GSI-based Hybrid Data Assimilation: Practical Implementation and impact on the NCEP GFS

Daryl Kleist

Univ. of Maryland-College Park, Dept. of Atmos. & Oceanic Science

with acknowledgements to Dave Parrish, Jeff Whitaker, John Derber, Russ Treadon, Wan-shu Wu, Kayo Ide, and many others

National Taiwan University, Taipei, Taiwan – 6 November 2014

Data Assimilation

- NWP is both an initial and boundary value problem. To integrate the numerical model forward in time, we need an estimate for the starting point (initial conditions). We never know the exact "truth", but have estimates to within various uncertainties.

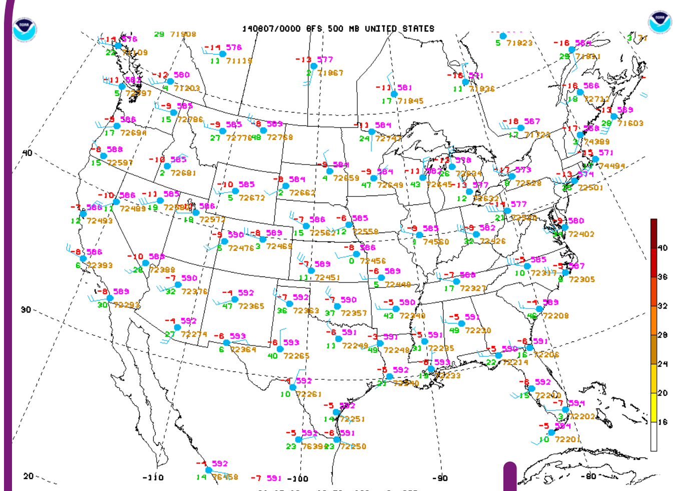
- **Data Assimilation** – Incorporating observations into a (numerical) model of a (geo) physical system
  - Analysis procedure linked to model of physical system
  - Critical component to Numerical Weather Prediction (NWP)
    - Also used in other fields for initializing prediction models
    - Other applications like reanalysis/climate monitoring
  - Notion of providing "balanced" information that will not be rejected by model
    - Initialization techniques
  - Many different algorithms available
    - Kalman Filter, ensemble, variational, and combinations thereof

Data Assimilation

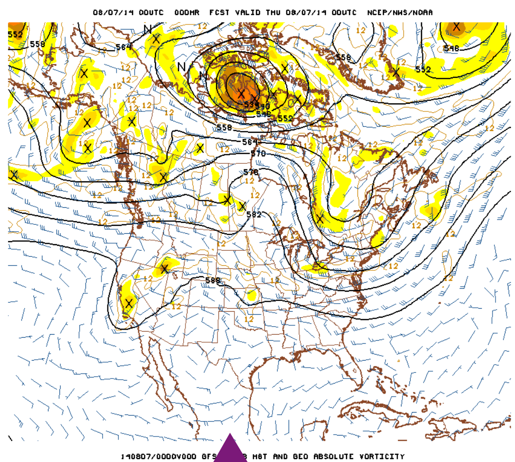
- **Data assimilation** is an iterative method for monitoring nature (**the process is cumulative since states are cycled**)

