Recent Development of Multiscale Data Assimilation for Numerical Weather Prediction

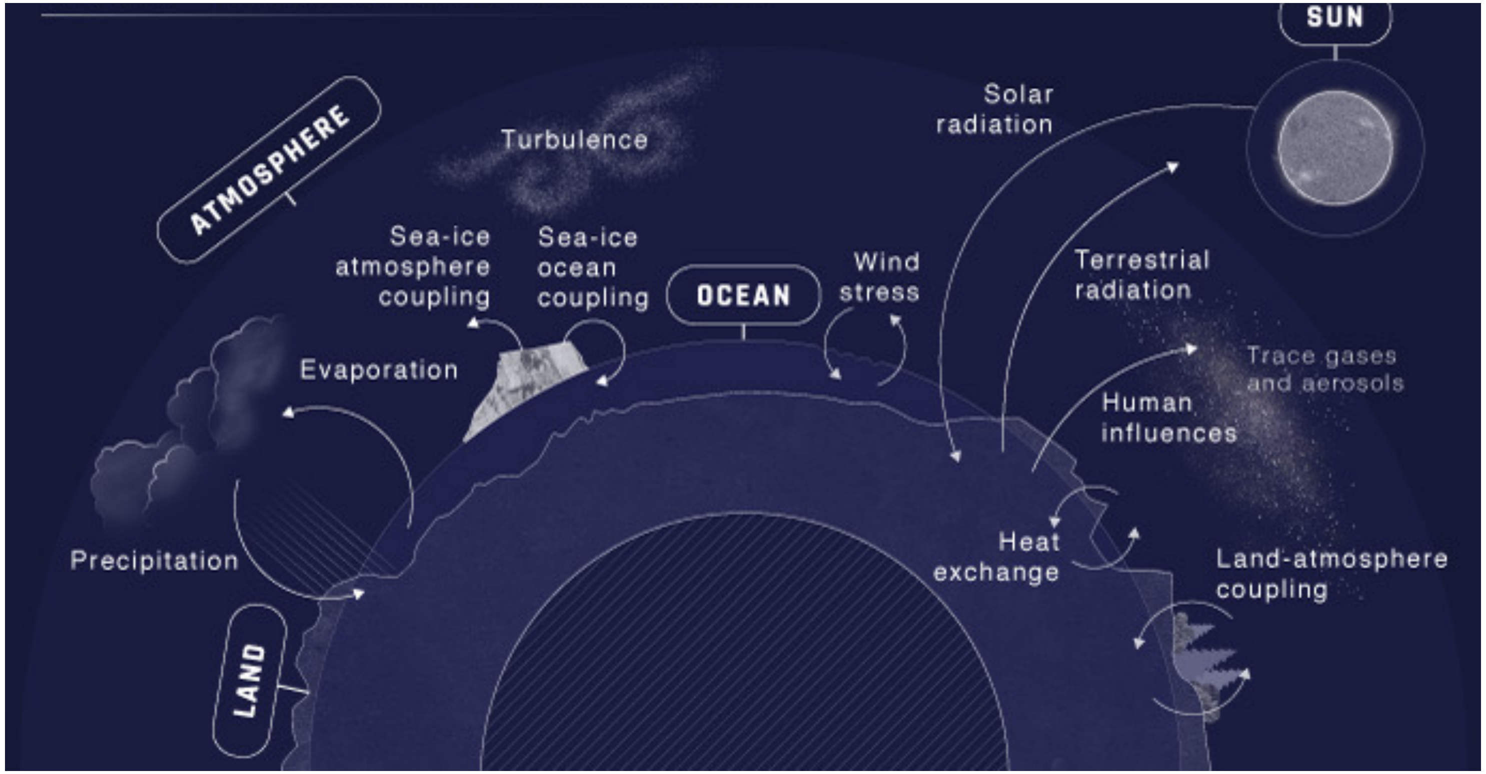
Xuguang Wang
Multiscale data **A**ssimilation and **P**redictability (**MAP**) lab
University of Oklahoma, Norman, USA

Acknowledgement
Yongming Wang*, Erin Jones*, Xu Lu*, Naicheng Xu*, Yue Yang* (OU MAP)
Collaborators from NOAA EMC, PSL, GSL, HRD, NSSL; NRL; Univ. Melbourne

*MAP lab students/early career scientists

ISDA, Bologna, Italy, 2023

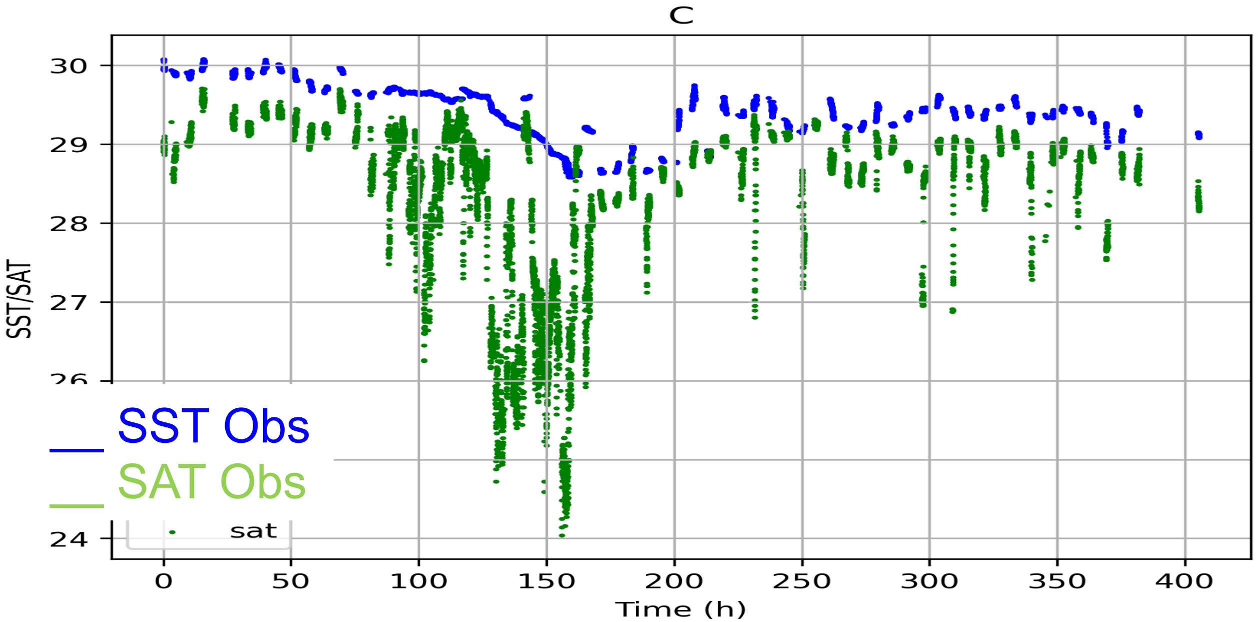
Why Multiscale Data Assimilation (MDA)



https://www.ecmwf.int/en/about/media-centre/focus/2021/fact-sheet-earth-system-modelling-ecmwf

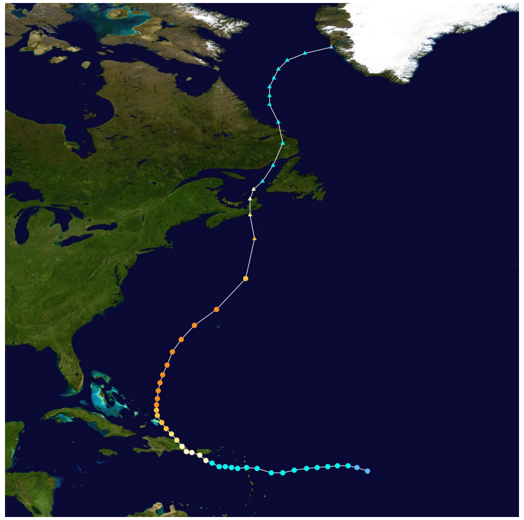
- In addition to the atmosphere, ocean, sea ice, and the land surface etc can have a significant impact on weather.
- These earth system components have **intrinsically different spatial and temporal scales**.
- Accurate numerical weather prediction requires proper initialization of these **multiscale** earth system components and their **interaction (coupling)** through data assimilation.

Why Multiscale Data Assimilation (MDA)



Lu* and Wang et al. 2023b

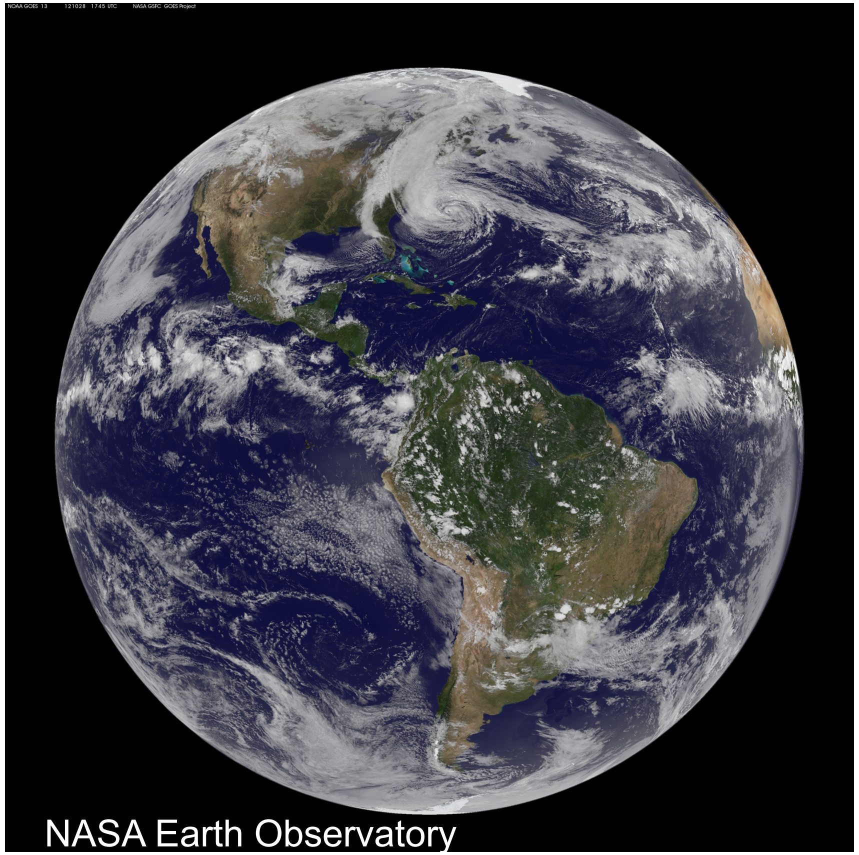
Saildrone.com

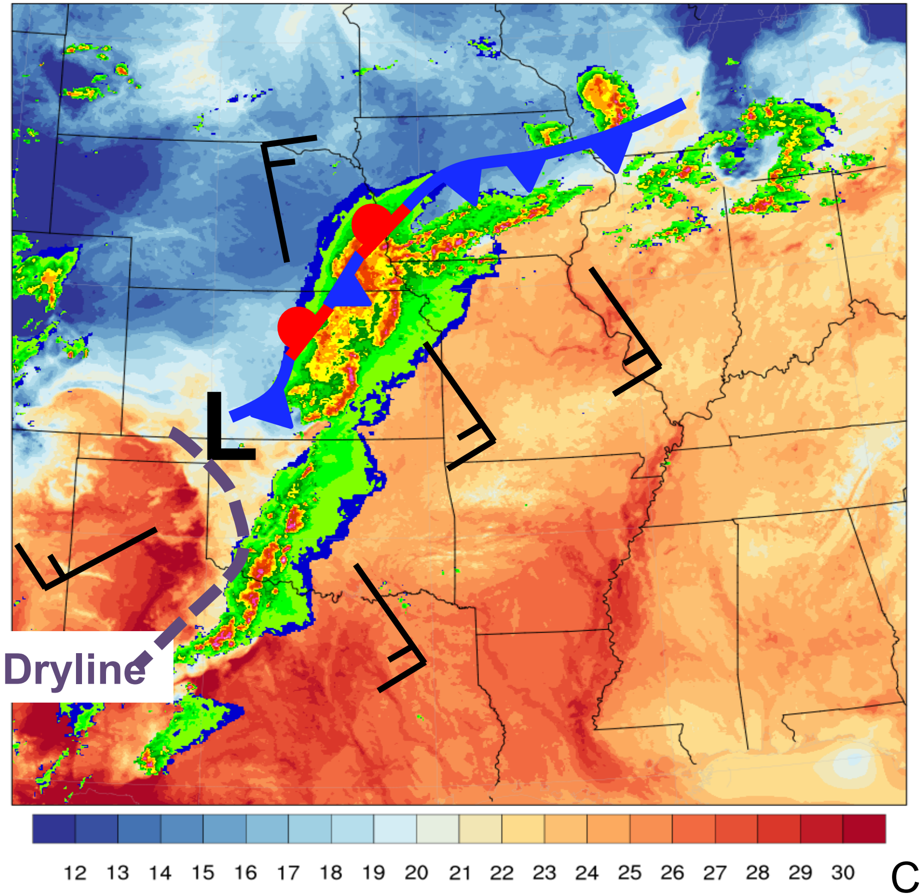


https://en.wikipedia.org/wiki/Hurricane_Fiona

- Proper initialization of earth system components and their interactions (coupling) through MDA not only applies for e.g. S2S prediction, but also for short term prediction.

Why Multiscale Data Assimilation (MDA)



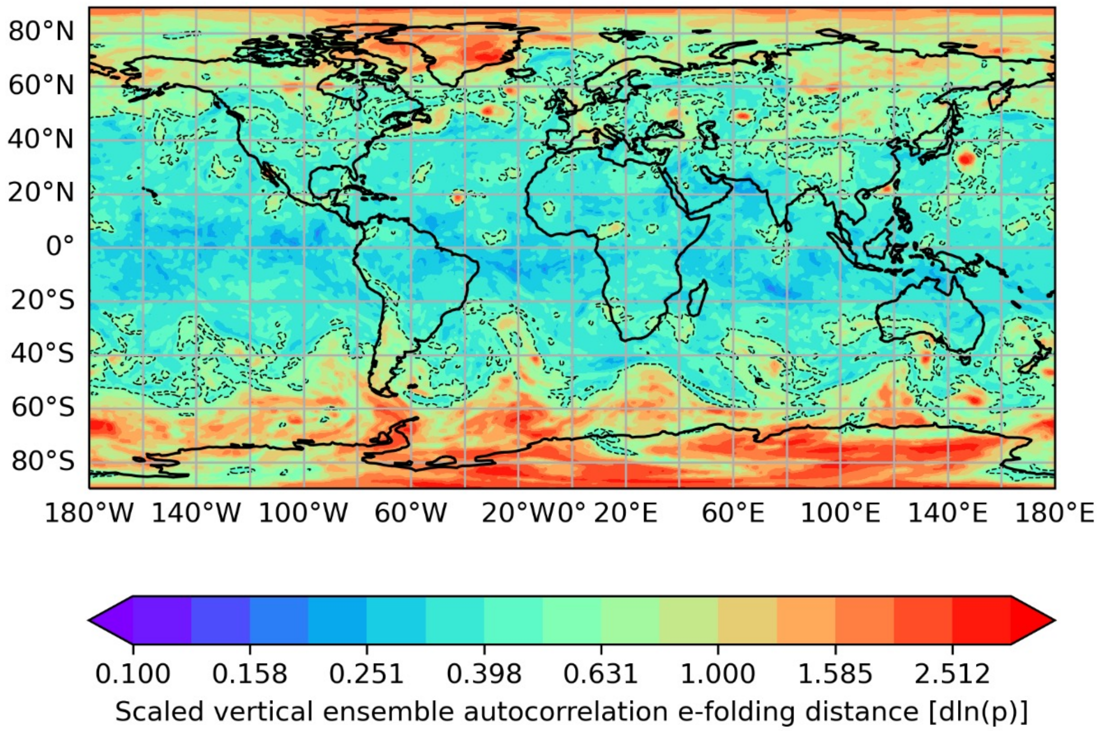


Wang* and Wang, 2023b

- The atmosphere, as one earth system component, itself is intrinsically multiscale as well, housing micro to planetary scales.
- Accurate prediction of storms (e.g. hurricanes, squall lines, supercells) require DA to initialize not only the storm but also its larger scale environment.

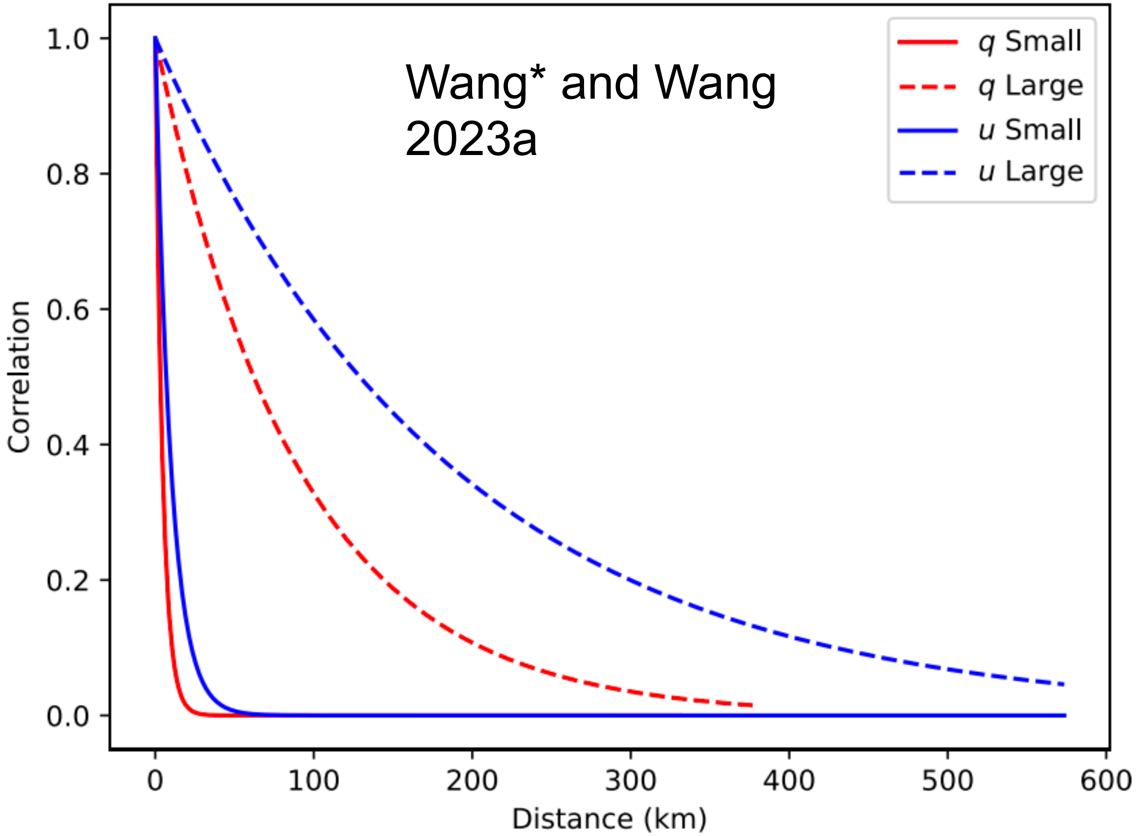
Why Multiscale Data Assimilation (MDA)

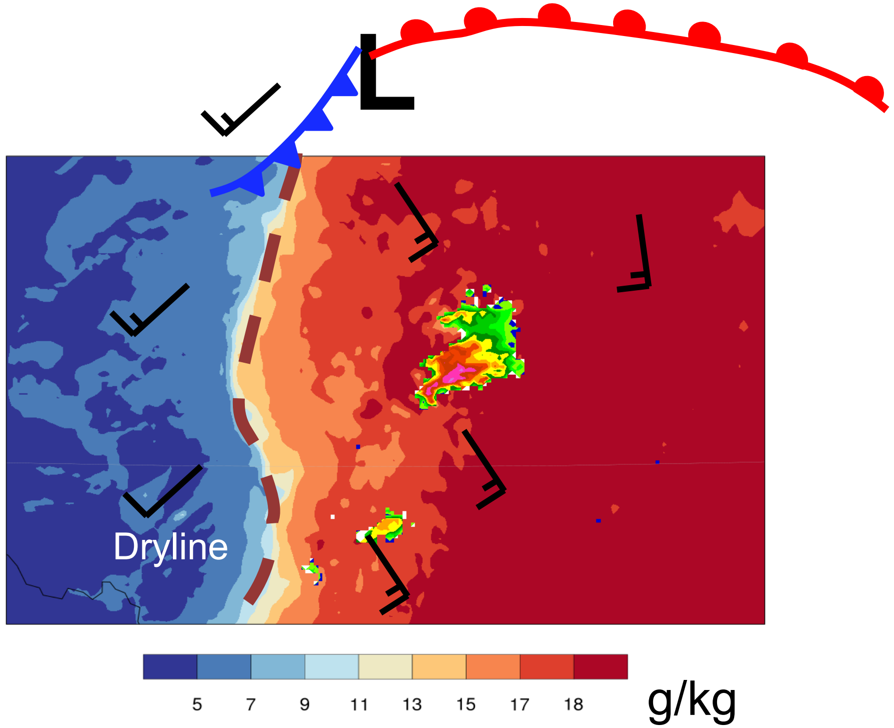
Jones* and Wang 2023b



- ❑ Deep convective systems have larger vertical correlation length scales than the mean atmospheric state (e.g., Ingleby 2001)
- ❑ Mean tropical atmosphere has smaller vertical correlation length than extratropics (e.g., Rabier et al. 1998)

Why Multiscale Data Assimilation (MDA)





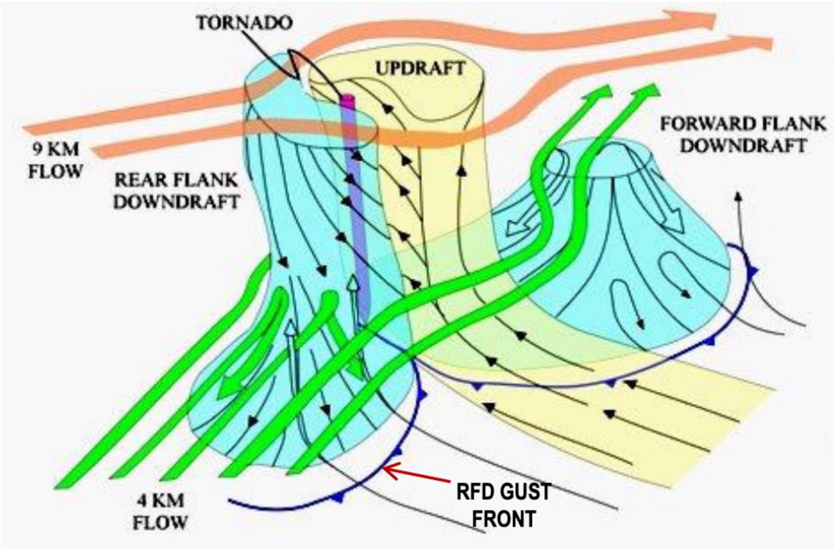
- Different variables can have intrinsically different scales.
- Considering intrinsic scale differences of variables are important for storm scale prediction (Wang, Y.* and X. Wang 2023)

