7th NOAA Testbed and Providing Grounds Workshop, 5-6 April, 2016, College Park, Maryland

Testing and Evaluation of Four-Dimensional Ensemble Variational Data Assimilation for Regional Weather Forecasts

Hui Shao[1,3], Ming Hu[2,3,4], Chunhua Zhou[1,3], Kathryn Newman[1,3], Xin Zhang[2,3,4], and Christina Holt[2,3,4]

[1] National Center for Atmospheric Research

[2] NOAA/ESRL Global Systems Division

[3] Developmental Testbed Center

[4]CU Cooperative Institute for Research in the Environmental Sciences

This project is sponsored by NOAA/OAR

Developmental Testbed Center

Outline

- Background
- Experiments and results
  - Global ensemble vs regional ensemble: ensemble representation
  - 6 hour cycling
  - First results with RAP system
- Summary and future plans

Methodology

- **Data assimilation**

  Assuming weather system under study is a random process, the solution of the data assimilation problem is the probability density function (PDF) of the system conditioned upon observations.

- **Variational (Var) data assimilation (DA)**

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

$\mathbf{x}$($\mathbf{x}_b$): analysis (background) state vector
$\mathbf{y}$: observation vector
B (R): background (observation) error covariance
H: observation (forward) operator for 3DVar (4DVar)

  - Analysis is achieved by finding the minimum $\mathbf{x}$ of the cost function, $J(\mathbf{x})$
  - Assuming background errors and observations are Gaussian, 3DVar and 4DVar solutions are the unbiased minimum variance estimates<=>the maximum likelihood estimates

- **EnVar**: a variational method using ensemble background covariances. The ensemble background covariance is estimated through an additional control variable added to the cost function – incorporate flow-dependent errors

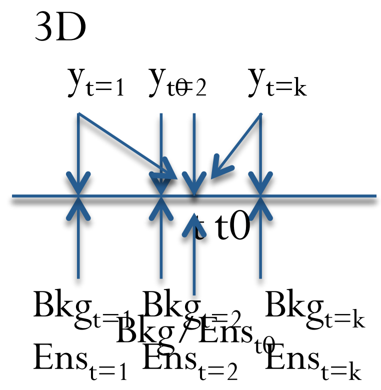
- **Hybrid**: a variational method using a combination of static and ensemble covariances – alleviate ensemble localization issues/allow for more efficient error estimation with necessarily small ensembles

Hybrid 4D EnVar DA

Static background error covariance

Incorporating ensemble background-error information through extended control variable

$$J(\mathbf{x}'_{\mathrm{f}},\boldsymbol{\alpha})=\beta_{\mathrm{f}}\frac{1}{2}(\mathbf{x}'_{\mathrm{f}})^{\mathrm{T}}\mathbf{B}_{\mathrm{f}}^{-1}(\mathbf{x}'_{\mathrm{f}})+\beta_{\mathrm{e}}\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{\alpha}^{n})^{\mathrm{T}}\mathbf{L}^{-1}(\boldsymbol{\alpha}^{n})+\frac{1}{2}\sum_{k=1}^{K}(\mathbf{H}_{k}\mathbf{x}'_{k}-\mathbf{y}'_{k})^{\mathrm{T}}\mathbf{R}_{k}^{-1}(\mathbf{H}_{k}\mathbf{x}'_{k}-\mathbf{y}'_{k})$$



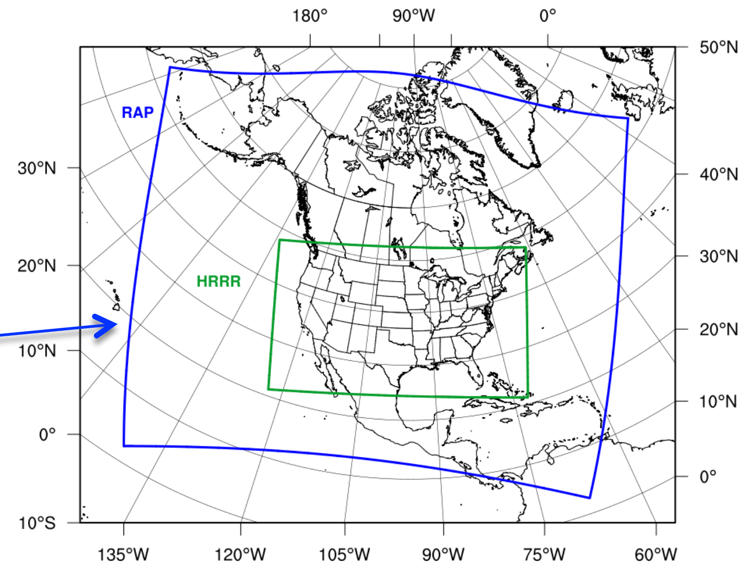
4D increment is prescribed through linear combination of 4D ensemble perturbations plus static contributions

$$\mathbf{x}'_{k}=\mathbf{x}'_{\mathrm{f}}+\sum_{n=1}^{N}\left(\boldsymbol{\alpha}^{n}\circ(\mathbf{x}_{\mathrm{e}})_{k}^{n}\right)$$

