Research and Development of GSI-based EnVar for Convective Scale Radar Data Assimilation: Challenge and Progress

Xuguang Wang

Multiscale data Assimilation and Predictability (MAP) Lab
School of Meteorology

University of Oklahoma, Norman, OK, USA

6th International Symposium on Data Assimilation,
Munich, Germany, Mar 5-9, 2018

MAP Lab DA research and development

MAP Lab

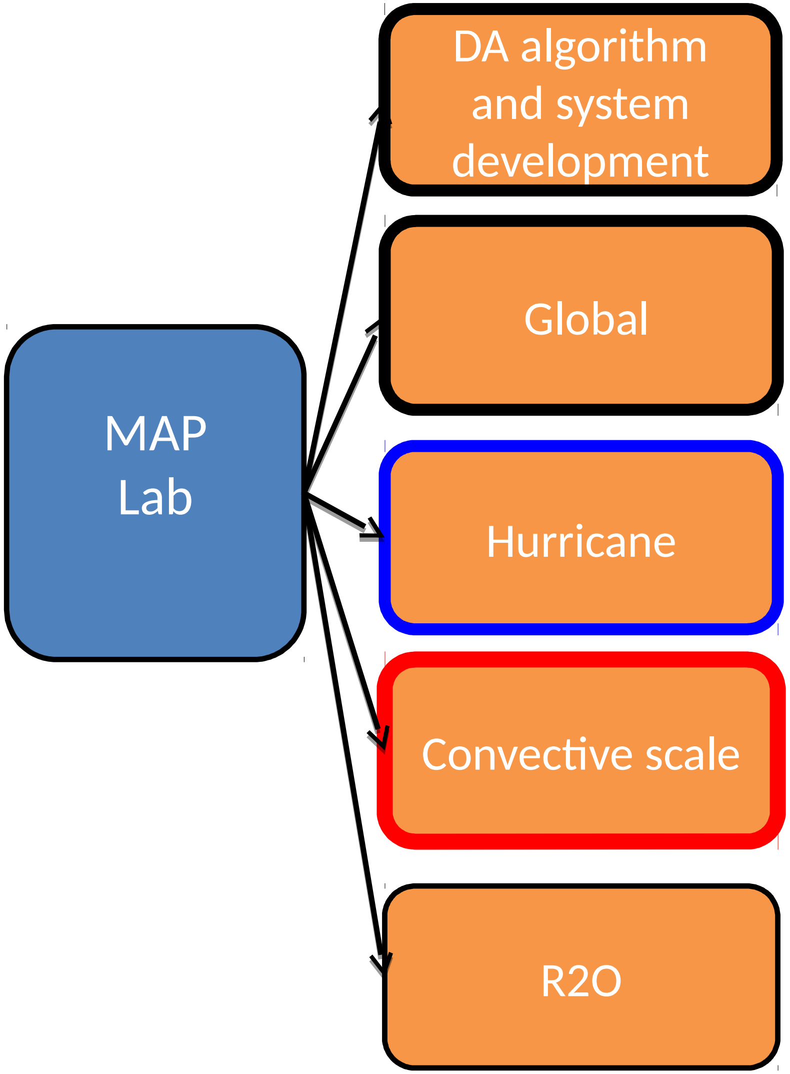
- DA algorithm and system development
- Global
- Hurricane
- Convective scale
- R2O

MAP Lab DA research and development

MAP Lab

- DA algorithm and system development
- Global
- Hurricane
- Convective scale
- R2O

MAP Lab DA research and development



Sample of recent work:

GSI 3DEnVar:
Wang, Parrish, Kleist, Whitaker, 2013, MWR

GSI 4DEnVar:
Wang and Lei, 2014, MWR

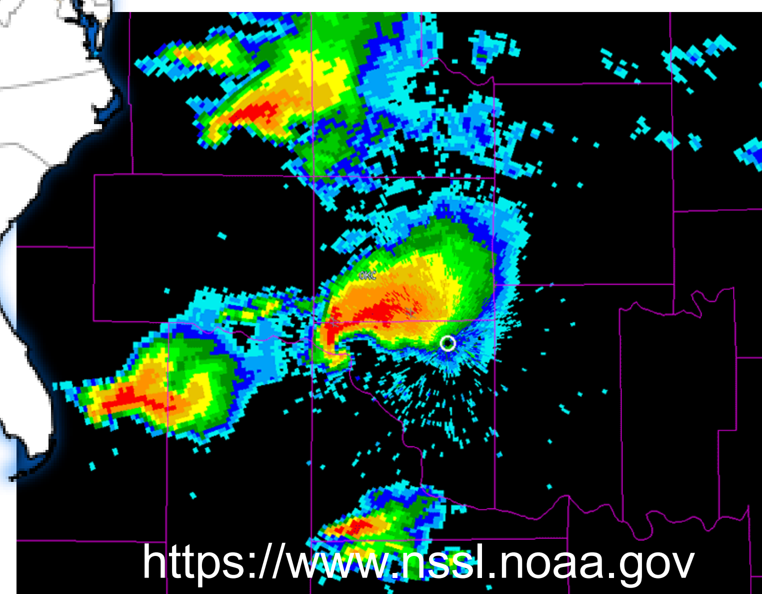
Cost effective method to increase ensemble size in GSI EnVar:
Huang and Wang 2018, MWR

Developed fully cycled GSI hybrid DA system for US operational convection allowing hurricane prediction system HWRF:
Lu, Wang, Tong and Tallapragada, 2017, MWR
Lu, Wang, Li, Tong, Ma, 2016, QJRMS

What got us there?

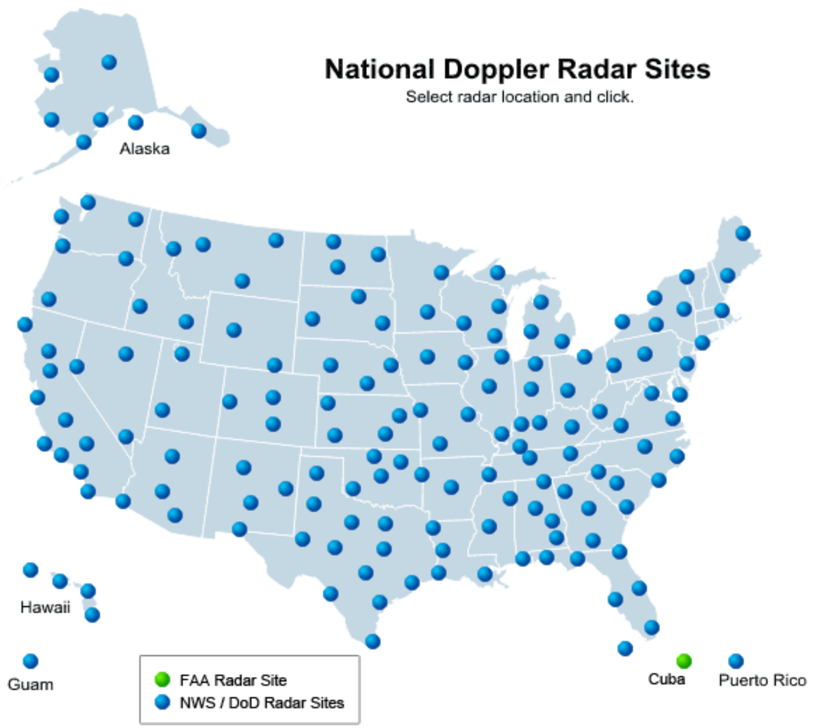
https://en.wikipedia.org

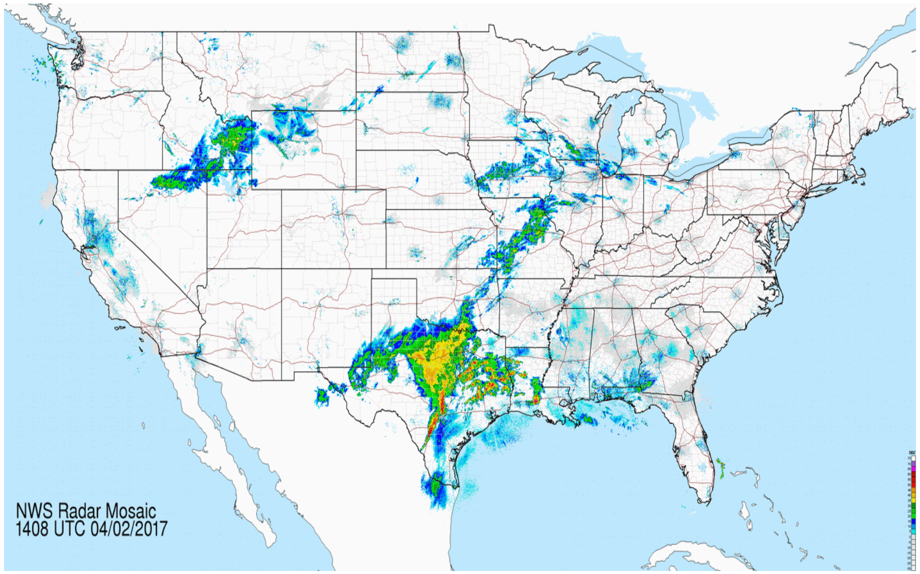
ttps://www.stormtours.com



https://www.nssl.noaa.gov

Importance of radar for convective scale observations and NWP

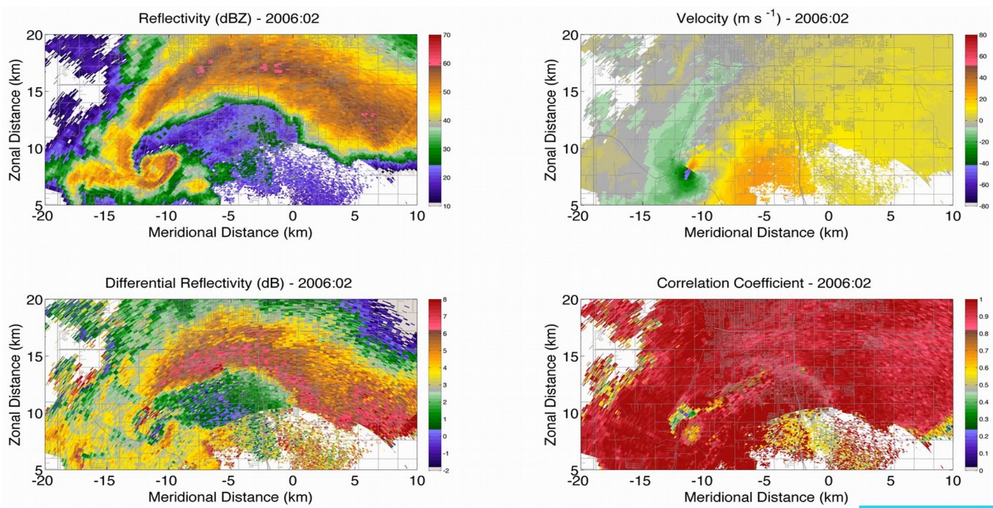




One of the most important observation platforms that provide high temporal and spatial resolution sampling of convective scales.

New radar technology pushes limits for radar DA

20 May 2013 Moore, Oklahoma Tornado



- PX solid-state, dual-polarization X-band radars by OU ARRC (Kurdzo et al. 2015)
- 30-m range resolution and 1.4 – 1.8° beamwidths
- 10 elevation angles in 10 – 20 s
- Assimilating movies?!

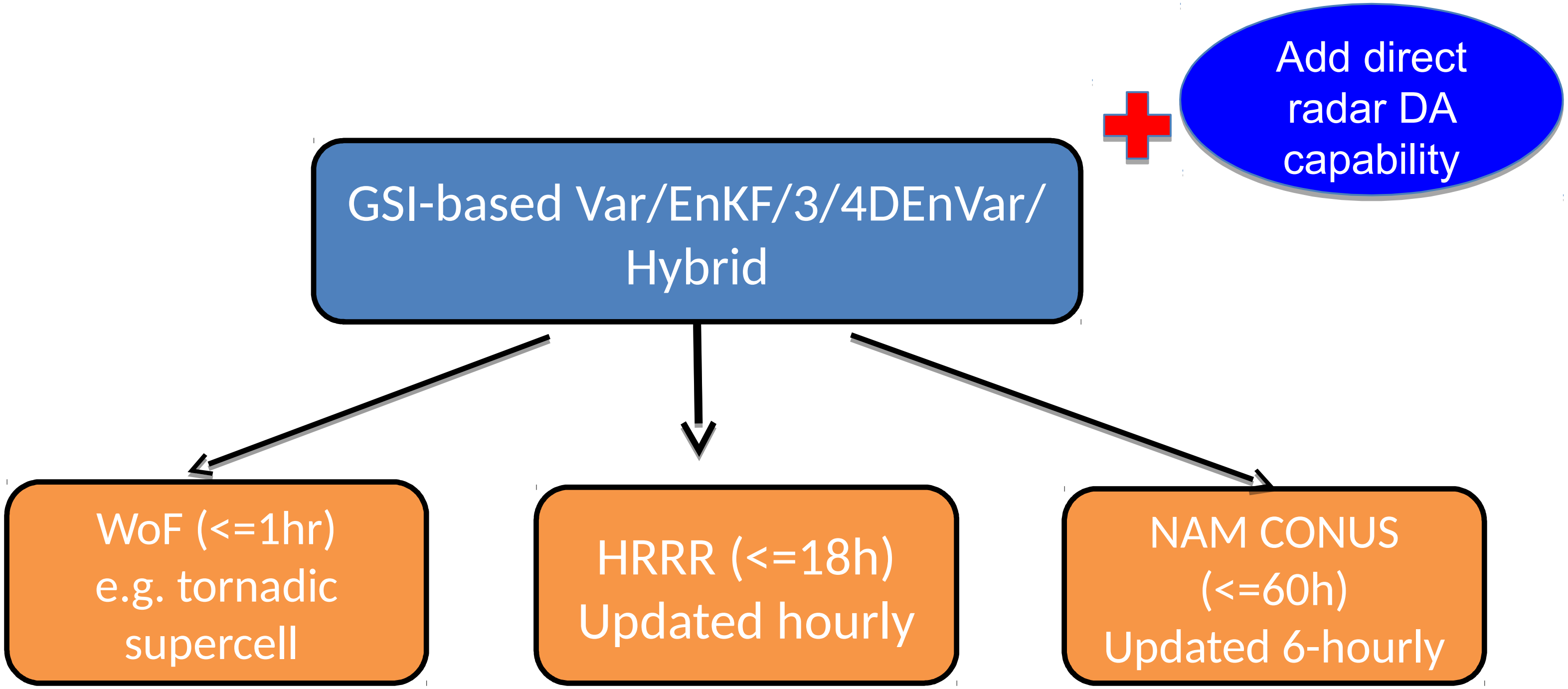
- Multi-function phased array radar replaces uniform scanning and enables targeted observations

https://www.nssl.noaa.gov

Challenges for convective scale radar data assimilation

- Require unique observation operators that are often complex and nonlinear (e.g., reflectivity, Dual pol radar variables)
- Both prior (e.g. hydrometeors) and observation errors are highly non-Gaussian
- Accurate cross-variable covariance is especially important
- Heavily rely on quality of microphysics schemes in numerical models – treatment of model errors
- Data are in much higher spatial resolution than the typical NWP model and in much higher temporal resolution than typical DA frequency – push limit
- New radar technology such as PAR allows adaptive sampling – targeted observations for radar and convective scale NWP push limit
- Systems shorter lived and with shorter predictability
- Convective scale prediction is a multi-scale problem, requiring an accurate estimate of both the convective scale details and the supporting mesoscale/synoptic scale environment

