Improving Global and Hurricane Predictions by Using Minimum-Cost Large Ensembles in GFS 4DEnVar Hybrid Data Assimilation System

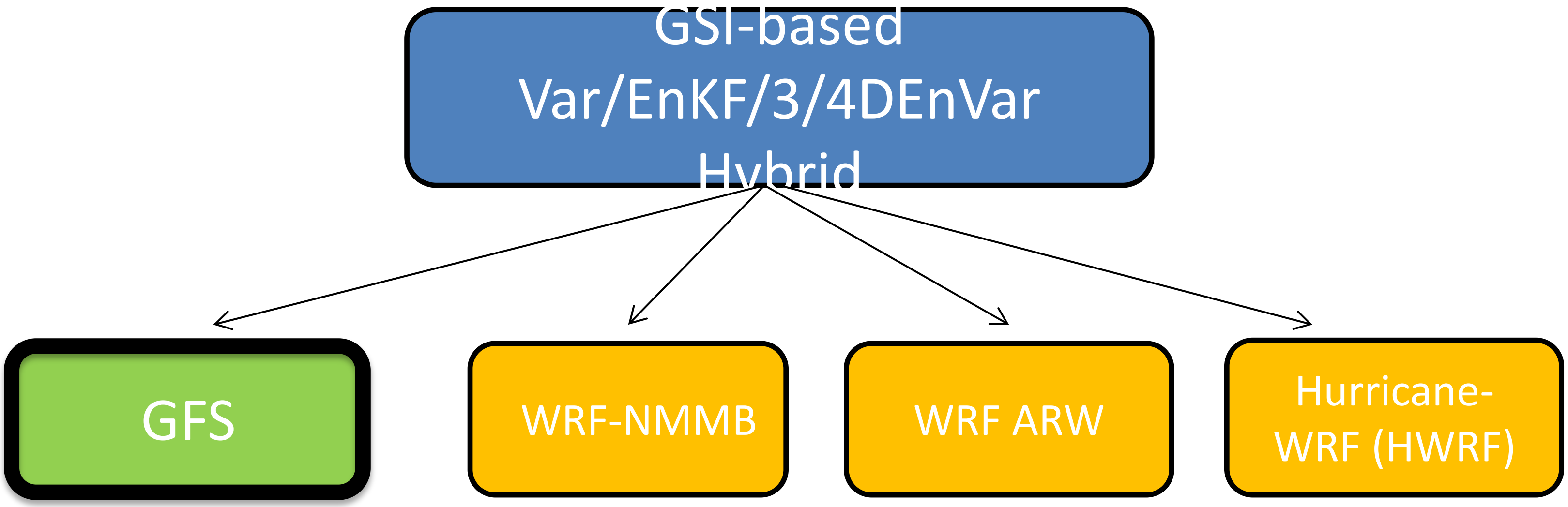
PI: Xuguang Wang

School of Meteorology
University of Oklahoma, Norman, OK, USA

Collaborators:
Jeff Whitaker (NOAA/ESRL)
Daryl Kleist (UMD)
Rahual Mahajan, Yuejian Zhu, John Derber (NOAA/EMC)

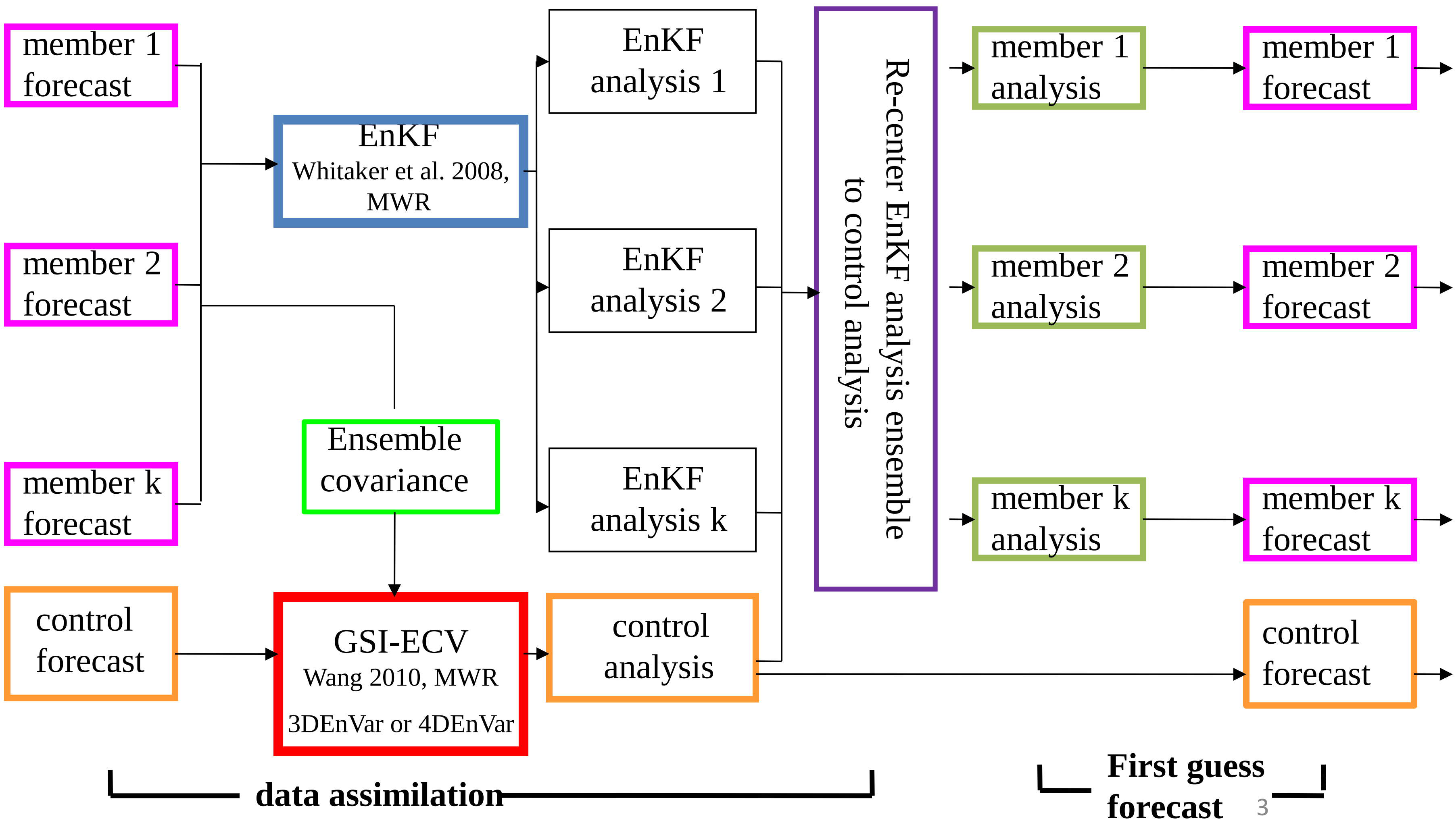
NGGPS external PI meeting, July 16-17, 2015

