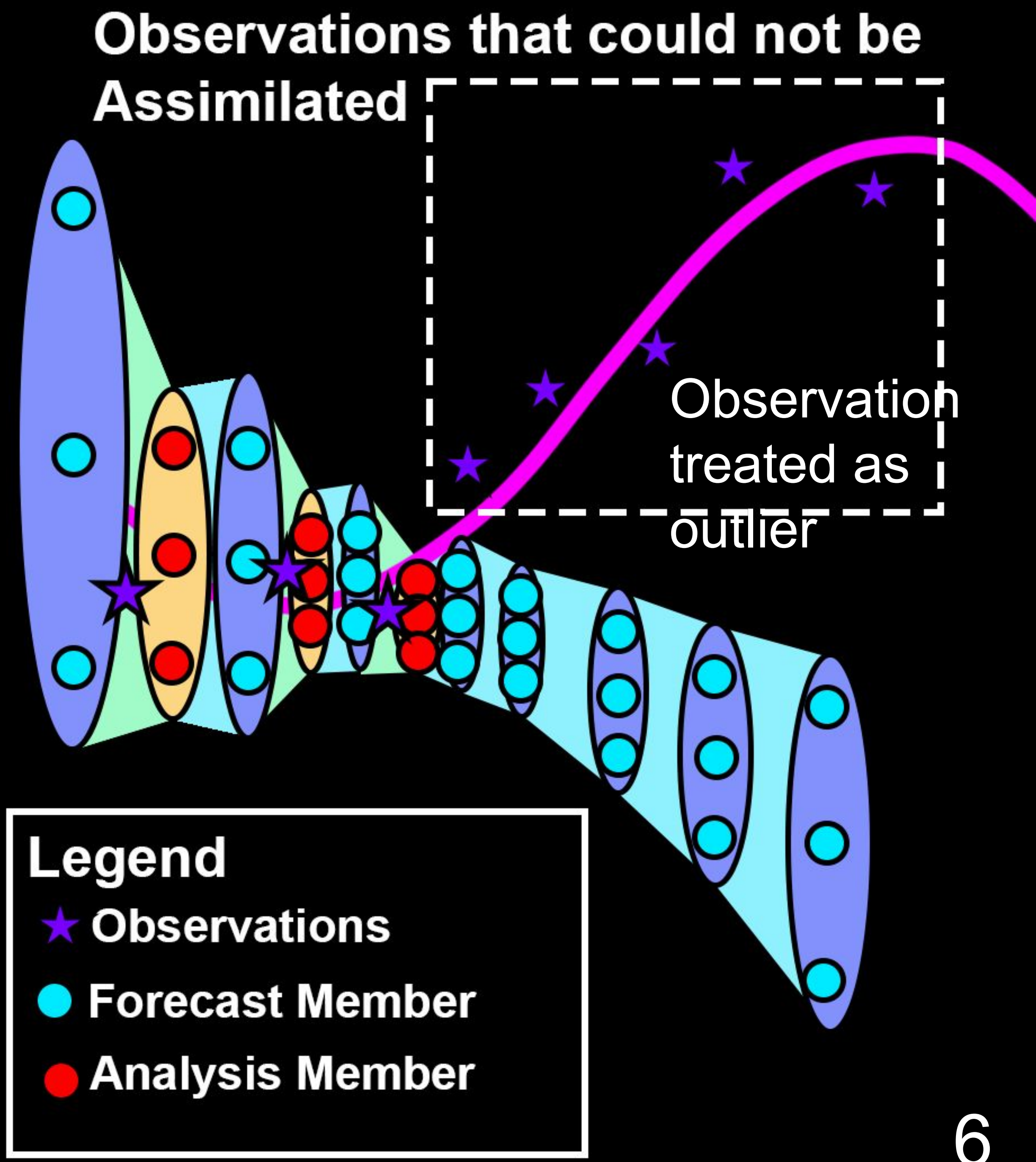
Image Credit
https://www.weathernerds.org/tc_guidance/AL01.html

The Ens DA-Model Cycle

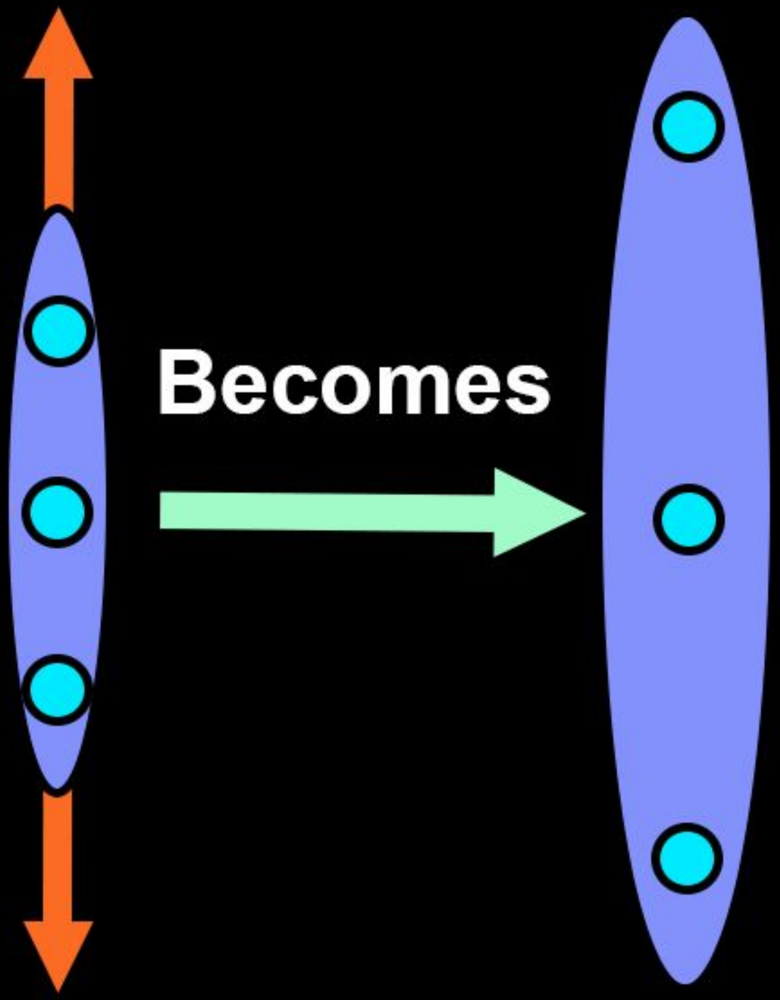


Ensemble Collapse

- DA strictly contracts spread, but not always in the right place
- Ensemble spread collapse can kill DA
- Occurs due to sampling error and model error



