



# Using error characteristics of the initial conditions to evaluate the likelihood of accurate forecasts

**Craig Schwartz**

NCAR Earth Systems Laboratory

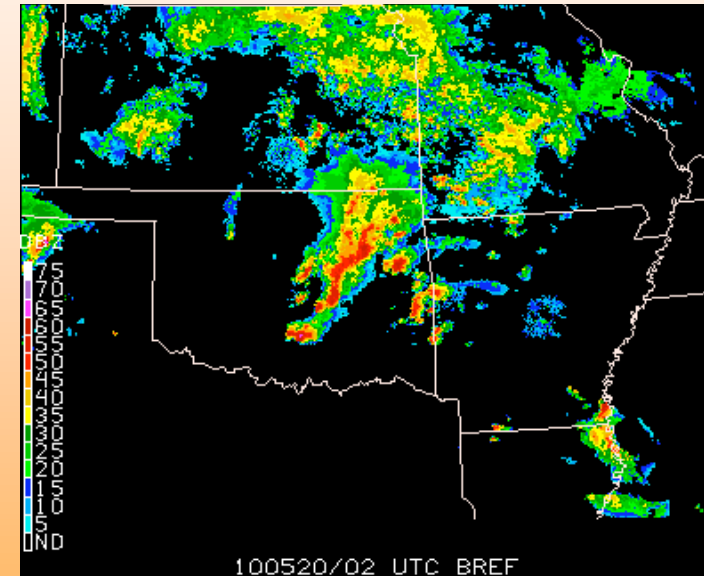
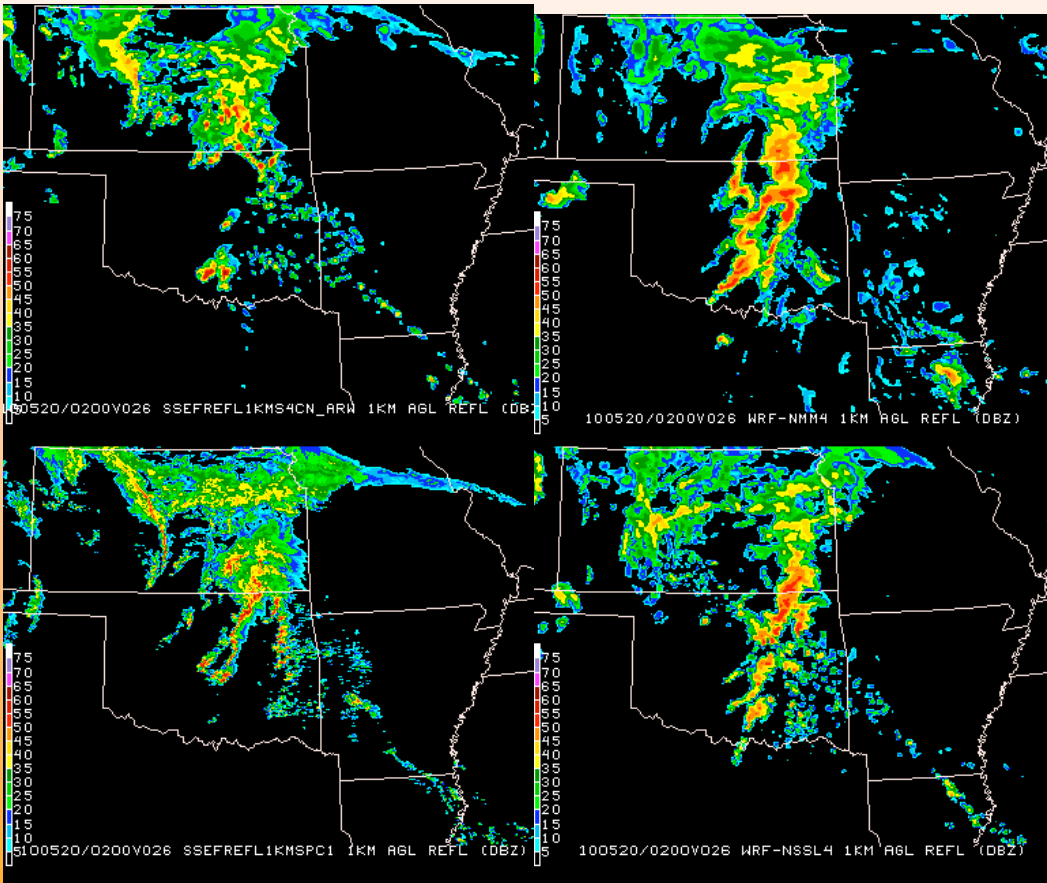
Thanks to Morris Weisman, Wei Wang

NCAR is sponsored by the National Science Foundation

# Motivation

- Sometimes, convection-allowing model (CAM) forecasts are quite good:

All panels valid 0200 UTC May 20, 2010



Observations

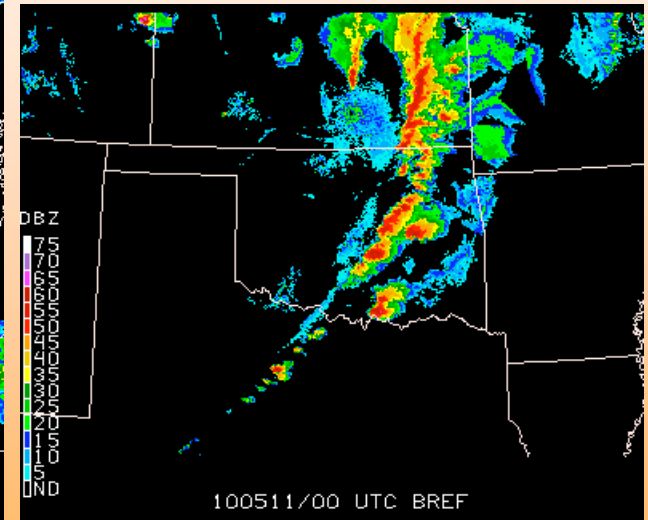
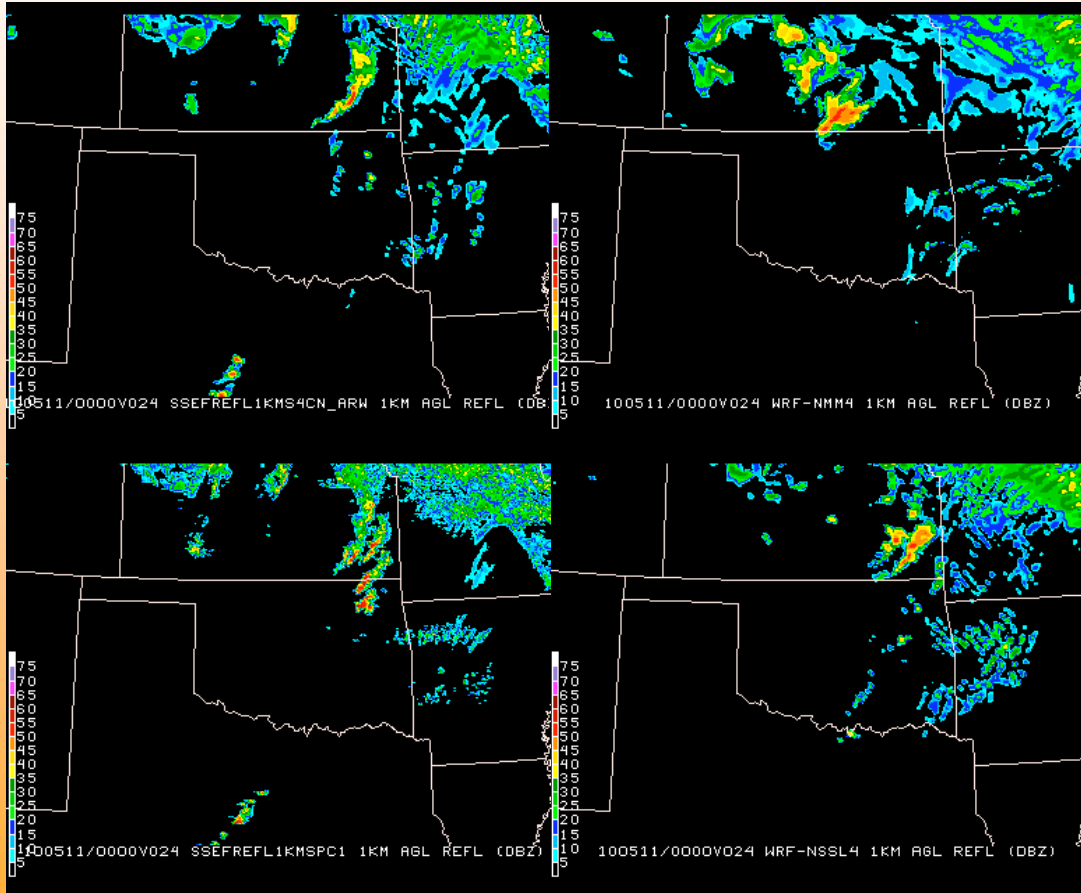
Various CAM forecasts

Graphics courtesy NOAA  
Hazardous Weather Testbed

# Motivation

- Other times, CAM forecasts are off the mark:

All panels valid 0000 UTC May 11, 2010



Observations

Various CAM forecasts:

# Motivation

**Is there a way to determine those days when  
CAM output is more likely to be accurate?**

# Parents and children

- Currently, CAMs are necessarily regional and derive their initial conditions (ICs) and lateral boundary conditions (LBCs) from a coarser resolution, larger domain model (the “parent”)
  - The CAM is the “child”
- Has been demonstrated that the parent’s ICs exert considerable influence on CAM solutions

