Challenges and Advances of Ensemble-Variational (EnVar) Hybrid Data Assimilation for Convective Scale Weather Prediction

Xuguang Wang

Multiscale data Assimilation and Predictability (MAP) Lab
School of Meteorology
University of Oklahoma, Norman, OK, USA

xuguang.wang@ou.edu

http://weather.ou.edu/~map

Acknowledgement:
*OU MAP students and early career scientists
NOAA EMC, ESRL, GSD, NSSL, HRD colleagues

CIMMS Predictability Workshop, University of Oklahoma, Norman, OK
Oct. 31-2, 2018

Advances of EnVar hybrid DA

2000

Development of theory

Research with simple model and simulated data

System development for real NWP model and test real data

2010

Operational implementation at NWP centers for global NWP, US NWS, Env. Canada, US Navy, UK Met, ECMWF

Active R&D and operational implementation for convective scale NWP

Theory/algorithm development

- Combining static and ensemble covariance in variational framework (Hamill and Snyder 2000)
- Extended control variable (ECV) method (Lorenc 2003; Wang et al. 2007b, 2008a; Wang 2010, etc.)
- Proved equivalence of ECV to direct combination of static and ensemble covariances (Wang et al. 2007b)
- 4D extension (Tian et al. 2008; Liu 2008; Buehner 2010)

Advances of EnVar hybrid DA

2000

2010

Development of theory

Research with simple model and simulated data

System development for real NWP model and test with real data

Operational implementation at NWP centers for global NWP, US NWS, Env. Canada, US Navy, UK Met, ECMWF

Active R&D and operational implementation for convective scale NWP

Simple model studies: e.g.
Zupanski 2005
Wang et al 2007a, 2009

Early development of EnVar for real regional NWP models:
Wang et al. 2008ab
Wang 2011
Li* et al. 2012
Zhang and Zhang 2012

Early development of EnVar for real global model:
Buehner 2005
Buehner et al. 2010
Bishop and Hodyss 2011
Wang et al. 2013

These studies show hybrid combines the best aspects of EnKF and Var (summarized in Wang 2010)

Advances of EnVar hybrid DA

2000

Development of theory

Research with simple model and simulated data

System development for real NWP model and test real data

2010

Operational implementation at NWP centers for global NWP, US NWS, Env. Canada, US Navy, UK Met, ECMWF

Active R&D and operational implementation for convective scale NWP

E,g.

US NWS
Wang 2010
Wang et al. 2013 (with Parrish, Kleist, Whitaker)
Wang and Lei 2014
Kleist and Ide 2015

US Navy
Kuhl et al 2013

Env. Canada
Buehner et al. 2010ab

UK Met
Clayton et al. 2013

etc

Advances of EnVar hybrid DA

2000

Development of theory

Research with simple model and simulated data

System development for real NWP model and test real data

2010

Operational implementation at NWP centers for global NWP, US NWS, Env. Canada, US Navy, UK Met, ECMWF

Active R&D and operational implementation for convective scale NWP

Advances of GSI EnVar for convective scales over CONUS

- Develop a algorithm to enable direct assimilation of radar reflectivity for GSI EnVar (Wang* and Wang 2017)
- Demonstration for 3 convective scale applications HRRR, NAM-CONUS, WoF (Wang* and Wang 2017, Wang* et al. 2018, Duda* et al. 2018, Wang* and Wang 2018ab)
- GSI EnVar for sub-kilometer DA (Wang* and Wang 2018a)
- Extend static covariance for convective scale EnVar hybrid (Wang* and Wang 2018b)

Work based on other systems: Li* et al. 2012, Caron et al, 2018, Kong et al. 2018 and Gao et al. 2016

Advances of EnVar hybrid DA

2000

Development of theory

Research with simple model and simulated data

System development for real NWP model and test real data

2010

Operational implementation at NWP centers for global NWP, US NWS, Env. Canada, US Navy, UK Met, ECMWF

Active R&D and operational implementation for convective scale NWP

GSI EnVar for convection allowing hurricane prediction

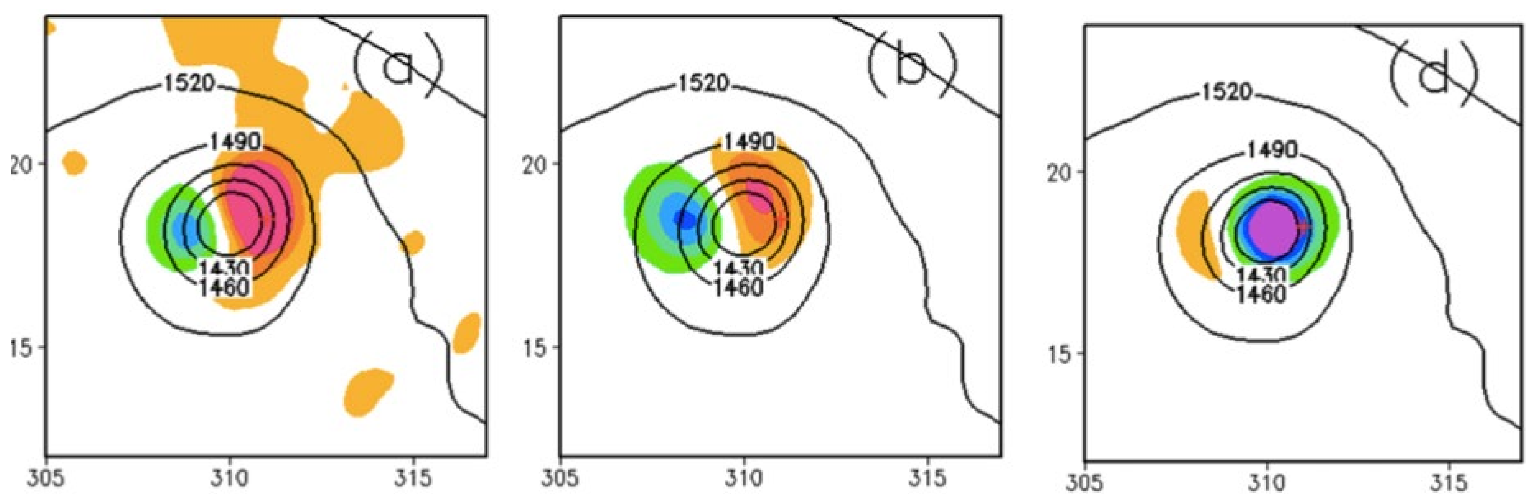
- Developed fully cycled GSI EnVar DA system for US operational convection allowing hurricane prediction system HWRF.
- Lu*, Wang, Tong and Tallapragada, 2017, MWR
- Lu*, Wang, Li*, Tong, Ma, 2016, QJRMS
- Operational implementation for HWRF since summer 2017
- Improve the assimilation and study the impact of variety in-situ/remote sensing inner core observations (Lu* and Wang 2018a)
- Reveal model physics errors (Lu* and Wang 2018b)

An example of operational implementation of EnVar for global NWP

- GSI-based 3DEnVar and 4DEnVar hybrid data assimilation system was operationally implemented for GFS at US NCEP in 2012 and 2016. Significant improvement was found for global analysis and forecasts (Wang et al. 2013; Wang and Lei* 2014; Kleist and Ide 2015ab) .

Example from Wang and Lei* 2014, MWR

Non-linear model integration | **4DEnVar** | **3DEnVar**



Challenges for convective scale data assimilation

- Require unique observation operators that are often complex and nonlinear (e.g., reflectivity, Dual pol radar variables, cloudy radiances)
- Both prior (e.g. hydrometeors) and observation errors are highly non-Gaussian
- Accurate cross-variable covariance is especially important
- Balance assumption in covariance for large scales do not fit any more
- Heavily rely on quality of numerical models (microphysics schemes, PBL schemes, etc.) – treatment of model errors is critical
- Observations can be in much higher spatial resolution than the typical NWP model and in much higher temporal resolution than typical DA frequency.
- 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.

Direct assimilation of radar reflectivity in GSI EnVar and demonstration in WoF and HRRR/NAM applications

Add direct radar DA capability

GSI-based 3/4DEnVar Hybrid

WoF (<=1hr)
e.g. tornadic supercell

HRRR (<=18h)
Updated hourly

NAM CONUS (<=60h)
Updated 6-hourly

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)

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

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

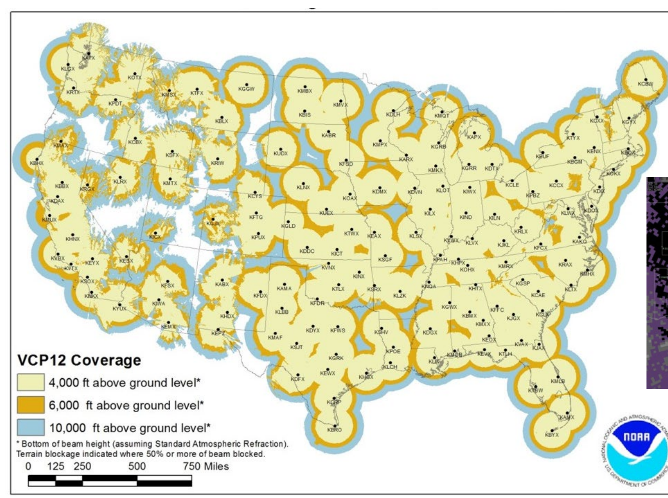
$$\mathbf{x}' = \sum_{k=1}^{K}(\mathbf{a}_k \circ \mathbf{x}_k^{e})$$