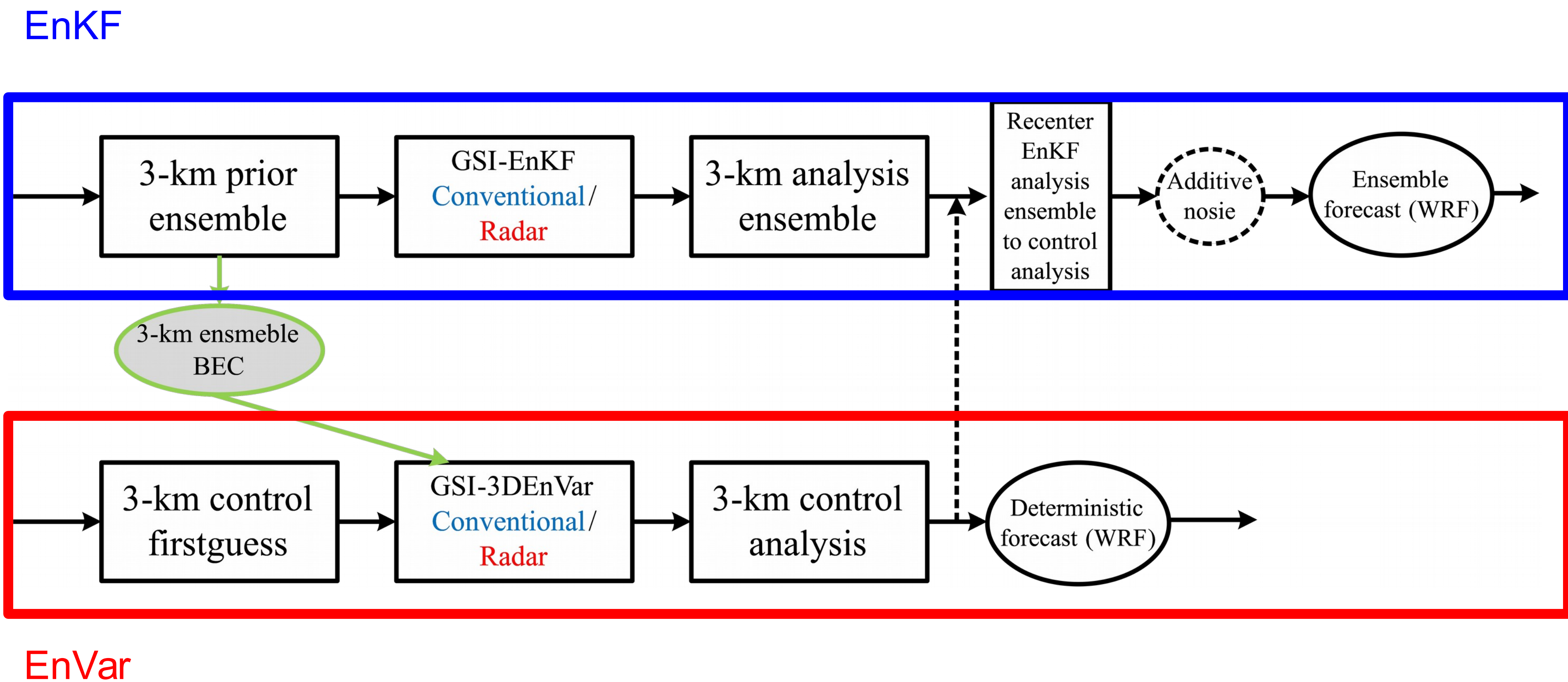
GSI-based EnKF-Var hybrid DA system further developed for various US convection-allowing prediction systems with direct radar DA capability



GSI: US operational data assimilation system
WoF: Warn on Forecast
HRRR: High Resolution Rapid Refresh
NAM CONUS: North American Mesoscale Forecast system - Continental US

In collaboration with NOAA EMC (Carley), GSD (Dowell) and NSSL (Wicker) colleagues

GSI-based EnKF-Var hybrid DA system further developed for various CONUS convection allowing prediction systems with direct radar DA capability



- Both the GSI EnKF and GSI EnVar components are extended
- Instead of using the coarser resolution global ensemble in EnVar like the operational HRRR, the new EnVar system ingests convection allowing model's own EnKF ensemble.

GSI-based EnKF-Var hybrid DA system further developed for various CONUS convection allowing prediction systems with direct radar DA capability

- Direct radar (Vr and dBZ) DA capability is developed for GSI EnKF and GSI EnVar

- Radial velocity and reflectivity observation operators are implemented.
- The hydrometeor-related prognostic variables (rainwater, snow, graupel mixing ratios) are added as state variables.
- The vertical velocity are added as state variables.
- Direct reflectivity assimilation capability is developed as opposed to using the operational Cloud Analysis (CA) method, which is a separate and empirical approach.

Outline

- Part I (Wang Y. and X. Wang, 2017, MWR)
- Issues of direct radar reflectivity assimilation in EnVar associated with the nonlinear operator
- Propose a method to directly assimilate radar reflectivity without tangent linear and adjoint of the nonlinear observation operator in GSI EnVar system
- Experiment with May 8th 2003 OKC tornadic supercell to demonstrate the method

- Part II (e.g. Duda, Wang, Wang, Carley, 2018)
- Systematic experiments to test the EnVar with the new direct reflectivity assimilation in the context of US CONUS operational convection allowing regional prediction system HRRR/NAM CONUS by comparing with the operational Cloud Analysis (CA)

- Part III (Wang Y. and X. Wang, 2018)
- Extension of dual resolution GSI based EnVar for sub-kilometer (<1km) analysis and prediction

- Part IV (Kerr and Wang, 2018)
- Ensemble based targeted observation applied for the Multi-function phase array radar (MPAR)

PART I

Issue with TL of nonlinear reflectivity operator in EnVar

Wang and Wang 2017, MWR, 145, 1447- 1471

- GSI-based EnVar cost function (Wang 2010, MWR)

$$J(\mathbf{a}) = 0.5(\mathbf{a})^{\mathrm{T}}\mathbf{A}^{-1}(\mathbf{a}) + 0.5(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}')^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}^{o'} - \underline{\mathbf{H}}\mathbf{x}')$$

$$\Delta_{\mathbf{a}} J_o = \mathbf{D}^{\mathrm{T}}\underline{\mathbf{H}}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}' - \mathbf{y}^{o'})$$

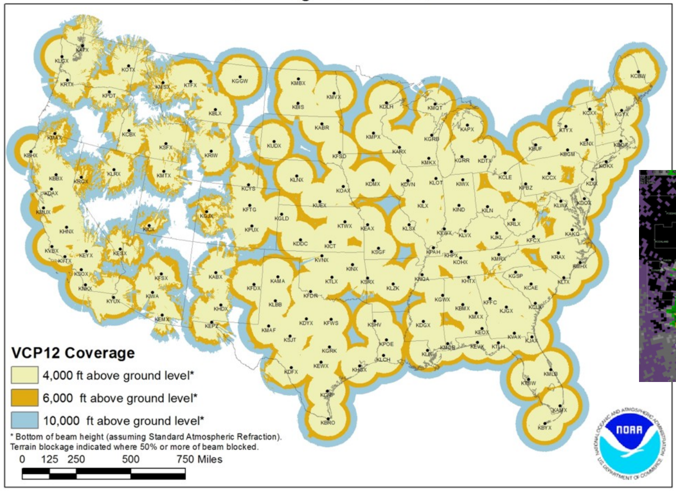
$$\mathbf{x}' = \sum_{k=1}^{K}(\mathbf{a}_k \circ \mathbf{x}_k^{e})$$

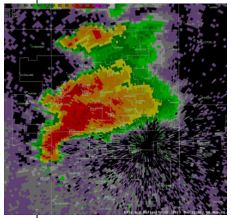
- Nonlinear radar reflectivity operator

$$H(q_r, q_s, q_g) = Z_{dB} = 10\log Z_e$$

$$Z_e = Z_r + Z_s + Z_g$$

$$Z_g = 4.33 \times 10^{10}(\rho q_g)^{1.75}$$

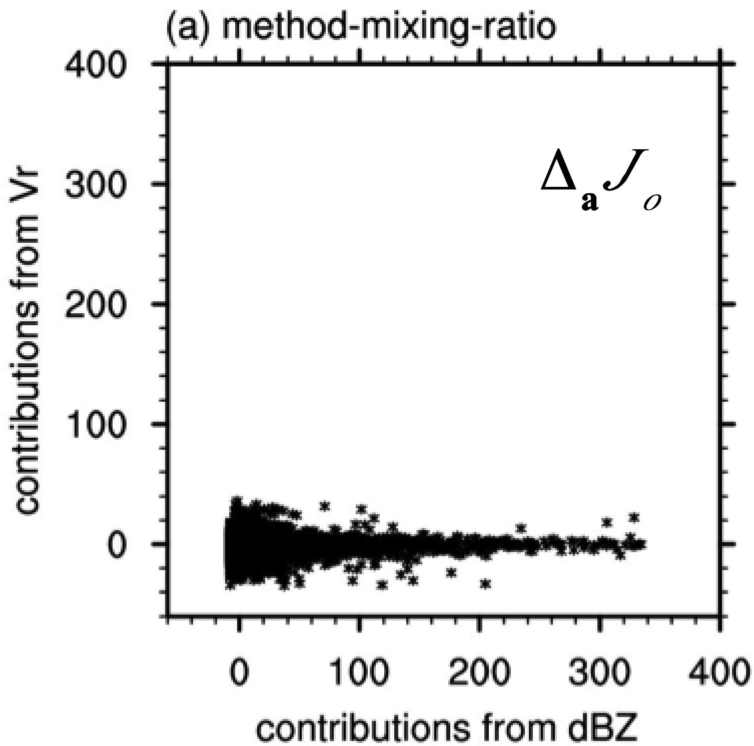


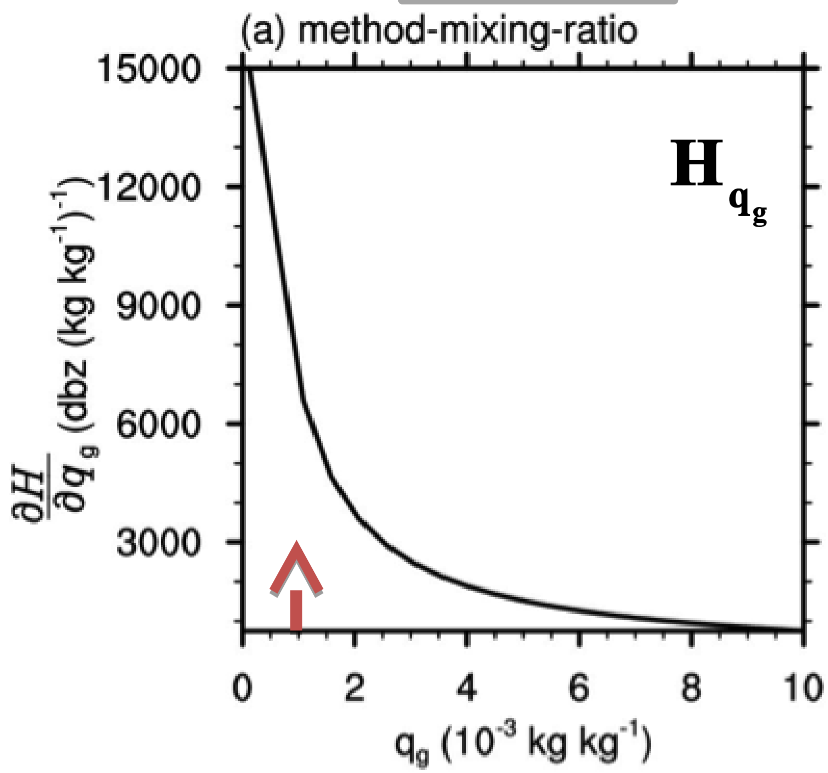


Issue with TL of nonlinear reflectivity operator in EnVar

Wang and Wang 2017, MWR

- Use hydrometeor mixing ratio as state variable $\boxed{H(q_s, q_r, q_g)}$



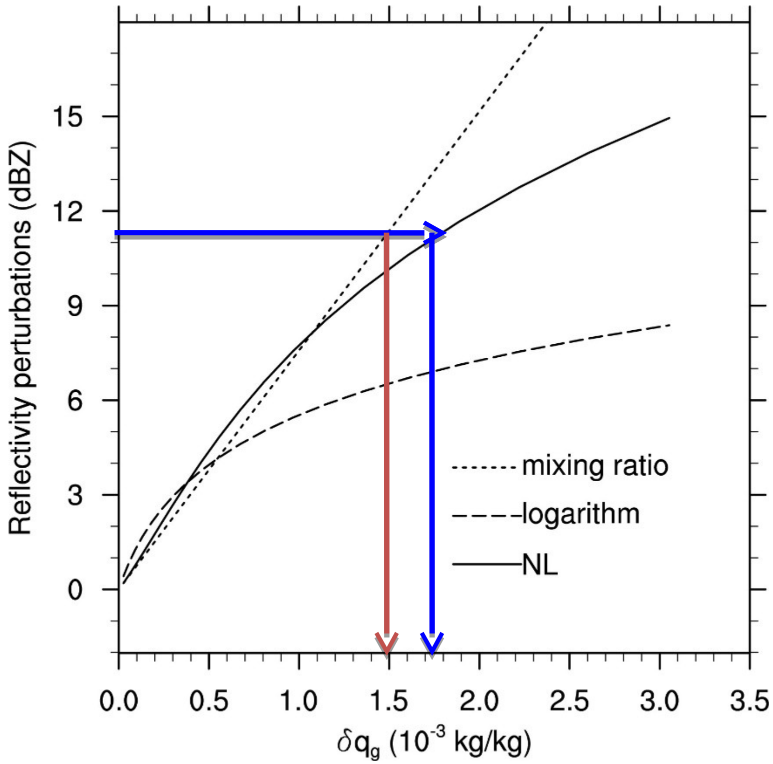


- ➢ Large values of TL of the nonlinear reflectivity associated with the small hydrometeor mixing ratios lead to large differences of cost function gradients contributed by Vr and Ref., which prevents efficient convergence and therefore under-estimates the hydrometeor increments.
- ➢ This issue disallows simultaneous assimilation of Vr and Ref.
- ➢ Issue true for Var (J. Sun mentioned in early 4DVar work), not just EnVar

Issue with TL of nonlinear reflectivity operator in EnVar

Wang and Wang 2017, MWR

- Use hydrometeor mixing ratio as state variable $\boxed{H(q_s, q_r, q_g)}$



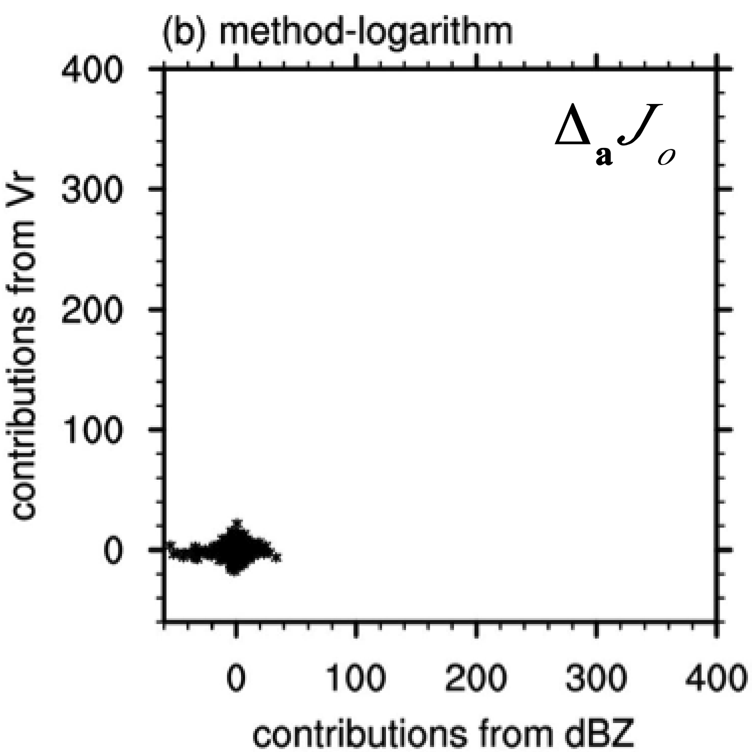
$$\Delta \mathbf{y} = H(\mathbf{x} + \Delta \mathbf{x}) - H(\mathbf{x}) = \mathbf{H}\Delta \mathbf{x}$$

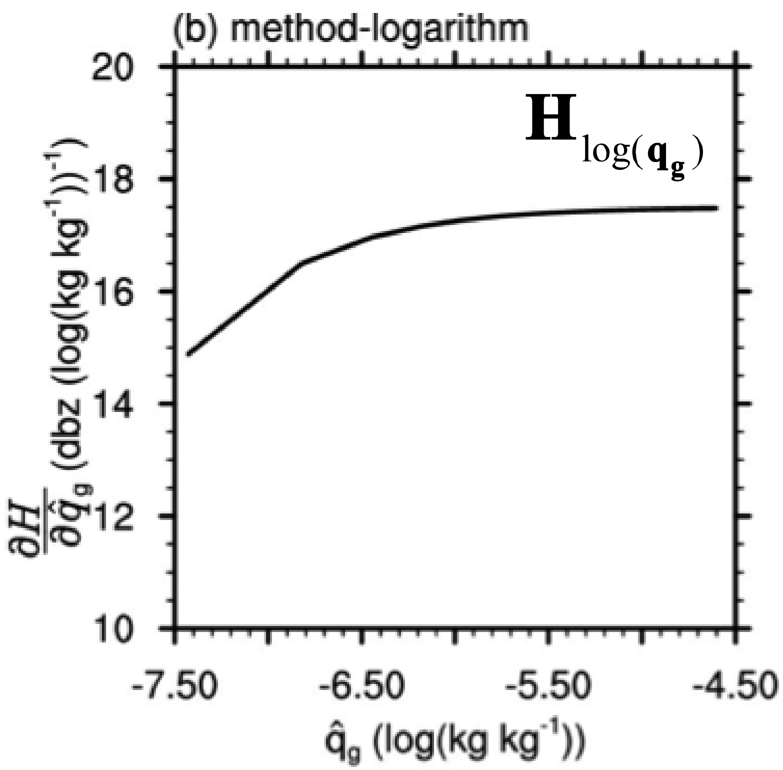
- ➢ The TL of the reflectivity operators itself further contributes to spuriously small hydrometeor increments