GSI based EnKF-Var hybrid system

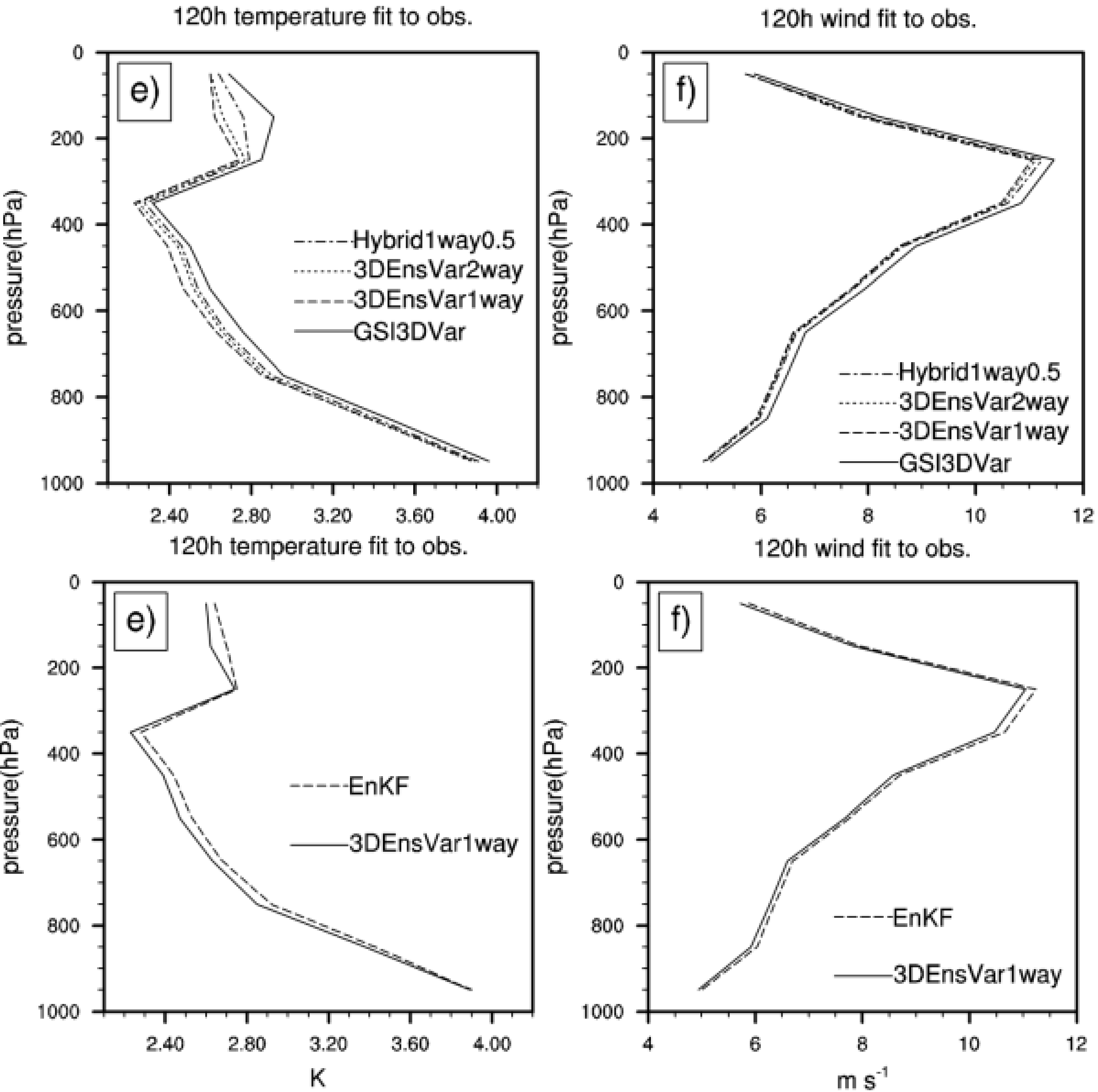


GSI-based EnKF-Var hybrid DA system

Wang, Parrish, Kleist, Whitaker 2013, MWR



GSI hybrid for GFS:
GSI 3DVar vs. 3DEnVar Hybrid vs. EnKF

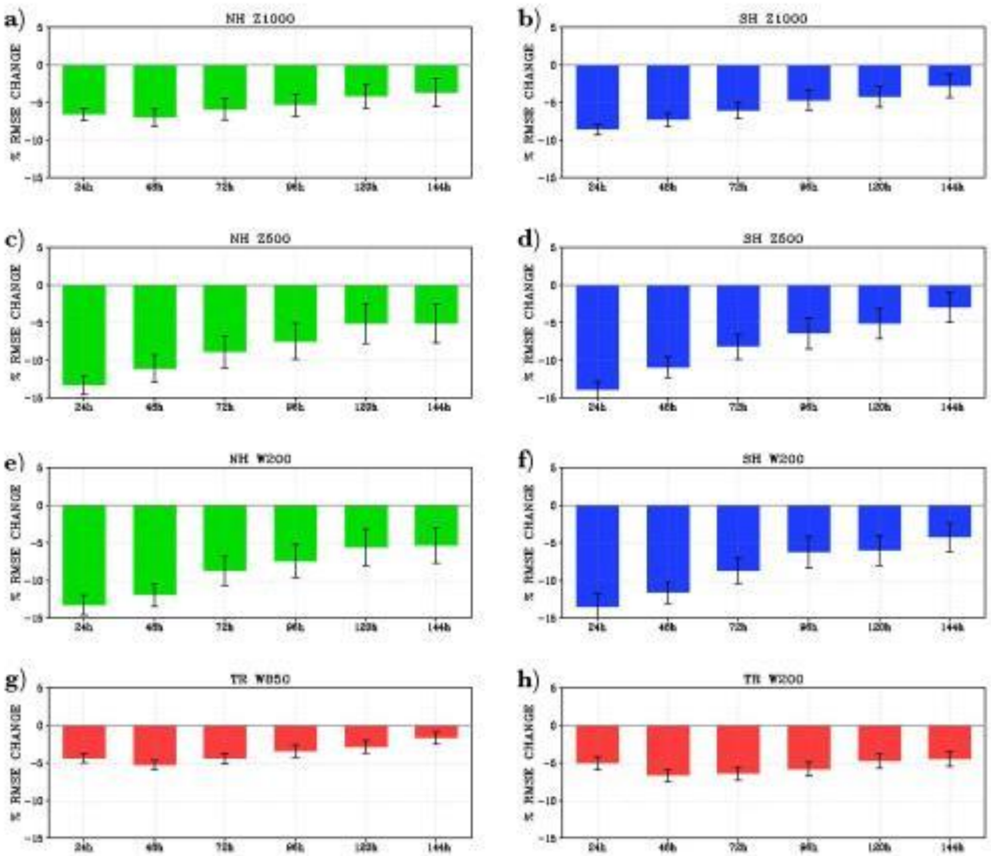


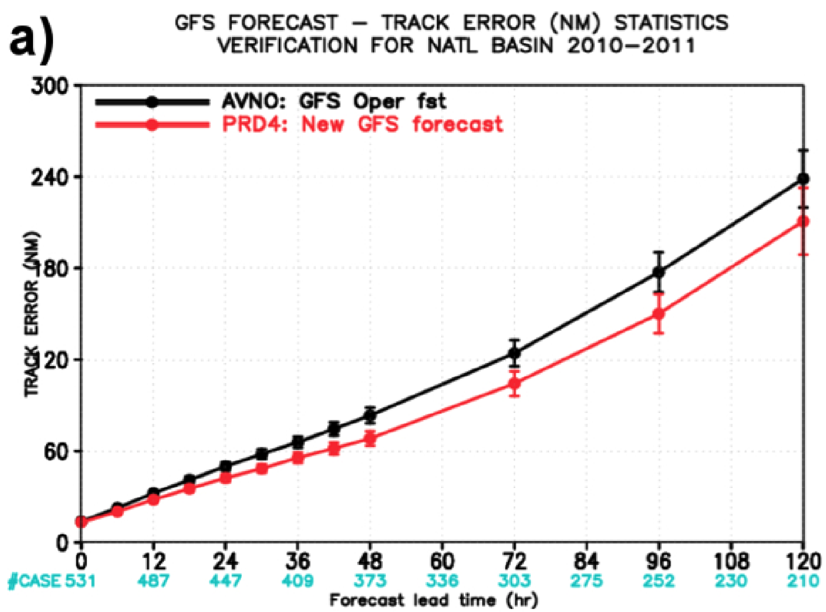
- 3DEnVar Hybrid was better than 3DVar due to use of flow-dependent ensemble covariance

- 3DEnVar was better than EnKF due to the use of tangent linear normal mode balance constraint (TLNMC)

Wang, Parrish, Kleist and Whitaker, MWR, 2013, 141, 4098-4117

GSI hybrid for GFS:
NCEP pre-implementation test for dual resolution 3DEnVar





Courtesy: Daryl Kleist

GSI hybrid for GFS: 3DEnVar vs. 4DEnVar

- **GSI-3DEnVar**: Extended control variable (ECV) method is used within GSI variational minimization (Wang 2010, MWR):

$$J(\mathbf{x}_1', \boldsymbol{\alpha}) = \beta_1 J_1 + \beta_2 J_e + J_o$$

$$= \beta_1 \frac{1}{2}\mathbf{x}_1'^T \mathbf{B}^{-1}\mathbf{x}_1' + \beta_2 \underline{\frac{1}{2}\boldsymbol{\alpha}^T \mathbf{C}^{-1}\boldsymbol{\alpha}} + \frac{1}{2}\left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}'\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}'\right)$$

Extra term associated with extended control variable

$$\mathbf{x}' = \mathbf{x}_1' + \underline{\sum_{k=1}^{K}\left(\boldsymbol{\alpha}_k \circ \mathbf{x}_k^e\right)}$$