Why Multiscale Data Assimilation (MDA)

- Control variables and intrinsic “balance” are different between large scales and convective scales

$$
\begin{pmatrix} u \\ v \\ t \\ \mathrm{ps} \\ \mathrm{rh} \\ w \\ \mathrm{ql} \\ \mathrm{qr} \\ \mathrm{qs} \\ \mathrm{qi} \\ \mathrm{qg} \\ \mathrm{dbz} \end{pmatrix}
=
\begin{pmatrix}
\mathbf{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{r}_{11} & \mathbf{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{r}_{21} & \mathbf{r}_{22} & \mathbf{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{r}_{31} & \mathbf{r}_{32} & \mathbf{r}_{33} & \mathbf{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{r}_{41} & \mathbf{r}_{42} & \mathbf{r}_{43} & \mathbf{r}_{44} & \mathbf{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{r}_{51} & \mathbf{r}_{52} & \mathbf{r}_{53} & \mathbf{r}_{54} & \mathbf{r}_{55} & \mathbf{I} & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{r}_{61} & \mathbf{r}_{62} & \mathbf{r}_{63} & \mathbf{r}_{64} & \mathbf{r}_{65} & \mathbf{r}_{66} & \mathbf{I} & 0 & 0 & 0 & 0 & 0 \\
\mathbf{r}_{71} & \mathbf{r}_{72} & \mathbf{r}_{73} & \mathbf{r}_{74} & \mathbf{r}_{75} & \mathbf{r}_{76} & \mathbf{r}_{77} & \mathbf{I} & 0 & 0 & 0 & 0 \\
\mathbf{r}_{81} & \mathbf{r}_{82} & \mathbf{r}_{83} & \mathbf{r}_{84} & \mathbf{r}_{85} & \mathbf{r}_{86} & \mathbf{r}_{87} & \mathbf{r}_{88} & \mathbf{I} & 0 & 0 & 0 \\
\mathbf{r}_{91} & \mathbf{r}_{92} & \mathbf{r}_{93} & \mathbf{r}_{94} & \mathbf{r}_{95} & \mathbf{r}_{96} & \mathbf{r}_{97} & \mathbf{r}_{98} & \mathbf{r}_{99} & \mathbf{I} & 0 & 0 \\
\mathbf{r}_{101} & \mathbf{r}_{102} & \mathbf{r}_{103} & \mathbf{r}_{104} & \mathbf{r}_{105} & \mathbf{r}_{106} & \mathbf{r}_{107} & \mathbf{r}_{108} & \mathbf{r}_{109} & \mathbf{r}_{1010} & \mathbf{I} & 0 \\
\mathbf{r}_{111} & \mathbf{r}_{112} & \mathbf{r}_{113} & \mathbf{r}_{114} & \mathbf{r}_{115} & \mathbf{r}_{116} & \mathbf{r}_{117} & \mathbf{r}_{118} & \mathbf{r}_{119} & \mathbf{r}_{1110} & \mathbf{r}_{1111} & \mathbf{I}
\end{pmatrix}
\begin{pmatrix} u \\ v_u \\ t_u \\ \mathrm{ps}_u \\ \mathrm{rh}_u \\ w_u \\ \mathrm{ql}_u \\ \mathrm{qr}_u \\ \mathrm{qs}_u \\ \mathrm{qi}_u \\ \mathrm{qg}_u \\ \mathrm{dbz}_u \end{pmatrix}.
$$

Convective scale static B
Wang Y.* and X. Wang 2021



[https://www.weather.gov/media/lmk/soo/Supercell_Structure.pdf]

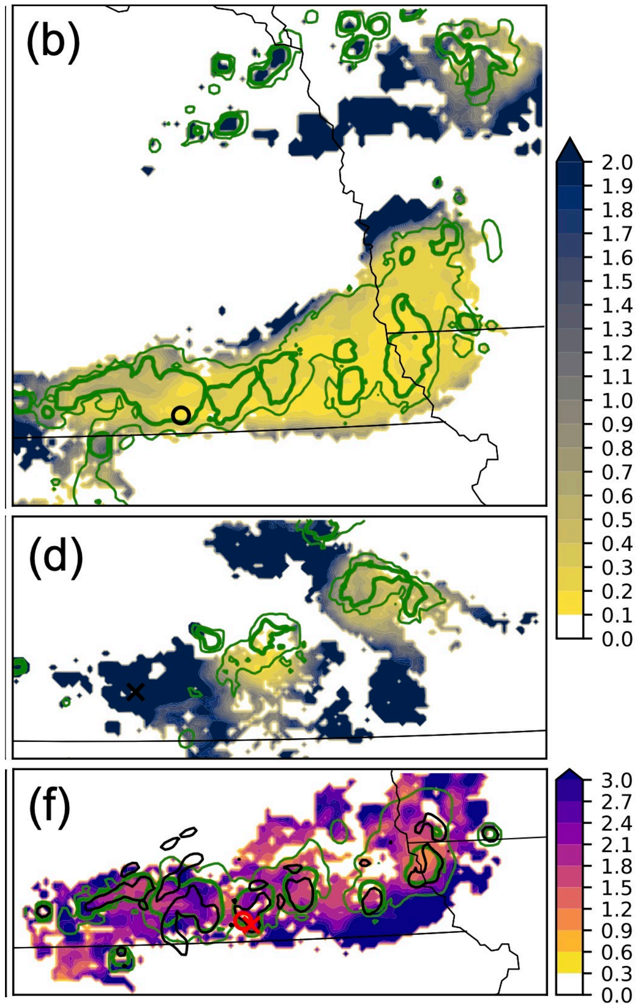
Why Multiscale Data Assimilation (MDA)

Non-Guassianity D_{KL}

Reflectivity

Reflectivity

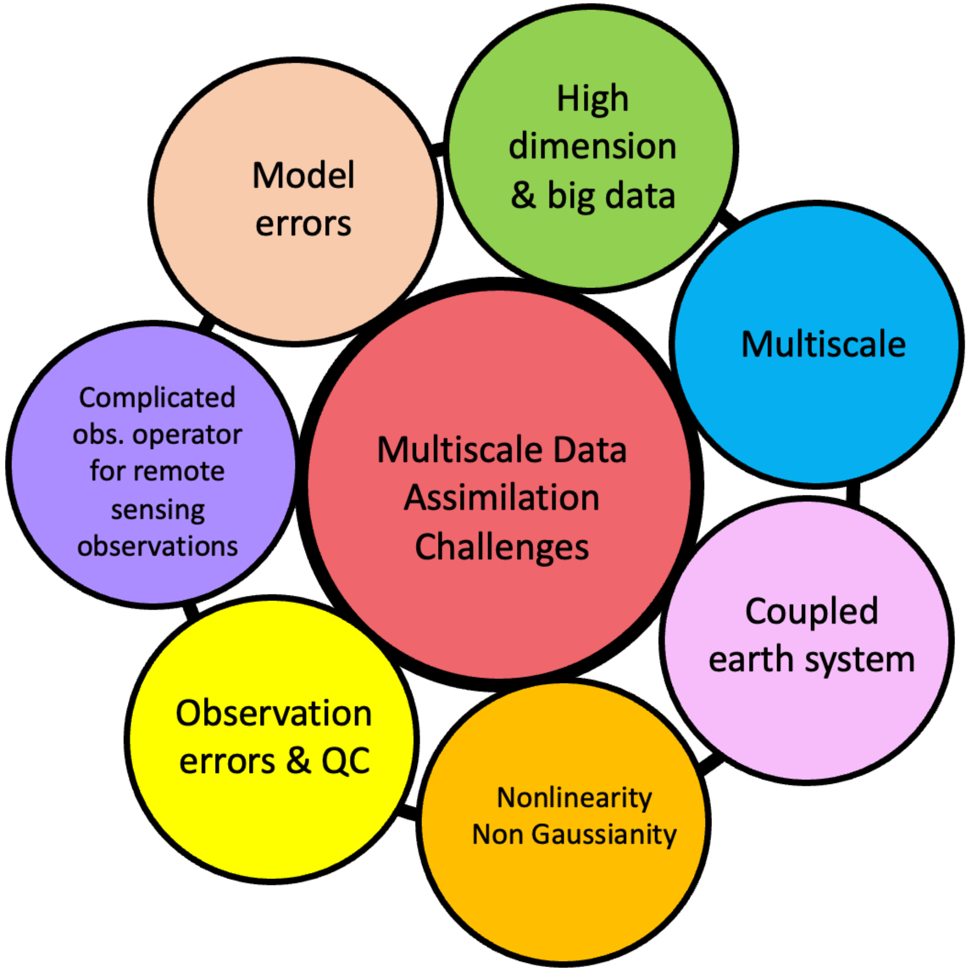
Vertical velocity



Yang* and Wang 2023

- Degrees of Non-Gaussianity/Non-linearity may be scale dependent

MDA definition and challenges



- ❑ The next generation data assimilation system is required to **effectively analyze the state and quantify its uncertainty across multiple scales**, termed as "multiscale data assimilation (MDA)".

OU MAP Lab current MDA efforts

❑ MDA methodology development

- **Develop new simultaneous MDA algorithm/solver**: e.g. MLGETKF (Wang, X. et al. 2021)
- **Develop methods to address ensemble deficiency for simultaneous MDA**

✓ Scale dependent inflation (SDI, Xu* et al. 2023)

✓ Scale dependent horizontal localization (SDL, Huang* et al. 2021, Lu* and Wang 2023)

✓ Scale dependent/flow dependent vertical localization (vFDL, Jones* and Wang 2023b)

✓ Variable dependent localization (VDL, Wang* and Wang 2023a)

✓ Multi-resolution (MR) background ensemble (Kay* and Wang 2020, Jones* and Wang 2023a)

✓ Optimizing coupled earth system component covariances, e.g. air-sea coupling for TC (Lu* et al. 2023)

OU MAP Lab current MDA efforts

❑ R&D of MDA for real NWP applications in US NOAA DA systems

- MR and SDL for **GFS** 4DEnVar (Kay* and Wang 2020, Huang* et al. 2021, Jones* and Wang 2023a)
- SDLVDL for **HRRR/RRFS and WoF** EnVar (Wang* and Wang 2023ab)
- SDL for **HAFS** EnVar (Lu* and Wang 2023)
- Coupled DA for **HAFS-MOM6** (Lu* et al. 2003)

GFS: US operational global model
HRRR/RRFS: US current and next generation convection allowing DA and modeling system over CONUS
HAFS: US next generation convection allowing hurricane DA and modeling system
WoF: US experimental DA and modeling system for tornado, severe thunderstorm, etc

Part I: A Multi-Resolution (MR) Ensemble 4DEnVar for GFS

Kay* and Wang, 2020; Jones* and Wang, 2023a

- Hybrid 4DEnVar (Wang and Lei* 2014)

$$J[(\mathbf{x}_1')_t, \boldsymbol{\alpha}]$$

Single-scale localization

$$= \frac{1}{2}\beta_1(\mathbf{x}_1')_t^{\mathrm{T}}\mathbf{B}_1^{-1}(\mathbf{x}_1')_t + \underbrace{\frac{1}{2}\beta_e(\boldsymbol{\alpha})^{\mathrm{T}}\boxed{\mathbf{A}^{-1}}(\boldsymbol{\alpha})}_{\text{single-res ensemble}}$$

$$+\frac{1}{2}\sum_{t=1}^{L}(\mathbf{H}\mathbf{x}_t' - \mathbf{y}_t^o)^{\mathrm{T}}\,\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_t' - \mathbf{y}_t^o)$$

- Multi-resolution ensemble (Kay* and Wang 2020) hybrid 4DEnVar

$$J[(\mathbf{x}_1')_t, \boldsymbol{\alpha}^L, \boldsymbol{\alpha}^H]$$

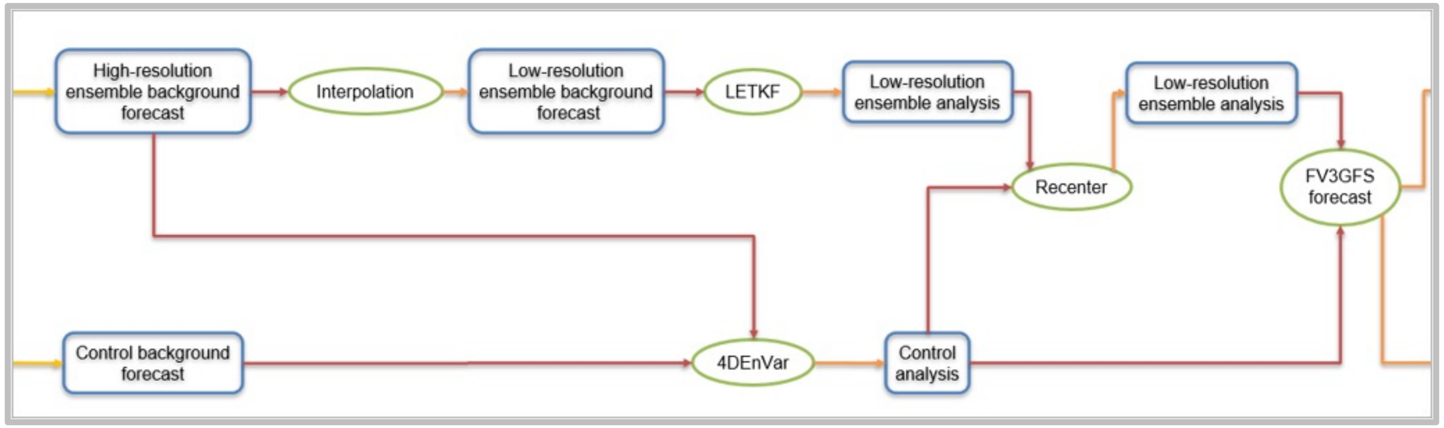
Large localization radius applied for low-res ensemble

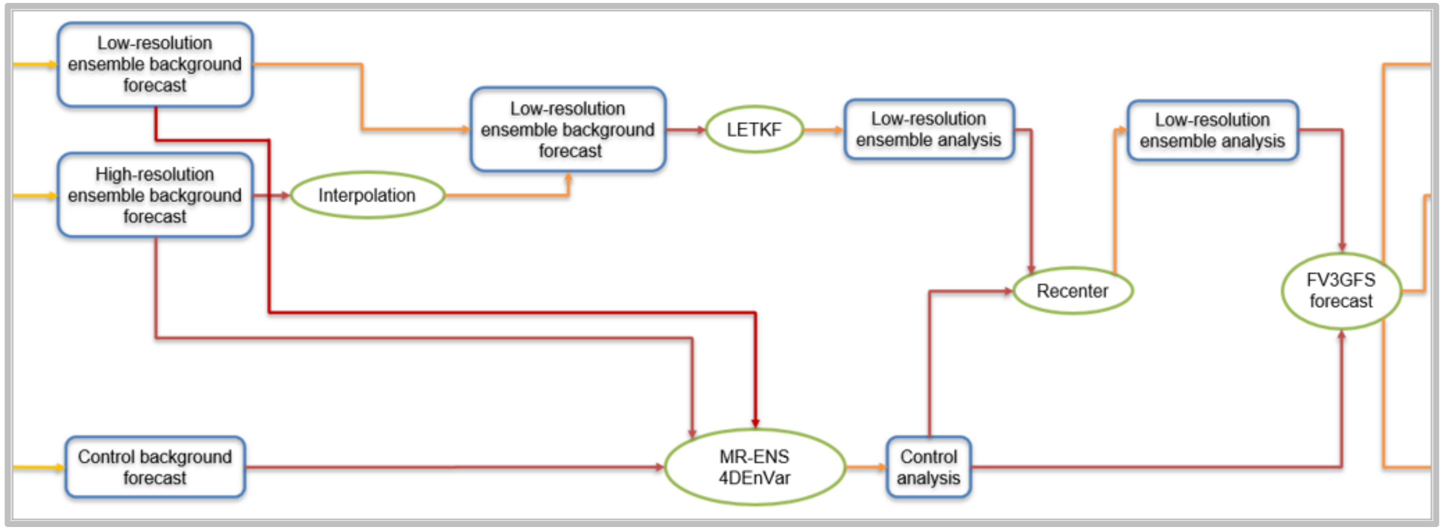
$$= \frac{1}{2}\beta_1(\mathbf{x}_1')_t^{\mathrm{T}}\mathbf{B}_1^{-1}(\mathbf{x}_1')_t + \underbrace{\frac{1}{2}\beta_L(\boldsymbol{\alpha}^L)^{\mathrm{T}}\boxed{\mathbf{A}_L^{-1}}(\boldsymbol{\alpha}^L)}_{\text{low-res ensemble}}$$

$$\underbrace{+\frac{1}{2}\beta_H(\boldsymbol{\alpha}^H)^{\mathrm{T}}\boxed{\mathbf{A}_H^{-1}}(\boldsymbol{\alpha}^H)}_{\text{high-res ensemble}}$$

Small localization radius applied for high-res ensemble

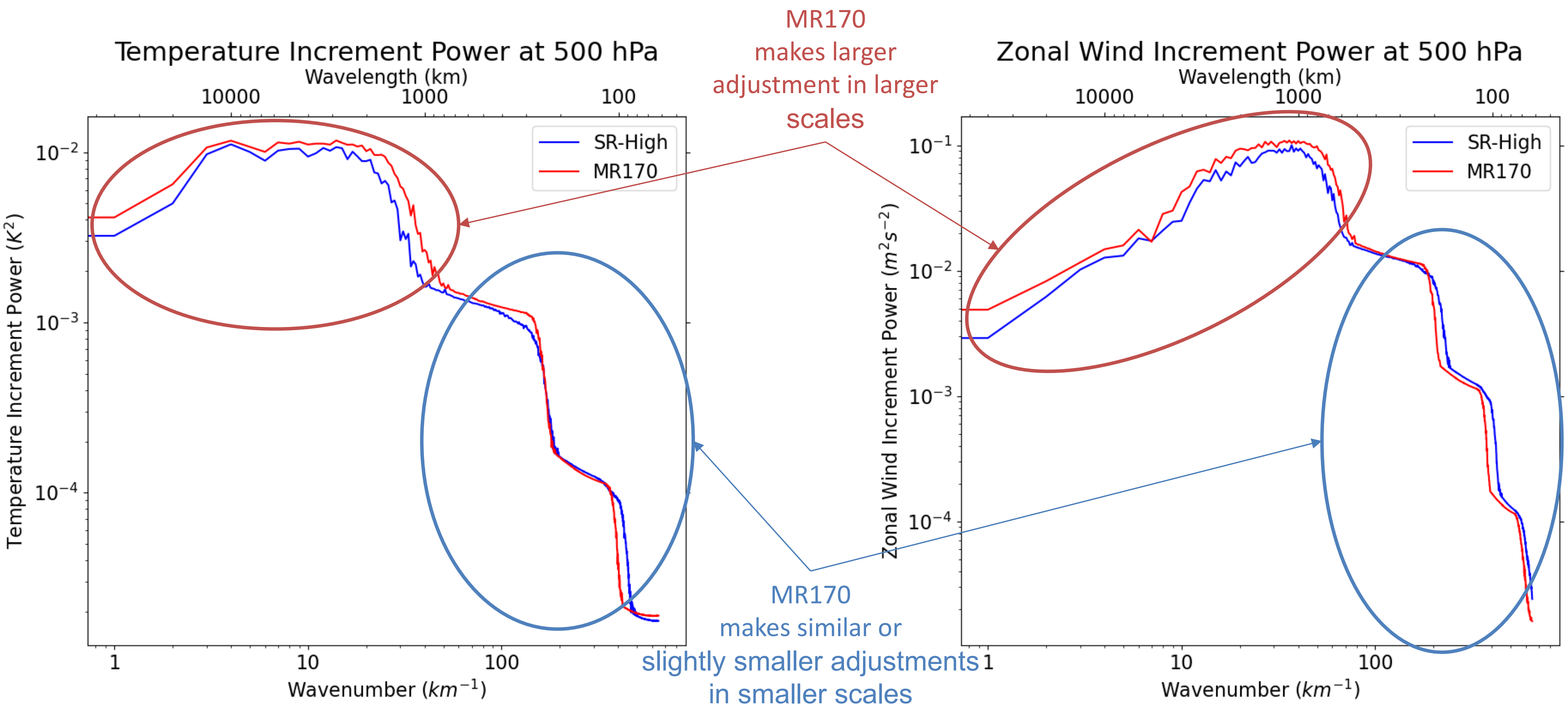
$$+\frac{1}{2}\sum_{t=1}^{L}(\mathbf{H}\mathbf{x}_t' - \mathbf{y}_t^o)^{\mathrm{T}}\,\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_t' - \mathbf{y}_t^o)$$





