Hybrid Variational/Ensemble Data Assimilation

Zhiquan Liu (liuz@ucar.edu)

NCAR/NESL/MMM

Outline

- Background
- Hybrid formulation in a variational framework
- Some results
- Introduction to hybrid practice

Motivation of Hybrid DA

- 3D-Var uses static ("climate") BE

$$J(\delta x) = \frac{1}{2}\delta x^{\mathrm{T}}\mathrm{B}^{-1}\delta x + \frac{1}{2}[\mathrm{H}\delta x - d]^{\mathrm{T}}\mathrm{R}^{-1}[\mathrm{H}\delta x - d]$$

- 4D-Var implicitly uses flow-dependent information, but still starts from static BE

$$J(\delta x) = \frac{1}{2}\delta x^{\mathrm{T}}\mathrm{B}^{-1}\delta x + \frac{1}{2}\sum_{i=1}^{I}[\mathrm{HM}_i\delta x - d_i]^{\mathrm{T}}\mathrm{R}^{-1}[\mathrm{HM}_i\delta x - d_i]$$

- Hybrid uses flow-dependent background error covariance from forecast ensemble perturbation in a variational DA system

What is the Hybrid DA?

- Ensemble mean is analyzed by a variational algorithm (i.e., minimize a cost function).
  - It combines (so "hybrid") the 3DVAR "climate" background error covariance and "error of the day" from ensemble perturbation.

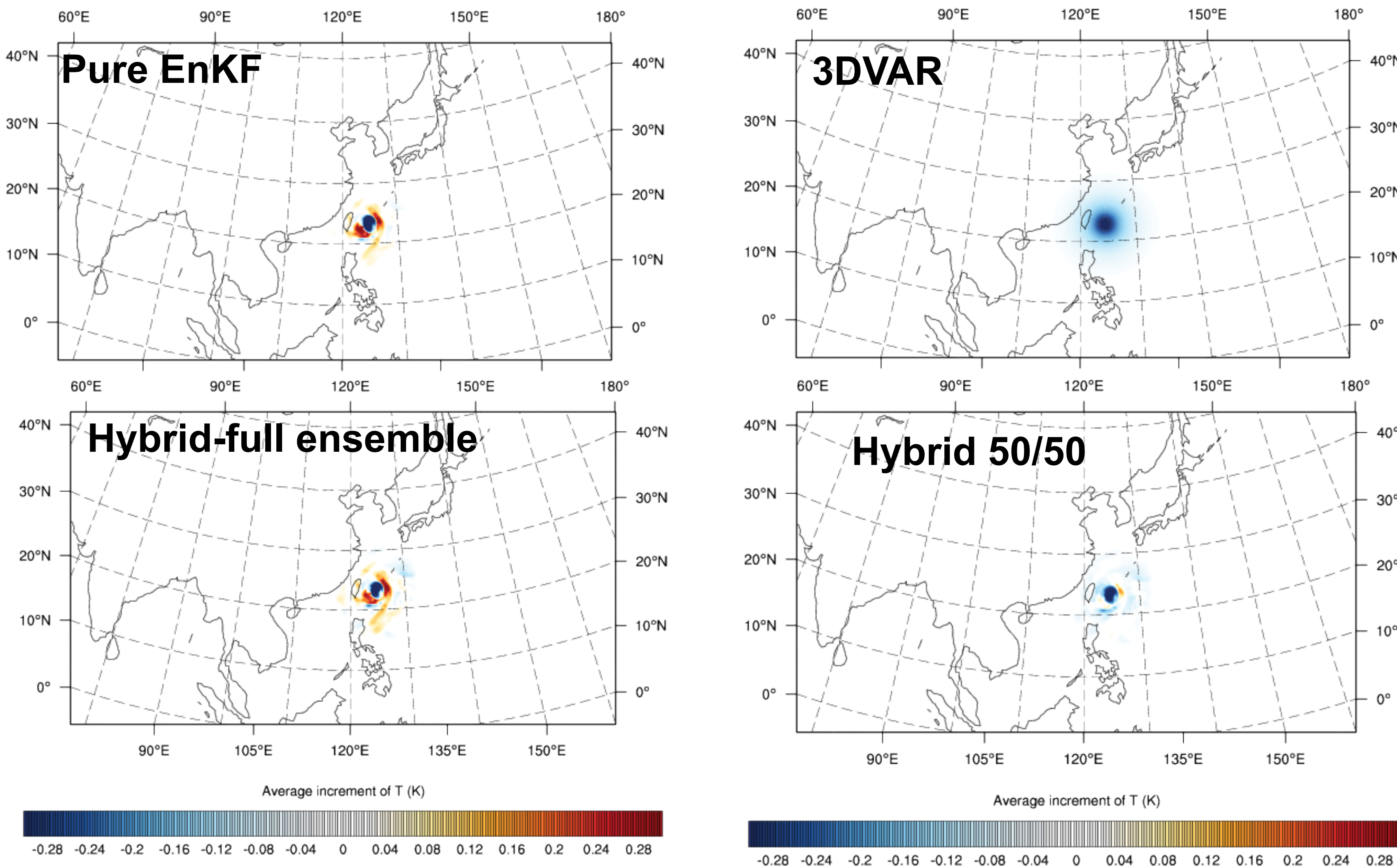
- Hybrid algorithm (again in a variational framework) itself usually does not generate ensemble analyses.