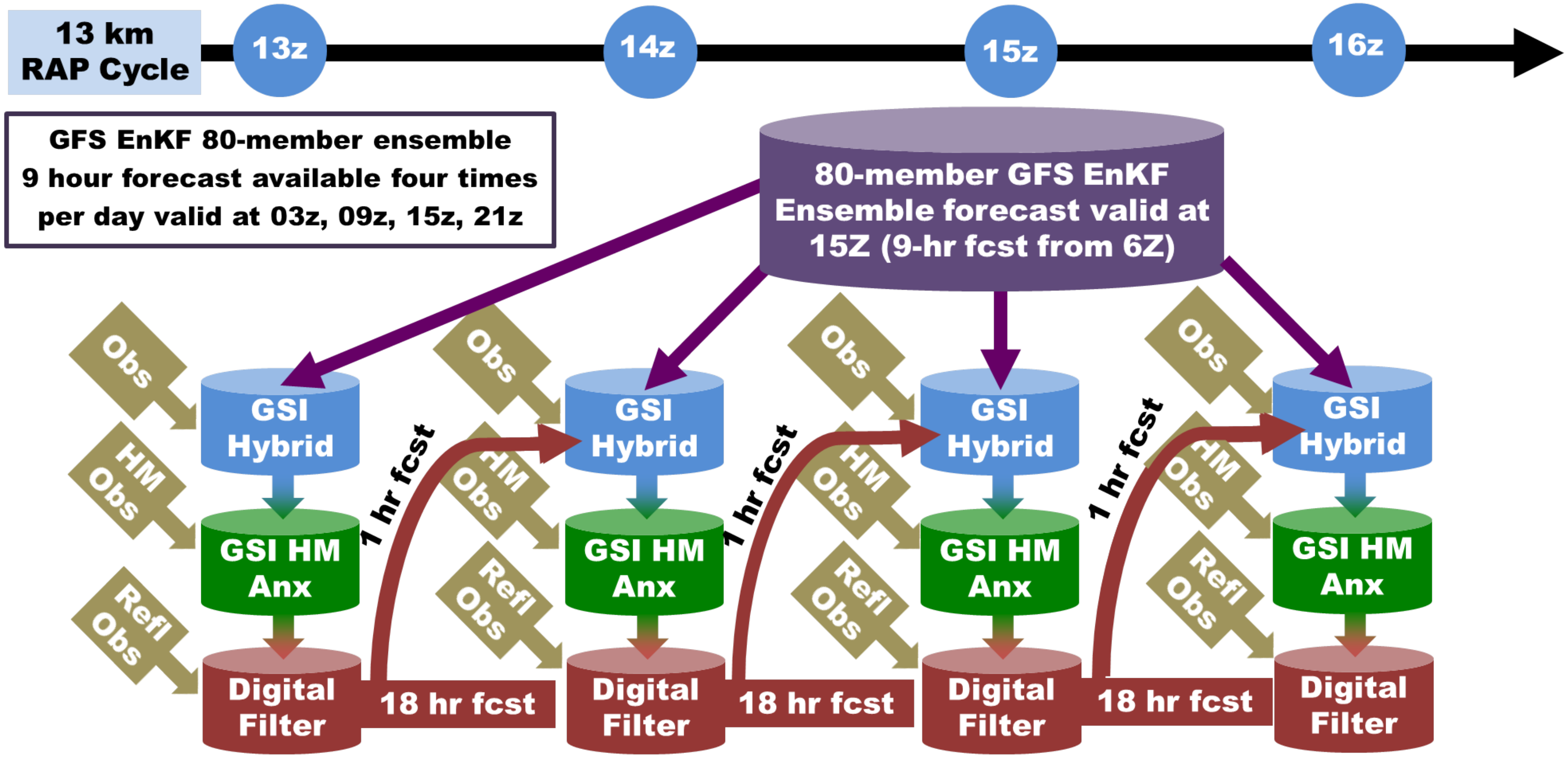
x': analysis increment vector
n: n-th ensemble member
k: k-th time bins
B_f: Static background error covariance
α: extended control variable
L: correlation matrix
H: foreword operator
y: observation vector
R: observation error covariance

Current Efforts

- NCEP Global Forecast System (GFS) hybrid 3D EnVar since 2012, using NOAA Gridpoint Statistical Interpolation (GSI) and Ensemble Kalman Filter (EnKF) systems
  - Shared observation operators
- NCEP plans to implement hybrid 4D EnVar for the upcoming GFS implementation
- This effort is applying the global hybrid 4D EnVar to regional NWP
  - NOAA Rapid Refresh (RAP) system:
    - Advanced Research WRF (ARW) dynamical core
    - 13km North American domain
    - Twice daily partial cycles initialized with GFS background
    - Hourly continuous cycled land-surface fields
  - High-Resolution RAP (HRRR)



RAP: Operational GSI 3D Hybrid DA system



- GSI hybrid 3D-VAR/GFS-ensemble data assimilation
- GSI non-variational cloud/precipitation hydrometeor (HM) analysis
- Diabatic Digital Filter Initialization (DDFI) using hourly radar reflectivity observations