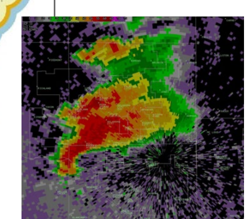
- Nonlinear radar reflectivity operator

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

$$Z_e = Z_r + Z_s + Z_g$$

$$Z_g = 4.33\times10^{10}(\rho q_g)^{1.75}$$

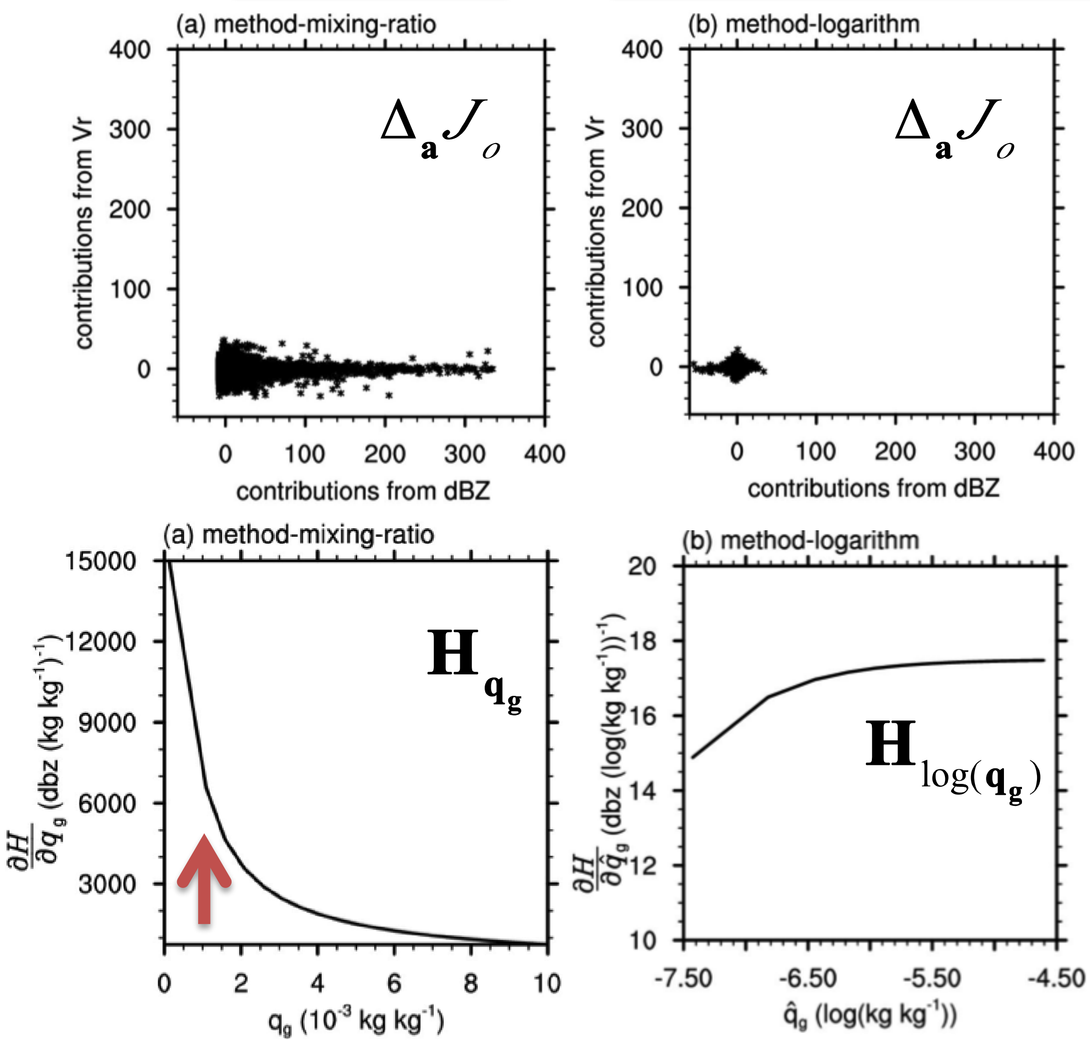




Issue with TL of nonlinear reflectivity operator in EnVar

Wang* and Wang 2017, MWR

$$\boxed{H(q_s, q_r, q_g)} \qquad \boxed{H(\log(q_s), \log(q_r), \log(q_g))}$$



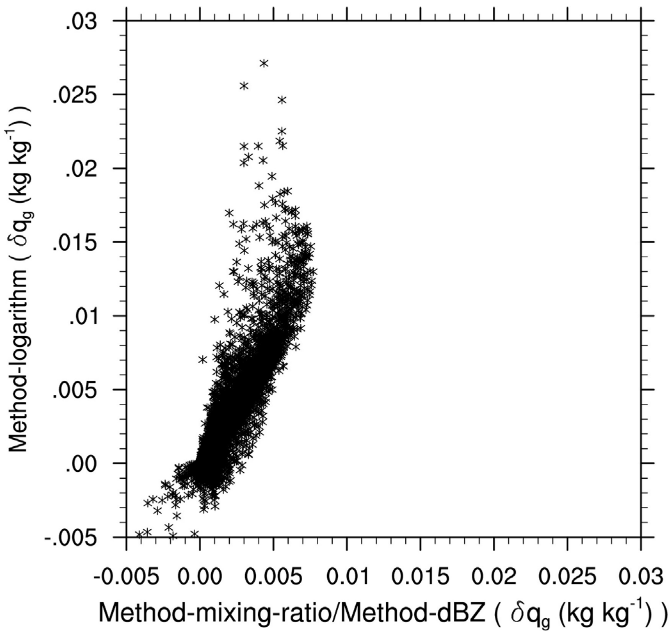
- When hydrometeor mixing ratio is used as state variables, large values of TL of the nonlinear reflectivity associated with the small hydrometeor mixing ratios lead to large differences of cost function gradients, which **prevents efficient convergence and therefore under-estimates the hydrometeor increments.**
- Using logarithm of hydrometeor mixing ratio as state variable fixes this issue, but incurs additional issues.

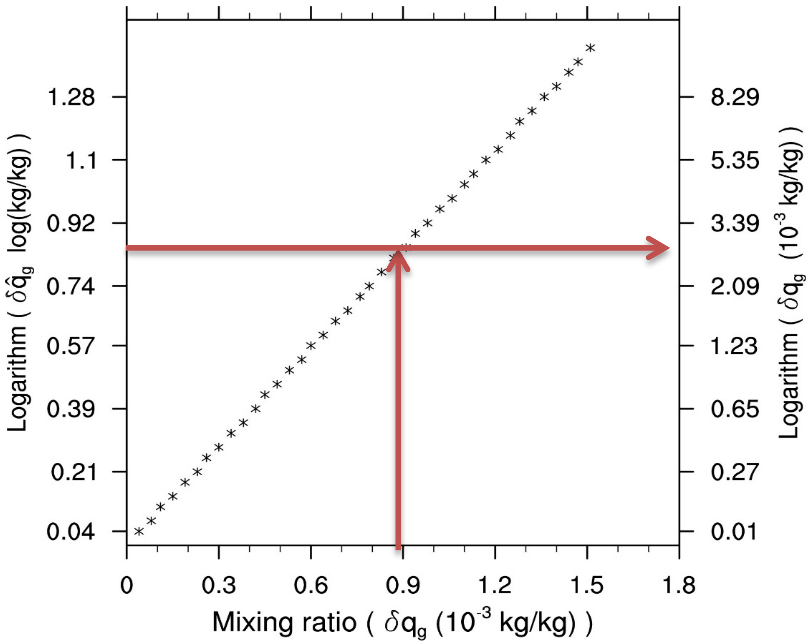
Issue with TL of nonlinear reflectivity operator in EnVar

Wang* and Wang 2017, MWR

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

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



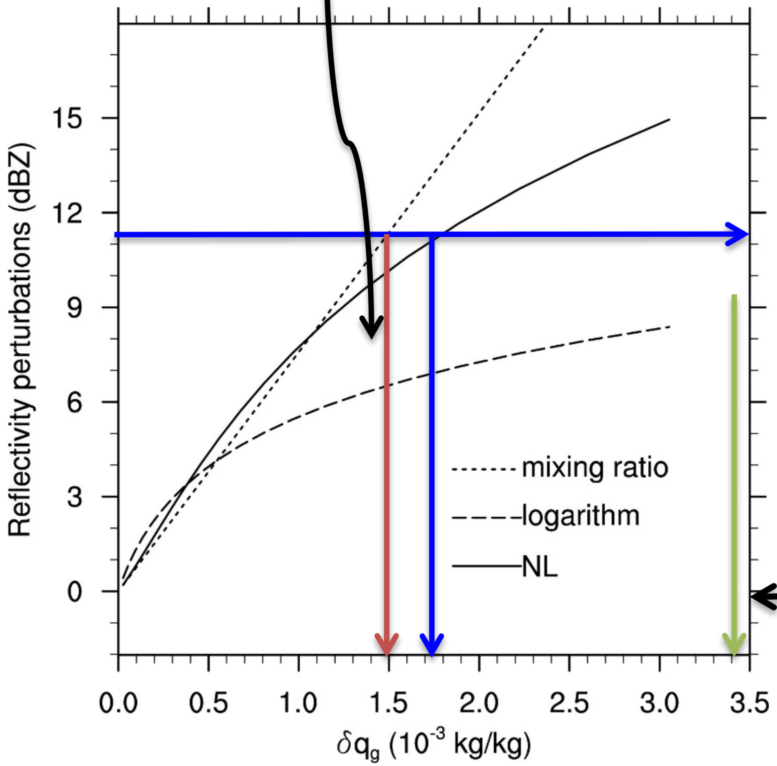


➢ However, it produces **anomalously large hydrometeor increment** partly due to the transform to and from the logarithmic space

Issue with TL of nonlinear reflectivity operator in EnVar

Wang* and Wang 2017, MWR

$\boxed{H(q_s, q_r, q_g)}$



$\boxed{H(\log(q_s), \log(q_r), \log(q_g))}$

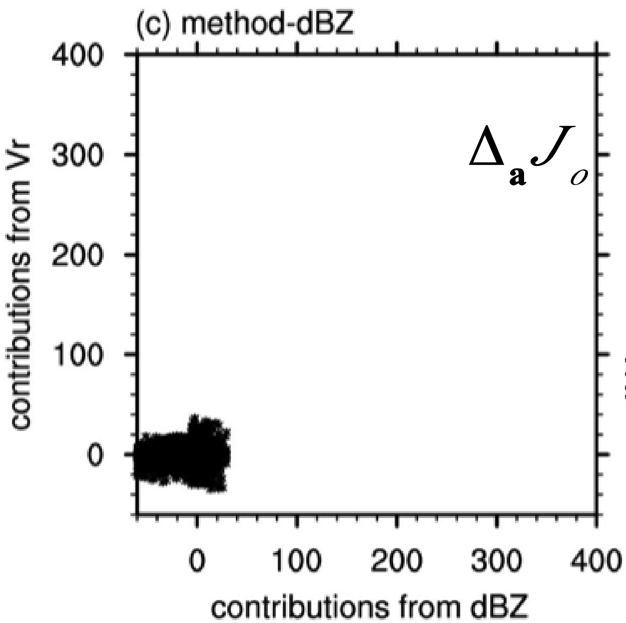
$$\Delta \mathbf{y} = H(\mathbf{x} + \Delta \mathbf{x}) - H(\mathbf{x}) = \mathbf{H}\Delta \mathbf{x}$$

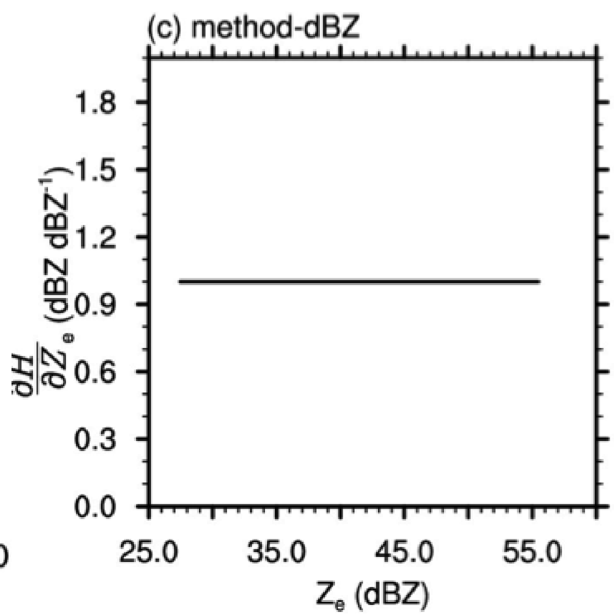
- The underestimation and overestimation of hydrometeor increments are **exacerbated by the TL assumption** of the nonlinear reflectivity operator itself.

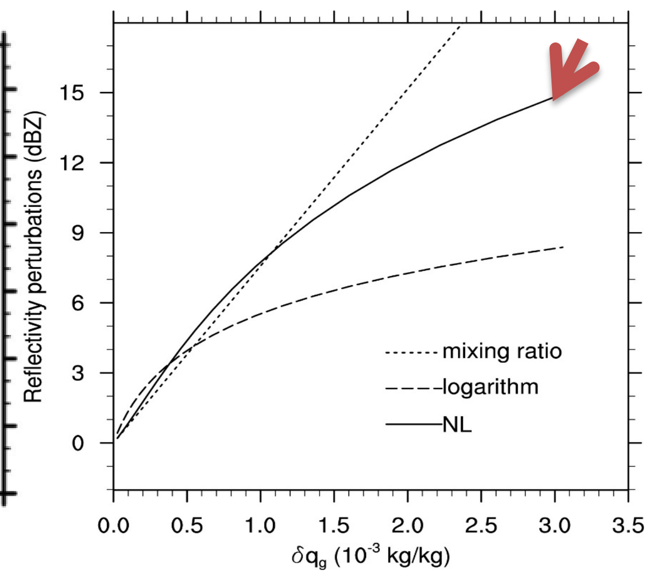
GSI-based EnVar without tangent linear (TL) and adjoint of the nonlinear reflectivity operator