- Need a separate system to update ensemble
  - Could be ensemble forecasts already available from NWP centers
  - Could be an Ensemble Kalman Filter-based DA system
  - Or multiple model/physics ensemble

- Ensemble needs to be good to well represent "error of the day"

single observation tests

Potential temperature increment, 21st model level



Hybrid formulation (1)
(Hamill and Snyder, 2000)

- 3DVAR cost function

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

- Idea: replace **B** by a weighted sum of static $\mathbf{B}_{\mathrm{s}}$ and the ensemble $\mathbf{B}_{\mathrm{e}}$

$$\mathbf{B} = a_s\mathbf{B}_{\mathrm{s}} + a_e\mathbf{B}_{\mathrm{e}},\ a_s = 1 - a_{\mathrm{e}}$$

  - Has been demonstrated on a simple model.
  - Difficult to implement for large NWP model.

Hybrid formulation (2): used in WRFDA
(Lorenc, 2003)

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

$$J(\mathbf{x},\alpha) = \beta_s \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^{\mathbf{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \overbrace{\beta_e \frac{1}{2}\sum_{i=1}^{\mathrm{N}}\alpha_i^{\mathbf{T}}\mathbf{C}^{-1}\alpha_i}^{\text{ensemble control variable } \alpha_i \ (M\times 1)}$$

$$+\frac{1}{2}[\mathbf{y}-H(\mathbf{x}+\mathbf{x}_e^{'})]^{\mathbf{T}}\mathbf{R}^{-1}[\mathbf{y}-H(\mathbf{x}+\mathbf{x}_e^{'})]$$

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

$\circ$ denote element - wise product. α_{i} is in effect the ensemble weight.

$\mathbf{C}$: correlation matrix (effectively loclization of ensemble perturbations)

• In practical implementation, α_i can be reduced to horizontal 2D fields (i.e., use same weight in different vertical levels) to save computing cost.

• β_s and β_e ($1/\beta_s + 1/\beta_e = 1$) can be tuned to have different weight between static and ensemble part.

Hybrid formulation (3)

- Equivalently can write in another form (Wang et al., 2008)

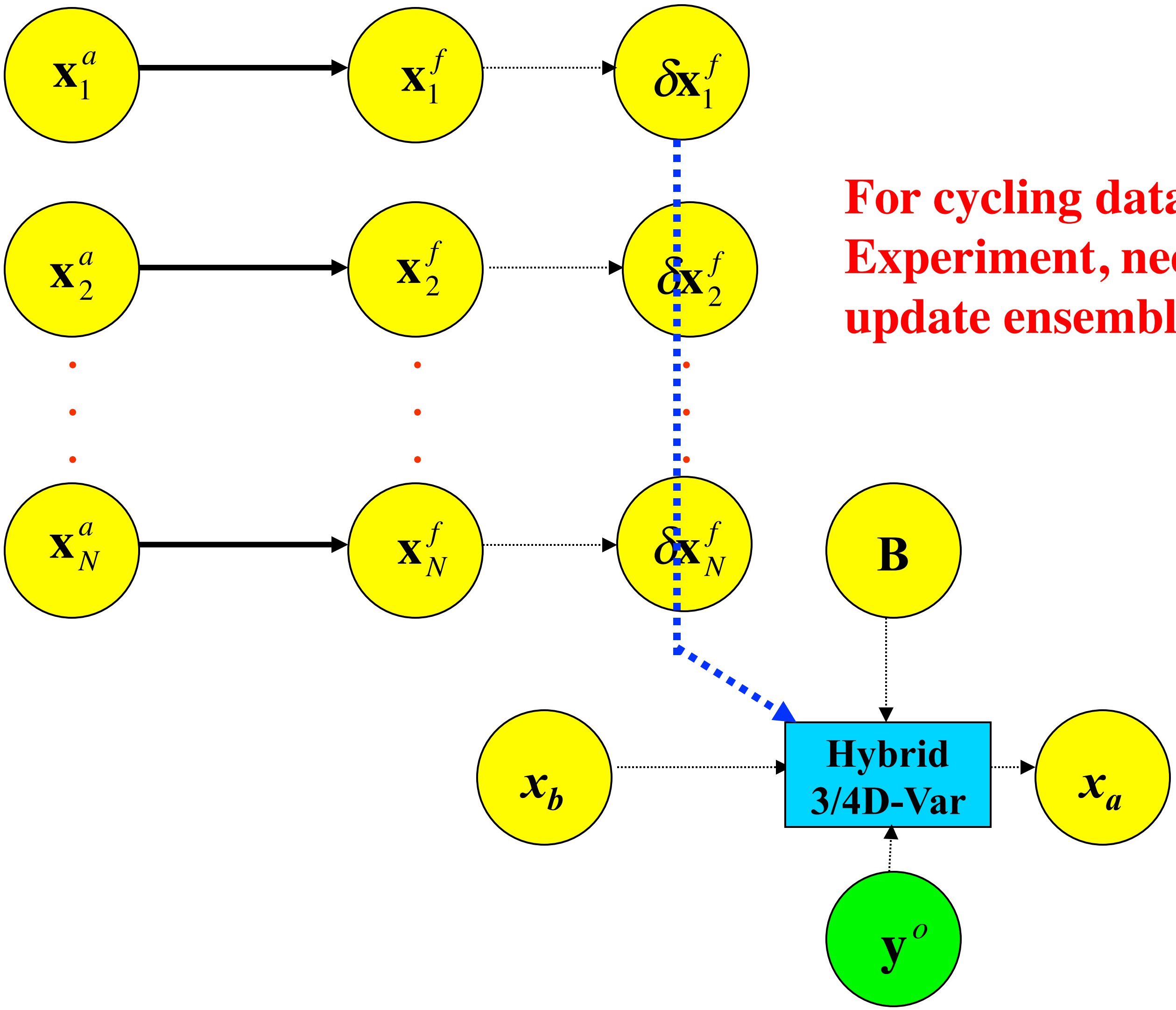
$$J(\mathbf{x},\alpha) = \frac{1}{2}(\mathbf{x} + \mathbf{x}_e - \mathbf{x}_b)^{\mathbf{T}}\left(\frac{1}{\beta_s}\mathbf{B} + \frac{1}{\beta_e}\mathbf{B}_e \circ \mathbf{C}\right)^{-1}(\mathbf{x} + \mathbf{x}_e - \mathbf{x}_b)$$
$$+ \frac{1}{2}[\mathbf{y} - H(\mathbf{x} + \mathbf{x}_e)]^{\mathbf{T}}\mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x} + \mathbf{x}_e)]$$

- **C** is "localization" matrix

Hybrid DA data flow

Ensemble Perturbations (extra input for hybrid)

For cycling data assimilation/forecast Experiment, need a mechanism to update ensemble.