Analysis Increment Power Spectrum

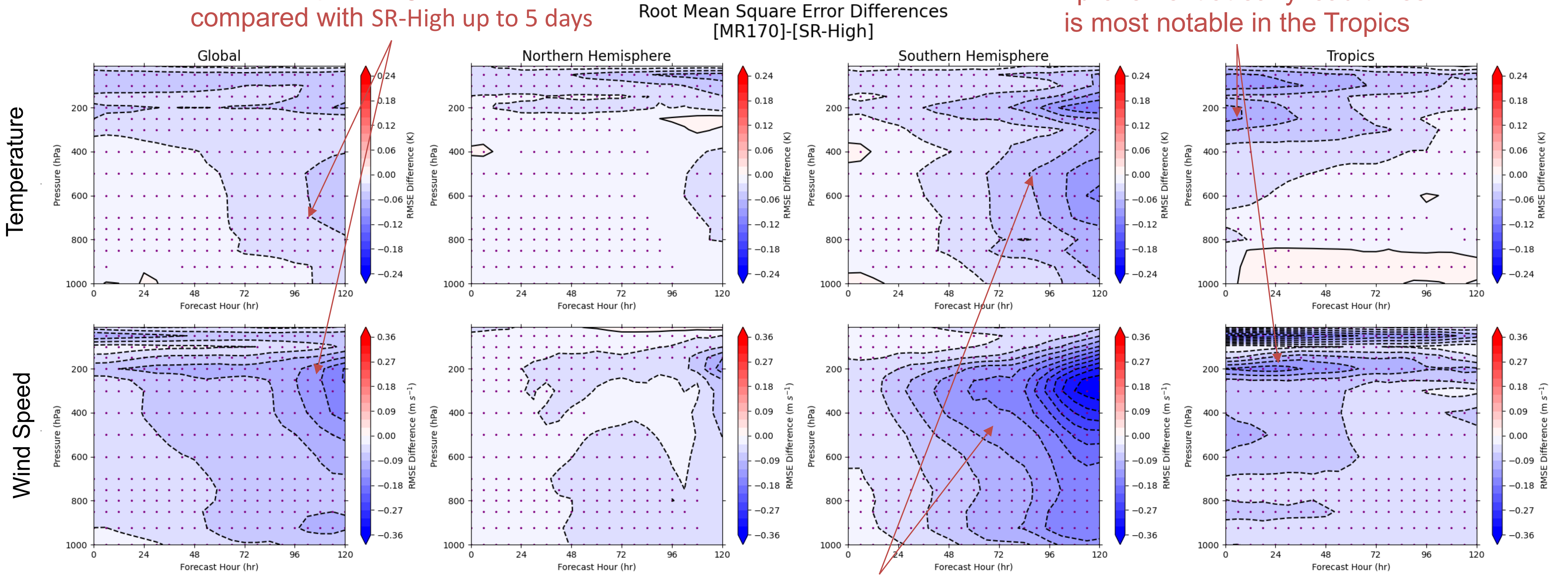
Jones* and Wang, 2023a



Impact on GFS forecast RMSE from cycled NH summer month long experiment

MR170 improves global forecast compared with SR-High up to 5 days

Improvement at early lead times is most notable in the Tropics



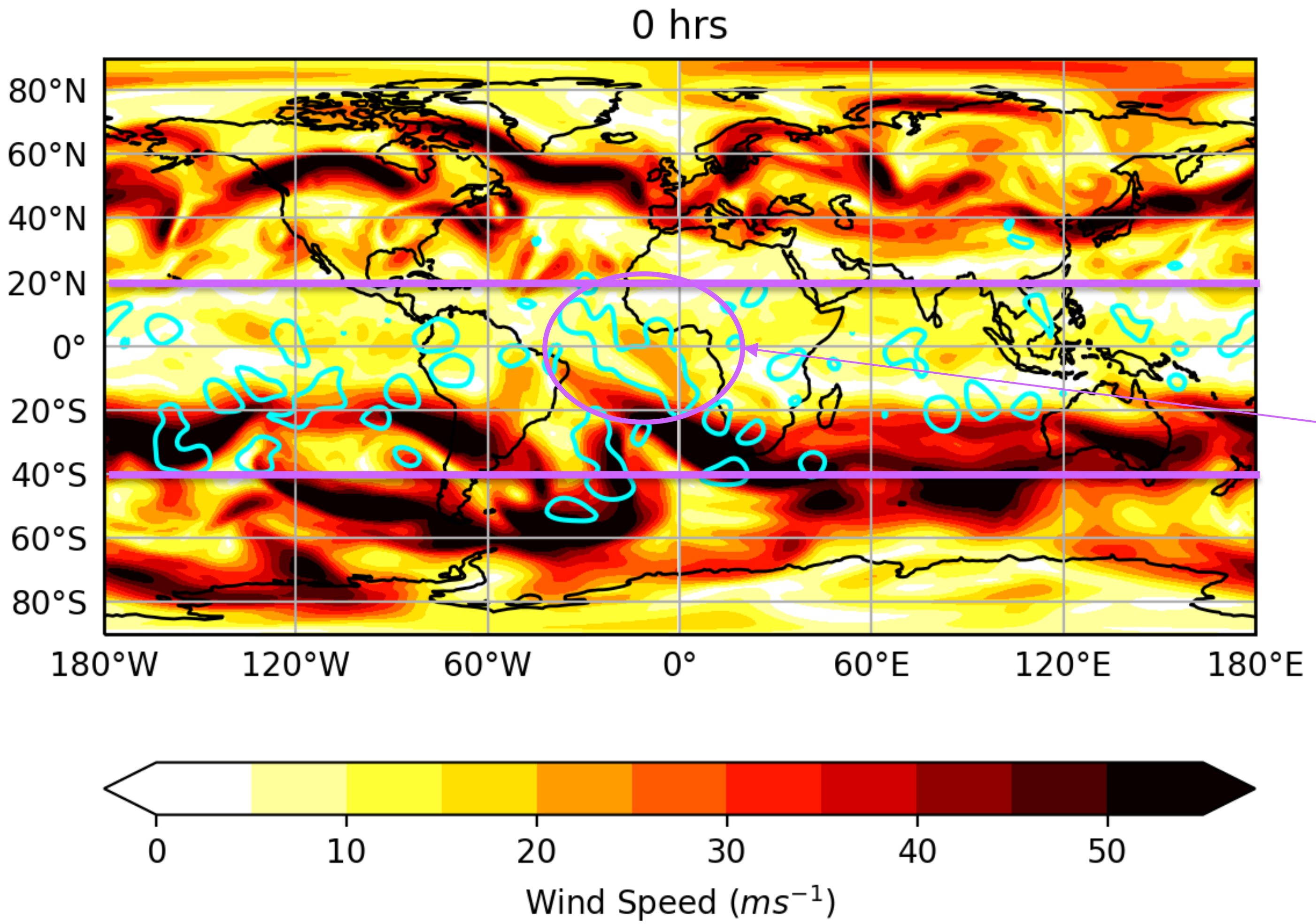
Improvement at longer lead times is most notable in the Southern Hemisphere

Error calculated using ERA-Interim as verification. Purple asterisks indicate 95% confidence using a paired *t*-test.

Why Multi-Resolution (MR) Ensemble 4DEnVar helps?

250 hPa Wind Speed and Difference in Total Energy Error

At analysis time, largest MR170 improvement in the tropics and SH subtropics

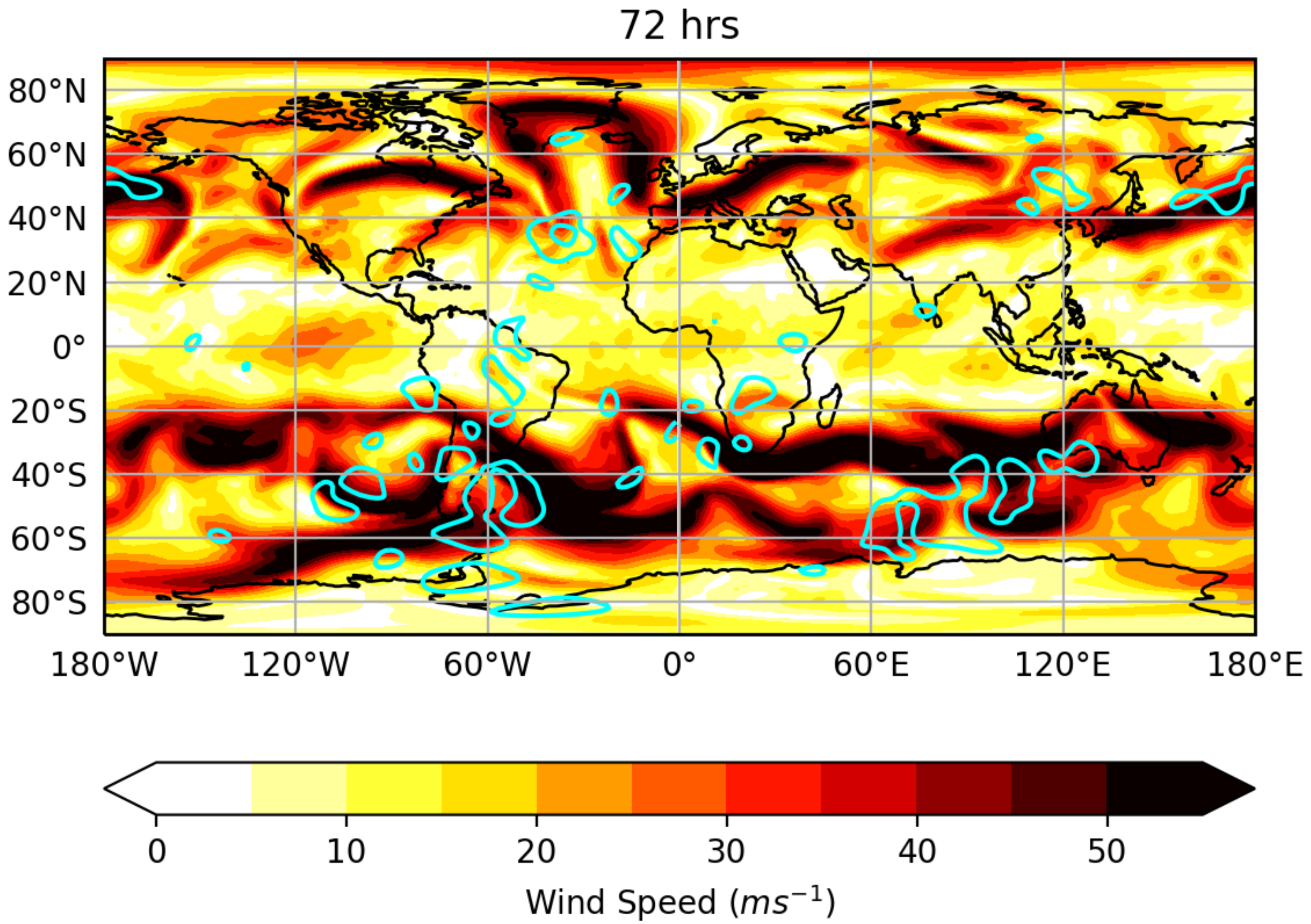


Largest area of improvement in region typically associated with Tropical Easterly Jet

Analysis time: 0000 UTC on 12 September 2017. Cyan contours indicate the 5% maximum improvement of total energy error filtered to include wavenumbers 5 to 25 for MR170 compared with SR-High

Why Multi-Resolution (MR) 4DEnVar helps?

Largest MR170 improvement shifts to extratropics, especially in SH



Largest areas of improvement tend to occur in regions influenced by jet interactions

Analysis time: 0000 UTC on 12 September 2017. Cyan contours indicate the 5% maximum improvement of total energy error filtered to include wavenumbers 5 to 25 for MR170 compared with SR-High

Part II: Simultaneous MDA with scale dependent localization (SDL) in GFS and HAFS 4DEnVar

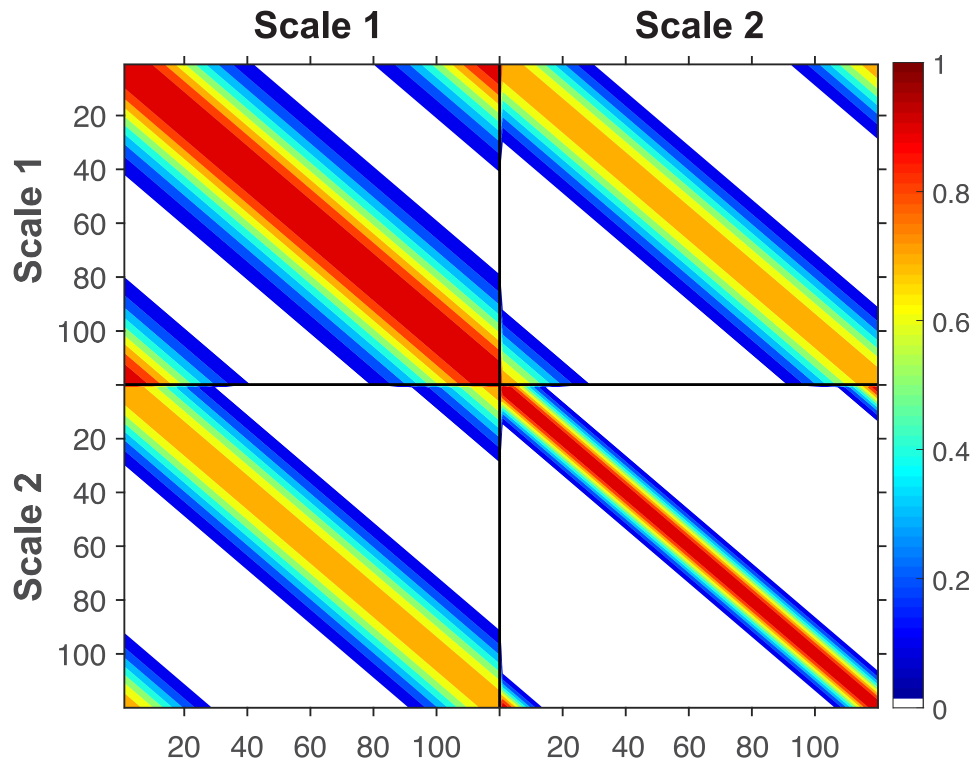
Huang* et al. 2021, Lu* and Wang 2023

- In the operational 4DEnVar, horizontal localization functions are scale-invariant at each level

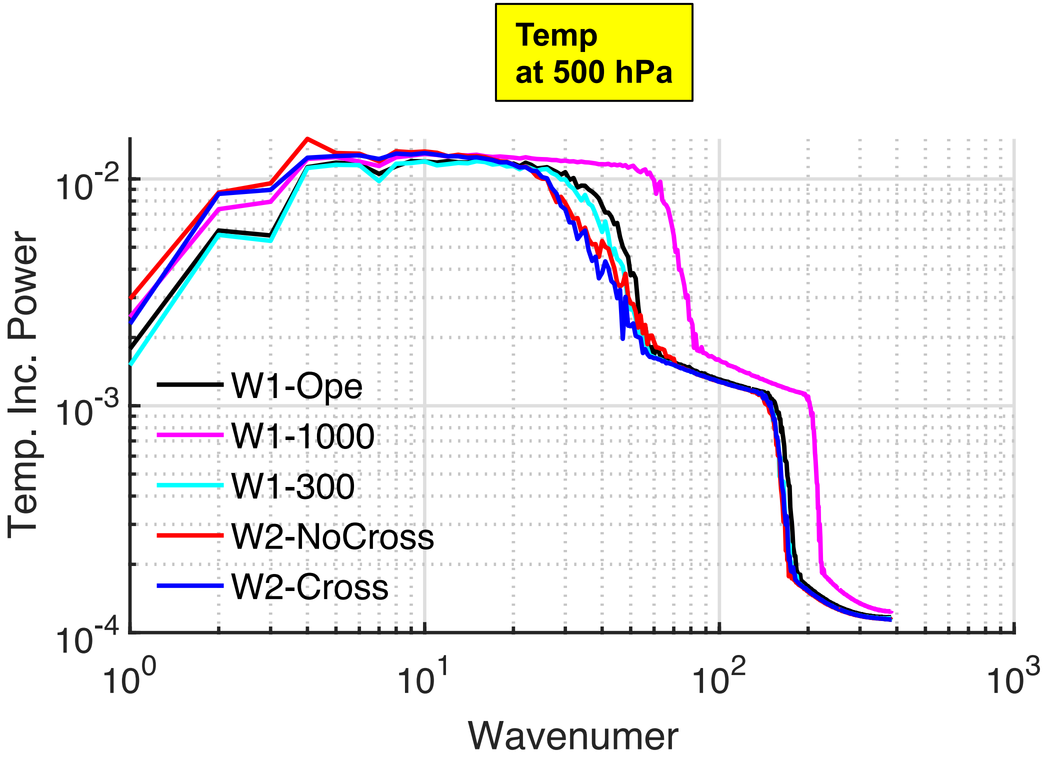
- A simultaneous multiscale DA using scale dependent localization (SDL, Buehner and Shlyaeva 2015) in 4DEnVar for NCEP FV3-based GFS is implemented

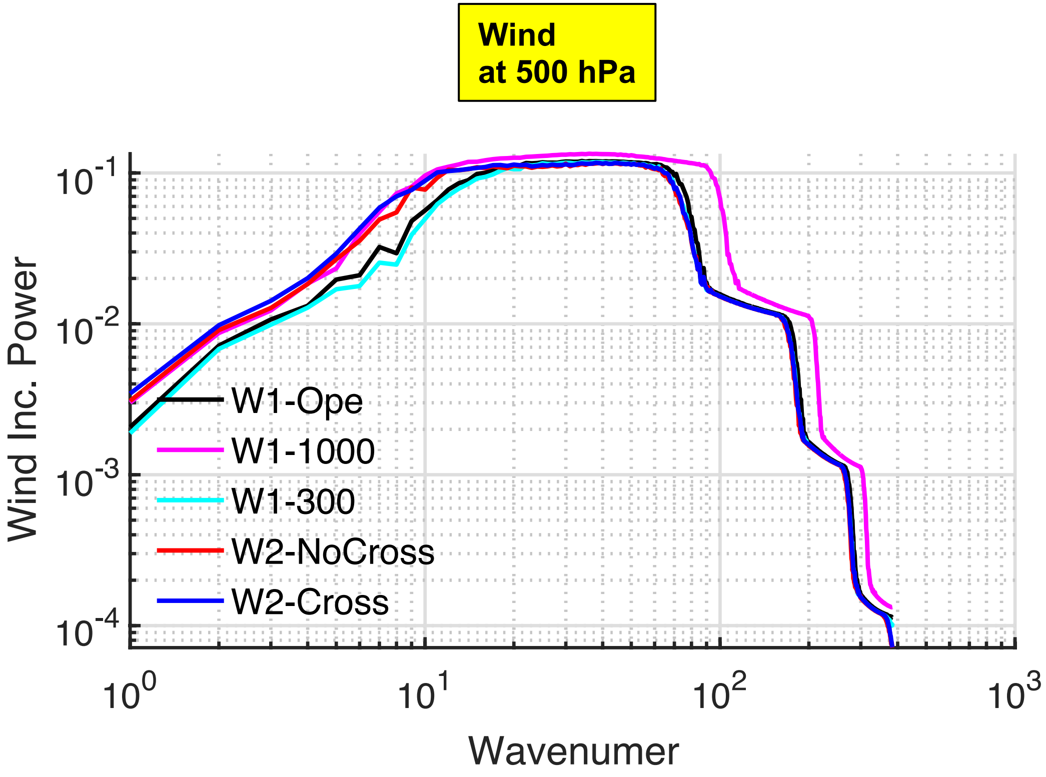
  - ensemble perturbation scale decomposition;
  - scale dependent and cross scale covariance localization

$$J(\mathbf{x}_1', \hat{\mathbf{a}}) = \frac{1}{2}\beta_1(\mathbf{x}_1')^{\mathrm{T}}\mathbf{B}_1^{-1}(\mathbf{x}_1') + \frac{1}{2}\beta_2(\hat{\mathbf{a}})^{\mathrm{T}}\boxed{\hat{\mathbf{A}}}^{-1}(\hat{\mathbf{a}})$$
$$+\frac{1}{2}\sum_{t=1}^{L}(\mathbf{y}_t^{o\prime} - \mathbf{H}_t\mathbf{x}_t')^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{y}_t^{o\prime} - \mathbf{H}_t\mathbf{x}_t'),$$



Analysis Increment Power

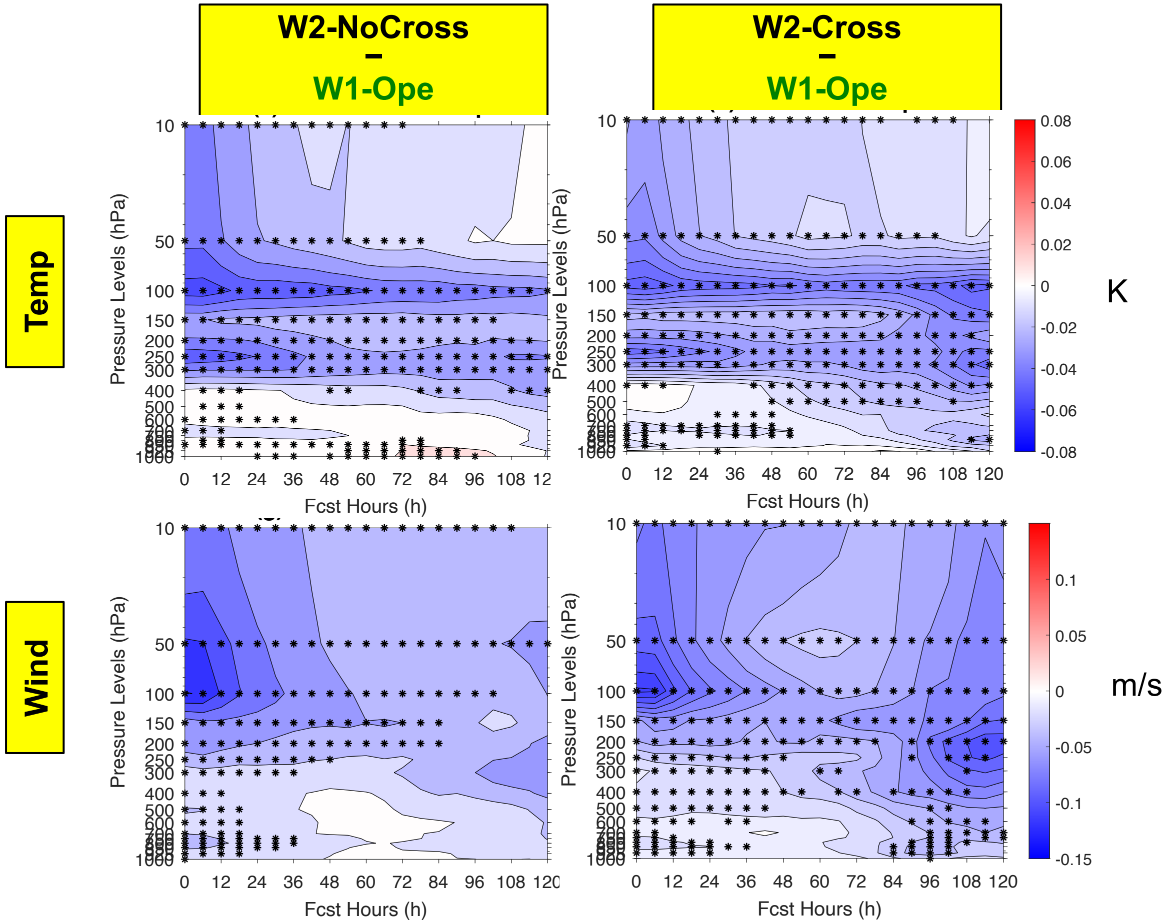




- By comparing W1 experiments, wider localization length results in larger analysis increment power.
- As expected, analysis increment power in W2-NoCross and W2-Cross is closer to W1-1000 (W1-300) at small (large) total wavenumbers **-> MDA with SDL can simultaneously update large and small scales**

Impact on GFS forecast from month long cycled experiments

RMSE difference (blue/red→ improvement/degradation relative to W1-Ope)

