# Direct Ensemble Assimilation

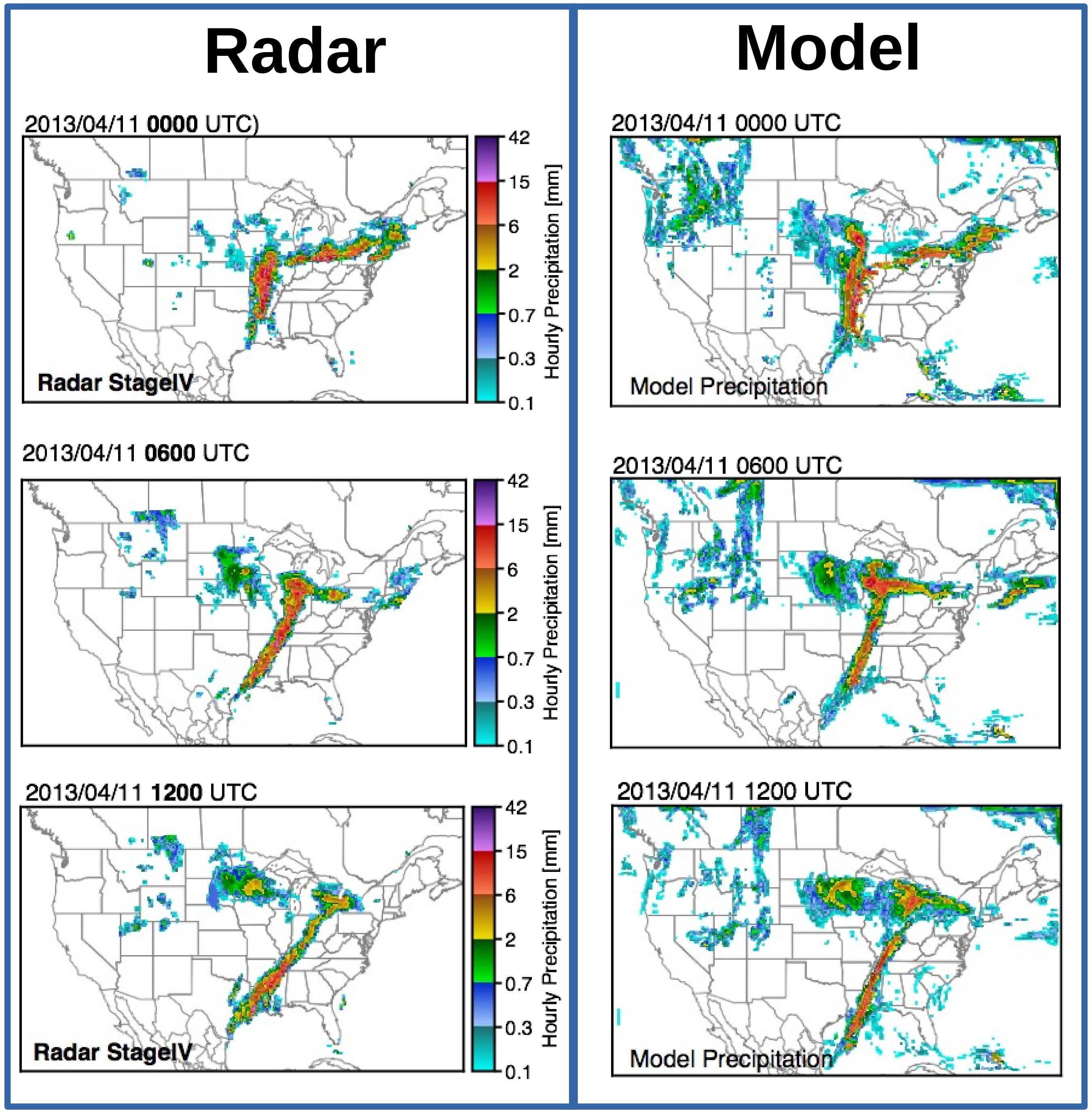
## A heuristic method for radar data assimilation

**Andrés Pérez Hortal**, Isztar Zawadzki and M.K. (Peter) Yau

McGill University, Montréal

**MOTIVATION: improve the nowcast of precipitation**

# Test case in an OSSE

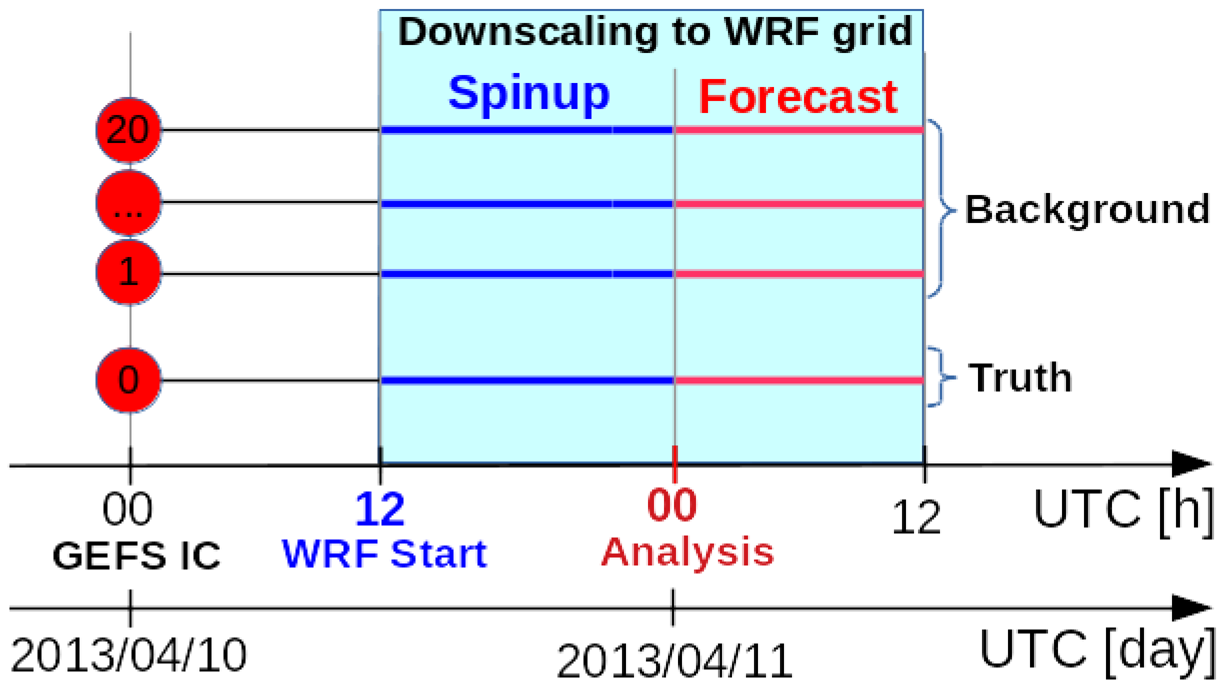



Good agreement between model forecast and observations!

Simulation of a precipitation system by WRF at 20 km resolution with

Microphysics: WRF single moment, 3 class (WSM3)
Boundary layer scheme: Yonsei University (YSU)
Cumulus parametrization: Kain-Fritsch
Shortwave Radiation: Dudhia scheme
Longwave radiation: Rapid Radiative Transfer Model (RRTM)

**An ensemble of 21 members was produced.**
Every member has different initial and boundary conditions




**In our OSSE one member of an ensemble forecast is taken as “truth” any other member is a background.**

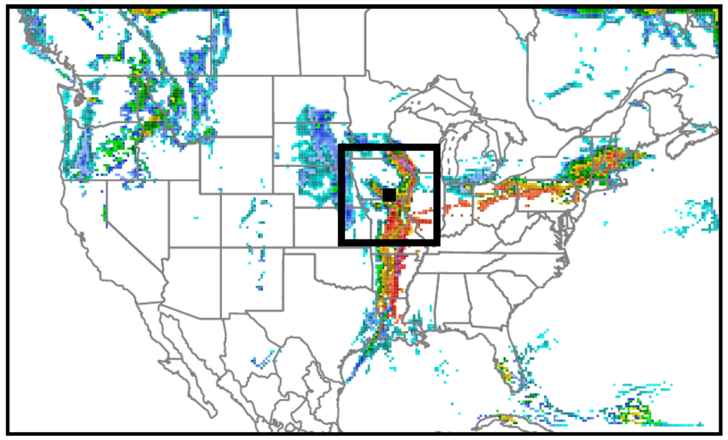
# The DA method in a nutshell

## Direct Ensemble Assimilation

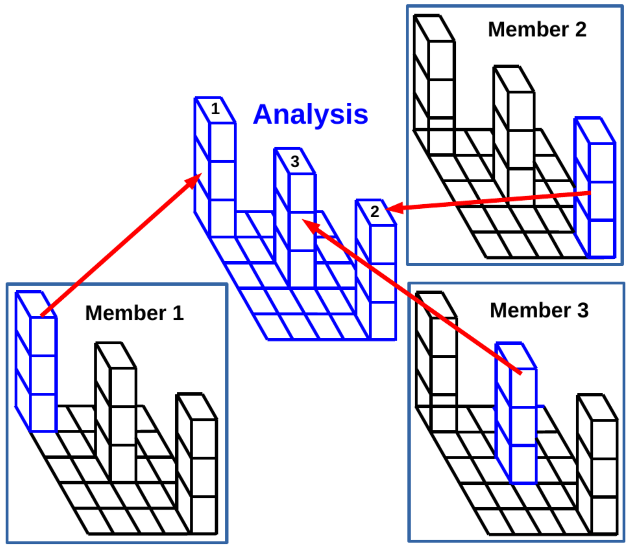
**CONSTRUCT THE ANALYSIS:**

1) For every grid point determine **M**ean **A**bsolute **D**ifference in precipitation (over a spatiotemporal window centred at the grid point) between each ensemble member and the "truth". Choose the member with the smallest MAD. **The window used is 820kmx820kmx30min**.