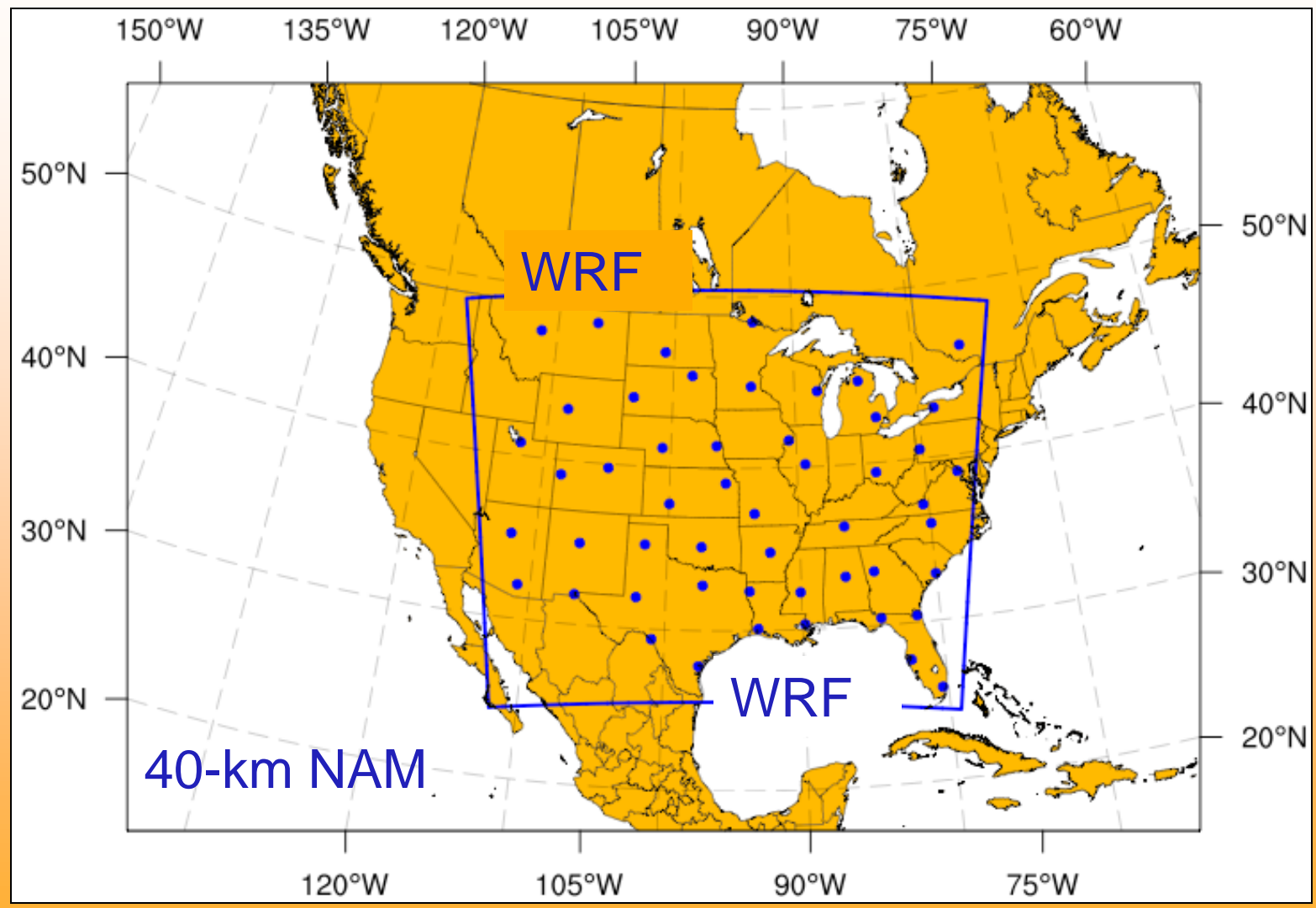


# Experimental Setup

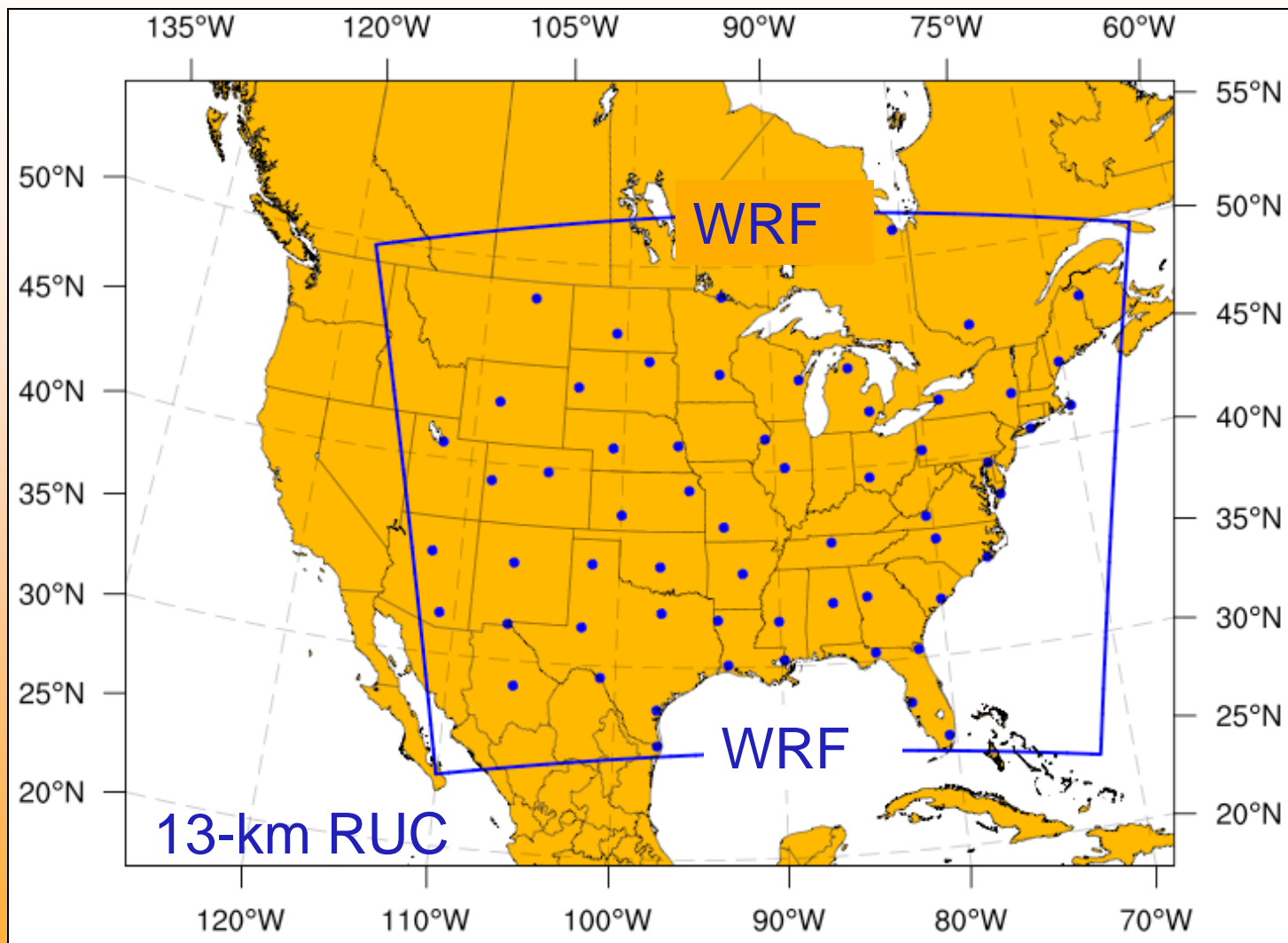
- Examine the parent-child relationship for forecasts from April 26 through June 7 in **both** 2007 and 2010 (43 days)
  - Parent in 2007 was 40-km NAM (LBCs from 40-km NAM)
  - Parent in 2010 was 13-km RUC (GFS LBCs)
- Child was a 3-km WRF-ARW configuration run at NCAR
  - Many differences between the 2007 and 2010 WRF configurations and parent initialization techniques



# 2007 Domains



# 2010 Domains



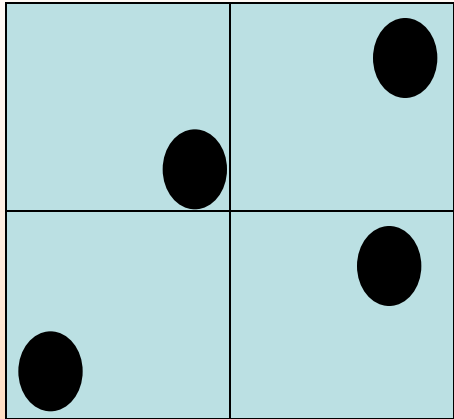
# Methods

- Simple quantitative strategy:
  - Consider soundings as “truth”
  - Verify the parent against soundings at initial time
  - Verify the child against soundings at initial time and forecast times
  - Compute desired error statistics (i.e., RMSE) for both parent and child (separately)
  - Examine the error-relationship between parent and child