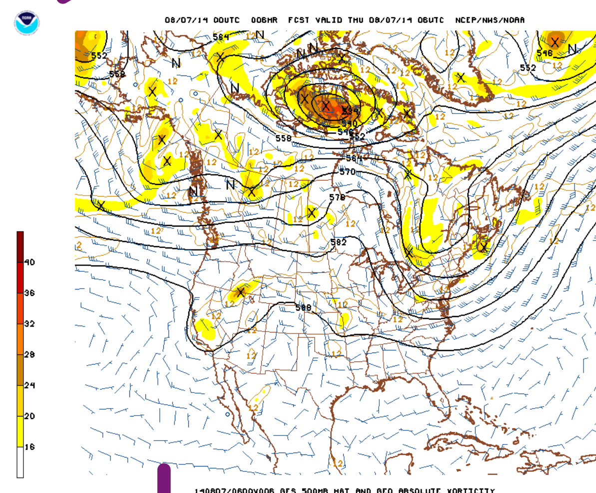
Step 1. Model forecast

past observation



Initial condition → future condition



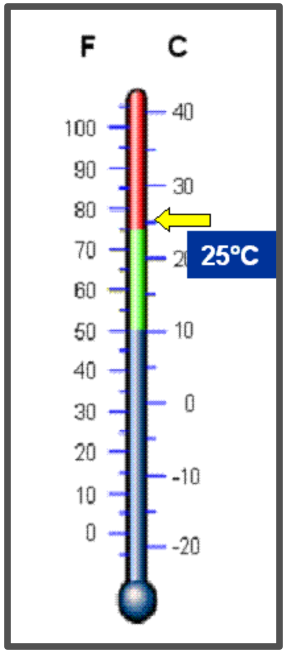


Step 2. Integration of observed info into the model condition

by optimally combining:

- Model
- Observations

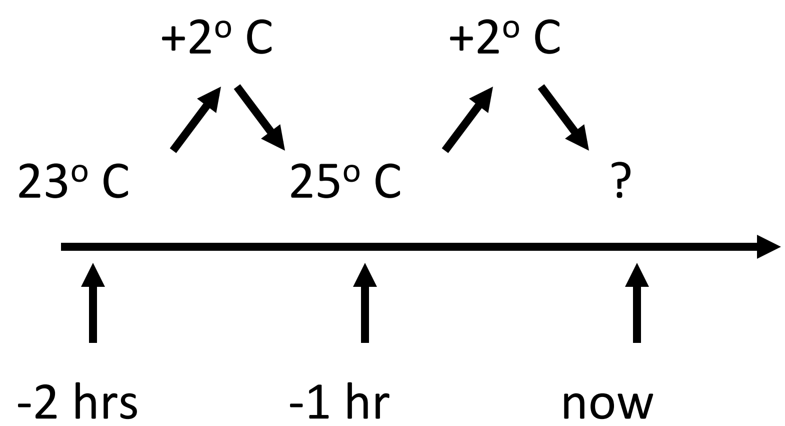
Simple (Scalar) Assimilation Example: Room Temperature



Observation (T_o):
25° C

$$\sigma_o = T_o - T_t$$

Trend Forecast:



"Model" (T_b):
27° C

$$\sigma_b = T_b - T_t$$

Assume that errors are known and unbiased

$$\sigma_o = 1°\mathrm{C}$$

$$\sigma_b = 2°\mathrm{C}$$

$$\bar{\sigma}_o = \bar{\sigma}_b = 0$$

Simple (Scalar) Assimilation Example: Room Temperature

Combine background and observation using Linear Unbiased Estimate to create analysis

$$T_a = \alpha T_o + (1-\alpha) T_b$$

It can be shown that the minimum variance estimate occurs when

$$\alpha = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

In our example, this results in α of 0.8, and **T_a=25.4ºC**

It can also then be shown that (analysis error is less than observation or background error):

$$\sigma_a^2 = \left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2} \right)^{-1}$$

This can also be extended to multiple dimensions

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H\mathbf{x}_b)$$

Data Assimilation Example

- Here is an example of combining ozone observations with a short term ozone forecast, using their assumed errors to come up with an "analysis" (with errors that are less than either the observations or the model forecast).

