GSI/ETKF Regional Hybrid Data Assimilation

Arthur P. Mizzi (mizzi@ucar.edu)

NCAR/MMM

2011 GSI Tutorial
June 29 – July 1, 2011
NCAR – FL2
Boulder, CO

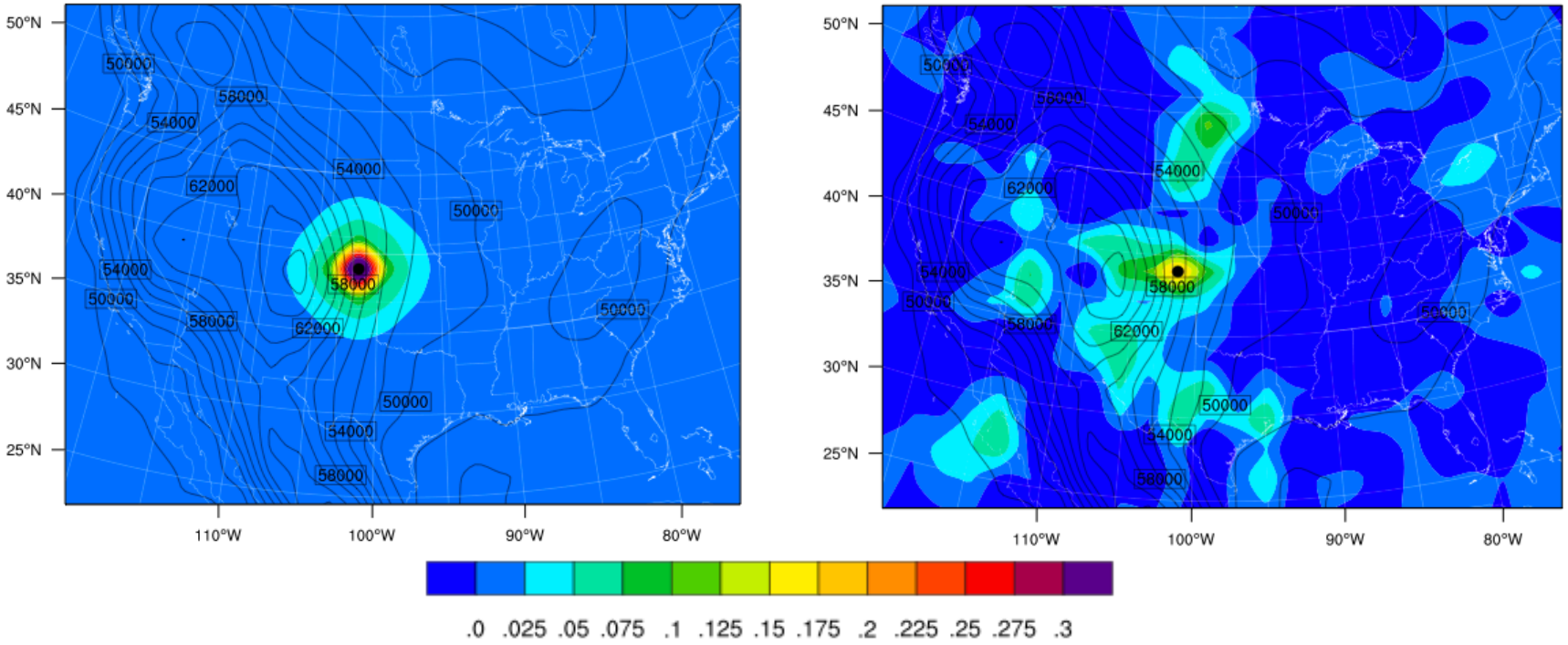
Why Use Hybrid Data Assimilation?

- Accurate error statistics are the key to improving variational data assimilation.
- Background errors are generally spatially isotropic and temporally invariant.
- Ensemble prediction systems (EPSs) provide flow-dependent background errors but large ensembles are computationally expensive.
- The hybrid provides benefit from whatever flow-dependent information a small ensemble can provide.
- The hybrid has been shown to produce improved forecasts at longer lead times.