# Schematic

Verify parent and child against soundings at the same locations

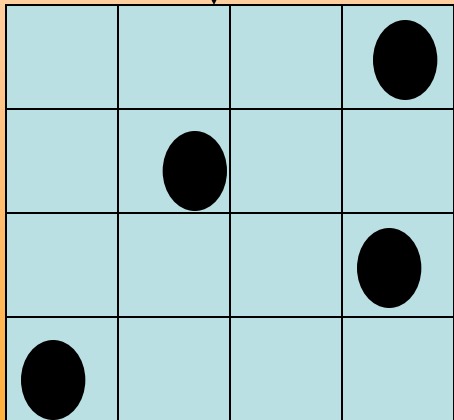
Parent



Examine error relationship

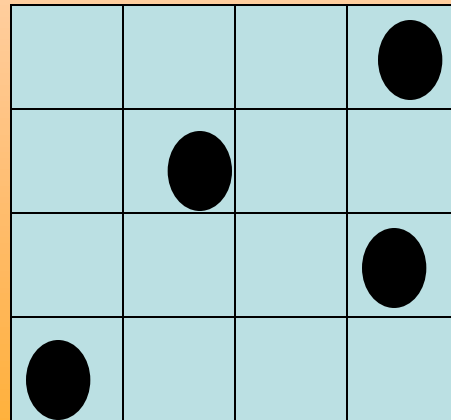


Child



$t = 0$

Examine error relationship



$t = 12 \text{ hrs}$

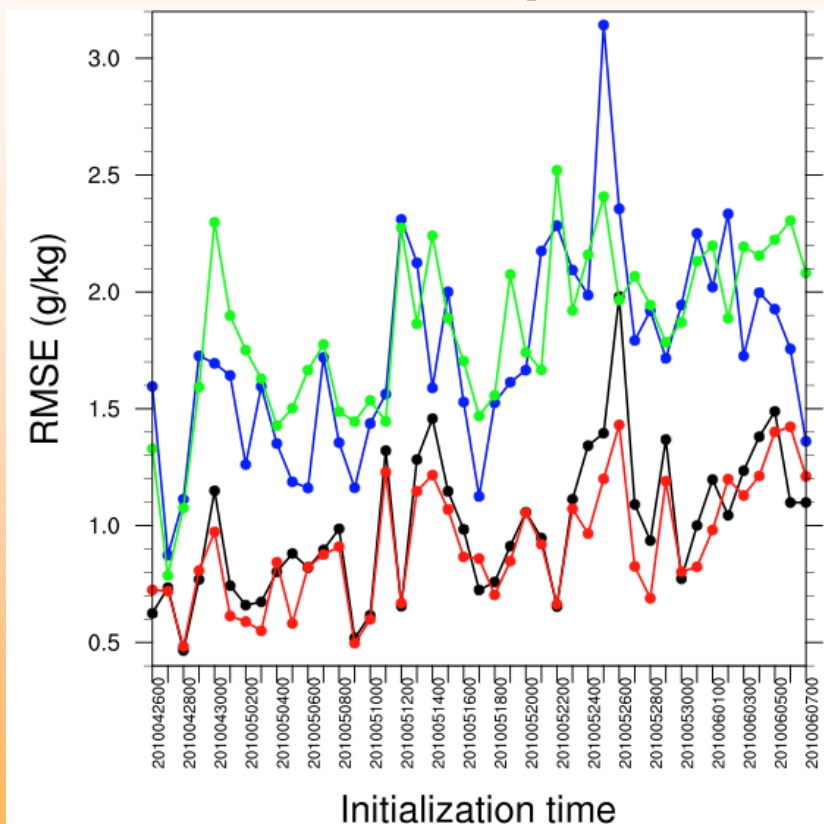
Examine error relationship



$t = 24 \text{ hrs}$

# 850 hPa specific humidity

## 2010: RUC parent



— WRF initialization

— 12-hr WRF fcsts

— parent initialization

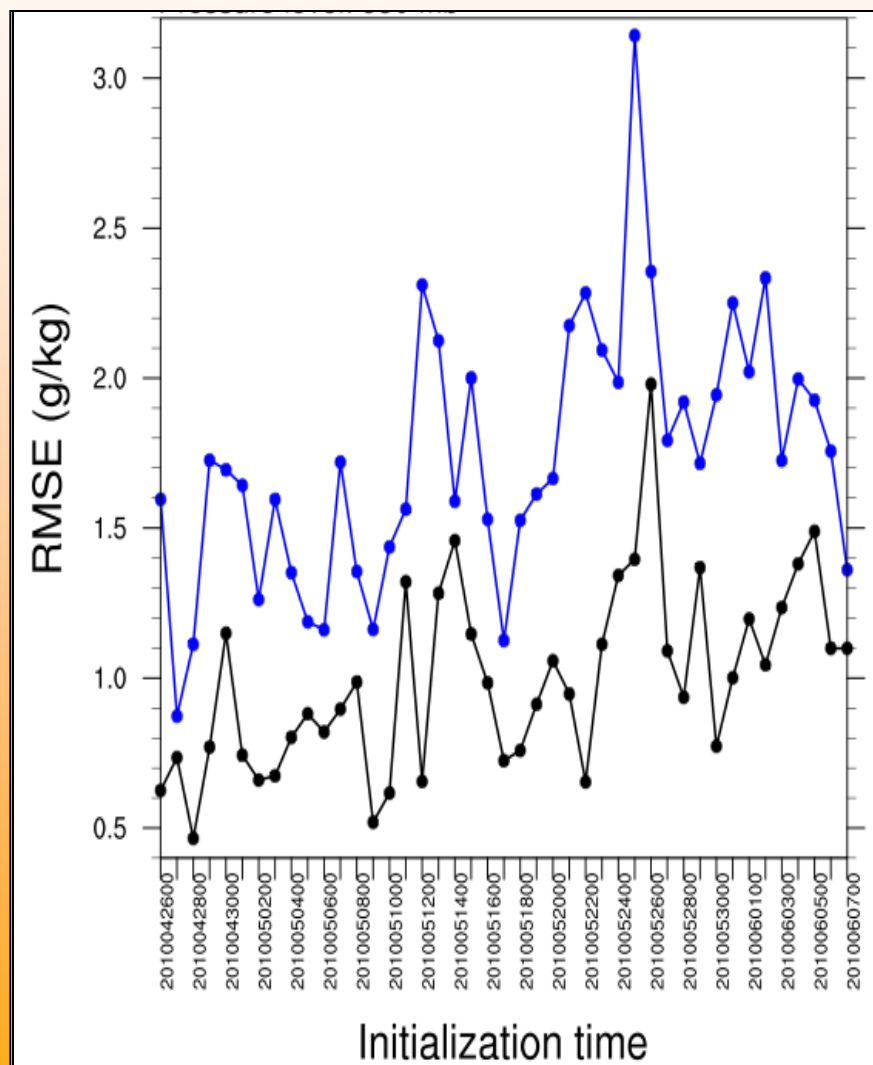
— 24-hr WRF fcsts

# Probabilistic analysis

- Create a joint probability density function (PDF) using error statistics from parent and child

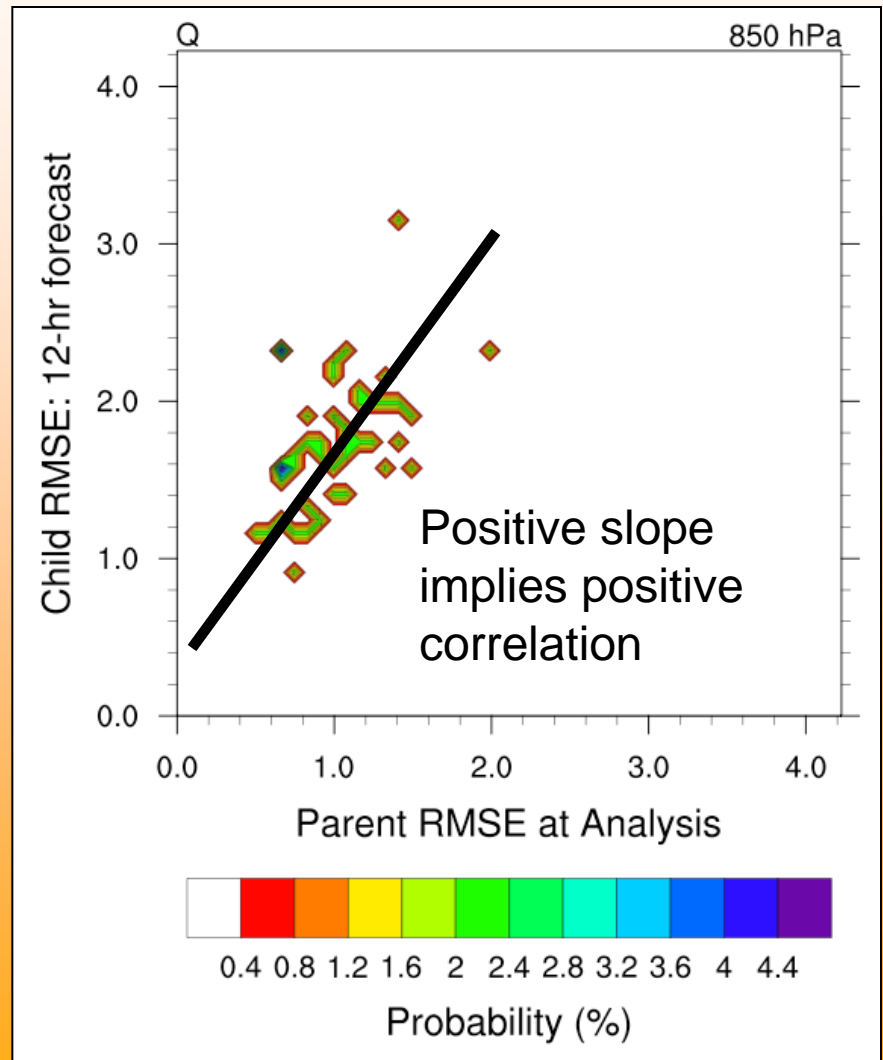
- 850 hPa specific humidity for 2010 dataset

—●— parent initialization  
—●— 12-hr WRF fcsts



# Probabilistic analysis

- Create a joint probability density function (PDF) using error statistics from parent and child
- Joint PDF of 850 hPa specific humidity for 2010 dataset
- X-axis: RUC RMSE at the analysis time
- Y-axis: WRF RMSE from 12-hr forecasts
- Color fill: probabilities (%)



# Probabilistic analysis

- Want to examine aspects of the child's error characteristics given the parent's initial error
- Use conditional probabilities:

$$P(B | A) = \frac{P(B \cap A)}{P(A)}$$

- Events A and B can be defined in any manner



# Model error climatologies

- Necessary to determine reference to which daily CAM forecasts can be compared
- Error climatologies:
  - Parent initializations have a characteristic error
  - Child initializations have a characteristic error
  - 12-hr child forecasts have their own characteristic error
  - etc...
- Use this idea to define events for computing probabilities

# Probabilistic analysis

- Notation/Symbols:

$A_0$  RMSE of the parent at the initial time

$\bar{A}_0$  Mean RMSE of the parent at the initial time

$B_f$  RMSE of the child at the  $f$ -th forecast hour

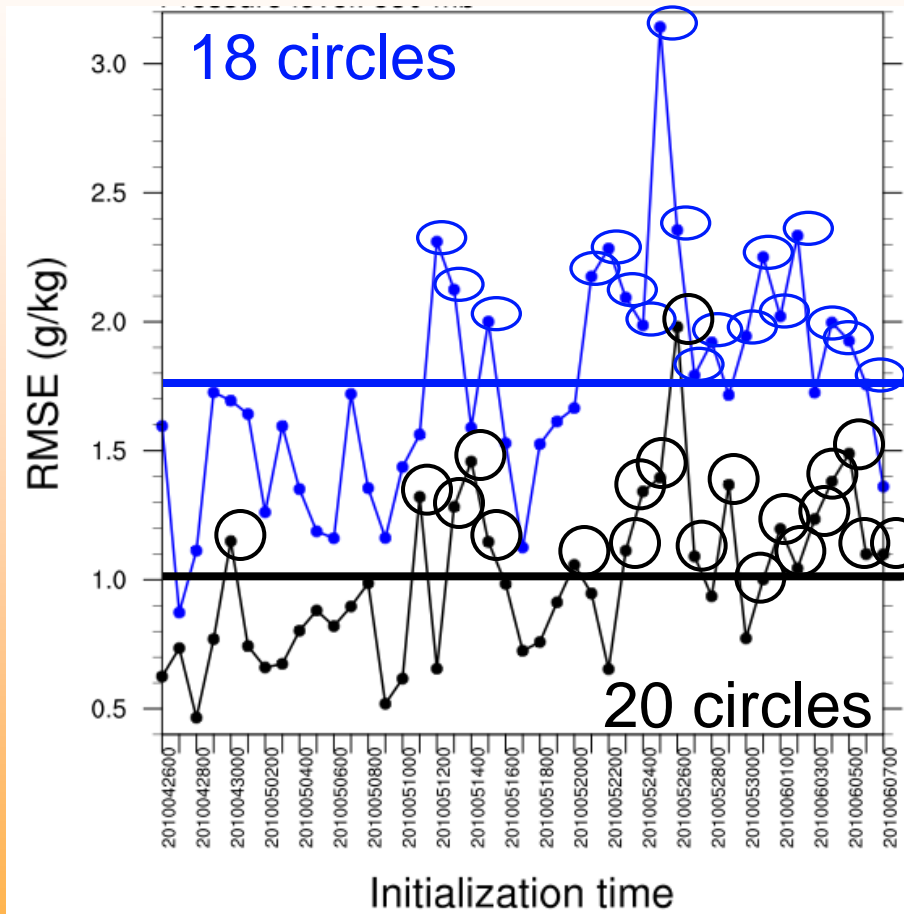
$B_{12}$ : RMSE of the child at the 12-th forecast hour

$\bar{B}_f$  Mean RMSE of the child at the  $f$ -th forecast hour

$$P\left(B_f \geq \bar{B}_f \mid A_0 \geq \bar{A}_0\right)$$

# 850 hPa specific humidity

**N = 43; 12-hr WRF forecast**



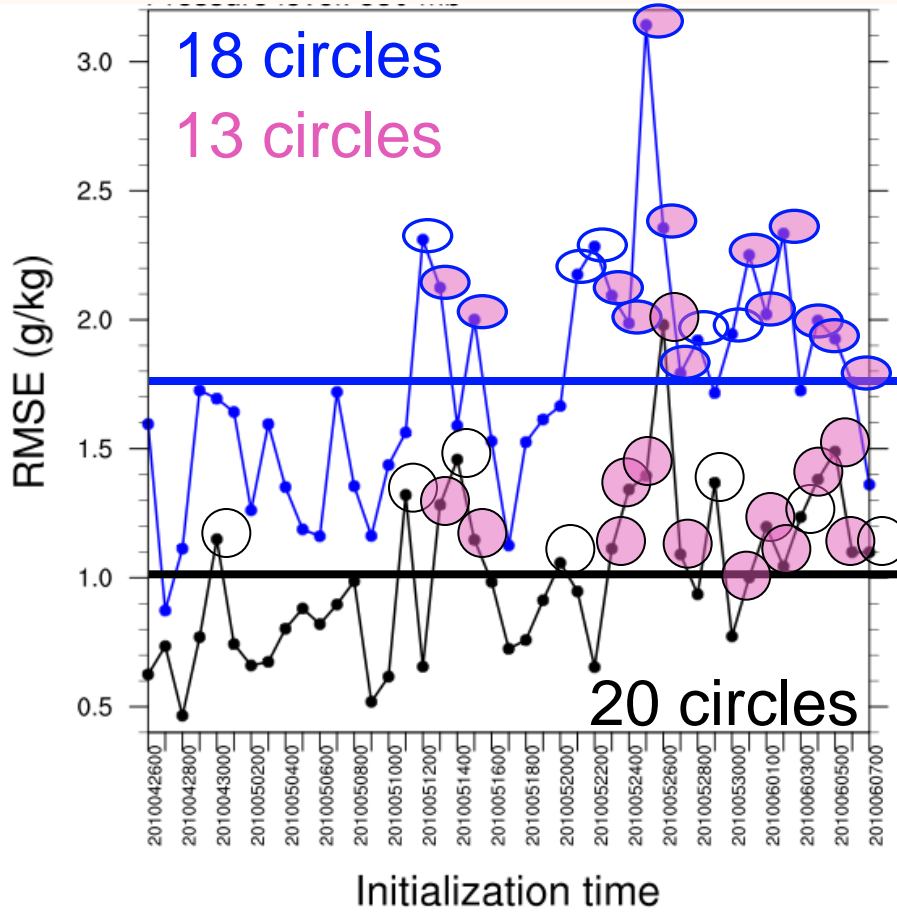
$$P(A_0 \geq \bar{A}_0) = \frac{20}{43} = 0.465$$

$$P(B_{12} \geq \bar{B}_{12}) = \frac{18}{43} = 0.419$$

— parent initialization  
 — 12-hr WRF fcsts

# 850 hPa specific humidity

**N = 43; 12-hr WRF forecast**



— parent initialization  
— 12-hr WRF fcsts

$$P(A_0 \geq \bar{A}_0) = \frac{20}{43} = 0.465$$

$$P(B_{12} \geq \bar{B}_{12}) = \frac{18}{43} = 0.419$$

$$P(B_{12} \geq \bar{B}_{12} \cap A_0 \geq \bar{A}_0) = \frac{13}{43} = 0.302$$

$\bar{B}_{12}$

$\bar{A}_0$

$$P(B_{12} \geq \bar{B}_{12} | A_0 \geq \bar{A}_0) = \frac{P(B_{12} \geq \bar{B}_{12} \cap A_0 \geq \bar{A}_0)}{P(A_0 \geq \bar{A}_0)}$$

$$P(B_{12} \geq \bar{B}_{12} | A_0 \geq \bar{A}_0) = \frac{0.302}{0.465} = 0.65$$



$$P\left(B_0 \geq \bar{B}_0 \mid A_0 \geq \bar{A}_0\right) \mathbf{f} = \mathbf{0} \text{ (child)}$$

Pressure Levels (hPa)

		850	700	500	300	
U wind component	U	2007	73.7	94.4	93.3	85.0
		2010	90.9	92.3	92.3	86.7
V wind component	V	2007	90.5	76.2	80.0	81.8
		2010	84.6	100	100	81.8
Temperature	T	2007	75.0	93.3	90.0	76.9
		2010	80.0	78.9	81.3	76.5
Specific Humidity	Q	2007	68.2	66.7	69.6	85.0
		2010	90.0	95.0	84.0	100

Probabilities (%)



$$P\left(B_{12} \geq \bar{B}_{12} \mid A_0 \geq \bar{A}_0\right) \mathbf{f} = \mathbf{12} \text{ (child)}$$

Pressure Levels (hPa)

		850	700	500	300	
U wind component	U	2007	50.0	44.4	26.7	35.0
		2010				
V wind component	V	2007	42.9	57.1	50.0	36.4
		2010				
Temperature	T	2007	50.0	53.3	45.0	58.8
		2010				
Specific Humidity	Q	2007	40.9	58.3	65.2	68.4
		2010				

Probabilities (%)



$$P\left(B_{12} \geq \bar{B}_{12} \mid A_0 \geq \bar{A}_0\right) \mathbf{f} = \mathbf{12} \text{ (child)}$$

Pressure Levels (hPa)

		850	700	500	300	
U wind component	U	2007	50.0	44.4	26.7	35.0
		2010	81.8	61.5	46.2	40.0
V wind component	V	2007	42.9	57.1	50.0	36.4
		2010	69.2	56.3	40.0	54.5
Temperature	T	2007	50.0	53.3	45.0	58.8
		2010	73.3	63.2	62.5	58.8
Specific Humidity	Q	2007	40.9	58.3	65.2	68.4
		2010	65.0	65.0	72.0	85.7

Probabilities (%)

# Interpretation

- Influence of ICs declines with time
- Accuracy of the parent's ICs associated more strongly with 12-hr child forecast errors in 2010
  - Information from the parent's ICs appear to be “retained” for longer in 2010
  - However, probabilities not always substantially greater than 50%
- Changes/advances in WRF model, initialization techniques, and parent resolution likely the biggest contributors to the differences between 2007 and 2010

# Final Thoughts



- Results (conditional probabilities) corroborate previous work
  - Close connection between the parent and child
  - Diminishing influence of initial conditions with forecast time
- This work concerns aspects of forecast accuracy only, not value
- Can use the child's ICs in place of the parent's ICs: useful for examining the impact of regional data assimilation
- Can likelihood of child precipitation accuracy be inferred from the goodness of the ICs?
- Need longer datasets to better define the error climatologies
- Is it desirable that accuracy of the child's forecasts be related to the accuracy of the parent's ICs?