EnKF-based Ensemble Generation

- EnKF with perturbed observations

- EnKF without perturbed observations
  - All based on square-root filter
  - Ensemble Transformed Kalman Filter (ETKF)
  - Ensemble Adjustment Kalman Filter (EAKF)
  - Ensemble Square-Root Filter (EnSRF)

- Most implementation assimilates obs sequentially (i.e., one by one, or box by box)
  - can be parallelized

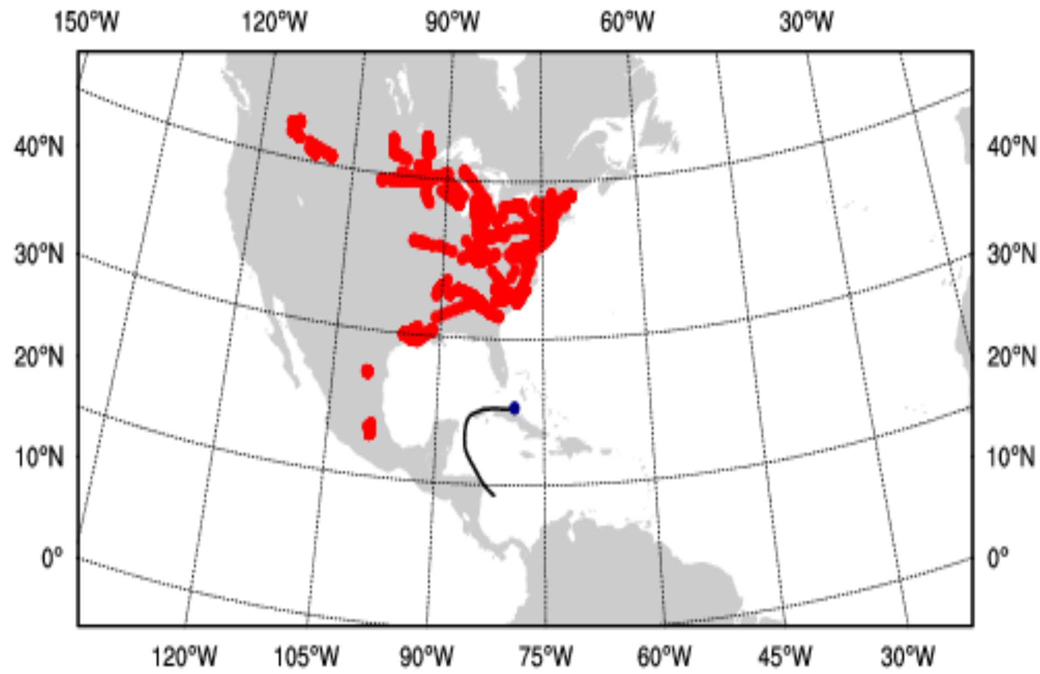
More information was given in 2012 slides.

Advantages of the Hybrid DA

- Hybrid localization is in model space while EnKF localization is usually in observation space.
- For some observations type, e.g., radiances, localization is not well defined in observation space
- Easier to make use of existing radiance VarBC in hybrid
- For small-size ensemble, use of static B could be beneficial to have a higher-rank covariance.

a Hurricane Case Study (Dongmei Xu)

- Paula case: 0600 UTC 10 October 2010 to 1200 UTC 15 October 2010;
- Background: 15km interpolated from GFS data;
- Resolution: 718x 373 (15km) and 43 levels;
- Observations: GTS and TAMDAR;
- Cycle frequency: 6 hours;
- Background error:CV5;
- Time widows: 2 hours;



- TAMDAR: a new Tropospheric Airborne Meteorological Data Reporting (TAMDAR) observing system that has been developed by AirDat company.