2) For each grid point the **analysis state** is given by the soundings of (*U, V, Θ,* $q_v$) of the ensemble member whose **MAD** - mean over a window centered at the grid point - is **closest to truth**. The other state variables are given by the background.

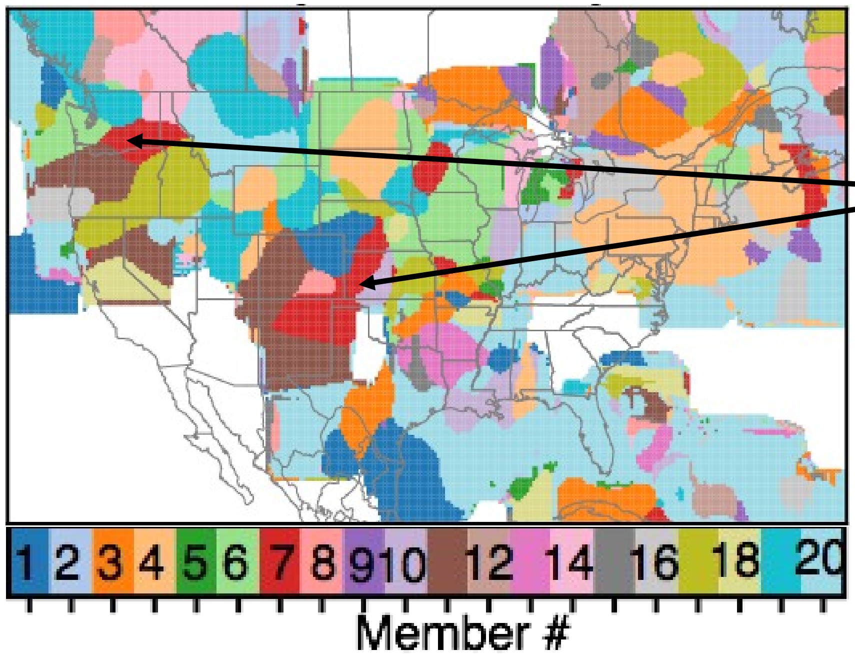



$$MAD_m(j) = \overline{\langle |dBZ - dBZ_t| \rangle}$$



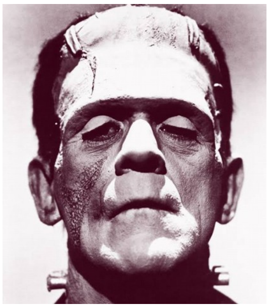

**The analysis state will have a decrease in precipitation errors (by construction).**
**We will see that there will be a decrease in the errors of (U, V, Θ, qv).**

# The DA method in a nutshell




**Example**: In all the red regions precipitation of M7 is closest to observations. The state variables of M7 in these zones contribute to the analysis at all heights: *U*(z), *V*(z), *Θ*(z), $q_v$(z).

**This mosaic of states will be called Frankenstate**



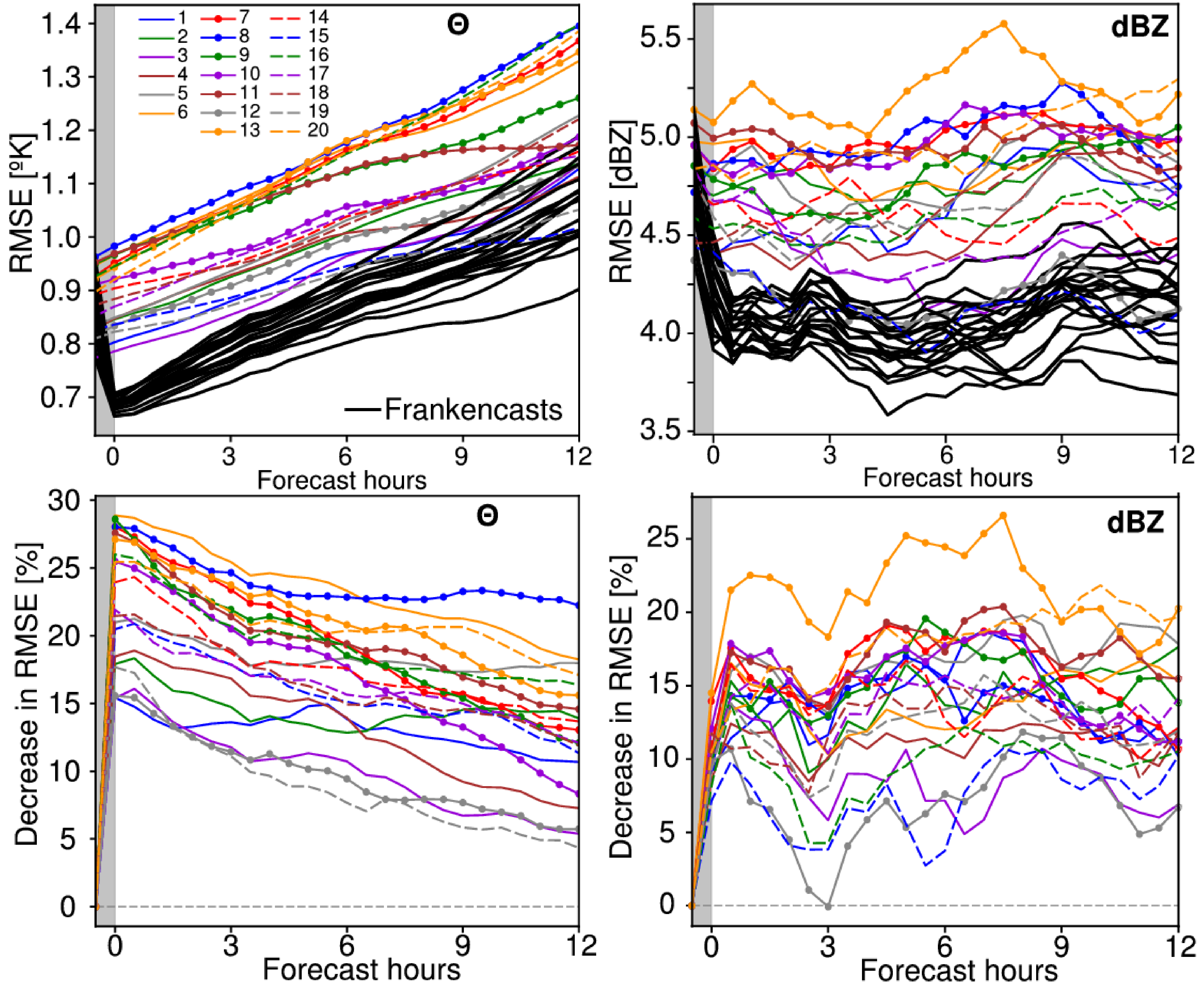
Whatever covariances (linear or not) between the observation and state variables exist they are implicit in the ensemble.

To **reduce the impact of imbalances** at border discontinuities between closest ensemble members, the initial conditions for the new ensemble forecast are obtained by nudging the background *U, V, Θ and* $q_v$ towards the analysis.

**After a 30 min nudging** of the 20 members toward the Frankenstate **we run the** 20-member **ensemble forecasts (Frankencasts)**:

# Example of Forecast

Similar results for the other state variables (U, V, and qv).




There is a **significant and persistent improvement** in the ensemble forecast.