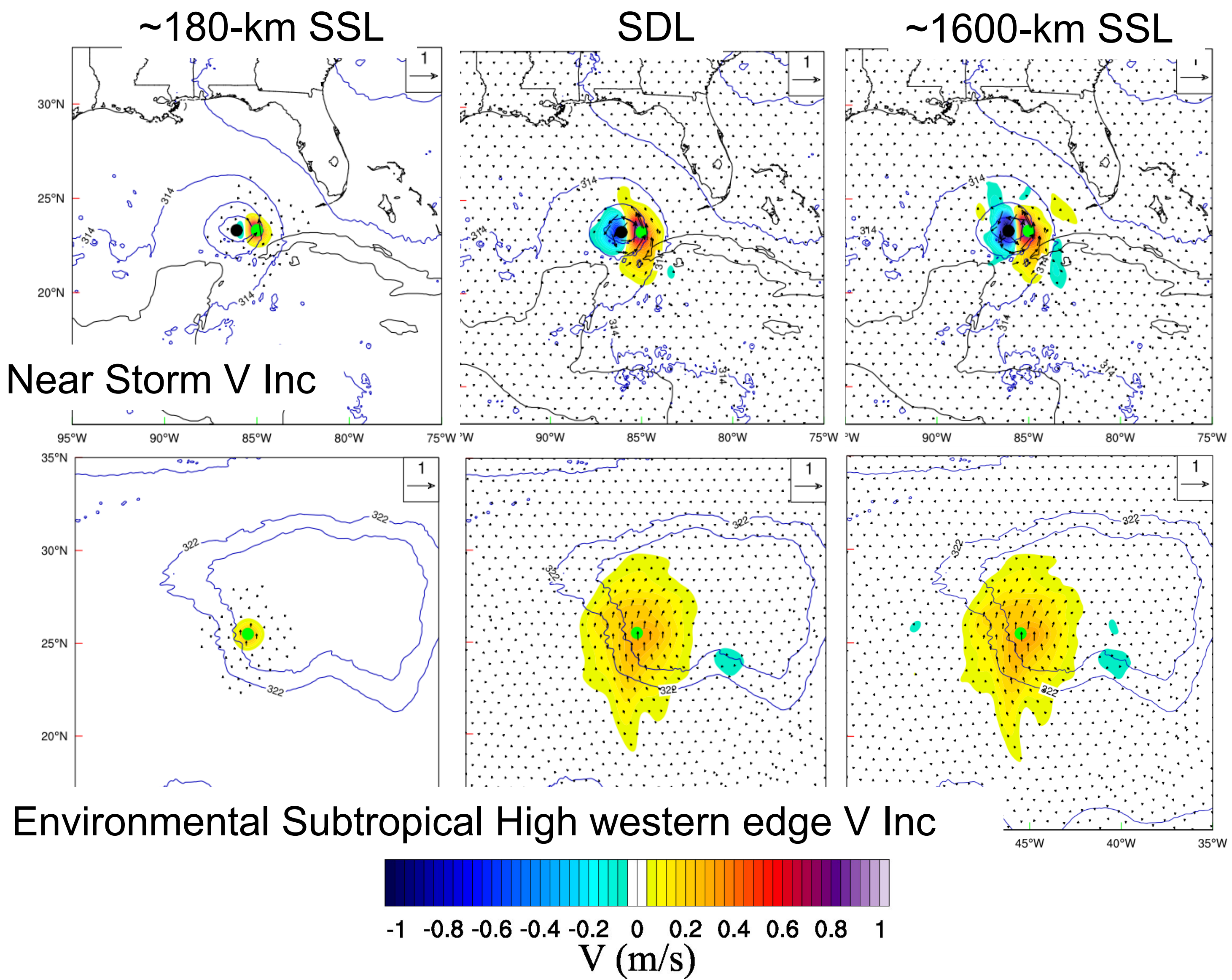


- ❑ MDA with SDL improves global forecasts almost at all pressure levels over operational approach.
- ❑ NOAA pre-implementation test shows similar global forecast improvement (Kleist et al. 2023)

19

Simultaneous MDA with SDL in HAFS 4DEnVar

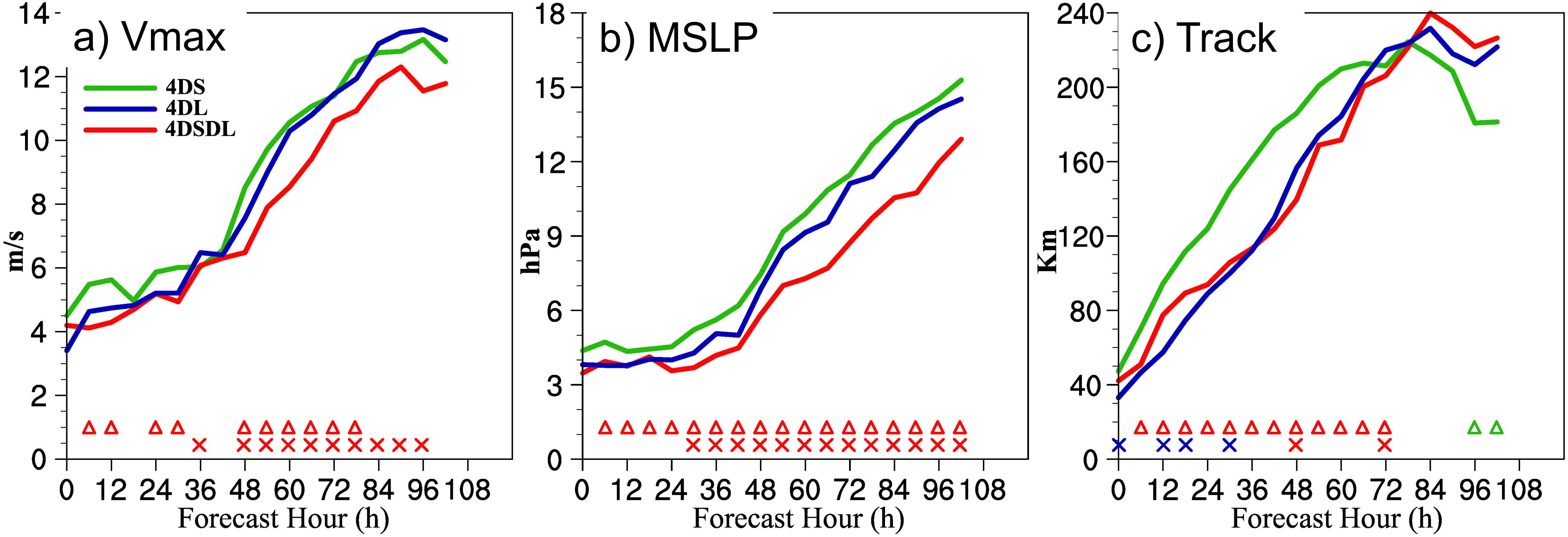
Lu* and Wang 2023a



- The simultaneous MDA with SDL is recently further extended and implemented for HAFS EnVar
- MDA with SDL can simultaneously properly correct both the TC and its large scale steering environment (subtropical high)

Impact on convection allowing hurricane forecasts

Mean Absolute Forecast Errors during Hurricane Laura (2020)

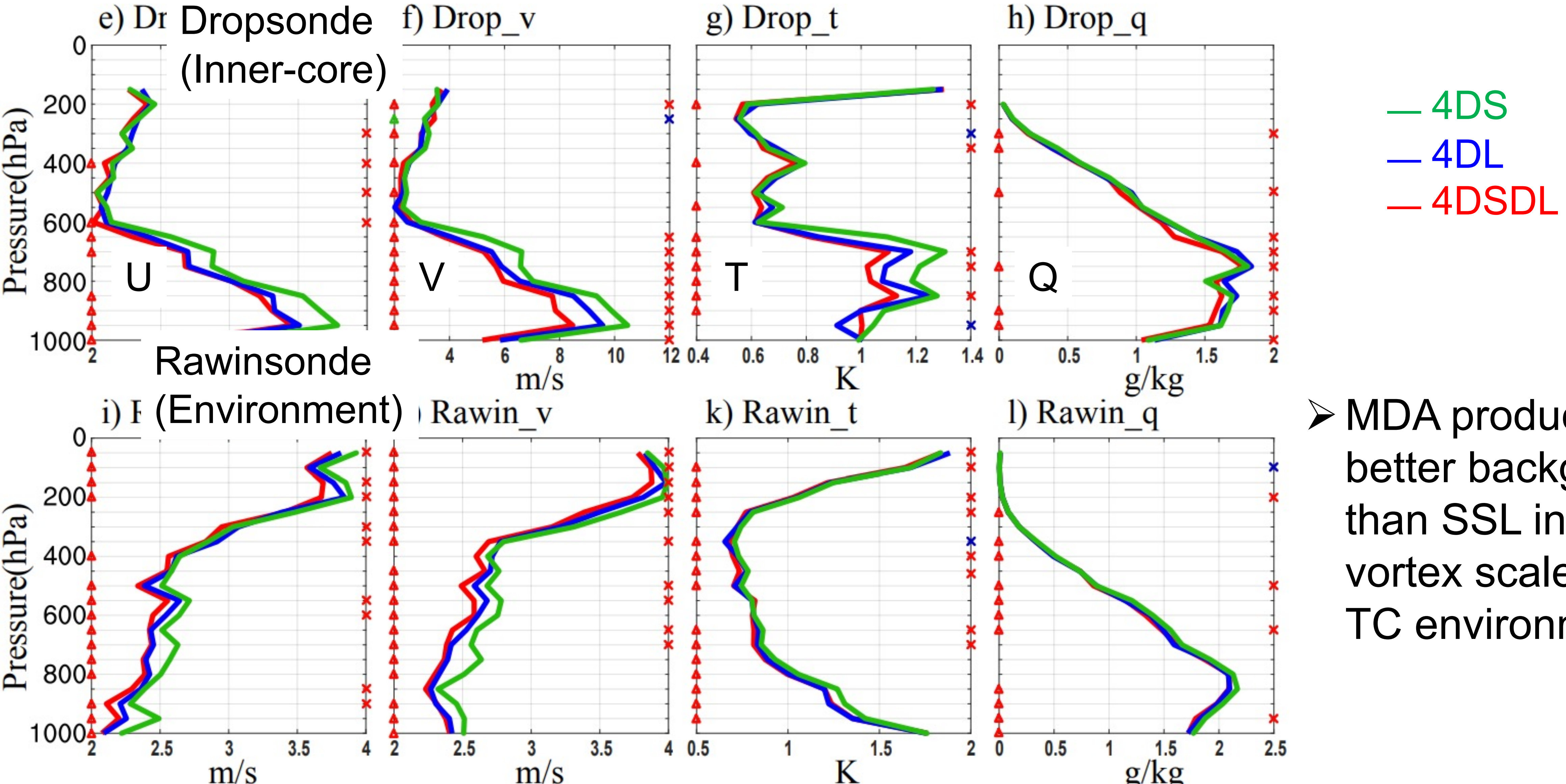


➢ ***Localization Impacts on Forecasts***

- 4DSDL outperforms 4DL and 4DS in intensity predictions
- 4DSDL outperforms 4DS in track prediction and has mixed results compared to 4DL

Impact of Simultaneous Multiscale DA

6-h Background Forecast Verification against Dropsonde & Rawinsonde

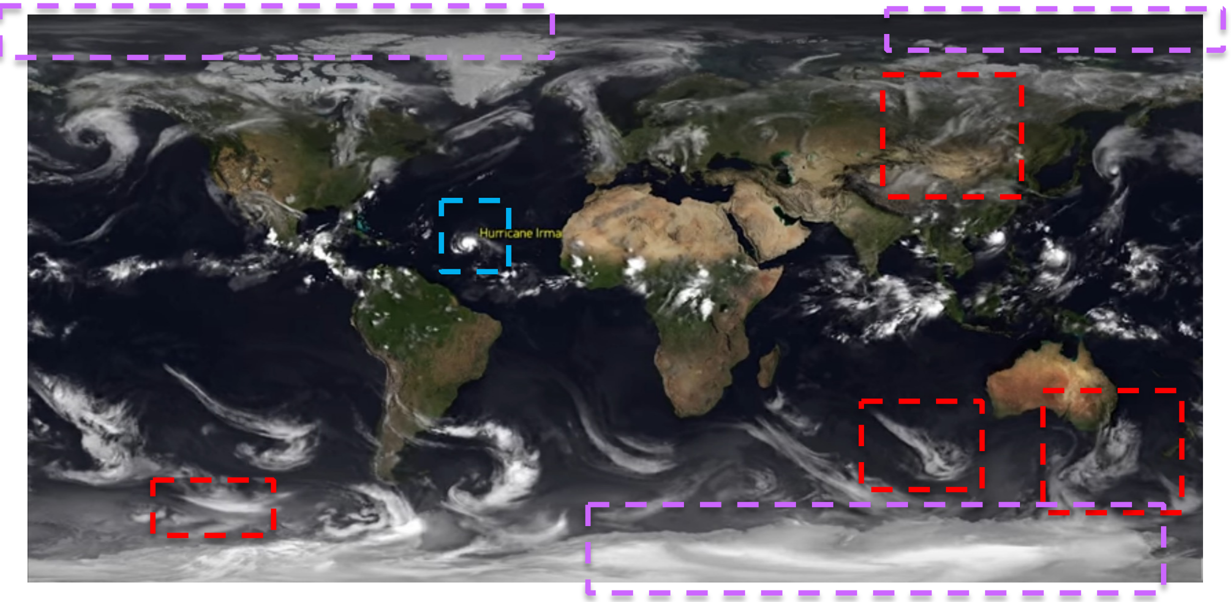


- MDA produces better background than SSL in both vortex scale and TC environment.

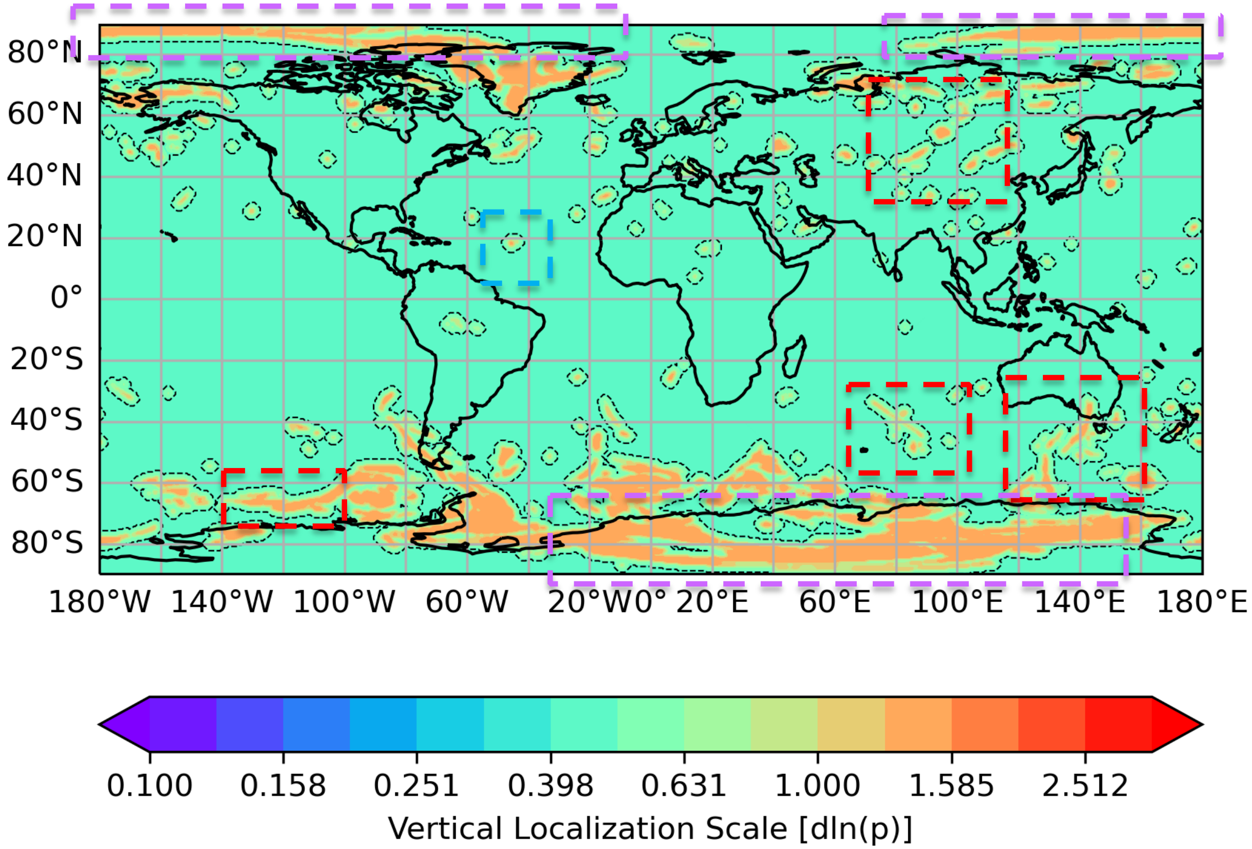
Part III: Simultaneous MDA with flow dependent vertical localization (vFDL) for GFS 4DEnVar

Jone* and Wang 2023b

- Areas identified for larger vertical localization lengths tend to be within:
  - Tropical cyclones
  - Frontal regions
  - Broad polar regions

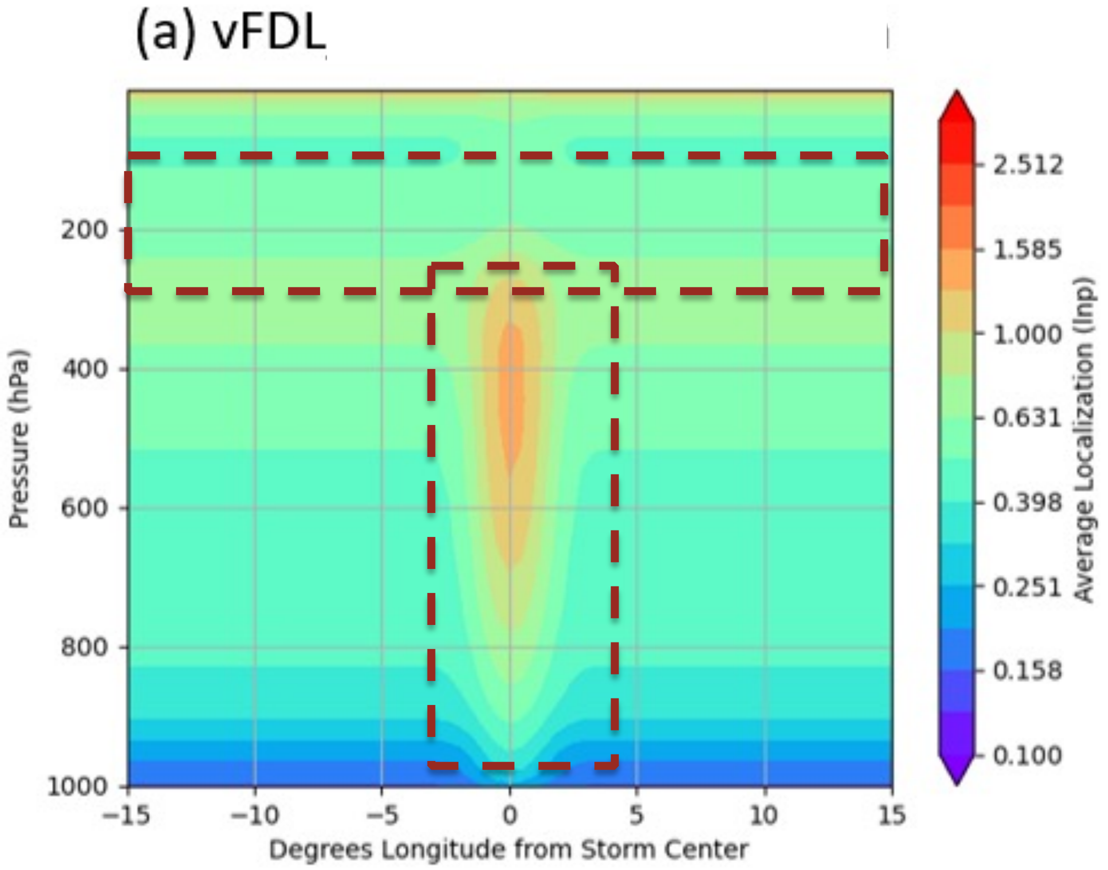


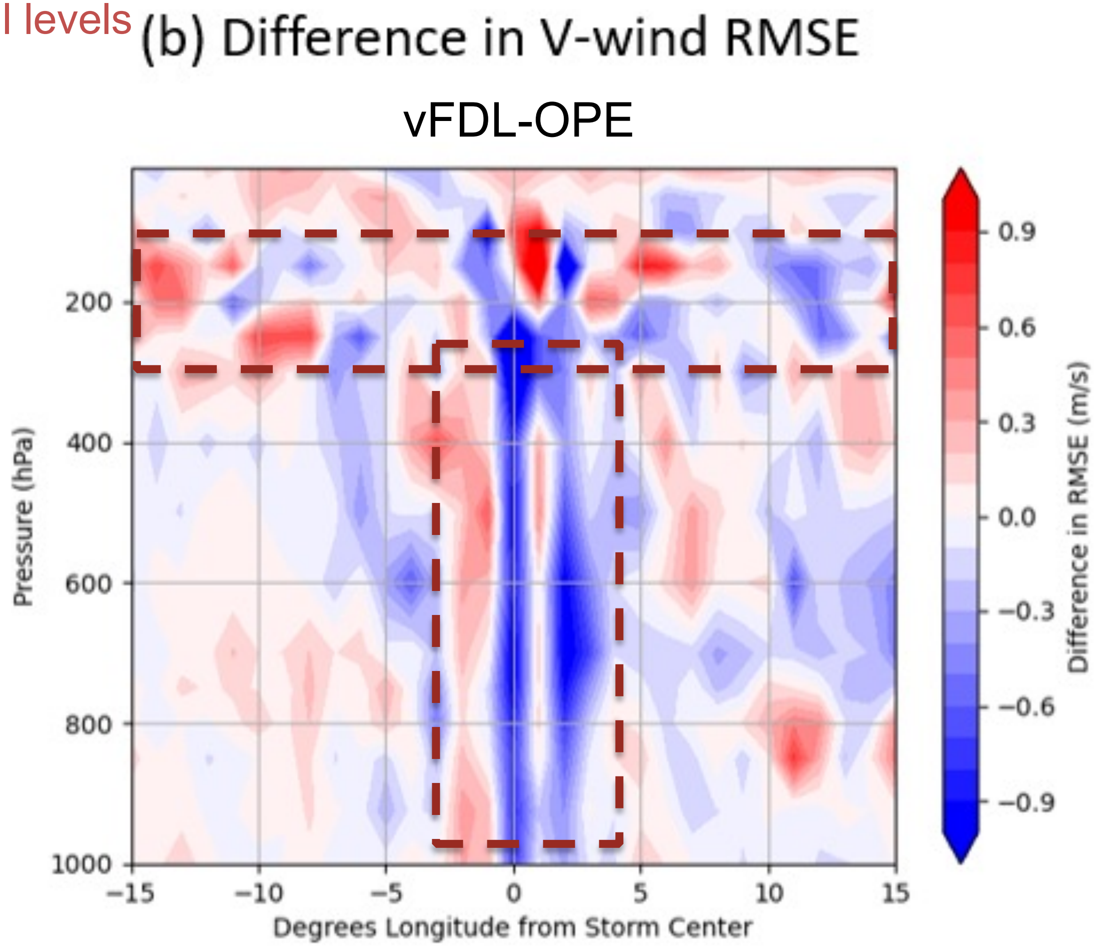
Courtesy of EUMETSAT



Hurricane Irma (2017) Composite Localization and Difference in RMSE Cross Sections