Wang* and Wang 2017, MWR

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

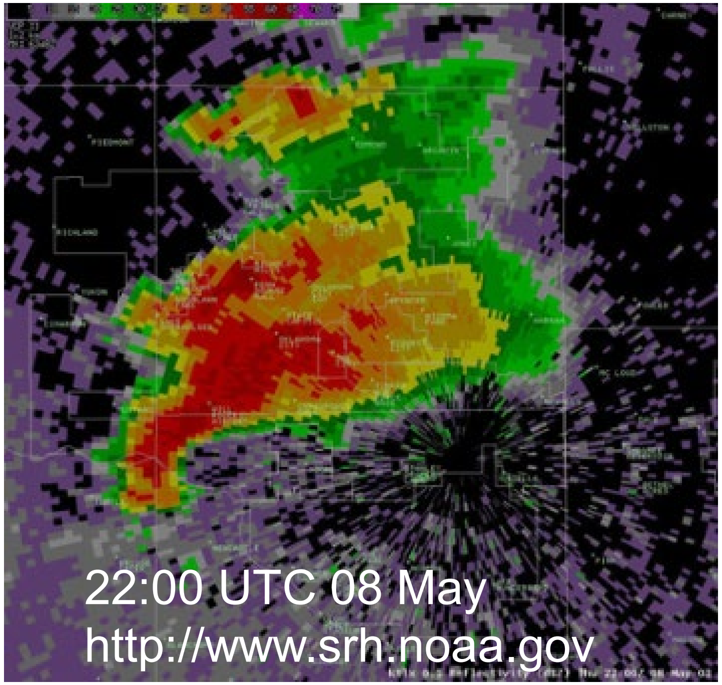


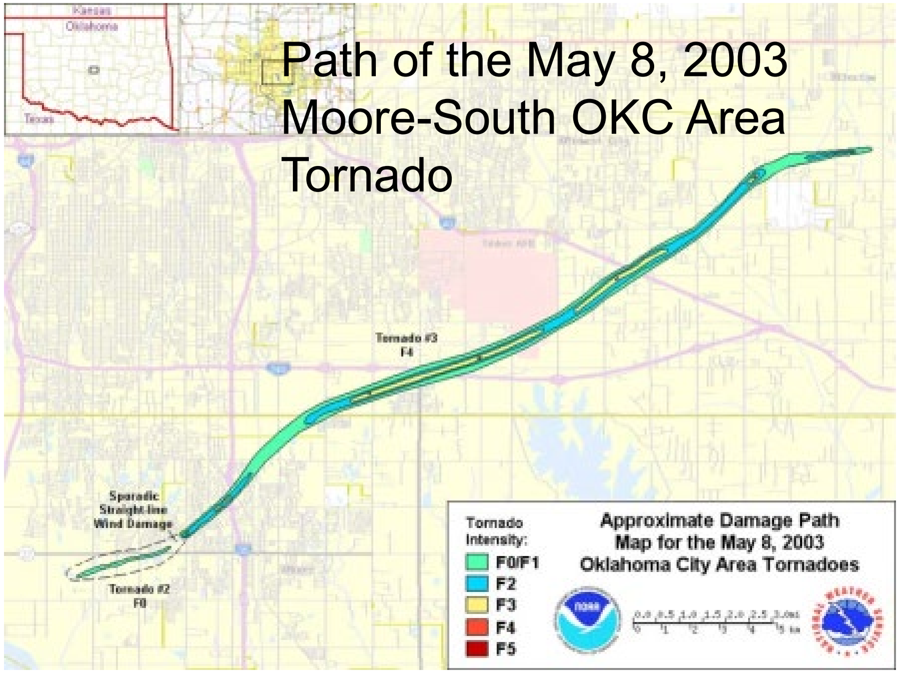




- Gradient issues fixed
- In this method, no TL of the reflectivity operator explicitly exists in variational minimization. Hydrometeor is related to reflectivity following the nonlinear relationship.

May 8th 2003 OKC Tornadic Supercell

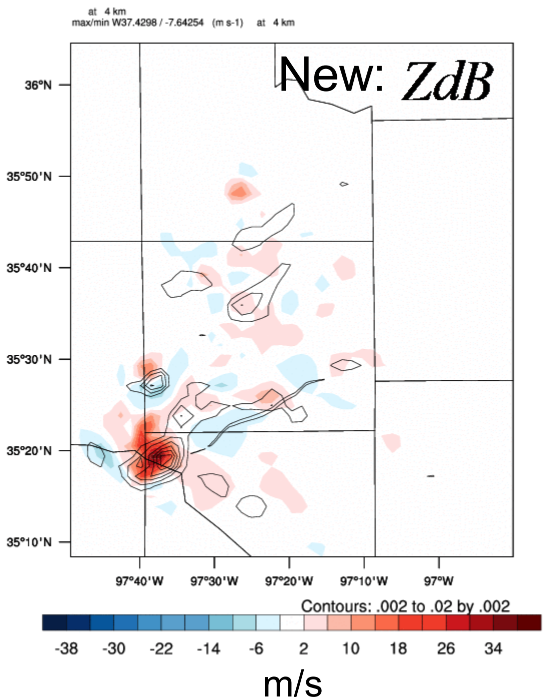




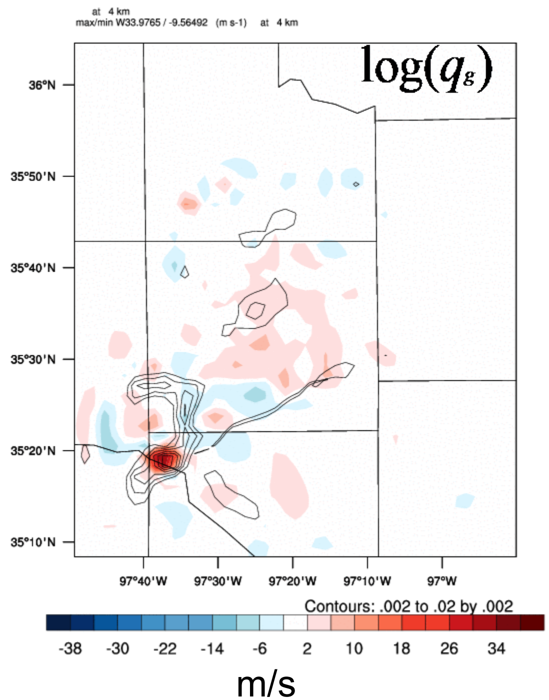
- An isolated supercell case that produced F-4 intensity tornadoes in Moore and Oklahoma City (OKC) during about 2210—2240 UTC.
- Supercell maintained well beyond 2300 until about 0000 UTC.

1 hour forecast: w and vorticity at 4km

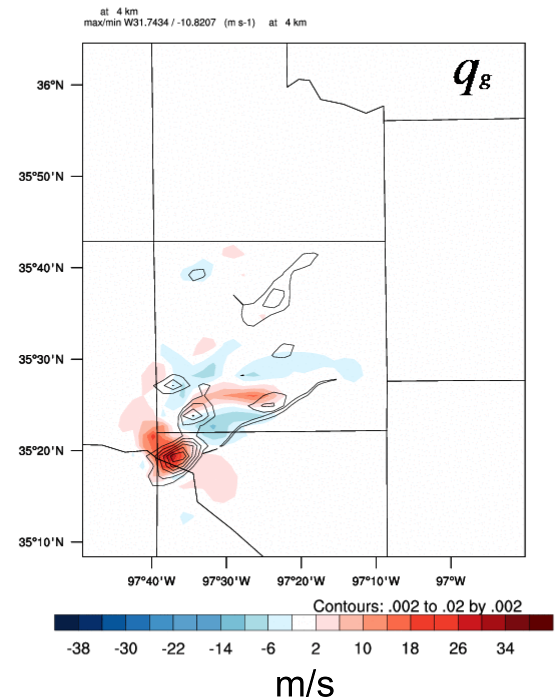
New: extend state variable with reflectivity



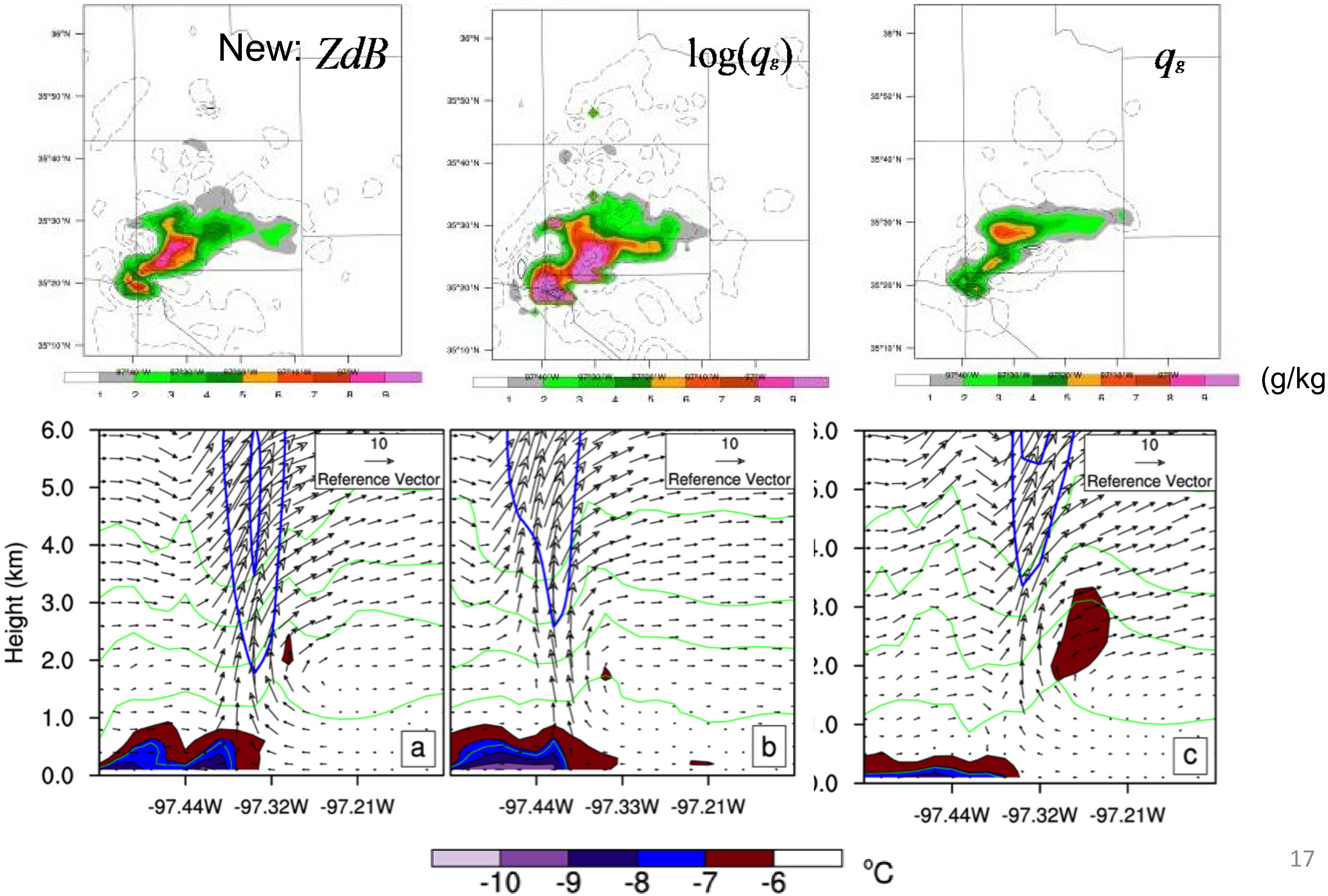
Use log transform (q_hydrometeor) as state variable