Single Observation T Increments

GSIDA

GSI/ETKF Hybrid



B=1.0, C=8, N=20,
ΔT=1K, σ_T=1K

B=0.0, C=8, N=20,
ΔT=1K, σ_T=1K

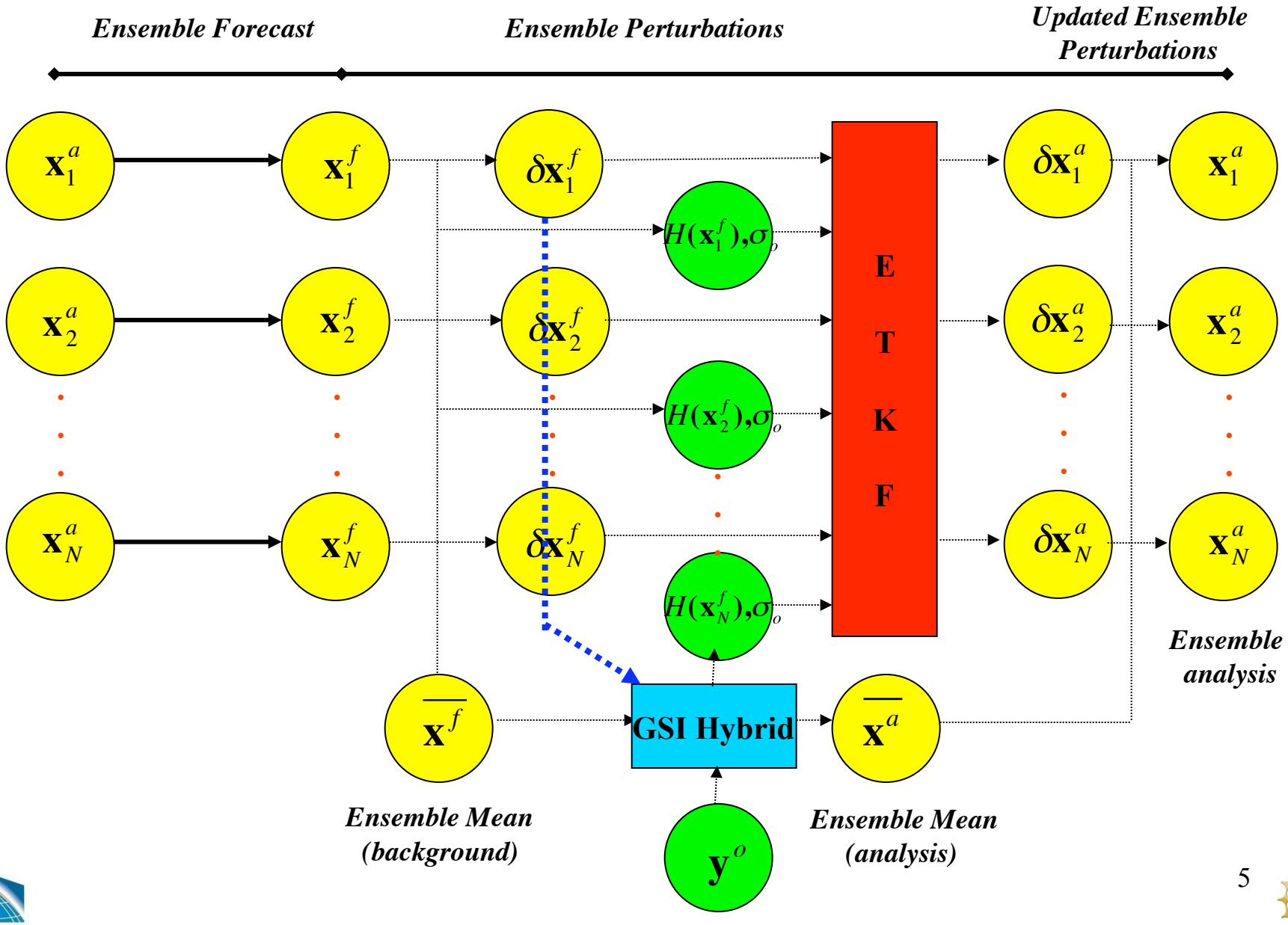
NESL

NSF

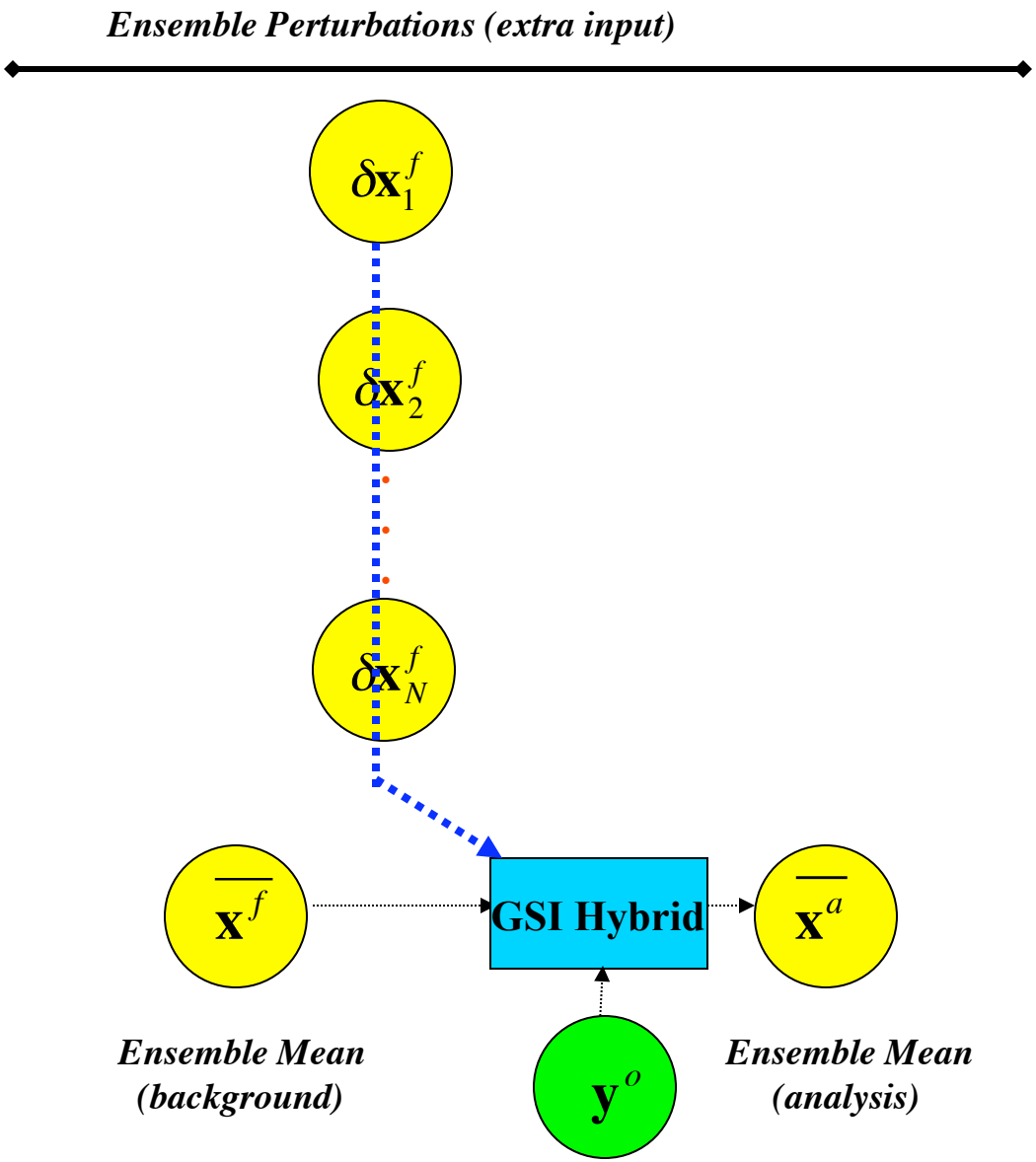
Steps for GSI Hybrid Data Assimilation

1. Generate initial ensemble.
2. Calculate ensemble mean and variance.
3. Update ensemble mean with GSI regional hybrid.
4. Update ensemble perturbations using ETKF, LETKF, EnKF, Inverse Hessian, PO, or BV.
5. Obtain total fields by adding updated mean and perturbation for each ensemble member.
6. Update the boundary conditions.
7. Run cycle time forecasts for each ensemble member.
8. Go to step 2 and repeat process with the ensemble forecasts from step 7.