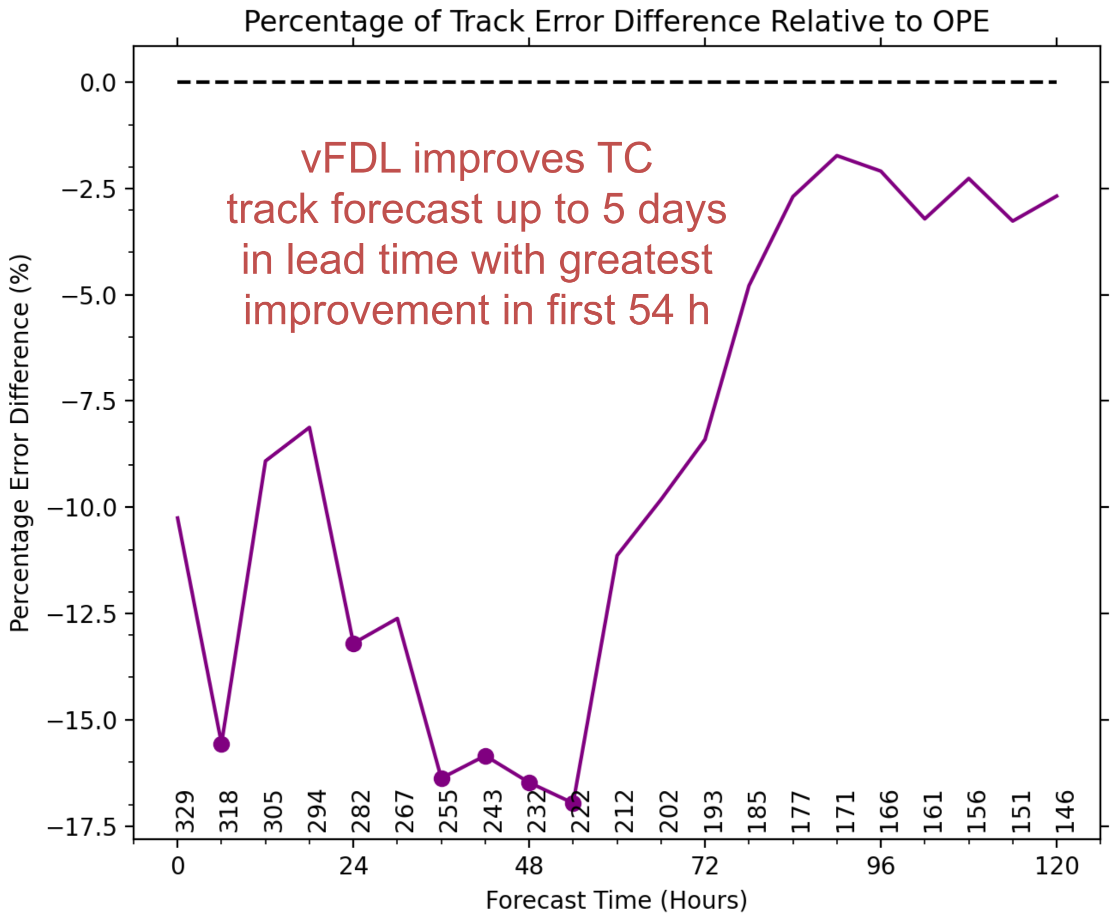
Largest differences in RMSE tend to occur at ~200 hPa and in inner core at all levels





Most impact regions correspond with areas Identified for increased vertical localization

Impact on month long GFS hurricane track forecasts

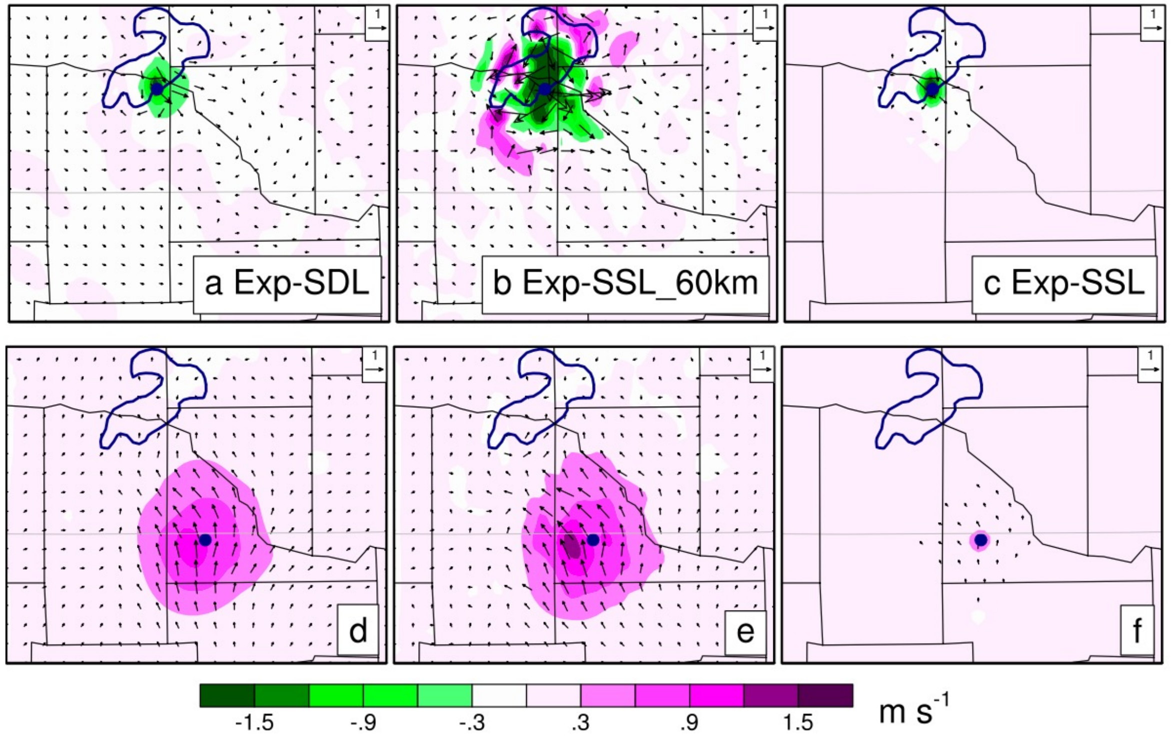


- Flow-dependent vertical localization show promises to improve GFS hurricane track forecasts
- Improved initial position leads to impro ved forecasts for the first 72 h
- Large scale improvement in zonal wind leads to improvement after 72 h

Tracking algorithm by Marchok (2002). Numbers above the x-axis denote how many tracks at each lead time. Filled dots indicate 95% confidence using a paired *t*-test.

Part IV: Simultaneous MDA with scale and variable dependent localization (SDLVDL) for WoF & HRRR/RRFS EnVar

Wang* and Wang 2023ab

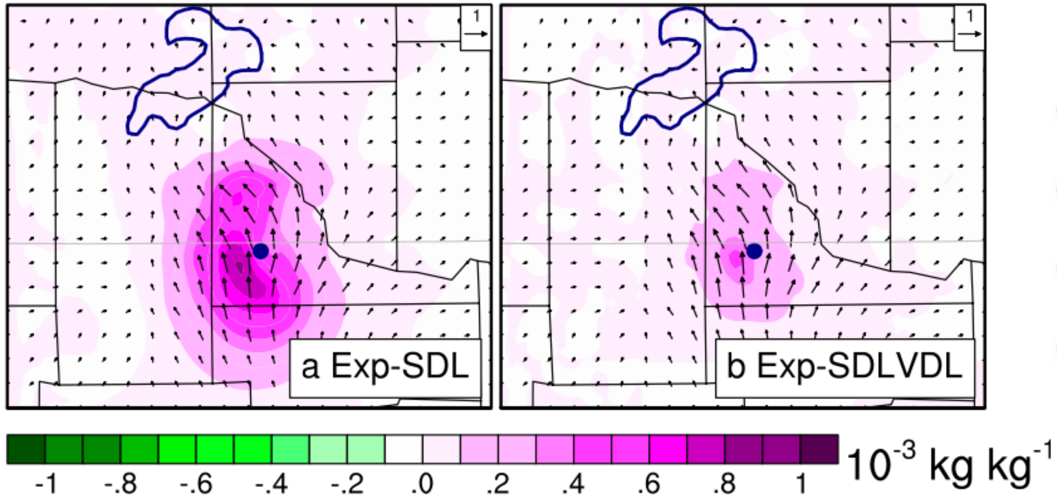


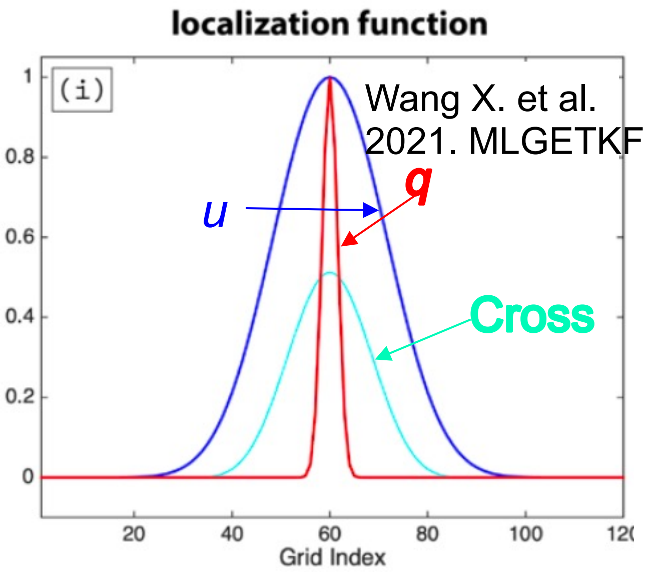
Analysis increments of wind (vector) and v-wind (shaded) through assimilating a single observation of Vr with an innovation of -30 m/s at 1 km AGL.

Wang, Y.* and X. Wang 2023a

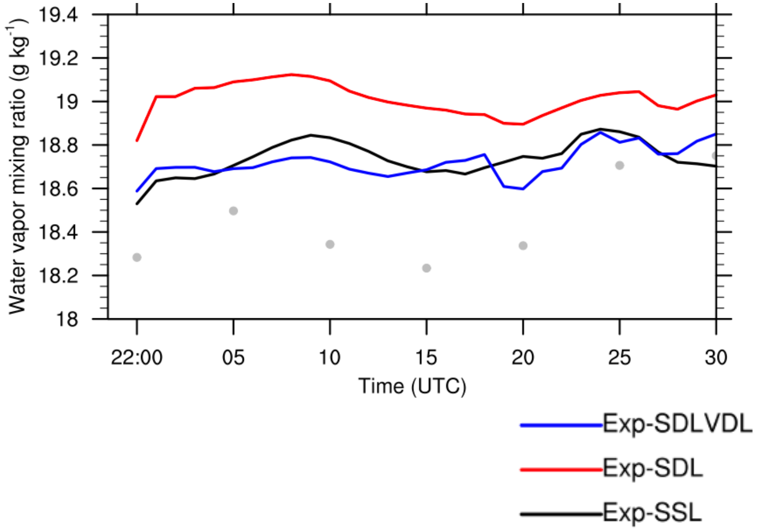
- MDA with SDL can simultaneously properly correct both the supercell storm and its ambient environment, which represent different scales

Impact of including VDL in MDA





Overestimated q in Exp-SDL

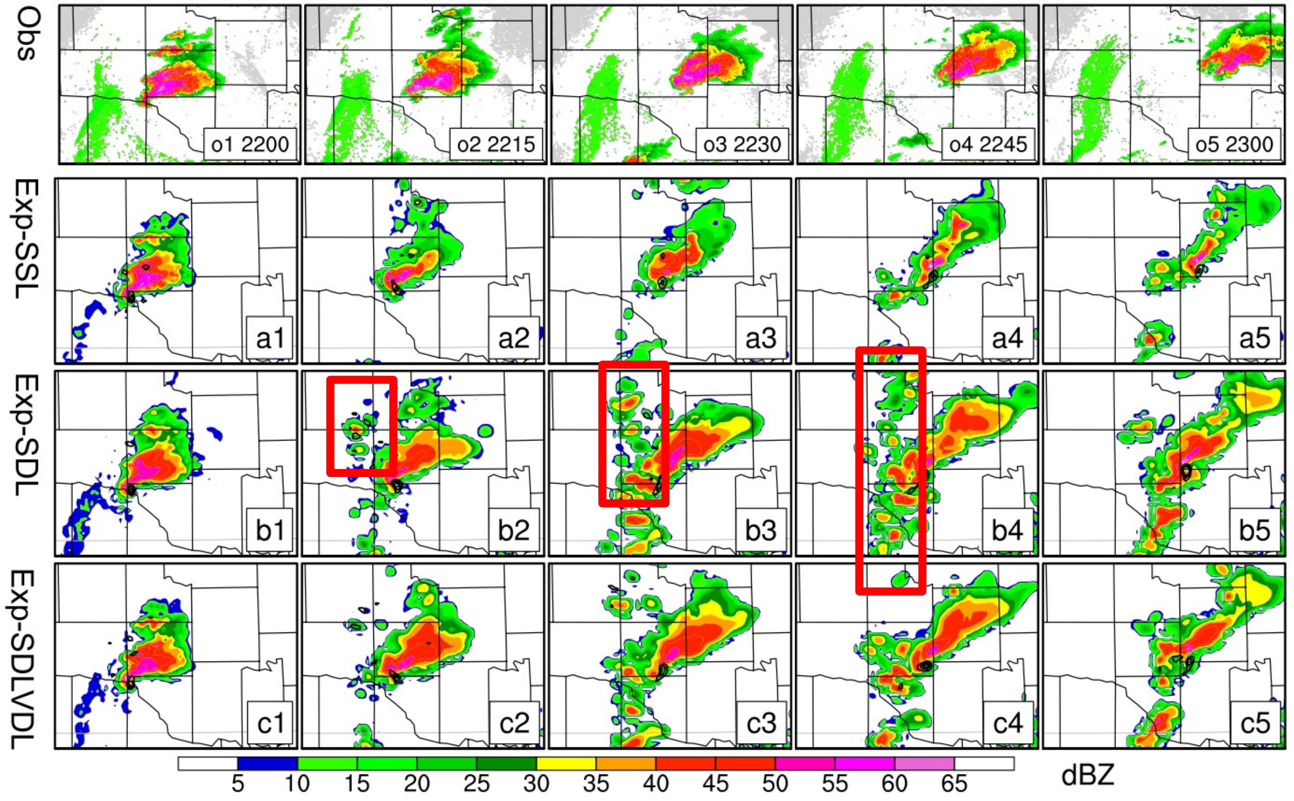


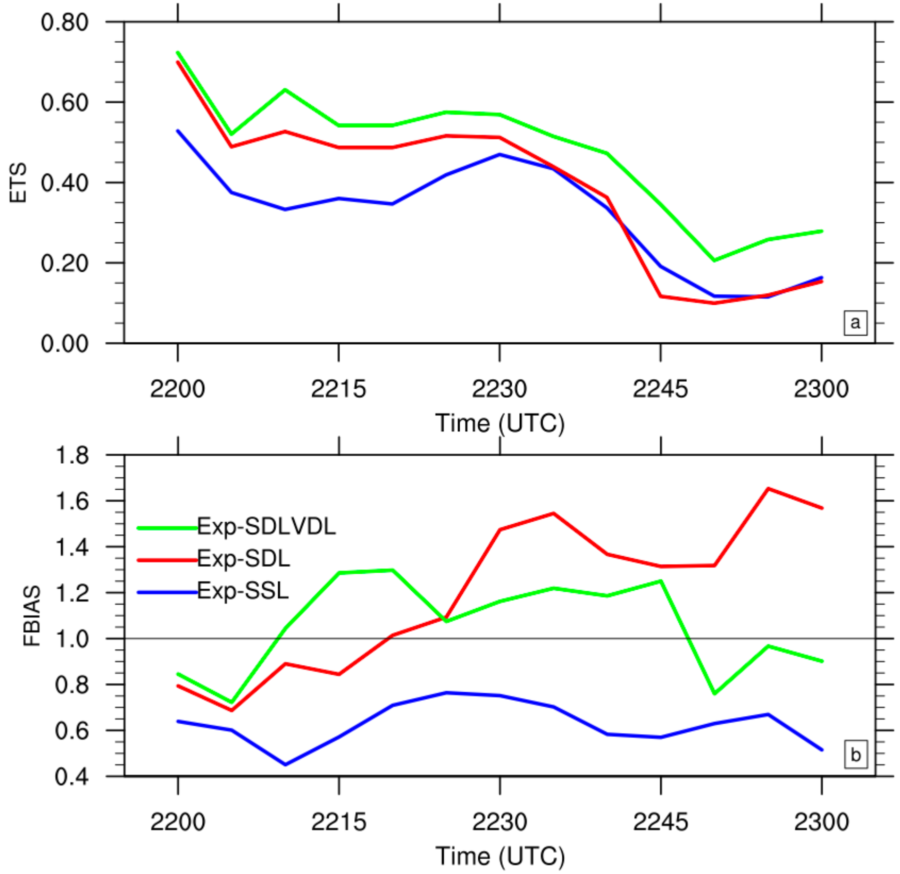
SDLVDL (Wang, Y* and. X. Wang 2023a)

$$\hat{\mathbf{L}}_{m,n} = \begin{bmatrix} \hat{\mathbf{L}}_{m_1,n_1} & \hat{\mathbf{L}}_{m_1,n_2} & \cdots & \hat{\mathbf{L}}_{m_1,n_V} \\ \hat{\mathbf{L}}_{m_2,n_1} & \hat{\mathbf{L}}_{m_2,n_2} & \cdots & \hat{\mathbf{L}}_{m_2,n_V} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{L}}_{m_V,n_1} & \hat{\mathbf{L}}_{m_V,n_2} & \cdots & \hat{\mathbf{L}}_{m_V,n_V} \end{bmatrix},$$

- Using VDL with an appropriately smaller localization for q greatly reduces intensity and coverage of moisture increments compared to using SDL only.

Impact on supercell forecasts
(Reflectivity @ 1km AGL)



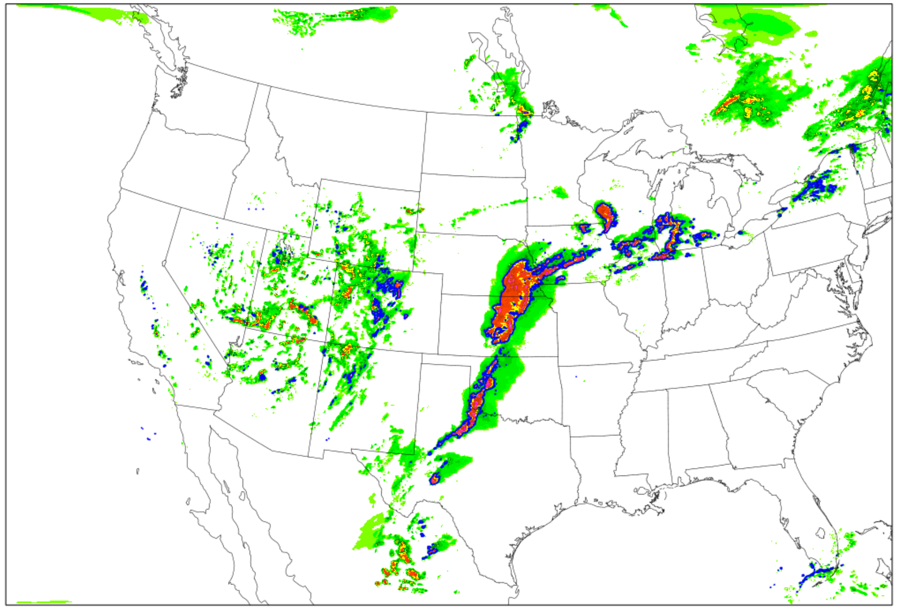


- Compared to Exp-SSL, Exp-SDL and Exp-SDLVDL improve the forecasts for the forward-flank distributions;
- Exp-SDLVDL has less spurious storms and further enhances supercell than Exp-SDL

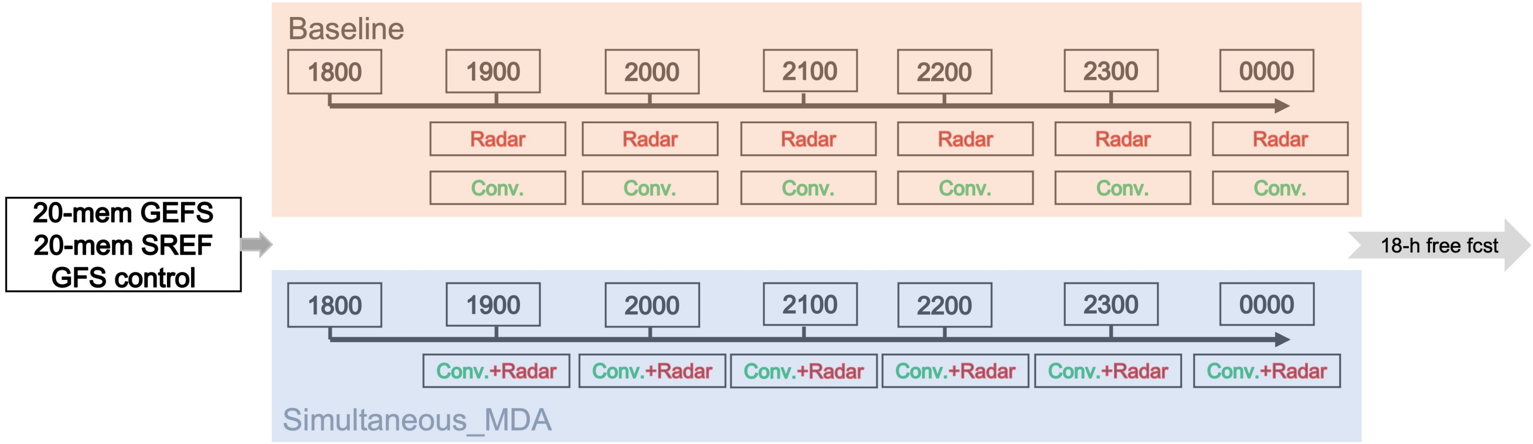
Experiment design (HRRR/RRFS)