Note that during the first hour the **decrease in dBZ error** is entirely **due to the improved model state (U,V,qv,T) .**

# Why and how does it work?

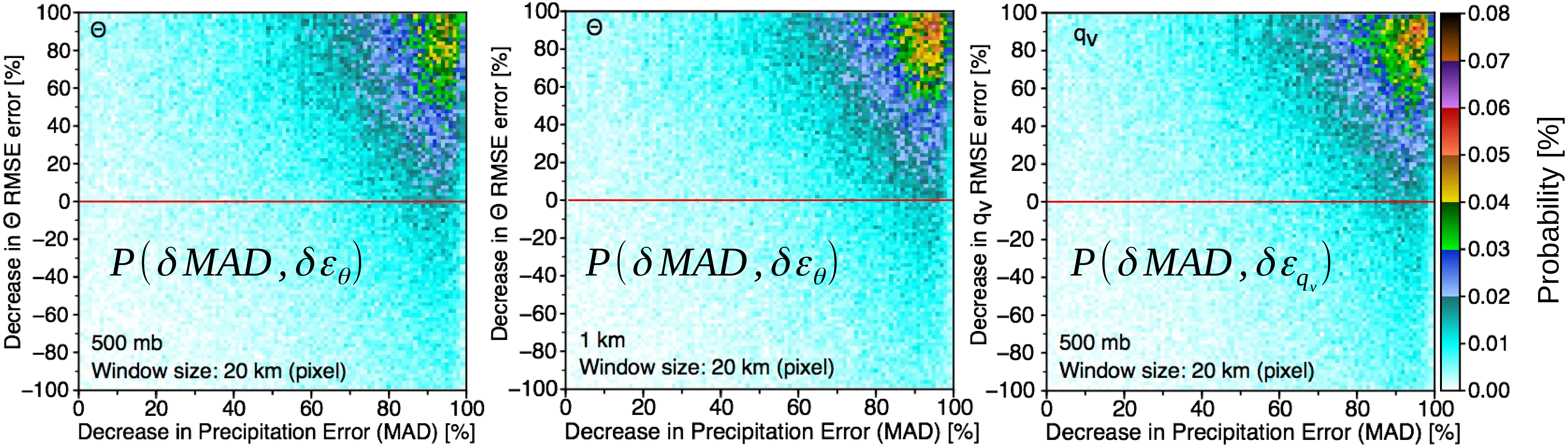
There is **one underlying hypothesis** to the method:

If **precipitation information** of an ensemble member is **locally closest** to the 'truth" there is a probability > 0.5 that the state ($U$, $V$, $\Theta$, $q_v$) of that member will be **locally closer** to the true values.

No other assumptions used:
linerizations, Gaussianity, or any other observation error probabilities

# Is the Hypothesis valid? (Frontal case)

Let us first consider the **pdf of the decrease in error of a state variable** when the error in **precipitation is decreased by assigning** at each pixel (**grid point**) the ensemble **member with the smallest difference to "true" precipitation**:




Similar results are for the all state variables (U, V, Θ, qv) at all heights, and larger observations windows.

The **hypothesis is verified in a probabilistic sense**: when precipitation of an ensemble member is closer to true value its state variables have a high probability of being closer to truth.

There are pixels of the domain where errors in state variables increase even when precipitation is pushed closer to true values. **The reasons for this are twofold**:

- Although the state variables in the K-F parametrization deterministically determine precipitation at ground the inverse is not true: the relationship of state variables to precipitation (our observations) is stochastic.
- The limited number of ensemble members is a source of noise.

# Definition of state error

*To **represent the state variables in a single parameter***
*we define column state by:*

$$\Psi = \left(\vec{\theta}, \vec{q}_v, \vec{U}, \vec{V}\right)$$

Consequently, the state distance to truth, state error is defined by:

$$\|\Psi - \Psi_t\| = \sqrt{\sum_{z=0}^{Nz} \left(\frac{\Delta\theta_z}{\sigma_{\theta,z}}\right)^2 + \left(\frac{\Delta q_{v,z}}{\sigma_{q_v,z}}\right)^2 + \left(\frac{\Delta U_z}{\sigma_{U,z}}\right)^2 + \left(\frac{\Delta V_z}{\sigma_{V,z}}\right)^2}$$ (Mahalanobis distance)

where $\Delta X_z$ denotes the difference between the background and the truth of any state variable $(\theta, q_v, U, V)$ at height $z$, $\sigma_{X,z}$ is its variance over the ensemble,and $N_z$ the upper model level within the troposphere.

# Is the Hypothesis valid? (Frontal case)

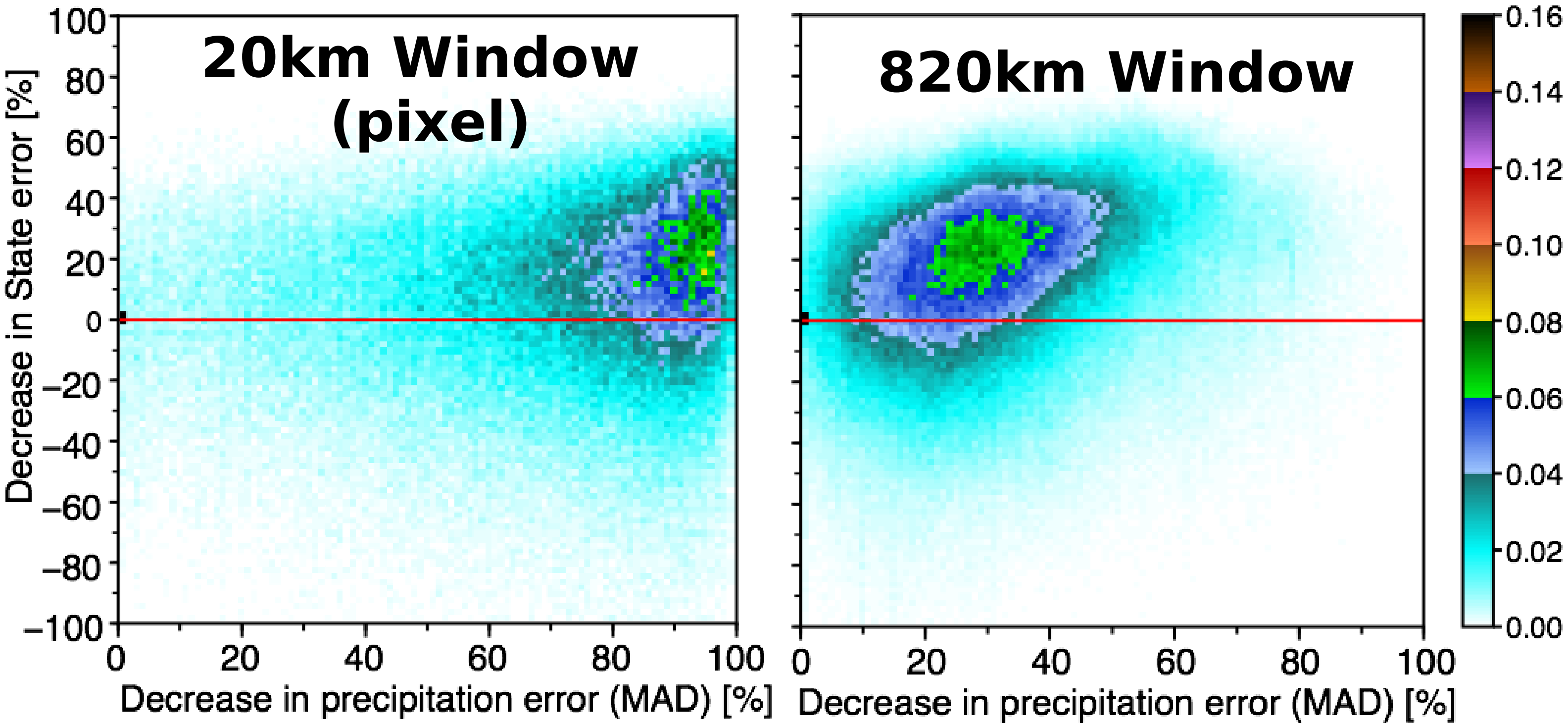
**Decrease in State error:**

$$100\frac{\|MemberState-TruthState\|-\|Frankenstate-TruthState\|}{\|MemberState-TruthState\|}=100\frac{\|\Psi_m-\Psi_t\|-\|\Psi_F-\Psi_t\|}{\|\Psi_m-\Psi_t\|}$$