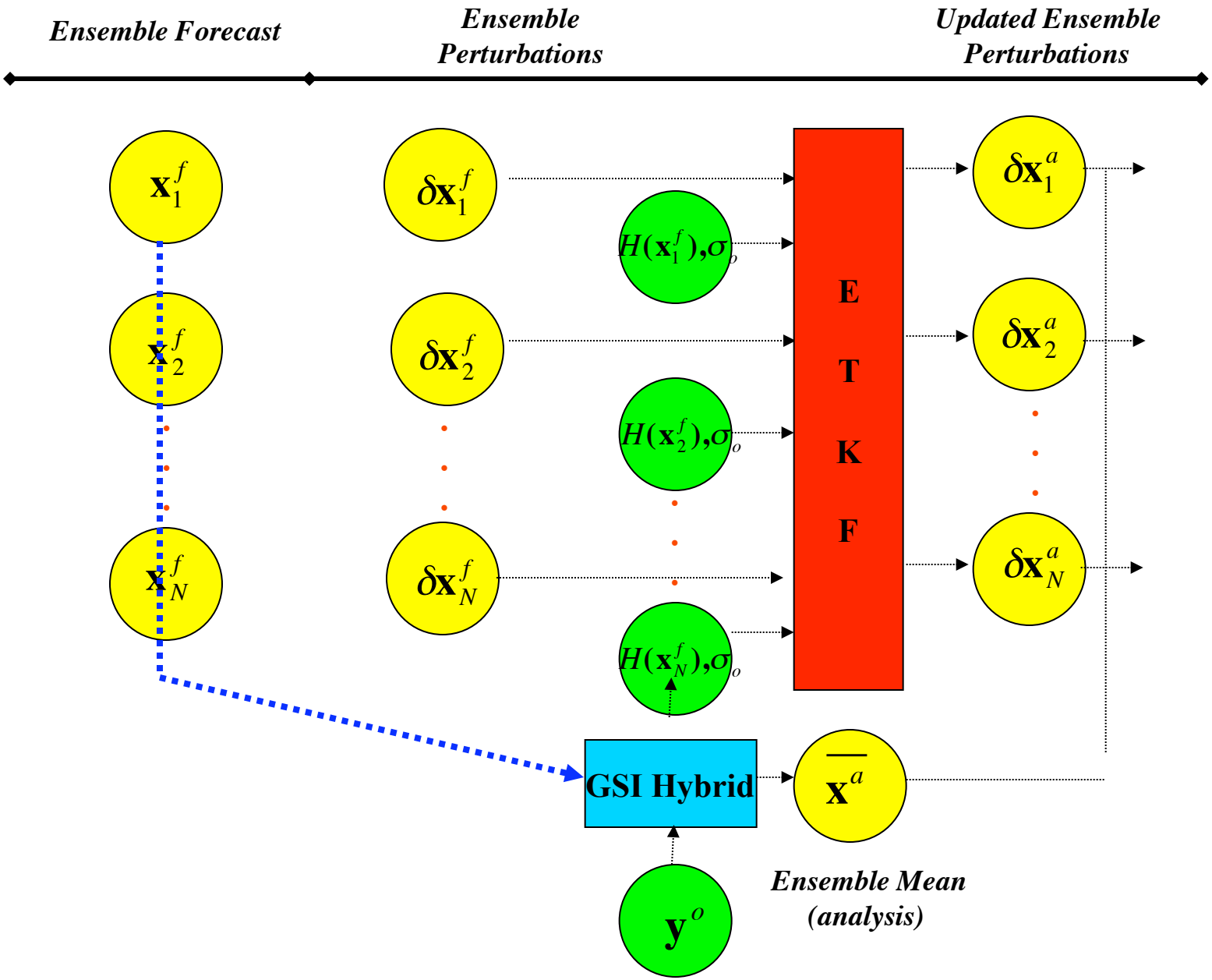
GSI/ETKF Regional Hybrid Cycling



GSI Hybrid DA: Variational Part



GSI Hybrid DA: Perturbation Part



GSI Hybrid Cost Function

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

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

Extra term associated with extended control variable

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

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

Extra increment associated with ensemble

B 3DVAR static covariance; **R** observation error covariance; K ensemble size;
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;
H linearized observation operator; β_1 weighting coefficient for static covariance;