Inflation



- Changes only the spread
- Better able to assimilate observations
- Mitigates model "over-confidence"

Without Inflation

Members (A, B, C)

Forecast (prior)

$$\begin{aligned} A&:(1 \quad 1 \quad 1) \\ B&:(2 \quad 2 \quad 2) \\ C&:(3 \quad 3 \quad 3) \end{aligned}$$

Variables

Observation

$$(? \quad ? \quad 2)$$

Members (A, B, C)

Analysis (posterior)

$$\begin{aligned} A&:(1 \quad 1 \quad 1.9) \\ B&:(2 \quad 2 \quad 2) \\ C&:(3 \quad 3 \quad 2.1) \end{aligned}$$

Variables

With Inflation

Members (A, B, C)

Forecast (prior)

$$A: (1 \quad 1 \quad 1)$$
$$B: (2 \quad 2 \quad 2)$$
$$C: (3 \quad 3 \quad 3)$$

Variables

Observation

$$(? \quad ? \quad 2)$$

Members (A, B, C)

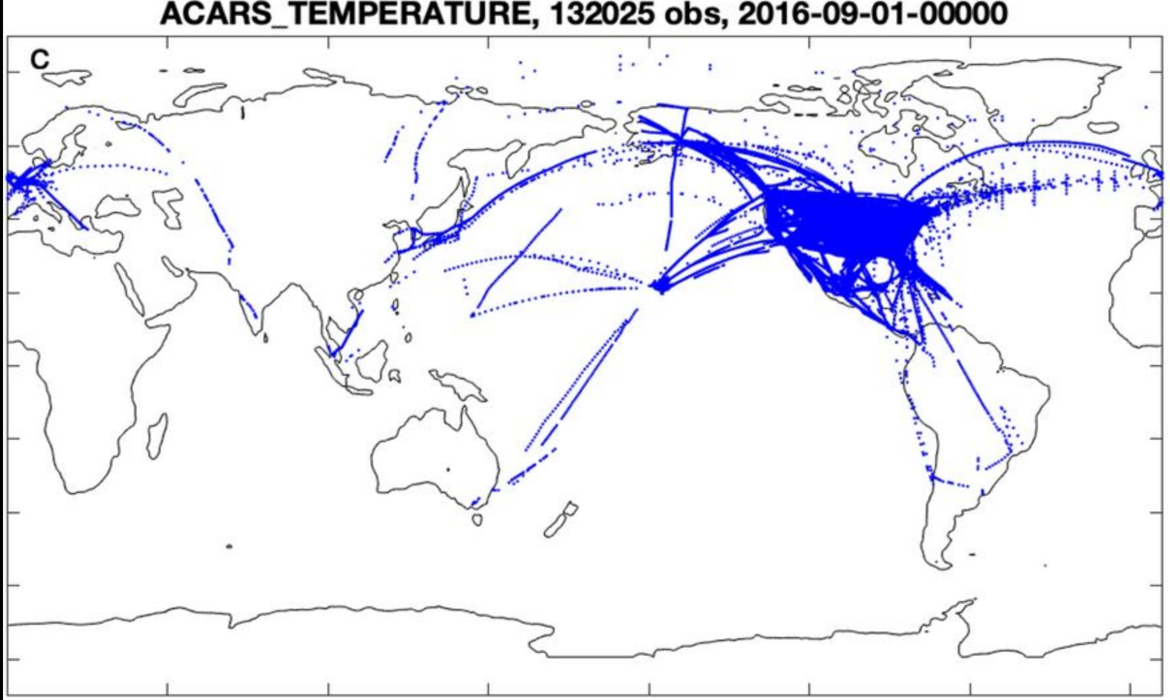
Analysis (posterior)

$$A: (? \quad ? \quad 1.8)$$
$$B: (? \quad ? \quad 2)$$
$$C: (? \quad ? \quad 2.2)$$