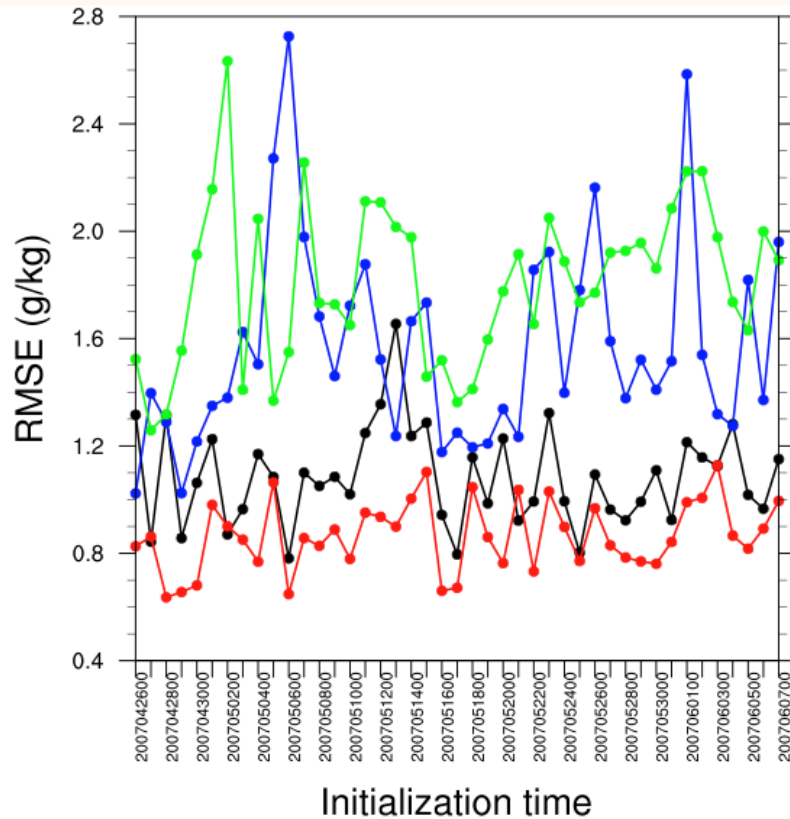




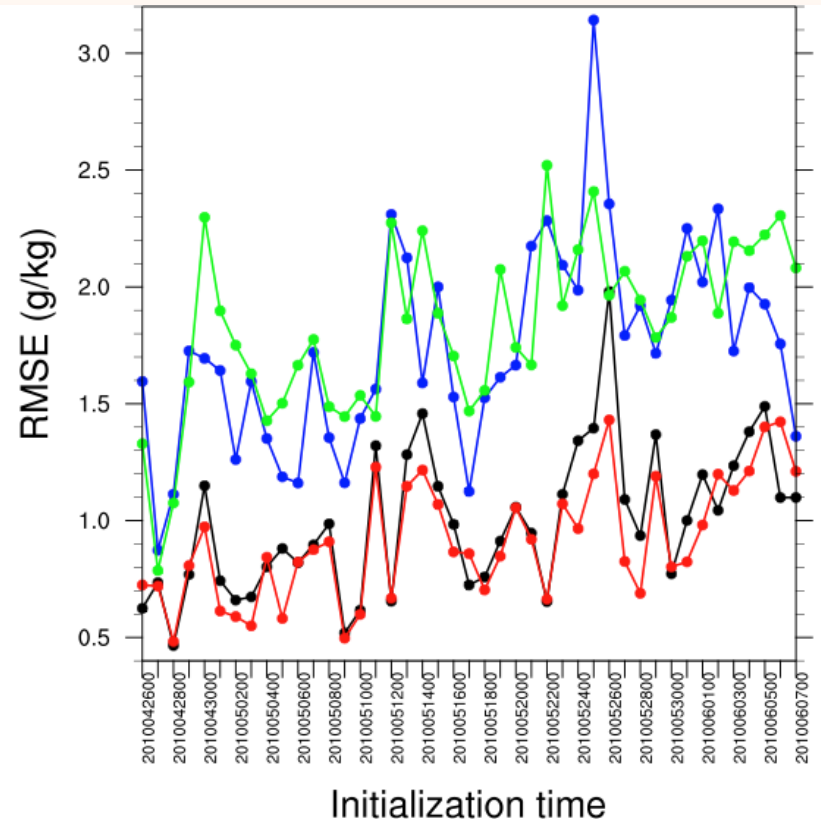


# 850 hPa specific humidity

## 2007: NAM parent



## 2010: RUC parent



— WRF initialization

— parent initialization

— 12-hr WRF fcsts

— 24-hr WRF fcsts

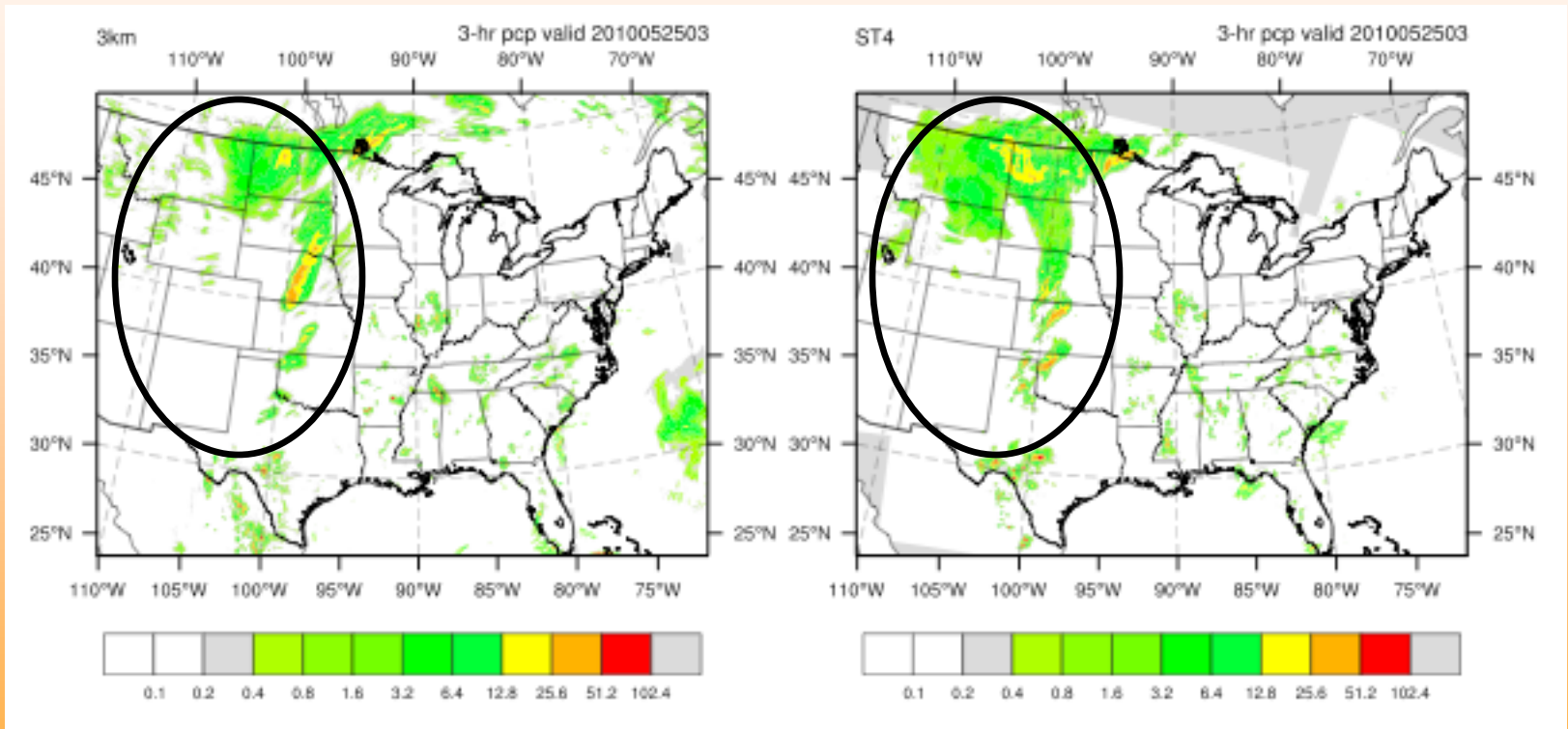


# Precipitation Forecasts

- When the parent more accurately initializes mass and velocity fields, should the child's precipitation forecasts be more accurate?
  - Seems like they should, but complexities arise regarding precipitation forecasts (e.g., microphysics parameterizations)
- Perhaps a subjective evaluation is the best way to examine the link (if any) between the child's precipitation forecasts and parent's ICs

# □ Precipitation Forecast

- 3-hr forecast of 3-hr accumulated precipitation initialized 0000 UTC May 25, 2010