β_2 weighting coefficient for ensemble covariance; $\boldsymbol{\alpha}$ extended control variable.

NESL

Ensemble Perturbation Generation

- EnKF
  - Computationally expensive
  - Undersampling
  - Requires inflation
  - Spurious correlations, requires localization
- ETKF
  - Computationally fast
  - Undersampling
  - Rank deficiency
  - Requires inflation
  - Spurious correlations, not easily localized

NESL

NSF

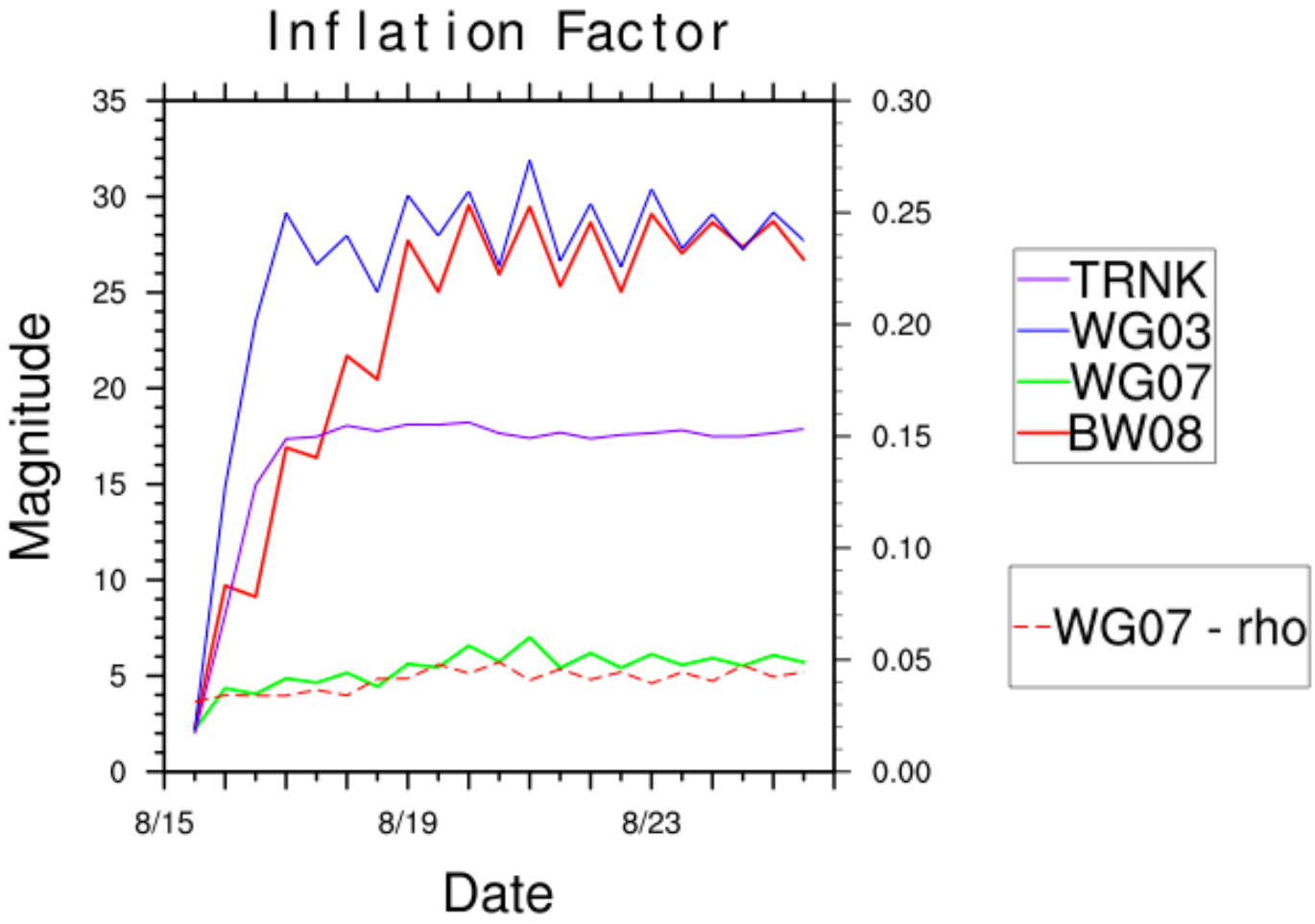
Ensemble Perturbation Generation

- LETKF
  - Computationally fast
  - Undersampling
  - Reduced rank deficiency
  - Localization eliminates spurious correlations
- Inverse Hessian methods
  - Under investigation