Variables

Spatiotemporally Varying Observations

- Observations are tied to specific locations
- DA increments are also localized



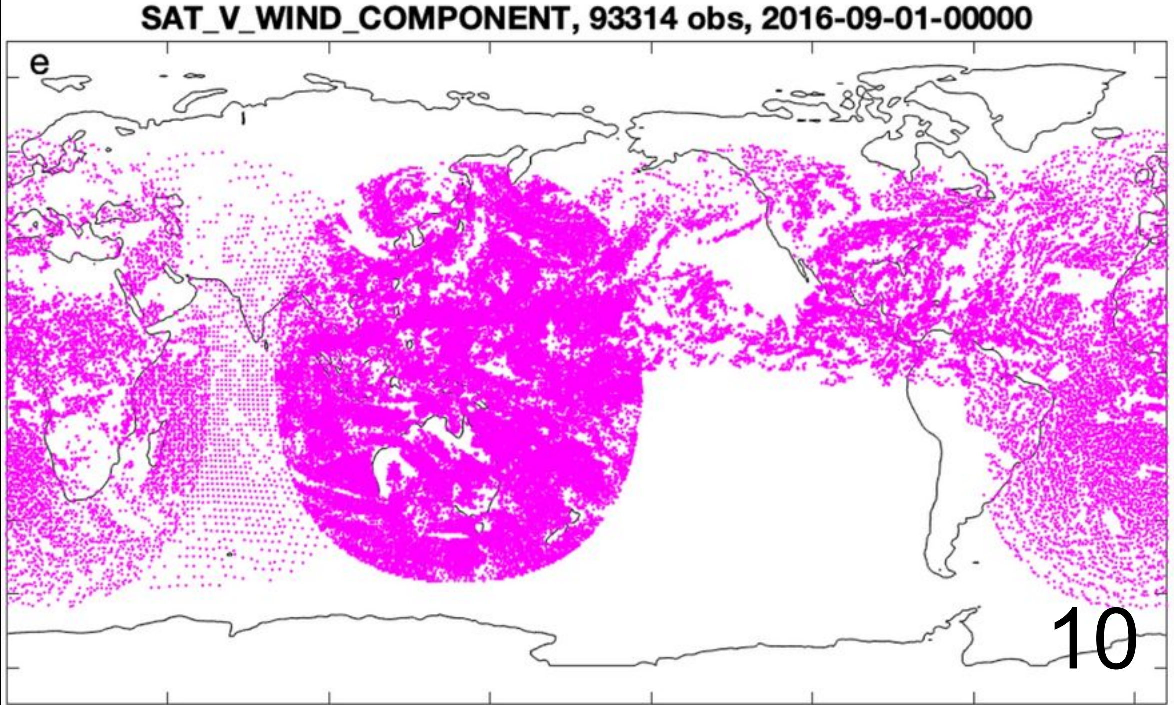


Image Credit
Raeder, K., et al., 2021, *Sci Rep*, **11**,

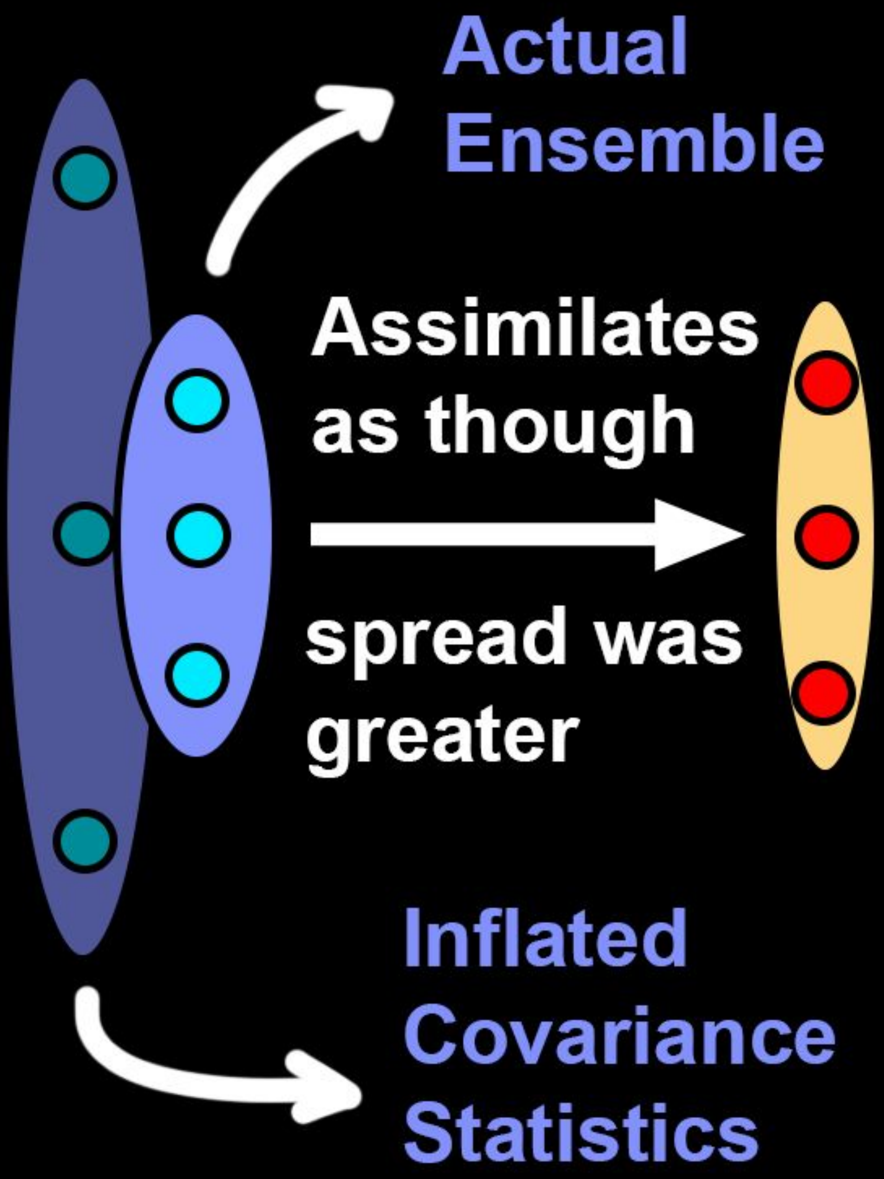
Spatiotemporally Varying Inflation Algorithms

- Relaxation to Prior Spread [1] / Perturbation [2] (RTPP / RTPS)
  - Good for sampling error, less effective for model error
  - Issues with bounded / non-linear variables

- Adaptive (Multiplicative) Inflation [3,4,5]
  - Issues with highly varying observation systems
  - Prior Inflation unaware of observations at current timepoint

1. Zhang, F., et al. 2004, *Mon. Weather Rev.*, **132,** 5, 1238-1253.
2. Whitaker, J. S., Hamill, T. M., 2012, *Mon. Weather Rev.*, **140**, 9, 3078-3089.
3. Anderson, J. L., 2007, *Tellus A*, **59**, 2, 210-224.
4. Anderson, J. L., 2009, *Tellus A*, **61**, 1, 72-83.
5. El Gharamti, JM., 2018, *Mon. Weather Rev.*, **146**, 2, 623-640.