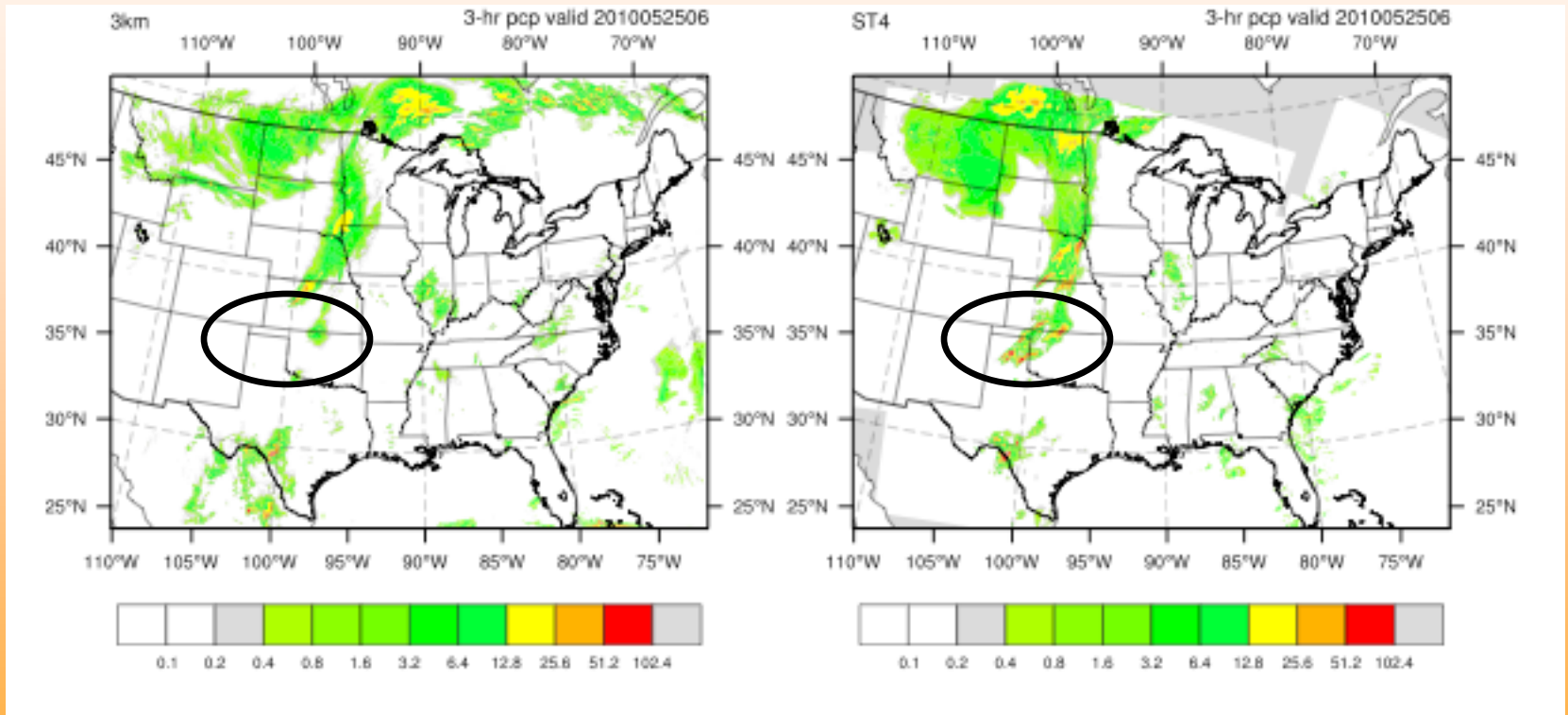


WRF

“Obs”

# □ Precipitation Forecast

- 6-hr forecast of 3-hr accumulated precipitation initialized 0000 UTC May 25, 2010

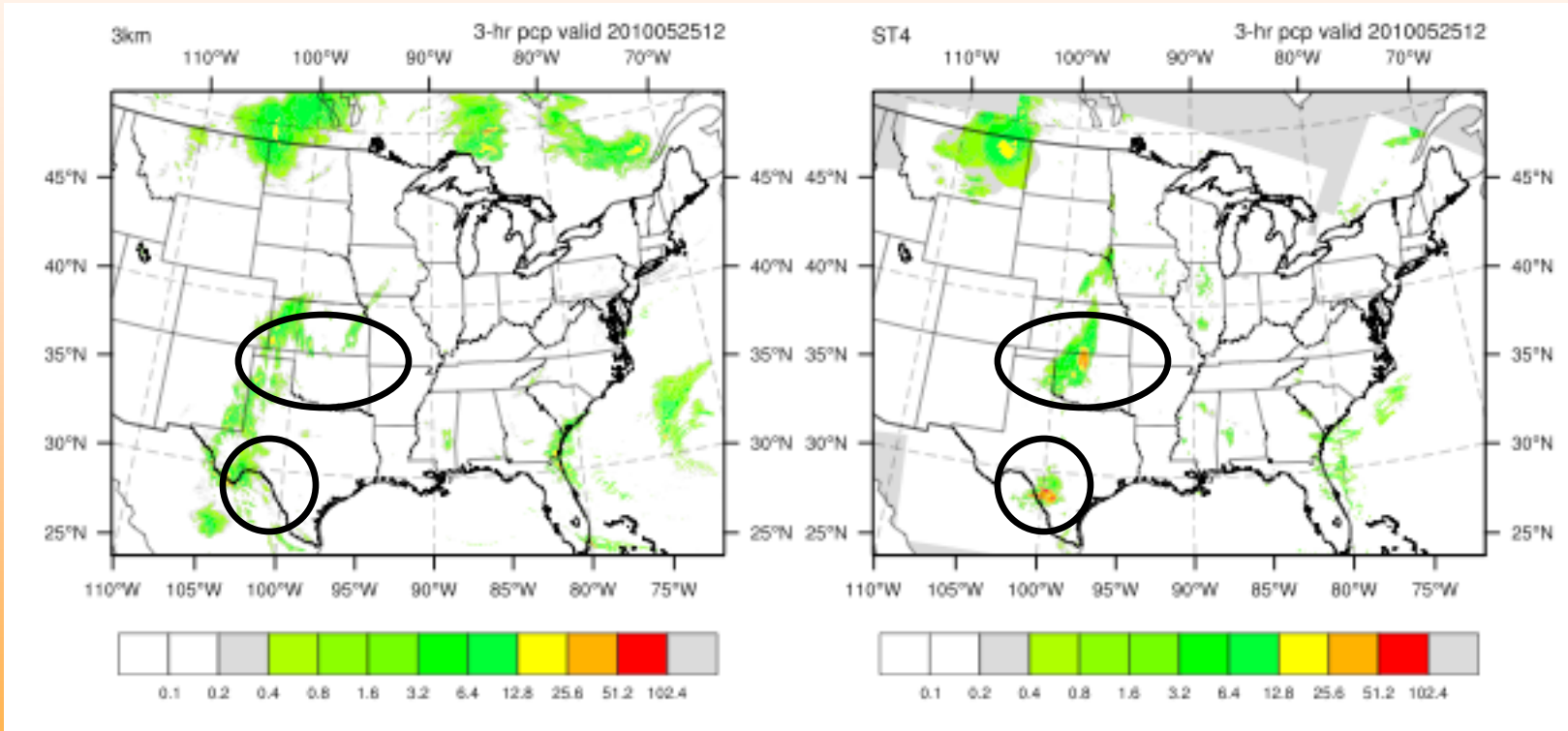


WRF

“Obs”

# □ Precipitation Forecast

- 12-hr forecast of 3-hr accumulated precipitation initialized 0000 UTC May 25, 2010

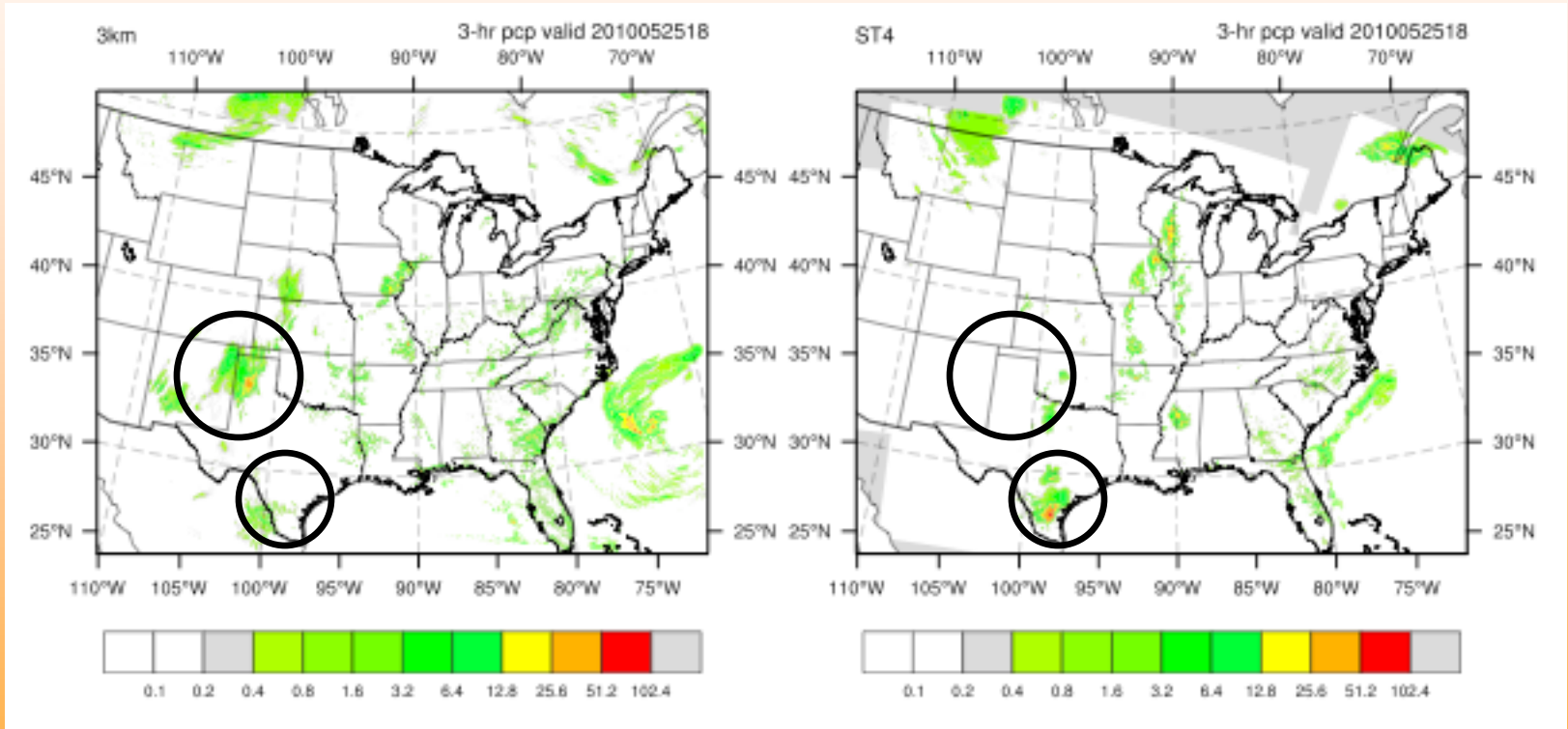


WRF

“Obs”