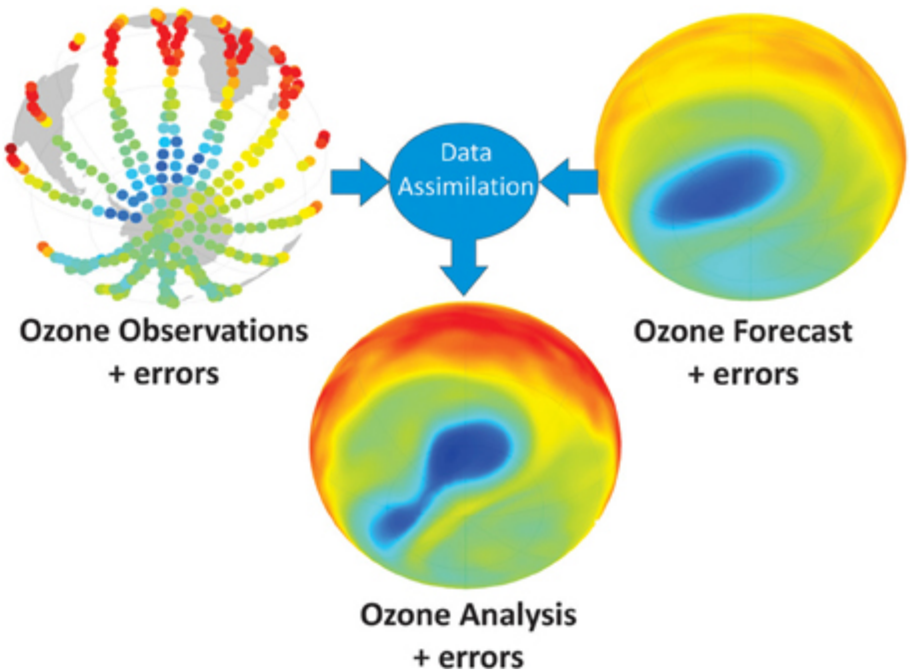


Image courtesy of: AUTHOR=Lahoz William Albert, Schneider Philipp, Data Assimilation: Making Sense of Earth Observation , Frontiers in Environmental Science (DOI=10.3389/fenvs.2014.00016)

Variational Data Assimilation

$$J_{\mathrm{Var}}\left(\mathbf{x}'\right)=\frac{1}{2}\left(\mathbf{x}'\right)^{\mathrm{T}}\mathbf{B}_{\mathrm{Var}}^{-1}\left(\mathbf{x}'\right)+\frac{1}{2}\left(\mathbf{H}\mathbf{x}'-\mathbf{y}_{\mathrm{o}}'\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{x}'-\mathbf{y}_{\mathrm{o}}'\right)+J_{\mathrm{c}}$$

J : Penalty (Fit to background + Fit to observations + Constraints)

x' : Analysis increment ($x_a - x_b$) ; where x_b is a background

B_{var}: Background error covariance

H : Observations (forward) operator

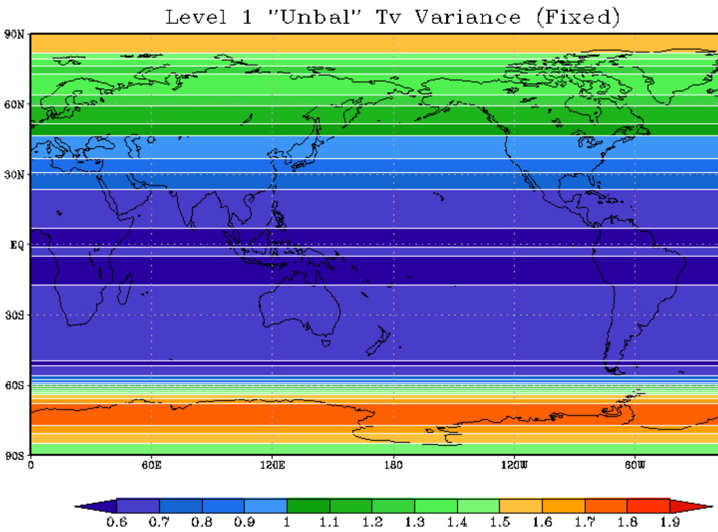
R : Observation error covariance (Instrument + representativeness)

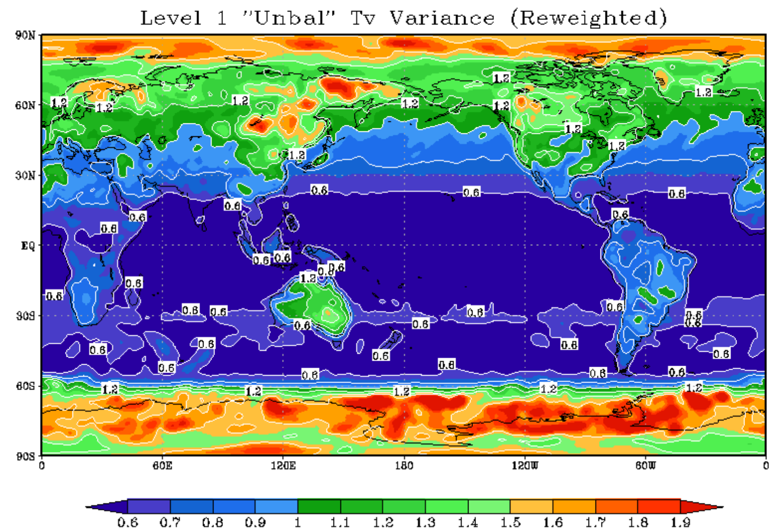
y_o' : Observation innovations

J_c : Constraints (physical quantities, balance/noise, etc.)