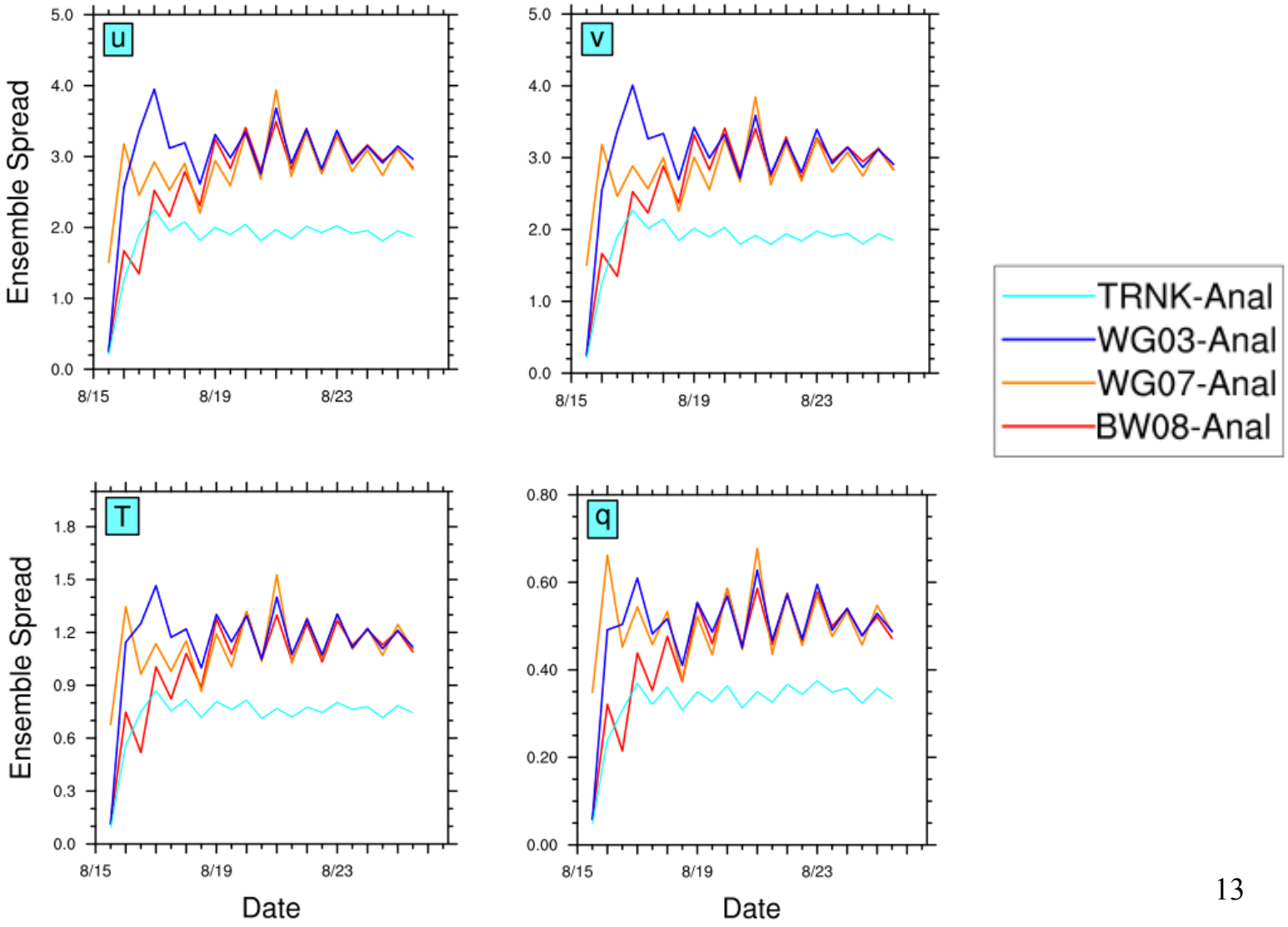
- **PROBLEM**: Under-sampling of forecast distribution results in underestimation of ensemble spread – need inflation.

