



Hybrid Variational/Ensemble Data Assimilation

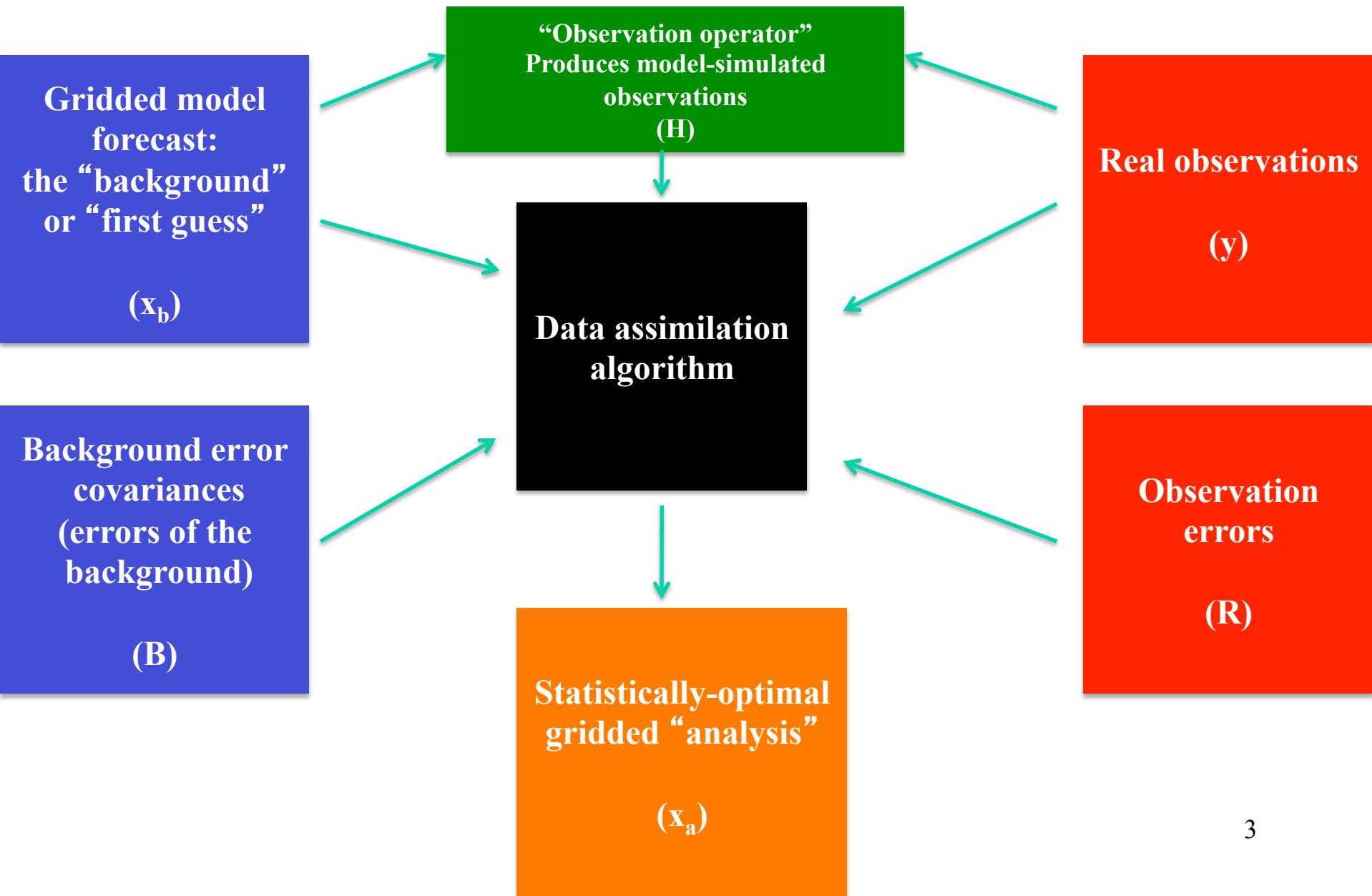
Craig Schwartz and Zhiqian Liu
(schwartz@ucar.edu)

NCAR/MMM

Outline

- Background
- Some results
- Introduction to hybrid practice

What is data assimilation?

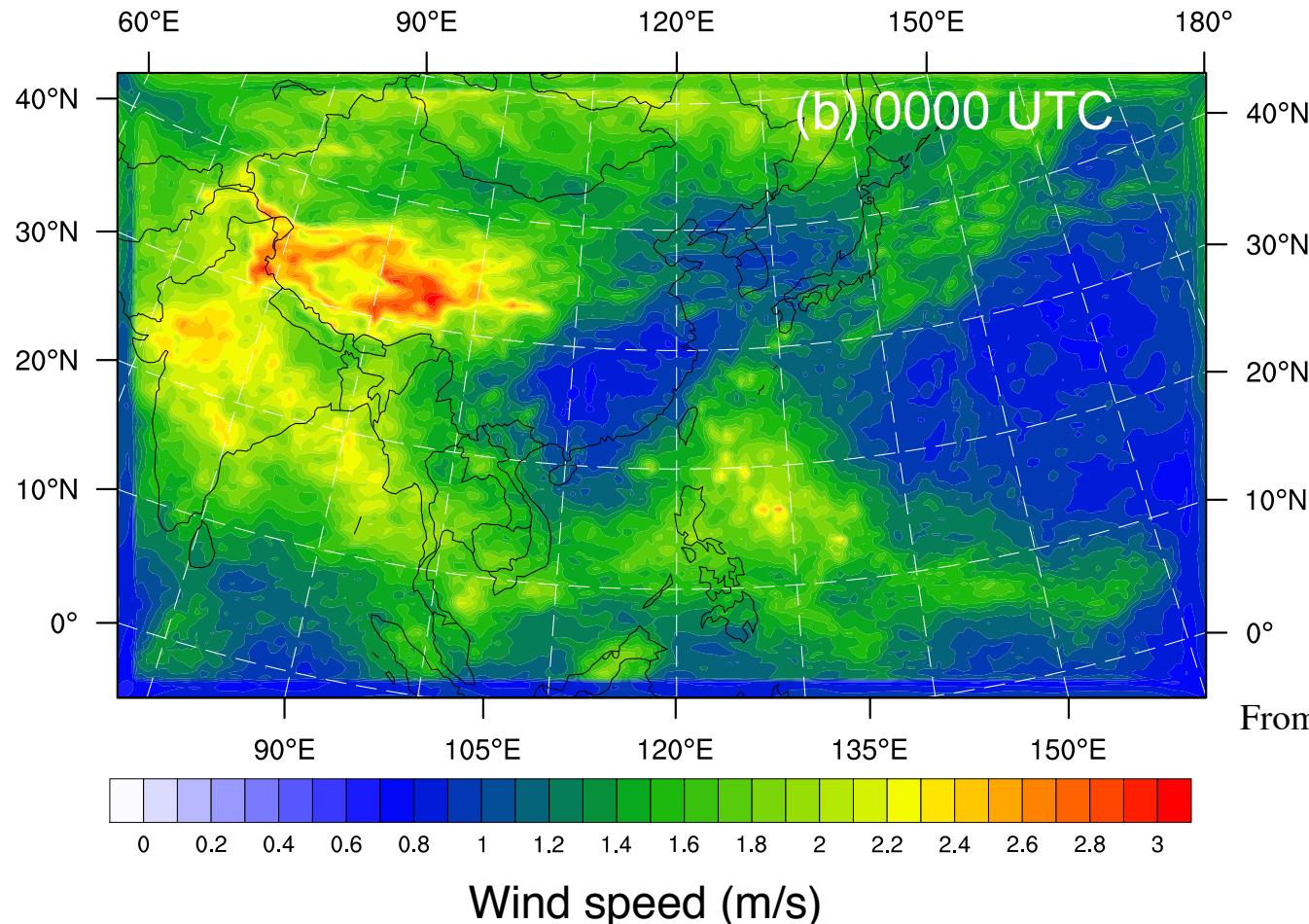


Some data assimilation methods

- Three-dimensional variational (3DVAR)
 - Background error covariances (BECs) typically fixed/time-invariant
 - May yield poor results when actual flow differs from that encapsulated within the fixed “climatology”
- Ensemble Kalman filter (EnKF)
 - Time-evolving, “**flow-dependent**” BECs estimated from a short-term ensemble forecast
 - Many different flavors (e.g., ETKF, EAKF)