Extra increment associated with ensemble

$\mathbf{B}$ stat 3DVAR static covariance; $\mathbf{R}$ observation error covariance; K ensemble size;
$\mathbf{C}$ correlation matrix for ensemble covariance localization; $\mathbf{x}_k^e$ kth ensemble perturbation;
$\mathbf{x}_1'$ 3DVAR increment; $\mathbf{x}'$ total (hybrid) increment; $\mathbf{y}^{o'}$ innovation vector;
$\mathbf{H}$ linearized observation operator; β_1 weighting coefficient for static covariance;
β_2 weighting coefficient for ensemble covariance; $\boldsymbol{\alpha}$ extended control variable.

GSI hybrid for GFS: 3DEnVar vs. 4DEnVar

- **GSI-4DEnVar**: Naturally extended from and unified with GSI-based 3DEnVar hybrid formula (Wang and Lei, 2014, MWR; Kleist and Ide 2015). Conveniently avoid TLA.

Add time dimension in 4DEnsVar

$$J(\mathbf{x}_1', \boldsymbol{\alpha}) = \beta_1 J_1 + \beta_2 J_e + J_o$$

$$= \beta_1 \frac{1}{2} \mathbf{x}_1'^T \mathbf{B}_{static}^{-1} \mathbf{x}_1' + \beta_2 \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{C}^{-1} \boldsymbol{\alpha} + \boxed{\frac{1}{2} \sum_{t=1}^{T} (y_t^o{}' - \mathrm{H}_t \mathbf{x}_t)^{\mathrm{T}} \mathrm{R}_t^{-1} (y_t^o{}' - \mathrm{H}_t \mathbf{x}_t)}$$

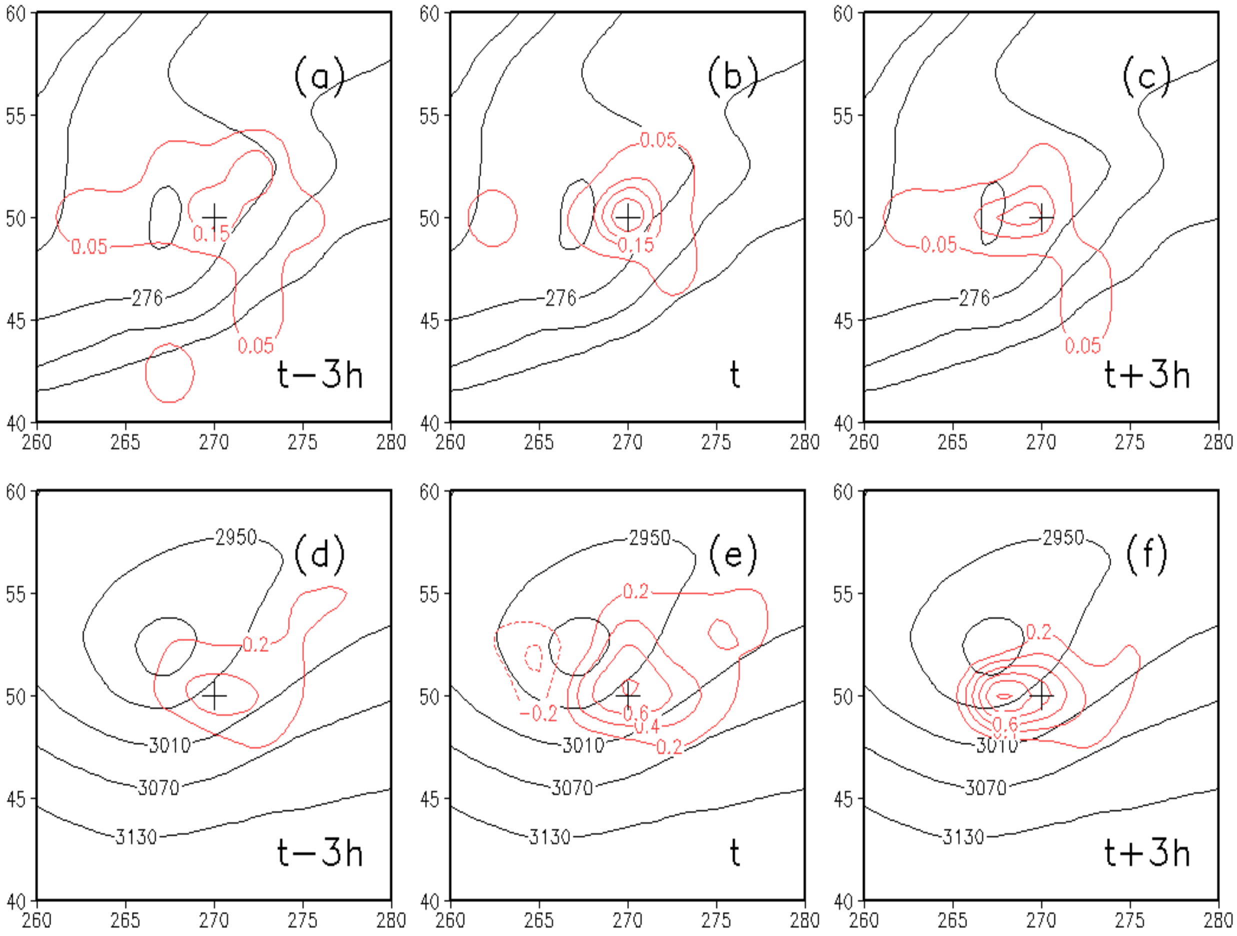
$$\mathbf{x}_t' = \mathbf{x}_1' + \boxed{\sum_{k=1}^{K} \left(\boldsymbol{\alpha}_k \circ (\mathbf{x}_k^e)_t \right)}$$

