How does covariance inflation impact EnKF-initialized convection-allowing ensemble forecasts?

Craig Schwartz, Glen Romine, and Kate Fossell

The National Center for Atmospheric Research

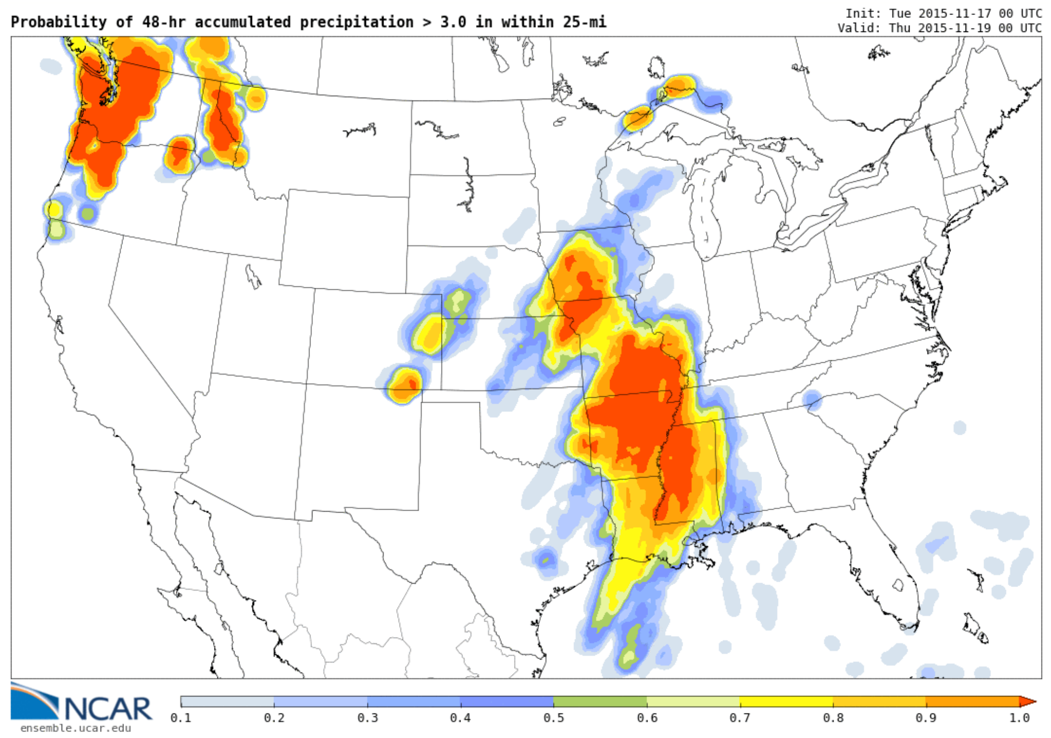
schwartz@ucar.edu

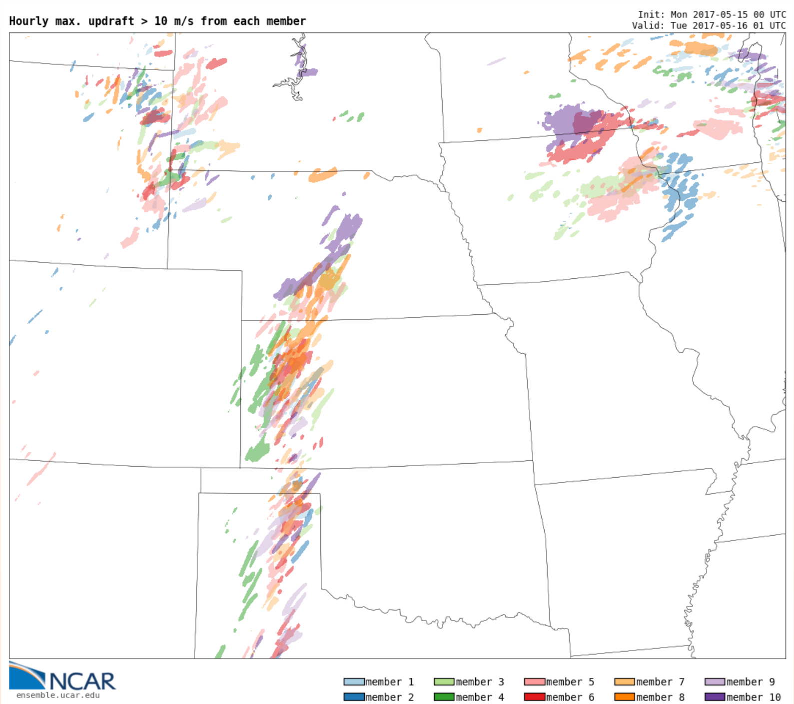
NCAR is sponsored by the National Science Foundation

This work partially supported by NOAA Grant No. NA17OAR4590182 and the NCAR STEP program

The NCAR ensemble

- Between April 2015 – Dec 2017, NCAR produced real-time, 48-h ensemble forecasts
  - 3-km horizontal grid spacing; 10 members
  - Initialized at 0000 UTC daily





http://ensemble.ucar.edu; Schwartz et al. (2015; doi:10.1175/WAF-D-15-0103.1)

Components of the NCAR ensemble

1) Ensemble analysis system

- Assimilated observations every 6 hours with a continuously-cycling ensemble Kalman filter (EnKF)
- 15-km horizontal grid spacing
- Self-contained system (other than boundary conditions)

2) Ensemble prediction system

- 10-member, 3-km ensemble forecasts
- 48-h forecasts initialized at 0000 UTC
- Single physics EPS

– Both components used the WRF-ARW model

Purpose of the NCAR ensemble

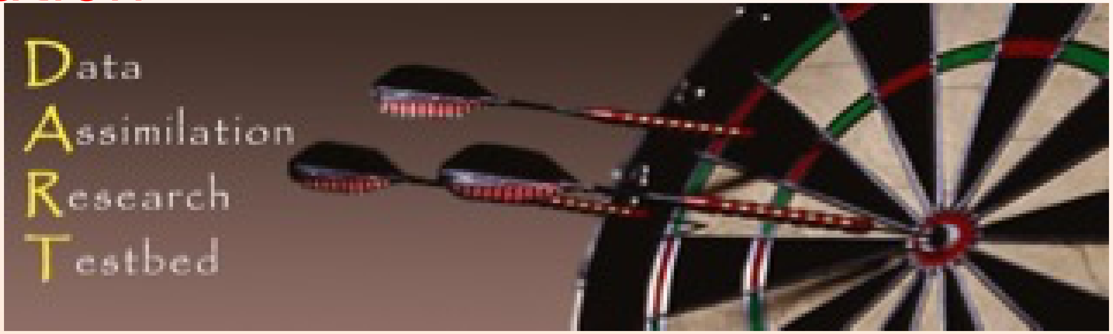
- Demonstrate a real-time convection-allowing ensemble spanning the conterminous United States
  - USA did not have an operational convection-allowing ensemble when project began