Ensemble BECs (i.e., spread)

- Average ensemble spread of wind speed over ~3 weeks at 0000 UTC



Ensemble BECs (i.e., spread)

- General definition of covariance:

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- In vector matrix form (here, n is ensemble size):

$$= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

$$= \frac{1}{n-1} \sum_{i=1}^n (\delta \mathbf{x}_i)(\delta \mathbf{x}_i)^T$$

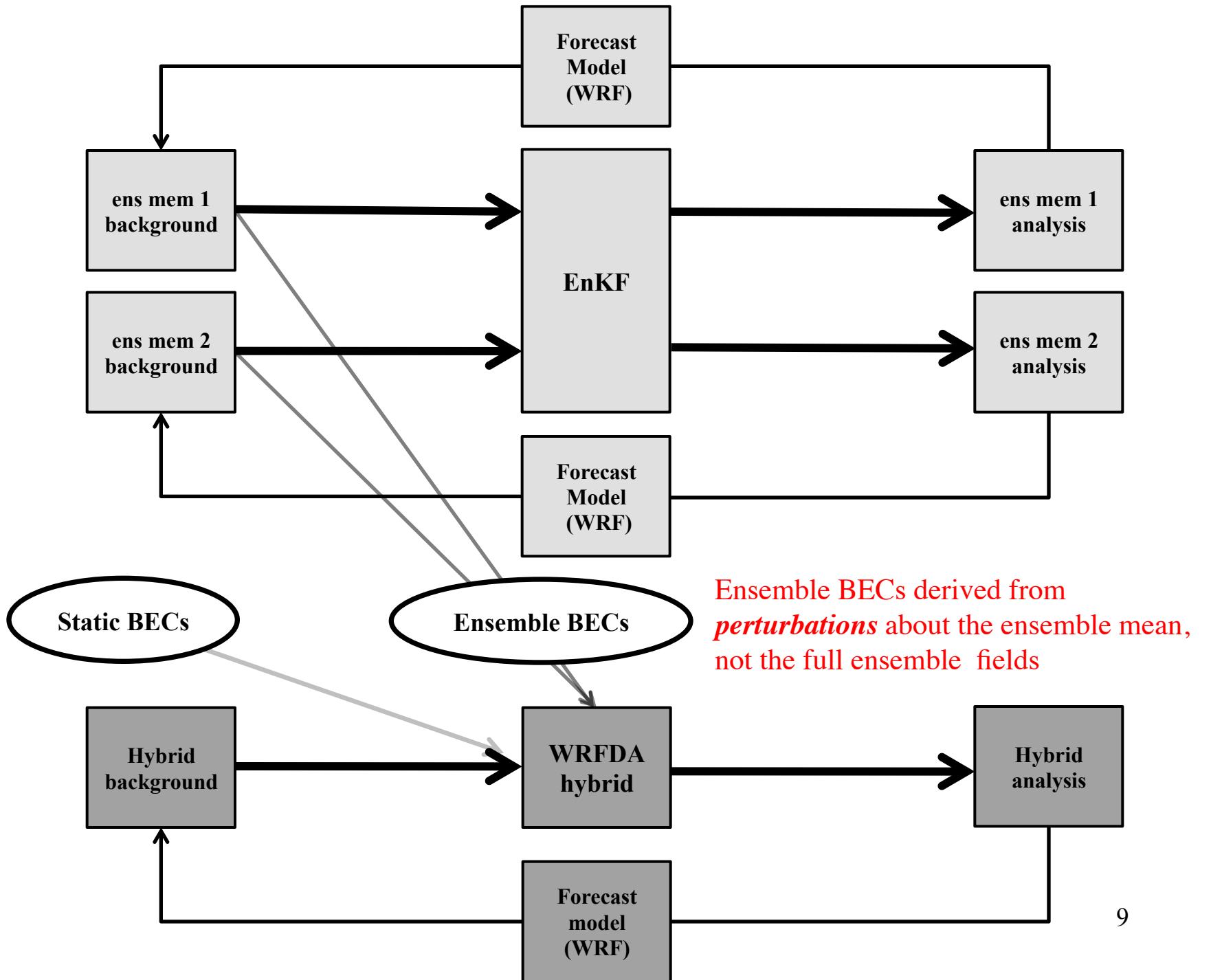
“Hybrid” variational/ensemble DA

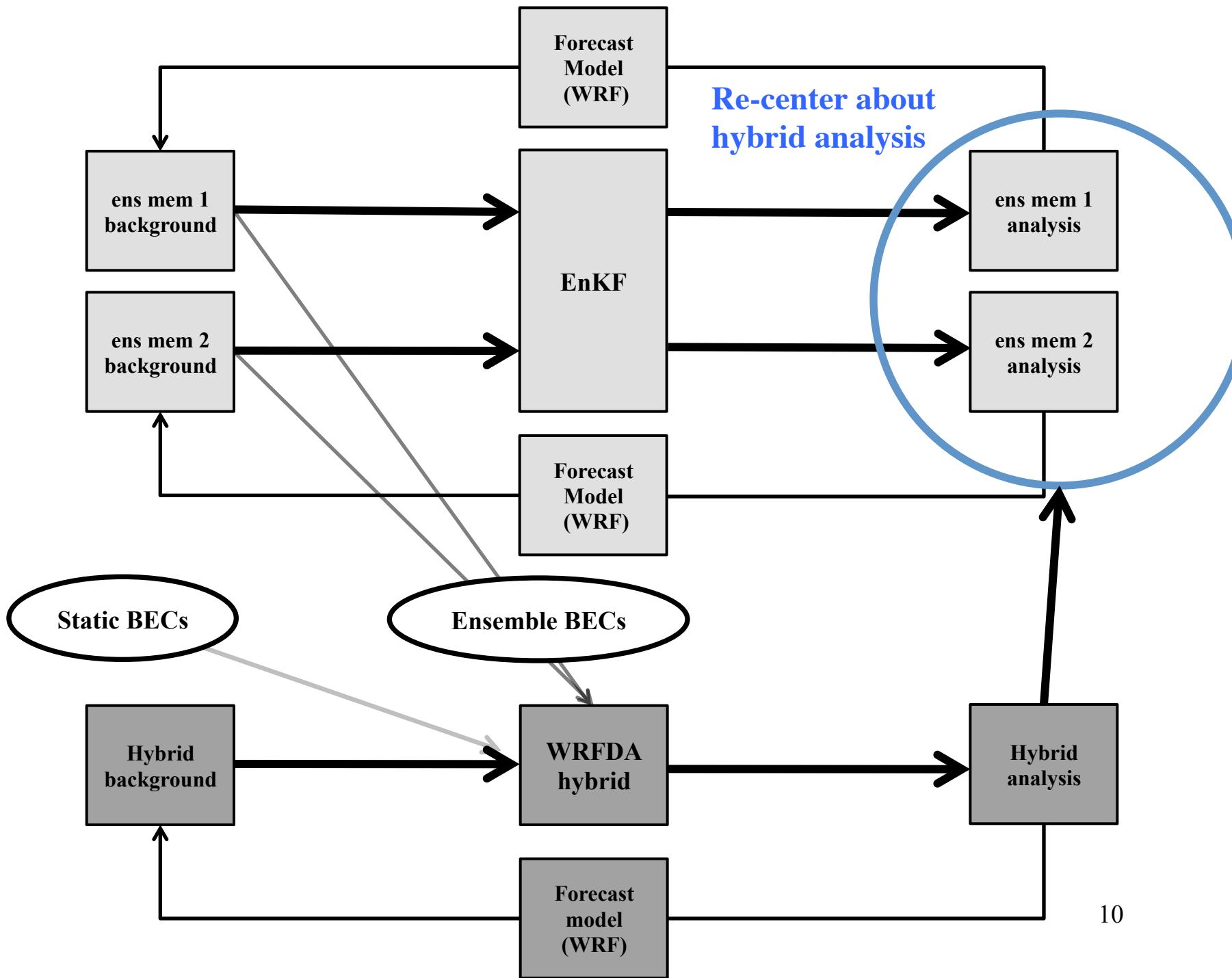
- “Hybrid” variational/ensemble
 - Incorporates ensemble background errors within a variational (e.g., 3DVAR) framework
 - Combination of fixed and time-evolving background errors
 - Main additional expense compared to 3DVAR is running an ensemble of forecasts



What is Hybrid DA?