Issue with TL of nonlinear reflectivity operator in EnVar

Wang and Wang 2017, MWR

- Using logarithm of hydrometeor mixing ratio as state variable $\boxed{H(\log(q_s), \log(q_r), \log(q_g))}$



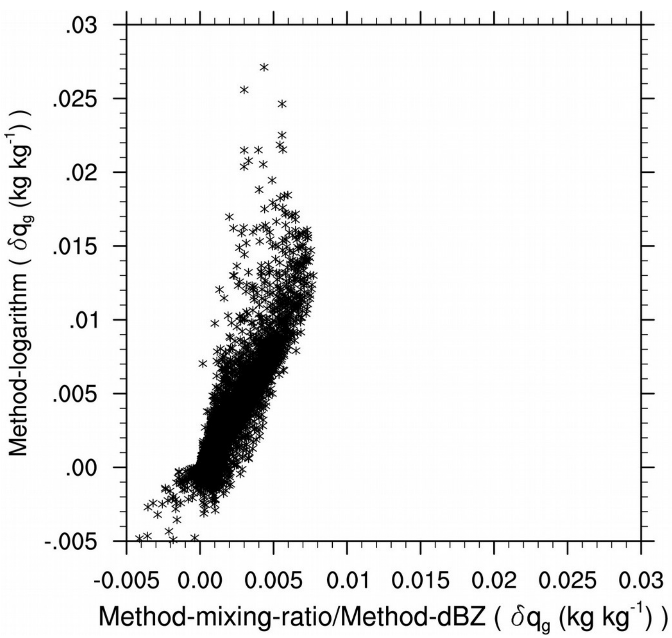


- Fixes the cost function gradient issue

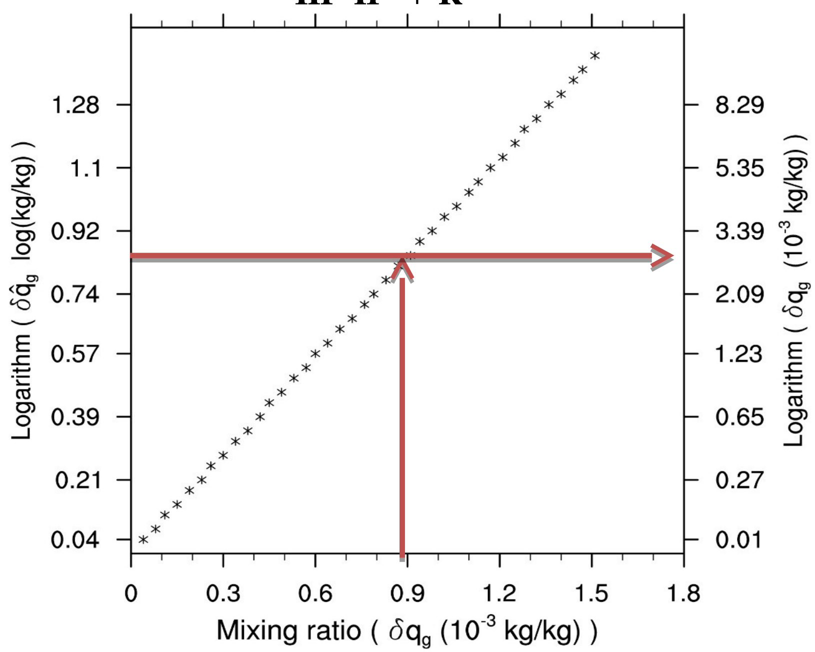
Issue with TL of nonlinear reflectivity operator in EnVar

Wang and Wang 2017, MWR

- Use logarithm of hydrometeor mixing ratio as state variable $\boxed{H(\log(q_s), \log(q_r), \log(q_g))}$



$$\delta \mathbf{x} = \frac{\mathbf{P}^b\mathbf{H}^T}{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

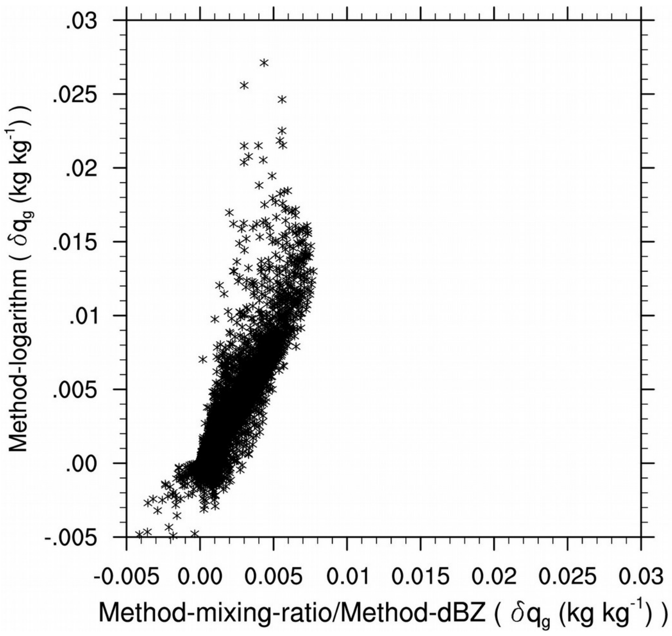


- ➢ However, it produces anomalously large hydrometeor increment partly due to the transform to and from the logarithmic space.
- ➢ Ad hoc thresholding may help, however solution is fundamentally incorrect.

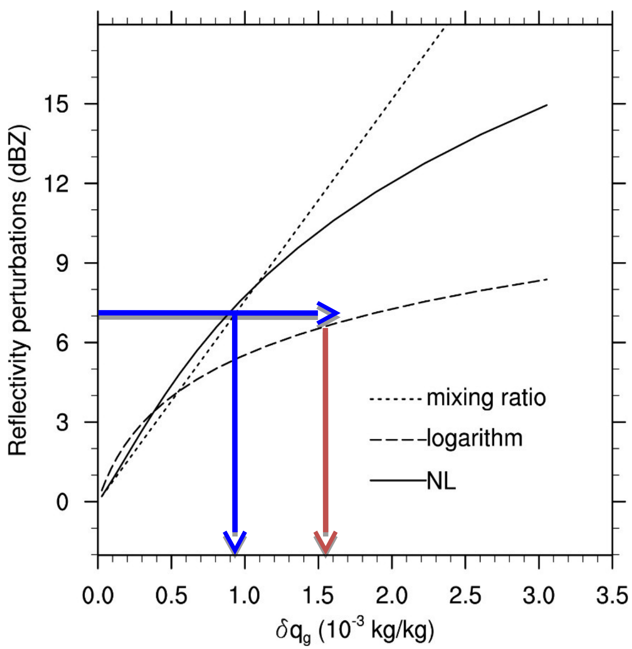
Issue with TL of nonlinear reflectivity operator in EnVar

Wang and Wang 2017, MWR

- Use logarithm of hydrometeor mixing ratio as state variable $\boxed{H(\log(q_s),\log(q_r),\log(q_g))}$



$$\Delta\mathbf{y} = H(\mathbf{x}+\Delta\mathbf{x}) - H(\mathbf{x}) = \mathbf{H}\Delta\mathbf{x}$$

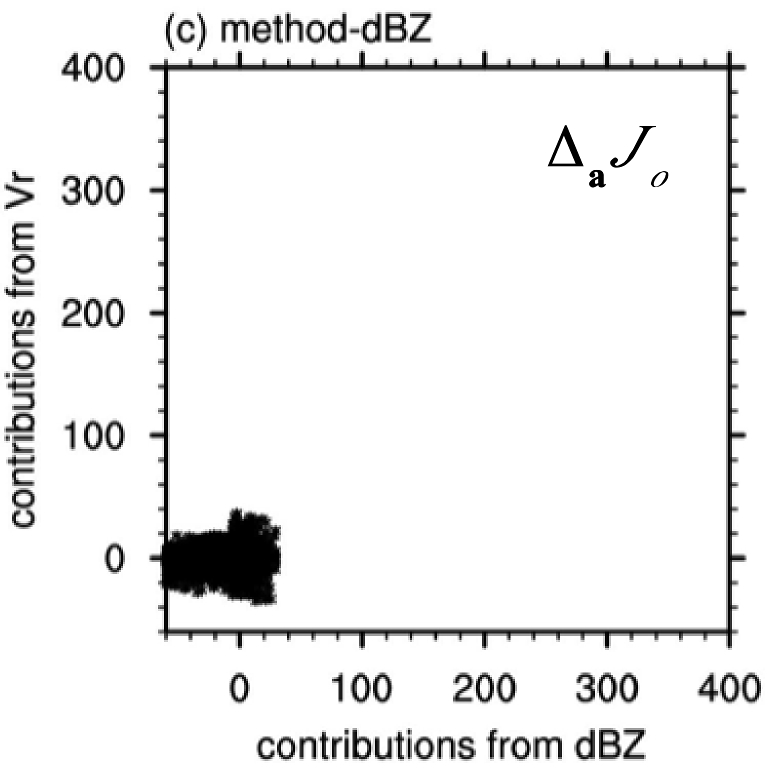


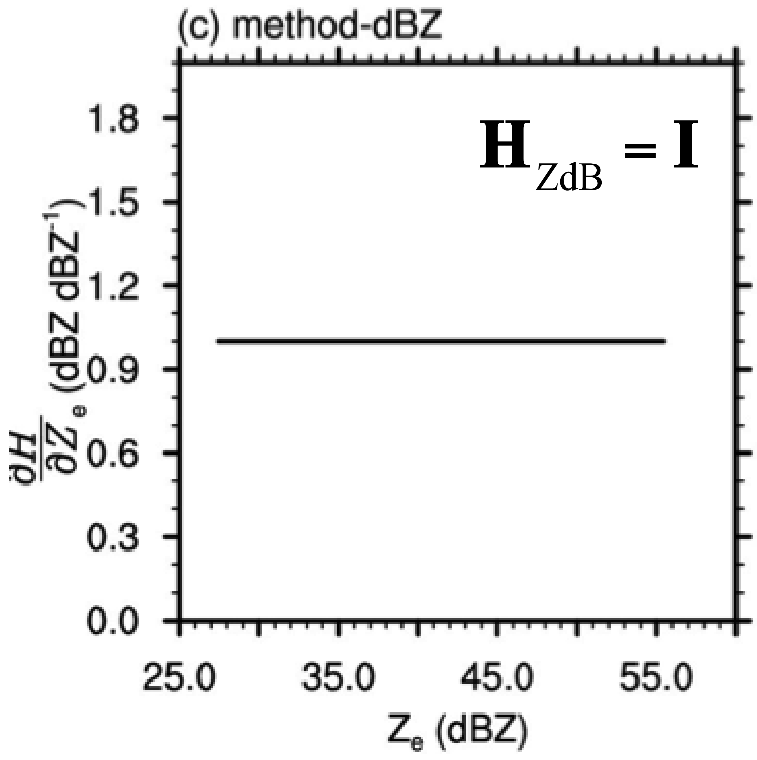
➢ The TL of the reflectivity operators itself further contributes to spuriously large hydrometeor increments

GSI-based EnVar without tangent linear (TL) and adjoint of the nonlinear reflectivity operator

Wang and Wang 2017, MWR

- A new method extending state variables by directly including reflectivity as state variable is proposed: $\boxed{H(ZdB)}$
- No reflectivity operator appears in cost function or $\mathbf{H}_{\mathrm{ZdB}} = \mathbf{I}$



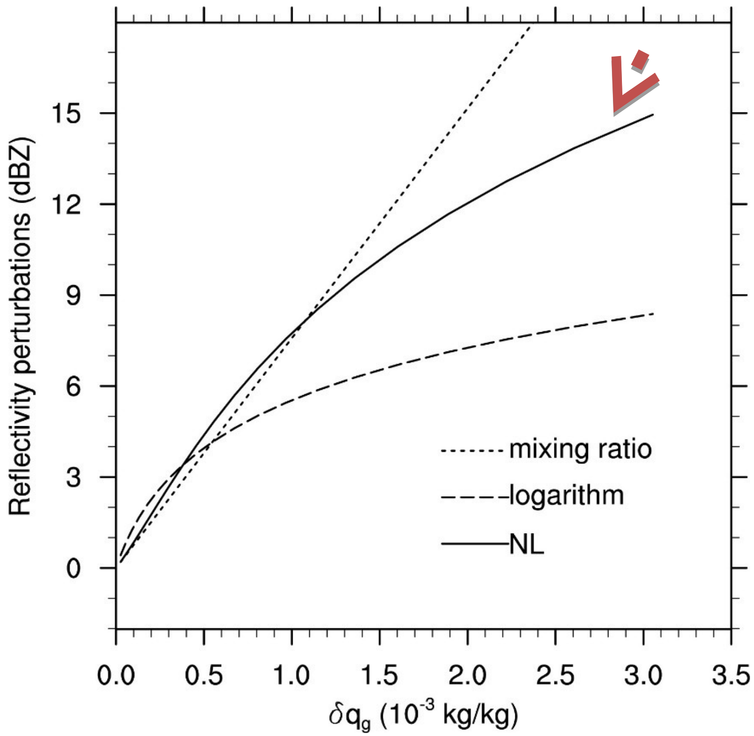


- Gradient issues fixed - allow simultaneous assimilation of Ref. with other observations

GSI-based EnVar without tangent linear (TL) and adjoint of the nonlinear reflectivity operator

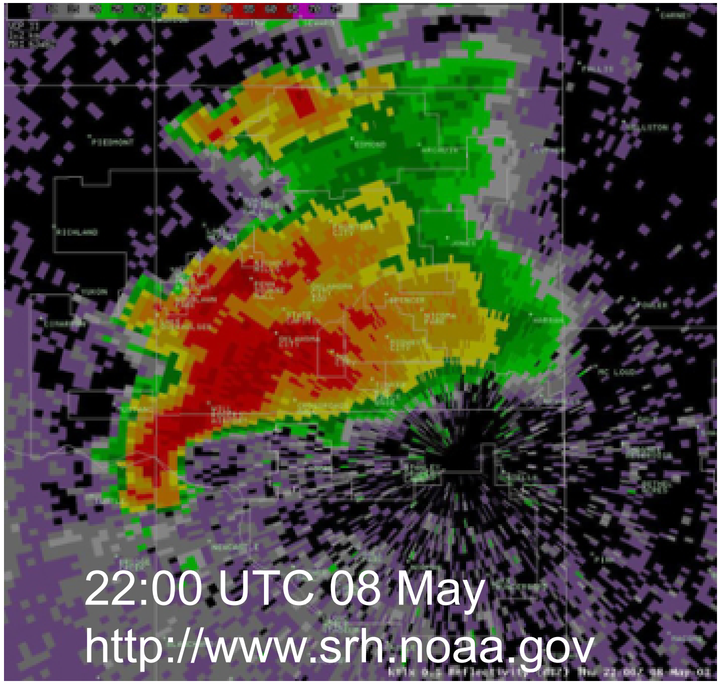
Wang and Wang 2017, MWR

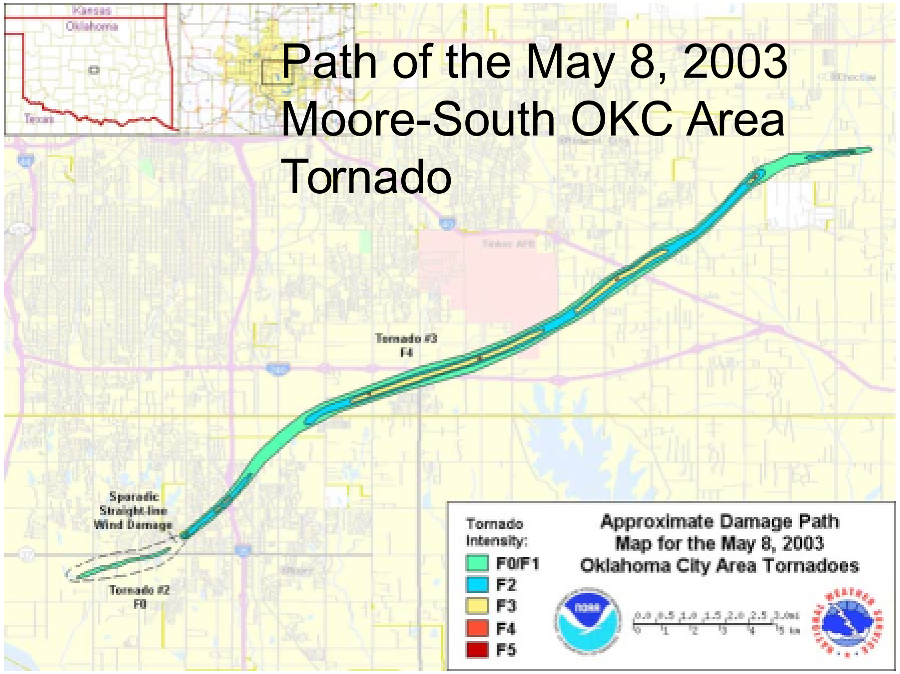
- A new method including reflectivity as state variable is proposed $\boxed{H(ZdB)}$ $\mathbf{H}_{\mathrm{ZdB}} = \mathbf{I}$



- In this method, no TL of the reflectivity operator exists . Hydrometeor is related to reflectivity following the nonlinear relationship.