Use q_hydrometeor as state variable

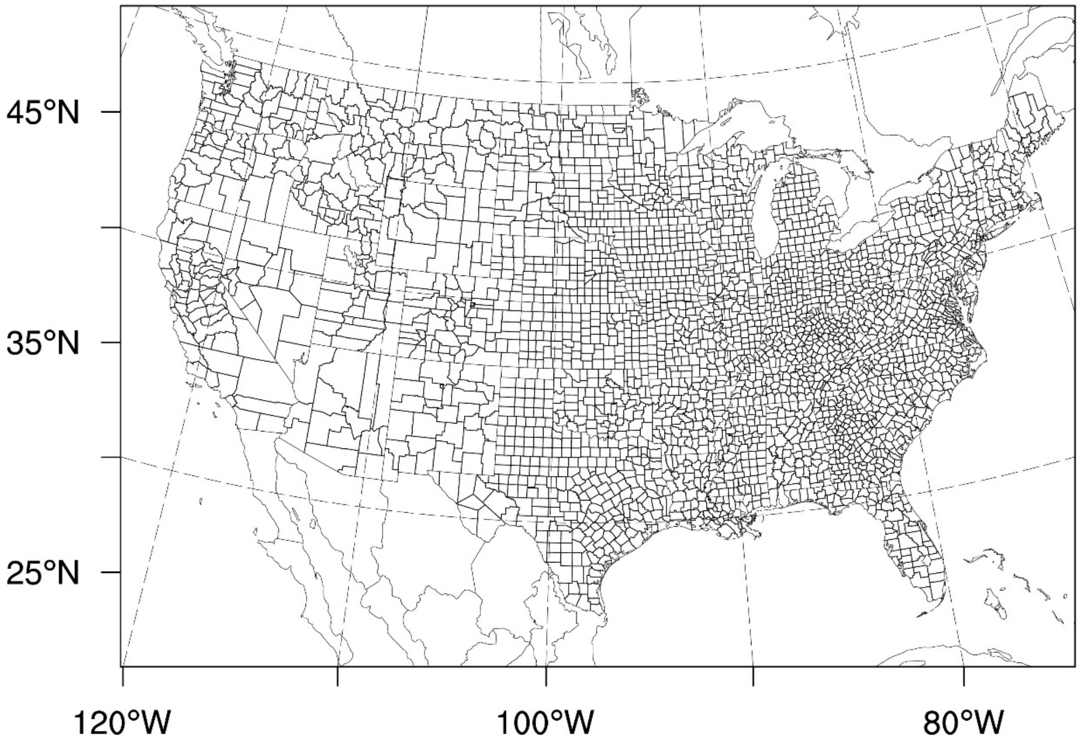


Graupel (q_g) analysis



Implementation and experiment in HRRR/NAM like applications over CONUS

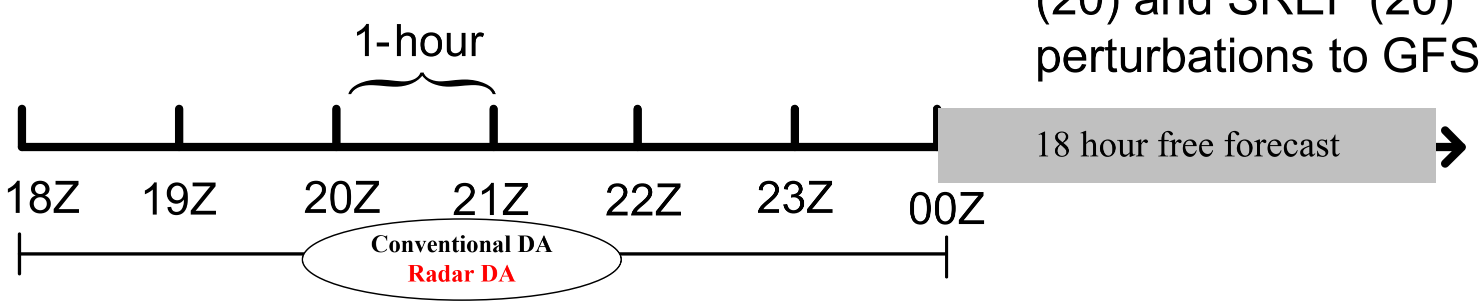
Wang * et al. 2018



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

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

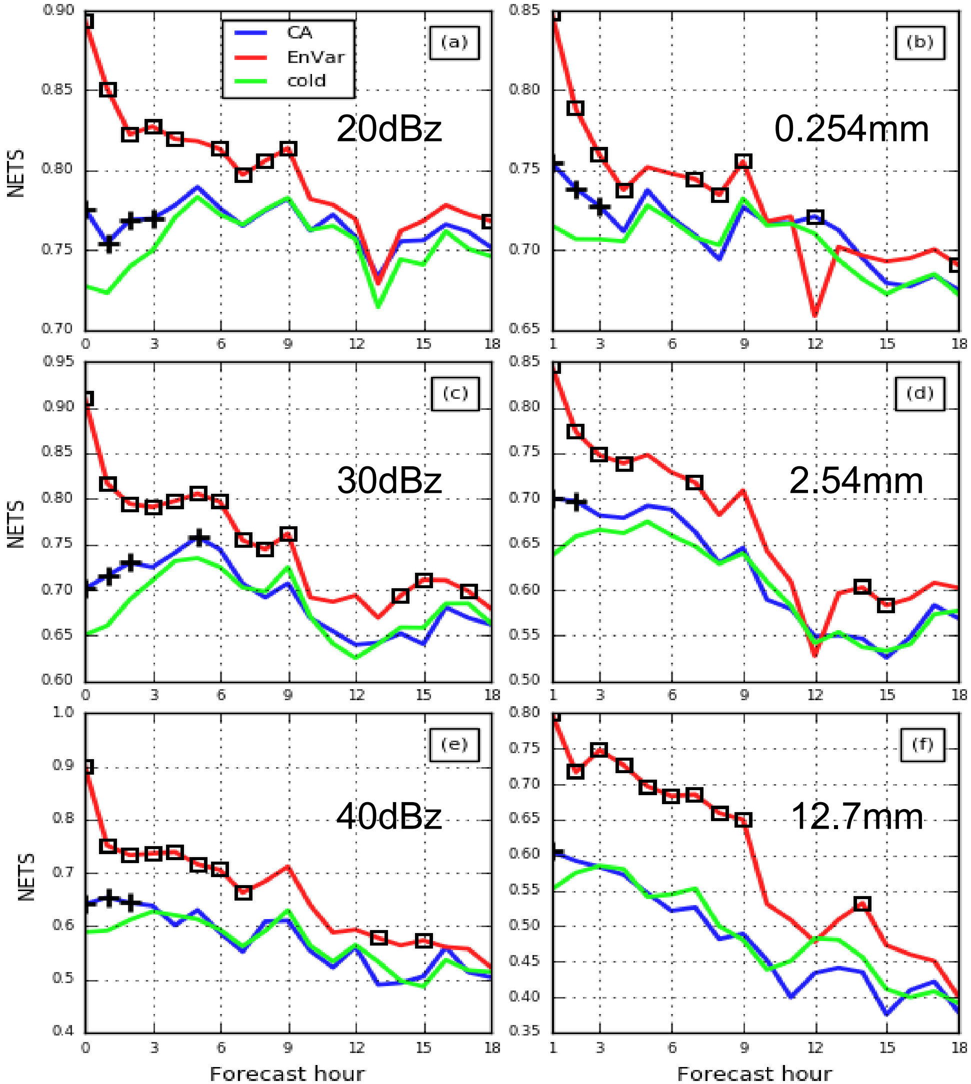
IC and LBC ensemble are provided by recentering GEFS (20) and SREF (20) perturbations to GFS-ctl



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

Duda*, Wang, Wang*, Carley, 2018, MWR

- ❑ Most operational system assimilating radar reflectivity uses empirical approach such as CA and diabatical initialization (e.g. HRRR, Hu et al. 2006)
- ❑ 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.
- ❑ Collaborating with GSD and EMC to transition the radar DA development into operations through HRRRv4 in 2020



GSI EnVar at sub-kilometer resolution

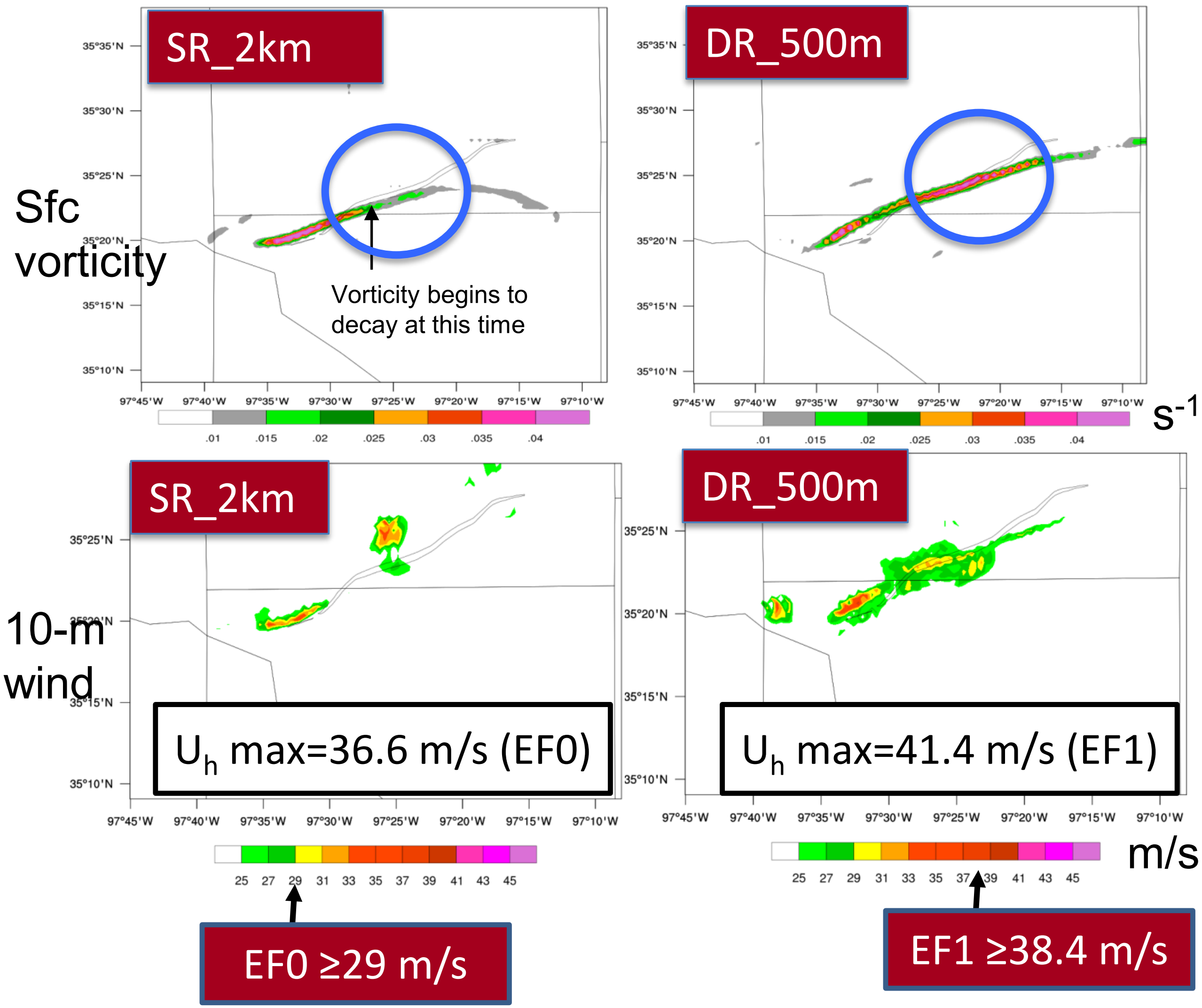
Kurdzo et al. 2015 from David Bodine (ARRC)

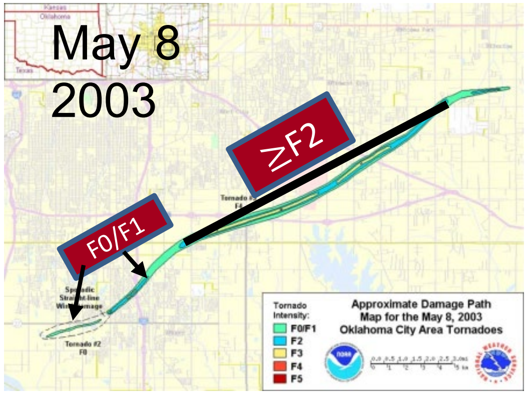
GSI-based dual resolution EnVar for sub-kilometer analysis and prediction

Wang Y. *and X. Wang 2018a, MWR

- State of the art radar provide measurements in very high resolution.
- Early study has demonstrated the need for ~100m possibly ~10's m grid spacing to fully resolve convective motions and explicit forecasting of tornado like vortices (e.g. Bryan et al. 2003).
- Many early studies simulate or predict tornado or tornado like vortices (TLV) by running sub-km model.
- Is there a need to run DA at finer resolution (<=1km)? 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).

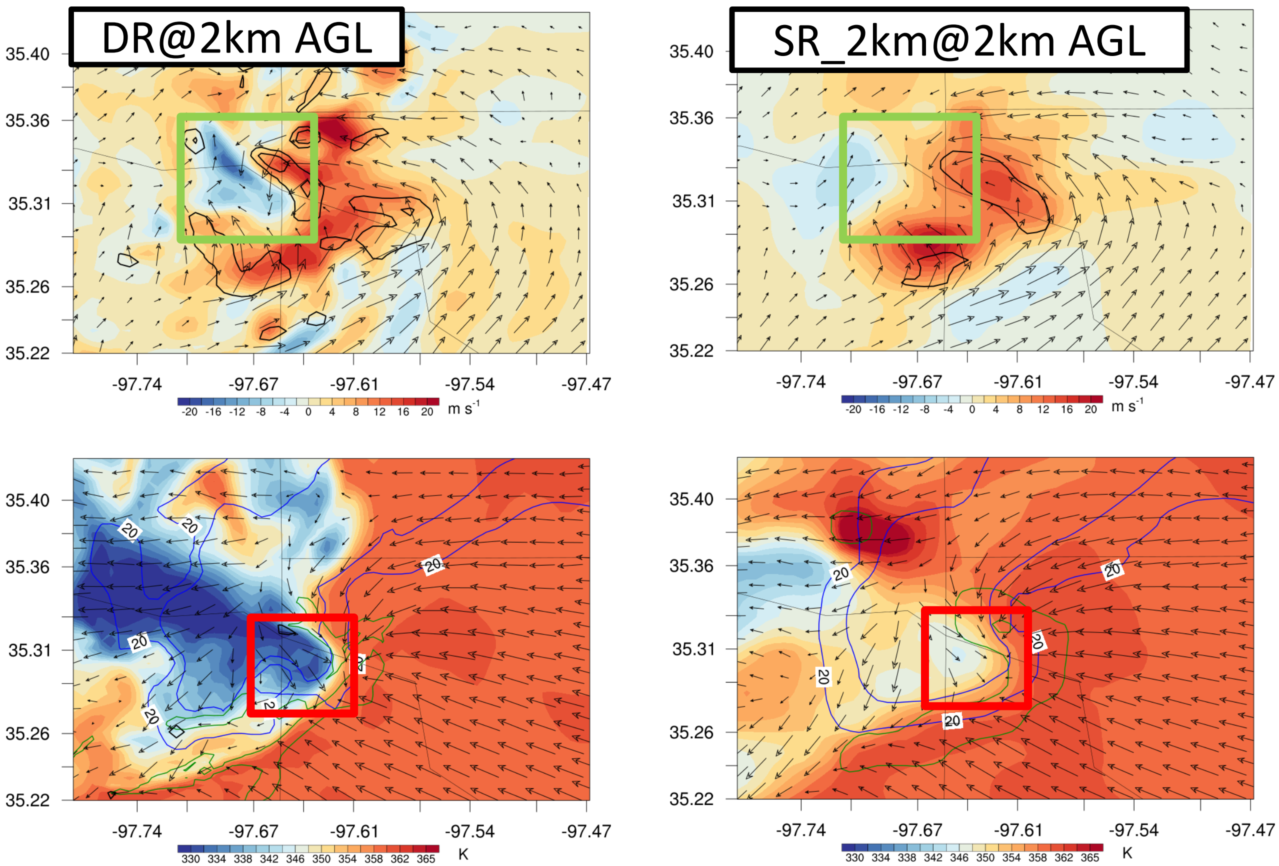
Composite maximum sfc vorticity and 10-m wind improved by dual resolution EnVar





- The predicted vorticity is enhanced after 20-min forecast in DR_500m. Its vorticity evolution is much more consistent with the reality than SR_2km.
- DR_500m 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_500m (left) than SR_2km (right) over the rear-flank region.
- Stronger outflow (red box) surge trailing the RFGF in DR_500m than SR_2km.