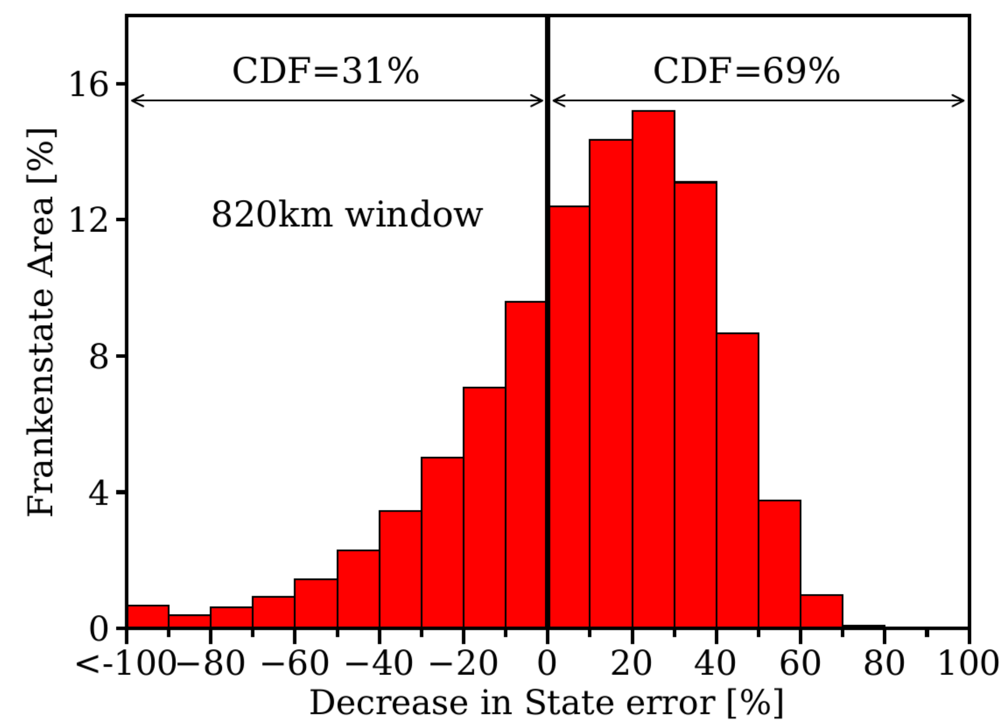
**Decrease in Precipitation error:**

$$100\frac{\|MAD_m-MAD_t\|-\|MAD_F-MAD_t\|}{\|MAD_m-MAD_t\|}$$

**Probability of occurrence as a function of error decrease in precipitation and error in state**




**Frankenstate area in terms of Decrease in state error**



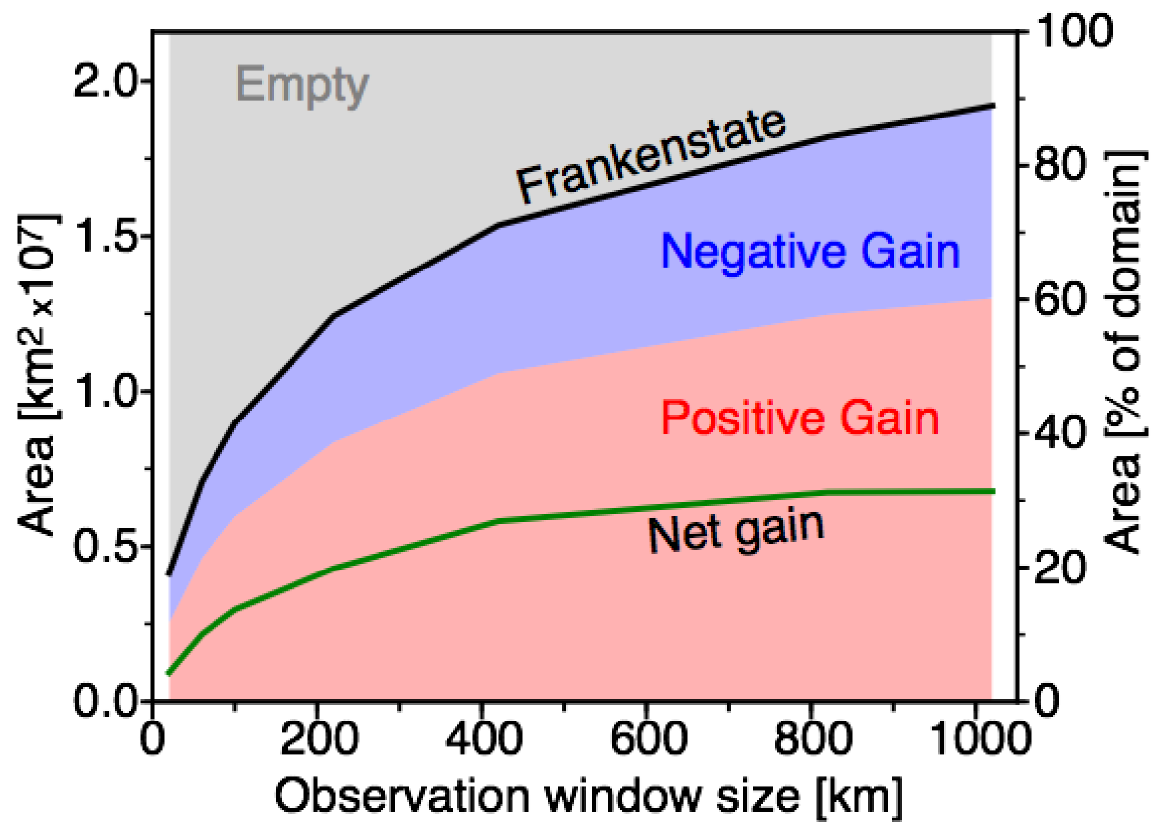

The area of decreased state error (~70% of the Frankenstate) increases with increasing size of observation window (**GOOD!:** the area of positive impact of precipitation information is extended).

Note the negative gain over 30% of the Frankenstate: closeness in precipitation does not always lead to closeness in state.

# Optimal window size

Extending the window size over which MAD is smoothed increases the area over which the Frankenstate can be constructed. Over some of this area we generate negative gain. We can determine the optimal window size by measuring the fraction of area with positive and negative gain for each window size:

The **blue region:** fraction of the domain where the state error increases due to the stochasticity of the relationship between state and precipitation.




Assume that an equal fraction of positive gain area results from stochasticity we get a **net gain area (green curve).**

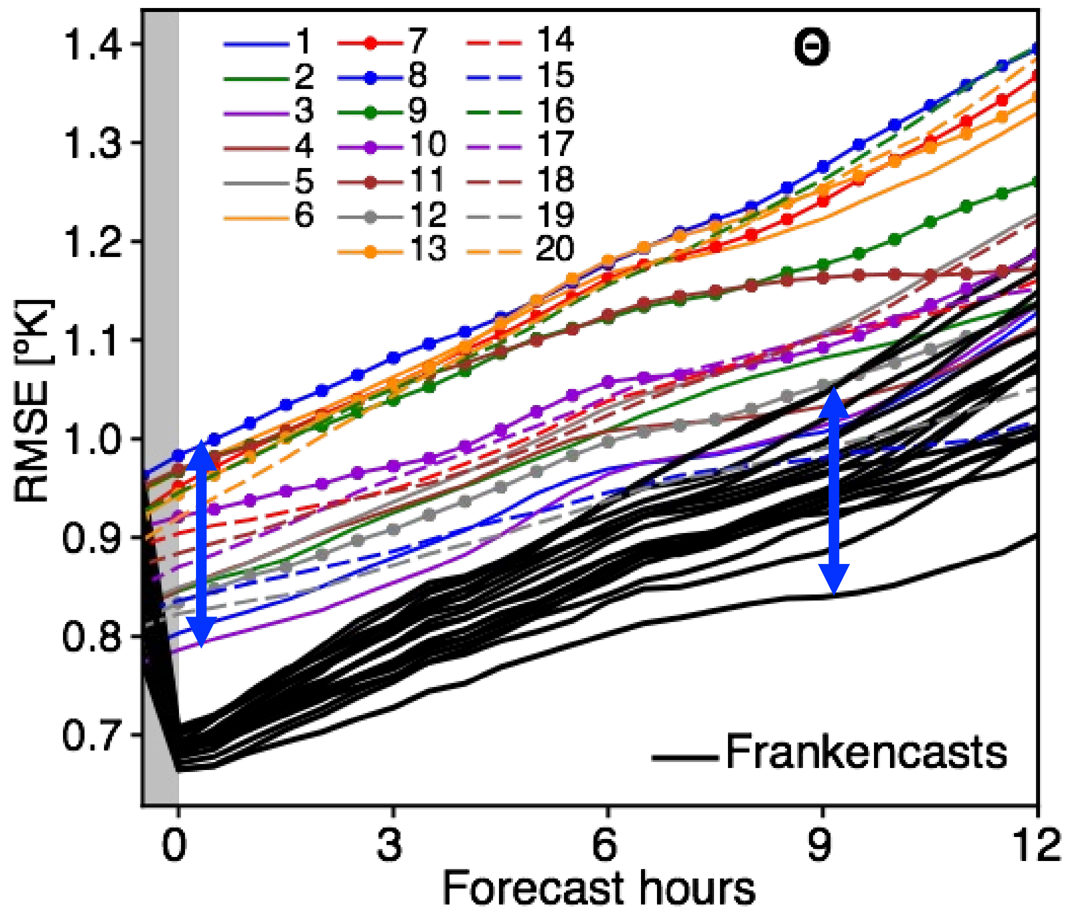
The **black curve** shows the **area over which Frankenstate can be constructed** (an ensemble member can be assigned to a grid point) as a function of the window size.