May 8th 2003 OKC Tornadic Supercell

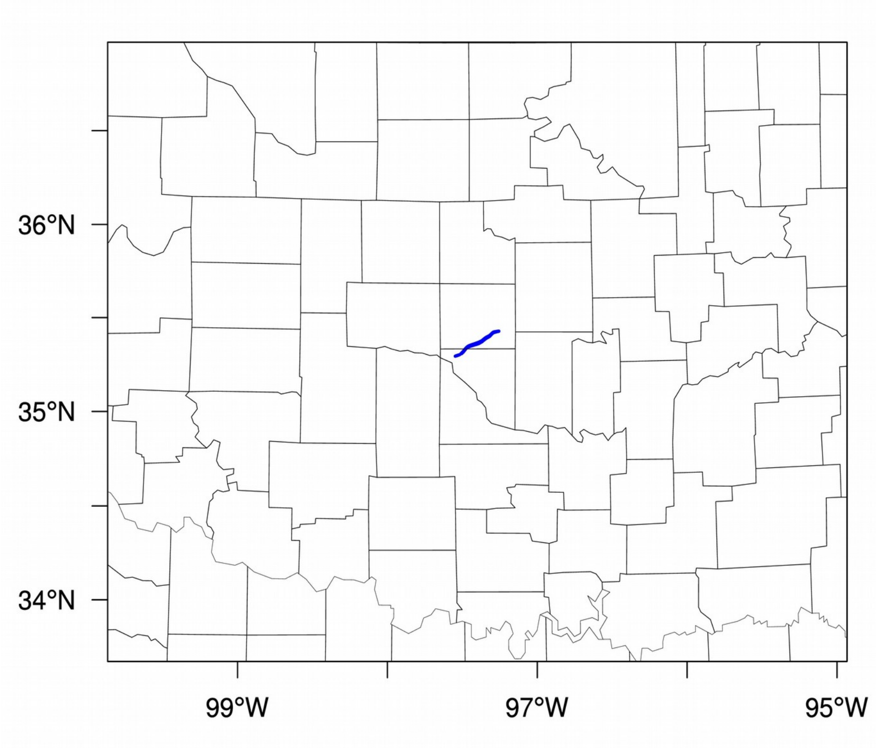




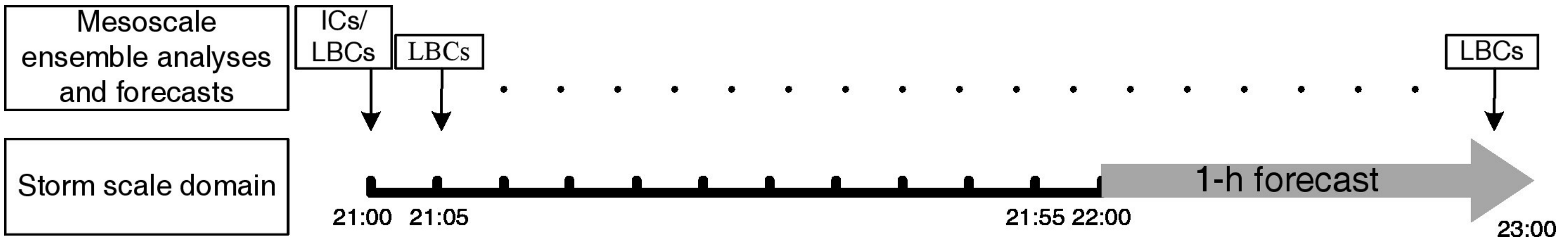
- An isolated supercell case that produced F-4 intensity tornadoes in Moore and Oklahoma City (OKC) during about 2210—2240 UTC.
- Supercell maintained well beyond 2300 until about 0000 UTC.

Experiment design

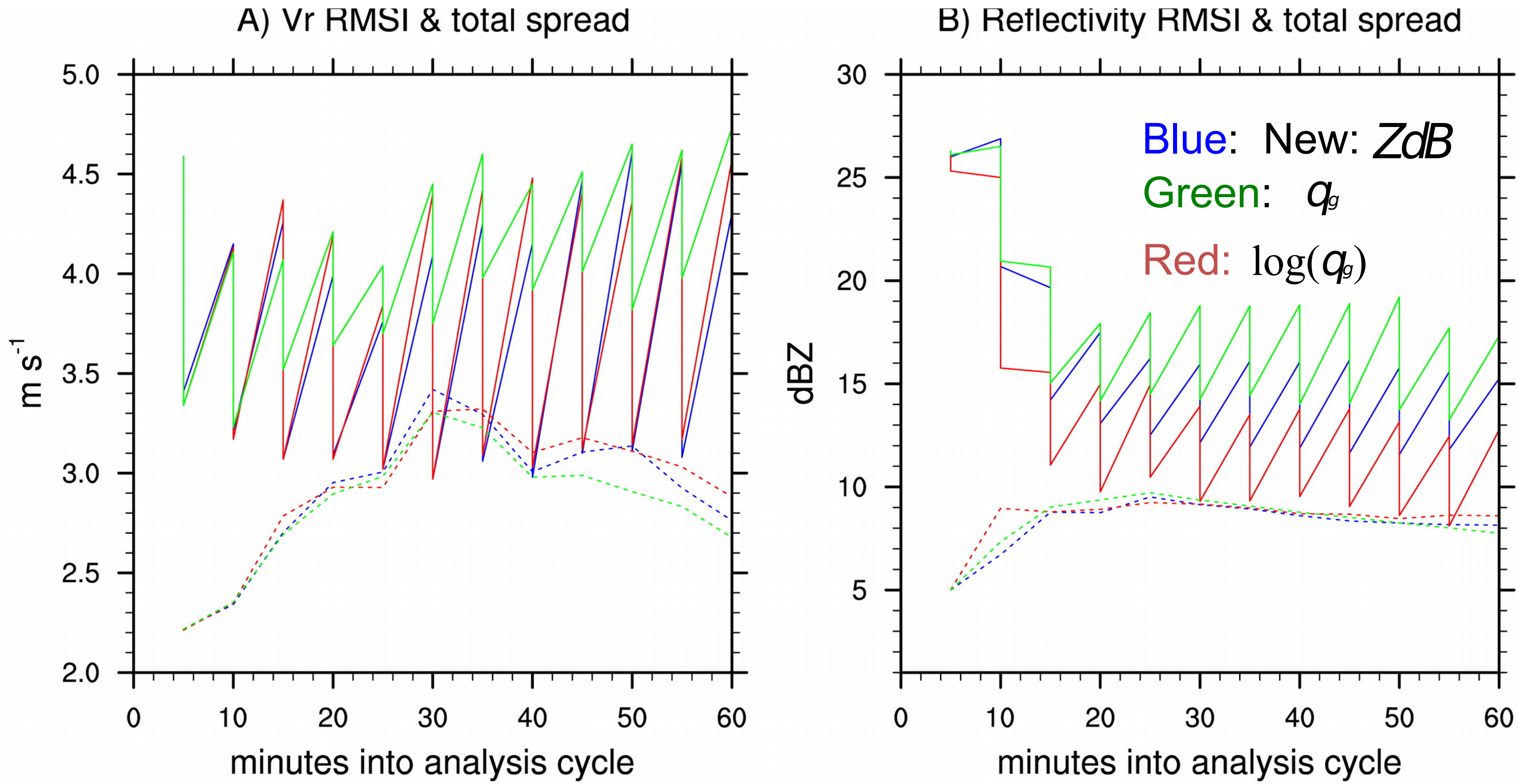
Wang and Wang, 2017, MWR



- **Model**: WRF-ARW 2km
- **Observation**: radar radial wind and reflectivity from KTLX
- **IC and LBC ensemble**: A 45-member ensemble downscaled from a mesoscale ensemble at 2100 UTC.



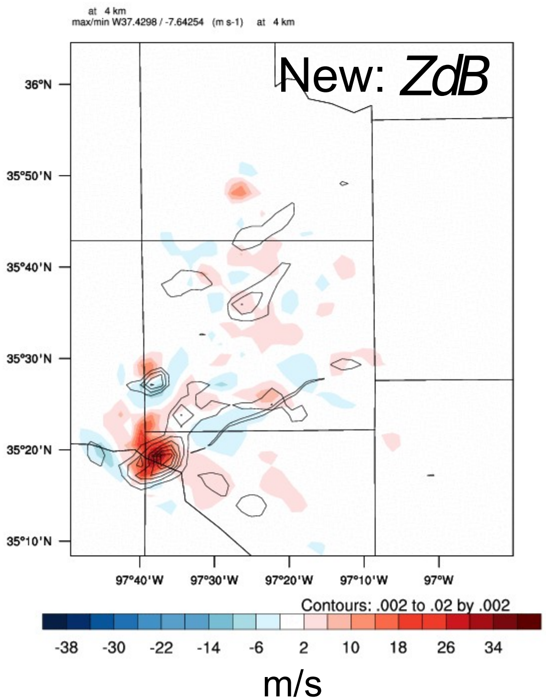
RMS fit to obs. during DA cycling



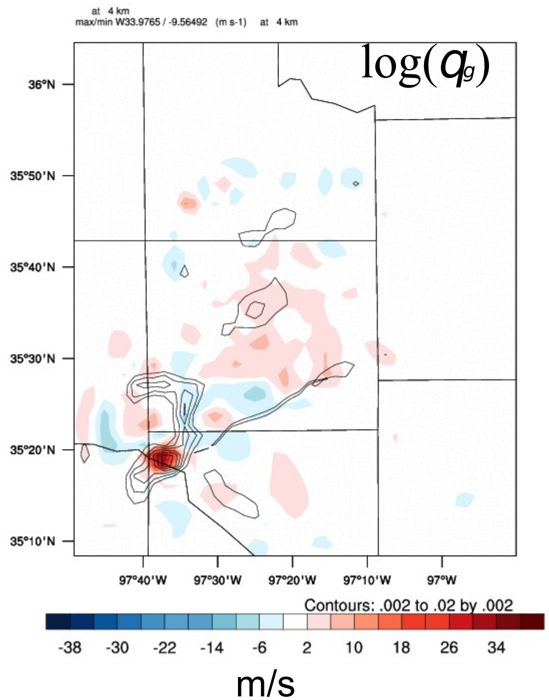
- Both radial velocity and reflectivity are under-fitted for using hydrometeor as state variables
- Reflectivity is over-fitted using the log(hydrometeor) as state variables

1 hour forecast: w and vorticity at 4km

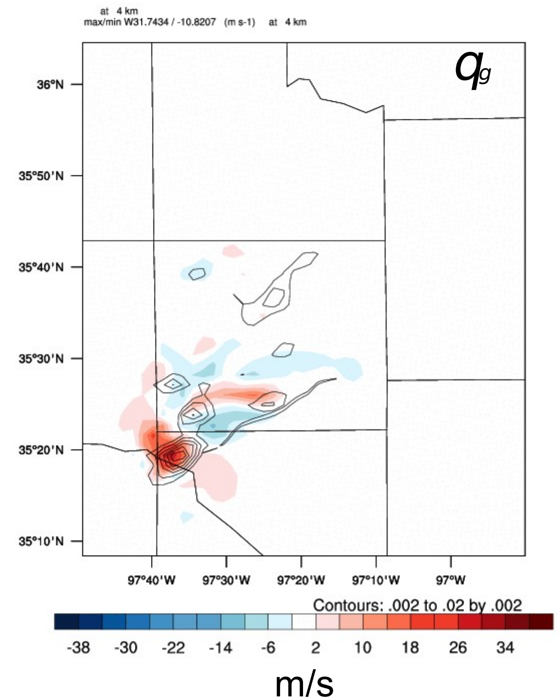
New: extend state variable with reflectivity



Use log transform (q_hydrometeor) as state variable

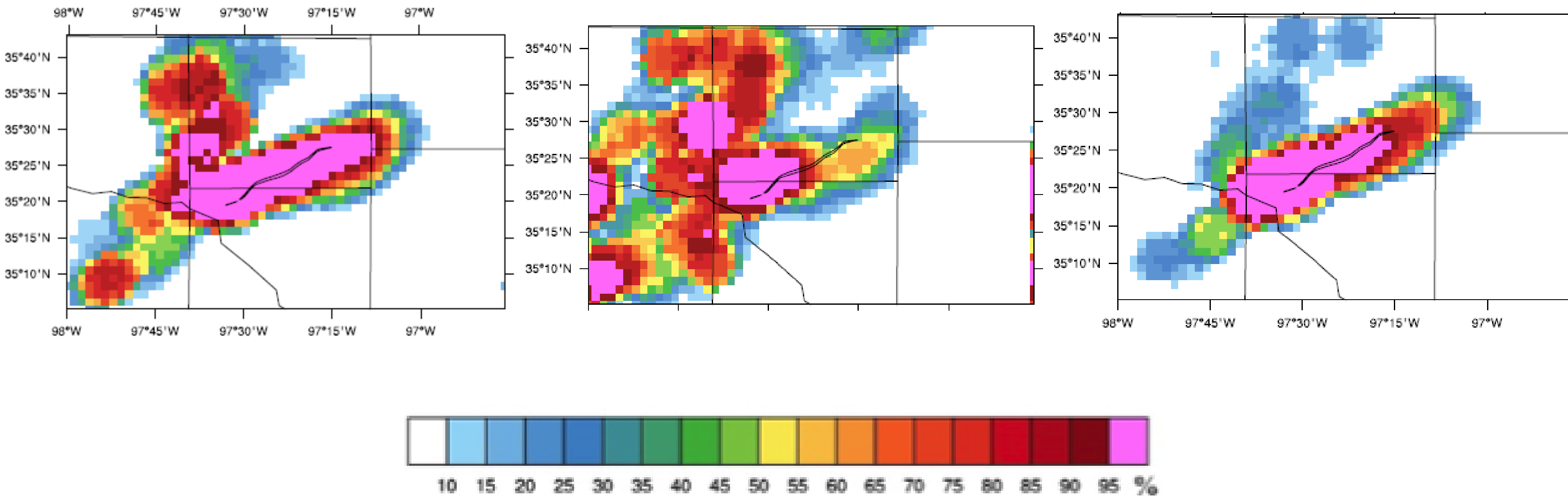


Use q_hydrometeor as state variable

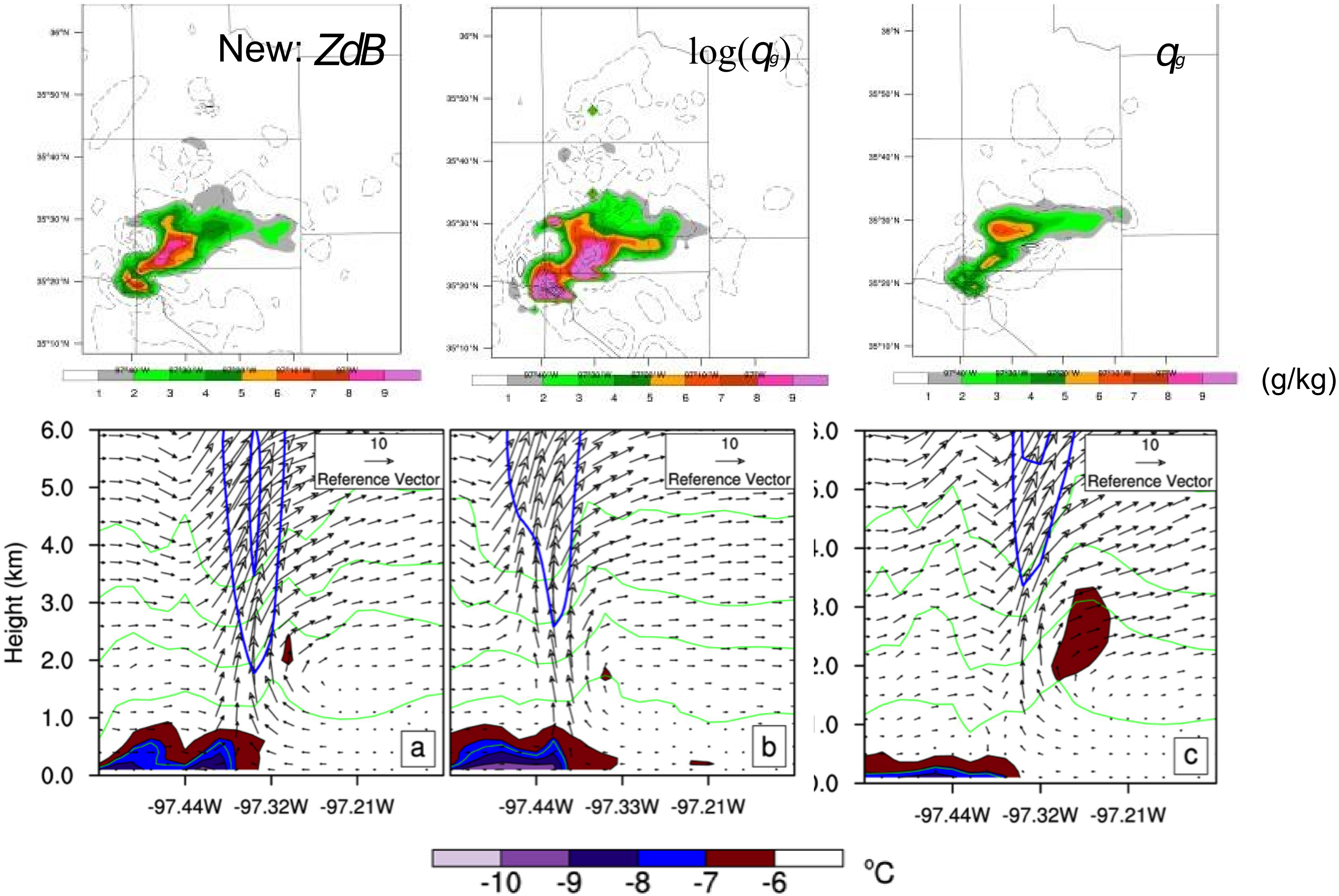


1 hour forecast of neighborhood ensemble probability (%) of vorticity at 150 m AGL