Extending static covariance for convective scales to treat background ensemble deficiency in GSI EnVar

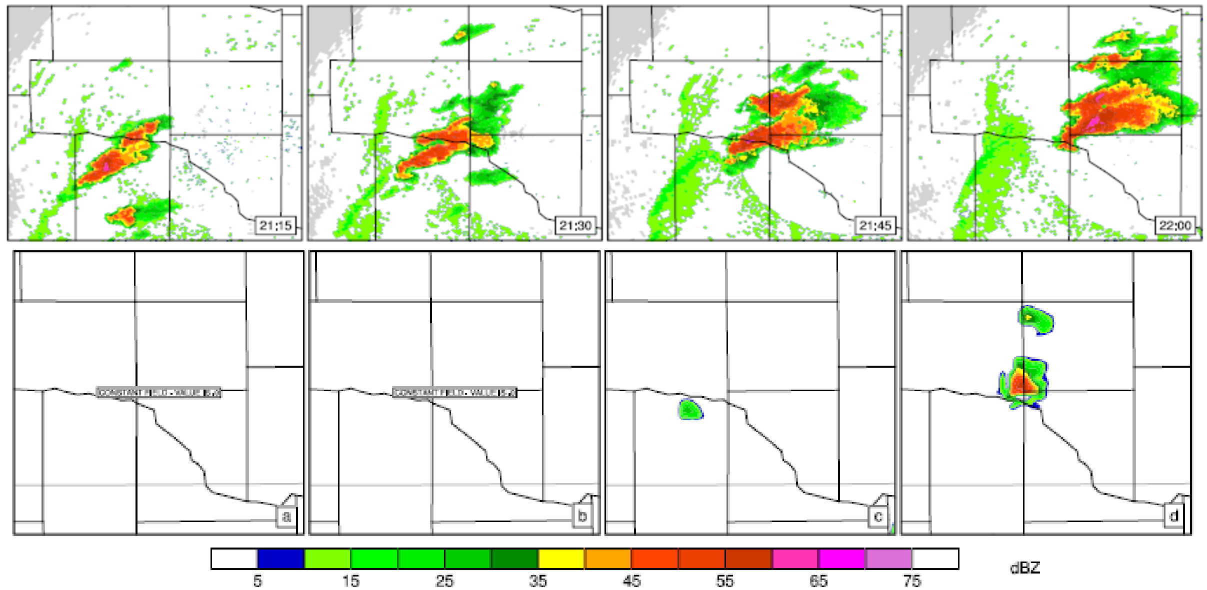
Ensemble background deficiencies

Wang* and Wang 2018b

DA cycling for May 8 2003 tornadic supercell

Obs.

Ensemble DA:
initial ensemble down scaled from meso. ensemble



- Ensemble background can be seriously deficit. For example, none of the members have the storm where in reality there is. In this case, obs. will not be used effectively to update the background since the background ensemble spread is zero.
- Random additive perturbation method was proposed (Dowell et al. 2004). However perturbations are not coherent among different variables and it does not add e.g. hydrometeors perturbations

Static covariance further extended for convective scale hybrid EnVar: impact on DA cycling