$\mathbf{B}$ stat 3DVAR static covariance; $\mathbf{R}$ observation error covariance; K ensemble size;
$\mathbf{C}$ correlation matrix for ensemble covariance localization; $\mathbf{x}_k^e$ kth ensemble perturbation;
$\mathbf{x}_1'$ 3DVAR increment; $\mathbf{x}'$ total (hybrid) increment; $\mathbf{y}^{o'}$ innovation vector;
$\mathbf{H}$ linearized observation operator; β_1 weighting coefficient for static covariance;
β_2 weighting coefficient for ensemble covariance; $\boldsymbol{\alpha}$ extended control variable.

Temporal evolution of error covariance by 4DEnVar

Temp.

Height



Downstream impact

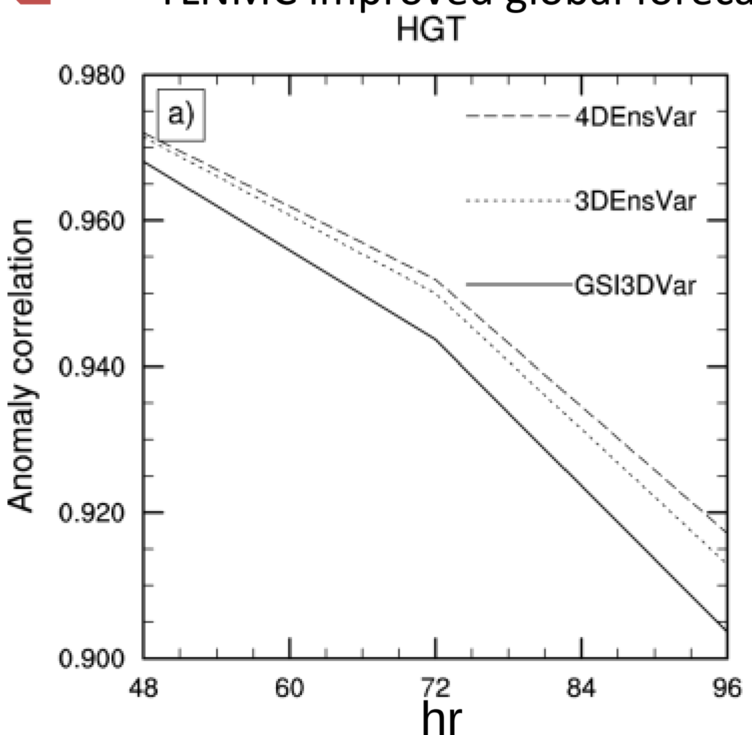
Upstream impact

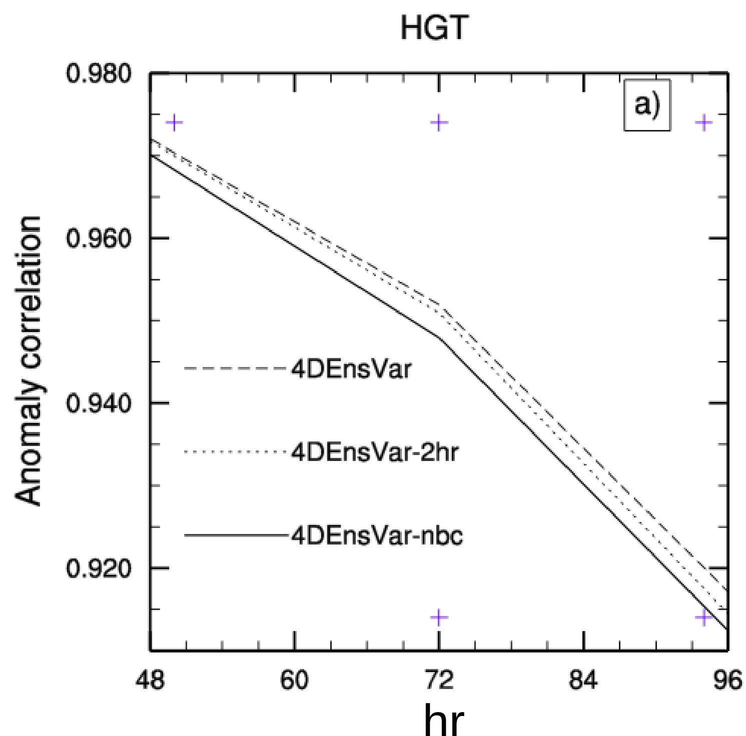
Wang and Lei, 2014, MWR