**We see that up to a window size of 820 km there is some net gain.**

# Forecast

The predominance of positive gain in state ☛ improved forecast of every member.

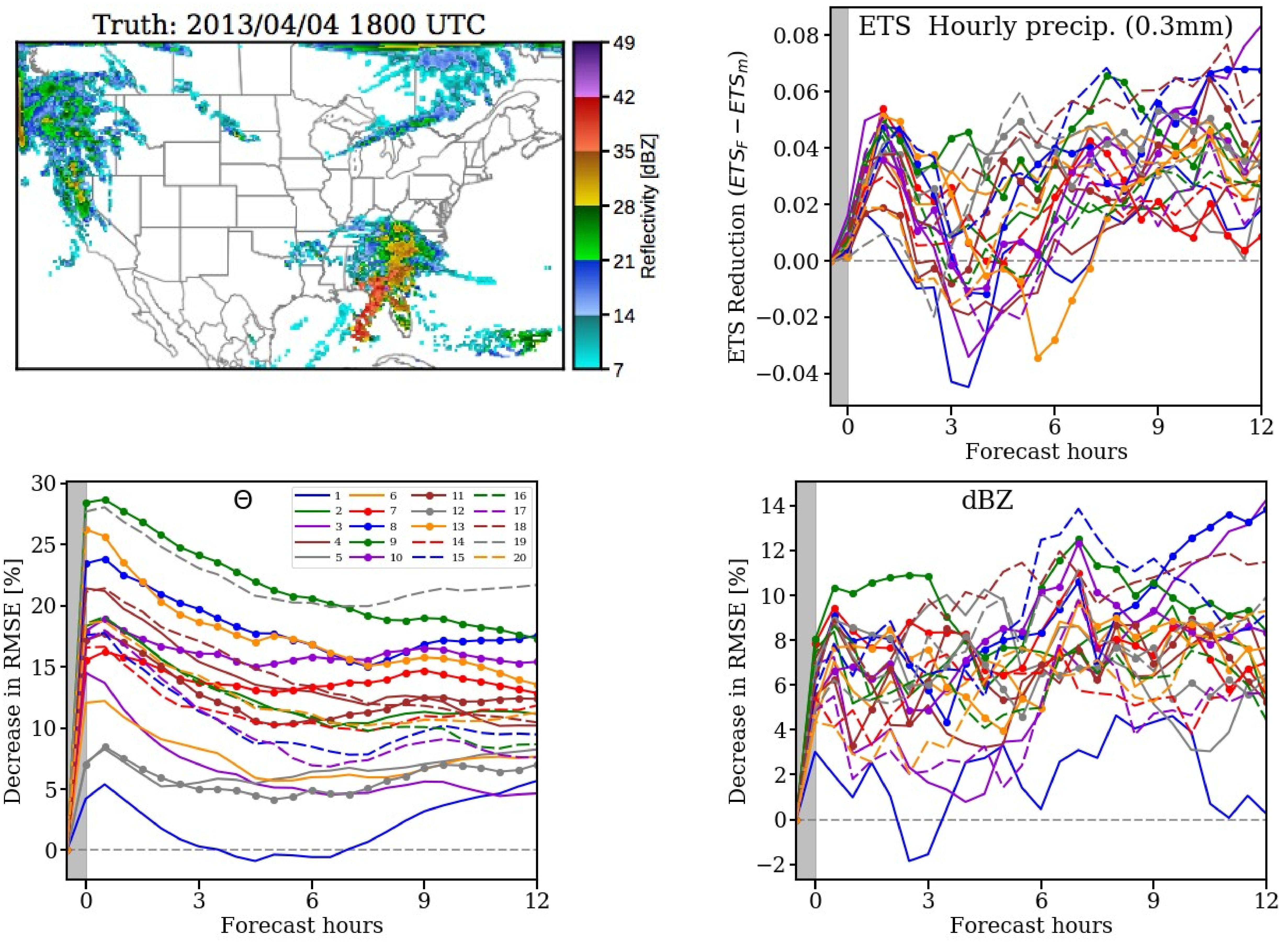
The dynamic initialization **(nudging) introduce perturbations** in the ensemble members.




**These perturbations are “natural**” in the sense that they are produced by

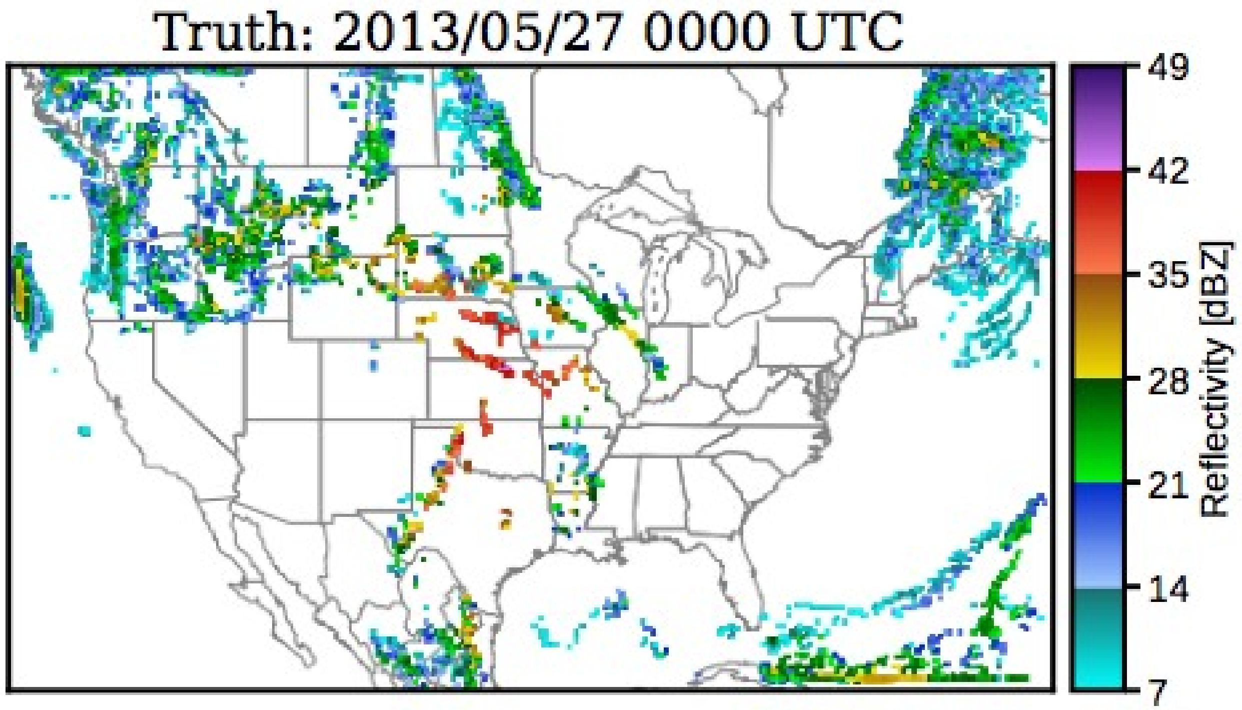
a) the non-deterministic dependence of state variables on precipitation “observations”, and
b) by the uncertainties due to the limited number of ensemble members.