# □ Precipitation Forecast

- 18-hr forecast of 3-hr accumulated precipitation initialized 0000 UTC May 25, 2010

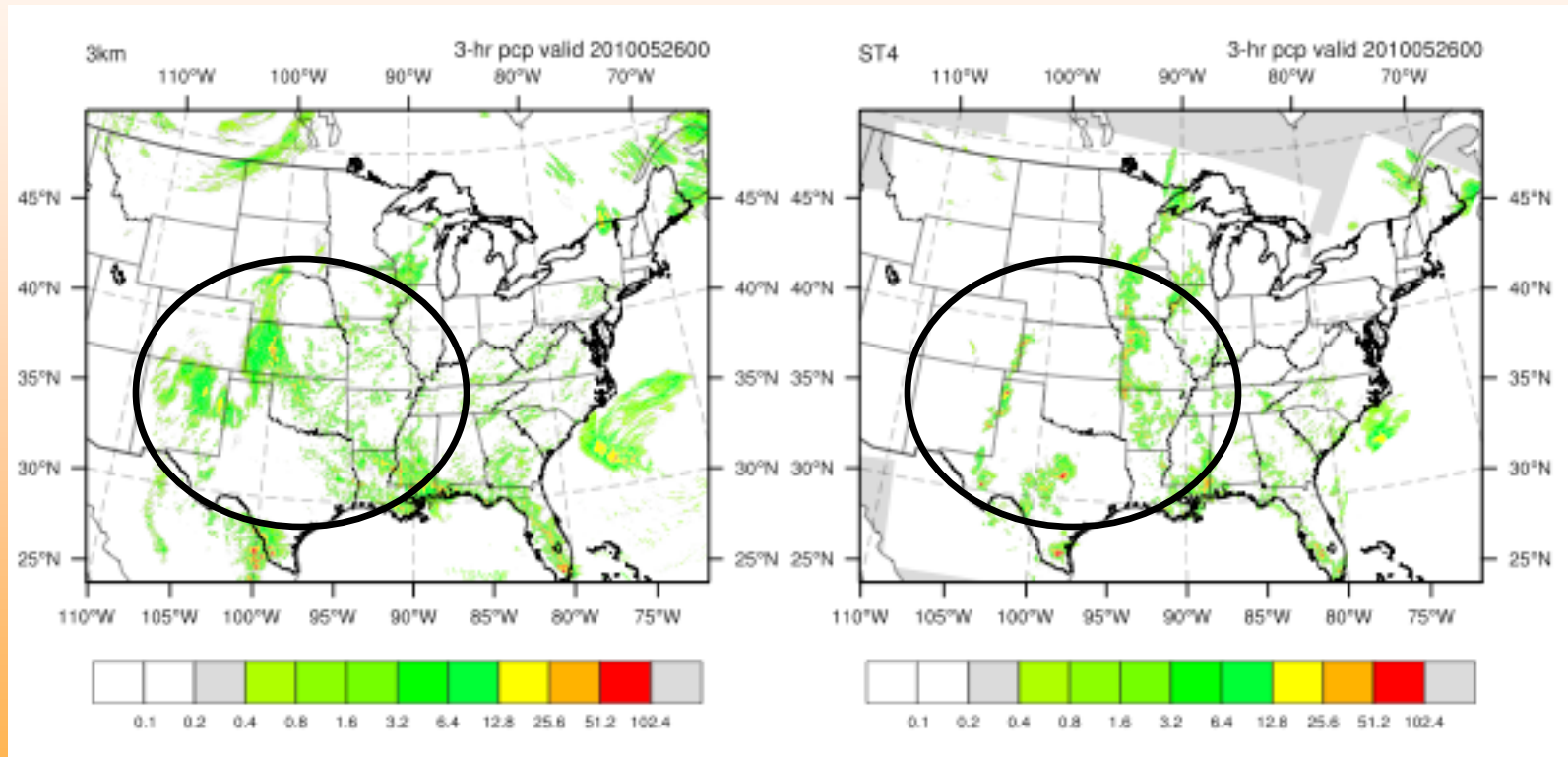


WRF

“Obs”

# □ Precipitation Forecast

- 24-hr forecast of 3-hr accumulated precipitation initialized 0000 UTC May 25, 2010



WRF

“Obs”



# □ Future Directions

- About accuracy...NOT value

**Is there a way to determine those days when  
CAM output is more likely to be accurate?**









# Revisiting Precipitation

- Subjectively noted that patterns of CAM precipitation imitate their parents
  - Locations of CAM precipitation entities are linked closely to their parent (though smaller-scale features present in CAM output are valuable)
- Can this subjective impression be confirmed quantitatively?
- RegridDED precipitation to a 20-km grid for comparison using standard metrics

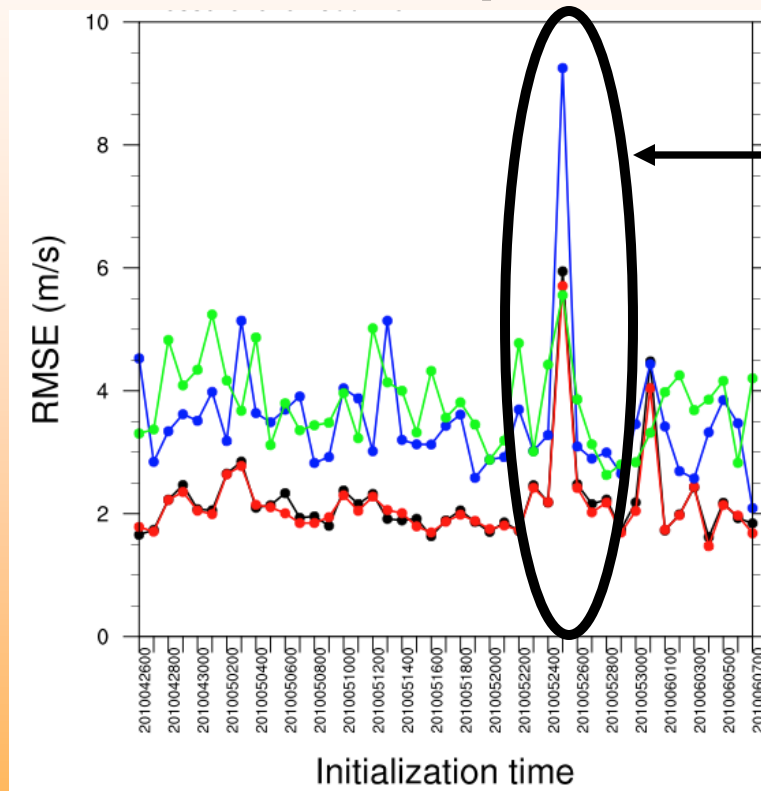
# WRF Differences

	2007	2010
Parent Resolution	40-km	13-km
WRF Model Version	2.2???	3.2
Parent data assimilation cycle	3-hrs	1-hr
Radar data in Ics?	No	Yes
Physical Parameterizations	Lots of differences	



# 500 mb zonal wind

## 2010: RUC parent



May 25, 2010

— WRF initialization

— 12-hr WRF fcsts

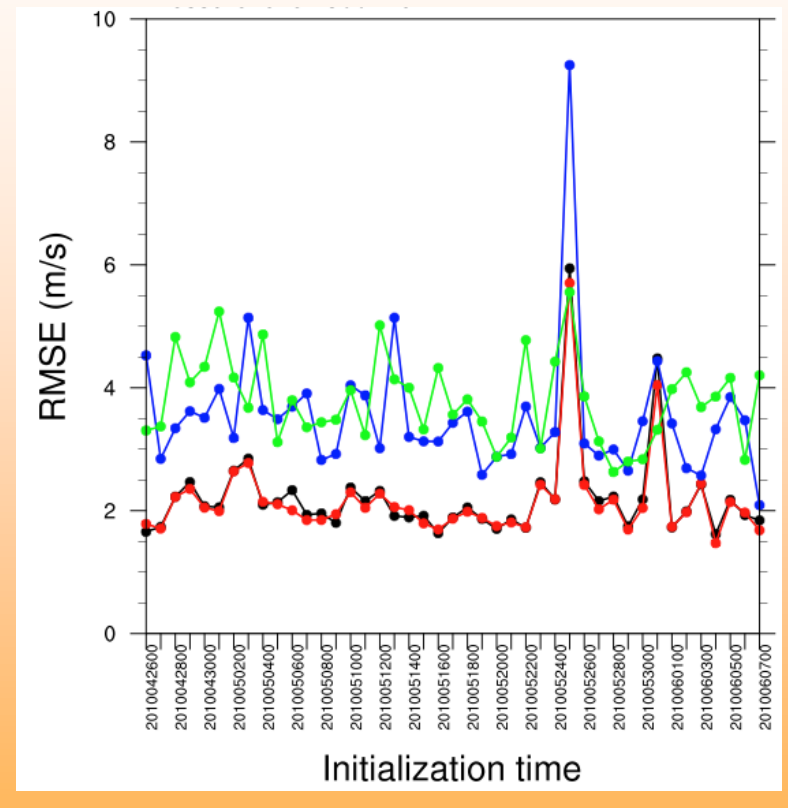
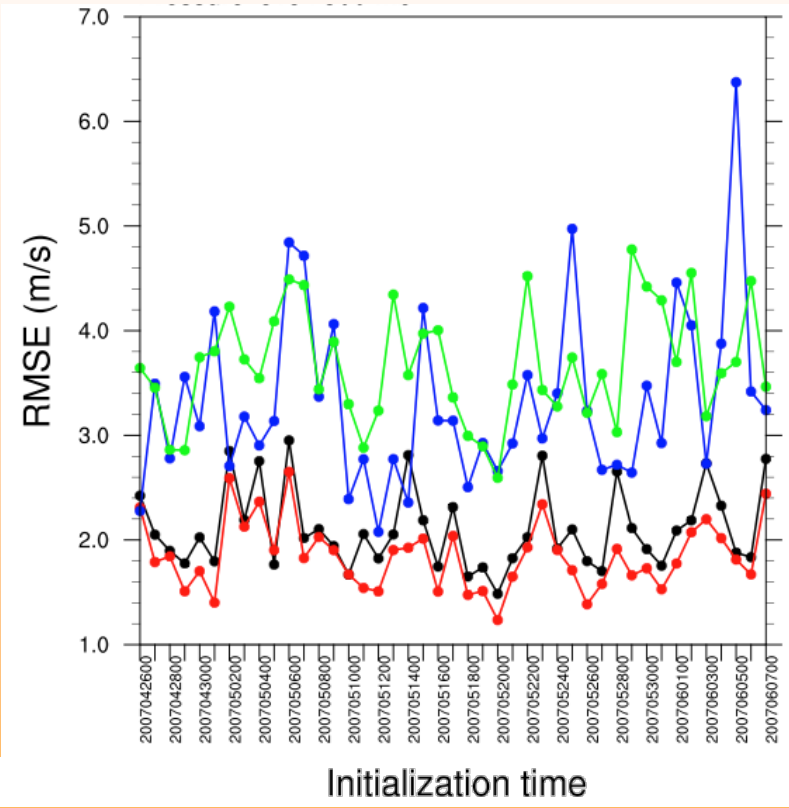
— parent initialization