Kalman Inflation



$$\bar{x}_j^a = \bar{x}_j^f + \left(\frac{\lambda_j \sigma_{jo}}{\sigma_{oo} + R}\overline{y^o - H(x)}\right) * \gamma_j$$

$$x_{ij}^{a\,\prime} = x_{ij}^{f\,\prime} + \left[(\lambda_j - 1)x_{ij}^{f\,\prime} - \lambda_j \frac{\sigma_{jo}}{\sigma_{oo}}\left(1 - \sqrt{\frac{R}{\sigma_{oo} + R}}\right)y_i'\right] * \gamma_j$$

- Simulates multiplicative inflation (by λ_j)
- Applied under DA localization (γ_j)

Variant 1

Mimics prior inflation

$$\bar{x}_j^a = \bar{x}_j^f + \left(\frac{\lambda_j \sigma_{jo}}{\sigma_{oo} + R}\overline{y^o - H(x)}\right) * \gamma_j$$

$$x_{ij}^{a\prime} = x_{ij}^{f\prime} + \left[(\lambda_j - 1)x_{ij}^{f\prime} - \lambda_j \frac{\sigma_{jo}}{\sigma_{oo}}\left(1 - \sqrt{\frac{R}{\sigma_{oo} + R}}\right) y_i^\prime\right] * \gamma_j$$

Variant 2

Mimics posterior inflation

$$\bar{x}_j^a = \bar{x}_j^f + \left(\frac{\sigma_{jo}}{\sigma_{oo} + R}\overline{y^o - H(x)}\right) * \gamma_j$$

$$x_{ij}^{a\prime} = x_{ij}^{f\prime} + \left[(\lambda_j - 1)x_{ij}^{f\prime} - \lambda_j \frac{\sigma_{jo}}{\sigma_{oo}}\left(1 - \sqrt{\frac{R}{\sigma_{oo} + R}}\right) y_i^\prime\right] * \gamma_j$$

Methodology

```
#ensemble_projectname=Default_DART_Ensemble
ensemble_projectname=Larger_DART_Ensemble
output_main_projectname=Larger_Test_Posterior_CovInf
#output_main_projectname=Larger_No_DA_Reference
#output_main_projectname=Larger_Moha_Prior_Reference
#output_main_projectname=Larger_Single_SS_Prior_Reference
ensemble_directory=/Users/xinguan/DATA
obs_repository=/Users/xinguan/DATA/obs_repository/L96
n_ensemble_members=20
is_regular_obs=1                        # 0 = No, 1 = Yes
is_regular_obs_2=1                      # 0 = No, 1 = Yes
is_irregular_obs=0                      # 0 = No, 1 = Yes
is_random_obs=0                         # 0 = No, 1 = Yes | Note: not fully random, consistent in time
                                        #                 |     : for fully random obs, generate manually, then set is_full_random_input=1
n_tasks=2
is_clean_output_directory=1             # 0 = No, 1 = Yes

is_perfect_model_obs=0                  # 0 = No, 1 = Yes
is_generate_obs_input=0                 # 0 = No, 1 = Yes
is_run_filter=1                         # 0 = No, 1 = Yes
is_save_obs_file=1                      # 0 = No, 1 = Yes
obs_save_filepath_header=obs_seq_dense
is_grab_obs_file=1                      # 0 = No, 1 = Yes
obs_grab_filepath_header=obs_seq_dense
is_full_random_input=0                  # 0 = No, 1 = Yes | REQUIRES is_generate_obs_input=0

prior_inflation_type=0                  # 0 = No Inflation, 3 = Single SS, 5 = Moha Inflation, 6 = My Inflation
posterior_inflation_type=6              # 0 = No Inflation, 3 = Single SS, 5 = Moha Inflation, 6 = My Inflation
initial_prior_inflation=1.00
initial_posterior_inflation=1.014
inflation_increment_prior=0.002
inflation_increment_posterior=0.002
experiment_count=10
is_tuning=0                             # 0 = False, 1 = True

lorenz_model_forcing=8.00
```