# Forecast of precipitation: Other cases

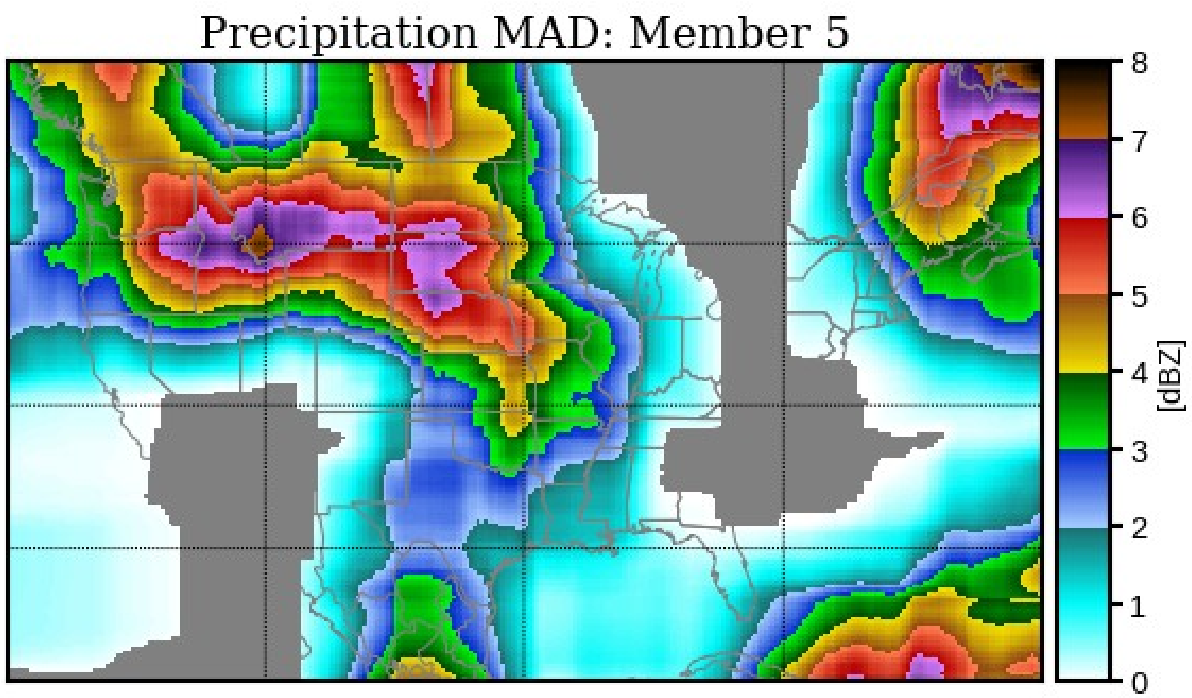


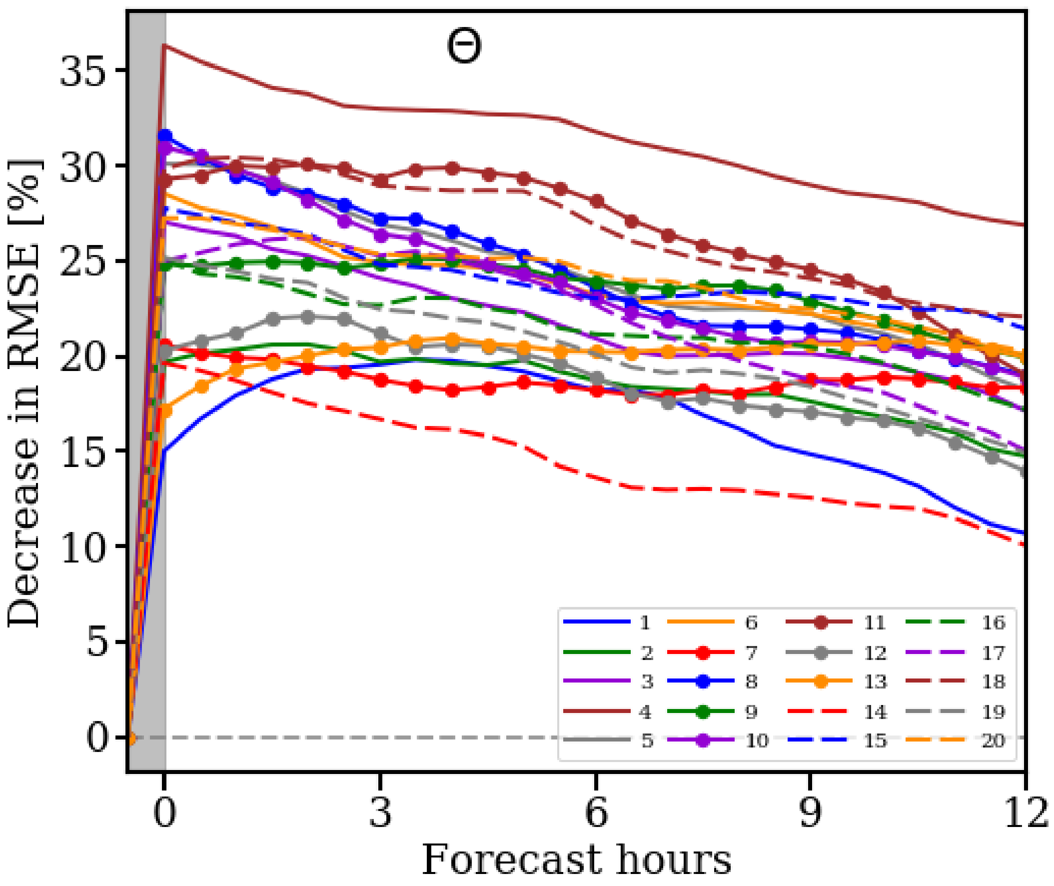
# Forecast of precipitation: Other cases