- One of the forefronts of NWP model research is high-resolution ensemble design
  - Initial condition perturbations and DA approaches
  - NCAR ensemble as framework for testing

- This study systematically examines inflation methods within NCAR ensemble framework

NCAR ensemble DA configurations

- Ensemble Adjustment Kalman Filter (EAKF)
  - From Data Assimilation Research Testbed (DART) software, which is heavily used for research with many models
  - Horizontal localization half-width: 640 km
  - Vertical localization half-width: 0.5 scale heights
  - 50 ensemble members (increased to 80 on 2 May 2016)
  - DART adaptive inflation



https://www.image.ucar.edu/DAReS/DART

Adaptive inflation

- Anderson (2009; doi: 10.1111/j.1600-0870.2008.00361.x)
  - El Gharamti (2018; *Mon. Wea. Rev.*) proposes enhancements
  - Each state variable has its own inflation value that is *updated during observation assimilation*
  - Designed to work with minimal tuning
  - Computationally efficient

$$\delta \boldsymbol{x}_i^* \leftarrow \delta \boldsymbol{x}_i \sqrt{\lambda}$$

These can be either prior or posterior quantities

$\delta \boldsymbol{x}_i$: *i*th member's perturbation about mean

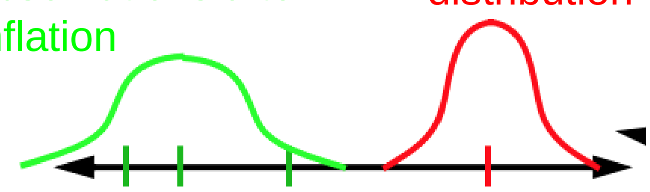
$\delta \boldsymbol{x}_i^*$: *i*th member's perturbation about mean after inflation

λ : inflation factor

$\delta \boldsymbol{x}_i$: *i*th member's perturbation about mean
$\delta \boldsymbol{x}_i^*$: *i*th member's perturbation about mean after inflation
λ : inflation factor

Adaptive inflation complexity

- Complexity arises from determining and updating λ
  - Temporal variation of λ can be difficult to understand
  - = Bayesian update based on mismatch between prior observations and observation distribution



from Kevin Raeder et al.

7. Appendix A: Computing the updated inflation distribution

The mean of the updated inflated distribution is estimated using an approximation of the mode of (6). One could also attempt to approximate the mean of (6) by evaluating the integral of the product of the prior and likelihood terms.

The likelihood (13) is first approximated by a Taylor expansion truncated at first order:

$$p(y^o|\lambda) \cong \bar{l} + l'(\lambda - \bar{\lambda}_p), \quad \text{((A.1))}$$

where

$$\bar{l} = p(y^o|\bar{\lambda}_p) = (\sqrt{2\pi}\bar{\theta})^{-1} \exp\left(-\tfrac{1}{2} D^2 \bar{\theta}^{-2}\right), \quad \text{((A.2))}$$

$$l' = \frac{\partial}{\partial\lambda}[p(y^o|\lambda)]_{\bar{\lambda}_p} = \bar{l}(D^2\bar{\theta}^{-2} - 1)\, \bar{\theta}^{-1} (\partial\theta/\partial\lambda)_{\bar{\lambda}_p}, \quad \text{((A.3))}$$

$$\bar{\theta} \equiv \sqrt{\bar{\lambda}^o \sigma_p^2 + \sigma_o^2}, \quad \text{((A.4))}$$

$$\bar{\lambda}^o \equiv [1 + \gamma(\sqrt{\bar{\lambda}_p} - 1)]^2 \quad \text{((A.5))}$$

and

$$(\partial\theta/\partial\lambda)_{\bar{\lambda}_p} = \tfrac{1}{2}\sigma_p^2 \gamma (1 - \gamma + \gamma\sqrt{\bar{\lambda}_p}) / (\bar{\theta}\sqrt{\bar{\lambda}_p}). \quad \text{((A.6))}$$

The approximate mode of the posterior (6) can be found by setting the derivative of the approximate posterior to 0 after removing the constant coefficient from the prior Gaussian

$$\frac{\partial}{\partial\lambda}\left\{[\bar{l} + l'(\lambda - \bar{\lambda}_p)] \exp\left[-\tfrac{1}{2}(\lambda - \bar{\lambda}_p)^2 \sigma_\lambda^{-2}\right]\right\} = 0 \quad \text{((A.7))}$$

and solving for λ. This results in a quadratic equation for λ:

$$\lambda^2 + (\bar{l}/l' - 2\bar{\lambda}_p)\lambda + (\bar{\lambda}_p^2 - \sigma_\lambda^2 - \bar{l}\bar{\lambda}_p/l') = 0. \quad \text{((A.8))}$$

Solutions to (A.8) are two real roots, and the one closest to $\bar{\lambda}_p$ is selected.

Anderson (2009)

Spread restoration

- Increases ensemble spread during assimilation

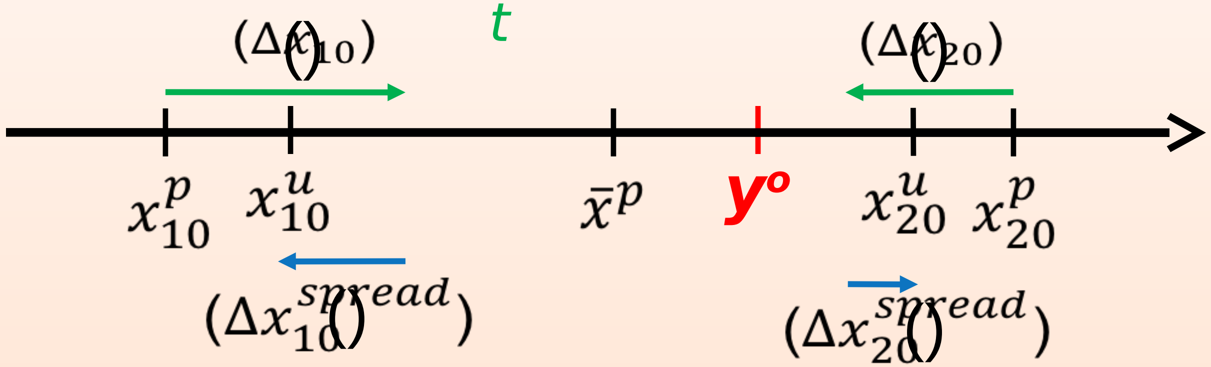
$$\text{Let } a = \left(\frac{\sigma_o^2}{\sigma_f^2+\sigma_o^2}\right)^{1/2} \text{ be constrained} \leq 1$$

$$\text{Then, } f = \left[1 + \frac{a-1}{1.63(N_e)-2.47}\right]^{-1} - 1 \geq 0 \text{ (since } a \leq 1\text{)}$$