- Deterministic background is analyzed by a variational algorithm (i.e., minimize a cost function)
 - It combines the 3DVAR “climatological” BECs and “errors of the day” from ensemble perturbations
- Traditionally generates a deterministic analysis (like 3DVAR)
- Need a separate system to update ensemble
 - Could be ensemble forecasts already available from operational centers
 - Could be an EnKF-based DA system
 - Could be a multiple model/physics ensemble
- Ensemble needs to be good to well-represent “errors of the day”





Hybrid formulation

(Hamill and Snyder, 2000)

- 3DVAR cost function

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2} [\mathbf{H}(\mathbf{x}) - \mathbf{y}]^T \mathbf{R}^{-1} [\mathbf{H}(\mathbf{x}) - \mathbf{y}]$$

- Idea: replace \mathbf{B} by a weighted sum of static \mathbf{B}_s and the ensemble \mathbf{B}_e

$$\mathbf{B} = a_s \mathbf{B}_s + a_e \mathbf{B}_e \circ \mathbf{C}, \quad a_s = 1 - a_e$$

- Term \mathbf{C} is localization for the ensemble
- Terms a_s and a_e can be tuned to determine how much \mathbf{B}_s and \mathbf{B}_e are weighted
- This form is difficult to implement for a large NWP model
 - Most systems use “extended control variables”

Hybrid formulation used in WRFDA

(Lorenc, 2003)

- Ensemble covariance is included in the 3DVAR cost function through augmentation of control variables

$$J(\mathbf{x}, \boldsymbol{\alpha}) = \beta_s \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \underbrace{\beta_e \frac{1}{2} \sum_{i=1}^N \overbrace{\boldsymbol{\alpha}_i^T \mathbf{C}^{-1} \boldsymbol{\alpha}_i}^{\text{ensemble control variable } \boldsymbol{\alpha}_i \text{ } (M \times 1)}} + \frac{1}{2} [\mathbf{y} - H(\mathbf{x} + \mathbf{x}'_e)]^T \mathbf{R}^{-1} [\mathbf{y} - H(\mathbf{x} + \mathbf{x}'_e)]$$

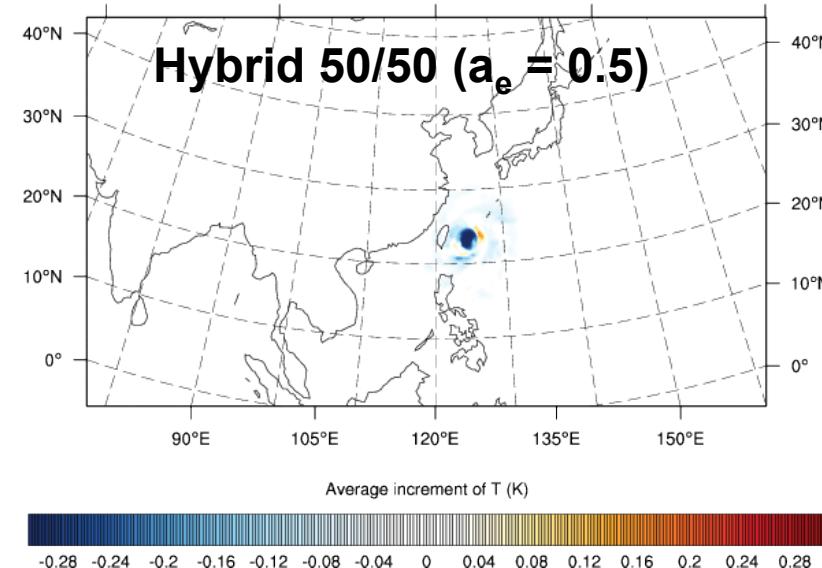
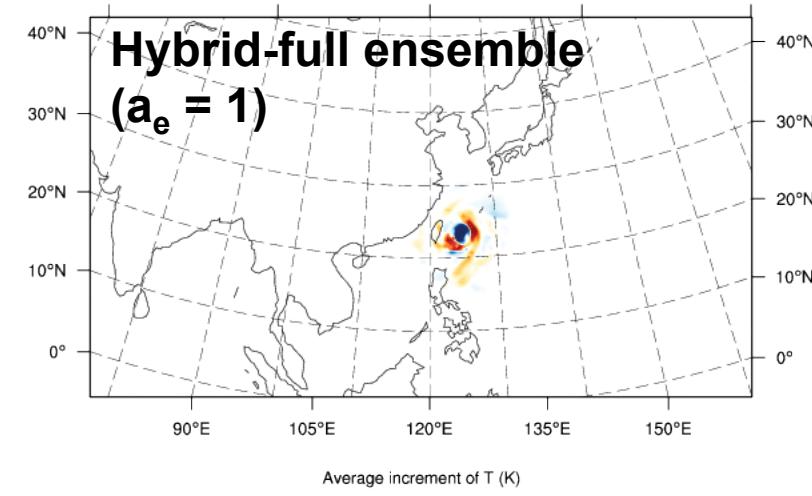
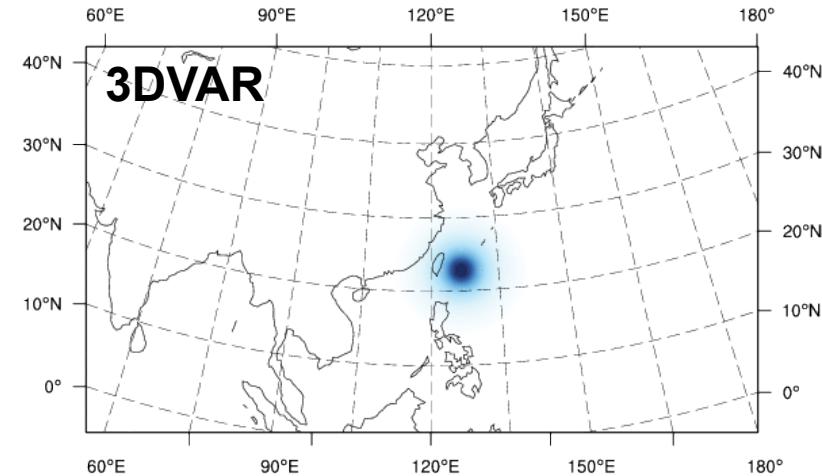
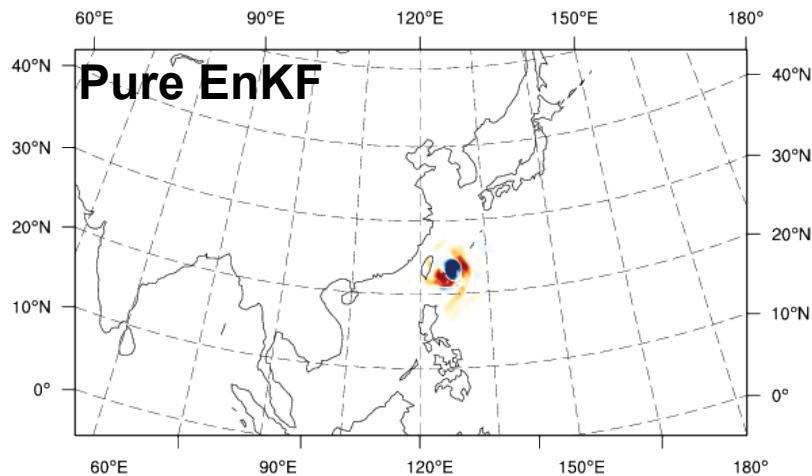
$\mathbf{x}'_e = \sum_{i=1}^N \boldsymbol{\alpha}_i \circ \mathbf{x}'_i$, where \mathbf{x}'_i is the ensemble perturbation for the ensemble member i.