New: *ZdB* | $\log(q_g)$ | q_g



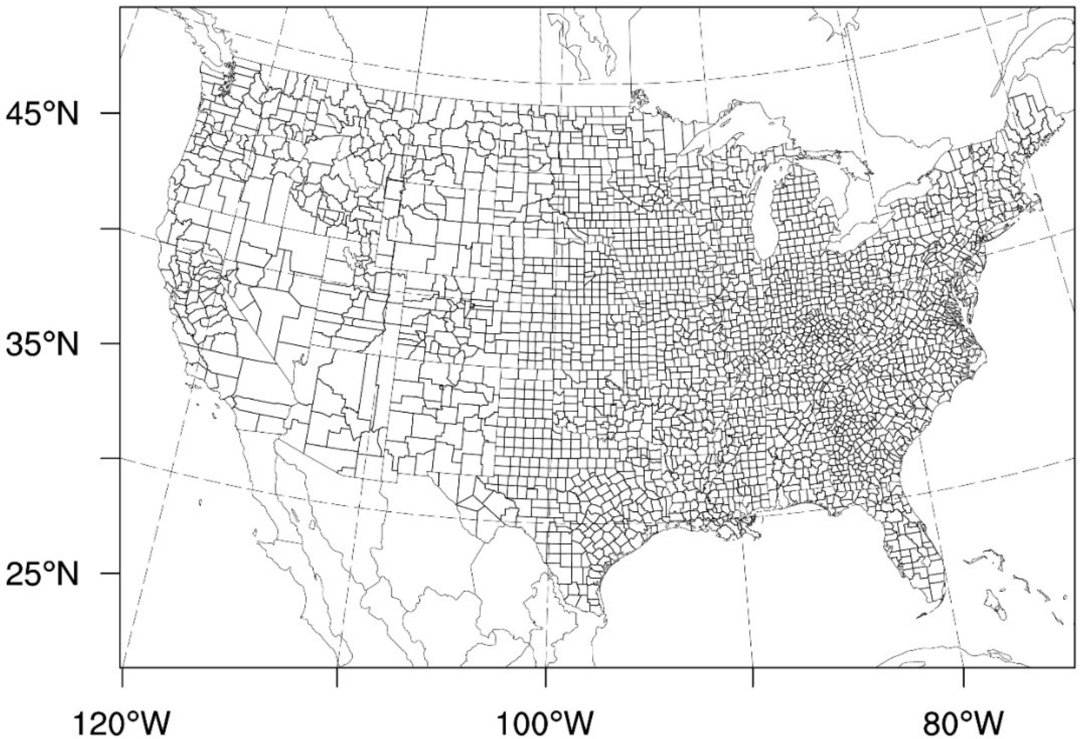
Graupel (q_g) analysis



PART II

Implementation and experiment in operational HRRR/NAM CONUS

Wang Y., Wang, Carley 2018, MWR



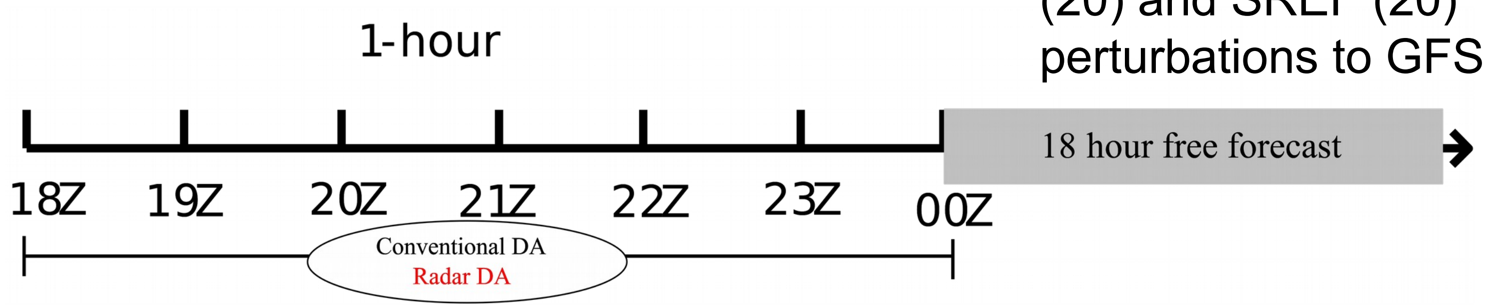
Domain:

- Resolution: 3 km
- Grid: 1621 X 1121 X 50
- Large CONUS domain in operational HRRR context

Observations:

- Conventional obs. are assimilated hourly for 6 hours
- Radar data are assimilated sub-hourly/hourly

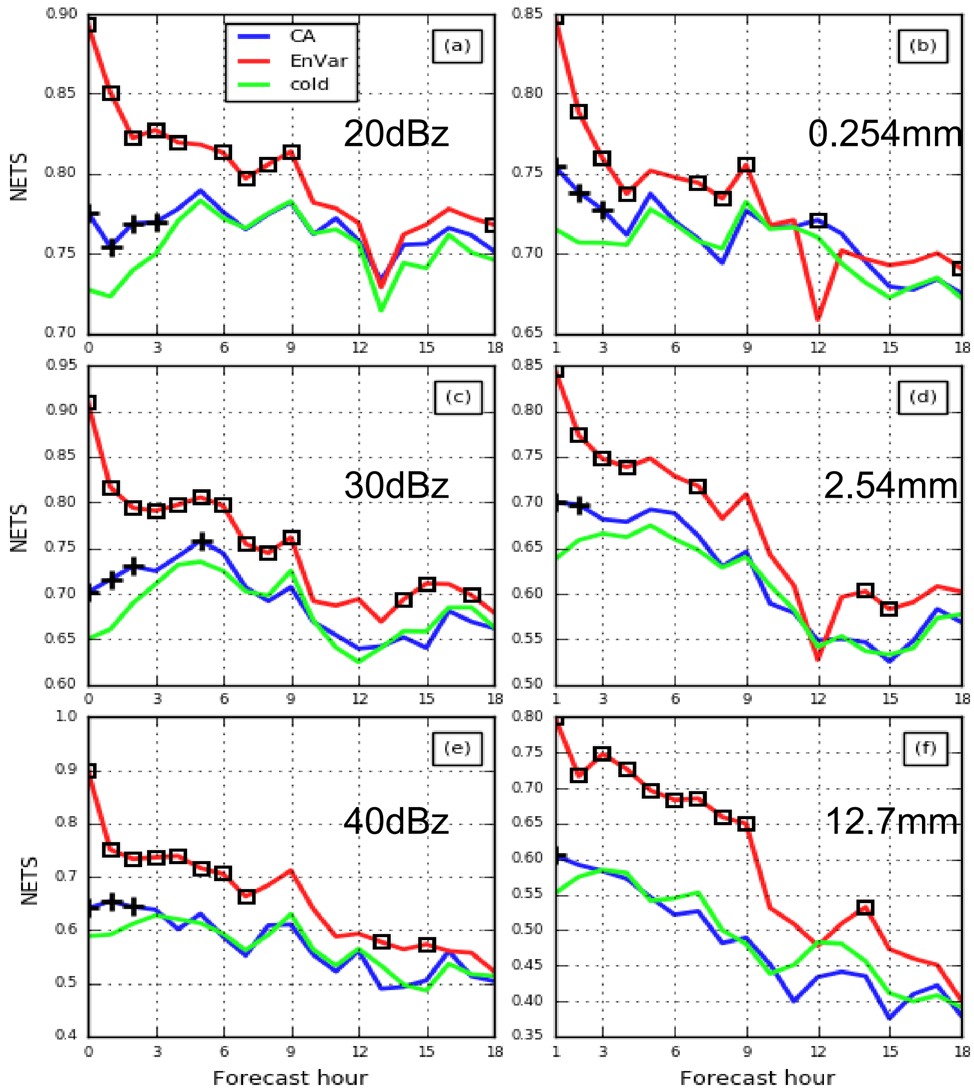
IC and LBC ensemble are provided by recentering GEFS (20) and SREF (20) perturbations to GFS-ctl



GSI-EnVar direct reflectivity assimilation vs cloud analysis (CA)

Duda, Wang, Wang, Carley, 2018a, MWR

- ❑ EnVar overall verifies much better than CA.
- ❑ CA does provide some benefit over not assimilating radar reflectivity at all, however, but only a few hours' worth.

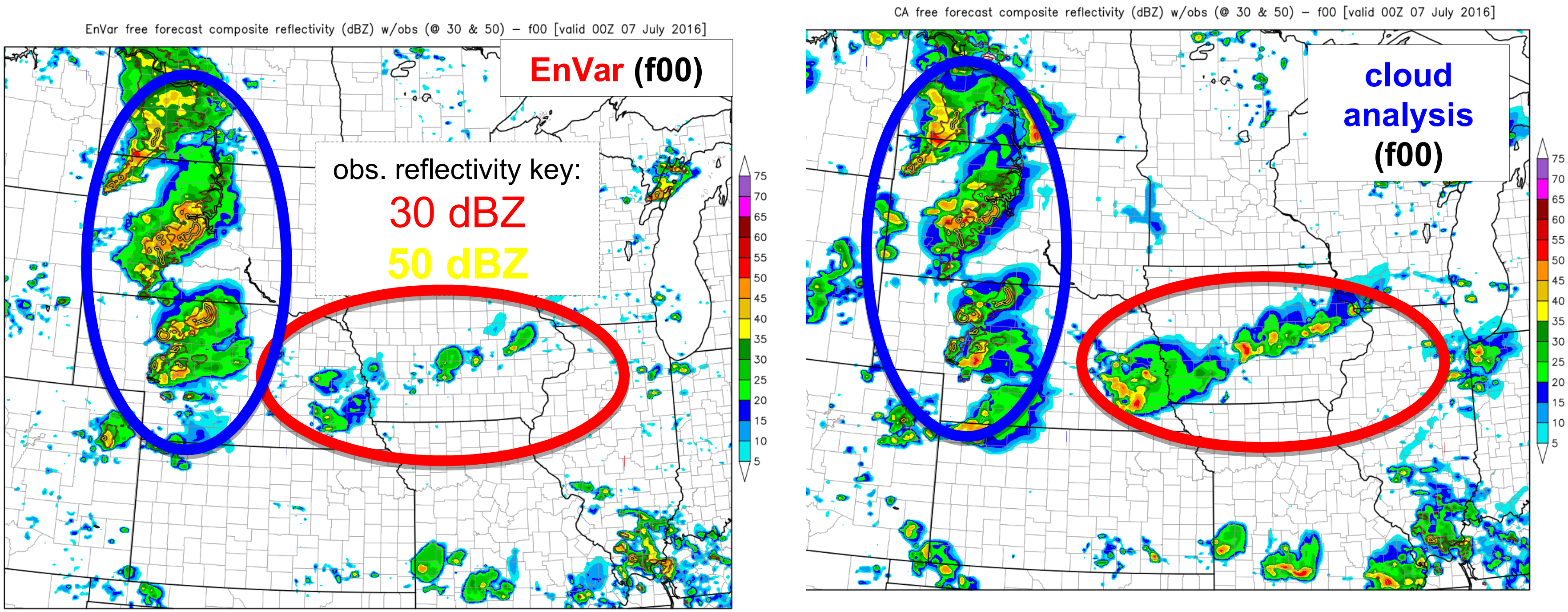


NETS – neighborhood version of ETS using 50 km circular radius

Why GSI EnVar is better than CA?

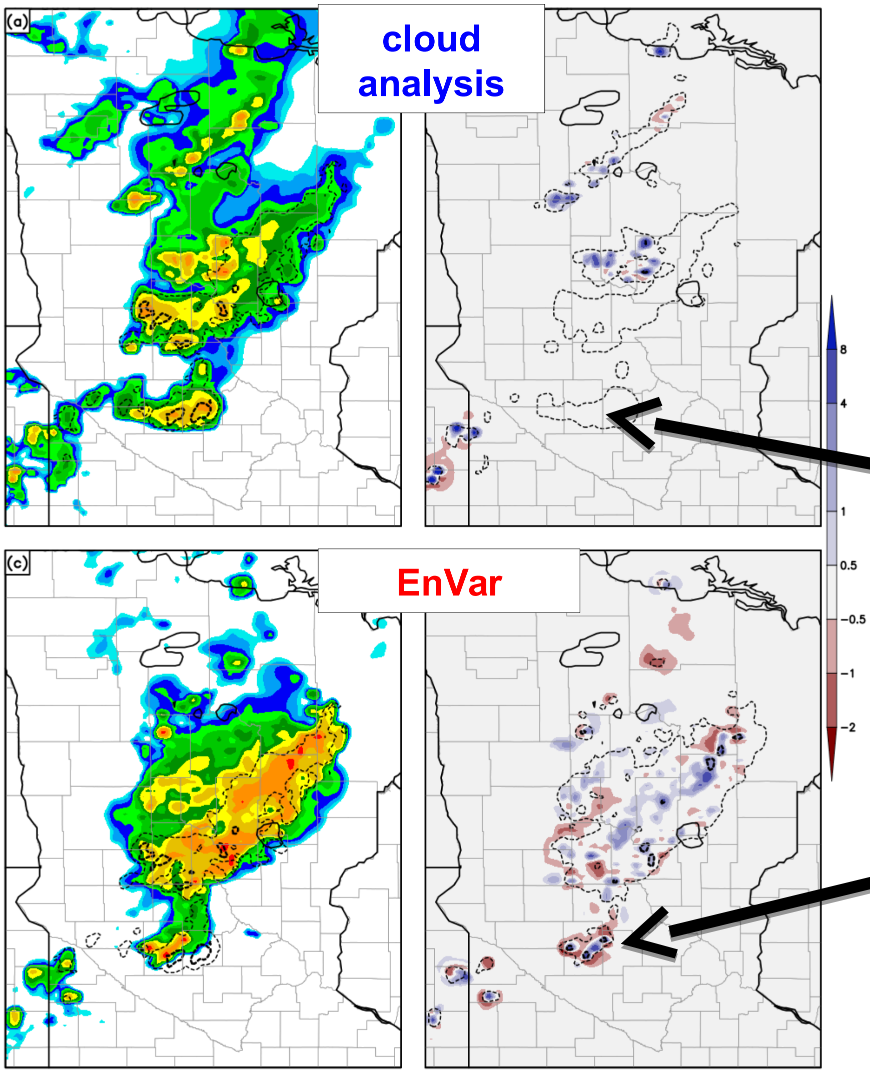
Duda, Wang, Wang, Carley 2018a, MWR

00 UTC 07 July 2016



Spurious convection is better suppressed by the EnVar.

Why GSI EnVar is better than CA?



17 June 2016

- Weak vertical motion from CA
- While CA is able to add reflectivity, it is incapable of providing a consistent update of the dynamical field (e.g., w) itself.