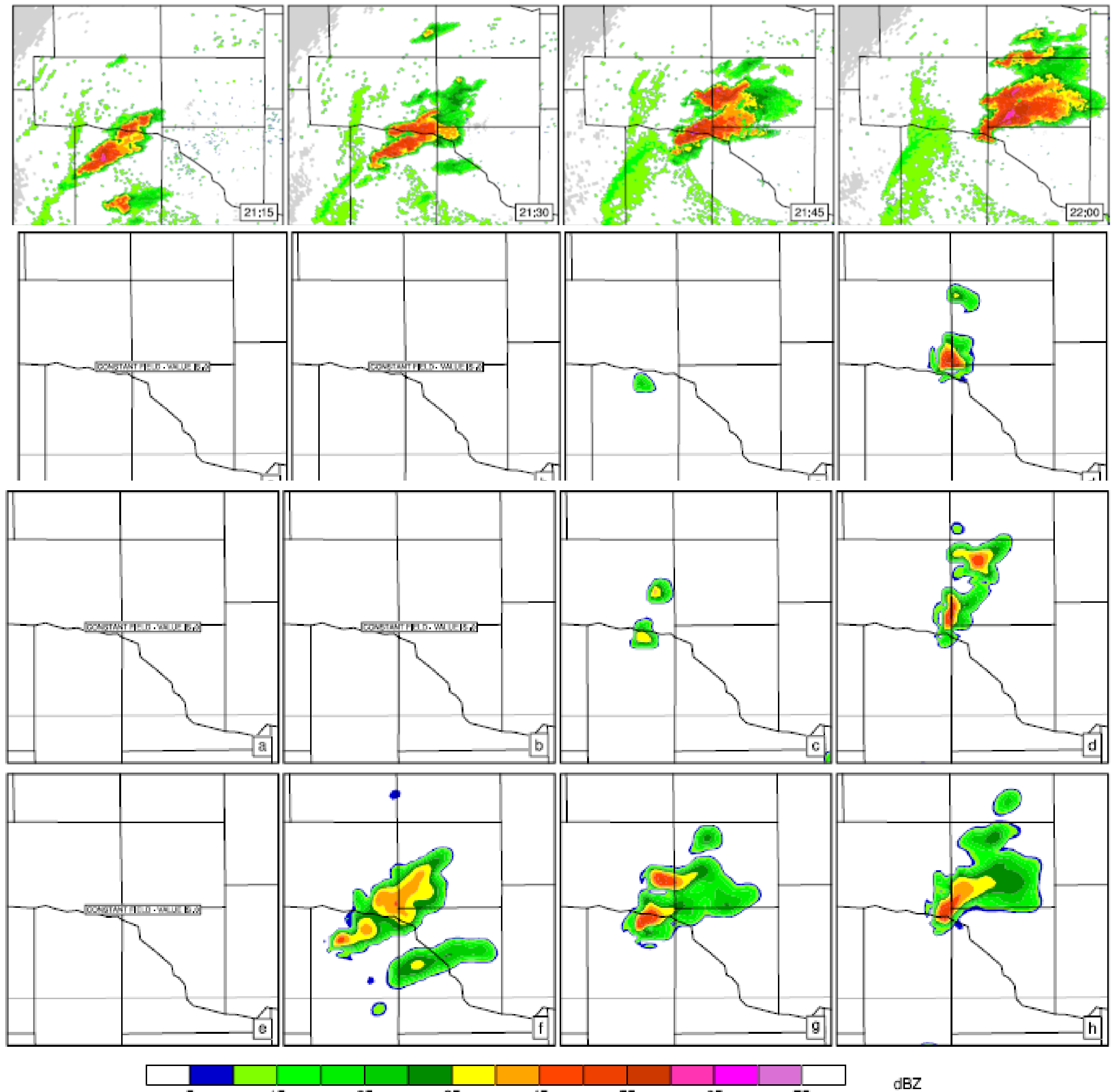
Wang* and Wang 2018b

Obs

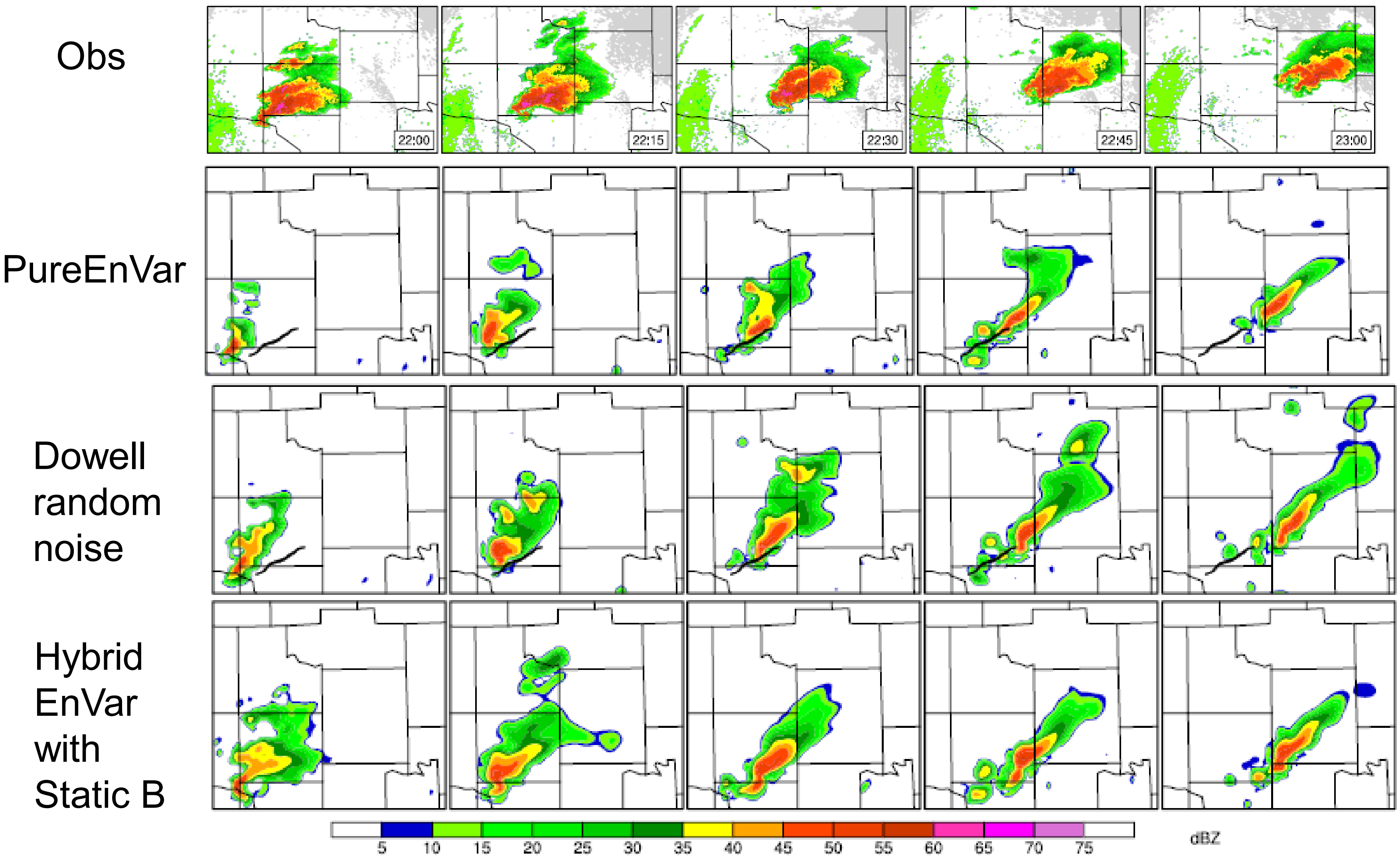
PureEnVar

Dowell random noise

Hybrid EnVar with Static B



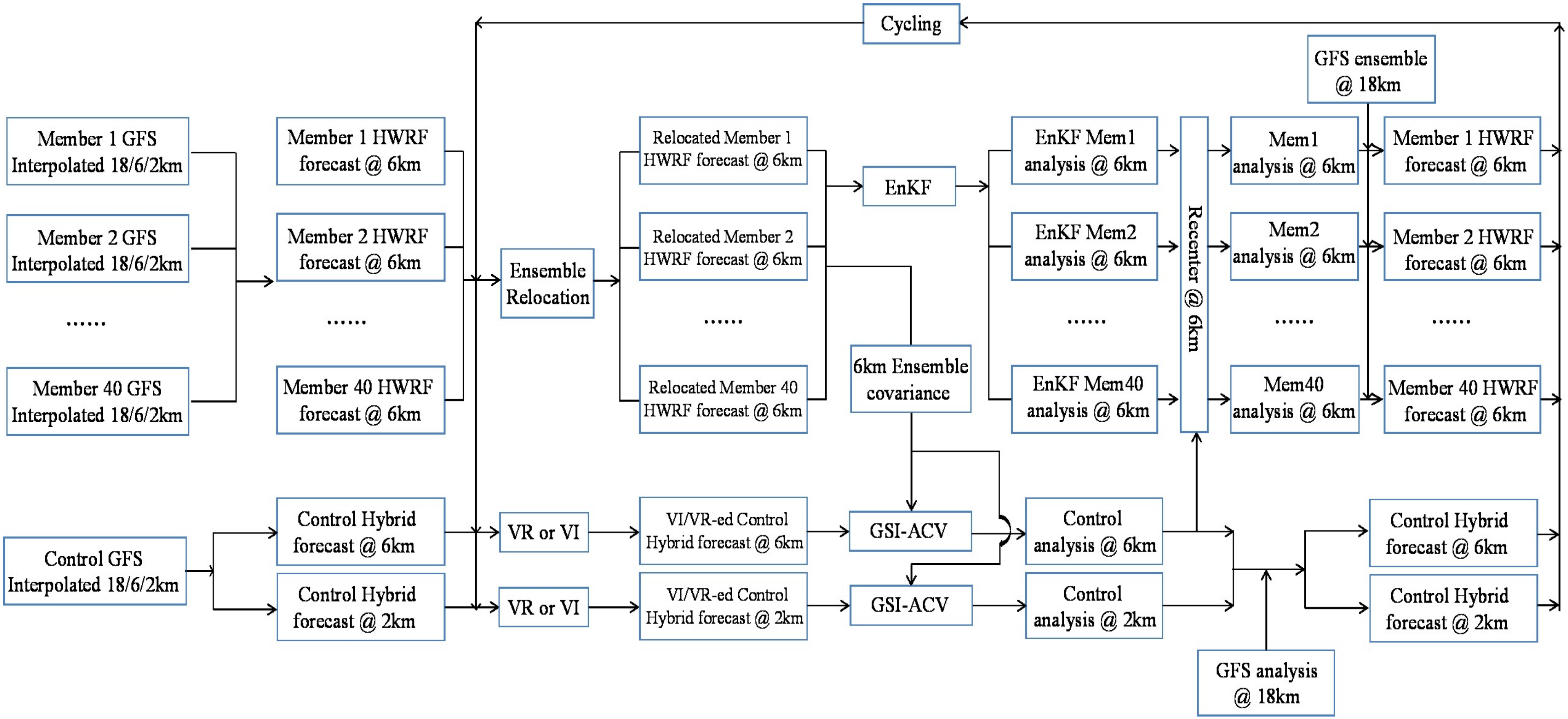
Static covariance further extended for convective sale hybrid EnVar: impact on forecasts



Use GSI EnVar DA to identify model deficiencies: an example from convection allowing hurricane prediction

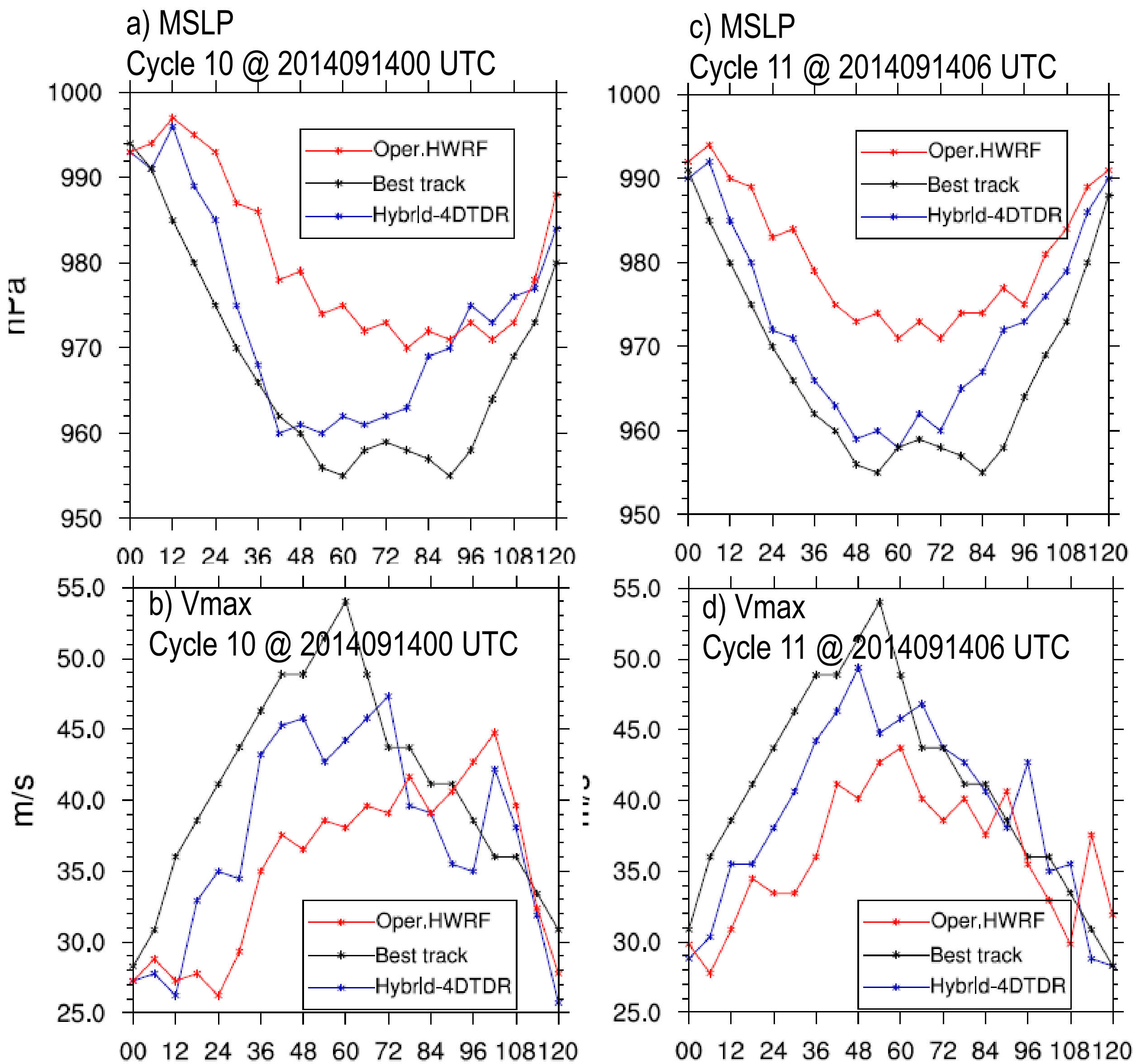
Continuously cycled, Dual-resolution, HWRF GSI hybrid DA system and its operational implementation in 2017

Lu*, Wang, Tong and Tallapragada 2017



- The GSI based hybrid DA system is developed with the following capabilities: (1) continuously cycling, (2) dual resolution, (3) 3DEnVar/4DEnVar, (4) assimilating all operational observations including TDR, HDOB, dropsonde, satellite radiances, etc., (5) integrated with VI (VR+VM)(Lu et al. 2017).

Alleviation of the “spin-down” issue relative to operational HWRF

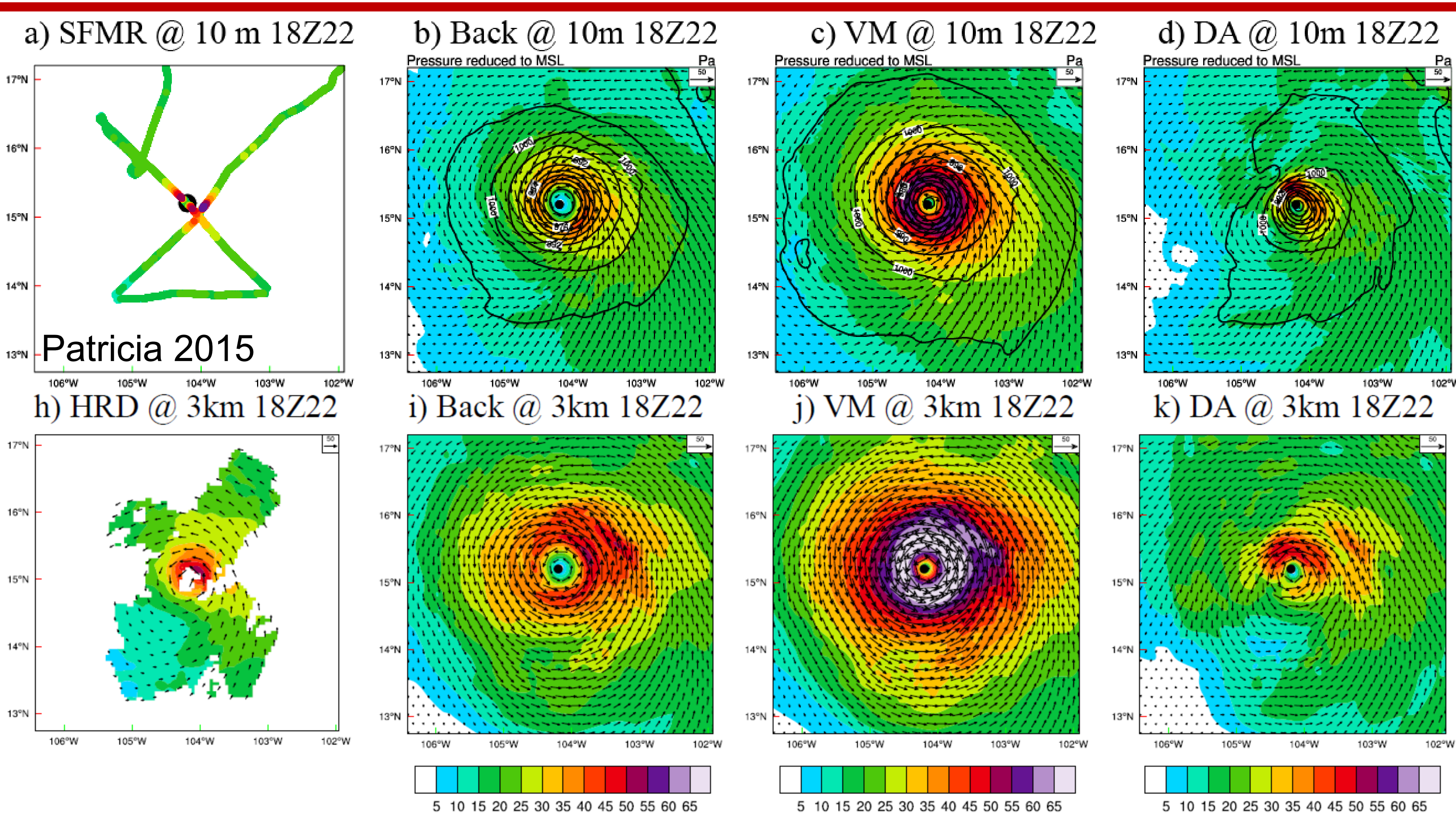


Edouard (2014)

- Improved analysis led to the improvements in the intensity forecasts through alleviation of the “spin down” issue presented in operational HWRF.