N_e: ensemble size

$$\Delta x_i^* = \Delta x_i + f(x_i^p - \bar{x}^p)$$

Total increment for ith member = *Traditional increment* + *Spread restoration increment* (Δx_i^{spread})



Relaxation to prior spread (RTPS)

- Whitaker and Hamill (2012; 10.1175/MWR-D-11-00276.1)
  - Easy to use; just one tunable parameter (α)
  - Applied to posterior state
  - Amount of inflation reflects observation network

$$\boldsymbol{\delta x}^*_{i,a} \leftarrow \boldsymbol{\delta x}_{i,a}\left(\alpha\,\frac{\boldsymbol{\sigma}_b - \boldsymbol{\sigma}_a}{\boldsymbol{\sigma}_a} + 1\right)$$

$$\boldsymbol{\sigma}^*_a \leftarrow \alpha\boldsymbol{\sigma}_b + (1-\alpha)\boldsymbol{\sigma}_a$$

Equivalent

$\boldsymbol{\delta x}_{i,a}$: ith member's posterior perturbation about mean

$\boldsymbol{\sigma}_a$: posterior spread (before inflation)

$\boldsymbol{\sigma}^*_a$: posterior spread (after inflation)

$\boldsymbol{\sigma}_b$: prior spread

α : tunable inflation parameter

Inflation experiments

1) Prior adaptive inflation

2) Prior adaptive inflation with spread restoration

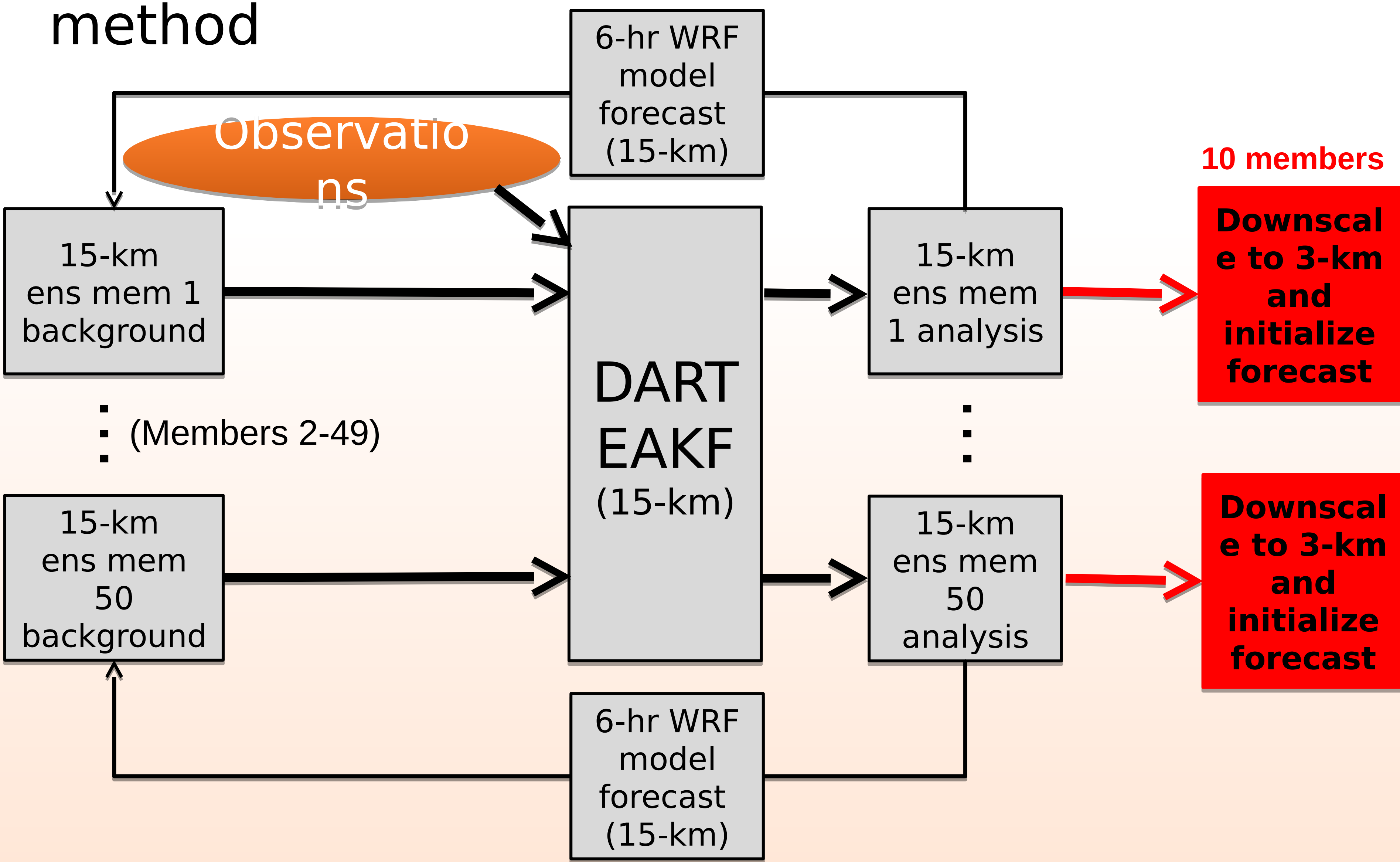
3) Combination of prior and posterior adaptive inflation

4) RTPS inflation (posterior; $\alpha = 1.12$)

- Otherwise, identical model and DA configurations
  - 50 member continuously-cycling EnKF (15-km)
  - 6-h cycles
  - 10-member, 3-km, 48-h “free forecasts”
  - Initialized 0000 UTC 1 – 31 May 2015
  - Identical observations available for assimilation

Experimental design

- Four identical DA systems, except for inflation method



NCAR ensemble analysis domain (15-km)