- EnVar can update other thermo/dynamical variables consistently through cross variable correlations

OU MAP **2017 HWT real time CONUS DA and ensemble**

Duda, Wang, Wang, Carley 2018b, MWR

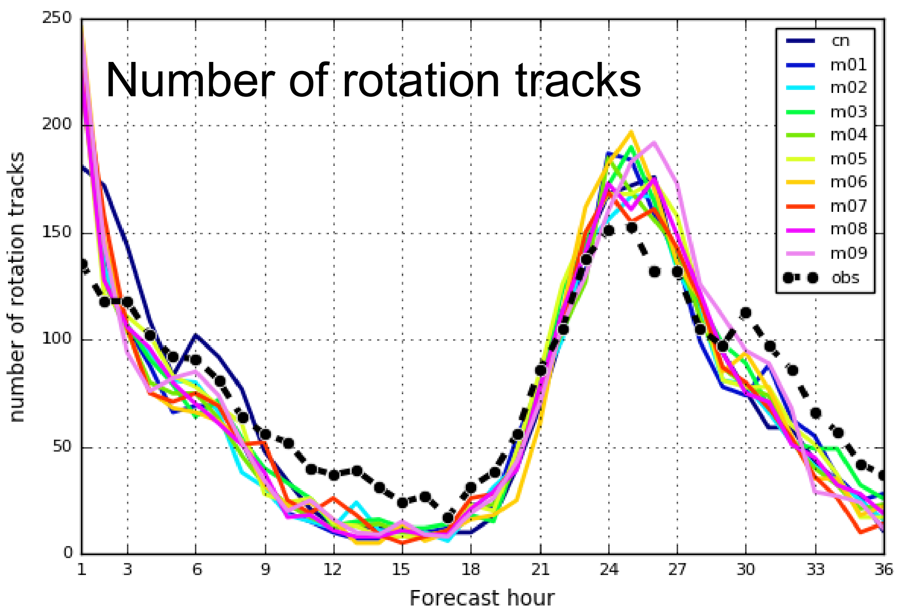
New verification: Object based verification of rotation tracks

- 1-hr rotation tracks

OBS: MRMS RotationTrackML60min product (3-6 km AGL maximum azimuthal shear from Doppler velocity, measured in 10^{-3} s^{-1}), integrated over the past hour
FCST: hourly-maximum UH

- Diurnal cycles of the number of rotation tracks verified well.

- More tracks than observed at early lead times



PART III

GSI-based dual resolution EnVar for sub-kilometer analysis and prediction

Wang Y. and X. Wang 2018, MWR

- Early study has demonstrated the need for ~100m possibly ~10's m grid spacing to fully resolve convective motions (Bryan et al. 2003). Explicitly forecasting of the tornado like vorticies needs to use a finer resolution (e.g., dx<1km).

- State of the art radar provide measurements in very high resolution.

- Most early studies simulate or predict tornado or tornado like vortices by running sub-km model initialized by downscaling a **coarse** resolution analysis (dx>= 1km).

- Is there a need to run DA at finer resolution? What is the impact of initializing with a finer resolution analysis (dx<1km)? Is there a cost effective way to do this?

- Given the large expense of running all ensemble members at sub-kilometers in EnVar, the **dual resolution EnVar** is further extended in GSI where the analysis is produced at sub-kilometer (e.g., 500m) whereas the ingested ensemble is still at lower kilometer resolution (e.g., 2km).

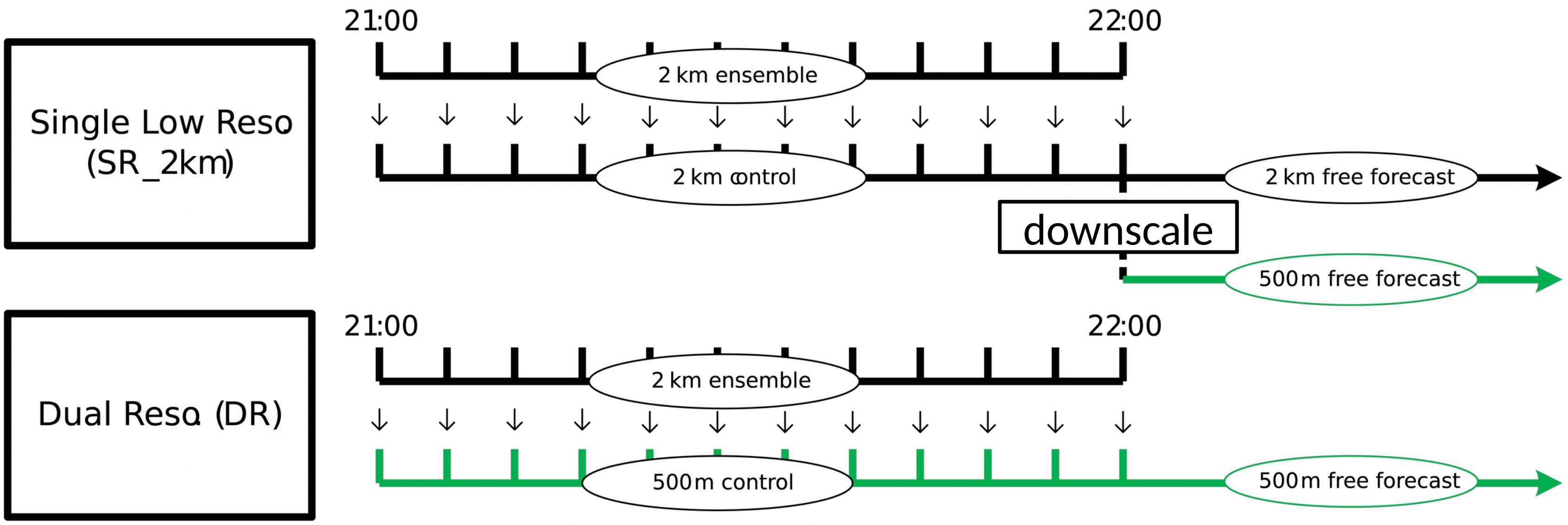
GSI-based dual resolution EnVar formula

- Dual resolution EnVar cost function

$$J(\mathbf{a}) = 0.5(\mathbf{a})^{T}\mathbf{A}^{-1}(\mathbf{a}) + 0.5\left(\mathbf{y}^{o'} - \mathbf{H}(\mathbf{LDa})\right)^{T}\mathbf{R}^{-1}\left(\mathbf{y}^{o'} - \mathbf{H}(\mathbf{LDa})\right)$$

L is the interpolation operator, which is used to interpolate from low-resolution (**LR, e.g. 2km**) to high-resolution (**HR, e.g. 500m**) space.

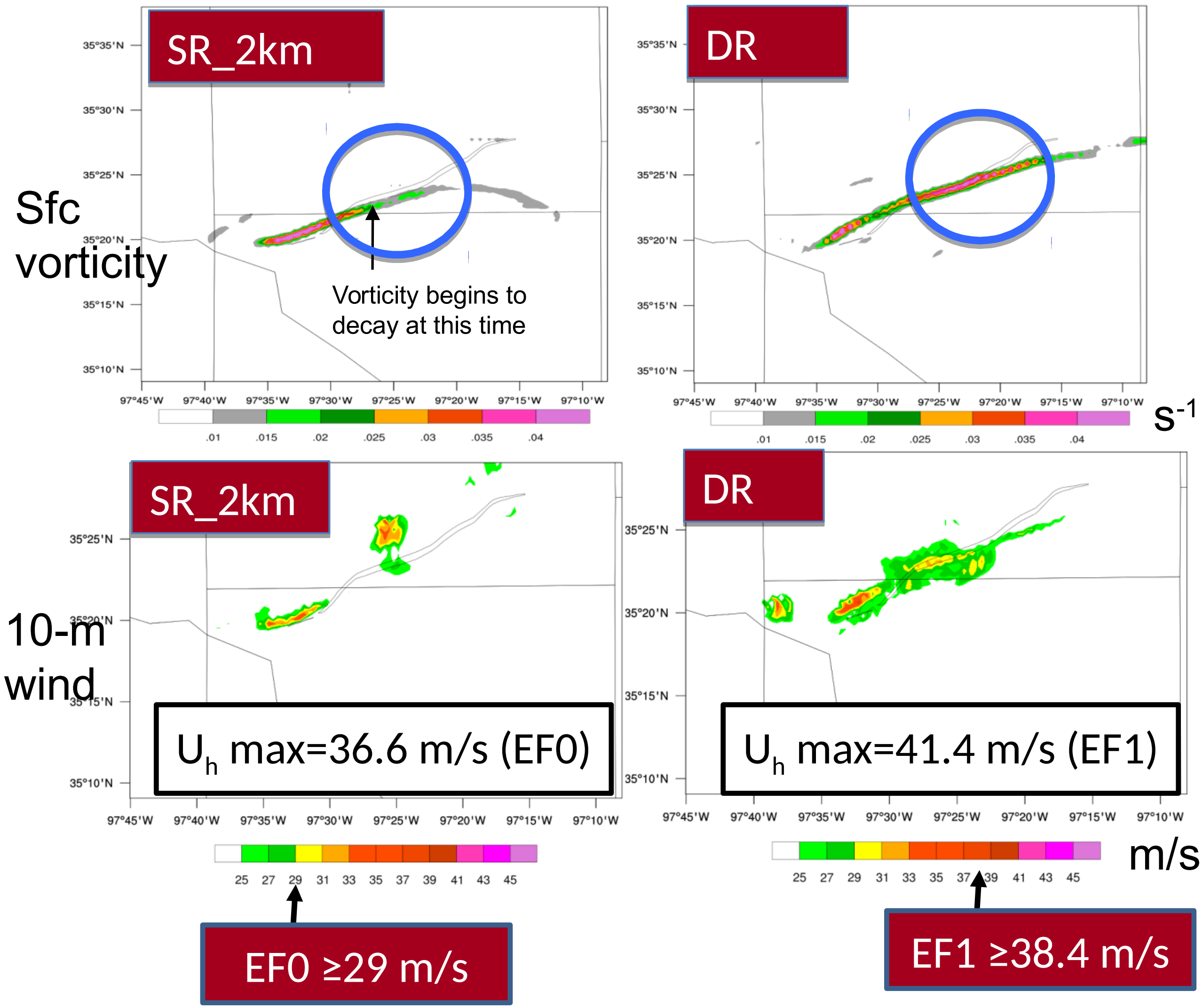
Experiment design

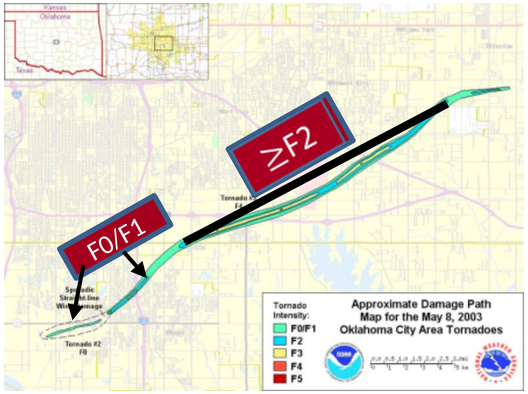


- Both SR_2km and DR ingest the ensemble at 2-km resolution during DA

Experiments	Description
SR_2km	Analysis produced at 2-km resolution ingesting 2-km ensemble. Free forecast at 500 m resolution initialized from downscaled 2-km analysis
DR	Analysis produced at 500-m ingesting 2-km ensemble through dual resolution capability. Free forecast at 500 m resolution

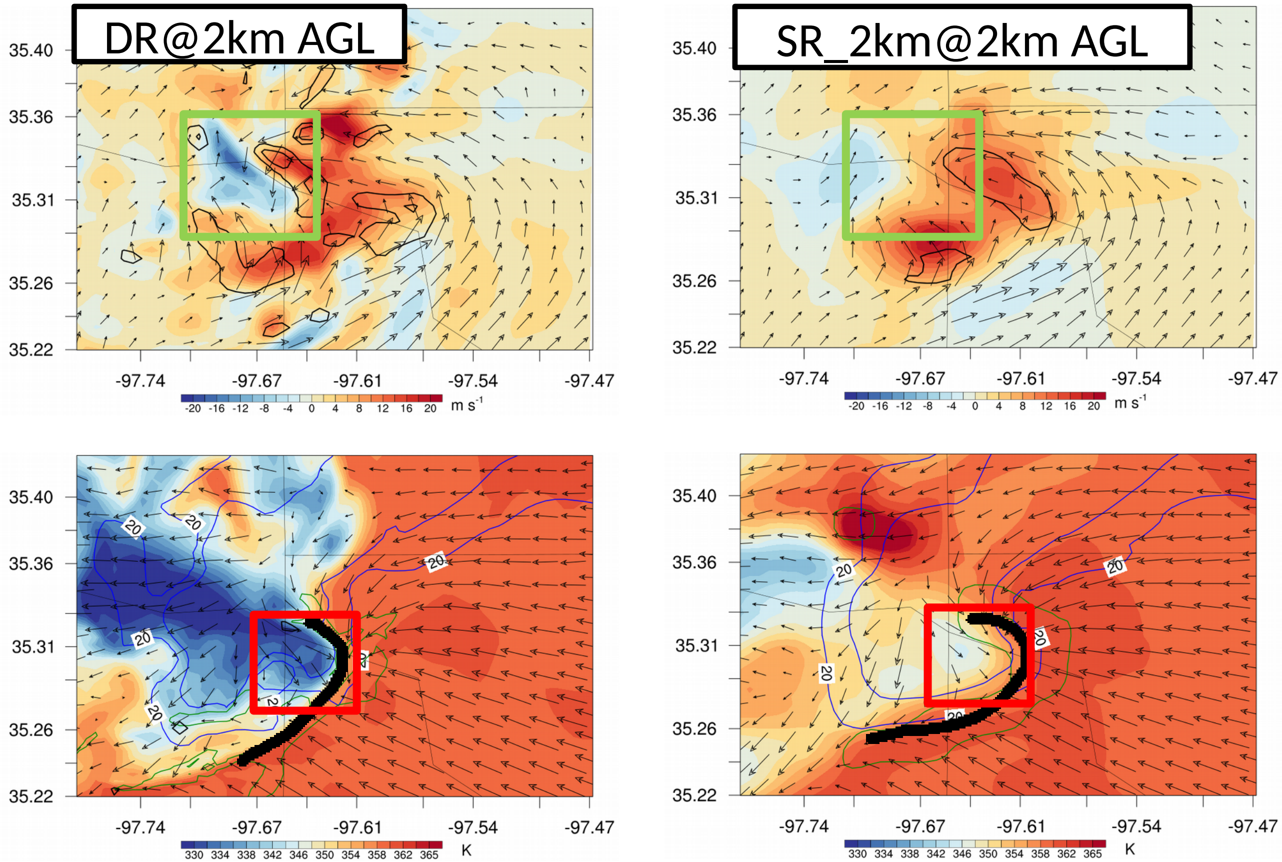
Composite maximum sfc vorticity and 10-m wind improved by dual resolution EnVar





- The predicted vorticity is enhanced after 20-min forecast in DR. Its vorticity evolution is much more consistent with the reality than SR_2km.
- DR is able to predict tornado strength sfc wind with longer duration and greater intensity (≥ EF1).

What are the differences in the final analysis?