Model and DA configuration

Wang, Y.* and X. Wang 2023b

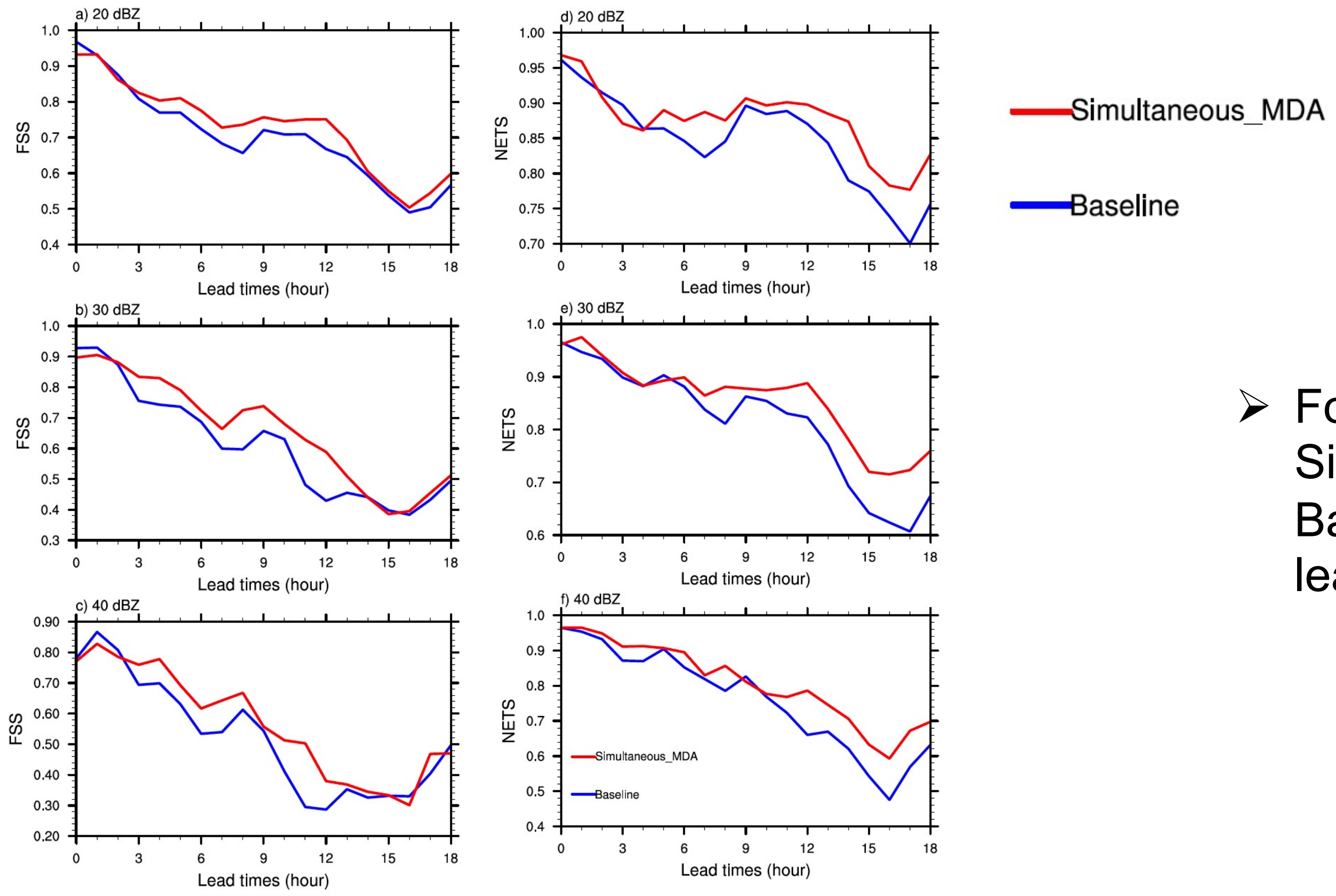


Model options	Specification
Grid size and resolution	1621×1121×51; 3 km
Microphysics	Thompson
PBL	MYNN
Radiation	RRTMG
Land surface model	Noah



Impacts of SDLVDL in EnVar on RRFS/HRRR forecasts

Wang Y.* and X. Wang 2023b

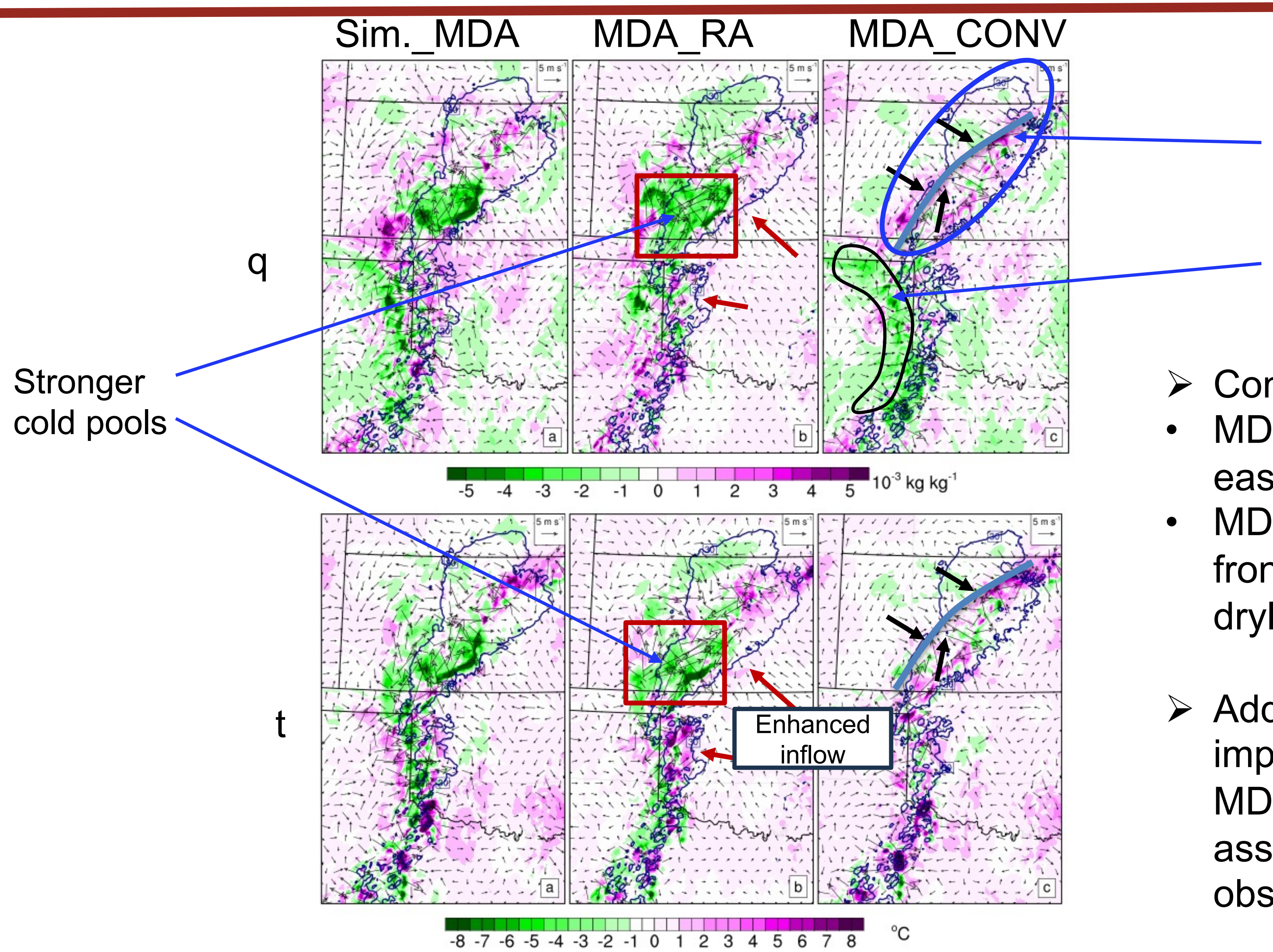


- For composite reflectivity forecast, Simultaneous_MDA outperforms Baseline in the majority of forecast leading time.

FSS and NETS for composite reflectivity

Understanding the impact of Simultaneous_MDA on RRFS/HRRR forecast

Analysis differences with Baseline

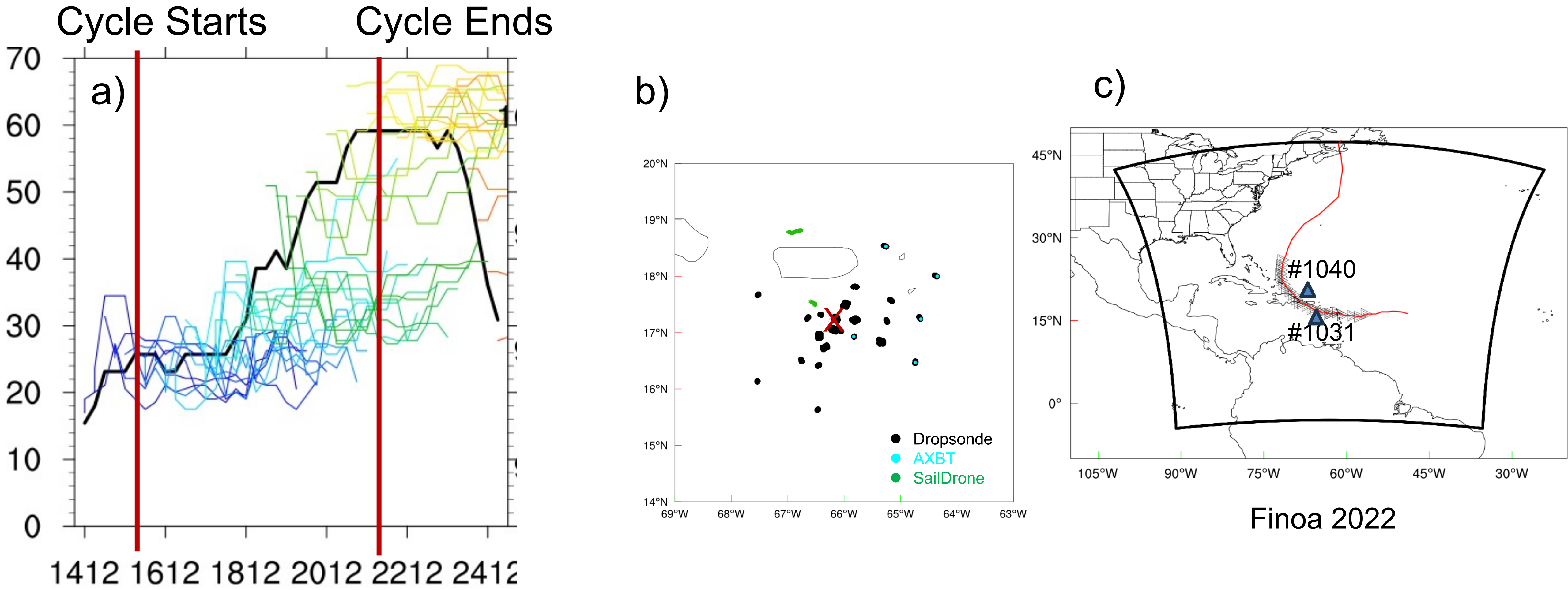


- ➢ Compared to Baseline,
- MDA has stronger cold pools and enhanced east-southeastern inflow;
- MDA produces greater convergence along the front and larger moisture gradient along the dryline.

- ➢ Additional experiments revealed such improvements are due to the effectiveness of MDA to correct multiple scales when assimilating radar or meso/synoptic in-situ observations alone

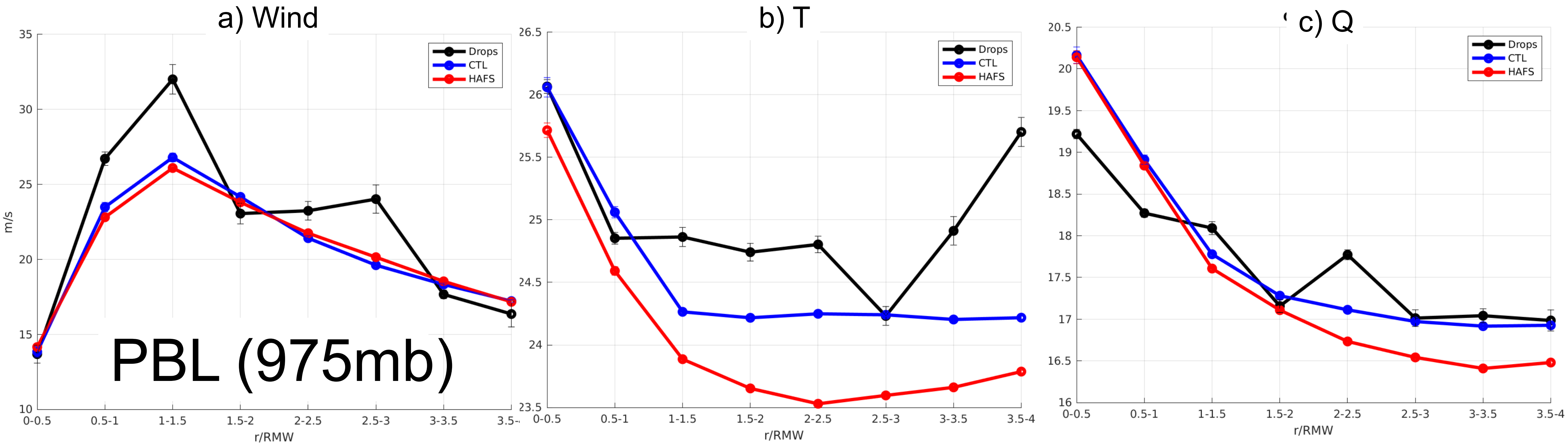
Part V: coupled ocean-atm DA for TC

Lu* and Wang et al. 2023b



Model Performance Verification

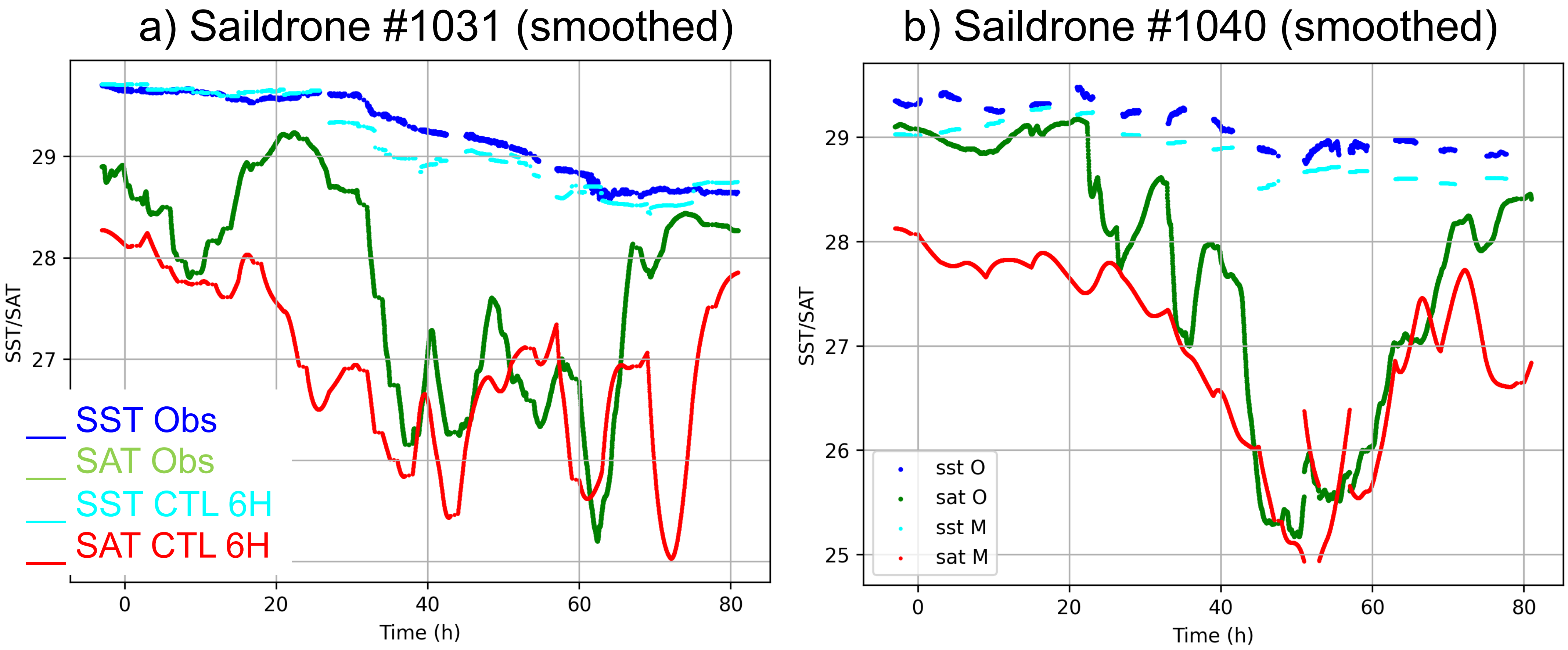
Azimuthal Verifications of PBL Structures from Dropsondes



- CTL experiment matches the dropsonde observations reasonably well.
- CTL performs even slightly better than the HAFS-A operational run in 2022.

Model Performance Verification

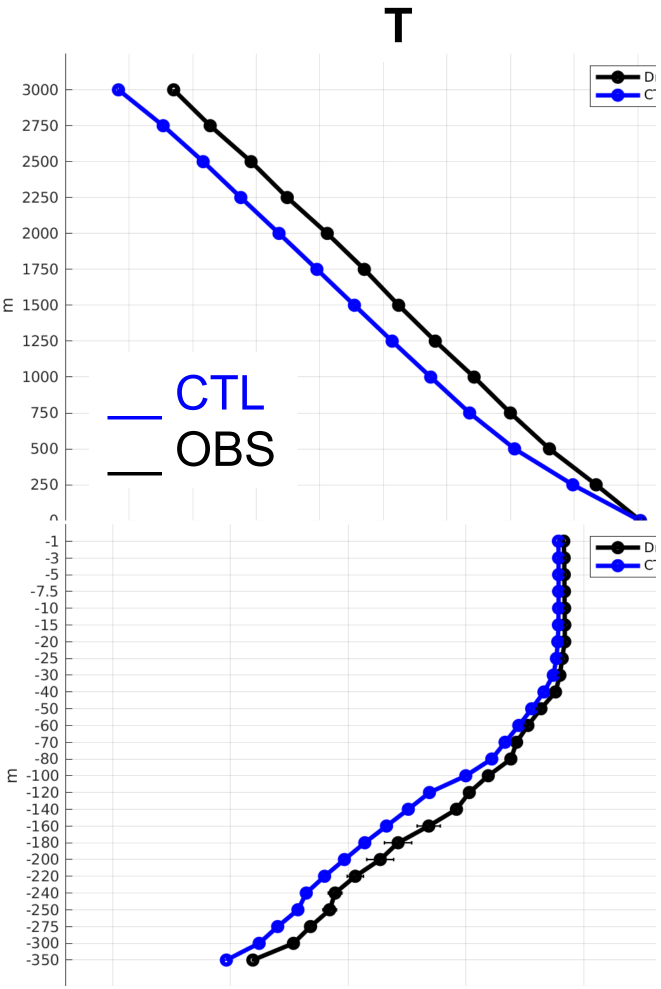
Air-Sea Interface verifications from SailDrones



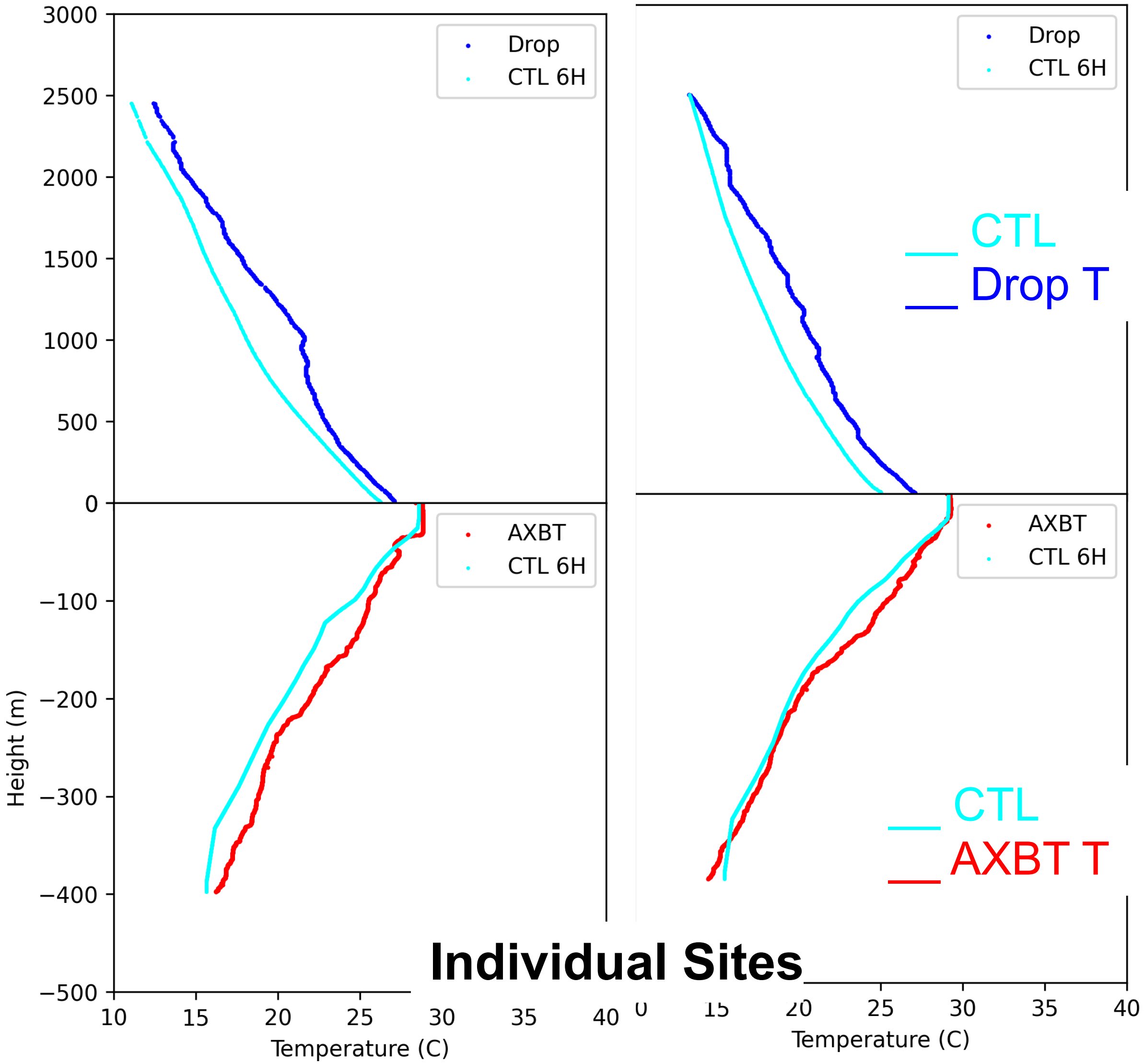
✓ Low bias in both SAT and SST from the model

Model Performance Verification

Air-Sea Interface verifications from Drops + AXBT



Avg @ 1.75-2.25 RMW



Individual Sites

- ✓ Slight negative model biases still exists in both air and sea.