Radiosonde

Aircraft

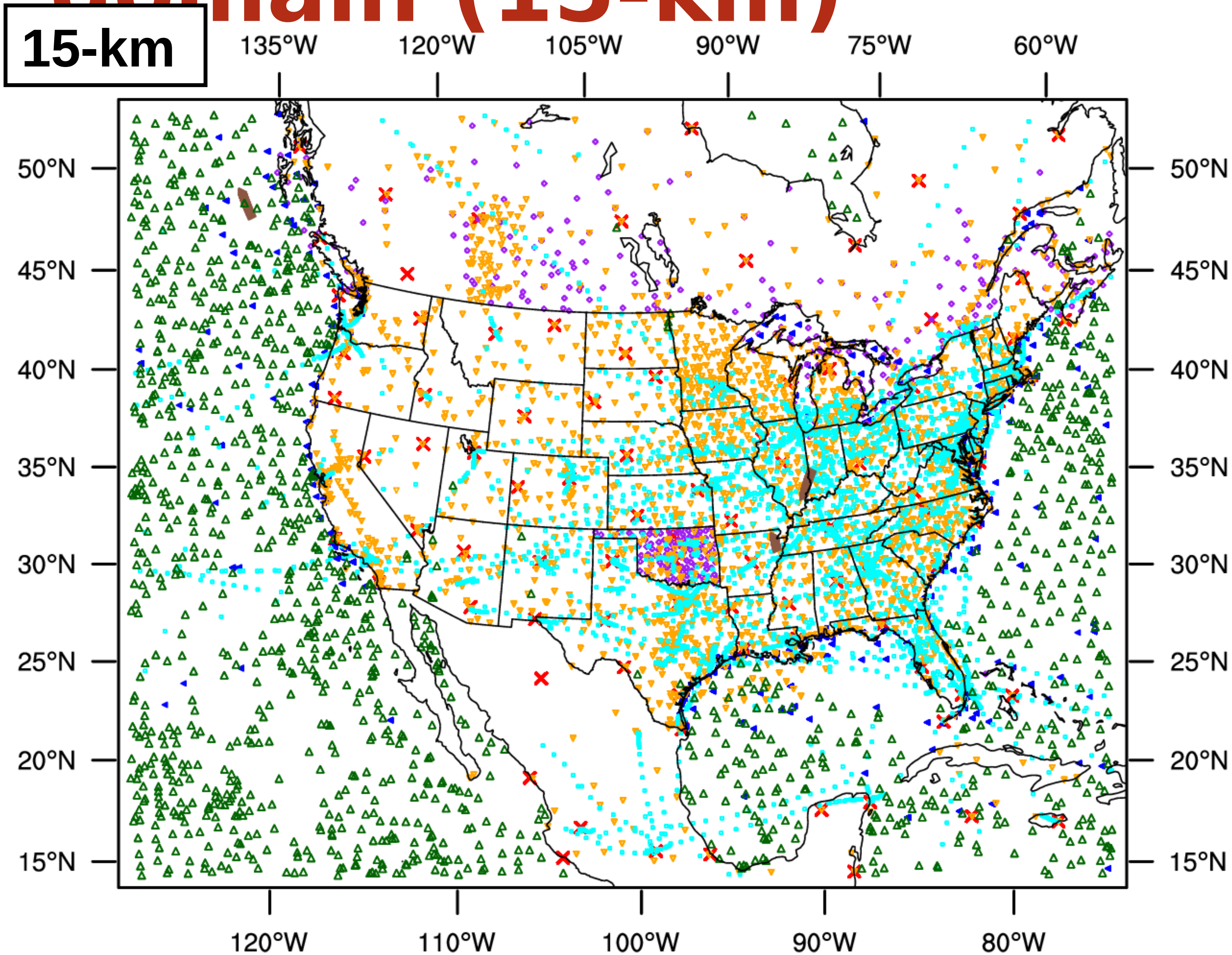
Satellite wind

METAR

MESONET

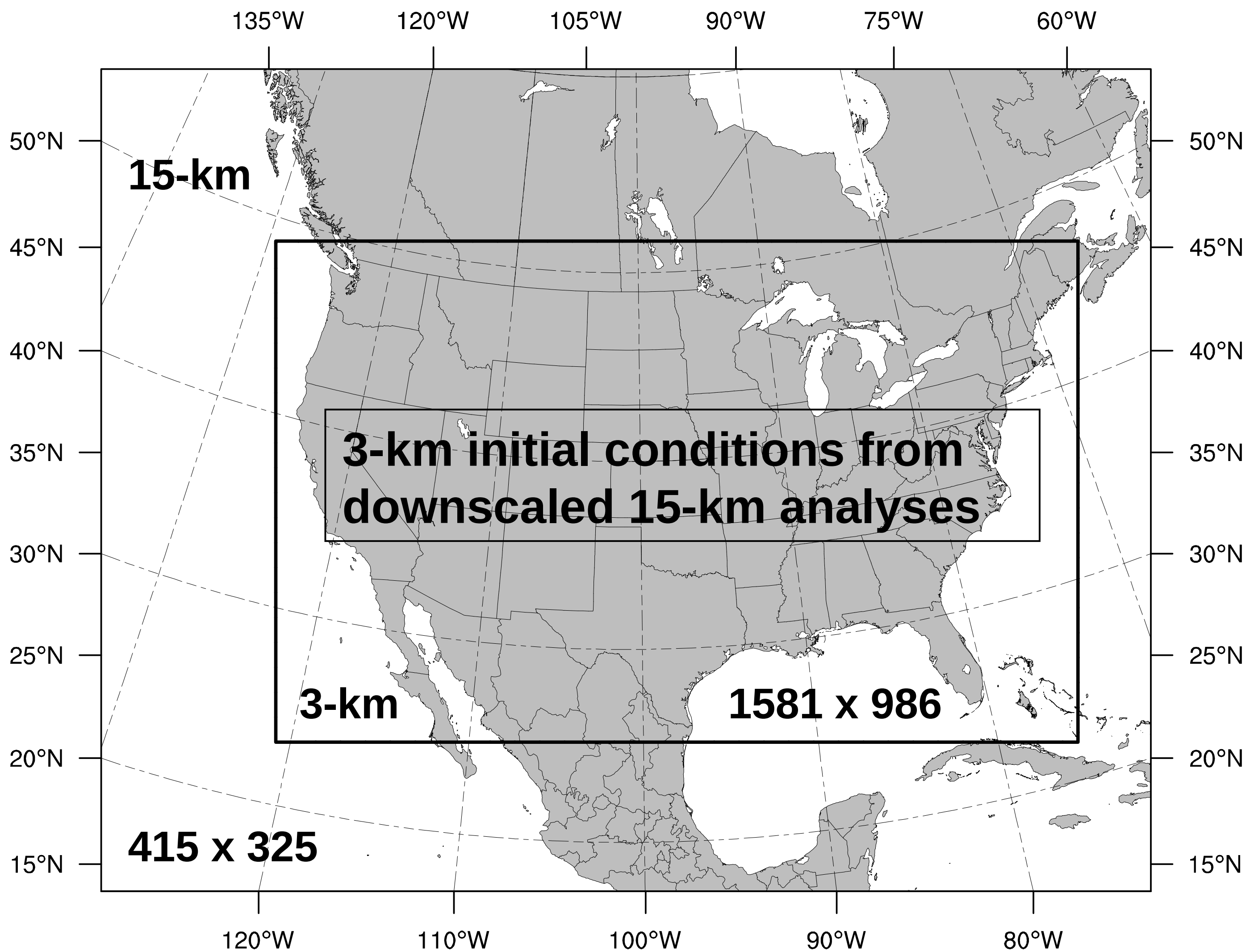
Marine

GPSRO



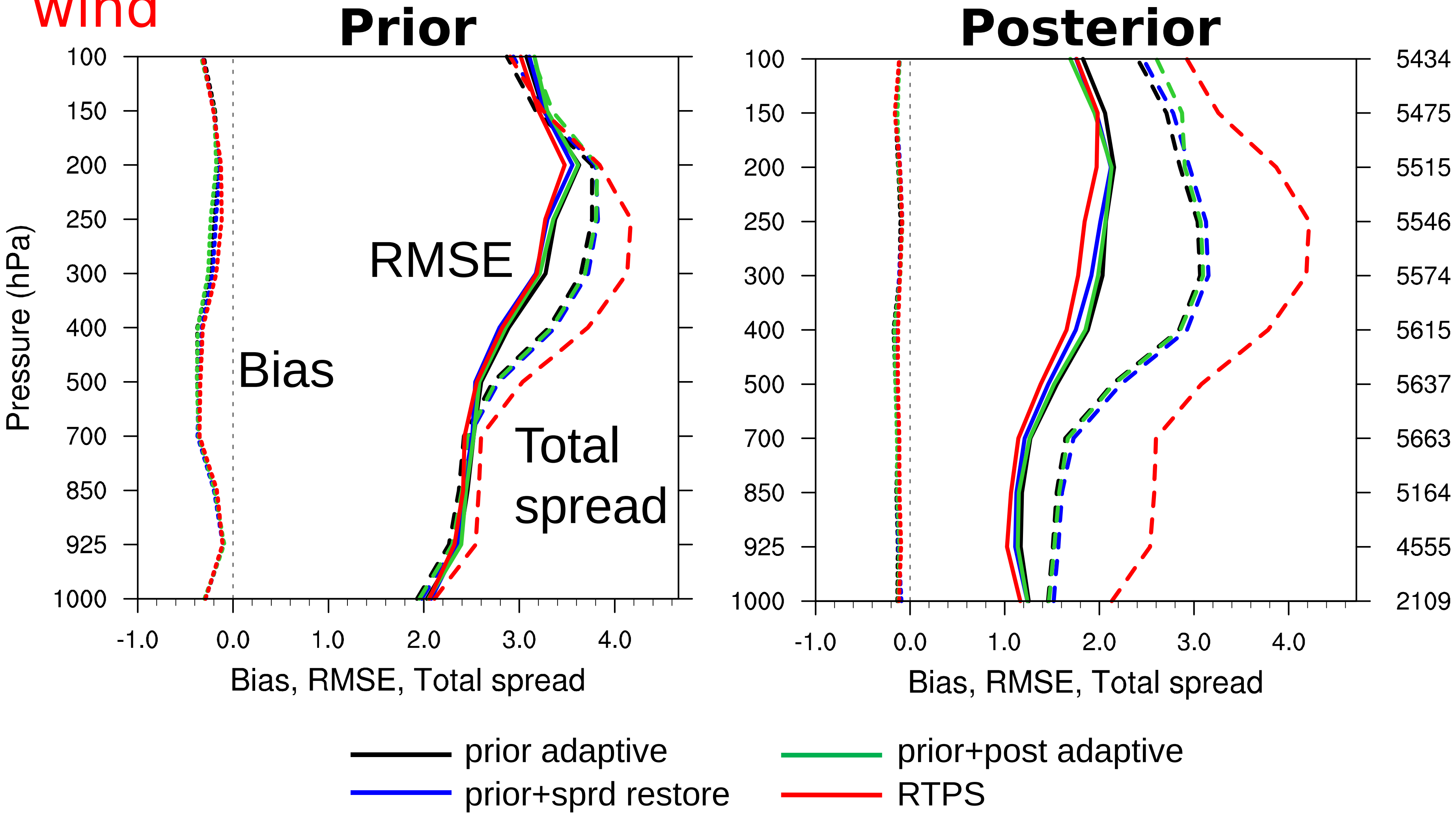
NCAR ensemble forecast domain

- 48-h, 10-member, 3-km forecasts



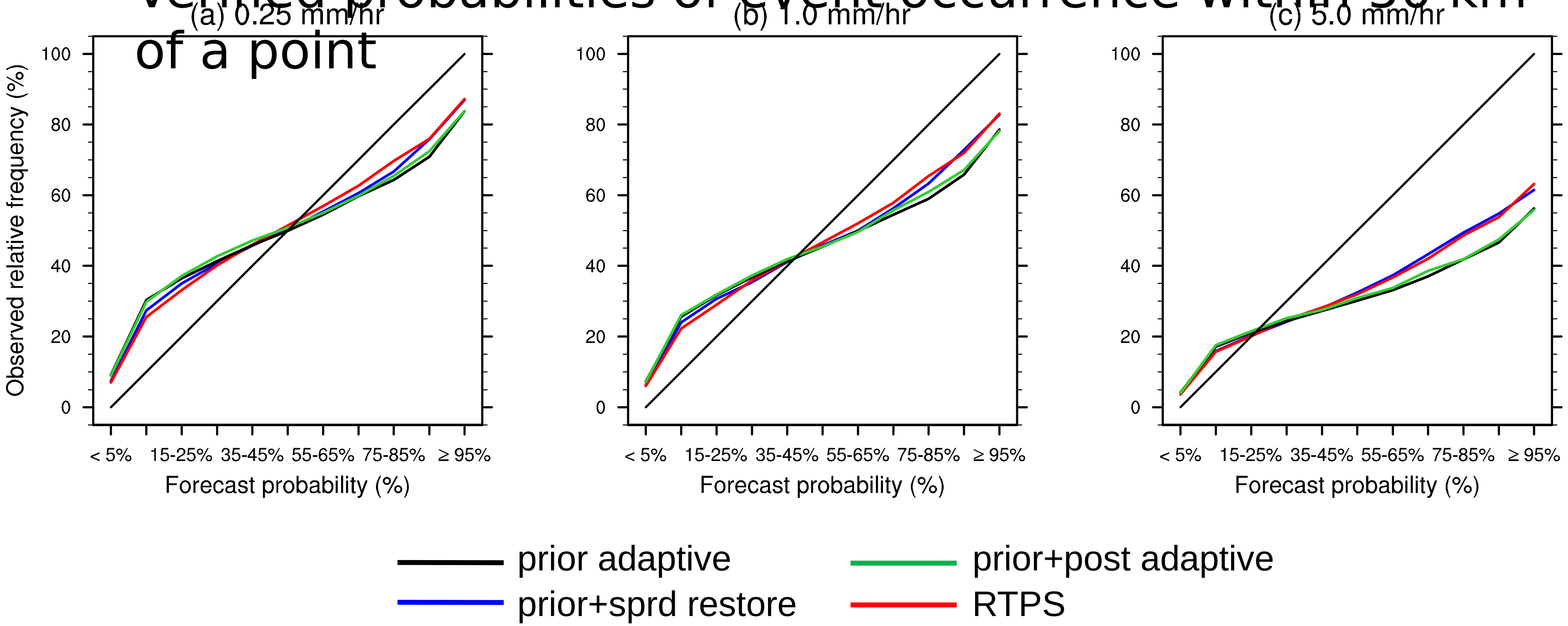
Prior/posterior EAKF statistics

- Prior/posterior stats (121 cycles) for radiosonde u-wind



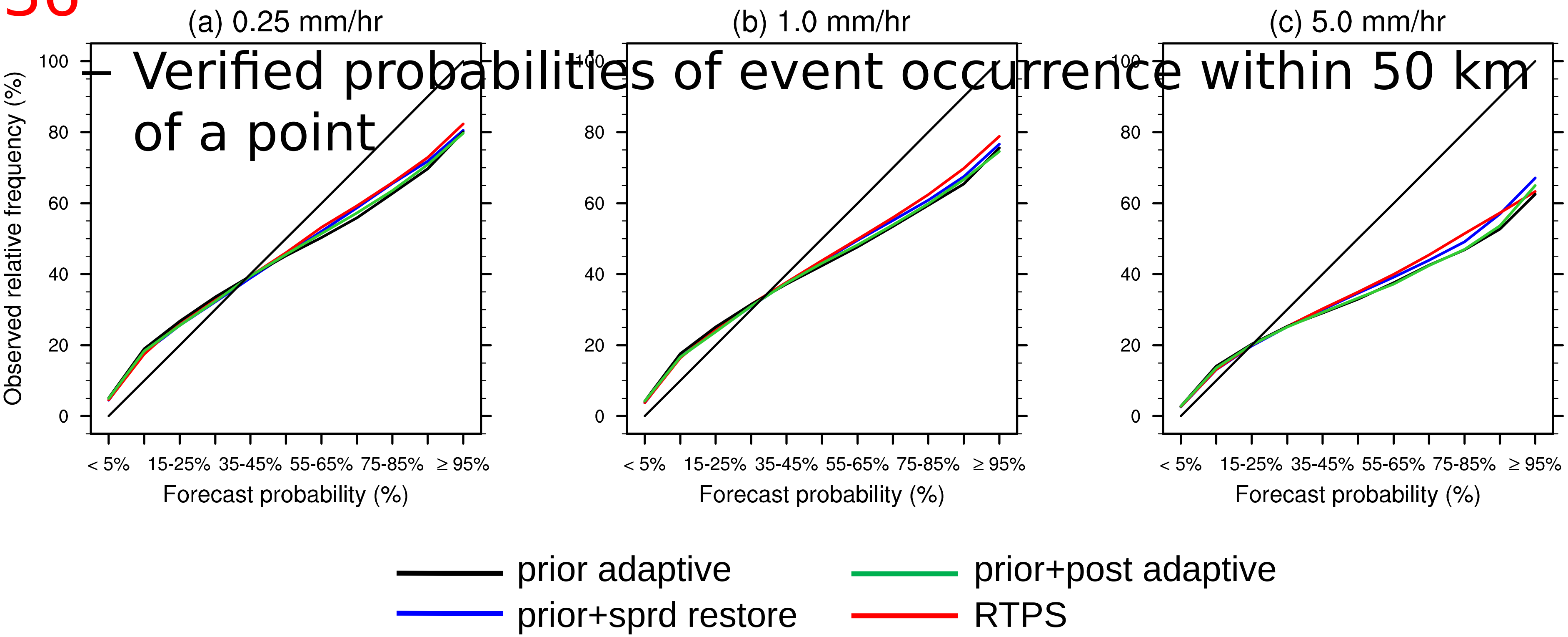
Reliability of precipitation forecasts