GSI hybrid for GFS: 3DEnVar vs. 4DEnVar

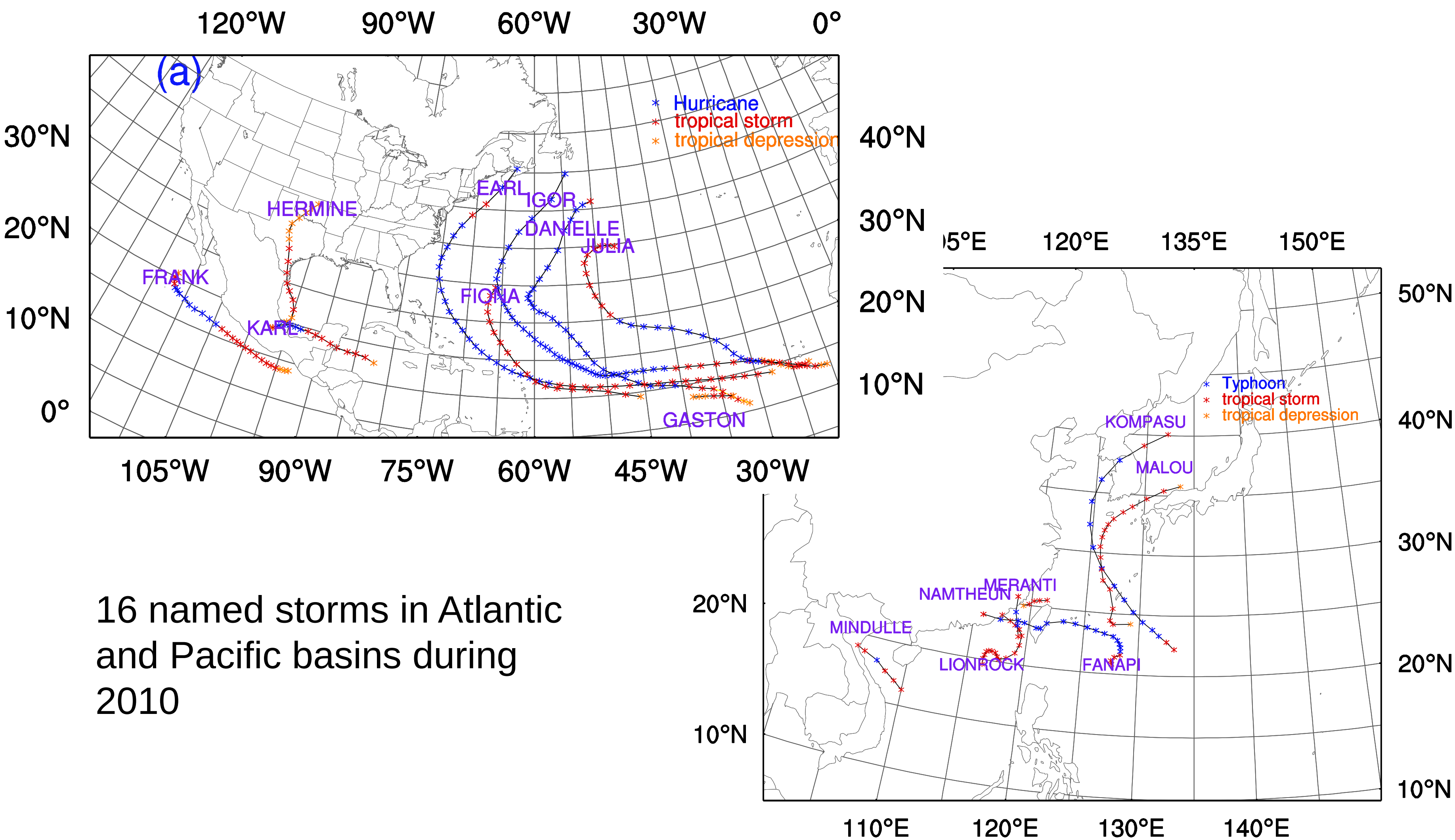
Results from Single Reso. Experiments (Wang and Lei 2014, MWR)

- 4DEnVar improved general global forecasts
- 4DEnVar improved the balance of the analysis
- Performance of 4DEnVar improved if more frequent ensemble perturbations used
- 4DEnVar approximates nonlinear propagation better with more frequent ensemble perturbations
- TLNMC improved global forecasts





GSI hybrid for GFS: 3DEnVar vs. 4DEnVar



16 named storms in Atlantic and Pacific basins during 2010

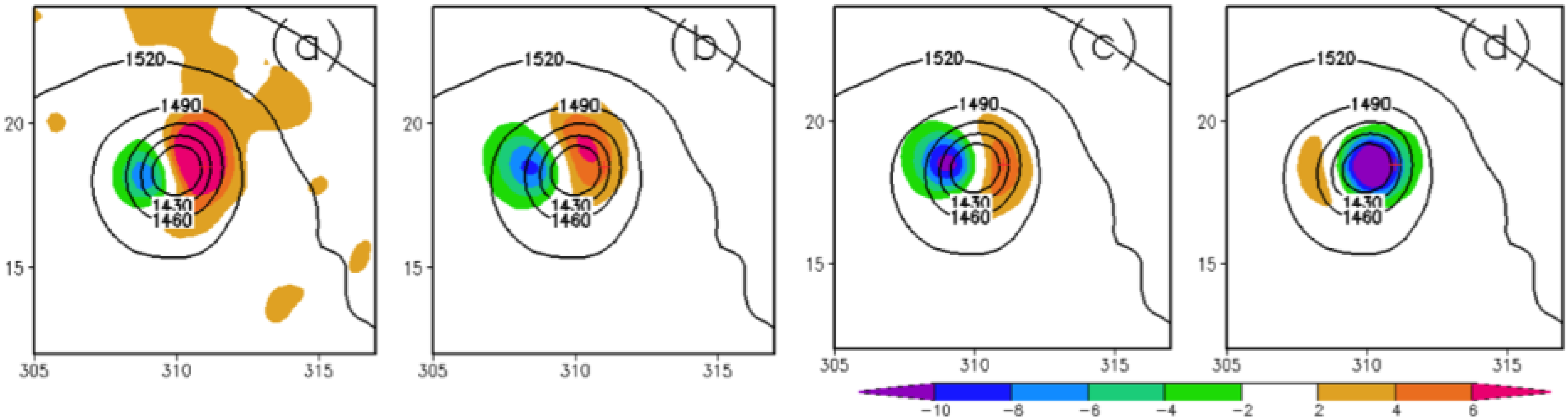
Approximation to nonlinear propagation

–3h increment propagated by model integration

4DEnVar (hrly pert.)

4DEnVar (2hrly pert.)