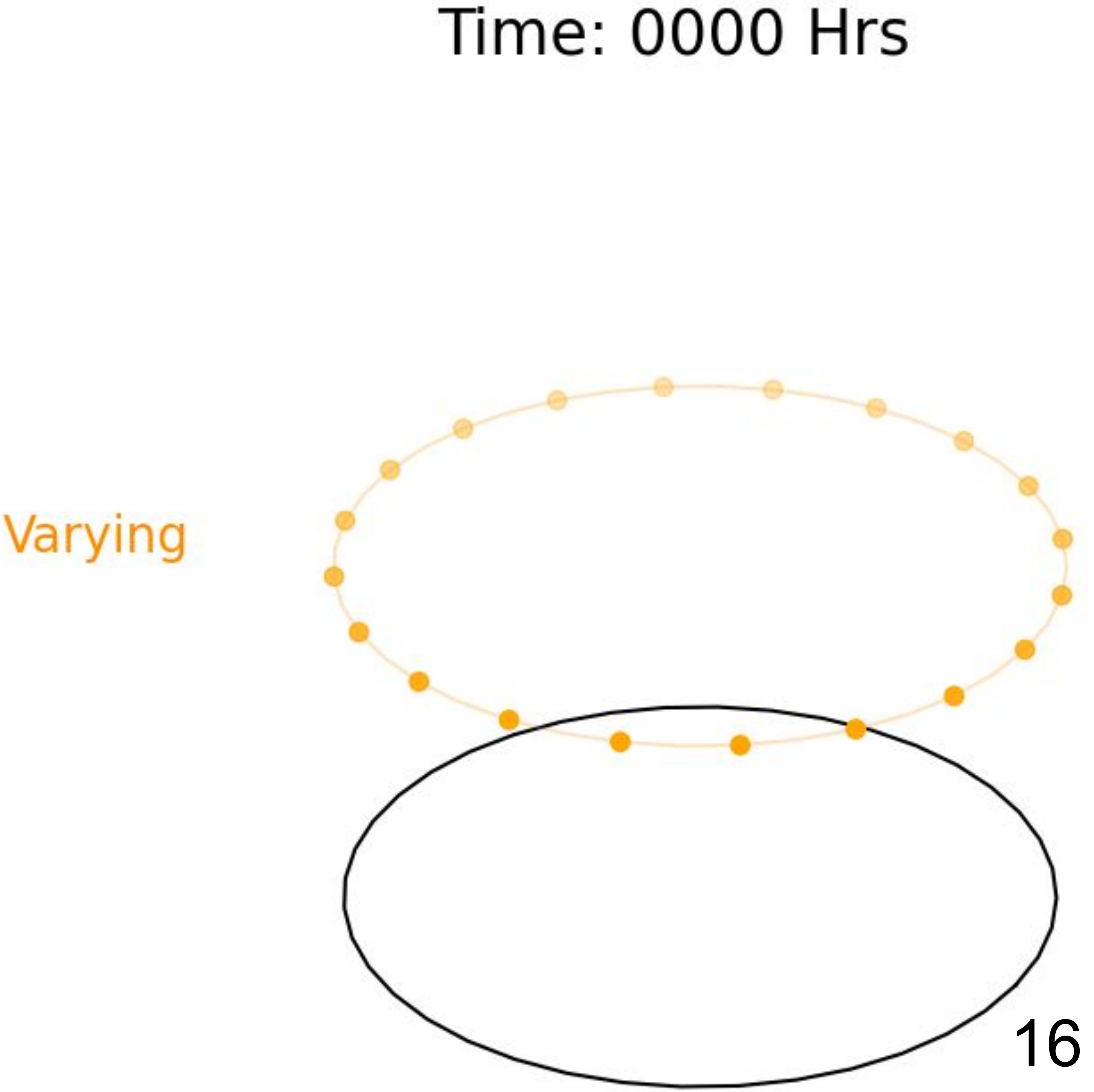
Experiment Overview

- Lorenz 96 [1]
- Data Assimilation Research Testbed (DART) [2]
- 2 Kalman Inflation variants

1. Lorenz, E. N., 1995, *ECMWF, Seminar on Predictability*.
2. El Gharamti, M. et al., 2025, *Bull. Am. Meteorol. Soc.*, **106**, 11

Observing Systems

- Mimic spatiotemporal variability
- Multiple Observing Networks
  - Everywhere
  - Varying



Experiment Setup

- Each set has 10 experiments
  - 10 different reference truths as basis for observations
- Each experiment lasts 30000 days
  - Either 30000 or 60000 DA cycles
  - 720000 model timesteps
- Tuning primarily based on forecast RMSE

Inflation Types Compared

Implementation Details

Kalman Inflation	**Version 1** **Version 2**	**Limited spatiotemporal variability**	**Flavor 6**
Adaptive Inflation	**Prior**	**Temporally Varying**	**Flavor 3**
		Spatiotemporally Varying [1]	**DART DEFAULT, Flavor 5**
No Inflation	**NA**	**NA**	**Flavor 0**

1. El Gharamti, M. (2018). *Mon. Weather Rev.*, **146**, 2, 623-640

Inflation Tuning

- Manual Tuning over one reference truth trajectory
- Inflation value not spatially varying
- Most likely not fully optimal

Observing Network	Version 1 (10 mems)	Version 2 (10 mems)	Version 1 (20 mems)	Version 2 (20 mems)
Everywhere	1.050	1.052	1.020	1.022
Sparse	1.004	1.008	1.002	1.002
Varying	1.038	1.038	1.016	1.014
Varying + Small Model Error	1.040	1.048	1.016	1.018
Varying + Large Model Error	1.080	1.092	1.058	1.062

Results

```
PE 0: filter: Main assimilation loop, starting iteration59998
PE 0: move_ahead Next assimilation window starts    at:  day=   29998 sec= 84601
PE 0: move_ahead Next assimilation window ends      at:  day=   29999 sec=  1800
PE 0: filter: Ready to run model to advance data ahead in time
PE 0: filter: Ready to assimilate up to      20 observations
PE 0: filter_assim: Processed      20 total observations
PE 0:
PE 0: filter: Main assimilation loop, starting iteration59999
PE 0: move_ahead Next assimilation window starts    at:  day=   29999 sec= 41401
PE 0: move_ahead Next assimilation window ends      at:  day=   29999 sec= 45000
PE 0: filter: Ready to run model to advance data ahead in time
PE 0: filter: Ready to assimilate up to      20 observations
PE 0: filter_assim: Processed      20 total observations
PE 0:
PE 0: filter: Main assimilation loop, starting iteration60000
PE 0: filter: No more obs to assimilate, exiting main loop
PE 0: filter: End of main filter assimilation loop, starting cleanup
PE 0: write_obs_seq  opening formatted observation sequence file "obs_seq.final"
PE 0: write_obs_seq  closed observation sequence file "obs_seq.final"

--------------------------------------
Finished ... at YYYY MM DD HH MM SS =
```

Table of Forecast RMSEs

Observing Network	10 members			20 members		
	Best Kalman Inflation	Best Adaptive Inflation	No Inflation	Best Kalman Inflation	Best Adaptive Inflation	No Inflation
Everywhere						
Varying						
Varying + Large Model Error						

Observing Network	Best Kalman Inflation (10 members)	Best Adaptive Inflation (10 members)	No Inflation (10 members)	Best Kalman Inflation (20 members)	Best Adaptive Inflation (20 members)	No Inflation (20 members)
Everywhere						
Sparse						
Varying						
Varying + Small Model Error						
Varying + Large Model Error						