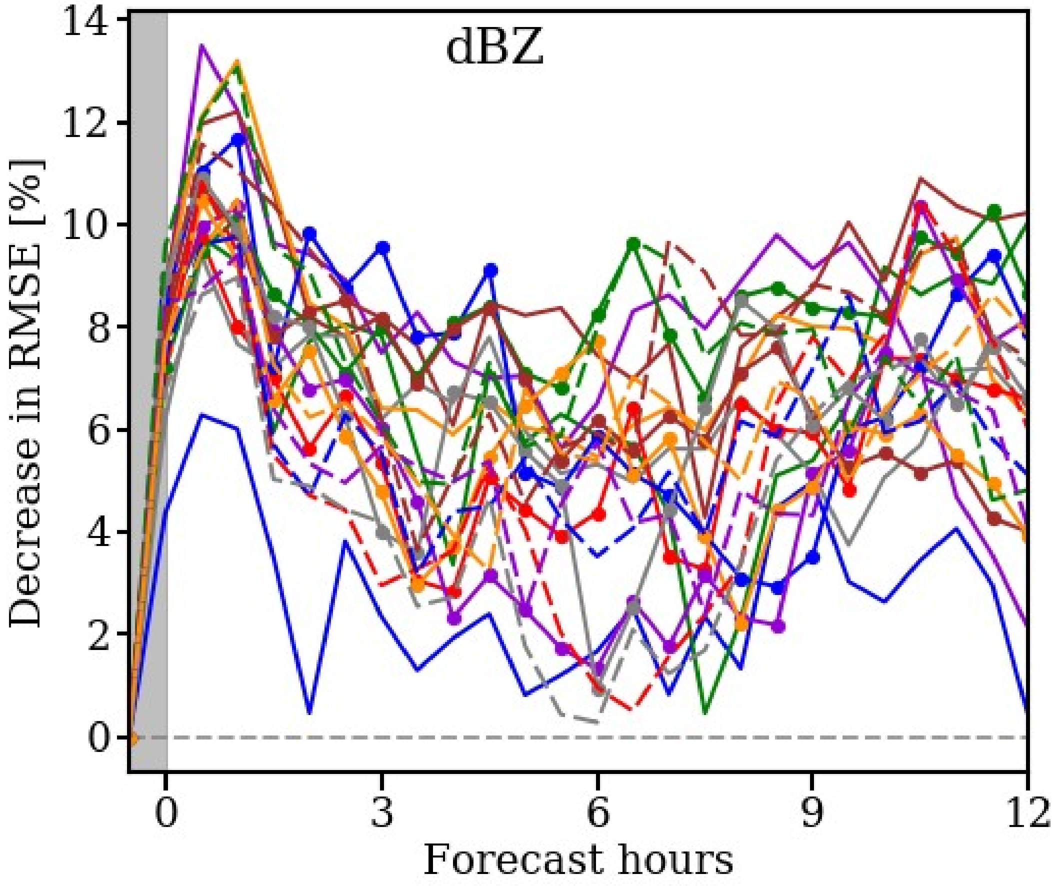
**This are our actual observations**



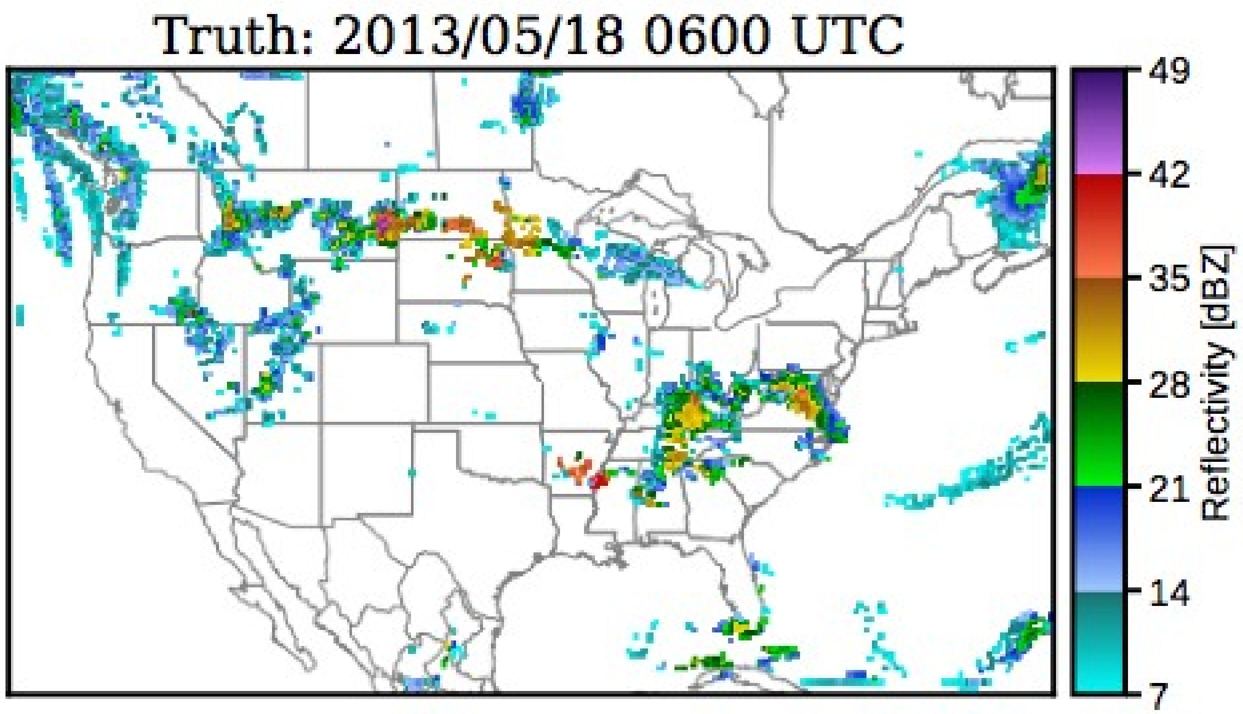

**This is what we use to derive the Frankenstate. Only the large scale component.**

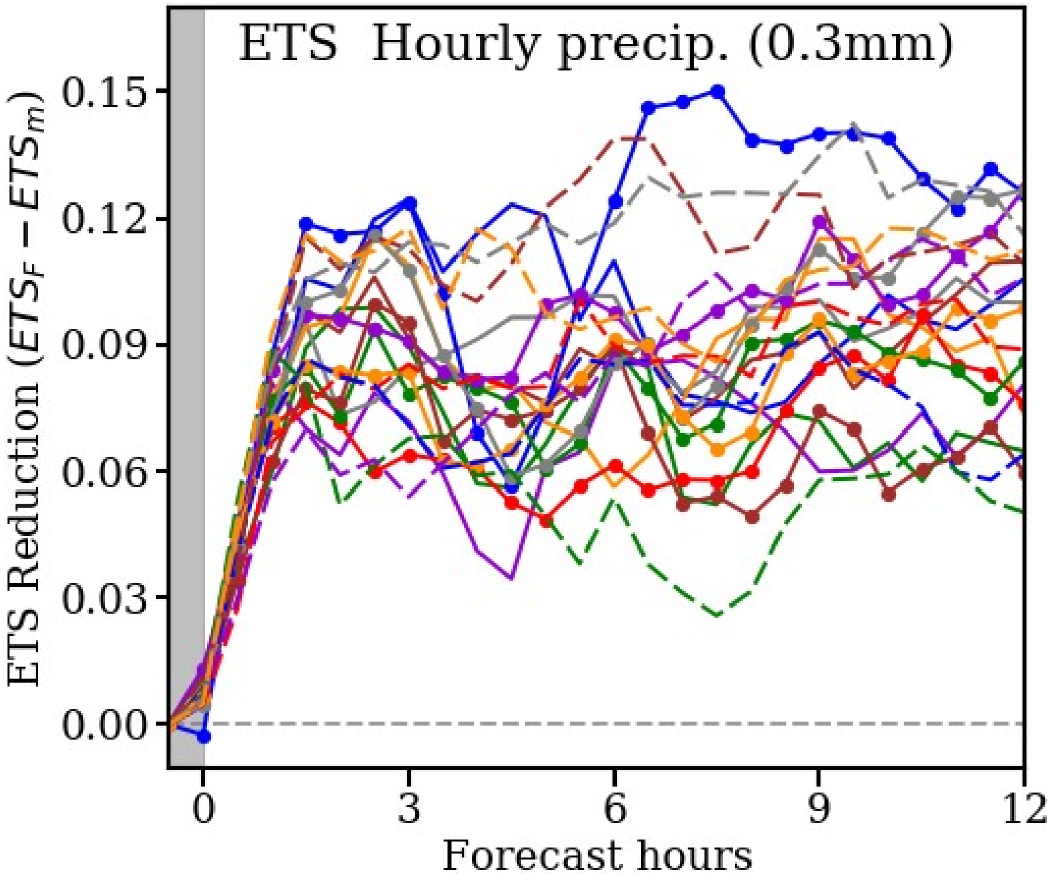


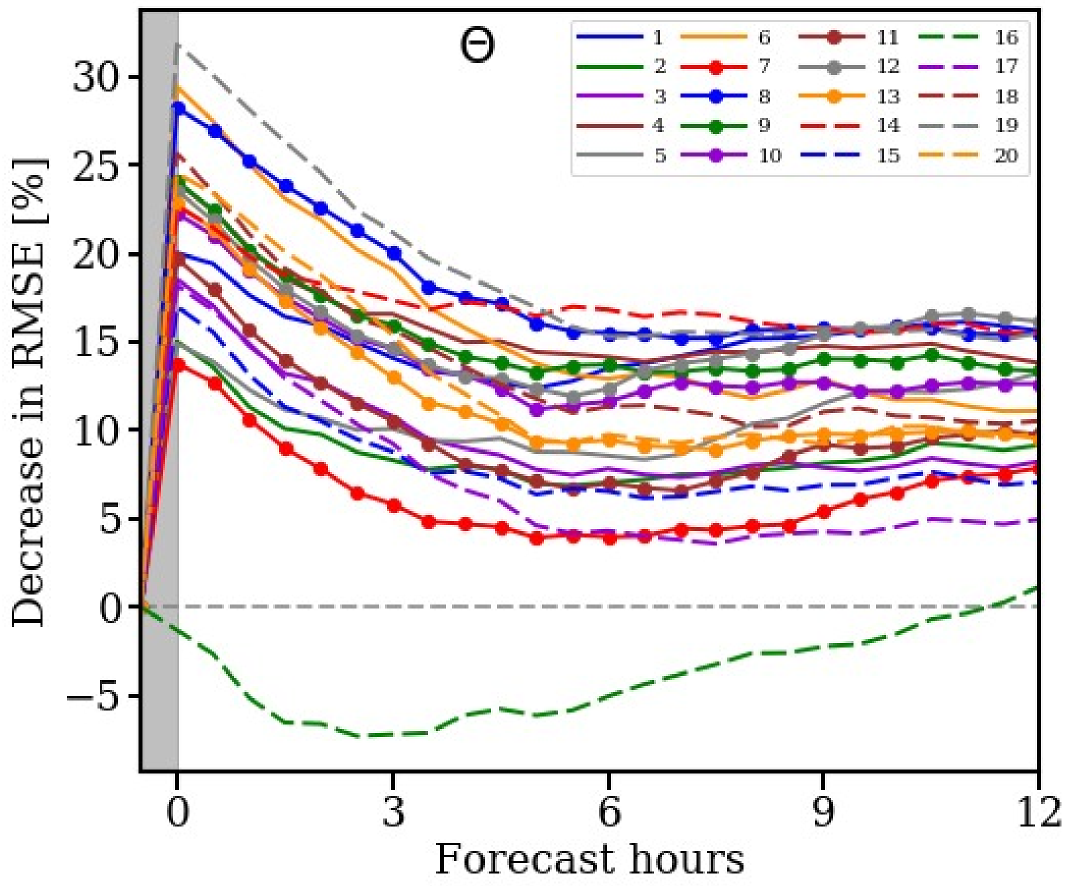



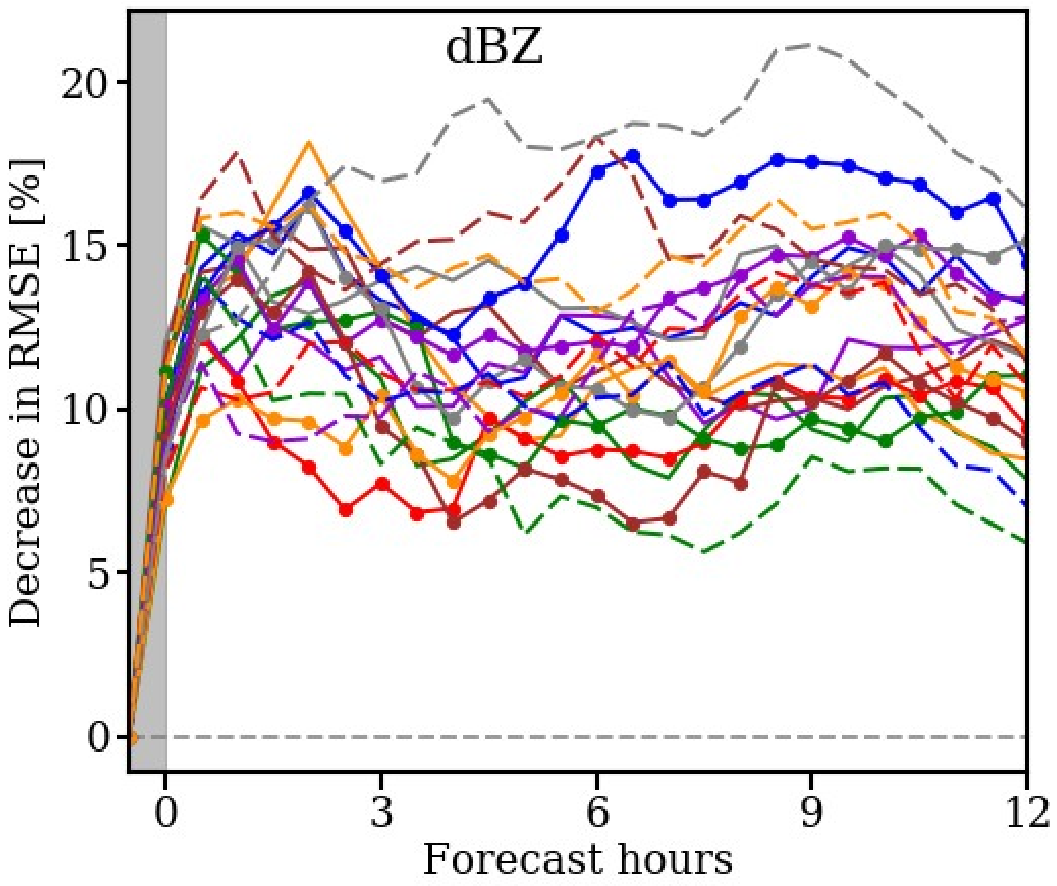



# Forecast of precipitation: Other cases



# Summary of encouraging results:

- ✓ **Direct Ensemble Assimilation is conceptually simple and sound.**

- ✓ **It uses a large scale observation area (two radar-coverage) over which observation errors are smoothed.**

- ✓ **At the same time the smoothing extends the coverage of positive impacts, from 17% of precipitation coverage to 30% of area of net positive impact.**

- ✓ **It introduces perturbations in each member of the ensemble that lead to a spread similar to one produced by the GEFS IC.**