- denotes element-wise product. $\boldsymbol{\alpha}_i$ is in effect the ensemble weight.
- C: correlation matrix (effectively localization of ensemble perturbations)

- More simply: $J(\mathbf{x}, \boldsymbol{\alpha}) = J_b + J_e + J_o$
- β_s and β_e ($1/\beta_s + 1/\beta_e = 1$) can be tuned to have different weight between static and ensemble part

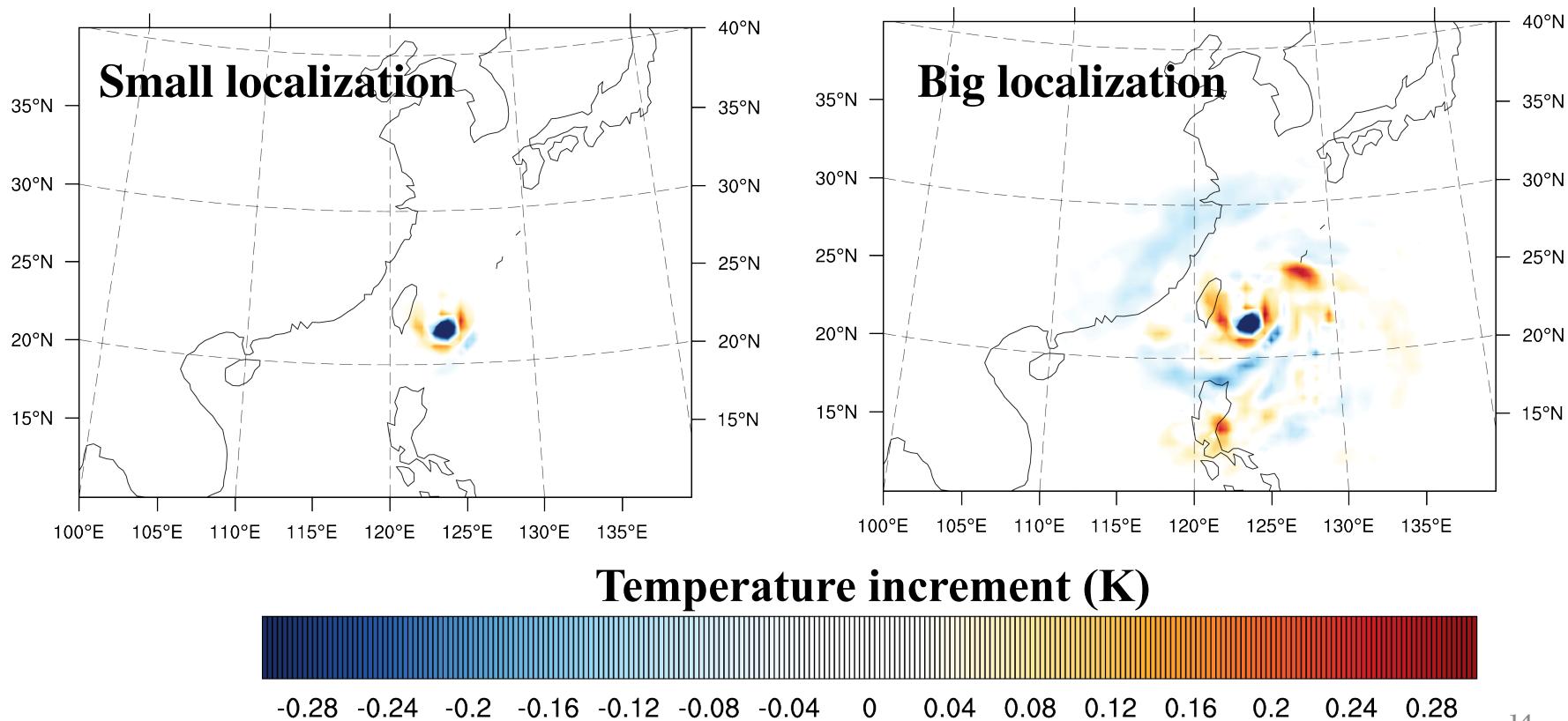
Single observation tests

- Potential temperature increment, 21st model level



Meaning of localization

- Localization defines the extent to which an observation can produce an analysis increment
- In this example, 100% of the BECs are from ensemble

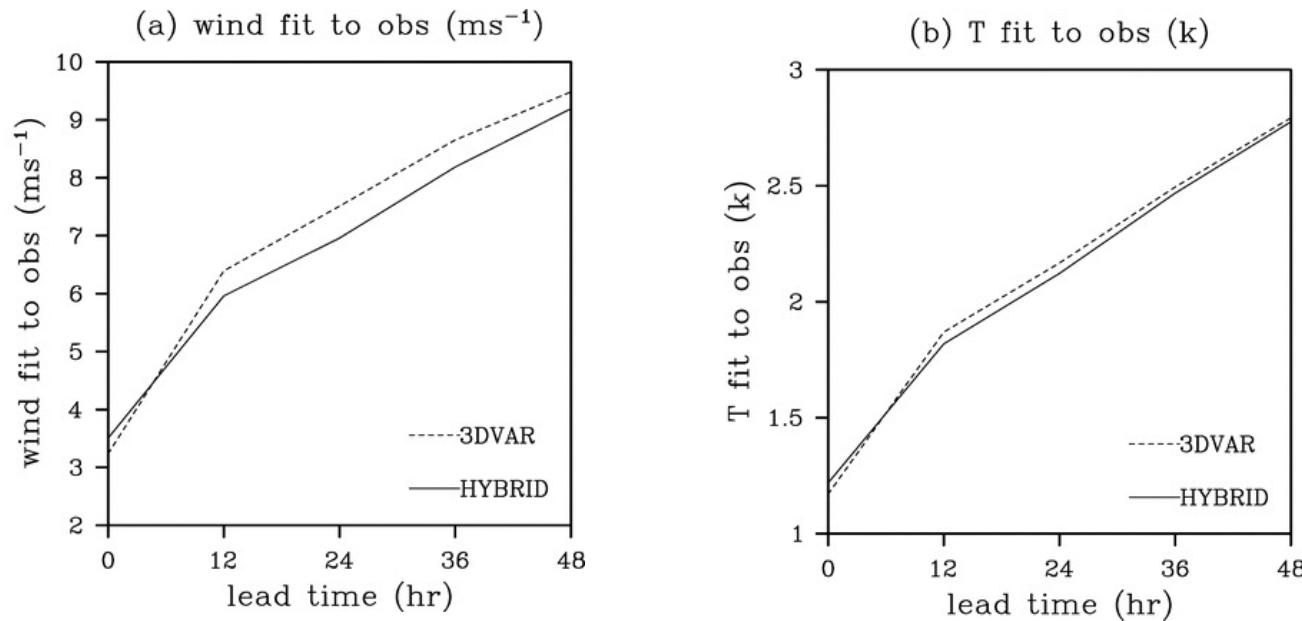


Advantages of Hybrid DA

- Hybrid localization is in model space while EnKF localization is usually in observation space
- For some observation types (e.g., radiances), localization is not well defined in observation space
- Easier to make use of existing radiance VarBC in hybrid
- For small ensembles, use of static \mathbf{B} could be beneficial to have a higher-rank covariance.