— 24-hr WRF fcsts

# 500 hPa zonal wind

## 2007: NAM parent

## 2010: RUC parent

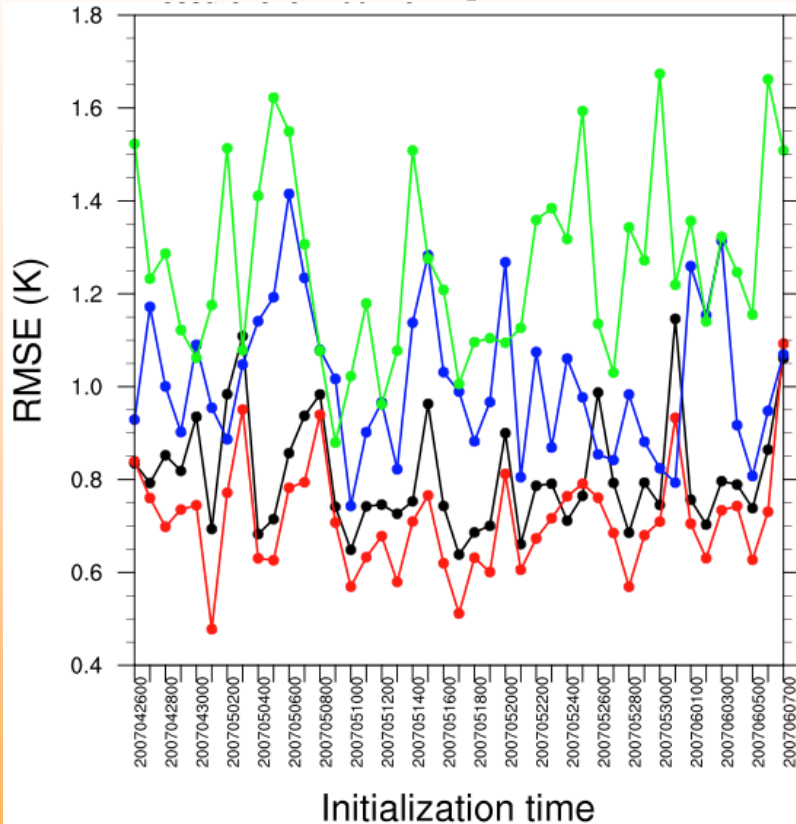


— WRF initialization  
— parent initialization

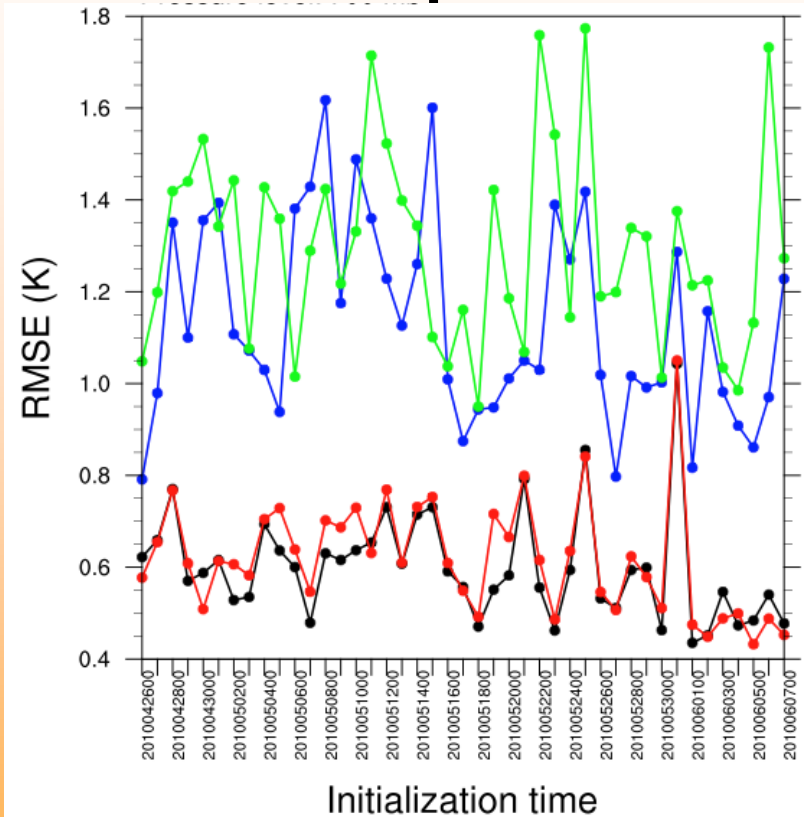
— 12-hr WRF fcsts  
— 24-hr WRF fcsts

# 700 hPa temperature

## 2007: NAM parent



## 2010: RUC parent



— WRF initialization  
— parent initialization

— 12-hr WRF fcsts  
— 24-hr WRF fcsts



# Spearman Correlations at analysis

- Correlation between the red and black lines

$$corr_{xy} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}}$$

$$corr_{xy} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}} = \frac{COV_{xy}}{\sigma_x \sigma_y}$$



$$P\left(B_{24} \geq \bar{B}_{24} \mid A_0 \geq \bar{A}_0\right) \text{Fhr} = 24$$

Pressure Levels (hPa)

		850	700	500	300	
U wind component	U	2007	52.6	44.4	37.5	45.0
		2010	72.7	46.2	61.5	60.0
V wind component	V	2007	52.4	50.0	65.0	31.8
		2010	46.2	50.0	70.0	45.5
Temperature	T	2007	55.0	53.3	65.0	58.8
		2010	46.7	66.7	43.8	29.4
Specific Humidity	Q	2007	59.1	41.7	56.5	60.0
		2010	85.0	55.0	72.0	68.4

Probabilities (%)



# Spearman Correlations at analysis

		Pressure Levels (hPa)				
		850	700	500	300	
U wind component	U	2007	.802	.886	.843	.879
		2010	.948	.958	.913	.960
V wind component	V	2007	.824	.844	.887	.896
		2010	.913	.955	.965	.971
Temperature	T	2007	.600	.856	.892	.950
		2010	.923	.892	.851	.767
Specific Humidity	Q	2007	.485	.680	.596	.766
		2010	.892	.960	.713	.788



# Spearman Correlations f00-f12

		Pressure Levels (hPa)				
		850	700	500	300	
U wind component	U	2007	.162	.049	.185	.175
		2010				
V wind component	V	2007	.308	.129	.377	.065
		2010				
Temperature	T	2007	.047	.162	.282	.181
		2010				
Specific Humidity	Q	2007	.114	.076	.200	.425
		2010				



# Spearman Correlations f00-f12

		Pressure Levels (hPa)				
		850	700	500	300	
U wind component	U	2007	.162	.049	.185	.175
		2010	.549	.452	.412	.401
V wind component	V	2007	.308	.129	.377	.065
		2010	.101	.404	.428	.419
Temperature	T	2007	.047	.162	.282	.181
		2010	.369	.422	.283	.478
Specific Humidity	Q	2007	.114	.076	.200	.425
		2010	.424	.531	.235	.589



# Spearman Correlations f00-f24

		Pressure Levels (hPa)				
		850	700	500	300	
U wind component	U	2007	.036	.012	.269	.313
		2010	.631	.259	.123	.436
V wind component	V	2007	.097	.024	.504	-.125
		2010	.115	.179	.407	.336
Temperature	T	2007	.124	.145	.278	.252
		2010	.234	.290	.221	.158
Specific Humidity	Q	2007	.312	-.252	.092	.187
		2010	.347	.284	.370	.328







# Probs from median f00-f12

		Pressure Levels (hPa)				
		850	700	500	300	
U wind component	U	2007	54.5	45.5	52.2	54.5
		2010	72.7	68.2	53.6	59.1
V wind component	V	2007	50.0	54.5	68.2	50.0
		2010	59.1	63.6	56.5	54.5
Temperature	T	2007	63.6	54.5	59.1	54.5
		2010	68.2	63.6	63.6	59.1
Specific Humidity	Q	2007	54.5	59.1	63.6	72.7
		2010	68.2	63.6	54.5	81.8

















# Correlations at analysis

		850	700	500	300
U	2007	.786	.897	.853	.909
	2010	.995	.993	.993	.999
V	2007	.811	.910	.912	.925
	2010	.991	.986	.997	.997
T	2007	.590	.865	.890	.938
	2010	.959	.942	.978	.902
Q	2007	.512	.690	.631	.850
	2010	.914	.932	.785	.911

# Correlations from 00-12

		850	700	500	300
U	2007	.186	.148	.118	.093
	2010	.662	.782	.813	.802
V	2007	.319	.132	.347	.012
	2010	.571	.657	.570	.662
T	2007	-.045	.174	.154	.302
	2010	.344	.328	.560	.458
Q	2007	-.110	.030	.179	.499
	2010	.459	.502	.484	.552