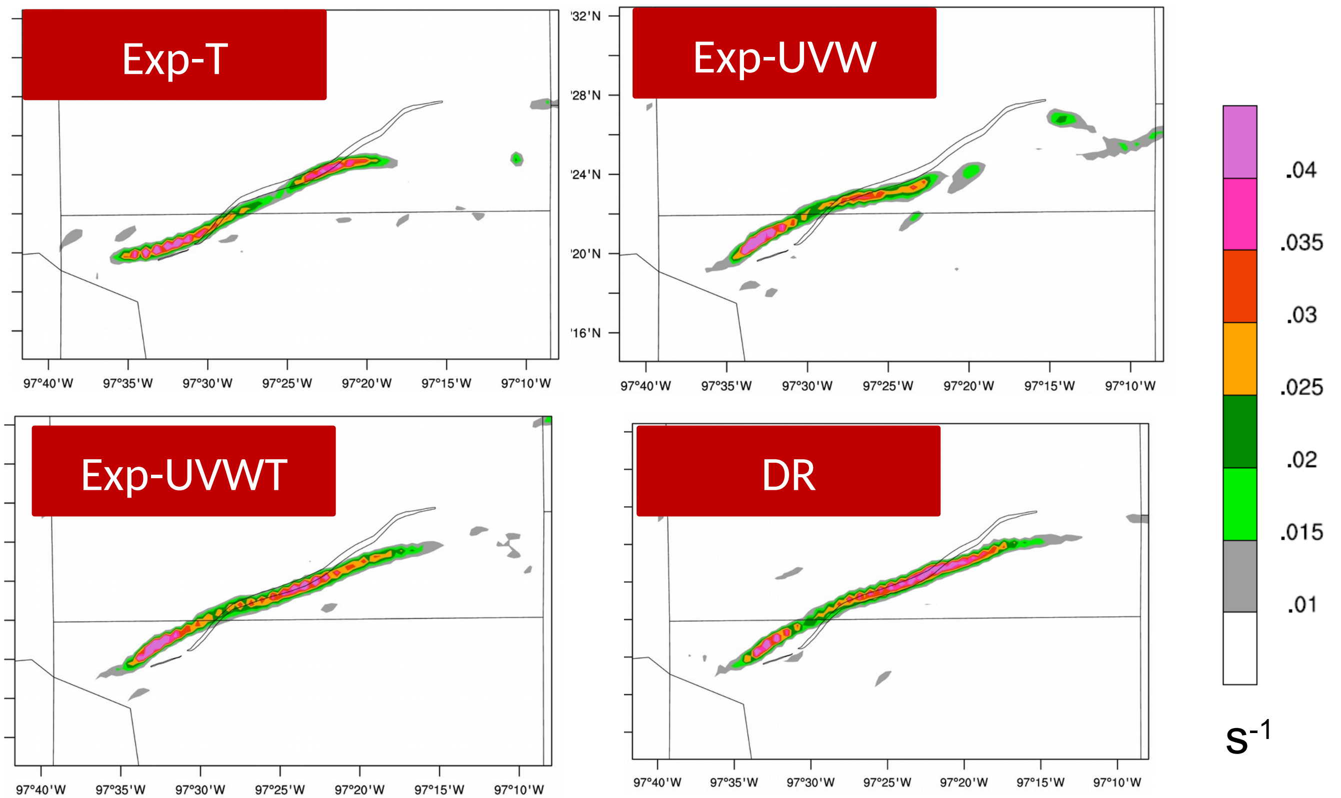


Vertical velocity (shaded) and vertical vorticity (contour) at 2 km AGL

Surface equivalent potential temperature (shaded), reflectivity (blue contour), rear flank gust front (RFGF; black thick line)

- Stronger and broader midlevel downdraft (green box) in DR (left) than SR_2km (right) over the rear-flank region.
- Stronger outflow (red box) surge trailing the RFGF in DR than SR_2km.

How do the differences in the analysis contribute to sfc vorticity forecast difference?

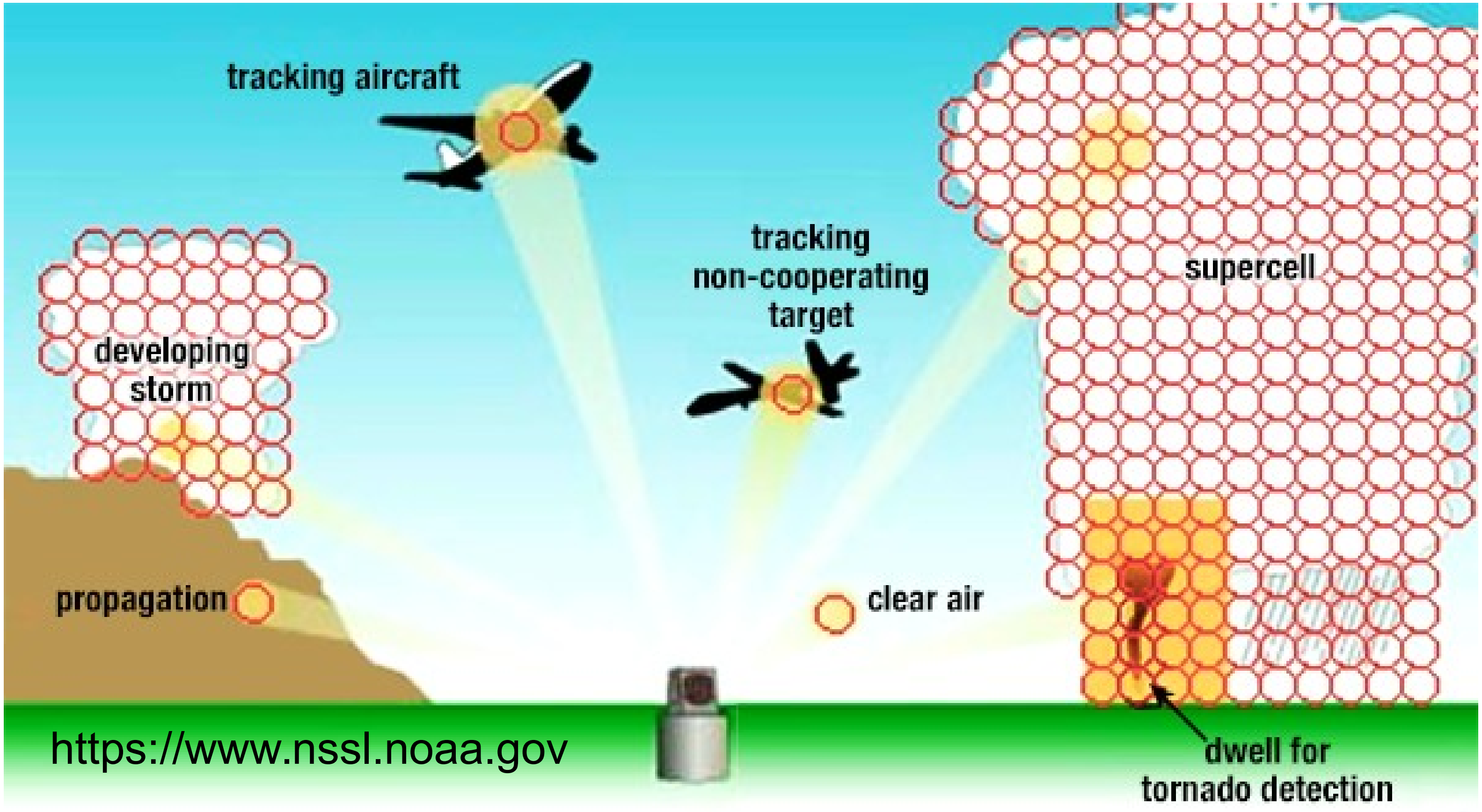


❑ Detailed dynamical diagnostics and sensitivity experiments suggest

- the timing of the weakening and re-intensification is affected by the strength of the downdraft in the analysis
- the longevity and strength of the sfc vorticity is determined by the magnitude of cold pool in the analysis;

PART IV

Multi-function PAR and targeted observations



Compared to the conventional, uniform scanning by WSR-88D radars, the flexibility inherent in MPAR enables “adaptive sampling” or alternatively termed as “targeted observations”

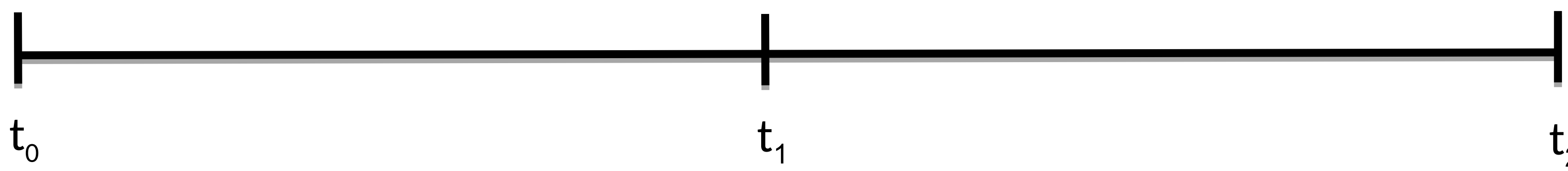
Radar Targeted Observations – push limit for DA!

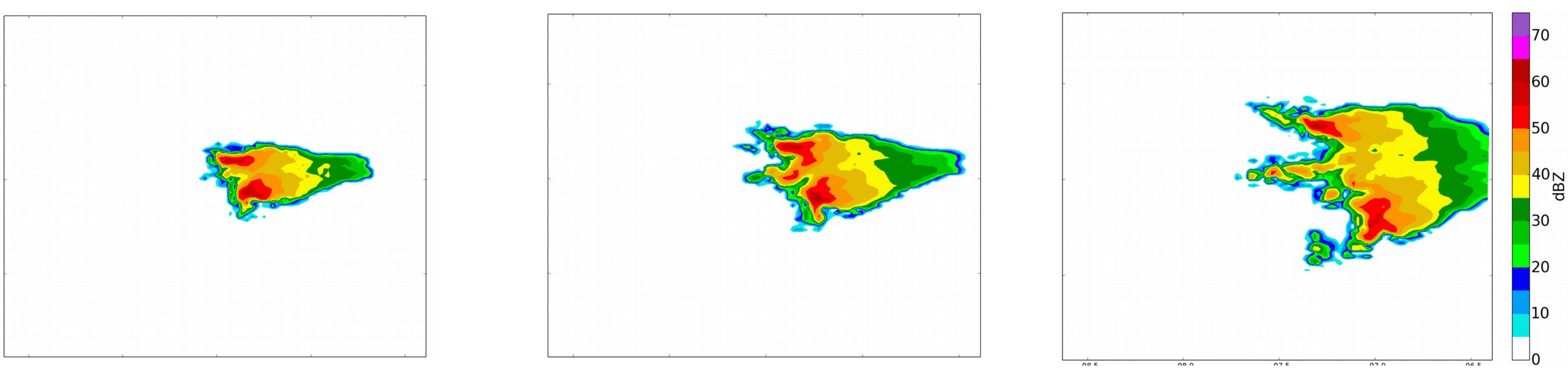
Kerr and Wang 2018

Optimal observation strategy at t_1 is **predicted** at t0

Desired time of observation collection and assimilation

Forecast time of interest



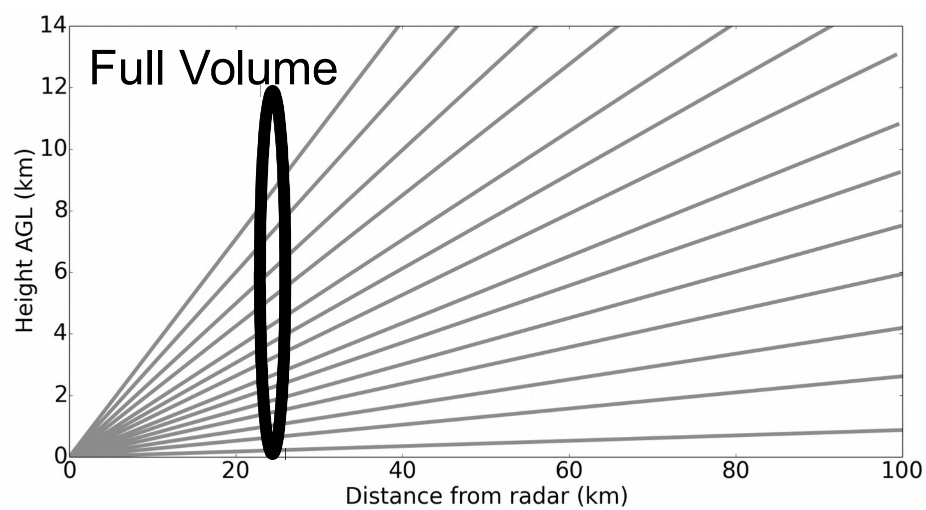


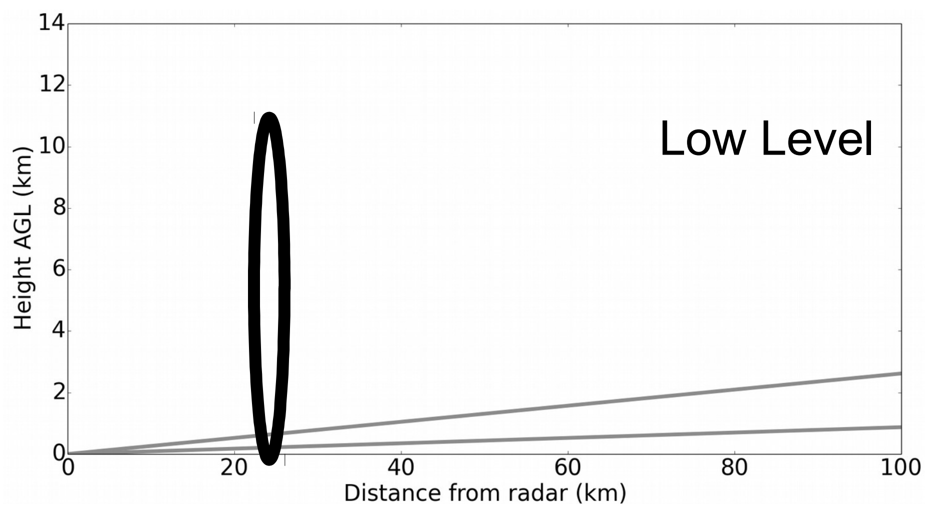
OSSE test-bed: Three Scan Strategies

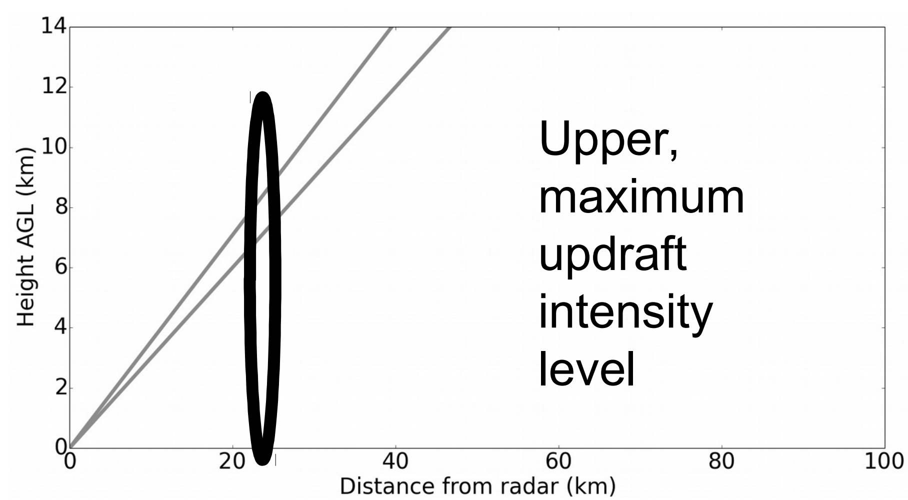
To what extent, assimilating the full volume scan improves upon assimilating partial scans?

For the two partial scans, when/where assimilating one strategy is better than the other?

Would targeting algorithm be able to predict their impacts?







Ensemble targeted observation algorithm

- The algorithm is derived from ensemble DA theory. The idea is to select the observation strategy that is predicted to reduce the forecast error variance the most. One key element is to use ensemble to estimate correlation over time (Bishop et al. 2001; Torn 2014)

- Push the limit for ensemble based radar DA:
  - Nonlinearity
  - Sampling errors to estimate time correlation

$$\delta\sigma_J = -\boldsymbol{J}\,(\boldsymbol{Hx})^{\mathrm{T}}\left(\boldsymbol{HP}^{\boldsymbol{b}}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}\right)^{-1}\boldsymbol{Hx}\,\boldsymbol{J}^{\mathrm{T}}$$

$\delta\sigma_J$: reduction of forecast error variance

$\boldsymbol{J}$: ensemble perturbation forecast metric of interest

$\boldsymbol{Hx}$: ensemble perturbation of observation priors

$\boldsymbol{P}^{\boldsymbol{b}}$: background error variance

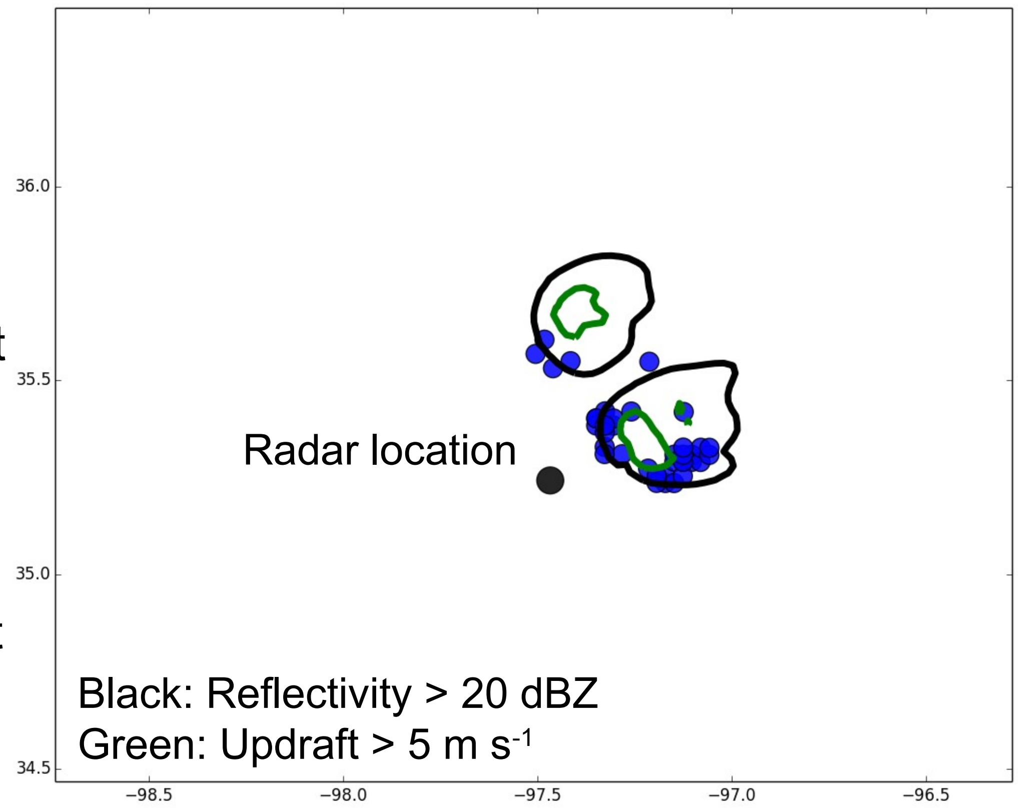
$\boldsymbol{R}$: observation error variance