Sample results

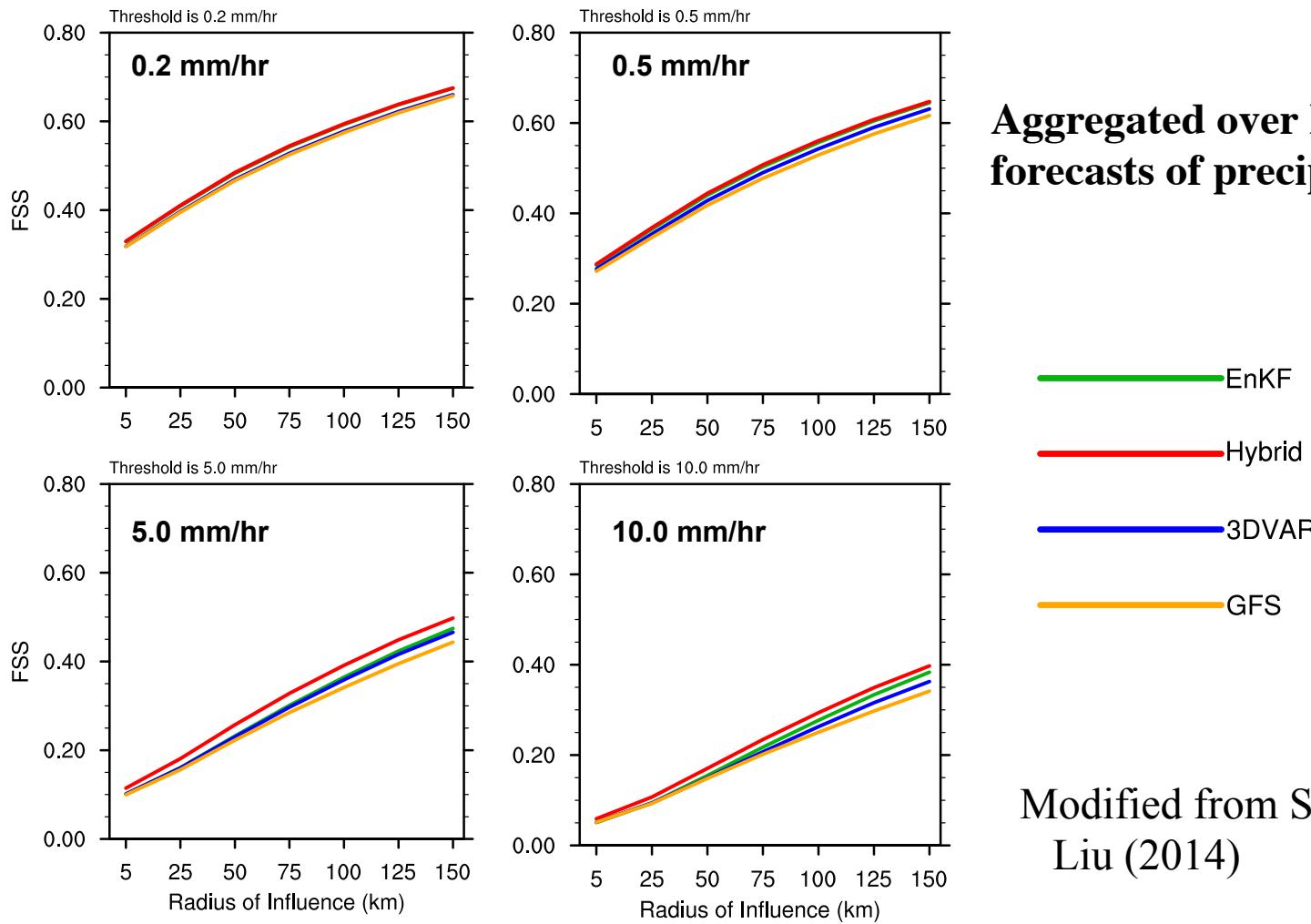
- Example over North America at coarse grid spacing
- Similar results have been obtained by many studies



From Wang et al. (2008)

Hybrid vs. 3DVAR and EnKF

- Fractions skill scores for rainfall (higher is better)



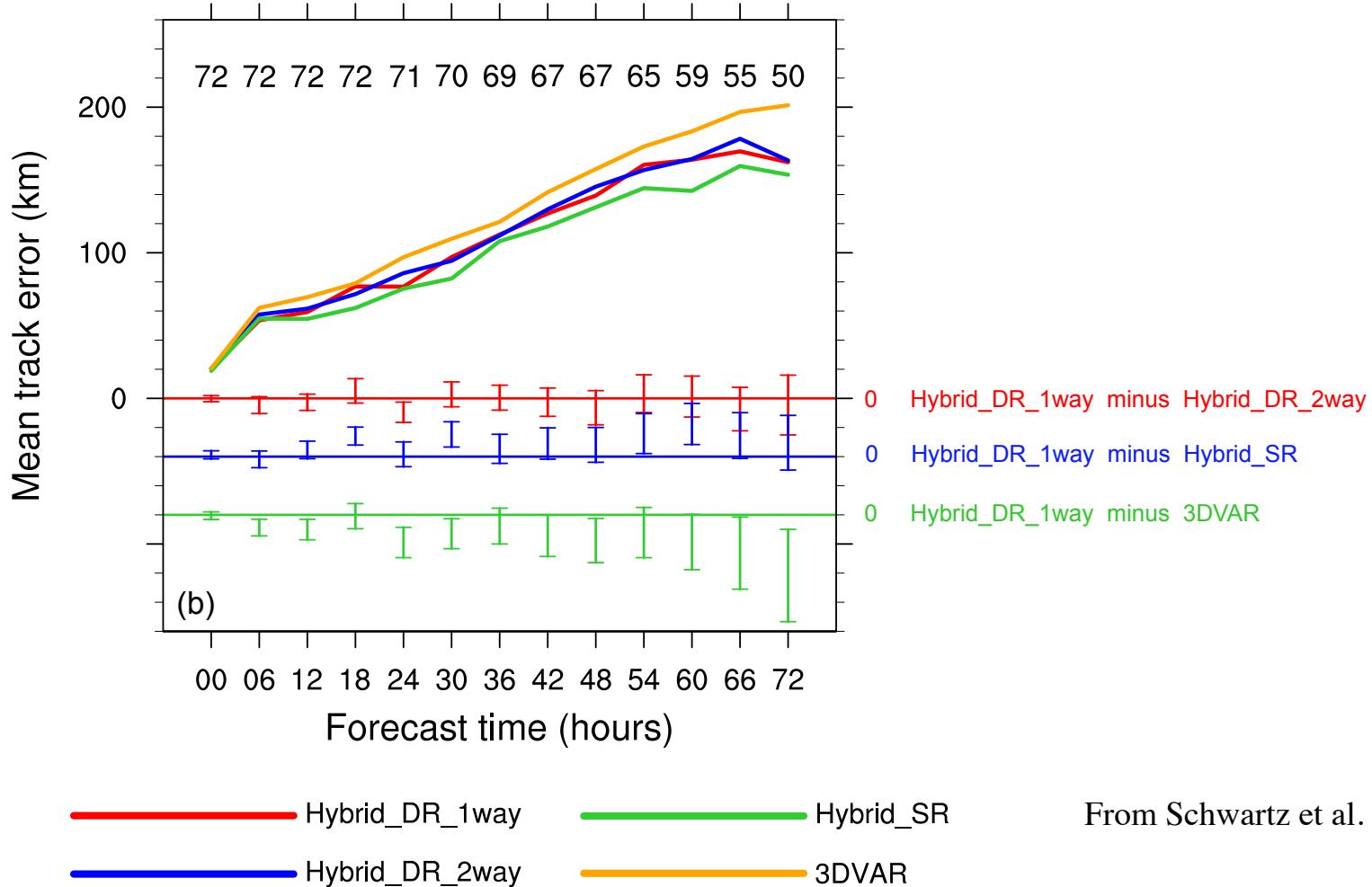
Aggregated over hourly 18-36-hr forecasts of precipitation

EnKF
Hybrid
3DVAR
GFS

Modified from Schwartz and Liu (2014)