Table of forecast RMSEs

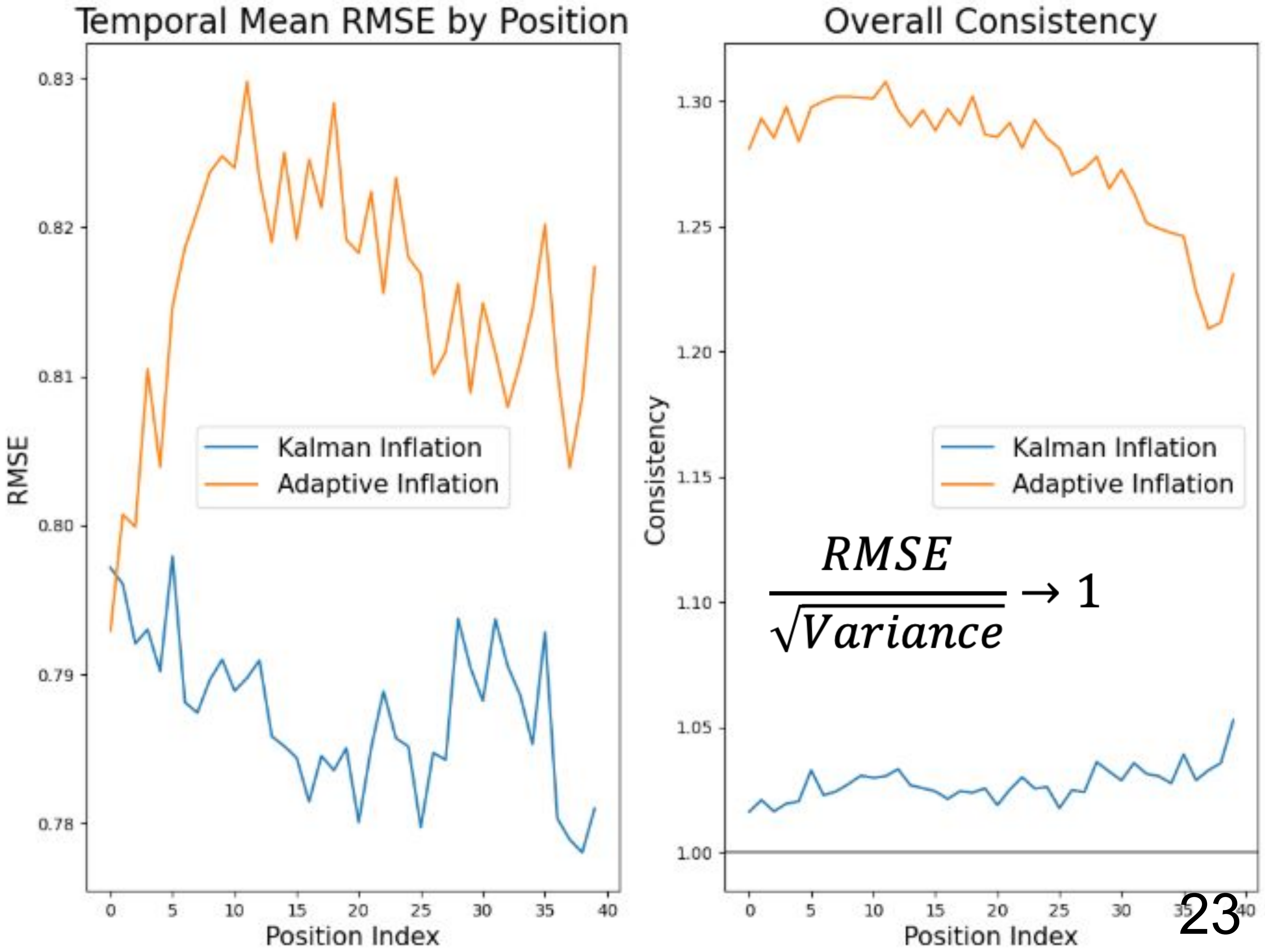
Kalman Inflation, has an opposite trend

Adaptive Prior Inflation, is more strongly affected by the *order* of assimilation