Use ensemble to predict if/how observations will impact a forecast metric

Forecast metric: 30min 0-1km UH forecast

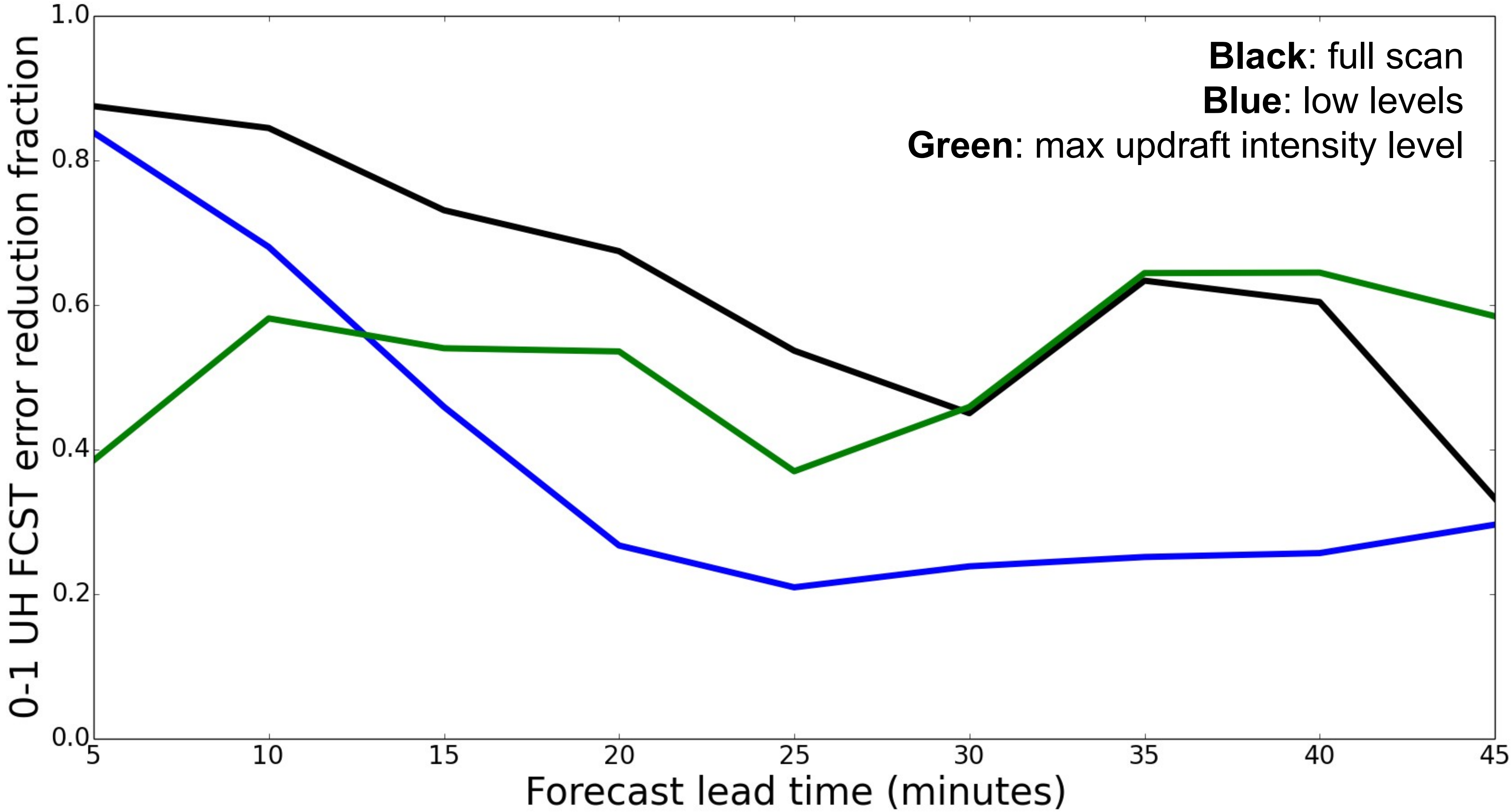
Blue dots represent locations of radial velocity observations from a max updraft velocity region scan that are determined by the method as being impactful

Note how the observations are cluttered along the edges of the right mover updraft "tangential" to radar beam



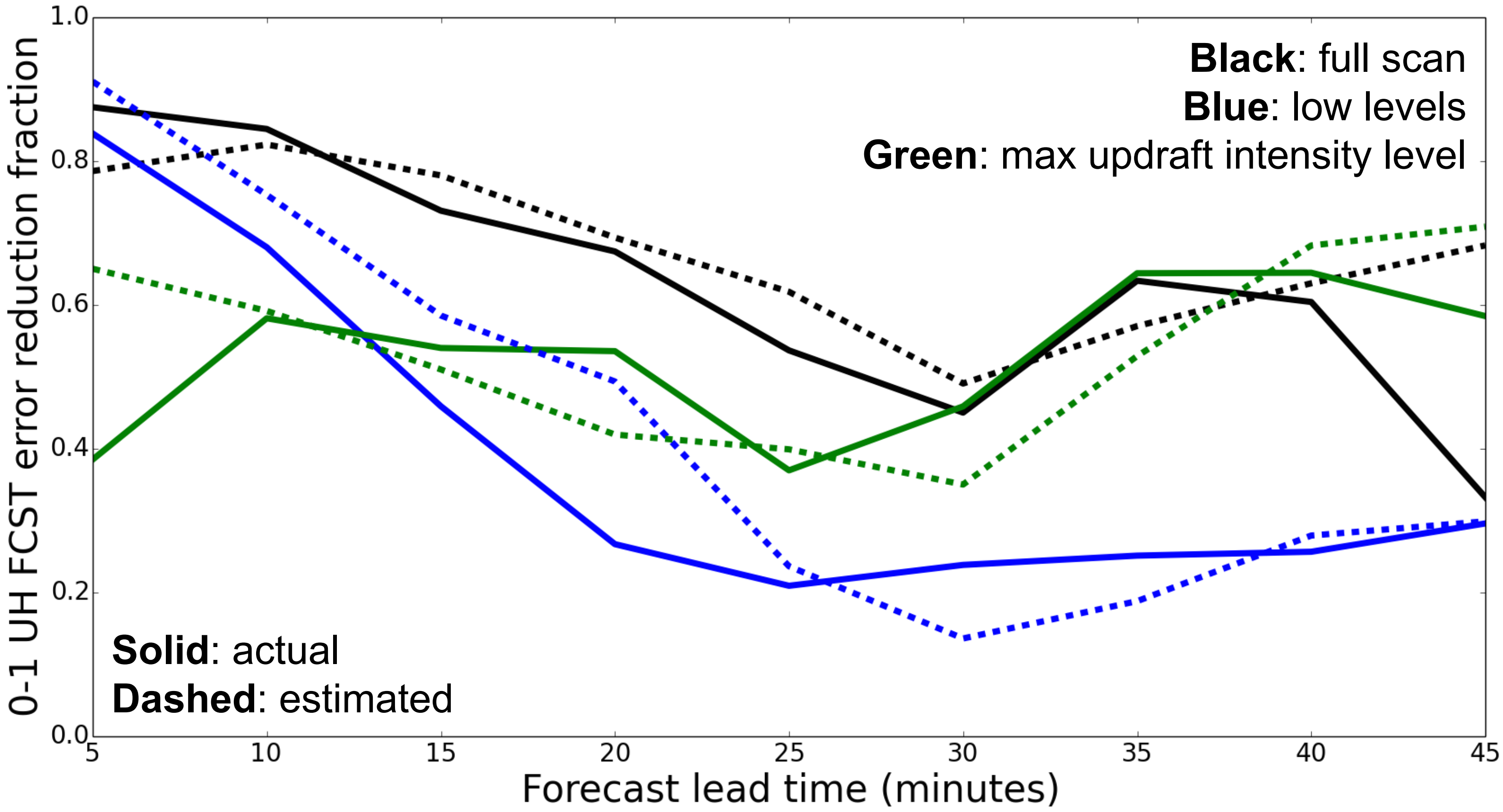
Actual error variance reduction

This shows how various observation sets affect the forecast metric, 0-1 km updraft helicity ("actual" error variance reduction)



How well does the targeting algorithm predict observation impacts?

At t_0, do we have an idea of these future observation impacts?



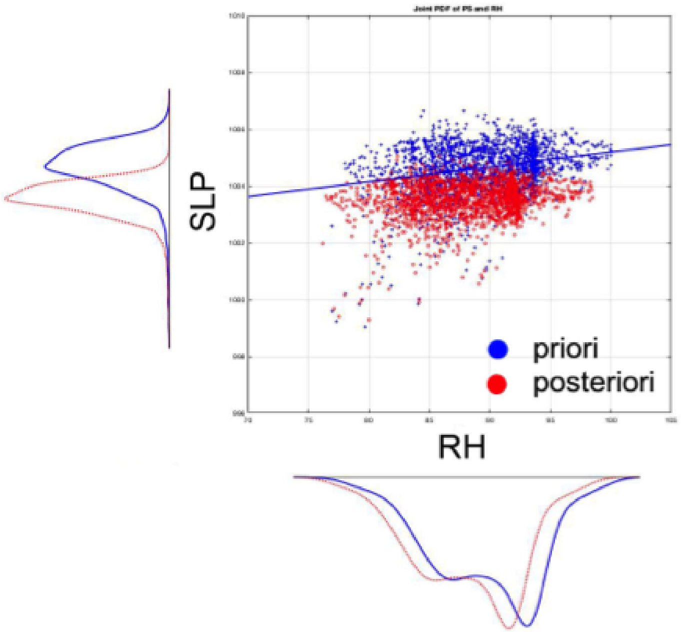
Summary and Remarks

- For direct reflectivity assimilation in EnVar, a method without tangent linear (TL) and adjoint of the nonlinear operator is developed to solve the issues associated with the TL of the reflectivity operator in EnVar.
- Idea maybe useful for observations with complicated operators where TLA may not be easy to develop or observation operators sharing similar issue as described here (e.g. space radar, dual pol variables, precipitation).
- With this approach, 4DEnVar is not only TLA free for forecast model, but also TLA free for nonlinear obs. operator.
- Issue is specific for Var, not applied for EnKF.
- Idea of extending state to include observed variables analogous to state augmentation of parallel implementation of serial EnKF but for addressing different issues.
- Experiment with the May 8 tornadic supercell case shows that strong updraft and vorticity are better maintained using the new method than using hydrometeor mixing or log transformed hydrometer mixing ratio as state variables.
- The method is implemented in operational HRRR and NAM-CONUS and found to improve precipitation forecast as compared to the operational cloud analysis.
- Sub-km analysis is useful and dual-resolution GSI-EnVar provides a cost effective means: Dual resolution GSI EnVar is further extended for sub-km analysis, which is found to be critical on the timing of weakening and re-intensification, and on the longevity and strength of the TLV for the May 8 case.
- Real time targeted observations for radar and convective scale NWP is possible.

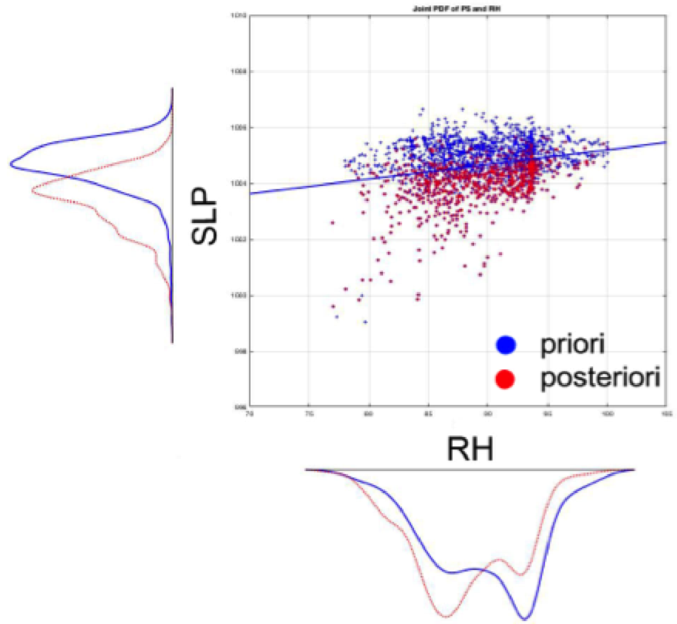
Ongoing and future work

- Research and development on multi-scale data assimilation.
- Complementary assimilation of GOES-R cloudy radiances and ground based radar observations.
- Continue research and development to treat nonlinearity issue.

EnKF



Particle Filter



Kay and Wang, 2018, MWR

References

Convective scale radar DA

Wang, Y. and X. Wang, 2017: Direct Assimilation of Radar Reflectivity without Tangent Linear and Adjoint of the Nonlinear Observation Operator in GSI-Based EnVar System: Methodology and Experiment with the 8 May 2003 Oklahoma City Tornadic Supercell, *Mon. Wea. Rev*., 145, 1447- 1471.

Wang Y., X. Wang, and J. Carley, 2018: GSI-based EnKF-Variational Hybrid Data Assimilation for NCEP NAMRR: System Development and Initial Testing. *Mon. Wea. Rev.* submitted.

Duda, J., X. Wang, Y. Wang and J. Carley, 2018: Comparing GSI-based ensemble-variational method for direct radar reflectivity assimilation with cloud analysis in convection-allowing forecasts over the continental US. Mon. Wea. Rev., submitted.

Duda, J., X. Wang, Y. Wang and J. Carley, 2018: Verification of a CONUS Storm-Scale Ensemble using an Ensemble-Variational Data Assimilation Scheme to directly assimilate multiscale in-situ and radar observations Participating in the 2017 Hazardous Weather Testbed Spring Forecasting Experiment. Mon. Wea. Rev., submitted

Wang Y., and X. Wang, 2018: Impact of the analysis resolution on the prediction of tornado like vorticity and surface wind using the dual resolution GSI-based EnVar data assimilation system: methodology and experiment with the May 8th 2003 Oklahoma City tornado case. Mon. Wea. Rev., to be submitted.

Kerr, C. and X.Wang, 2018: Ensemble based targeting observation applied for multi-function phased array radar. Mon. Wea. Rev., to be submitted.

Global DA

Wang, X., 2010: Incorporating ensemble covariance in the Gridpoint Statistical Interpolation (GSI) variational minimization: a mathematical framework. *Mon. Wea. Rev*., 138, 2990-2995.

Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP Global Forecast System: single resolution experiments. *Mon. Wea. Rev.,* 141, 4098-4117.

Wang, X. and T. Lei, 2014: GSI-based four dimensional ensemble variational data assimilation (4DEnsVar): formulation and single resolution experiments with real data for NCEP GFS. *Mon. Wea. Rev*., 142, 3303-3325.

Huang, B., and X. Wang, 2018: On the use of cost effective valid time shifting ensembles to increase ensemble size in the GFS hybrid 4DEnVar system. Mon.Wea. Rev., submitted

Hurricane DA

Lu, X., and X. Wang, M. Tong and V. Tallapragada, 2017: GSI-based, fully cycled, dual resolution hybrid ensemble-variational data assimilation system for HWRF: system description and experiment with Edouard (2014), *Mon. Wea. Rev.,* 145, 4877-4898.

Lu, X., X. Wang, Y. Li, M. Tong and X. Ma, 2016: GSI-based ensemble-variational hybrid data assimilation for HWRF for hurricane initialization and prediction: impact of various error covariances for airborne radar observation assimilation. *Q. J. R. Meteo. Soc.*, 143, 223-239.

Backup slides

How well does the targeting algorithm predict observation impacts?

Scatter plot illustrates trend in actual error reduction with estimated error variance reduction for various t_1 and t_2

Results suggest the targeting algorithm capable of distinguishing low impact strategy vs high impact strategy

Further testing parameters and increasing sample size