Part VI: MLGETKF and Scale Dependent Inflation

Wang, X. et al., 2021, Xu* et al, 2023

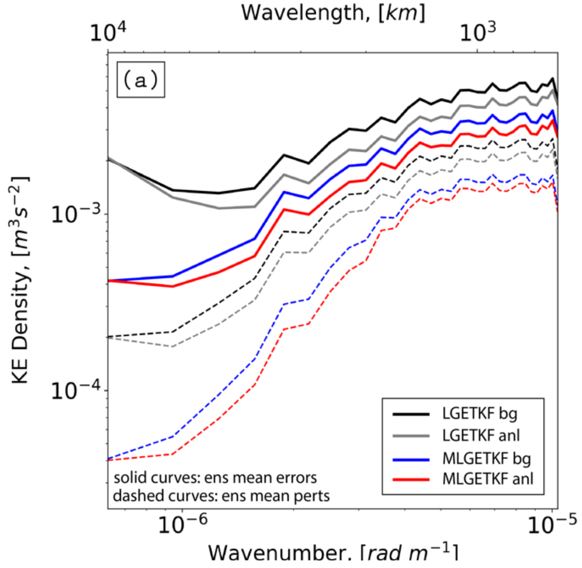
- A new ensemble-based, multiscale data assimilation (MDA) method, MLGETKF (Multiscale Local Gain Form Ensemble Transform Kalman Filter, Wang et al. 2021), was developed.

$$\boxed{\mathbf{Z}^{ML}} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_{K_{ML}}) = (\mathbf{I} \quad \mathbf{I} \quad \cdots \quad \mathbf{I}) \left(\boxed{\mathbf{X}^{MS}} \Delta \left[\left(\boxed{\mathbf{L}^{MS}} \right)^{\frac{1}{2}} \right] \right)$$

Modulated/expanded pseudo ensemble perturbations

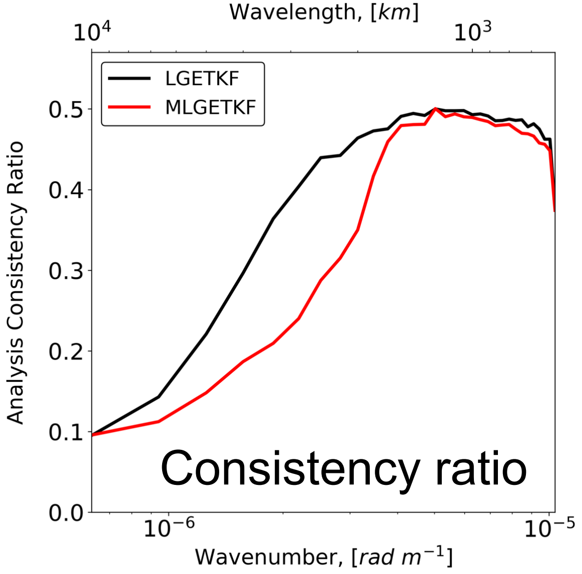
Scale decomposed raw perturbations

Multiscale model space localization



- MLGETKF reduces analysis and background errors relative to scale unaware LGETKF for all scales, especially towards the large scales.

- The same study also reveals that the common issue of background ensemble under-sampling can be scale dependent



Methodology

Scale dependent RTPS inflation (RTPS-SDI)

Naicheng Xu Thur. poster

- The RTPS inflation (Whitaker and Hamill 2012) relaxes the posterior spread toward the prior spread

- Inflation factor $\boldsymbol{g} = \alpha * \frac{\sqrt{\boldsymbol{P^b}} - \sqrt{\boldsymbol{P^a}}}{\sqrt{\boldsymbol{P^a}}} + 1$

- In this study, RTPS is further developed and implemented separately for different scales (hereafter "RTPS-SDI")

- Inflation factor $\mathbf{g}_m = \alpha_m * \frac{\sqrt{\boldsymbol{P_m^b}} - \sqrt{\boldsymbol{P_m^a}}}{\sqrt{\boldsymbol{P_m^a}}} + 1$

Methodology

Scale dependent inflation based on sampling error (SE-SDI)

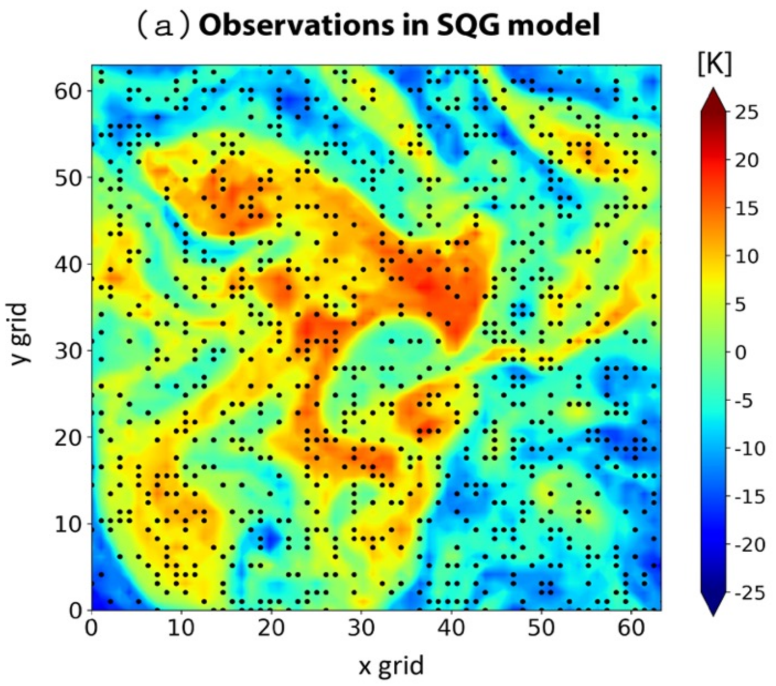
- Hodyss et al. (2016), using the ensemble Kalman filter theory, derived a posterior inflation that accounts for the sampling error deficiency (hereafter “SE” inflation)

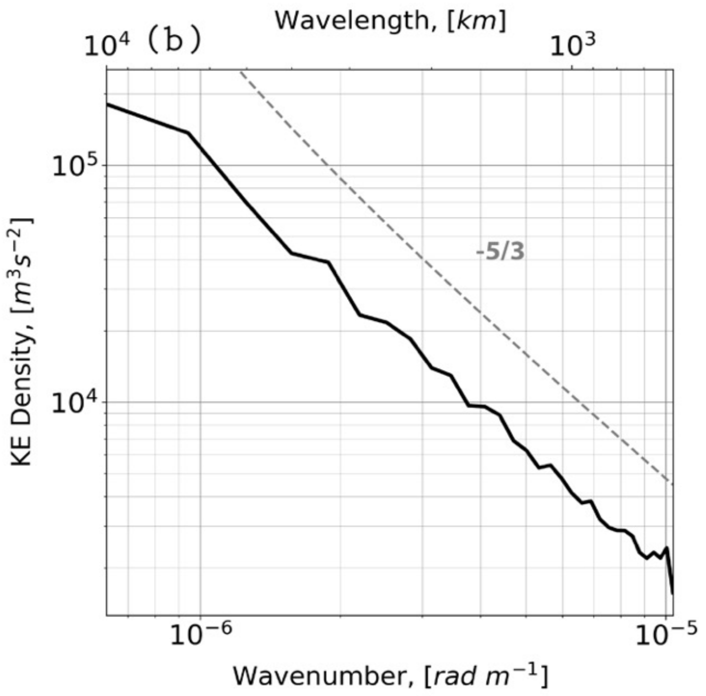
- $\boldsymbol{S} \approx \mathrm{a} * \boldsymbol{P}^a + \boxed{\left(\frac{\boldsymbol{P}^a}{\boldsymbol{P}^b}\right)^2 \left(b\,\frac{\boldsymbol{P}^b}{N_e} + c\,\frac{2}{N_e-1}\left(\overline{\boldsymbol{x}^a} - \overline{\boldsymbol{x}^b}\right)^2\right)}$ Inflation factor: $\boldsymbol{g} = \sqrt{\frac{\boldsymbol{S}}{\boldsymbol{P}^a}}$

Additional analysis error due to limited ensemble members, i.e., sampling error

- Scale dependent SE inflation (hereafter “SE –SDI” inflation):
- Starting with the multiscale analysis equation (equivalent to MLGETKF if assuming no cross scale covariance), rederive SE inflation for different scales of analysis
- $\boldsymbol{S}_m \approx a_m \boldsymbol{P}_m^a + \left[\left(\frac{\boldsymbol{P}_m^a}{\boldsymbol{P}_m^b}\right)^2 \left(b_m\,\frac{\boldsymbol{P}_m^b}{N_e} + c_m\,\frac{2}{N_e-1}\left(\overline{\boldsymbol{x}_m^a} - \overline{\boldsymbol{x}_m^b}\right)^2\right)\right]$ Inflation factors: $\boldsymbol{g}_m = \sqrt{\frac{\boldsymbol{S}_m}{\boldsymbol{P}_m^a}}$

Experiment Design



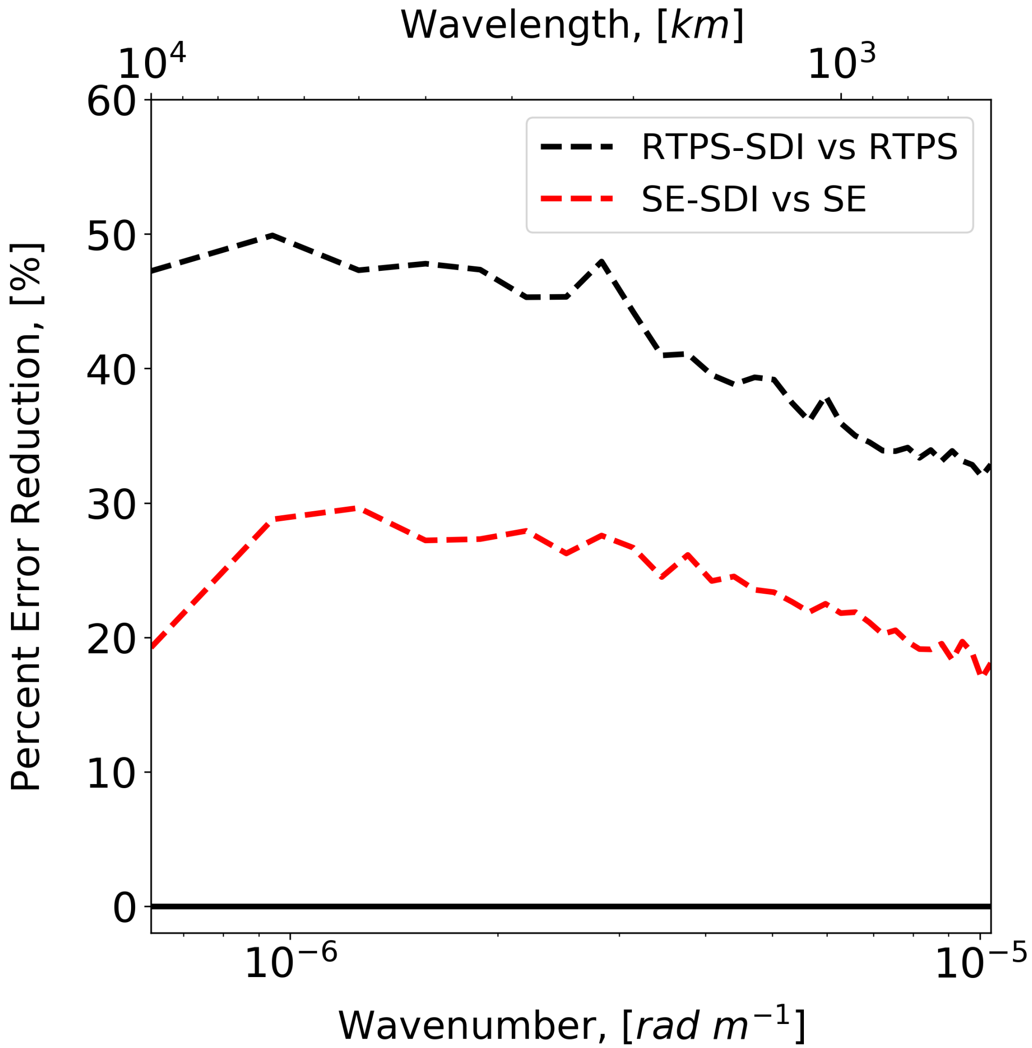


Exp. Name	Description
RTPS	Use RTPS for MLGETKF posterior inflation (Wang X. et al. 2021)
RTPS-SDI	Use scale dependent RTPS for MLGETKF posterior inflation
SE	Use sampling error (SE) derived inflation for MLGETKF posterior inflation
SE-SDI	Use scale dependent SE inflation for MLGETKF posterior inflation

- SQG turbulence model follows Tulloch and Smith (2009) and mimics mesoscale of the atmosphere
- The four inflation methods are implemented on MLGETKF (Wang X. et al. 2021)
- Simulated obs.: potential temperature on both model surfaces with a standard deviation of 1K
- Ensemble size: 20-member
- 3-hourly data assimilation is performed for 400 cycles

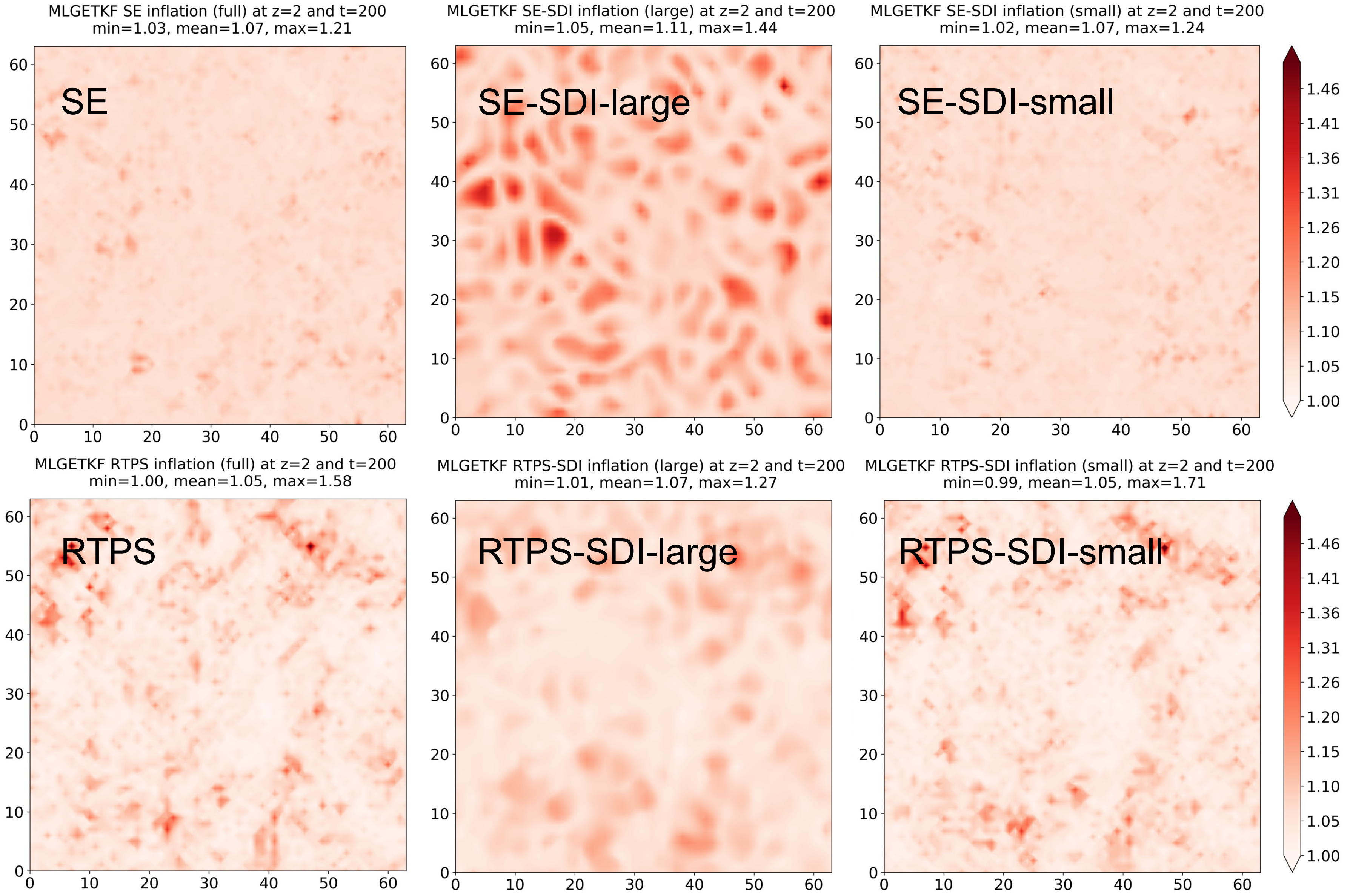
Development of scale dependent inflation (SDI) methods

Xu* et al. 2023



- Both SDI methods show improved analysis accuracy across all resolved scales relative to their own scale independent inflation counter parts.

Temporal and spatial behaviors of the inflation



- In the two SDI experiments, the inflation factors at large and small scales reflect the corresponding scale structure.

Summary and Remarks

- Great challenges exist to achieve effective multiscale DA for next generation NWP
  - Individual earth system component (e.g. atmosphere)
  - Coupled earth system components

- R&D on simultaneous MDA performed **using operational model and DA system**, including GFS, RRFS/HRRR, HAFS, WoF, demonstrate great potential of such approach to better utilize observations and to improve NWP

- **Examples of research on MDA methodology development are introduced**
  - A new MDA solver (MLGETKF)
  - New methods to treat ensemble deficiency in MDA

- **Fundamental research is needed** to address challenges associated with the **multiscale DA** for **all NWP applications: short range, medium range and S2S predictions**

Main References

Wang, X., H. Chipilski*, C. H. Bishop, L. Satterfield, N. Baker and J. Whitaker, 2021: A Multiscale Local Gain Form Ensemble Transform Kalman Filter (MLGETKF), Mon. Wea. Rev., https://doi.org/10.1175/MWR-D-20-0290.1

Jones, E.* and X. Wang, 2023a: A multi-resolution ensemble 4DEnVar with variable ensemble sizes to improve global and tropical cyclone track numerical prediction. Mon. Wea. Rev., 151, 1145-1166.

Jones, E.* and X. Wang, 2023b: Flow-dependent vertical localization in hybrid 4DEnVar to improve global and tropical cyclone track numerical prediction. Mon. Wea. Rev., to be submitted.

Kay, J. K.* and X. Wang, 2020: A multi-resolution ensemble hybrid 4DEnVar for global numerical prediction. *Mon. Wea. Rev*., 148, 825-847.

Huang, B.*, X. Wang, D. Kleist, and T. Lei, 2021: Simultaneous Multi-scale Data Assimilation using Scale Dependent Localization in GSI-based Hybrid 4DEnVar for NCEP FV3-based GFS. *Mon. Wea. Rev.*, https://doi.org/10.1175/MWR-D-20-0166.1.

Lu, X.* and Wang, X. 2023: Simultaneous Multiscale EnVar with scale-dependent localization (SDL) in HAFS to improve hurricane predictions. *Mon. Wea. Rev.,* to be submitted.

Wang, Y.*, and X. Wang, 2017: Direct assimilation of radar reflectivity without tangent linear and adjoint of the nonlinear observation operator in the GSI-Based EnVar System: Methodology and experiment with the 8 May 2003 Oklahoma City tornadic supercell. *Mon. Wea. Rev.*, 145, 1447–1471, https://doi.org/10.1175/MWR-D-16-0231.1.

Wang, Y.* and X. Wang, 2021: Development of Convective-Scale Static Background Error Covariance within GSI-Based Hybrid EnVar System for Direct Radar Reflectivity Data Assimilation. Mon. Wea. Rev., published.

Wang, Y.*, and X. Wang, 2023a: A Simultaneous Multiscale Data Assimilation using Scale- and Variable-Dependent Localization in EnVar to Improve Storm-Scale Ensemble-Based Data Assimilation and Forecasts: Methodology and Experiments for a Tornadic Supercell. *Journal of Advances in Modeling Earth Systems*, https://doi.org/10.1029/2022MS003430.

Wang, Y.* and X. Wang, 2023: Improving CONUS Convective-Scale Forecasting with Simultaneous Multiscale Data Assimilation: A Squall-Line Case Study, Journal of Geophysical Research – Atmospheres, submitted

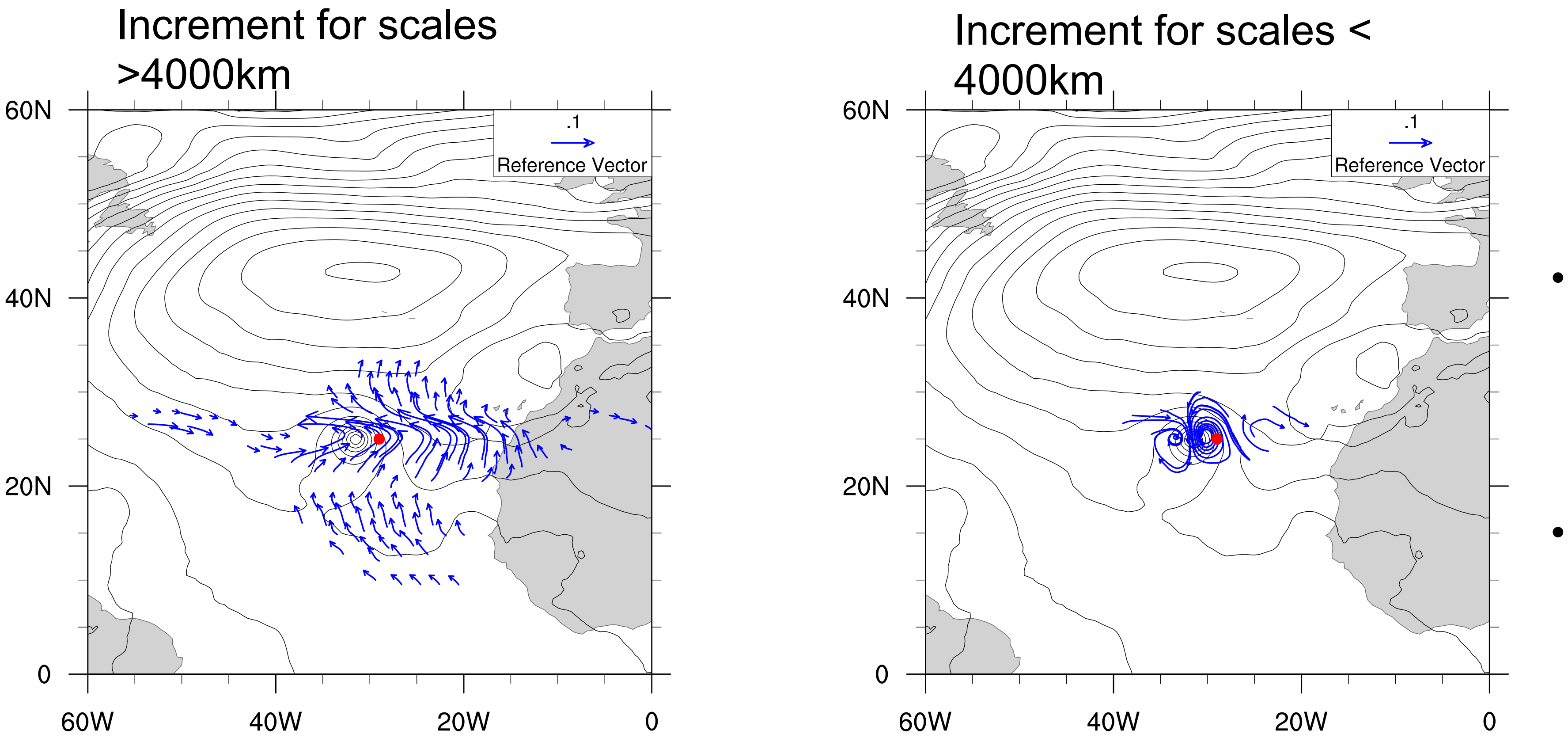
Xu*, N., X. Wang and Y. Wang*, 2023: Scale dependent inflation in multiscale ensemble based data assimilation. Mon. Wea. Rev., to be submitted

Yang, Y.* and X. Wang, 2023: A Comparison of 3DEnVar and 4DEnVar for Convective-Scale Direct Radar Reflectivity Data Assimilation in the Context of Filter and Smoother, Mon. Wea. Rev., accepted.