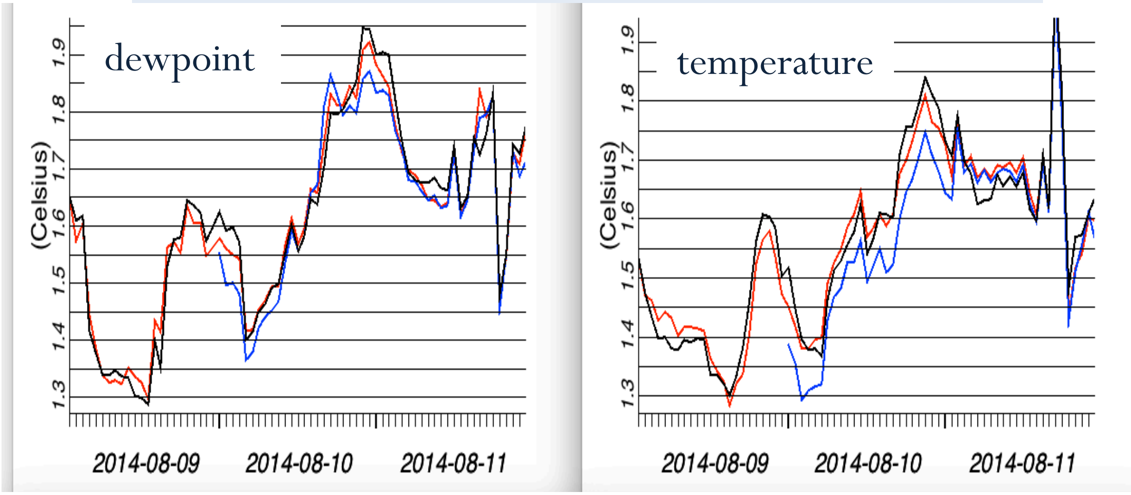
Ensemble Representativeness:

- Global or regional ensemble?

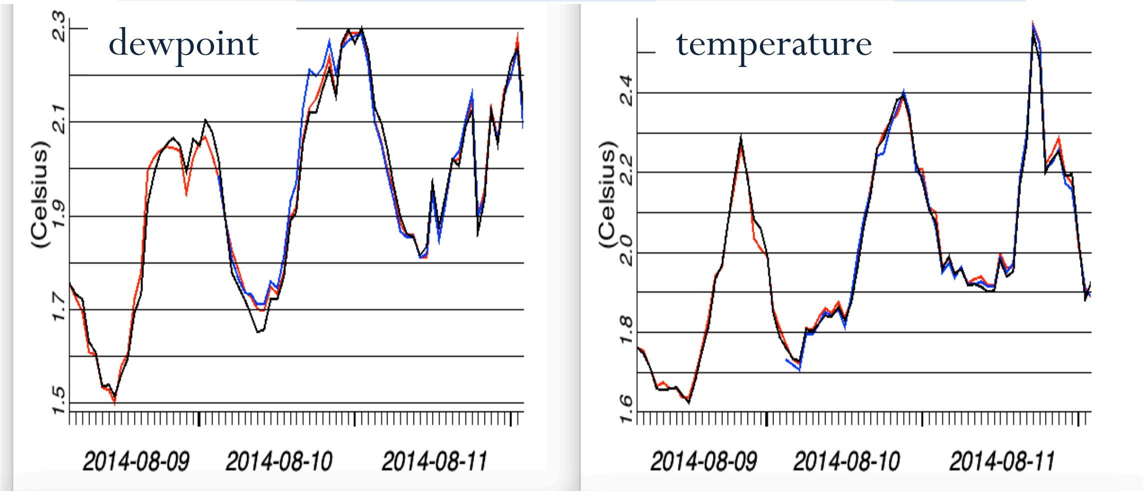
- GFS ensemble
  - Operational configuration
  - 30 km resolution
  - 80 members
  - Updated operationally through the NOAA Ensemble Kalman Filter (EnKF) DA system (recentered with GSI-hybrid analysis input)
- RAP ensemble
  - ARW forecasts initialized with the 30km GFS ensemble members
  - 13km resolution
  - 80 members

3D Hybrid with GFS Ensemble vs RAP Ensemble

Surface RMS time series for 0 hour forecast



Surface RMS time series for 6 hour forecast



GFS ensemble

RAP ensemble

RAP ensemble with increased inflation

Ensemble forecasts show minimal preference, therefore GFS ensemble is selected for the following EnVar tests