- Reliability diagrams for 1-h accumulated precipitation aggregated over 31 forecasts over hours 1-12
  - Verified probabilities of event occurrence within 50 km of a point



Reliability of precipitation forecasts

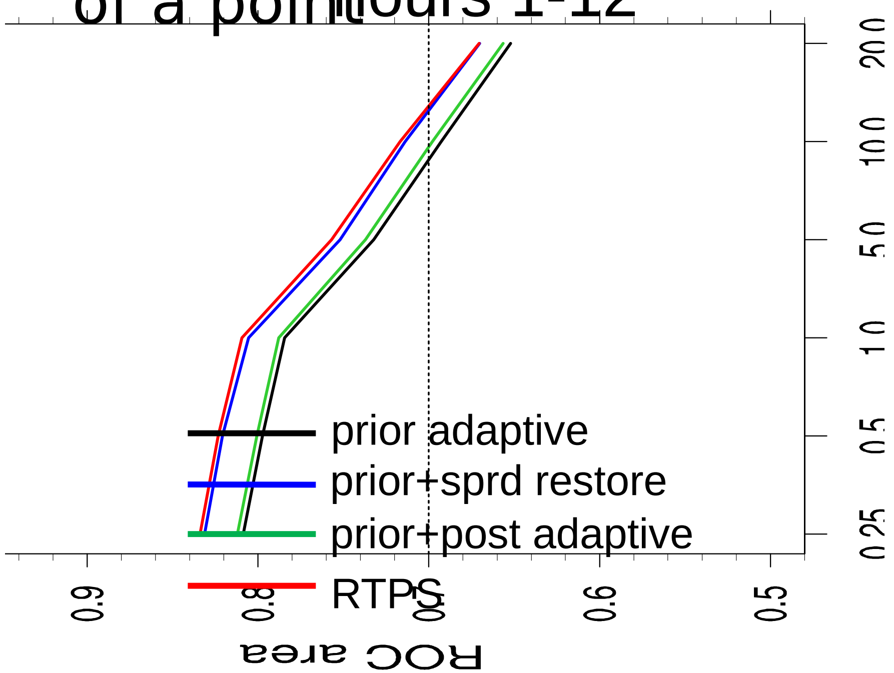
- Reliability diagrams for 1-h accumulated precipitation aggregated over 31 forecasts over hours 18-36
  - Verified probabilities of event occurrence within 50 km of a point



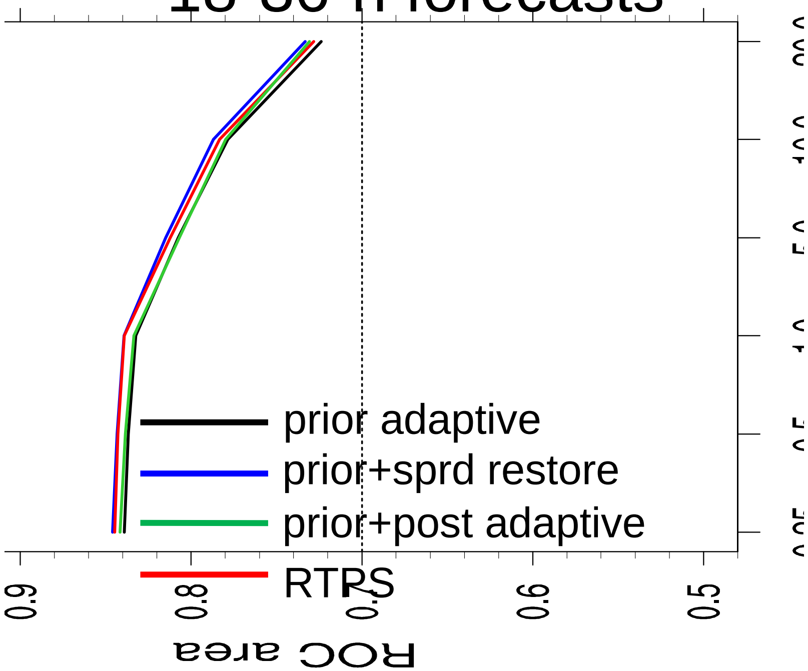
ROC areas for precipitation forecasts

- ROC areas for 1-h accumulated precipitation aggregated
  over 31 forecasts
  - Verified probabilities of event occurrence within 50 km of a point

hours 1-12



18-36-h forecasts



Some thoughts

- Can prior adaptive and RTPS inflation be tuned to achieve comparable results?
- Does spread restoration somewhat act like posterior inflation?
  - Should spread restoration be used with traditional posterior inflation?
- How is the EnKF “spread/skill” relationship related to high-resolution ensemble *forecast* spread?