GSI/ETKF Regional Hybrid Example Mean and Spread

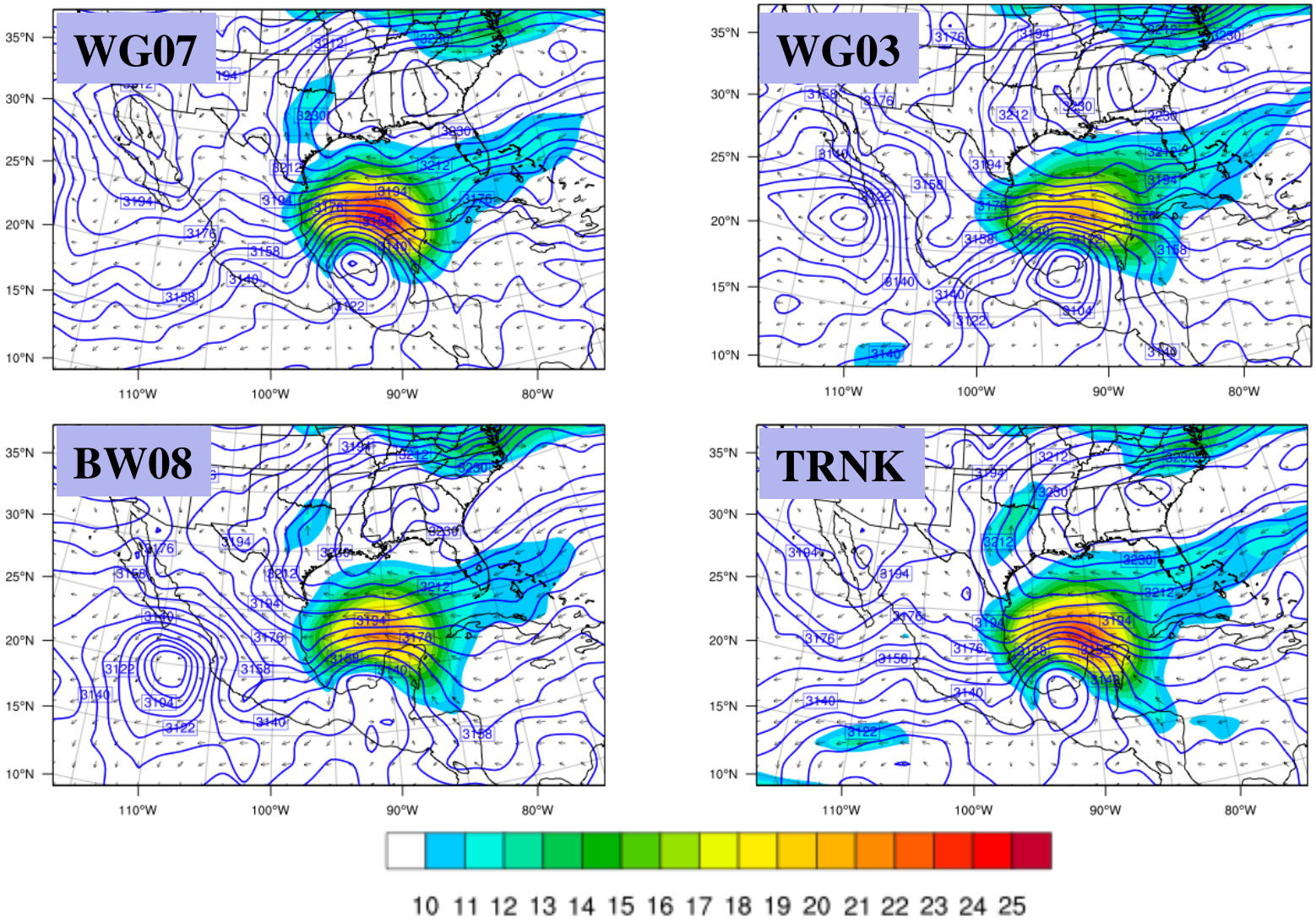
- Ensemble size: 20
- Study Period: Aug. 15 – Aug. 25, 2007 (Hurricane Dean Test Case).
- Cycle time: 12 hr.
- Domain: Same as single observation experiments.
- Observations: GTS conventional observations.
- ICs/BCs: GFS forecasts.
- Ensemble ICs/BCs: Produced by adding spatially correlated Gaussian noise to GFS forecasts.


Inflation Factor
Magnitude
35
30
25
20
15
10
5
0
0.30
0.25
0.20
0.15
0.10
0.05
0.00
8/15
8/19
8/23
Date
TRNK
WG03
WG07
BW08
WG07 - rho