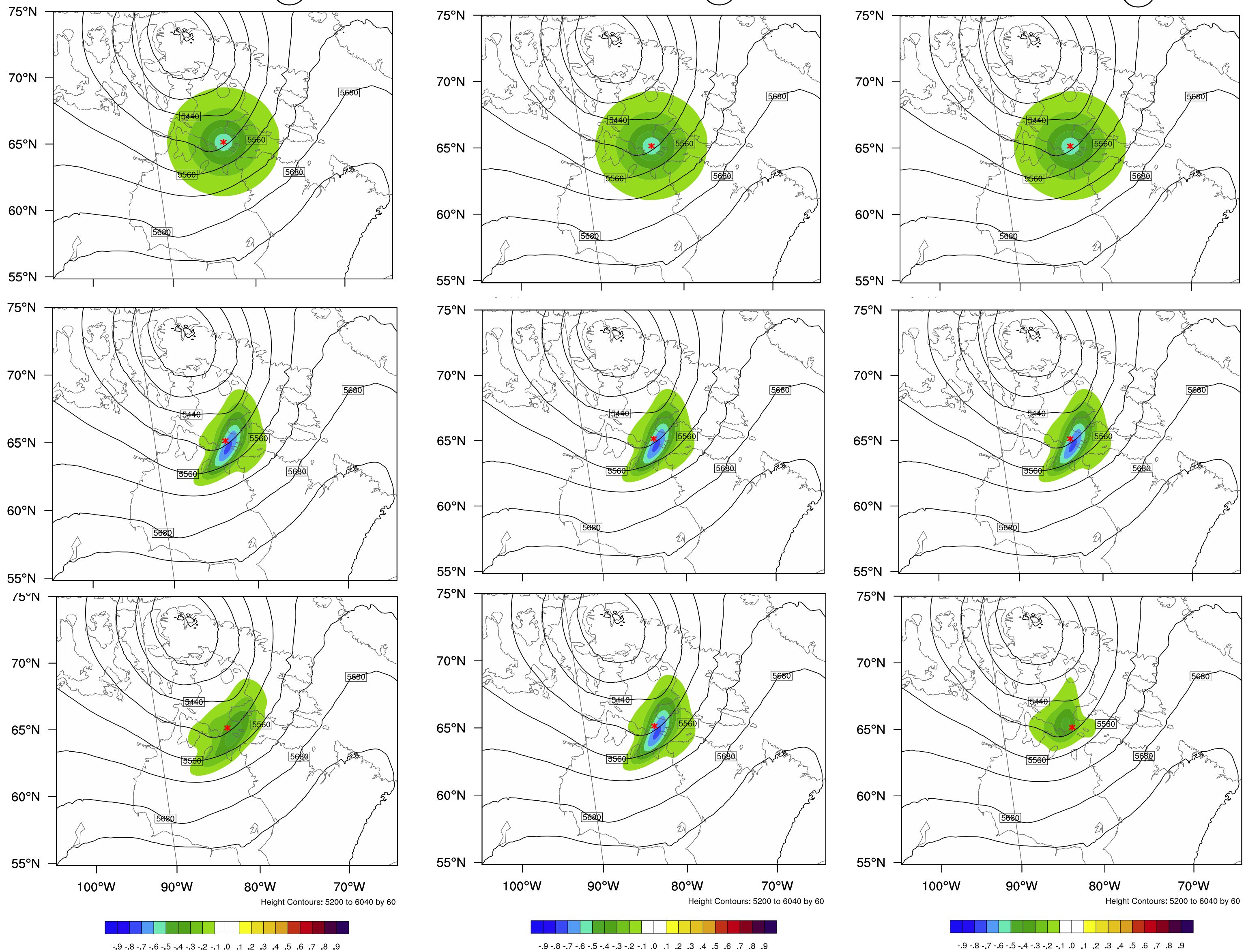
Pseudo Single Observation Tests: Analysis Increments