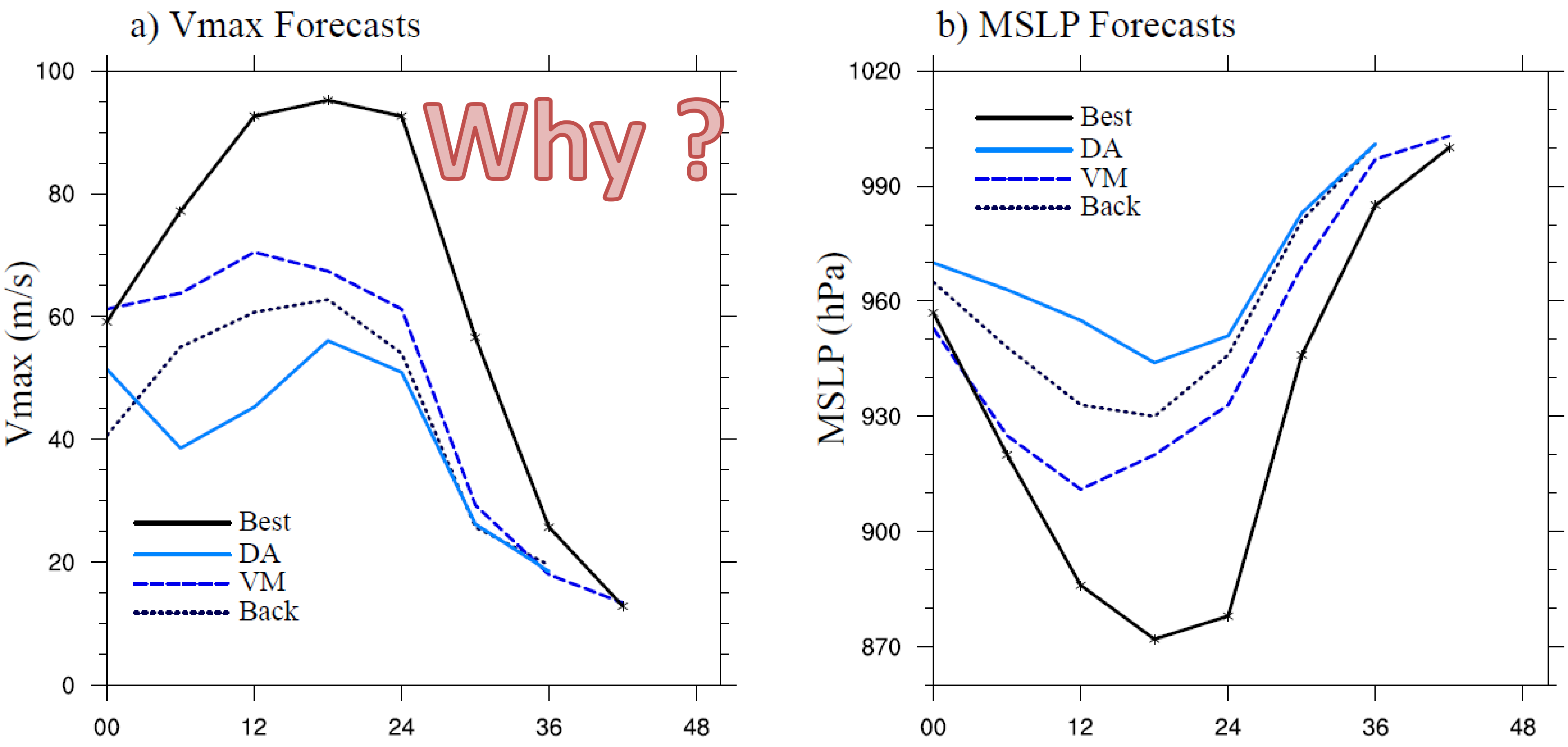
Use EnVar DA to reveal model physics errors

Lu* and Wang 2018a



- Back storm is large and weak as compared with observations.
- VM (Vortex Modification scheme) produces spurious strong and large storms.
- Inner core structures are much improved upon the background through DA.

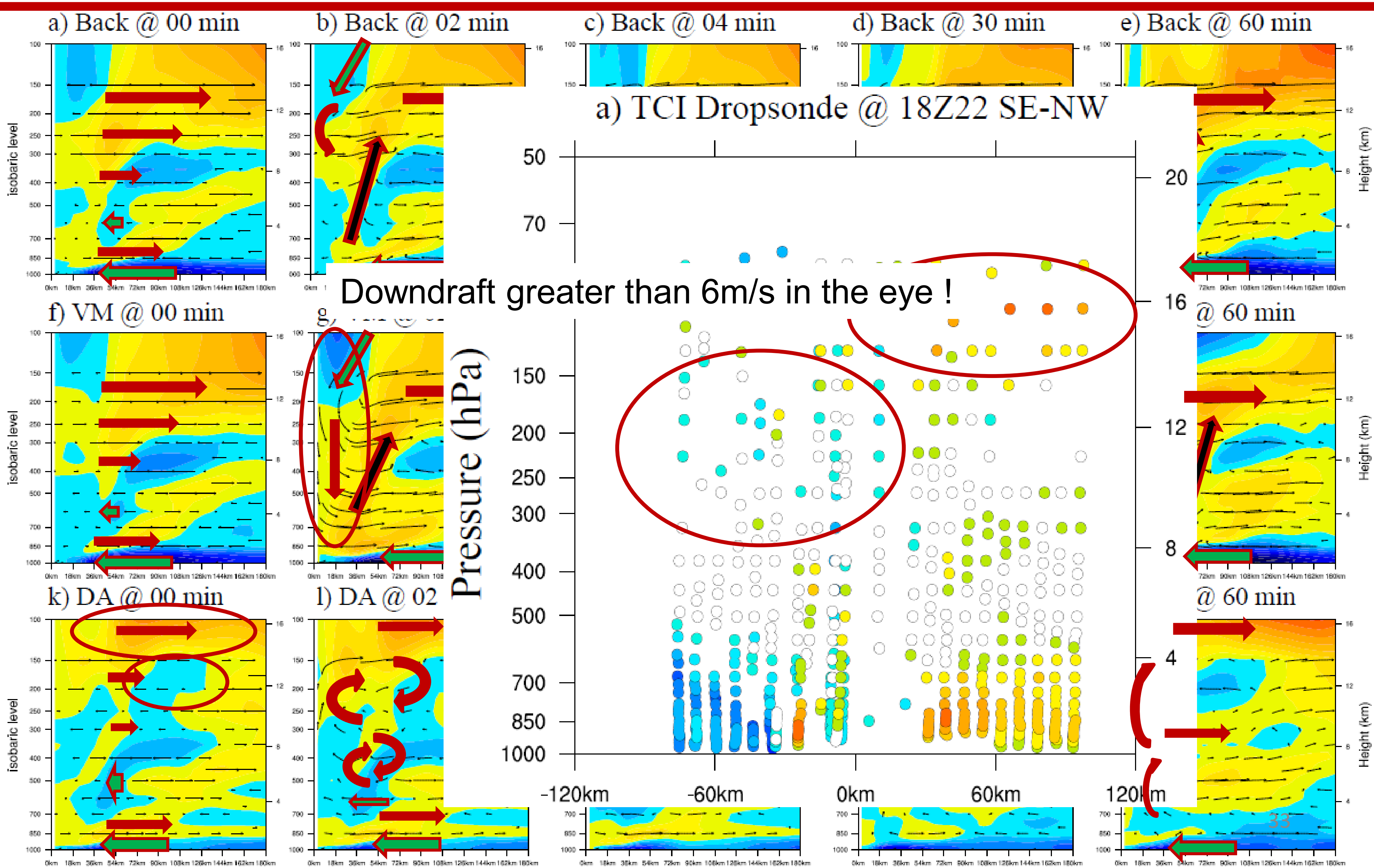
DA vs VM Intensity Forecasts



- Spin-down occurred in the experiments where inner-core wind structures are well captured in the analysis through DA.
- Background and VM analyses do not show spin-down.

Why TC spin-down with the more realistic DA analyses?

-- Secondary Circulation for the first hour



Why TC spin-down with the more realistic DA analyses?
Model physics issue 1: Horizontal diffusion too strong

Bryan and Rotunno, 2009

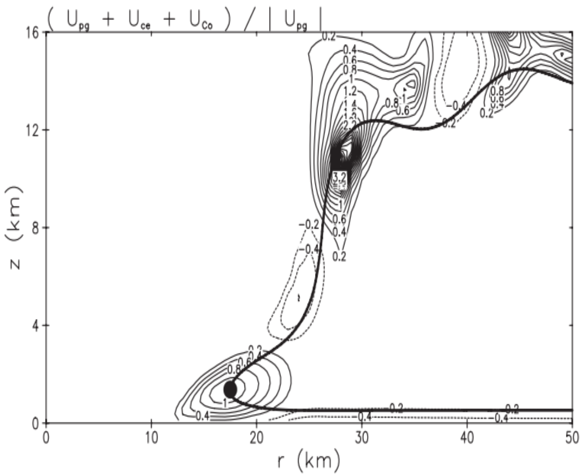
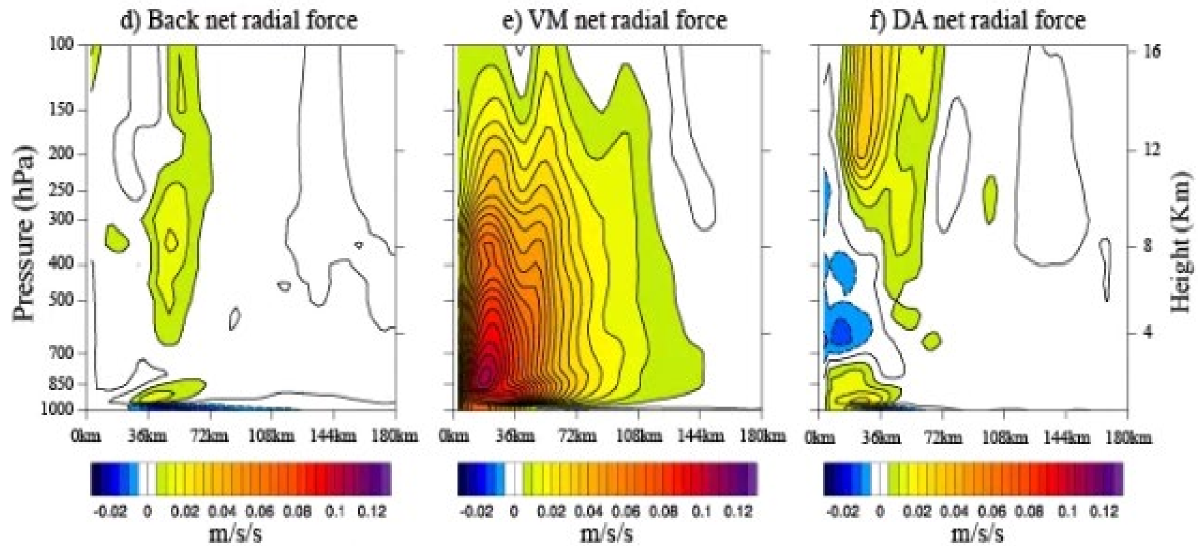


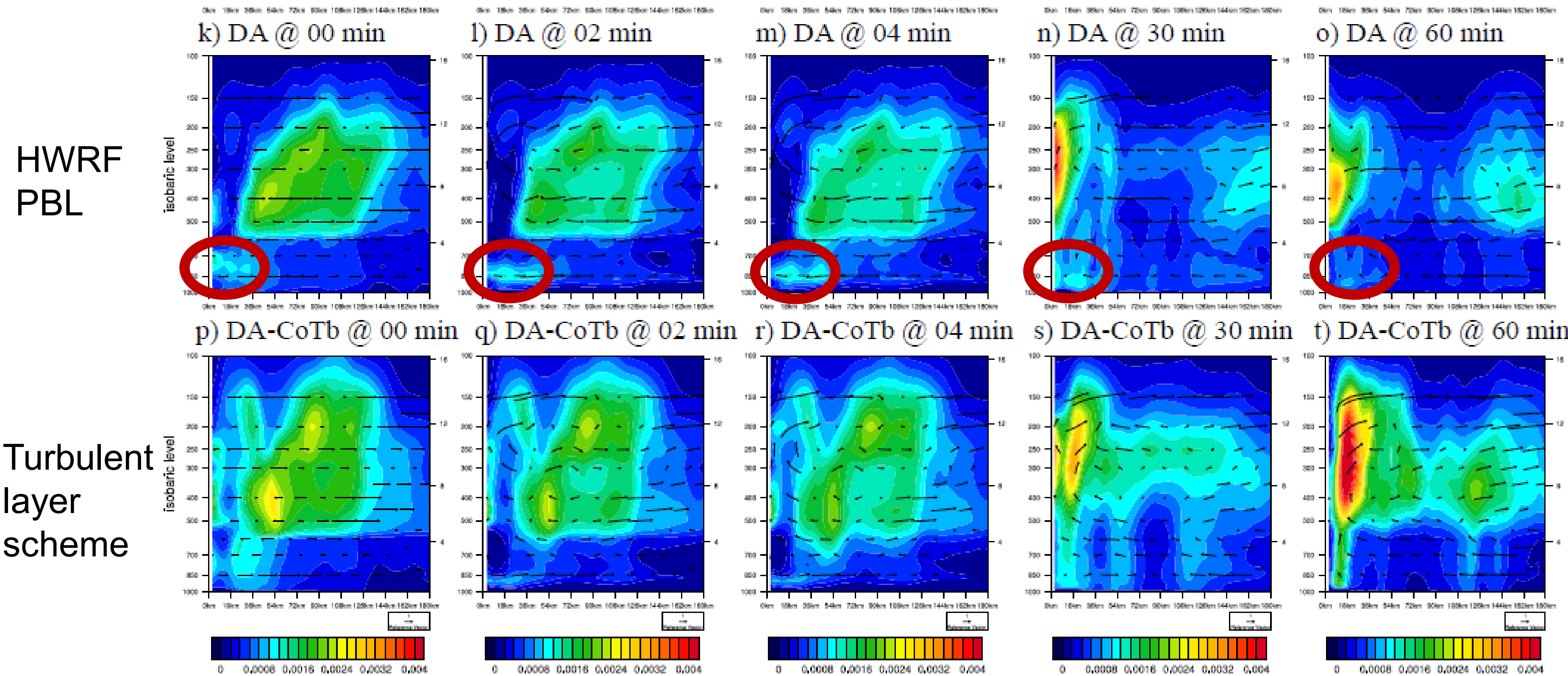
FIG. 8. Analysis of gradient-wind balance in the control simulation. Contoured is the sum $U_{pg} + U_{ce} + U_{Co}$ normalized by the magnitude of U_{pg}, with a contour interval of 0.2. The zero contour is excluded. The trajectory that passes through v_{max} is illustrated by the thick line, and the dot denotes the location of v_{max}.