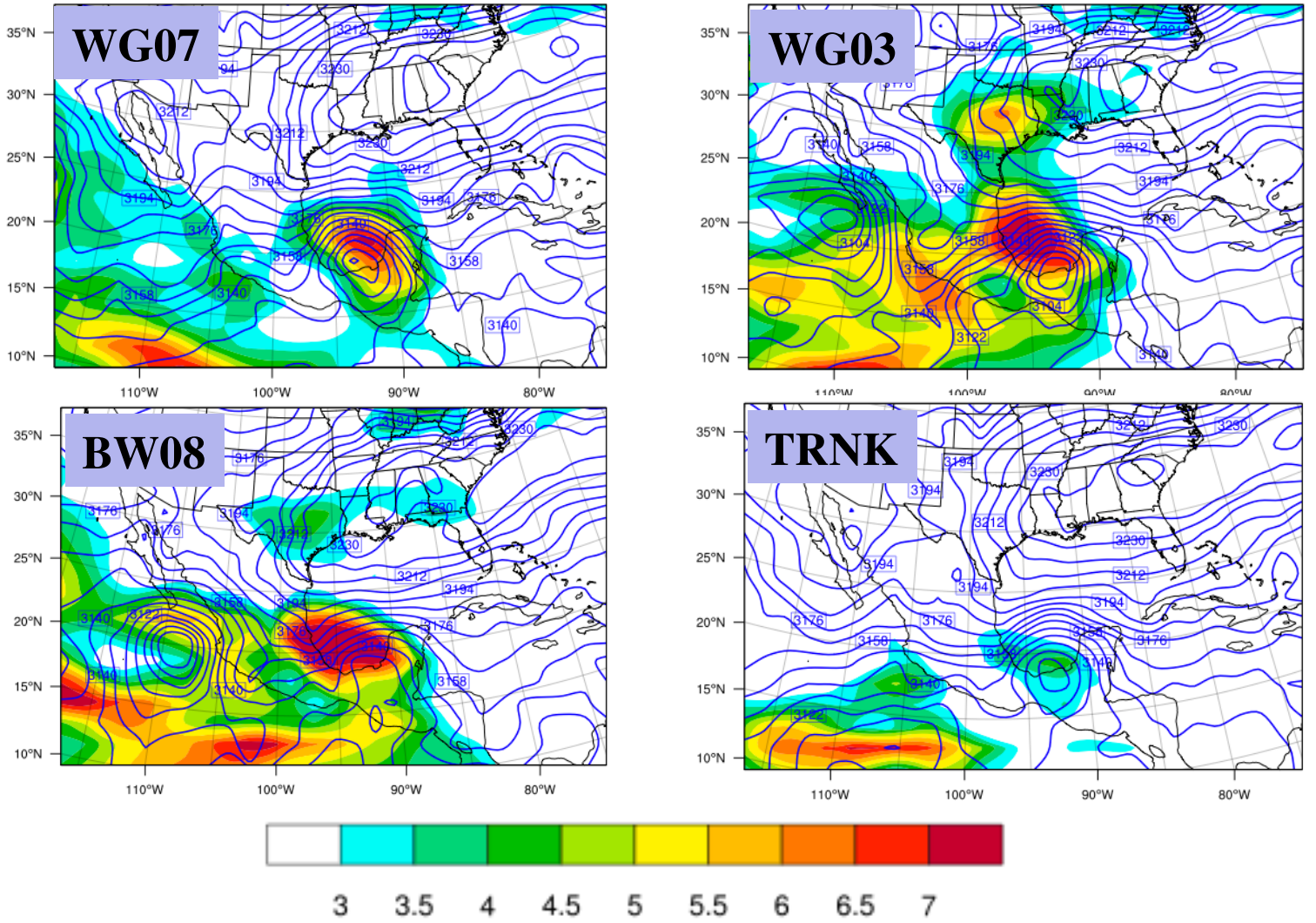
Posterior Ensemble Spread Time Series



Ensemble Mean Wind Speed (m/s)
Aug 22, 2007 00Z 700 hPa



Ensemble Spread: u-wind (m/s)
Aug 22, 2007 00Z 700 hPa



GSI Hybrid Practice Session

- **GOAL:**
  - Run GSI hybrid for a suite of single observation sensitivity experiments and plot the increments.
- **SCRIPTS:**
  - da_run_gsida_single_obs.ksh – produces the hybrid single observation analyses.
  - plot_gsida_single_hori.ncl – produces horizontal contour plots of the single observation hybrid increment.
  - plot_gsida_single_vert.ncl – produces vertical cross-section contour plots of the single observation hybrid increment.

Ensemble Members for GSI:

- To run the GSI regional hybrid, you must copy the ensemble forecast members to the GSI run directory – *gsiprd*.

- The format for the ensemble member names in *gsiprd* is as follows:

wrf_mass_forecast.eXXX

where *XXX* ranges from 001 to 999.

- For this tutorial, the Run Script will copy the ensemble members to *gsiprd* automatically. The GSI regional hybrid works only with WRF-ARW ensembles.

GSI First Guess and Other Fixed Files:

- GSI operates on a first guess field. For an ensemble cycling experiment, GSI operates on the ensemble mean.

- For this tutorial, the Run Script automatically calculates the ensemble mean and copies it to the GSI run directory as the first guess field.

- The Run Scrip also copies all other fixed files, such as the static background error covariances, needed to run GSI to the GSI run directory.

GSI Hybrid Namelist Parameters:

1. **Namelist Identifer:**

   HYBRID_ENSEMBLE

   Namelist Parameters:

   l_hyb_ens – (logical) set to .TRUE. to run the hybrid.

   n_ens – (integer) the number of ensemble members.

   uv_hyb_ens – (logical) set to .TRUE. if ensemble wind fields are *u* and *v*, otherwise they are ψ and χ.

   beta1_inv – (float) the weighting between the variational and ensemble increment, set to 1.0 for variational increment, set to 0.0 for ensemble increment, set to (0.0 – 1.0) for the hybrid increment.

   s_ens_h – (float) the horizontal localization length scale (km).

   s_ens_v – (integer) the vertical localization length scale (vertical grid points).

GSI Hybrid Parameters cont.:

generate_ens – (logical) set to .TRUE. generate the ensemble members internally based on the existing background error.

aniso_a_en – (logical) set to .TRUE. to use anisotropic localization of the hybrid ensemble control variable.

nlon_ens – (integer) the east-west dimension for the ensemble members.

nlat_ens – (integer) the north-south dimension for the ensemble members.

jcap_ens – (integer) the global spectral truncation.

jcap_ens_test– (integer) the global spectral truncation for test the dual resolution option.

GSI Single Observation Namelist Parameters:

1. **Namelist Identifer:**

   SETUP

   Namelist Parameter:

   oneobtest – (logical) set to .TRUE. for single observation test.

2. **Namelist Identifer:**

   SINGLEOB_TEST

   Namelist Parameters:

   maginnov – (float) magnitude of the single observation innovation.

   magoberr – (float) magnitude of the single observation error.

GSI Single Observation Namelist Parameters cont.:

Namelist Parameter cont.:

oblat – (float) latitude of the single observation (degrees).

oblon – (float) longitude of the single observation (degrees).

obpres – (float) pressure level of the single observation (hPa).

obdattim – (integer) date and time of the single observation in the YYYYMMDDHH format.

obhourset – (integer) time difference between the observation and analysis times.

pctswitch – (logical) set to .TRUE. if innovation and observation error are relative (%) to the background value (ozone only).

GSI Single Observation Run Script:

SET RUN OPTIONS:

RUN_ENS_MEAN – set to "true" to calculate ensemble mean.