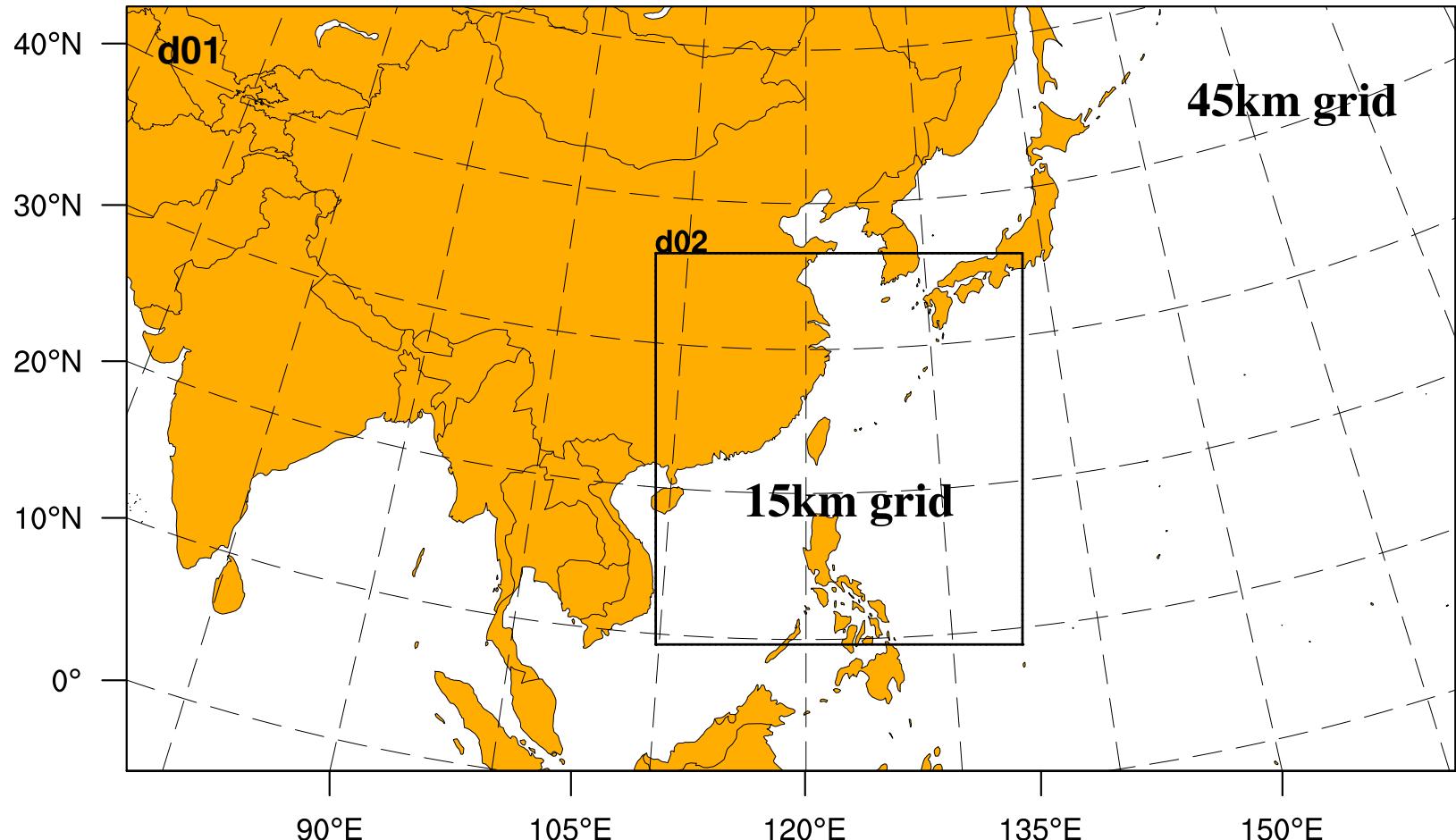
Typhoon example

- Mean tropical cyclone track errors



Dual-Resolution hybrid (V3.6)

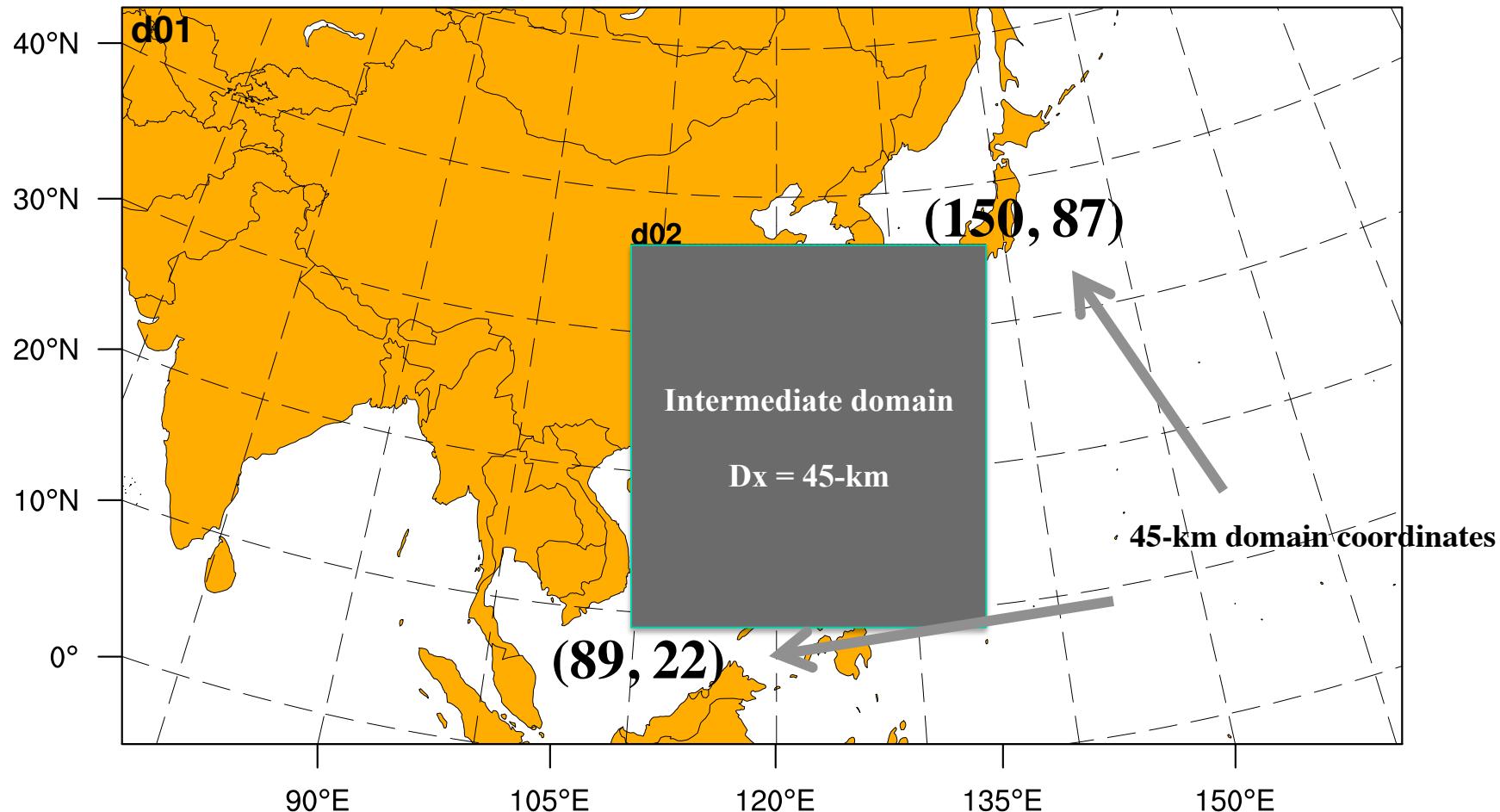
Schwartz et al. (2015; MWR)



Hybrid analysis on 15-km grid but with ensemble perturbation input from 45-km grid