Observation @-3h | Observation @0h | Observation @+3h

3D Var

Hybrid 3D EnVar

Hybrid 4D EnVar

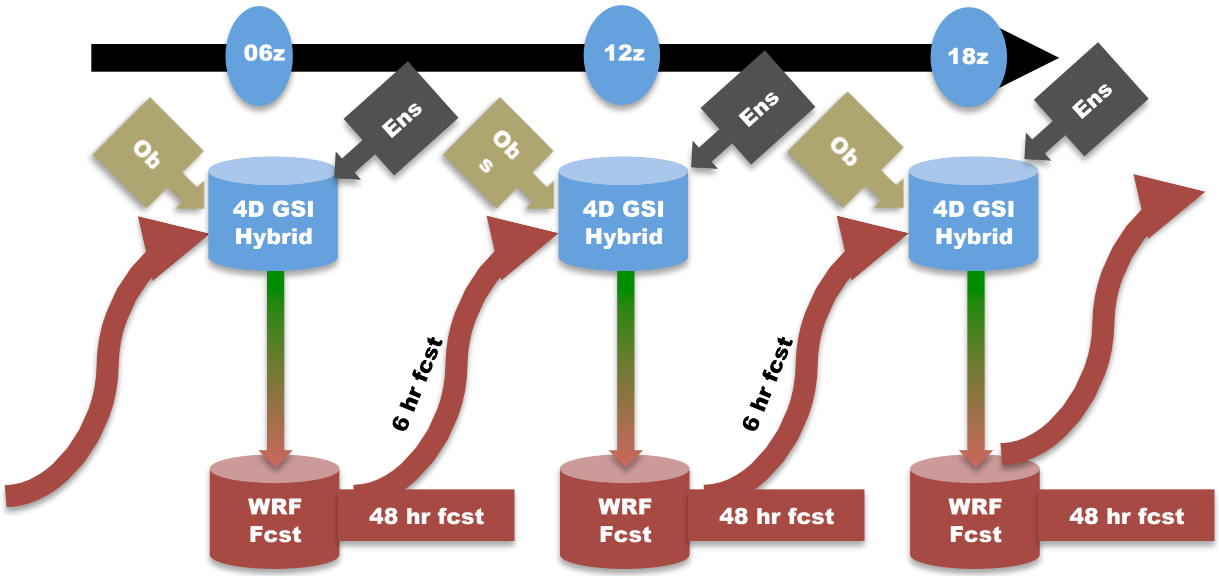


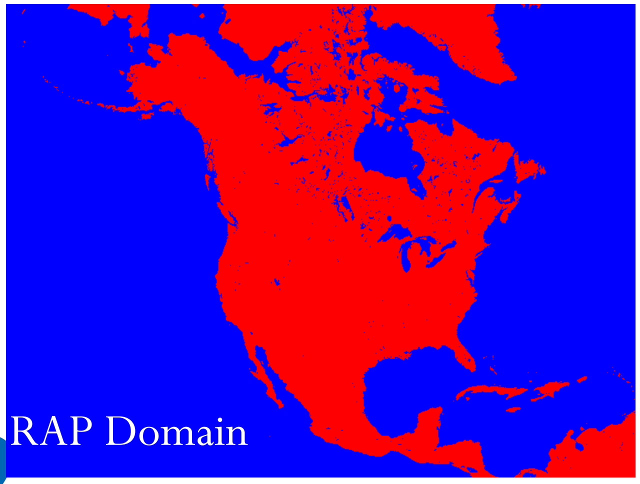
DTC Developmental Testbed Center

Experimental Design: 6 Hourly Cycling

- Proof of feasibility

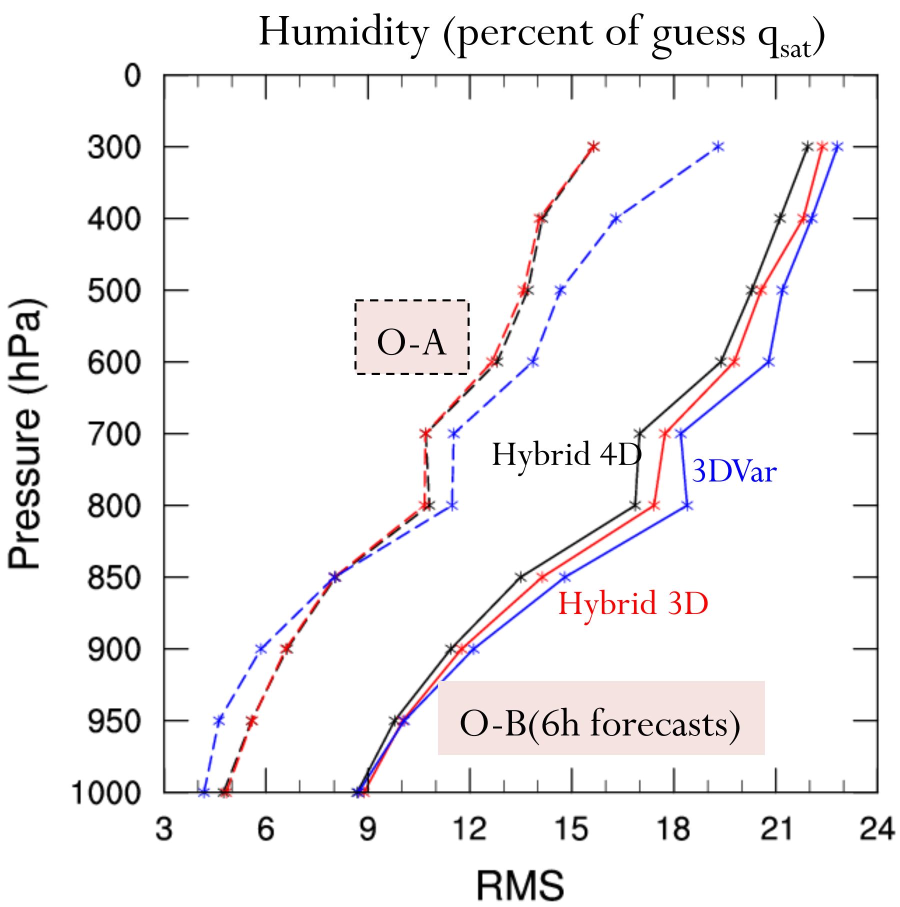
- Experiments using simplified RAP DA framework (e.g. no digital filter prior to forecasts):
  - 3DVAR
  - Hybrid 3D EnVar
  - Hybrid 4D EnVar
- Testing period: 2014080906-2014081600
- 6 hourly continuous cycling
- 6-hour data time window