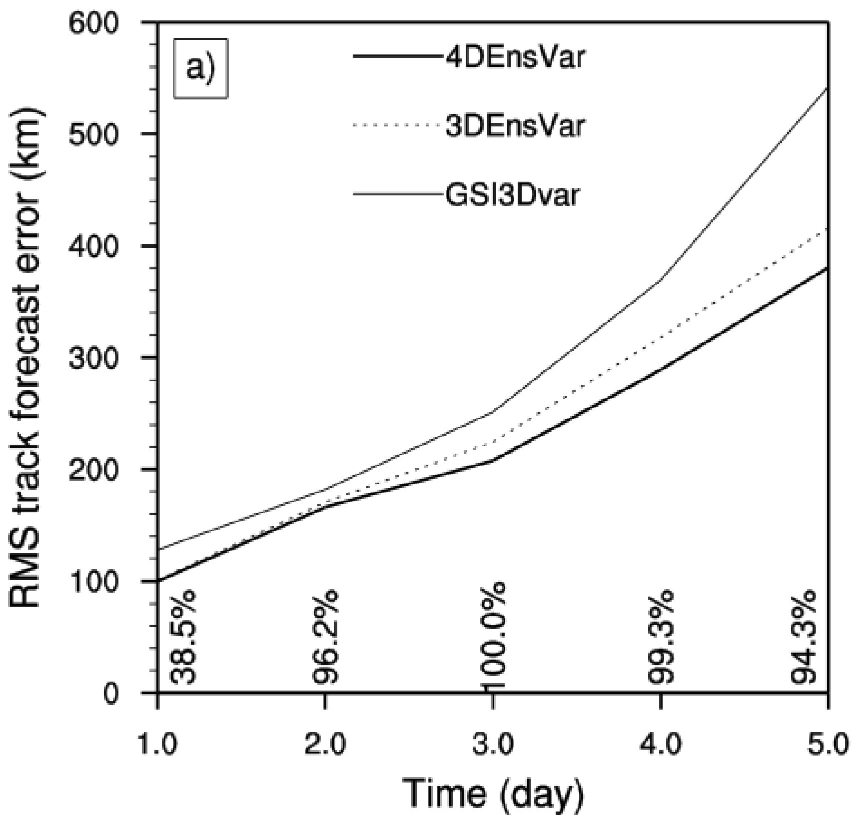
3DEnVar

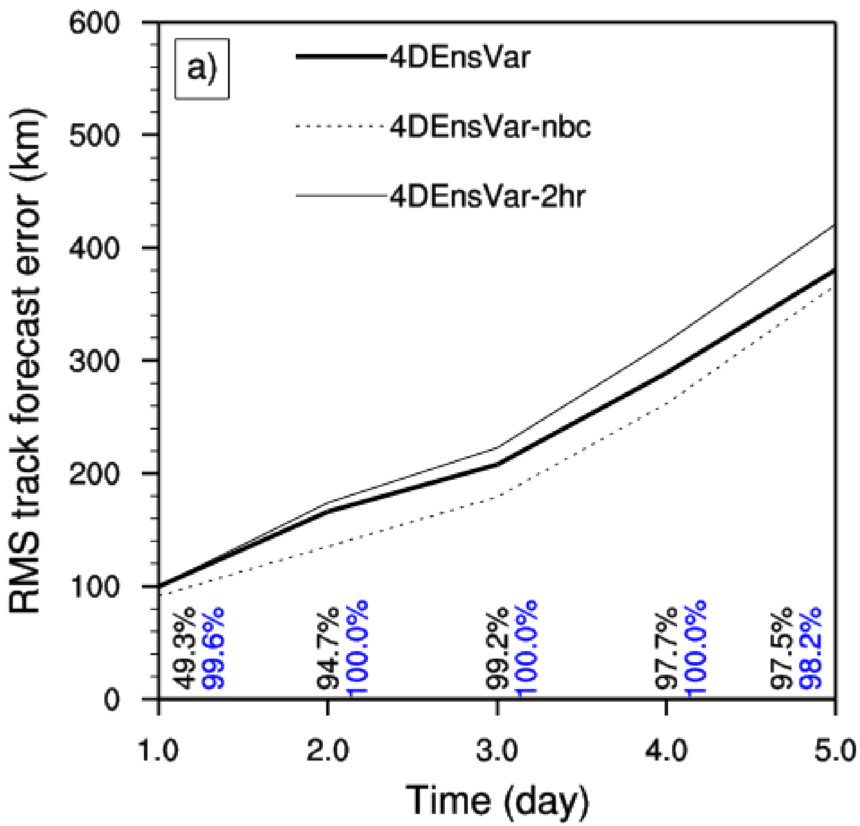


Hurricane Daniel 2010

Wang and Lei, 2014, MWR

Verification of hurricane track forecasts





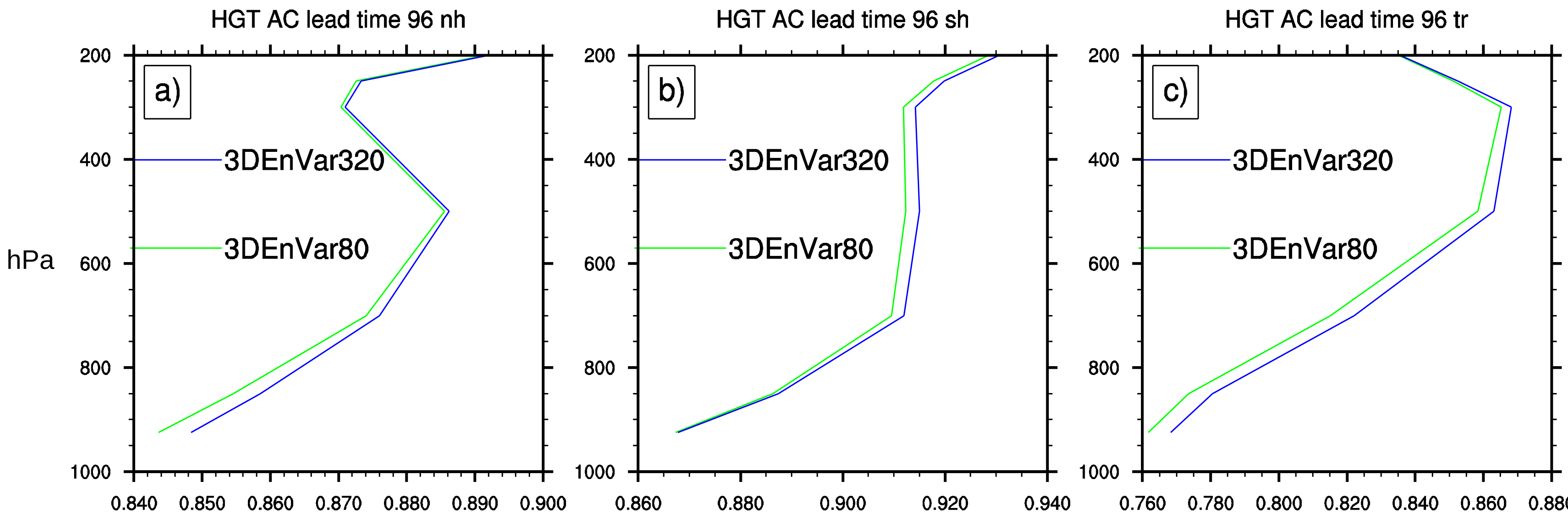
- 3DEnVar outperforms GSI3DVar.
- 4DEnVar is more accurate than 3DEnVar after the 1-day forecast lead time.
- Negative impact if using less number of time levels of ensemble perturbations.
- Negative impact of TLNMC on TC track forecasts. (IAU is implemented as an alternative to TLNMC in 4DEnVar under NCEP pre-implementation test)

Issues in ensemble background error covariance

- Sampling errors due to the use of limited ensemble members.
  - State of the art method e.g. covariance localization can cure some of them.
  - Increase ensemble size

- Mis-representation or under-representation of model errors in the ensemble. Model errors can be due to
  - Not enough resolution
  - Approximations in model numerics
  - Errors in physics parameterizations
  - Systematic (bias) or stochastic model errors

Motivation of 4DEnVar with time lagged ensembles: Impact of increasing ensemble size in GSI hybrid



- Reduced resolution experiments show quadrupling ensemble size improved analysis and subsequent forecast.
- Suggest great potential of using a larger number of ensemble members (beyond 80 members currently used in operations) to further improve the performance of GFS 4DEnVar