- ❑ The middle-level sub-gradient is very likely a direct response to the boundary layer super-gradient (Stern and Nolan, 2011). The oscillation roots in the PBL.
- ❑ Unbalanced flow effects have a nonnegligible effect on intensity in some cases and stronger radial diffusion damps the unbalanced flow effects (Bryan and Rotunno, 2009).

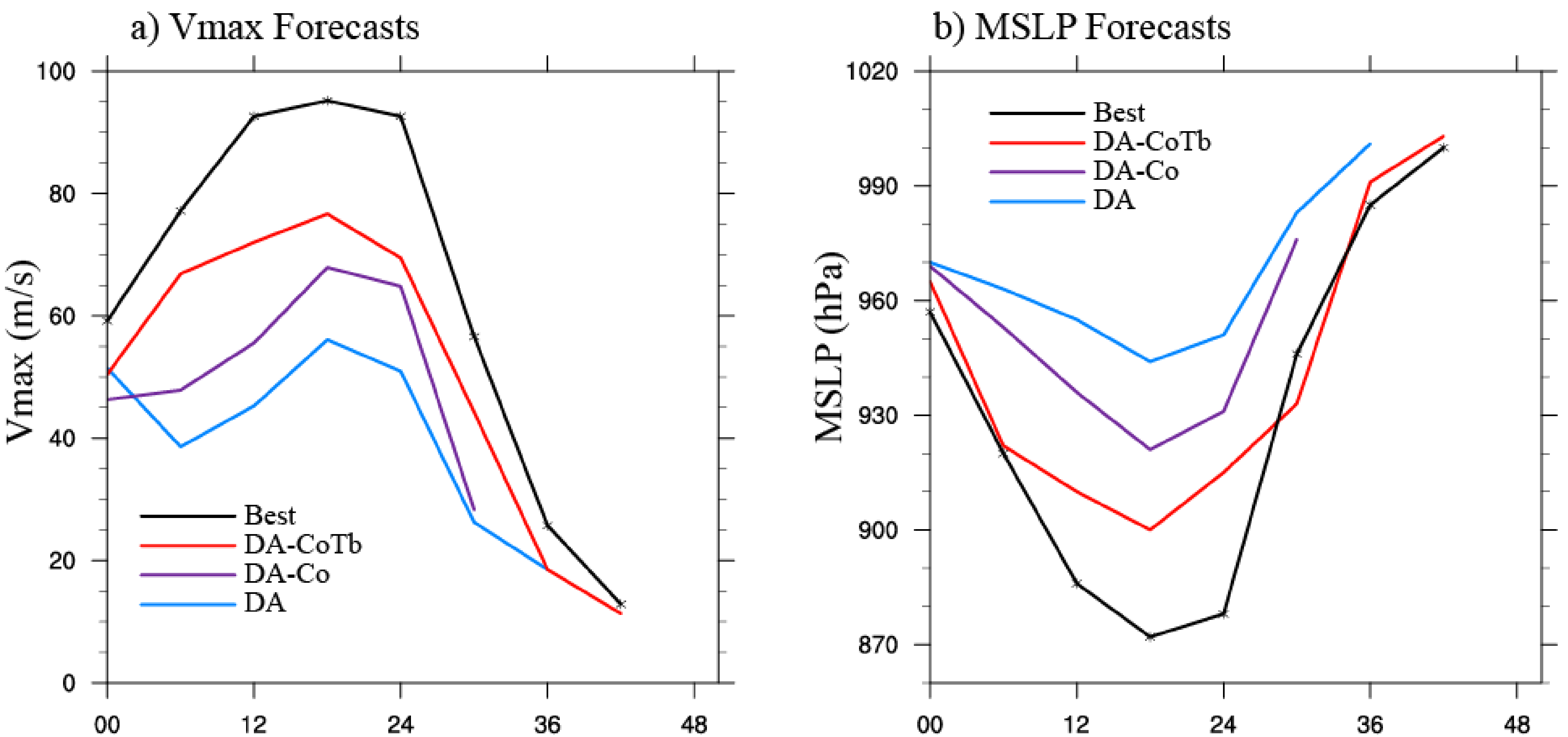
Why TC spin-down with the more realistic DA analyses?

Model physics issue 2: Lack of Mixing in HWRF PBL



- In the original HWRF PBL scheme, the discontinuity of turbulent mixing at the boundary layer top tends to constrain the communication of moisture and heat below and above the boundary layer top.
- Turbulent layer mixing (Zhu et al. 2016) allows more moisture and heat to be transported to the free atmosphere, facilitating establishing secondary circulation.

Advanced DA needs to be accompanied by advanced model Physics



- Reducing the horizontal diffusion in DA-Co shows improved MSLP forecast and apparent alleviation in the Vmax spin-down.
- Further using a modified turbulent mixing scheme (DA-CoTb) shows significant improvement in both Vmax, MSLP and track forecasts.

Remaining challenges of data assimilation for NWP

- ❑ Multiscale data assimilation

- ❑ Treat nonlinearity and non-gaussianity in high-dimensional system
- ✓ Parametric or non-parametric approach

- ❑ Accurate representation of the background errors
- ✓ Advancing methods to treat sampling errors and represent model errors

- ❑ Observation operator development for new instruments. Accurate representation of the observation errors and their correlation

- ❑ Reveal, correct and quantify model errors using DA.

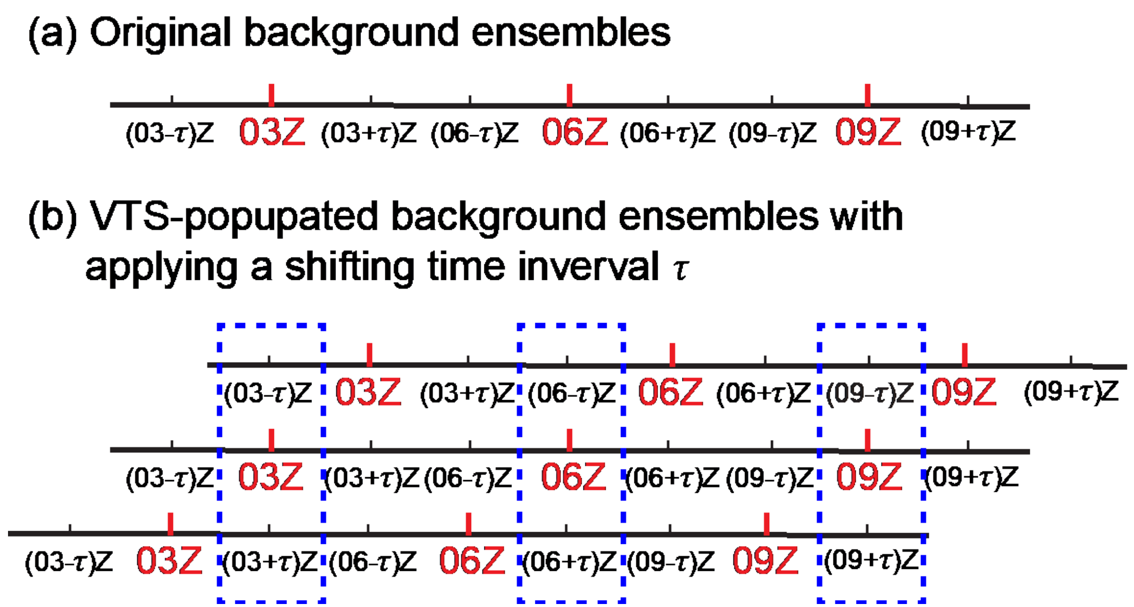
- ❑ Big data assimilation (huge amount of remote sensing and in-situ obs., increased model resolution and increased # of ensemble members)

Requires collaboration among data assimilation (DA) developers, model physics developers, observation/instrument experts and data/machine learning scientists

VTS method – a cost effective method to increase background ensemble size in 4DEnVar

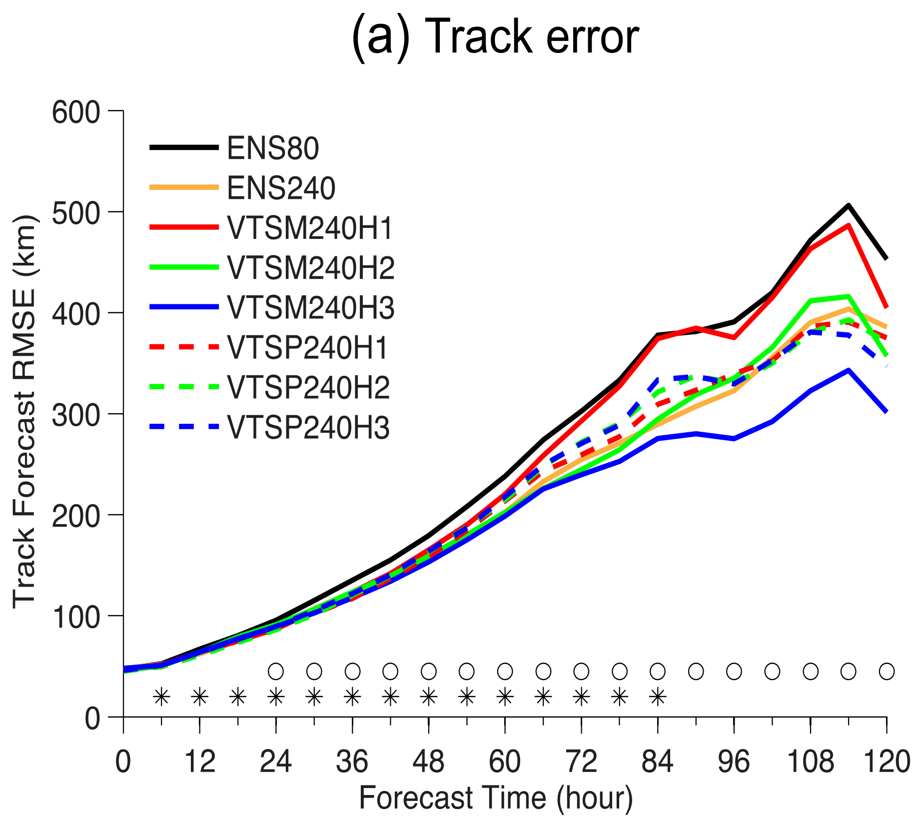
Huang* and Wang 2018, MWR

- VTS increases ensemble size by shifting the ensembles valid around the analysis time to the analysis time.



- VTSM by applying VTS to **the ensemble members**
- VTSP by applying VTS to the **ensemble perturbations**.

Evaluation on TC track forecasts



- All VTS experiments show smaller track errors than ENS80.
- VTSM shows smaller errors with larger lagging time interval, while VTSP is not very sensitive to the lagging time interval.
- VTSP performs similarly to ENS240. VTSM240H3 even produces smaller track errors than ENS240.