Obs: Everywhere

Order: Left to Right

Ensemble: Analysis

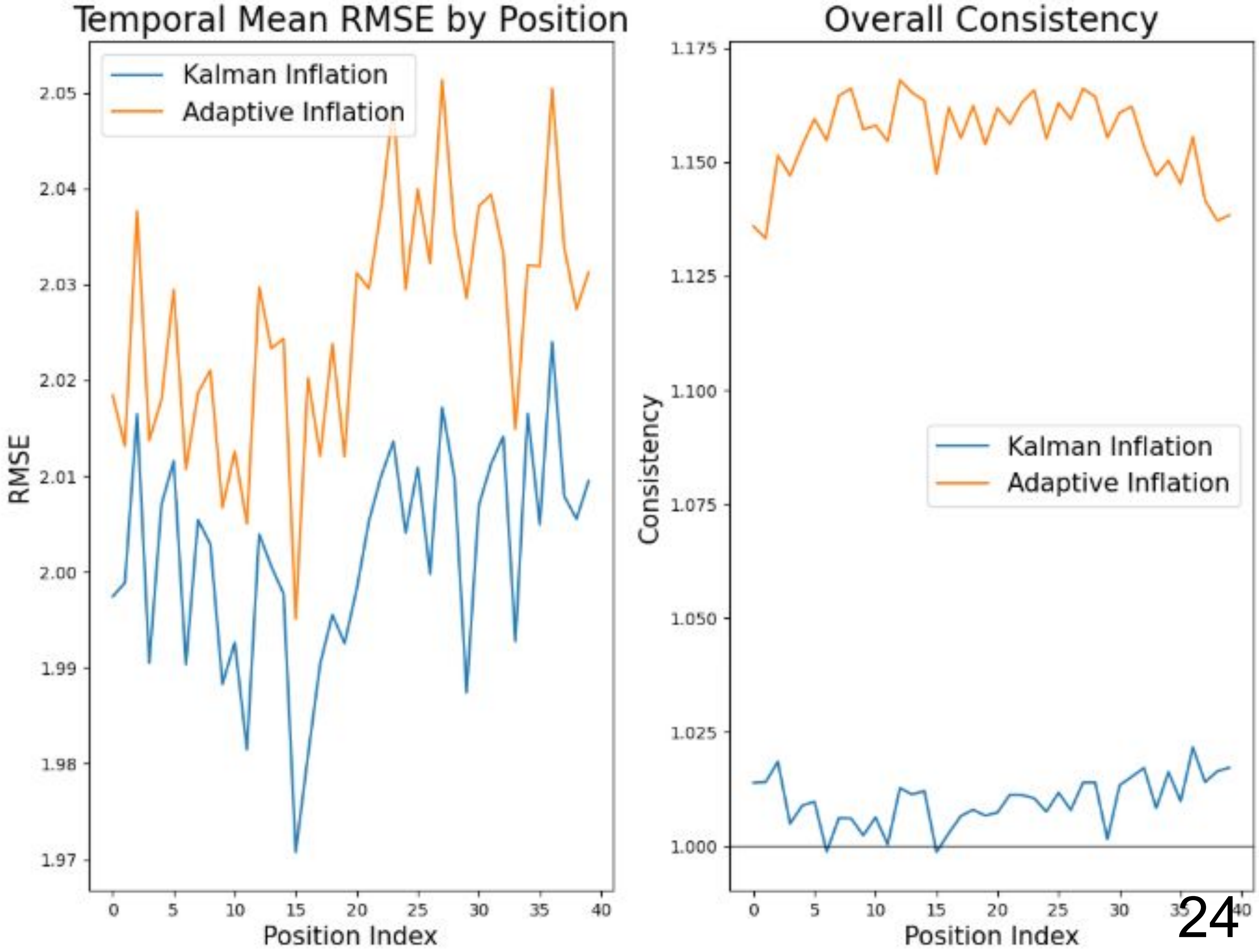


Forecast Position shift likely due to the motion of waves

Obs: Everywhere

Order: Left to Right

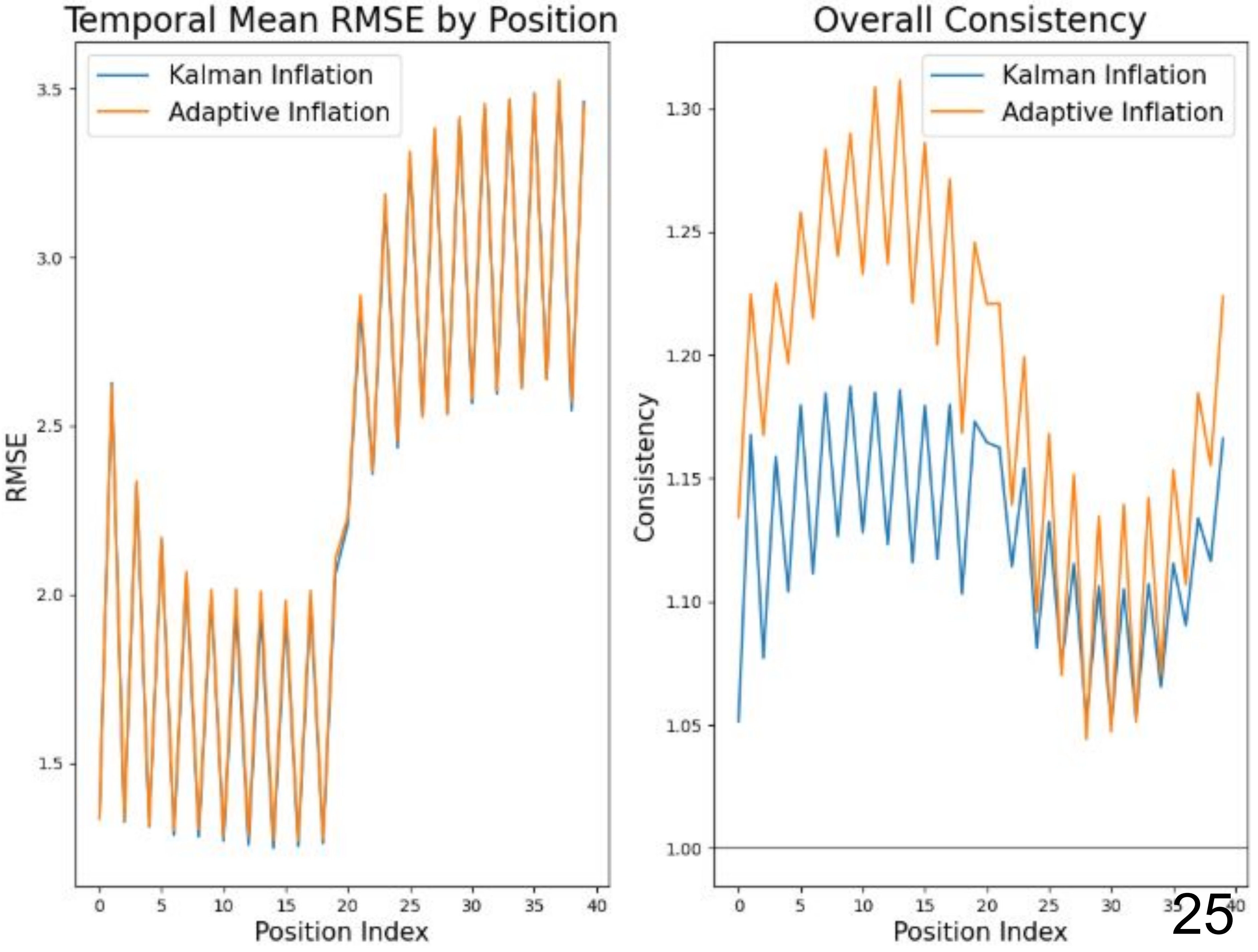
Ensemble: Forecast



Kalman Inflation is more sane for varying observation systems

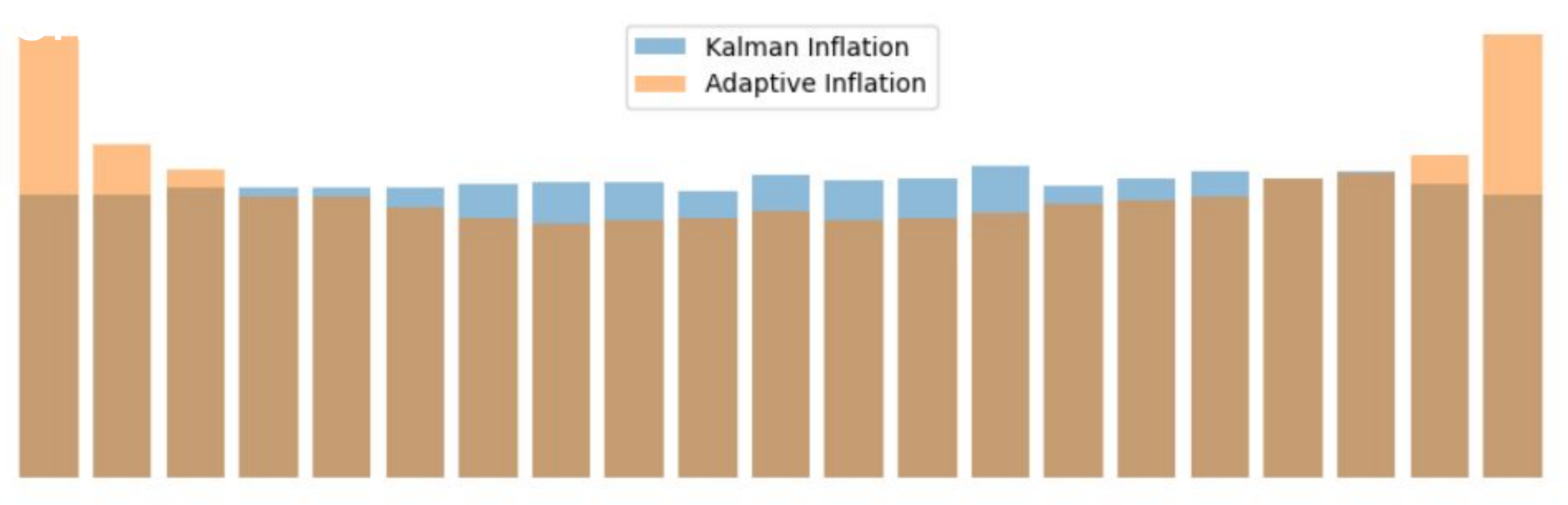
Obs: Varying
0000 | Even indices
1200 | First 20

Ensemble: Analysis



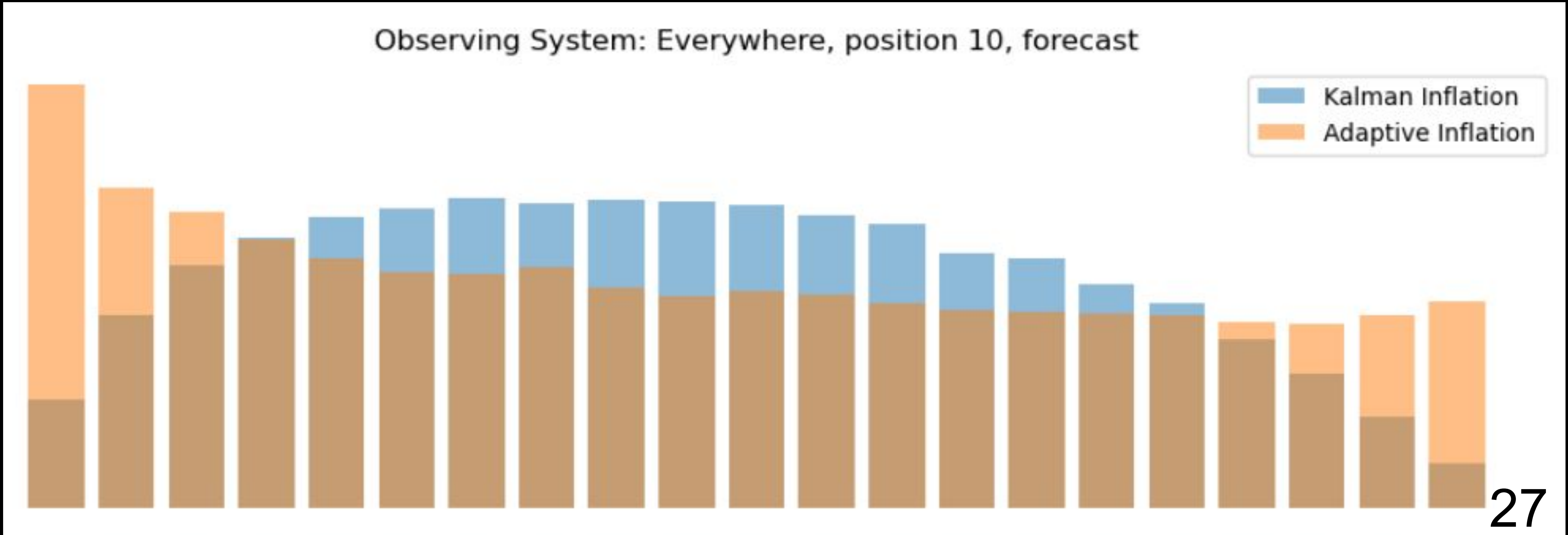
Rank histograms look better for covariance inflation

- True state more appropriately random draw from



Rank histograms look better for covariance inflation

- More stable for varying observation systems



Future Work

- Can be applied to non-gaussian / non-linear contexts
  - Non-linear observation operators (e.g. satellite radiances)
  - DA algorithms that use transport maps

- Can be adaptive as well

- Can be applied to more complex models

Thank you for listening!

Bibliography

1. El Gharamti, M. (2018). Enhanced Adaptive Inflation Algorithm for Ensemble Filters. *Mon. Weather Rev.*, **146**, 2, 623-640. DOI: 10.1175/MWR-D-17-0187.1
2. E. N. Lorenz (1995). Predictability: a problem partly solved. *ECMWF, Seminar on Predictability, 4-8 September 1995*. Accessible at https://www.ecmwf.int/sites/default/files/elibrary/1995/10829-predictability-problem-partly-solved.pdf
3. El Gharamti, M., Kerhsaw, H., Raeder, K., Raczka, B., Johnson, B., Smith, M., Anderson, J. L., Amrhein, D., Collins, N., Grooms, I., Kugler, L. (2025). The Data Assimilation Research Testbed: A Robust, Scalable Software Facility with Groundbreakding Capabilities for Model-Data Integration. *Bull. Am. Meteorol. Soc.*, **106**, 11. DOI: 10.1175/BAMS-D-24-0214.1

QnA