Lei and Wang 2015

Motivation of 4DEnVar with time lagged ensemble

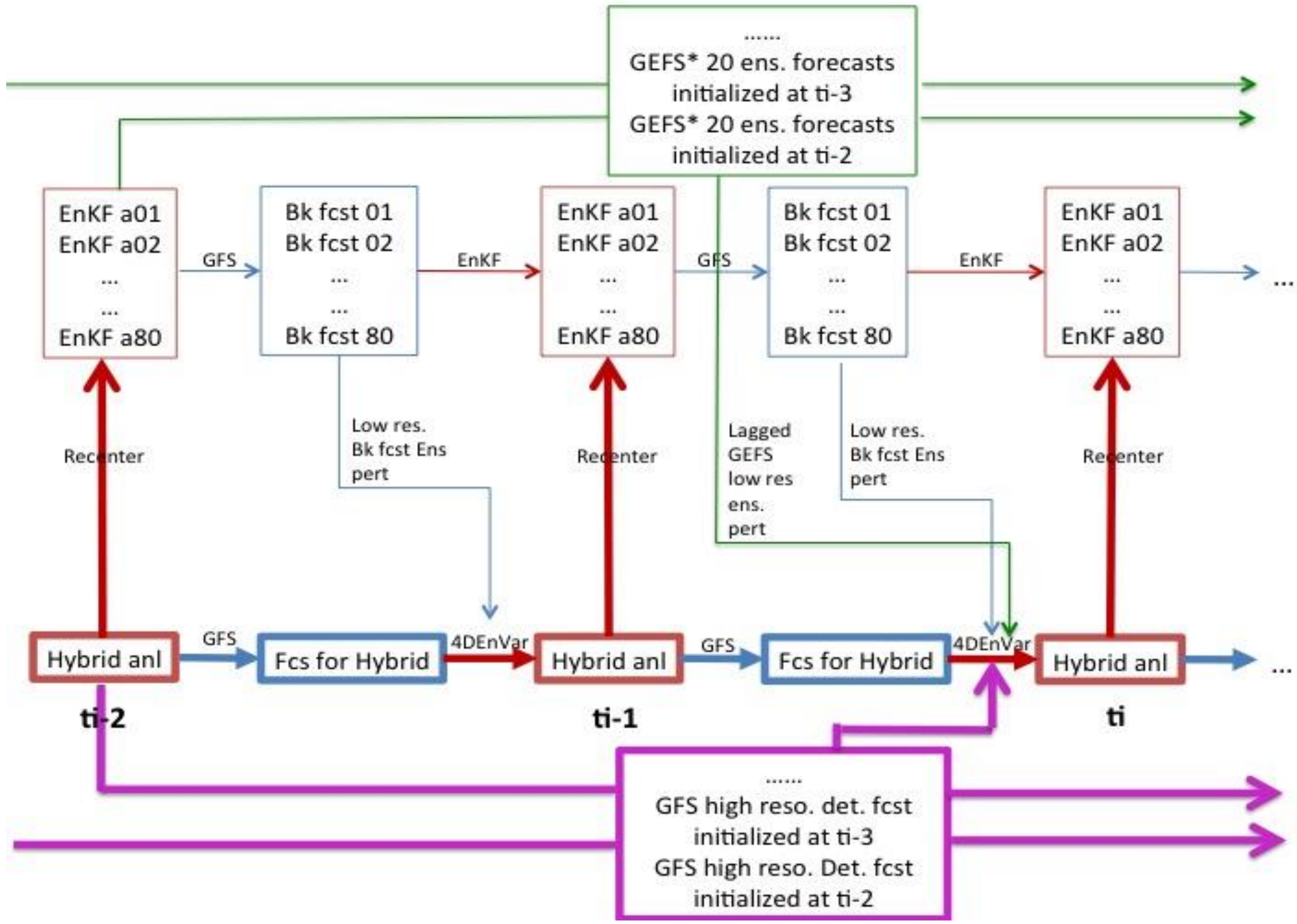
- One direct method to cure sampling error is to increase the number of ensemble members in the EnKF component. Significant increase of computational cost is expected.

- An alternative method to effectively increase the number of the ensemble members to be used for 4DEnVar while incurring minimum computational costs is proposed.

➢ Time lagged sampling approach
➢ Take advantage of lagged GEFS ensemble and lagged GFS control forecasts which are freely available for DA

- Expanding 4DEnVar with lagged ensemble alleviates sampling errors and help to sample model errors

GSI 4DEnVar ingesting time lagged ensemble



Primary objectives

- The primary goal of the project is to improve the GFS 4DEnVar hybrid data assimilation and subsequent global and hurricane forecasts through exploring cost-effective methods to increase the size of the ensemble ingested by 4DEnVar. Methods to optimally use lagged forecasts will be explored.

- Extend the infrastructure of the GSI-based 4DEnVar system for GFS to ingest lagged forecasts;
- Investigate the optimal way to use the lagged reduced resolution GEFS ensemble in 4DEnVar;
- Investigate the optimal way to use lagged full resolution GFS forecasts in 4DEnVar and the optimal way to combine lagged forecasts at different resolutions including lagged full resolution GFS forecasts and lagged reduced resolution GEFS forecasts in 4DEnVar;
- Systematically explore the impact of using lagged forecasts to increase ensemble size in 4DEnVar on the analyses and subsequent global and hurricane predictions;
- Working with EMC to transition the development and research on lagged forecasts 4DEnVar to NCEP operational global 4DEnVar hybrid DA system if results warrant.

Optimal use of GEFS lagged ensemble in GSI 4DEnVar

- lagged ensembles initialized at earlier times may not have equal skills to represent the background forecast errors at *ti.*

- Methods to optimally combine the lagged GEFS ensemble perturbations with the regular 80-member background ensemble perturbations will be explored.

- Equal weighting
- Skill weighting
- Spatially dependent weighting

Optimal use of multi resolution lagged ensemble in GSI 4DEnVar

- lagged ensembles include GEFS ensemble forecasts at the reduced resolution and the full resolution GFS forecasts.