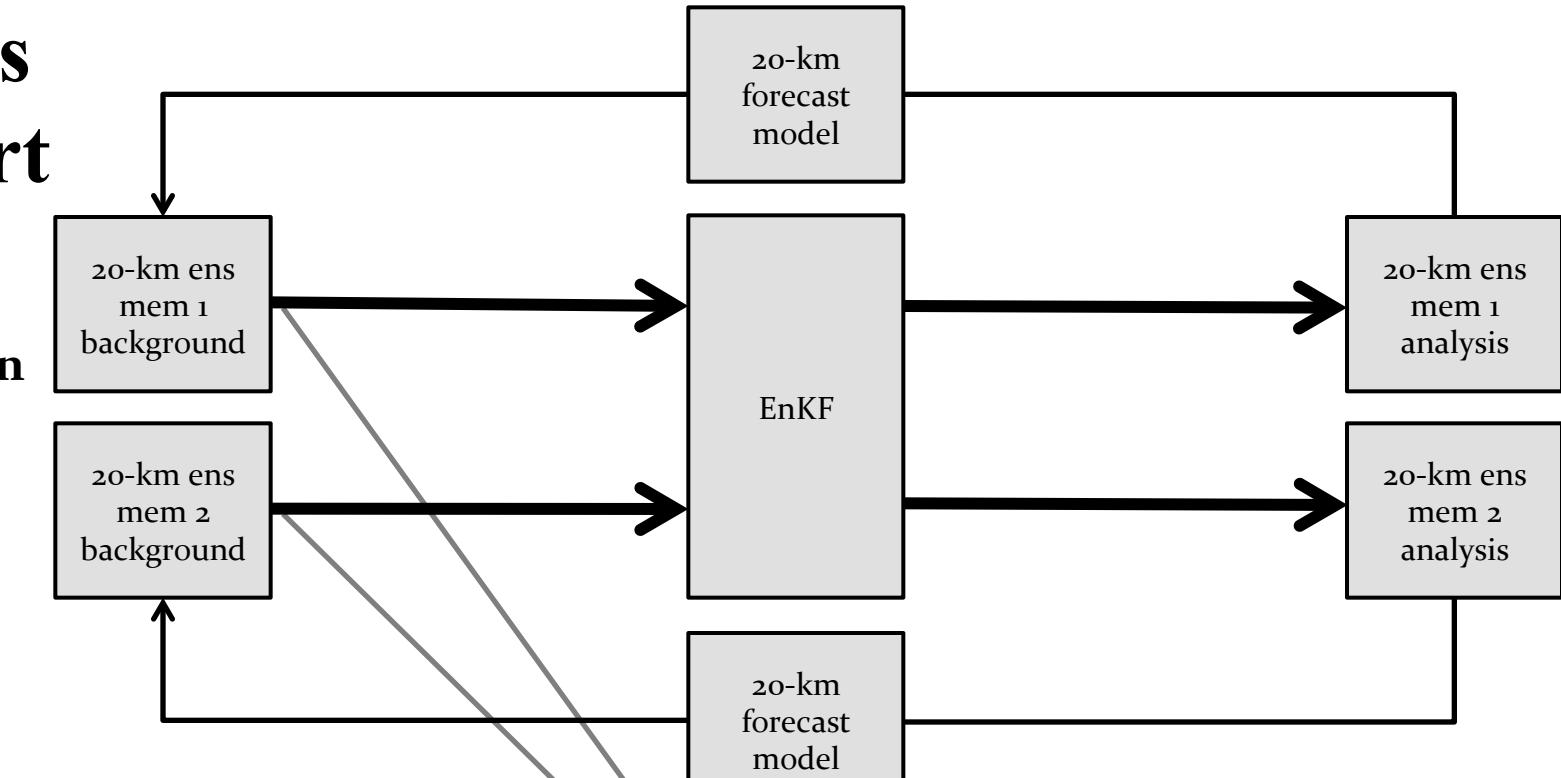
Intermediate domain

- WRFDA directly reads in d01 ensembles, then cuts to d02 size (**making use of WRF model nest namelist setting**)



Dual-res flowchart

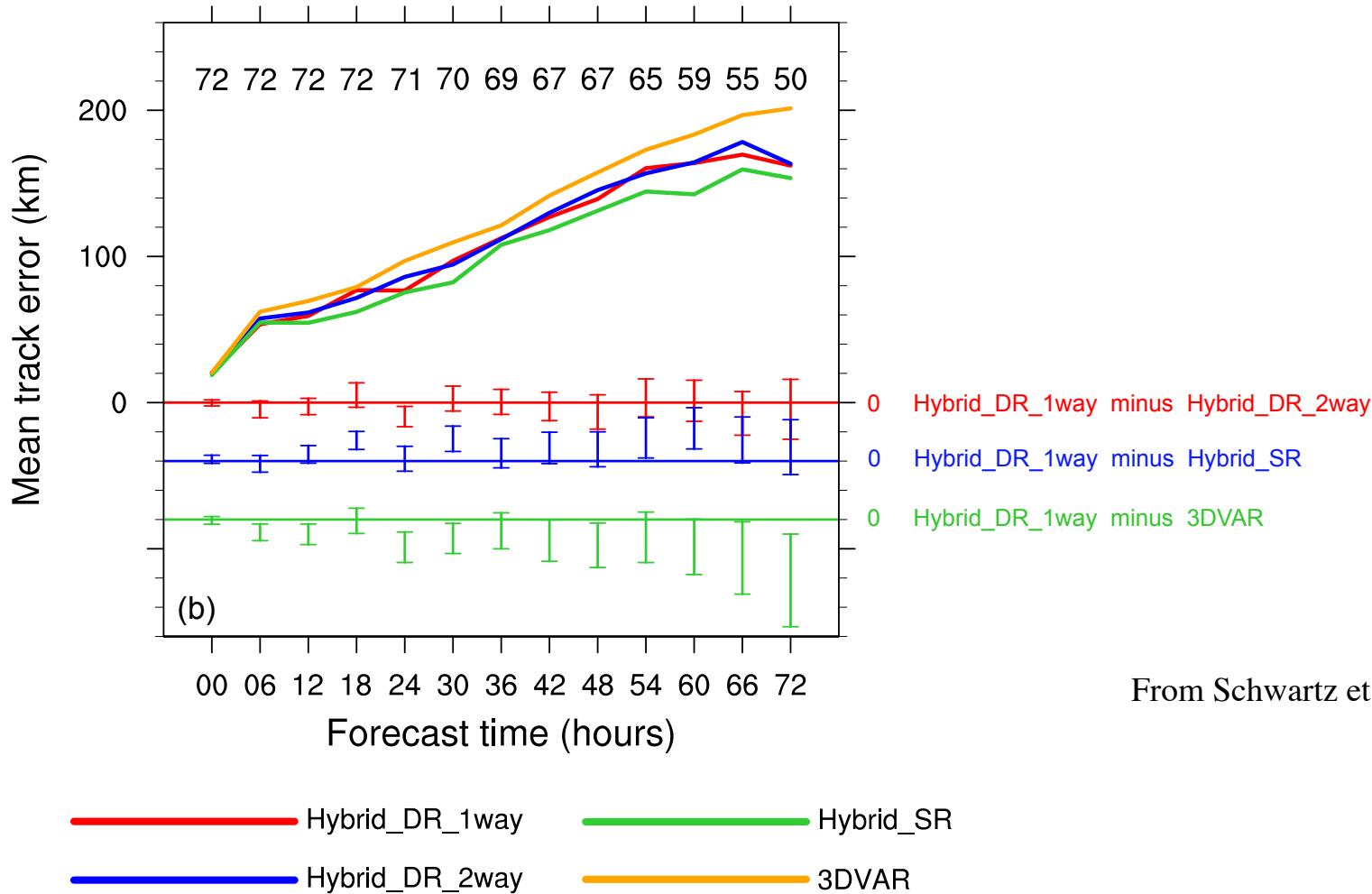
Low-resolution
(20-km)