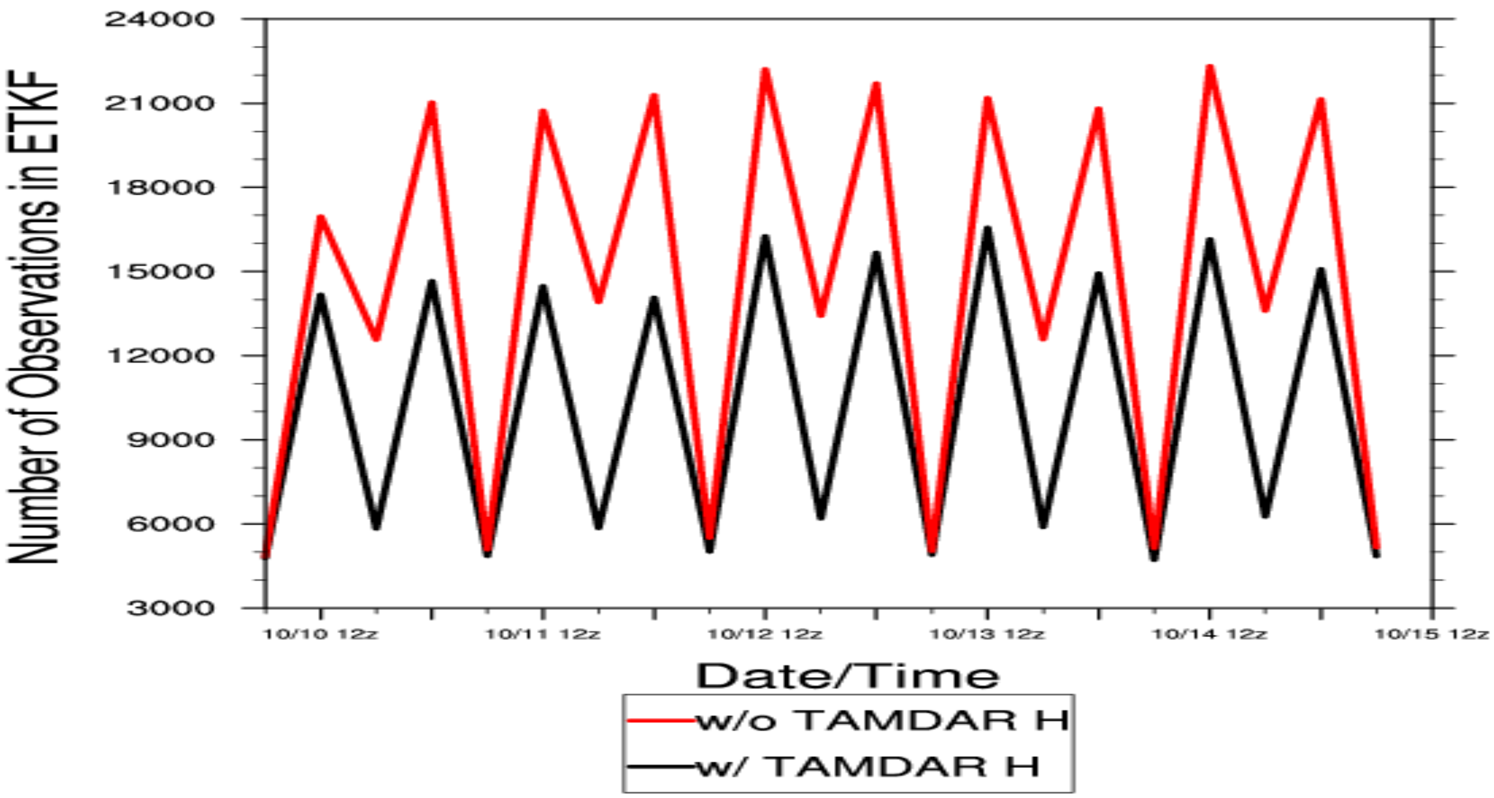
Experimental design

Experiments:

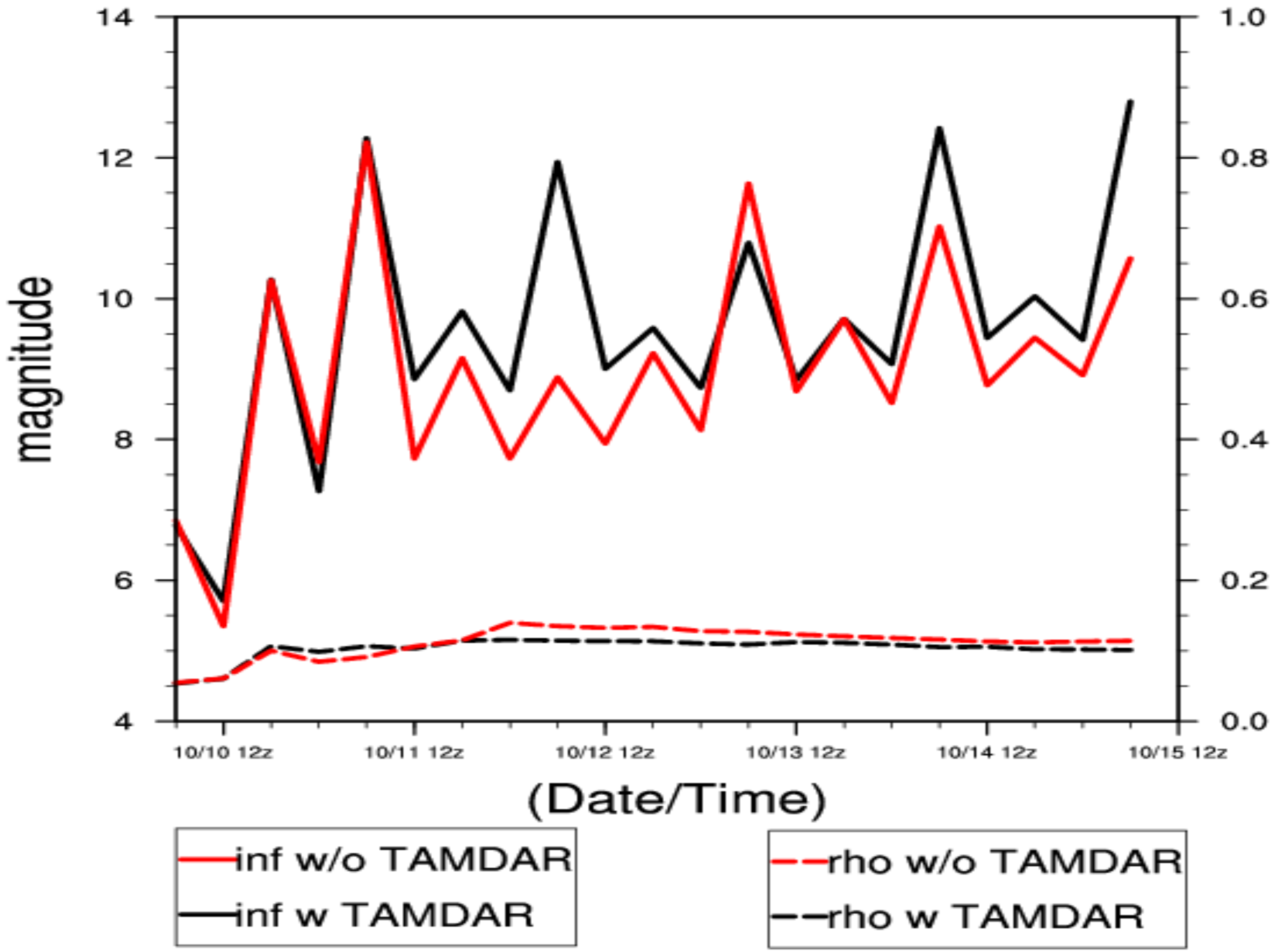
CYC1:assimilate GTS and TAMDAR with Hybrid (w/ TAMDAR H);