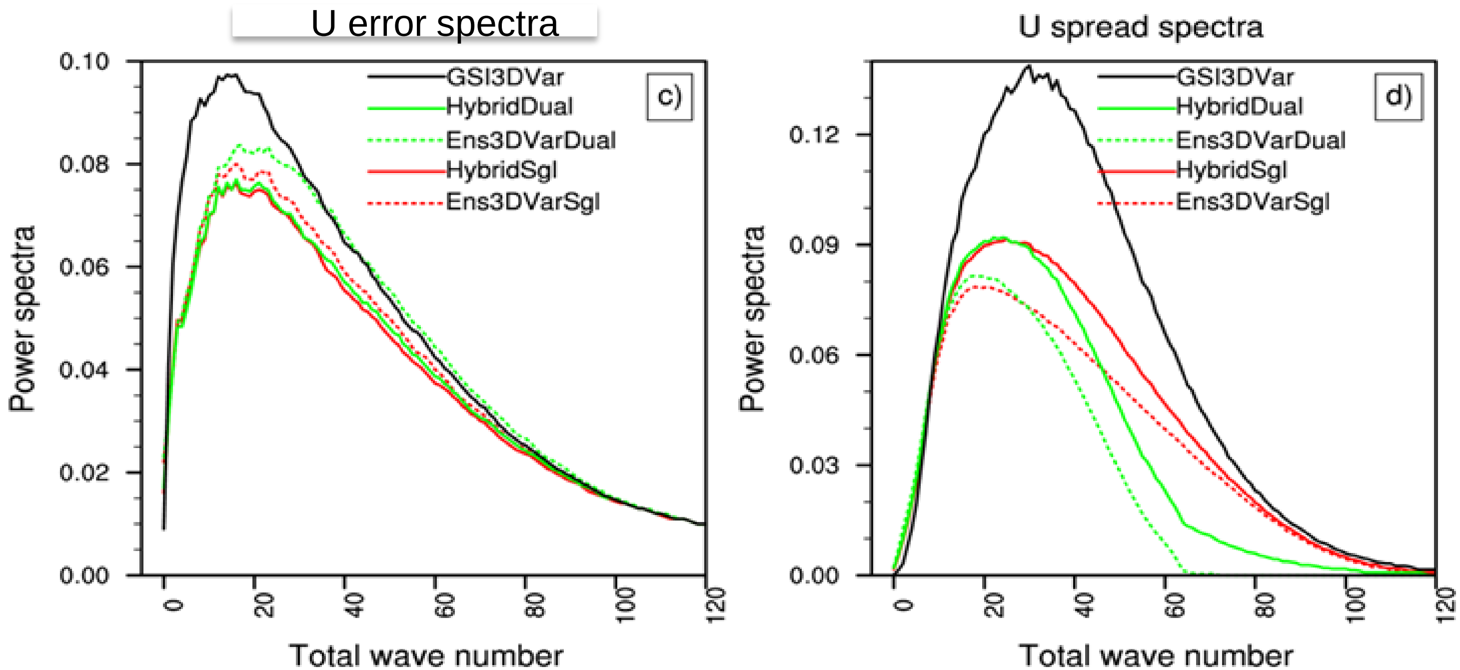
80 EnKF members
Lagged GEFS members
Lagged GFS deterministic forecasts

Reduced resolution

Full resolution

- Optimally using the lagged ensembles with members at different resolutions therefore involves exploring methods to optimally combine perturbations resolving different ranges of scales.

Optimal use of multi resolution lagged ensemble in GSI 4DEnVar

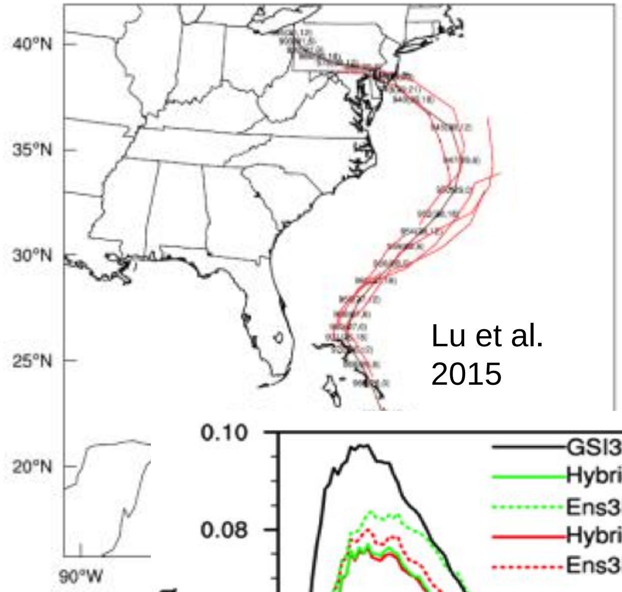


- Reduced resolution ensemble spread underestimate errors at large wave number
- Scale dependent weight needed for ensemble resolving different scales

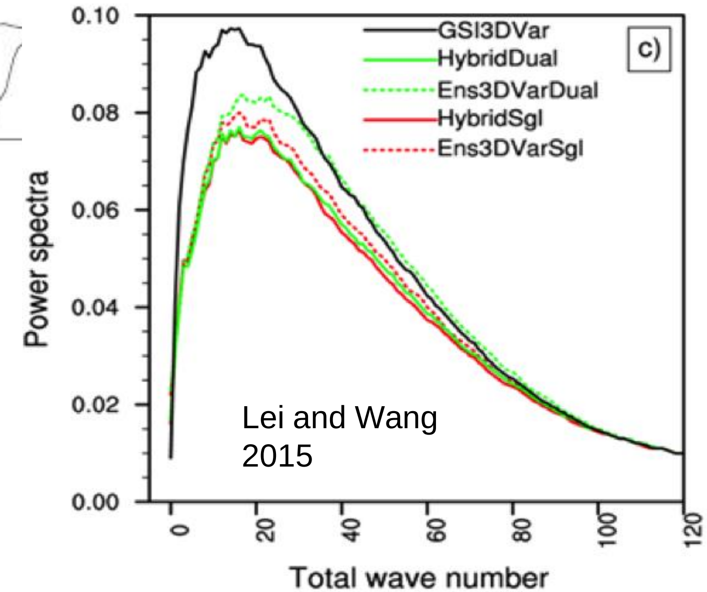
Lei and Wang 2015

Experiments and Verification metrics

- Use NCEP pre-implementation dual resolution 4DEnVar version to conduct DA cycling experiments for both winter and summer months
- Standard state space verifications for single deterministic and ensemble forecasts
- Spectral space verification
- Verification of hurricane track and intensity forecasts and compare analyzed hurricane structure and synoptic environment



Lu et al. 2015



Lei and Wang 2015

Address NGGPS goals

- Support development of NGGPS by improving data assimilation techniques
- Integrate NGGPS' data assimilation with ensemble prediction
- Direct benefit operational global and hurricane forecasts
- Method can be applied to other models of choice

References

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. S. 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 (4DEnsVar) data assimilation: formulation and single resolution experiments with real data for NCEP Global Forecast System. *Mon. Wea. Rev*., 142, 3303-3325.

Lei, T. and X. Wang, 2015: A comparison of GSI-based hybrid data assimilation for NCEP GFS at single and dual resolutions. *Mon. Wea. Rev*., to be submitted.

Johnson, A., X. Wang, J. Carely, L. Wicker and C. Karstens, 2015: A Comparison of Multi-scale GSI-based EnKF and 3DVar Data Assimilation using Radar and Conventional Observations for Mid-latitude Convective-scale Precipitation Forecasts. *Mon. Wea. Rev.* , in press.

Lu, X., X. Wang, Y. Li, M. Tong, X. Ma and H. Winterbottom, 2015: GSI-based ensemble-variational hybrid data assimilation for HWRF using airborne radar observations for hurricane initialization and prediction. *Q. J. R. Meteo. Soc., t*o be submitted.

Wang, Y., X. Wang, 2015: Assimilation of radar reflectivity data in GSI-based hybrid data assimilation System using various control variables for the 8 May 2003 Oklahoma City tornadic supercell storm, *Mon. Wea. Rev* , to be submitted.

Whitaker, J.S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev*., 136, 463–482.

Kleist, D. T. and K. Ide, 2015: An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part II: 4DEnVar and Hybrid Variants. *Mon. Wea. Rev.*, **143**, 452–470.