What about non-EnKF initialization?

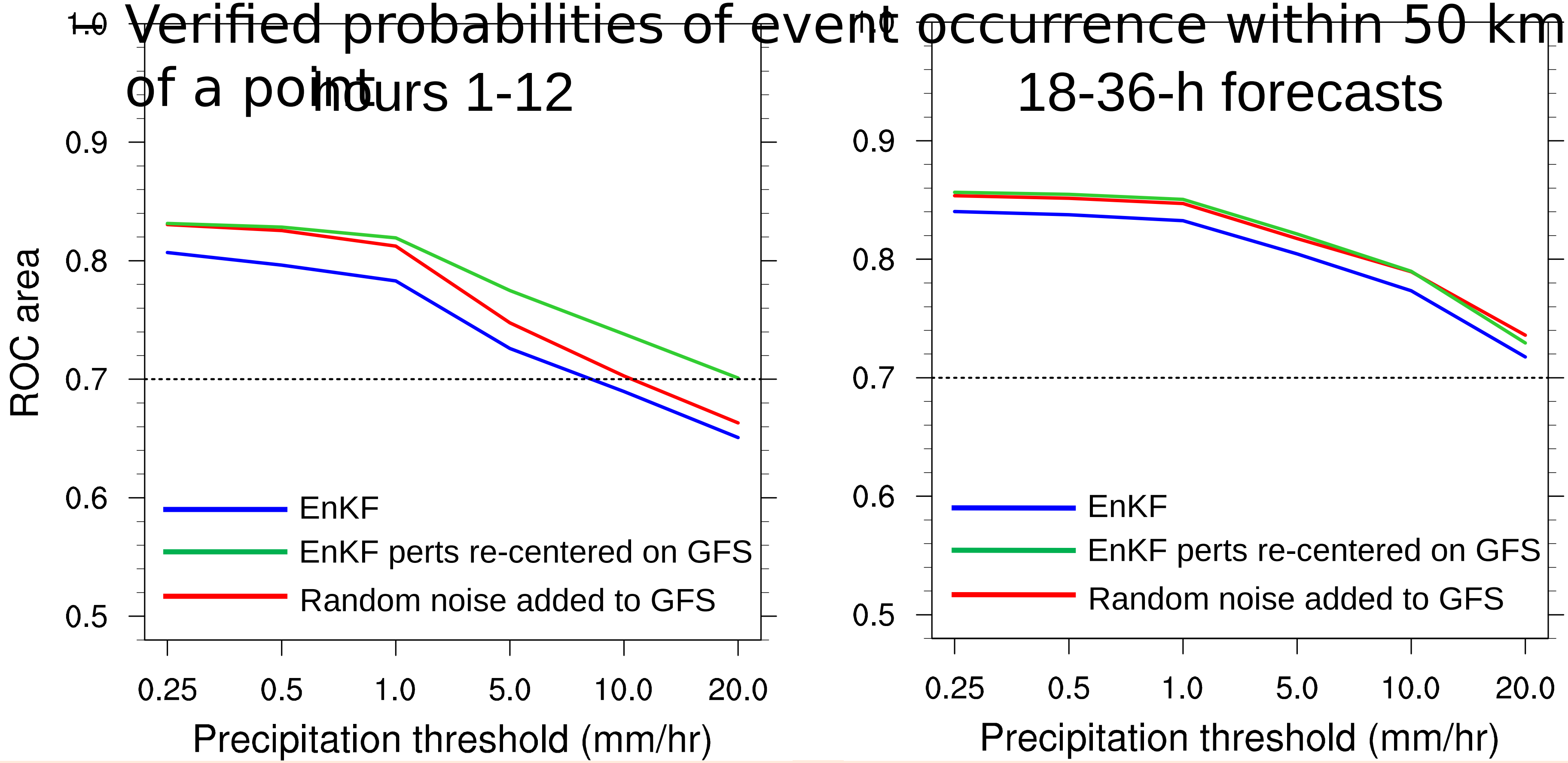
- Are EnKFs the best way to initialize high-resolution ensembles?

- Ran more 10-member ensemble forecasts with non-EnKF initial condition perturbations:
  - Added random, correlated noise to GFS analyses
  - Re-centered EnKF perturbations about GFS analyses

- Same configurations as inflation experiments

ROC areas for precipitation forecasts

- ROC areas for 1-h accumulated precipitation aggregated over 31 forecasts
  - Verified probabilities of event occurrence within 50 km of a point