CYC2:same to CYC1,but no TAMDAR (w/o TAMDAR H)

CYC3:assimilate GTS and TAMDAR with standard 3DVAR (Deterministic WRFDA)

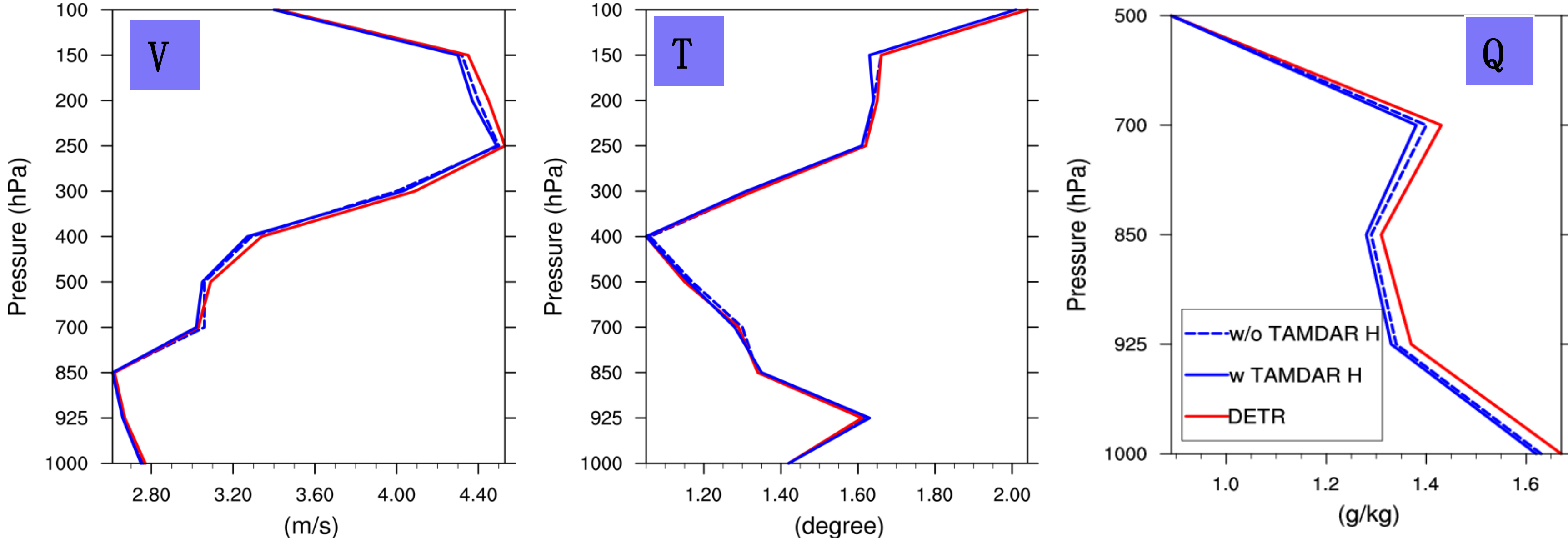


inflation and fraction factor♪

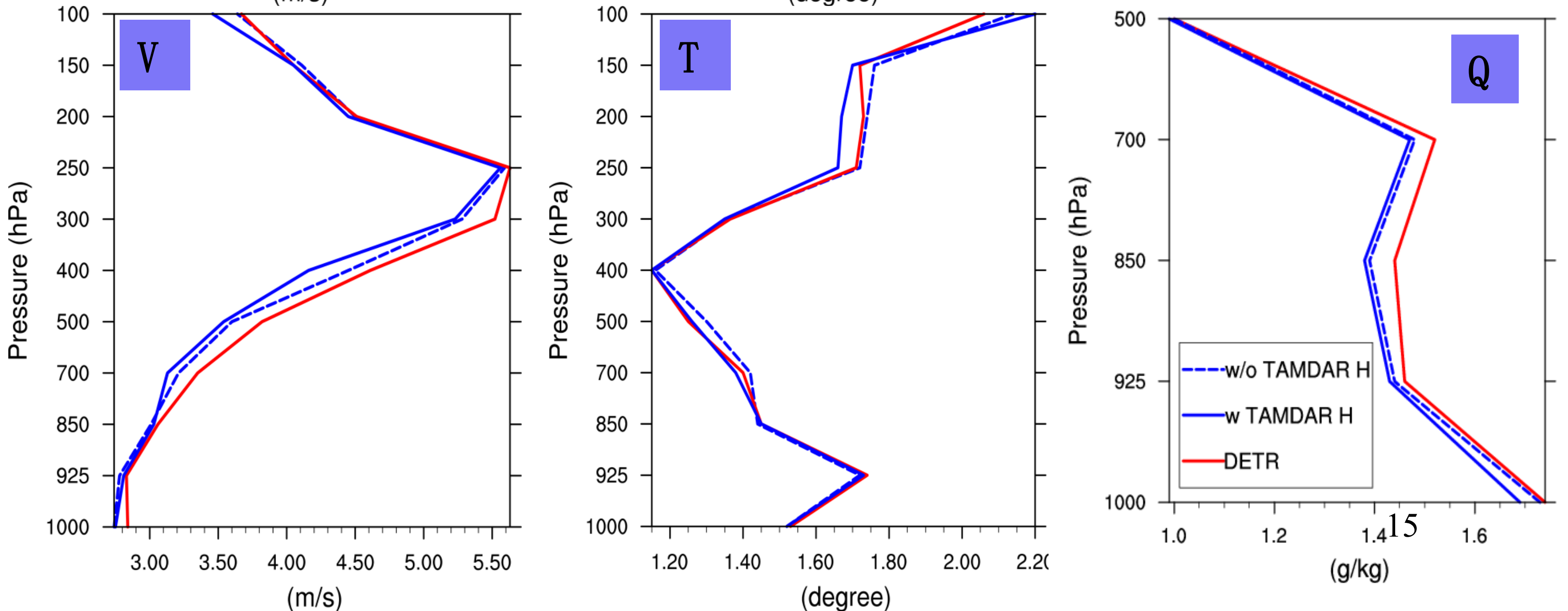


Forecast Verification: RMSE♪

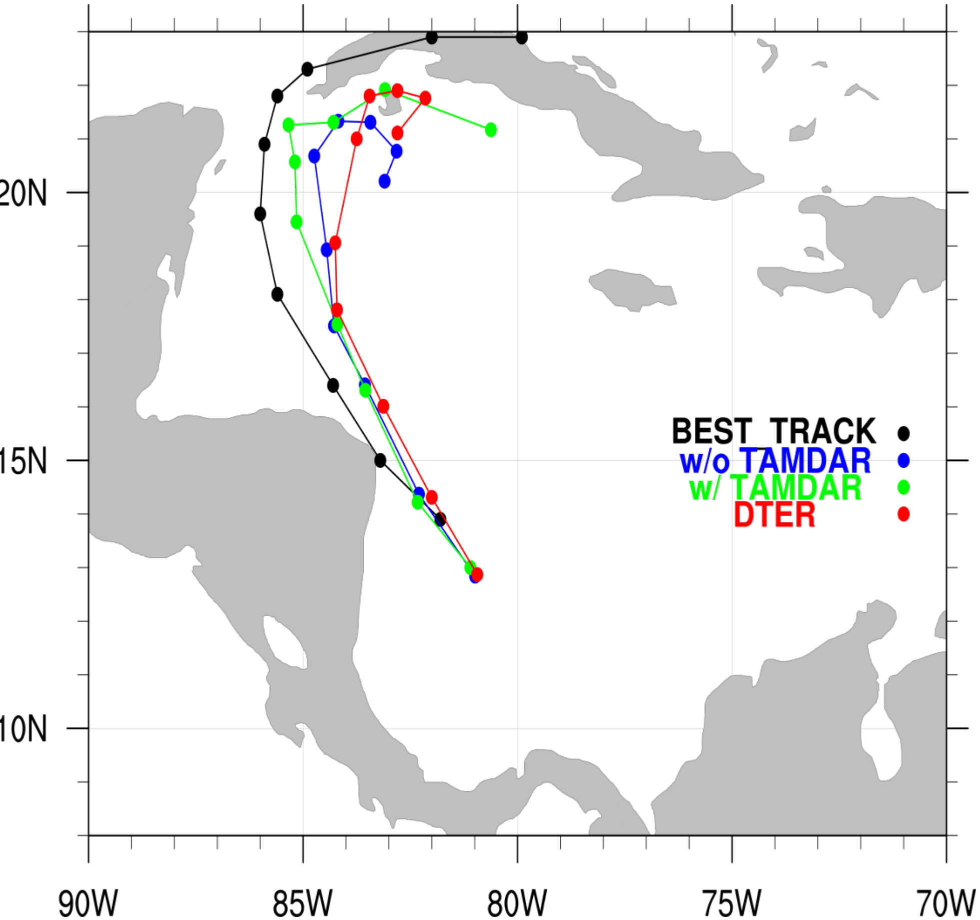
+12hr

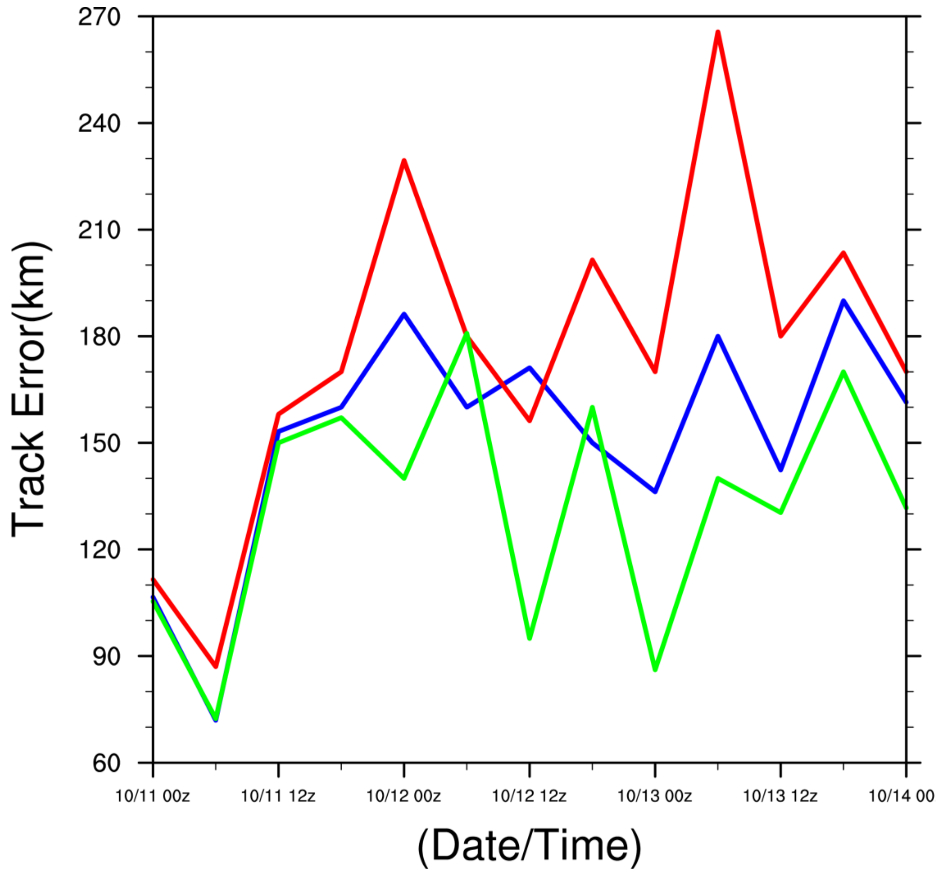