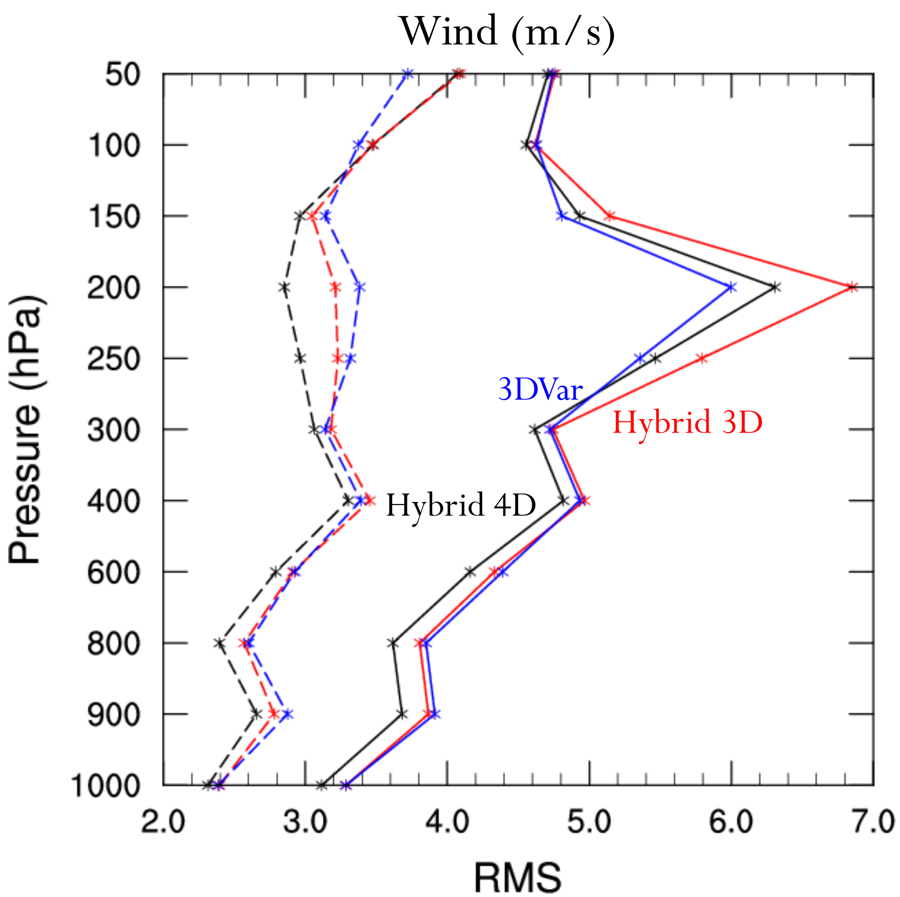




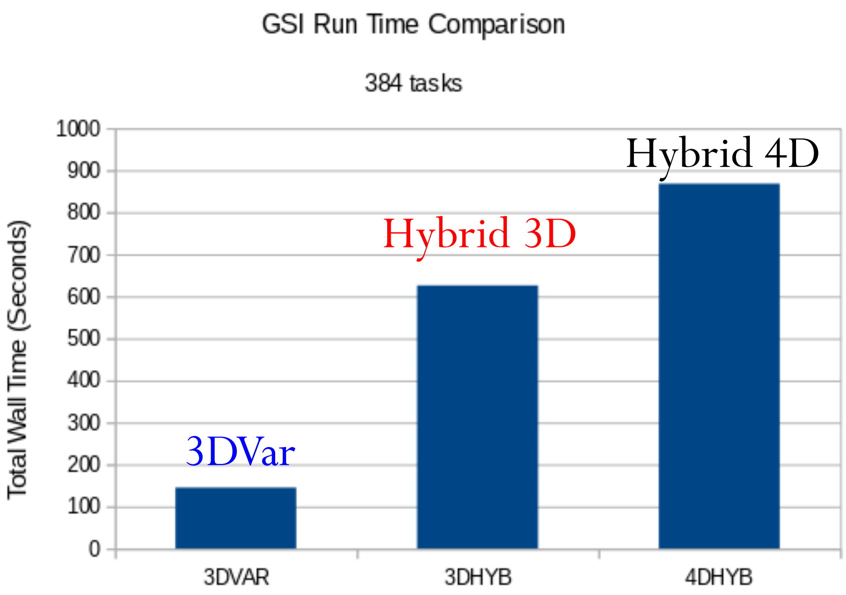
- GFS Data Assimilation System (GDAS) observations:
  - conventional data
  - GPS Radio Occultation
  - Radiance (AMSU-A, , MHS, HIRS4)
- RAP regional BE
- Warm start cycling radiance bias correction from previous RAP cycle
- 80-member
- 3,6 and 9hr GFS ensemble used (3 hour time bins)