High-res
(4-km)

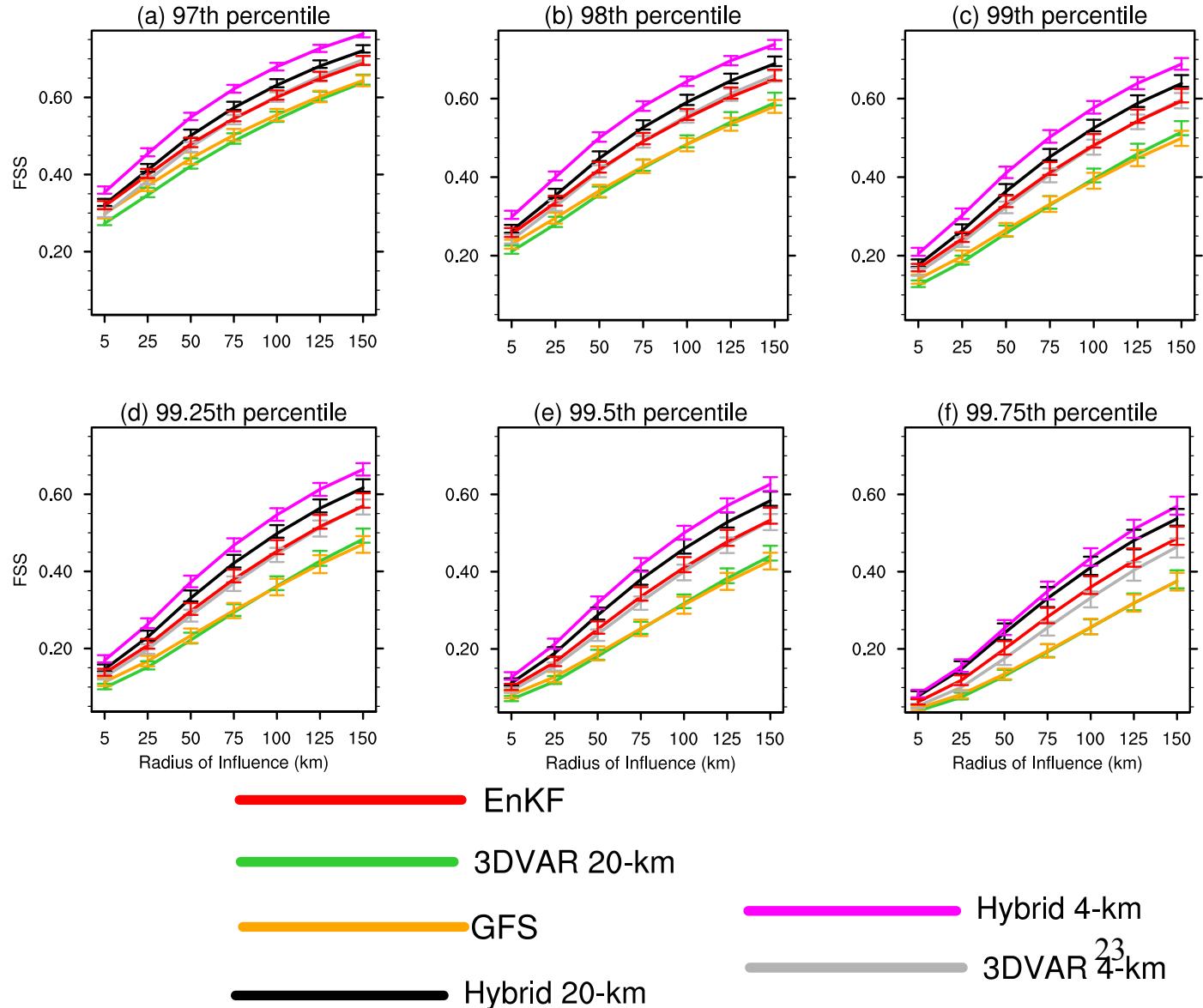
Impact of dual-resolution

- Mean tropical cyclone track errors



Impact of dual-resolution

- Fractions skill score (FSS) aggregated over the first 12 forecast hours and 55 4-km forecasts



Hybrid practice

- **Computation steps:**
 - Compute ensemble mean (**gen_be_ensmean.exe**)
 - Extract ensemble perturbations (**gen_be_ep2.exe**)
 - Run WRFDA in “hybrid” mode (**da_wrfvar.exe**)
 - Display results for: ens_mean, std_dev, ensemble perturbations, hybrid increments, cost function
 - If time permits, play with different namelist settings: “je_factor” and “alpha_corr_scale”
- **Scripts to use:**
 - Some NCL scripts to display results
- **Ensemble generation part not included in current practice**

Namelist for WRFDA in hybrid mode

```
&wrfvar7
je_factor=2,    # half/half for ensemble and static B weightings (tunable parameter)
&wrfvar16
alphacv_method=2,      # ensemble part is in model space (u,v,t,q,ps)
ensdim_alpha=10,       # ensemble size
alpha_corr_type=3,      # 1=Exponential; 2=SOAR; 3=Gaussian
alpha_corr_scale=750.,  # correlation scale in km (tunable parameter)
alpha_std_dev=1.,
alpha_vertloc=true, (use program “gen_be_vertloc.exe 42” to generate file)
hybrid_dual_res = .false. # If true, hybrid is in “dual-resolution” mode
```

Namelist for dual-resolution hybrid

- Dual-resolution hybrid uses WRF nesting to define grids, so also need to specify **nested domain** geometry in the namelist
- Analysis on the nested domain (i.e., “d02”), but using the ensemble from the parent domain (i.e., “d01”)
- When running in dual-resolution mode, also need to link “d01” file to run directory as “./fg_ens”:

ln -sf \${dir}/wrfinput_d01 ./fg_ens (ensemble grid)

ln -sf \${dir}/wrfinput_d02 ./fg (high-res background)

```
&wrfvar16  
hybrid_dual_res = .true.  
  
&domains  
e_we = 222, 316  
e_sn = 128, 274  
s_vert = 1,1  
e_vert = 45, 45  
dx = 45000, 15000,  
dy = 45000, 15000,  
hypsometric_opt = 2  
max_dom = 2  
grid_id = 1, 2,  
parent_id = 0, 1  
i_parent_start = 0, 74,  
j_parent_start = 0, 17,  
parent_grid_ratio = 1, 3
```

References

- T. M. Hamill and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, 128, 2905–2919.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, 126, 796–811.
- Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, 230, 112–126.
- Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-VAR. *Quart. J. Roy. Meteor. Soc.*, 129, 3183–3203.
- Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev.*, 136, 5116-5131.
- Schwartz, C. S., and Z. Liu, 2014: Convection-Permitting Forecasts Initialized with Continuously Cycling Limited-Area 3DVAR, Ensemble Kalman Filter, and “Hybrid” Variational–Ensemble Data Assimilation Systems. *Mon. Wea. Rev.*, 142, 716–738.
- Schwartz, C., Z. Liu, and X. Huang, 2015: Sensitivity of limited-area hybrid variational-ensemble analyses and forecasts to ensemble perturbation resolution. *Mon. Wea. Rev.* doi:10.1175/MWR-D-14-00259.1, in press.

References

- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Anderson, J. L., 2003: A Local Least Squares Framework for Ensemble Filtering. *Mon. Wea. Rev.*, **131**, 634–642.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Liu, Z., C. S. Schwartz, C. Snyder, and S.-Y. Ha, 2012: Impact of assimilating AMSU-A radiances on forecasts of 2008 Atlantic tropical cyclones initialized with a limited-area ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 4017–4034.
- Schwartz, C. S., Z. Liu, X.-Y. Huang, Y.-H. Kuo, and C.-T. Fong, 2013: Comparing limited-area 3DVAR and hybrid variational-ensemble data assimilation methods for typhoon track forecasts: Sensitivity to outer loops and vortex relocation. *Mon. Wea. Rev.*, **141**, 4350–4372.