RUN_HYBRID – set to ".true." to run GSI in hybrid mode.

SET EXPERIMENT OPTIONS:

DATE – set to date and time in YYYYMMDDHH format, sets the *obsdate* parameter (for this tutorial the range is 2007081512 to 2007082512).

NUM_MEMBERS – number of ensemble members, sets the *l_ens_hyb* parameter (for this tutorial the range is 2 to 80).

HOR_SCALE – the hybrid horizontal localization length scale (km), sets the *s_ens_h* parameter (for this tutorial suggested values ranges between (200 and 3000 km).

GSI Run Script cont:

SET EXPERIMENT OPTIONS:

VER_SCALE – the hybrid vertical localization length scale (grid points), sets the *s_ens_v* parameter (for this tutorial suggested values ranges between (2 and 30).

GSI_BETA – the weighting between the variational and ensemble increment, sets the *beta1_inv* parameter (for this tutorial this ranges between 0.0 and 1.0).

SET DIRECTORY PARAMETERS:

ENSEMBLE_MEM_DIR – path to the directory containing the ensemble data (for this tutorial do not change).

TRUNK_DIR – path to the directory containing the GSI trunk code (for this tutorial do not change).

WORK_DIR – path to the runtime and output directory (for this tutorial set to _____).

GSI Run Script cont:

GSI SINGLE OB DATA:

NL_ONEOBTEST – set to ".true." to run single observation test, set the *oneobtest* parameter.

NL_ONEOB_TYPE – the single observation type, sets the *oneobtype_type* parameter (possible values *u, v, t,* and *q*, for this tutorial suggest using "t" for consistency with the plotting script).

NL_MAGINNO – the single observation innovation magnitude, sets the *maginno* parameter (for this tutorial suggest using a value near 1.0 for consistency with the plotting script).

NL_MAGOERR – the single observation error, sets the *magoberr* parameter (for this tutorial suggest using a value near 1.0 for consistency with the plotting script).

GSI Run Script cont:

GSI SINGLE OB DATA cont.:

NL_OBLAT – the single observation latitude, sets the *oblat* parameter (for this tutorial use 39.0° for consistency with the plotting script).

NL_OBLON – the single observation longitude, sets the *oblon* parameter (for this tutorial use 258.4° for consistency with the plotting script).

NL_OBPRES – the single observation pressure level, sets the *obpres* parameter (for this tutorial use 507.1 hPa for consistency with the plotting script).

Single Observation Plotting Script:

DEFINE PARAMETERS:

date – (string) the single observation date (for this tutorial set to same value as *DATE* in the Run Script).

year – (string) the single observation year (for this tutorial set to "2007").

month – (string) the single observation month (for this tutorial set to "08").

day – (string) the single observation day (for this tutorial set according to the value of *DATE* in the Run Script).

hour – (string) the single observation hour (for this tutorial set according to the value of *DATE* in the Run Script).

GSI Plotting Script cont.:

DEFINE INPUT FILES:

file_type – (string) GSI output file template name (for this tutorial use "wrfinput_d01")

fg_directory – (string) the path to the first guess directory (for this tutorial set to the Run Script *WORK_DIR/gsida/gsiprd*).

an_directory – (string) the path to the GSI analysis directory (for this tutorial set to the Run Script *WORK_DIR/gsida/gsiprd*).

vr_directory – (string) the path to the ensemble variance directory (for this tutorial set the the Run Script *WORK_DIR/ens_mean/working*).

References

Anderson, J.L. , 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884-2903.

Anderson, J.L. , 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634-642.

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.

Bowler, N. E., A. Arribas, S.E. Beare, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. R. Meteor. Soc.*, **134**, 703–722.

Bueher, M., P.L. Houtekamer, C. Charette, H.L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single observation experiments. *Wea. Forecasting*, **138**, 1550-1566.

References cont.

Bueher, M., P.L. Houtekamer, C. Charette, H.L. Mitchell, and B. He, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: One-month experiments with real observations. *Wea. Forecasting*, **138**, 1567-1586.

Etherson, B.J. and C.H. Bishop, 2004: Resilence of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.*, **132**, 1065-1080.

Hamill, T.M. and C. Snyder, 2000: A hybrid 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.

Houtekamer, P.L. and H.L. Michell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

Lorenc, A.C., 2003: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-VAR. *Quart. J. R. Meteor. Soc.*, **129** 3183-3203.

References cont.

Ott, E., B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D.J. Patil, and J.A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415-428.

Wang , X., 2010: Incorporating ensemble covariance in the Gridpoint Statistical Interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990-2995.
Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116-5131.

Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008b: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiment. *Mon. Wea. Rev.*, **136**, 5132-5147.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140-1158.

NESL

NSF

References cont.

Wang, X., T.M. Hamill, J.S. Whitaker, and C.H. Bishop, 2007: A comparison of hybrid ensemble transform Kalman filter-optimum interpolation and ensemble square-root filter analysis schemes. *Mon. Wea. Rev.*, **135**, 1005-1976.

Wang, X., C. Snyder, and T.M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble-3DVAR hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222, 227.

Whitaker, J.S., and T.M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913-1924.

Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects., *Mon. Wea. Rev.*, **133**, 1710-1726.

Zupanski, M., I.M. Navron, and D. Zupanski, 2008: The maximum likelihod ensemble filter as a non-differentiable minimization algorithm. *Quart. J. R. Meteor. Soc.*, **134**, 1039-1050.