+24hr



Track Forecast Verification (+24hr)♪





Hybrid practice

- **Computation steps:**
  - Computing ensemble mean (**gen_be_ensmean.exe**).
  - Extracting ensemble perturbations (**gen_be_ep2.exe**).
  - Running WRFDA in "hybrid" mode (**da_wrfvar.exe**).
  - Displaying 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 Jb and Je term (tunable parameter)

&wrfvar16
alphacv_method=2,   # ensemble part is in model space (u,v,t,q,ps)

ensdim_alpha=10,

alpha_corr_type=3, # 1=Exponential; 2=SOAR; 3=Gaussian

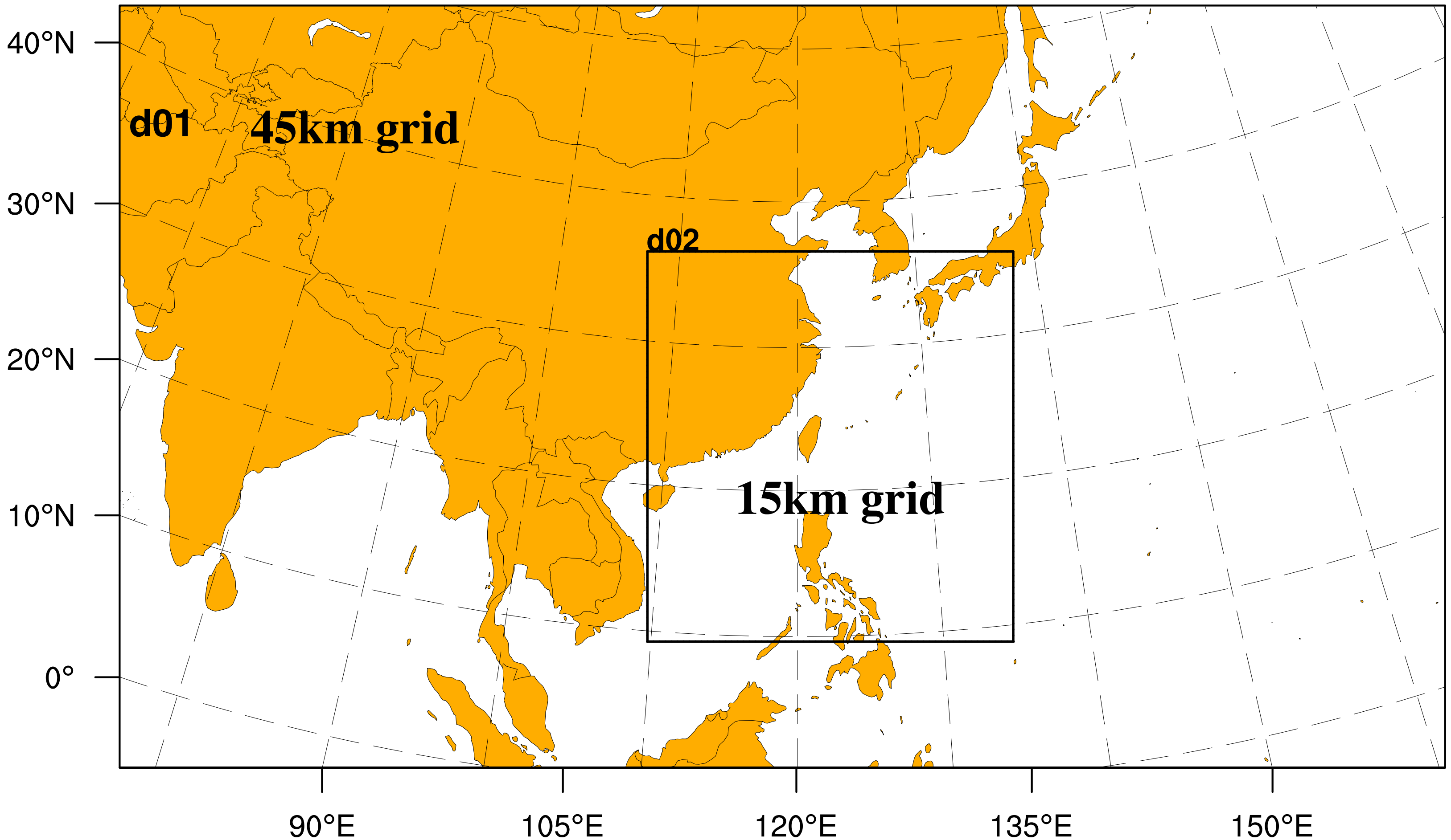
alpha_corr_scale=750., # correlation scale in km (tunable parameter)

alpha_std_dev=1.,

alpha_vertloc=true, (use program "**gen_be_vertloc.exe 42" to generate file**)