- ✓ **It gives persistent improvement of ~15% in precipitation forecast and better in state variables over the entire domain.**

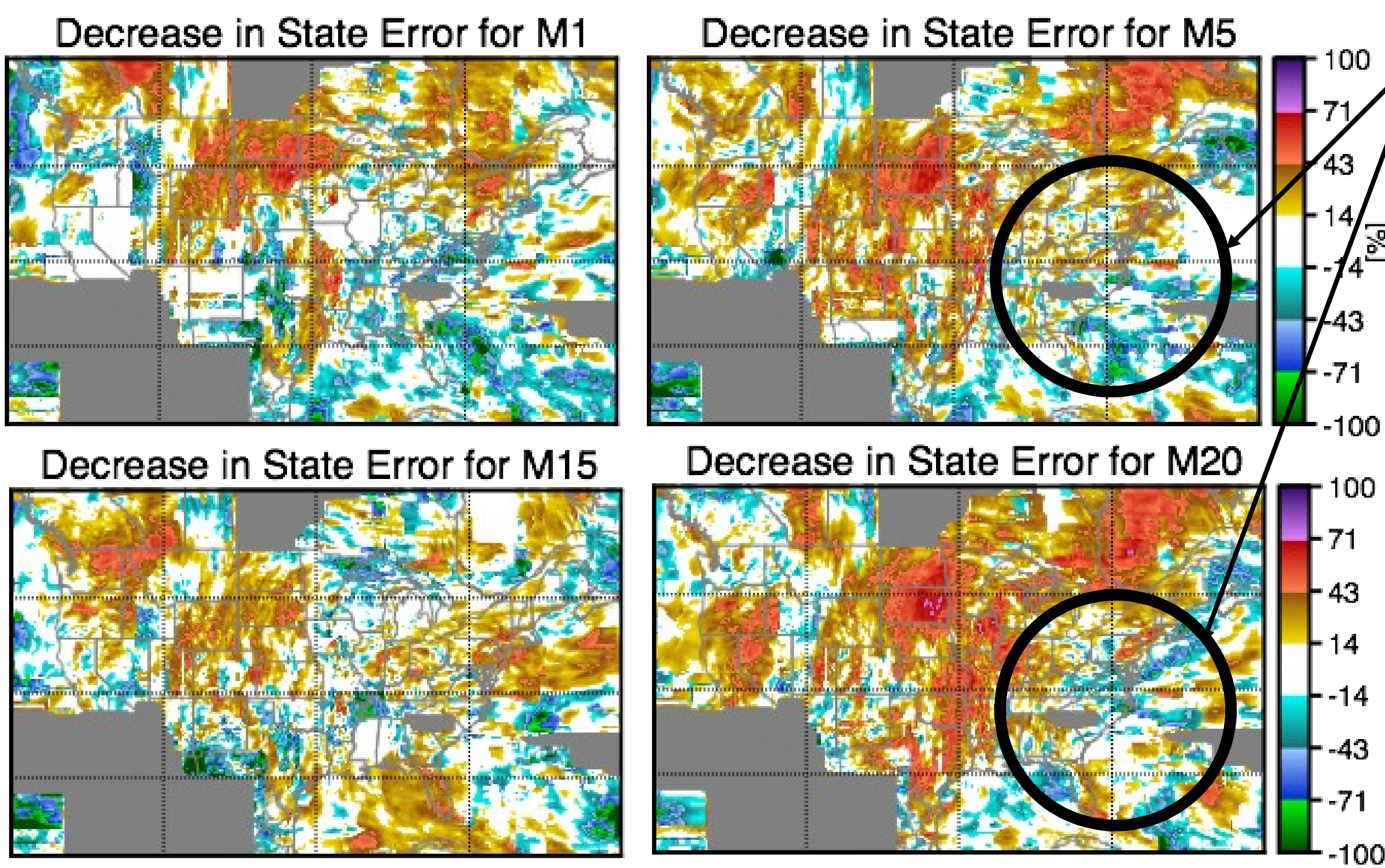
- ✓ **Under the same conditions it outperforms Latent Heat Nudging**

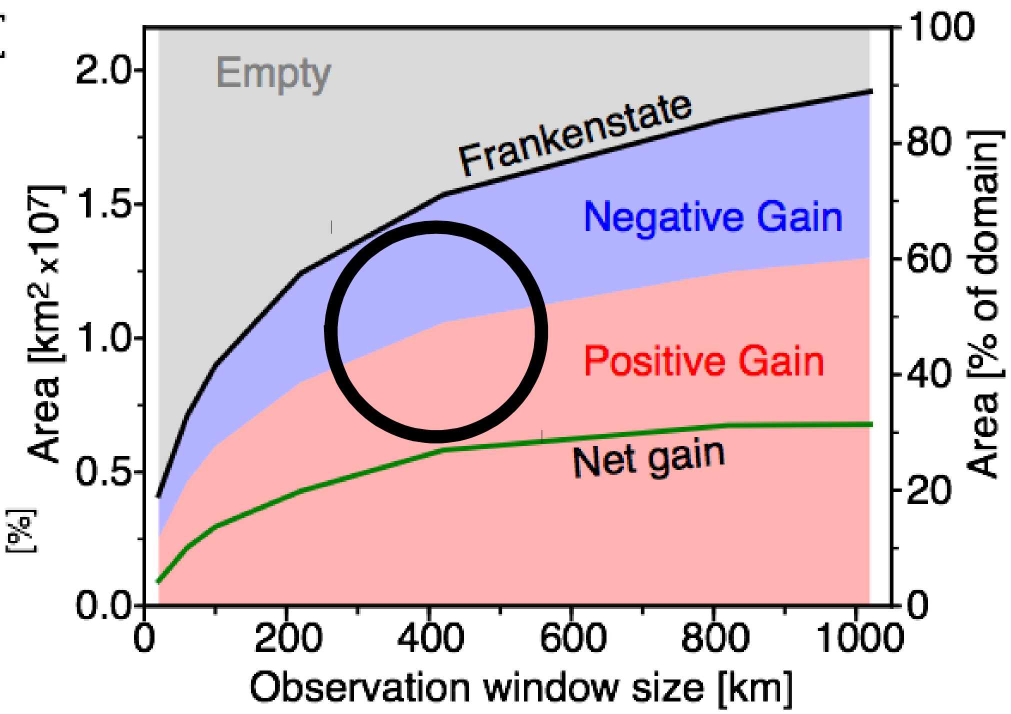
- ✓ **In a real situation (not OSSE) all illusions make break against model errors.**

# The negative gain regions

Members' Decrease in State Error: $100\frac{\|\Psi_m-\Psi_t\|-\|\Psi_F-\Psi_t\|}{\|\Psi_m-\Psi_t\|}$

Note that the positive-negative gains are different for each member.




**The regions of positive and negative gain due to stochasticity are different perturbations for each ensemble member. This, combined with the nudging generates ensemble dispersion during the forecast.**