* OU MAP students and early career scientists

Simultaneous Multiscale DA

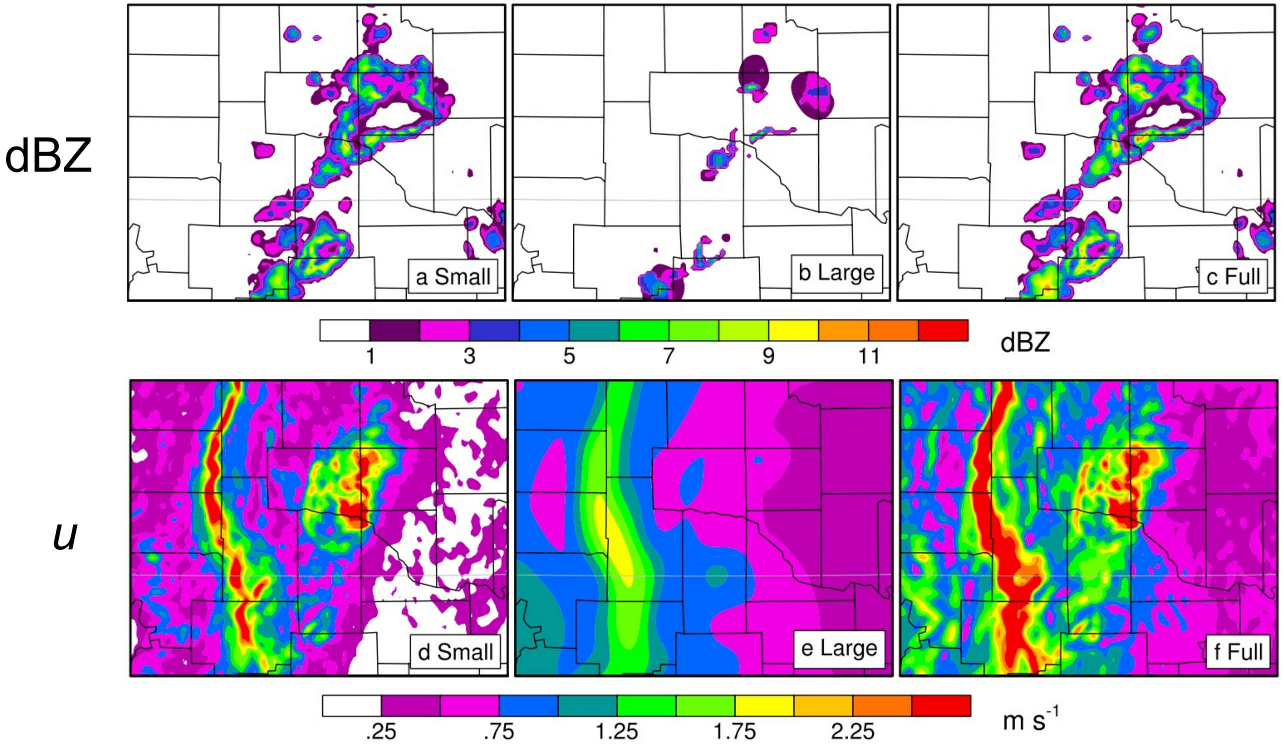
- A simultaneous multiscale DA approach, as opposed to a sequential approach, allows all observations to correct all resolved scales at once (Wang X., 2021).



- A single obs. can correct multiple scales in simultaneous MDA
- Simultaneous multiscale DA also defines cross scale band error correlation

Huang*, Wang et al. 2021

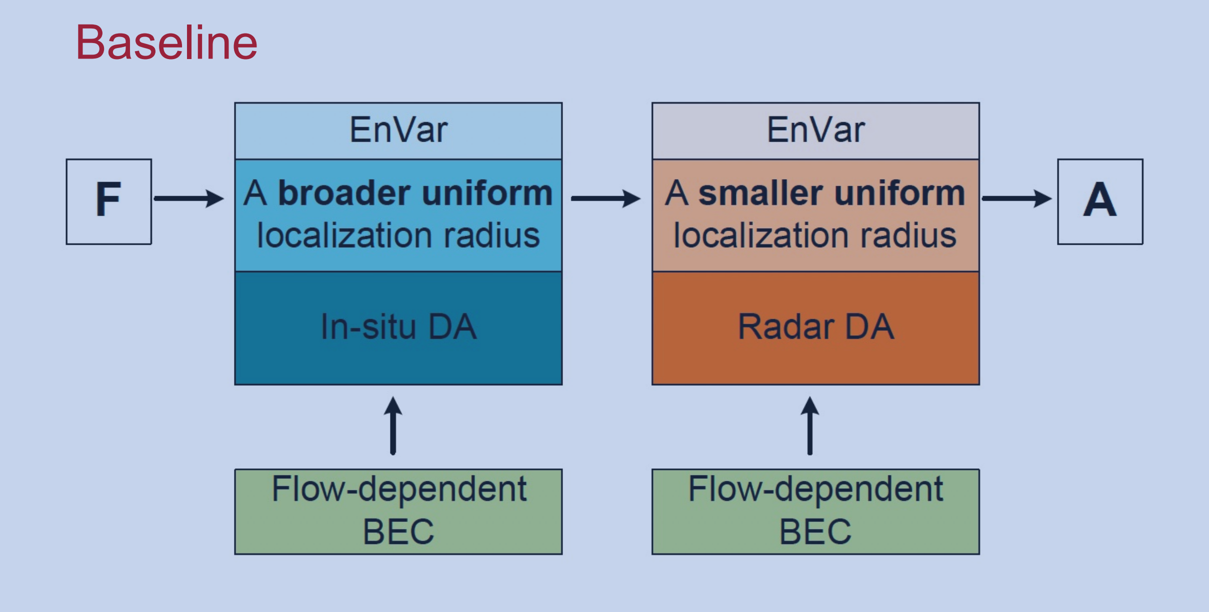
Implementation of SDLVDL within EnVar scale decomposition

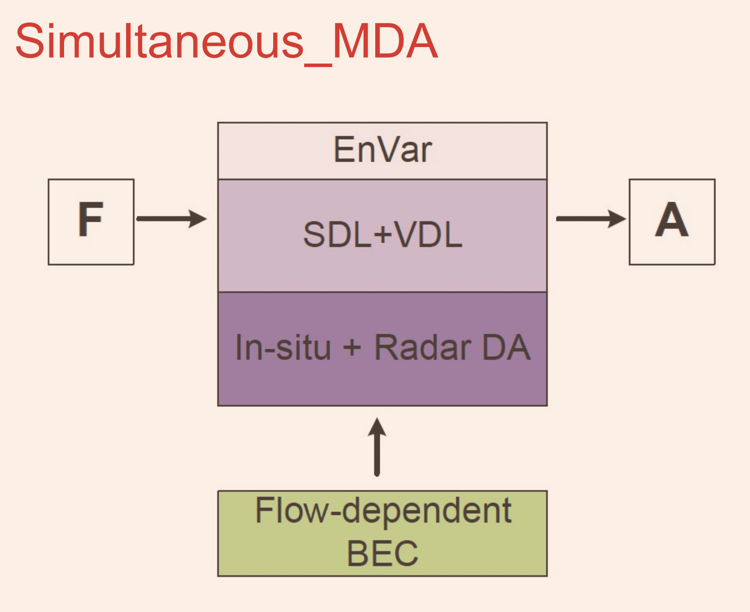


- A diffusion operator is applied to decompose scales for each ensemble perturbation.
- Uncertainty of the storm and small-scale low-level convergence are mostly concentrated at the decomposed small scale.
- Uncertainty of the environment-related fields are reflected by the large scale.

Experiment design

Model and DA configuration





Exps	Assimilation Strategy	Localization scales	
Baseline (Ope.-like)	Separate assimilation of in-situ and radar obs.	In-situ obs.	300 km
		Radar obs.	15 km
Simultaneous_MDA	Simultaneous assimilation of in-situ and radar obs.	Small scale	15 km: hydrometeors, and *w*
			60 km: *u*, *v*, *t*, *q*, and ps
		Large scale	60 km: hydrometeors, and *w*
			500 km: *u*, *v*, *t*, *q*, and ps

Table 2
List of Experiments and Their Configuration of Horizontal Localization Radius

Experiments	Horizontal localization radius
Exp-SSL	10 km for full-scale ensemble perturbations
Exp-SDL	10 and 60 km for all variables of small- and large-scale ensemble perturbations, respectively.
Exp-SDLVDL	10 km for all variables of small-scale ensemble perturbations; 15 km for *q*, *w*, ql, qr, qs, qi, qg, and dbz of large-scale ensemble perturbations; 60 km for *u*, *v*, *t*, ps of large-scale ensemble perturbations

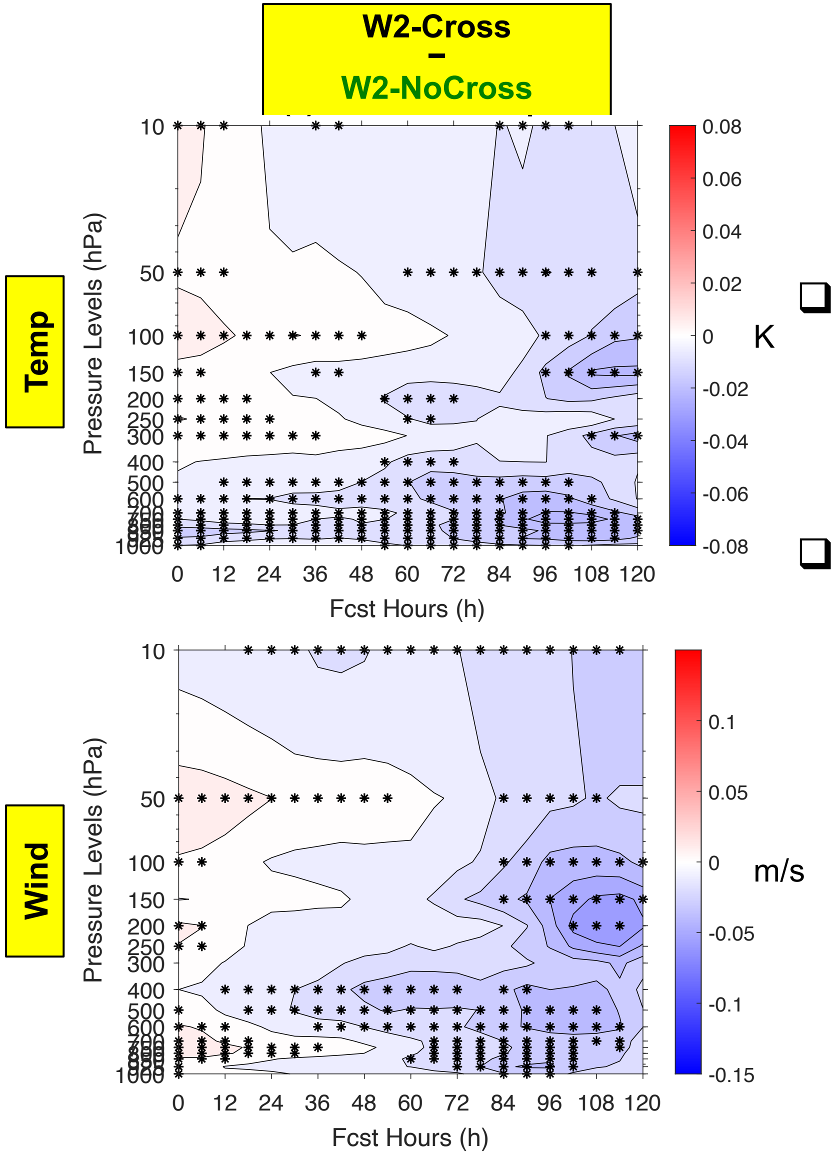
Key aspects of the algorithm

$$\boxed{\mathbf{Z}^{ML}} = \left(\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_{K_{ML}}\right) = (\mathbf{I} \quad \mathbf{I} \quad \cdots \quad \mathbf{I})\left(\boxed{\mathbf{X}^{MS}} \Delta \left[\left(\boxed{\mathbf{L}^{MS}}\right)^{\frac{1}{2}}\right]\right)$$

Modulated/expanded pseudo ensemble perturbations	Scale decomposed raw perturbations	Multiscale model space localization

- Rapid creation of many pseudo ensemble perturbations in a local volume via a multiscale ensemble modulation procedure.
- The modulated ensemble intrinsically includes multi-scale model space localization and is used to update ensemble mean and perturbations.
- Multi-scale model space localization adopts scale-aware localization. In addition, localization of the ensemble covariances between different scales are defined and can be further modulated.
- MLGETKF only updates and propagates the original number of ensemble members.

Impact of cross band correlation



RMSE difference between W2-Cross and W2-NoCross (blue/red→ better/worse forecasts in W2-Cross)

- W2-NoCross shows slightly better forecasts than W2-Cross within one day. This may benefit from the spatial averaging of ensemble covariances in W2-NoCross.

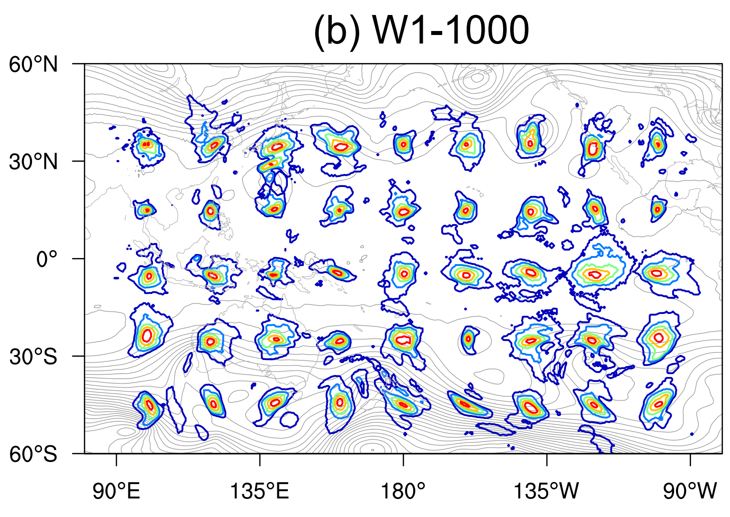
- Beyond one-day, W2-Cross in general shows more accurate forecasts than W2-NoCross, likely contributed by its higher degrees of retained heterogeneity of ensemble covariances and resultant analysis, and its more balanced analysis through partially including cross-waveband covariances.

Huang*, Wang, et al. 2021

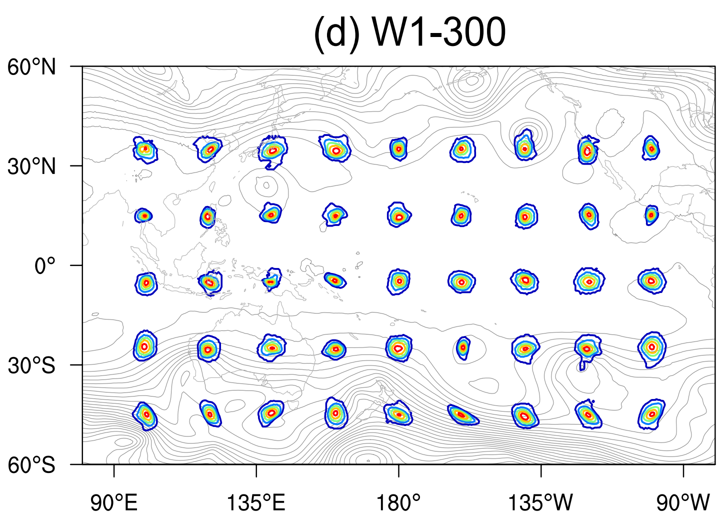
Multiscale vs Single scale ensemble based DA

Physical space visualization

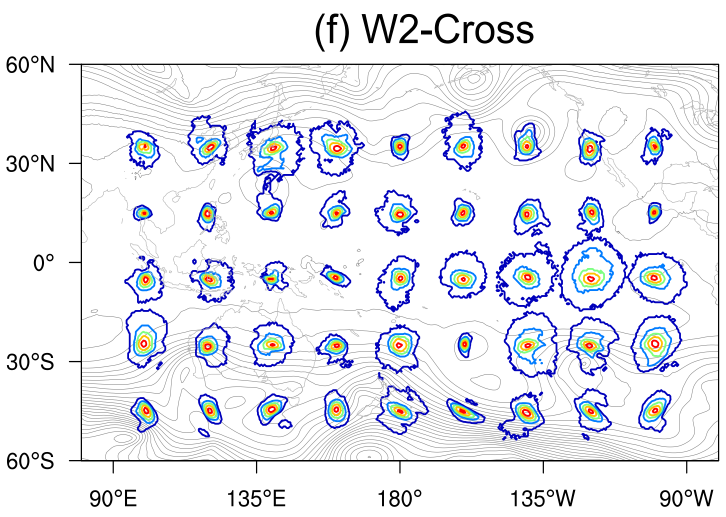
Single large



Single small

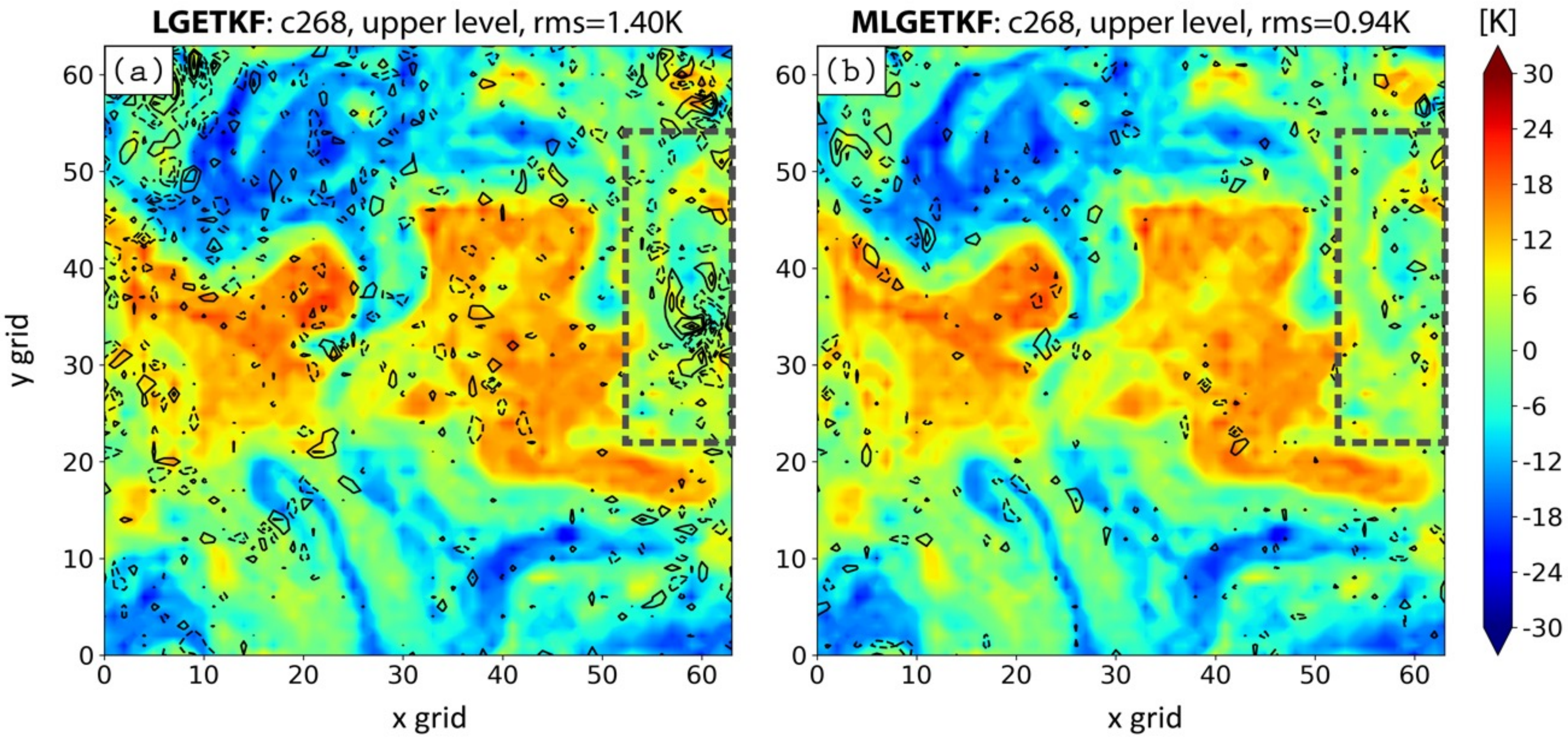


Multiscale (MDA)



Huang*, Wang et al. 2021

Analysis error in physical space



- ❑ MLGETKF not only shows skill in decreasing the small scale component of the analysis errors, but also is effective in suppressing the development of large scale, dynamical, high-amplitude analysis errors

Wang X. et al. 2021