Dual-Resolution hybrid (V3.6)

http://www2.mmm.ucar.edu/wrf/users/workshops/WS2014/ppts/6A.3.pdf



Doing Hybrid-Analysis at 15km d02 grid but with ensemble perturbation input from 45km d01 grid

Dual-resolution cost-function

- High-resolution (HR) variables:
  - $\mathbf{x}_1$, $\mathbf{B}$, $\mathbf{H}$, $\delta\mathbf{x}$
- Low-resolution (LR) variables:
  - $\mathbf{a}$, $\mathbf{A}$, $\mathbf{D}$

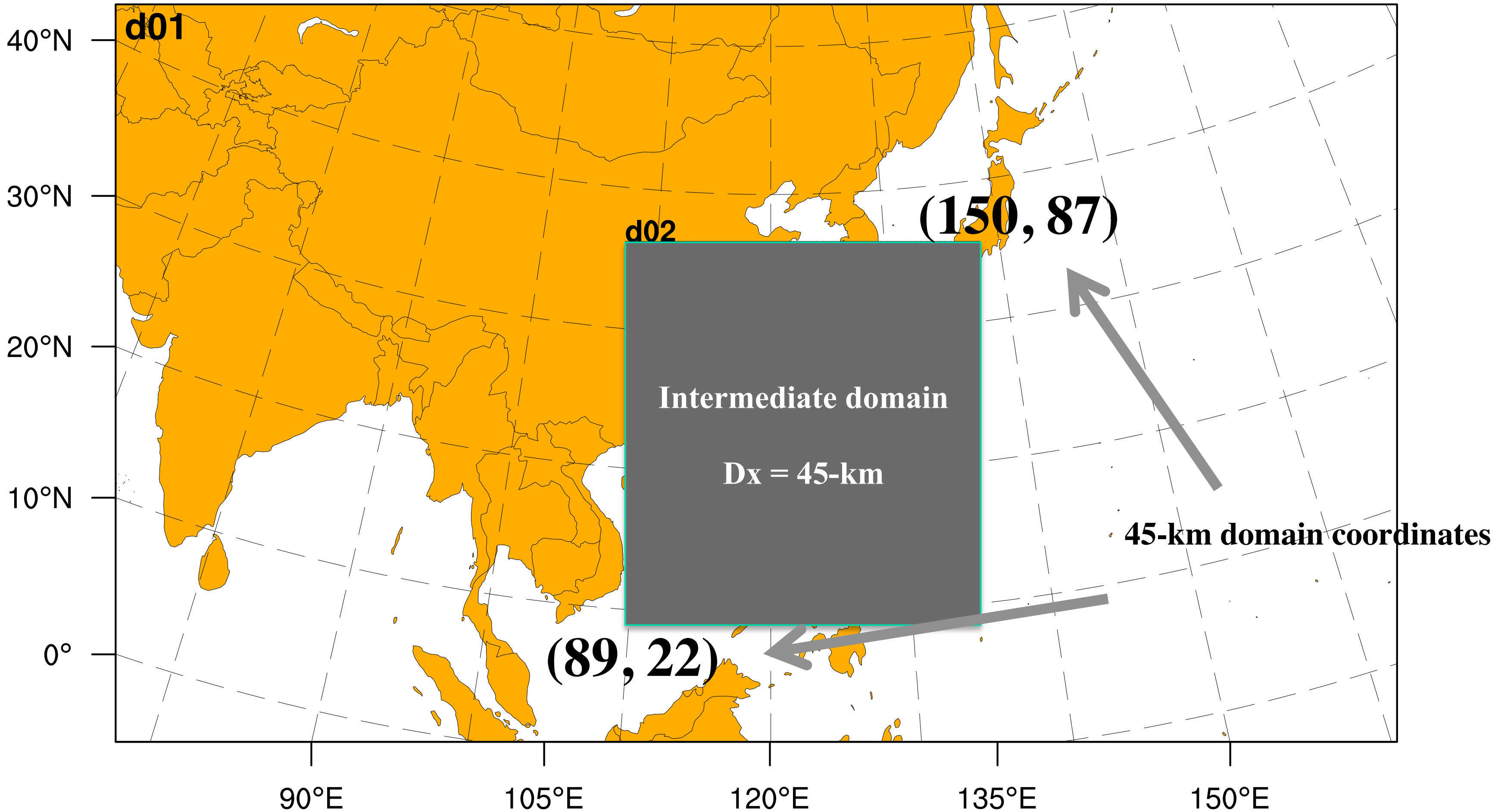
$$J(\mathbf{x}_1,\mathbf{a}) = \underbrace{\boxed{\frac{\beta_1}{2}(\mathbf{x}_1)^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x}_1}}_{\text{HR}}^{\text{Scalar!}} + \underbrace{\boxed{\frac{\beta_2}{2}\mathbf{a}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{a}}}_{\text{LR}}^{\text{Scalar!}} + \overbrace{\boxed{\frac{1}{2}(\mathbf{d}-\underbrace{\mathbf{H}\delta\mathbf{x}}_{\text{HR}})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{d}-\underbrace{\mathbf{H}\delta\mathbf{x}}_{\text{HR}})}}^{\text{Scalar!}}$$

$$\underbrace{\boxed{\delta\mathbf{x}}}_{\text{HR}} = \underbrace{\boxed{\mathbf{x}_1}}_{\text{HR}} + \underbrace{\boxed{\mathbf{Da}}}_{\text{LR}}$$

This term requires interpolation from low to high resolution.

Intermediate domain

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



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. Me teor. 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.

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. R ev., 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 tropic al 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-are a 3DVAR and hybrid variational-ensemble data assimilation methods for typhoon track forecas ts: Sensitivity to outer loops and vortex relocation. *Mon. Wea. Rev.*, **141, 4350-4372.**