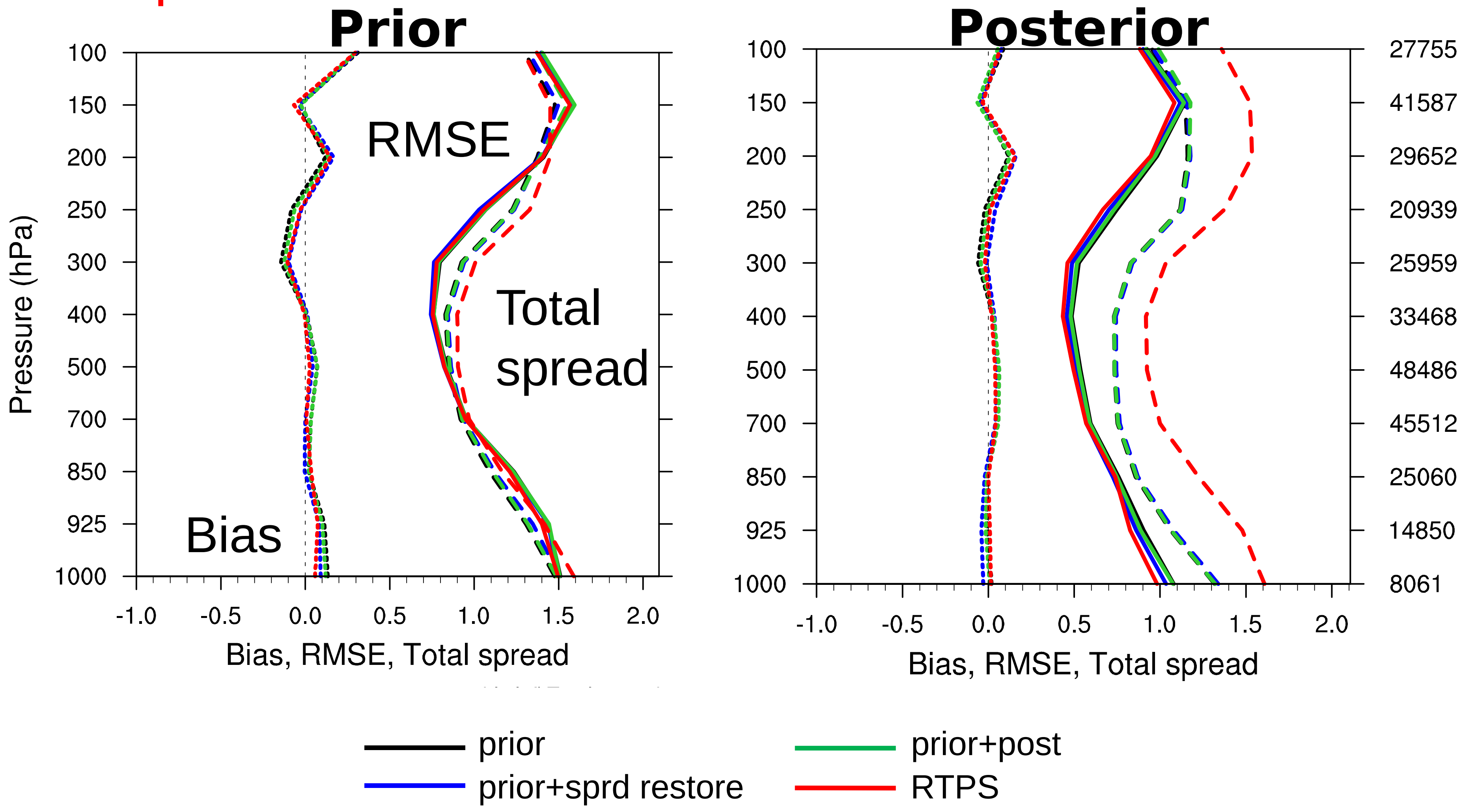


Future work

- Improve EnKFs so they initialize high-res ensemble forecasts better than ad hoc initialization methods over the conterminous United States
- Better understand character of initial perturbations
  – Spectral analysis
- Develop a 3-km, continuously-cycling EnKF over the conterminous United States
  – More frequent analyses (hourly)
  – Assimilate radar and geostationary satellite observations
  – Requires substantially more computational resources

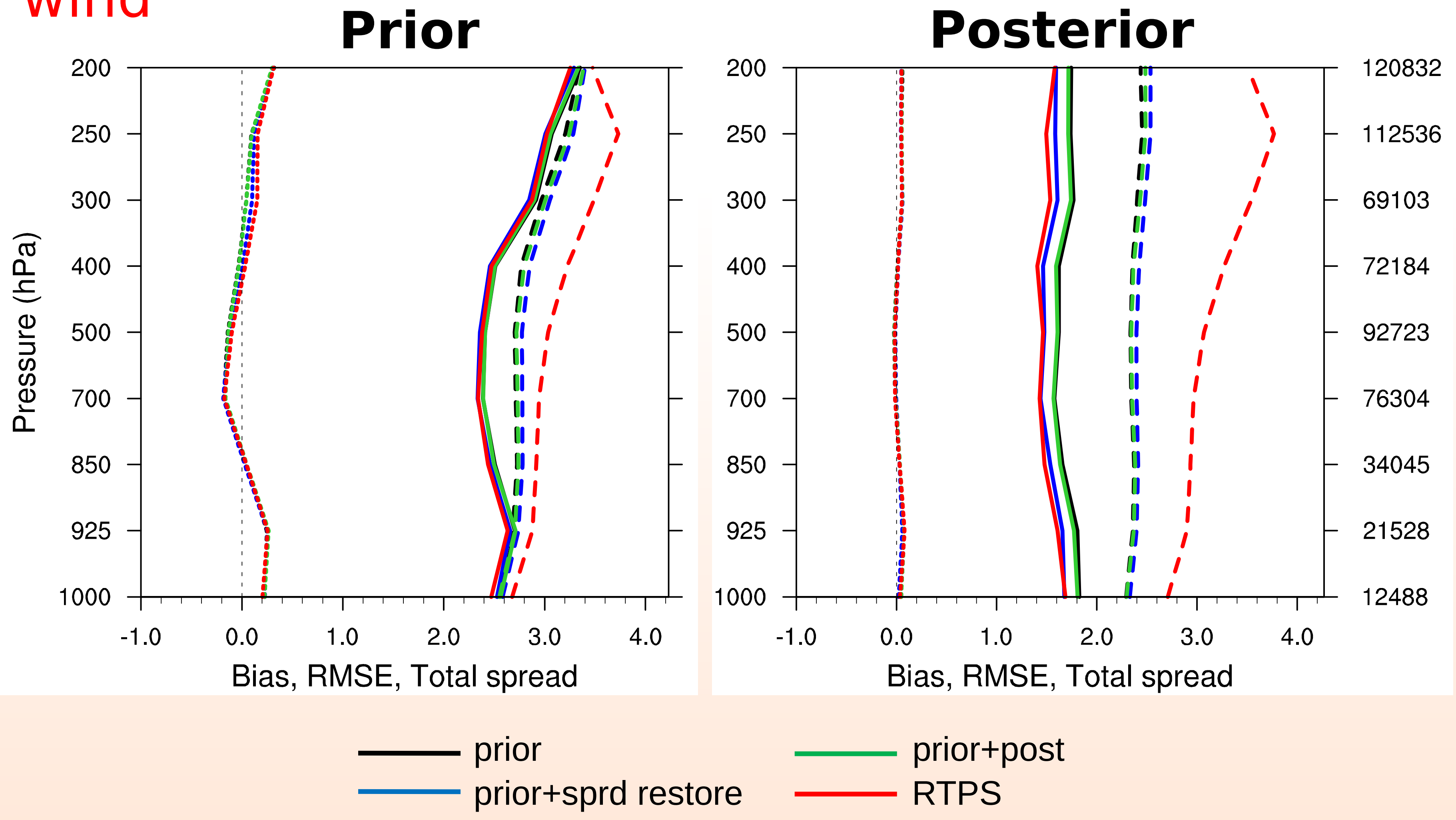
Prior/posterior EAKF statistics

- Prior/posterior stats (121 cycles) for radiosonde temperature



Prior/posterior statistics

- Prior/posterior statistics (121 cycles) for aircraft u-wind



Prior/posterior statistics

- Prior/posterior statistics (121 cycles) for aircraft temperature