Fit to Observations

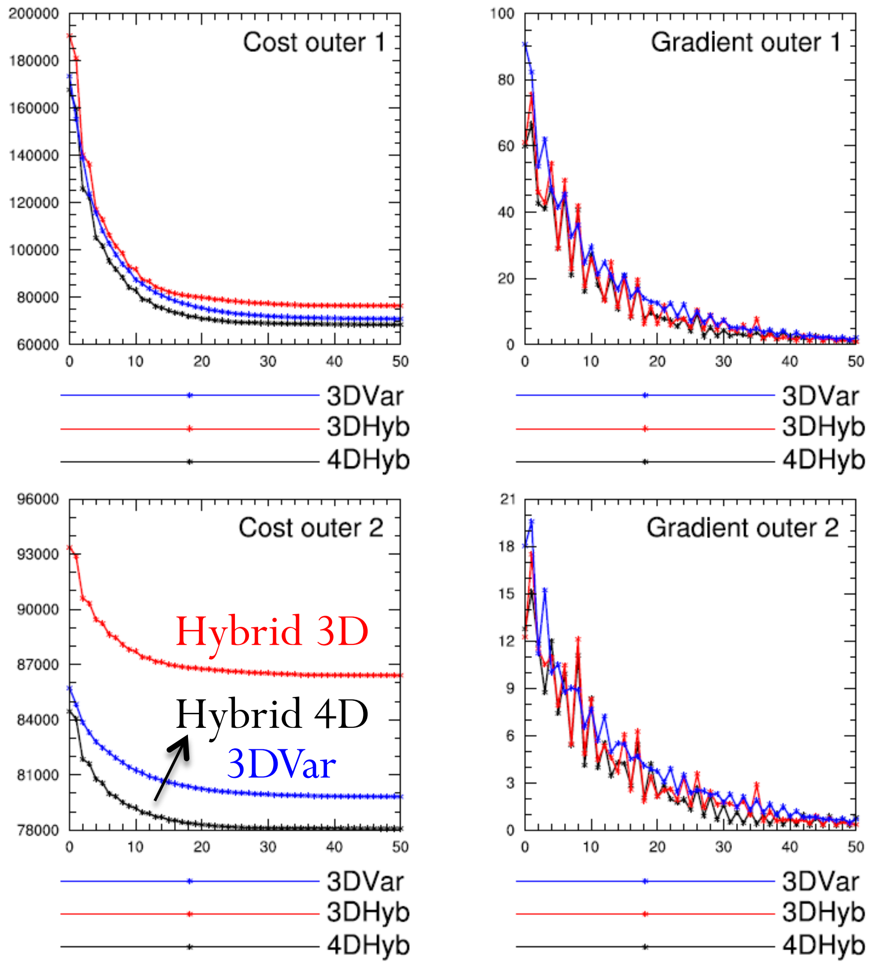




Minimization and Computational Cost



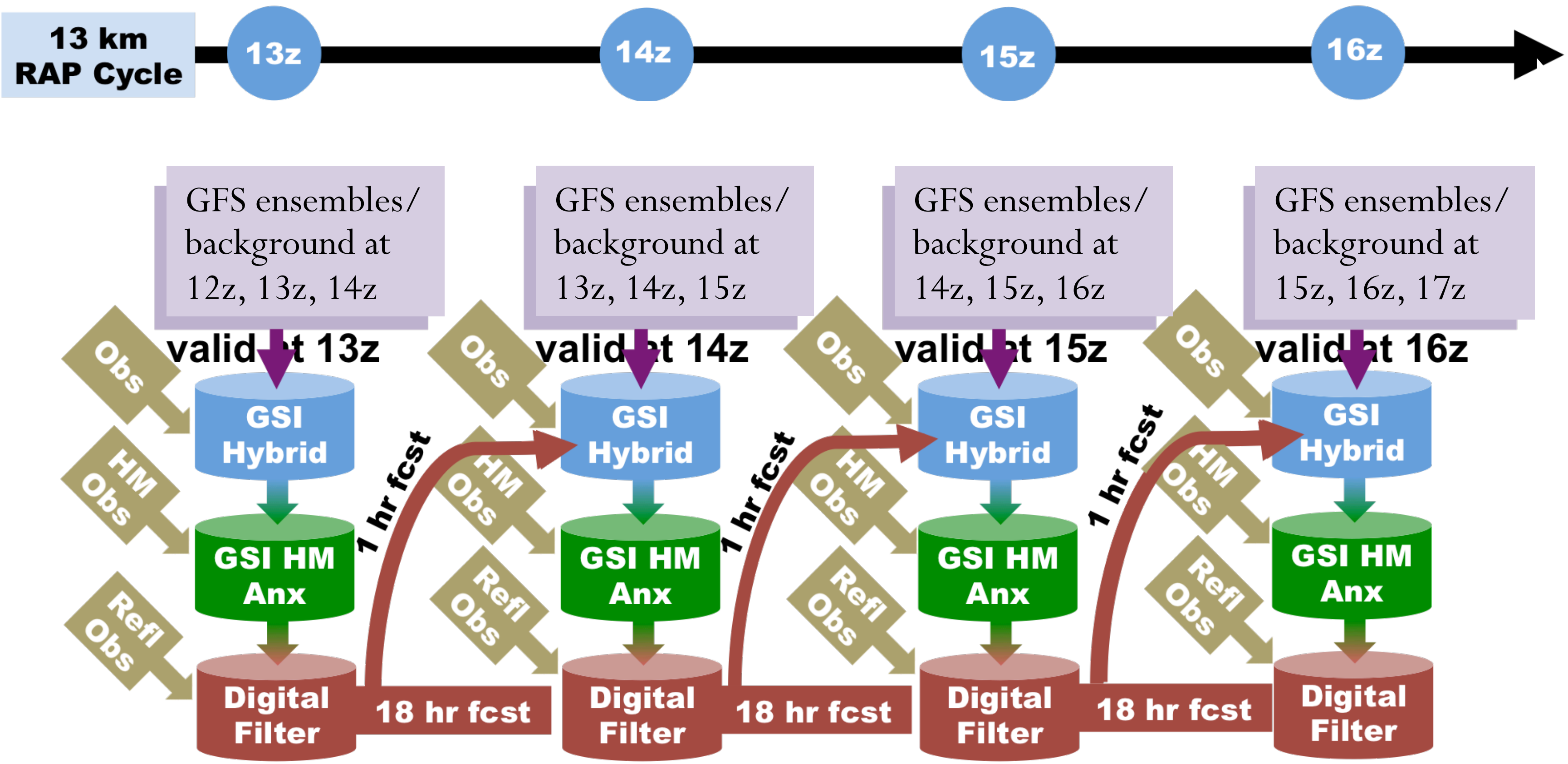
Using 384 processors



Cost function

Norm of cost function

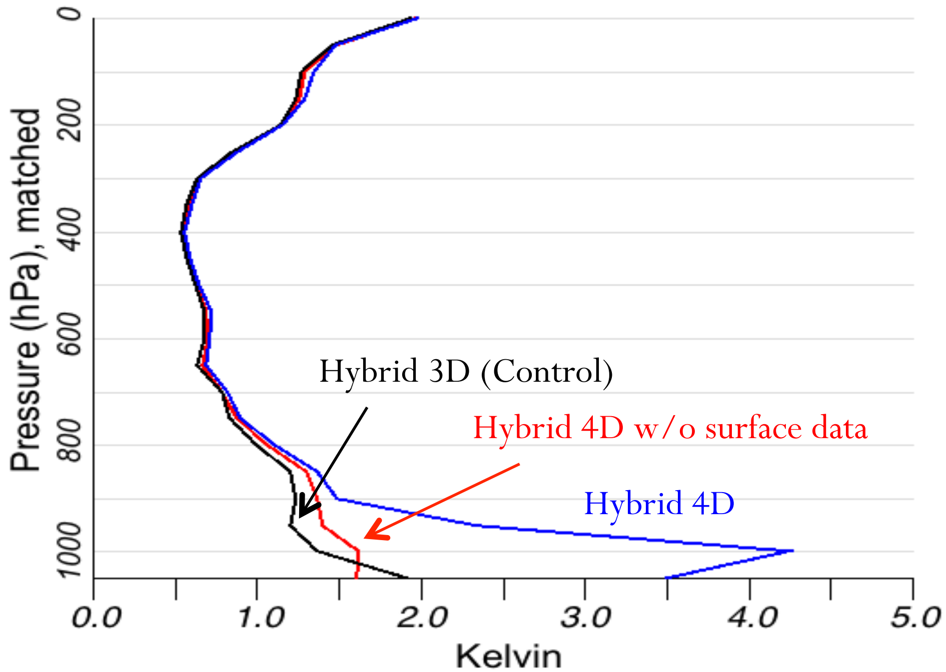
RAP: GSI 4D Hybrid with GFS Ensemble



- Hourly continuous cycling
- 2-hour data time window
- 1-hour time bins within each data time window

- RAP real-time observation files
- 80 GFS ensembles

Results from First Runs



- GFS observation files (used in 6 cycling experiments) reject most surface observations
- RAP observations files contain most available surface observations
- RAP DA uses 2m temperature/moisture and 10m wind background for surface data assimilation

  - Current GSI 4D EnVar capability does not handle time-bins correctly associated with RAP surface DA

Limitation and Technique Challenges

- 4D background/data handling issues
  - Surface DA being fixed inside GSI
  - Checking other data types
- Some GSI capabilities are limited to GFS or other regional models (e.g., NMM-B)
  - Dual resolution for ARW is not working
  - GSI can not read in multiple-time ARW ensembles (e.g., 4D EnVar cannot use regional ensembles directly)
    - Being fixed inside GSI
  - Full member handling (global ensemble mean as a member) namelist option is not available for ARW
    - Research capability developed by Wanshu Wu (NCEP-EMC, 2016) for NOAA NMM-B
- Other factors
  - Ensemble representations
  - Digital filters
  - Time windows/bins

Summary and Plans

- The 3D hybrid EnVar experiments using 13km RAP ensembles (downscaled from 30km GFS ensembles) present limited impacts on forecasts against those using GFS ensembles. However, regional ensembles with higher ensemble spread show potential to further improve analyses and forecasts
- Six-hour cycling of 4D EnVar experiments present great potential to improve regional weather analysis and forecasts
- RAP hourly cycling experiments examined the readiness of the 4D EnVar system for the rapid update system. It shows the technique is feasible, while still needed is thorough examination of 4D handling of background and observations in appropriate time windows/bins
- The DTC will continue to test and assist further development of 4D EnVar capability for regional applications, in collaboration with developers
  - Higher model and ensemble resolutions, e.g. HRRR
  - Fast-cycling (e.g., hourly or sub-hourly) of 4D EnVar for high-frequency observations
  - Other ensemble perturbation methods, e.g., stochastic physics