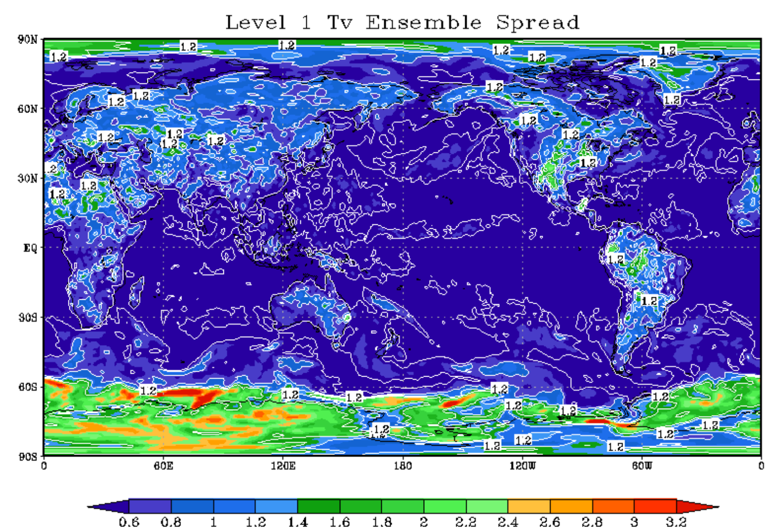
B is typically static and estimated a-priori/offline

Examples of **B** for Temperature





- Although flow-dependent ***variances*** are used, confined to be a rescaling of fixed estimate based on time tendencies
  - No multivariate or length scale information used
  - Does not necessarily capture 'errors of the day'



- Plots valid 00 UTC 12 September 2008

Kalman Filter in Var Setting

Forecast Step

$$\mathbf{x}^{\mathrm{b}} = M\left(\mathbf{x}^{\mathrm{a}}\right)$$

$$\mathbf{B}_{\mathrm{KF}} = \mathbf{M}\mathbf{A}_{\mathrm{KF}}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}$$

Analysis

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{K}\left(\mathbf{H}\mathbf{x}^{\mathrm{b}} - \mathbf{y}\right)$$

$$\mathbf{K} = \mathbf{B}_{\mathrm{KF}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}_{\mathrm{KF}}\mathbf{H}^{\mathrm{T}}\right)^{-1}$$

$$\mathbf{A}_{\mathrm{KF}} = \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\mathbf{B}_{\mathrm{KF}}$$

Extended Kalman Filter

- Analysis step in variational framework (cost function)

$$J_{\mathrm{KF}}\left(\mathbf{x}'\right) = \frac{1}{2}\left(\mathbf{x}'\right)^{\mathrm{T}}\mathbf{B}_{\mathrm{KF}}^{-1}\left(\mathbf{x}'\right) + \frac{1}{2}\left(\mathbf{y}'_{\mathrm{o}} - \mathbf{H}\mathbf{x}'\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}'_{\mathrm{o}} - \mathbf{H}\mathbf{x}'\right)$$

- $\mathbf{B}_{\mathrm{KF}}$: Time evolving background error covariance
- $\mathbf{A}_{\mathrm{KF}}$: Inverse [Hessian of $J_{\mathrm{KF}}(\mathbf{x}')$]

Motivation from KF

- ***Problem***: dimensions of $\mathbf{A}_{KF}$ and $\mathbf{B}_{KF}$ are huge, making this practically impossible for large systems (GFS for example).
- ***Solution***: sample and update using an ensemble instead of evolving $\mathbf{A}_{KF}/\mathbf{B}_{KF}$ explicitly

Forecast Step: $\mathbf{X}^{a} \rightarrow \mathbf{X}^{b}$ $\quad \mathbf{B}_{KF} \approx \mathbf{B}_{\mathbf{e}} = \frac{1}{K-1}\mathbf{X}^{b}\left(\mathbf{X}^{b}\right)^{T}$

Analysis Step: $\mathbf{X}^{b} \rightarrow \mathbf{X}^{a}$ $\quad \mathbf{A}_{KF} \approx \mathbf{A}_{\mathbf{e}} = \frac{1}{K-1}\mathbf{X}^{a}\left(\mathbf{X}^{a}\right)^{T}$

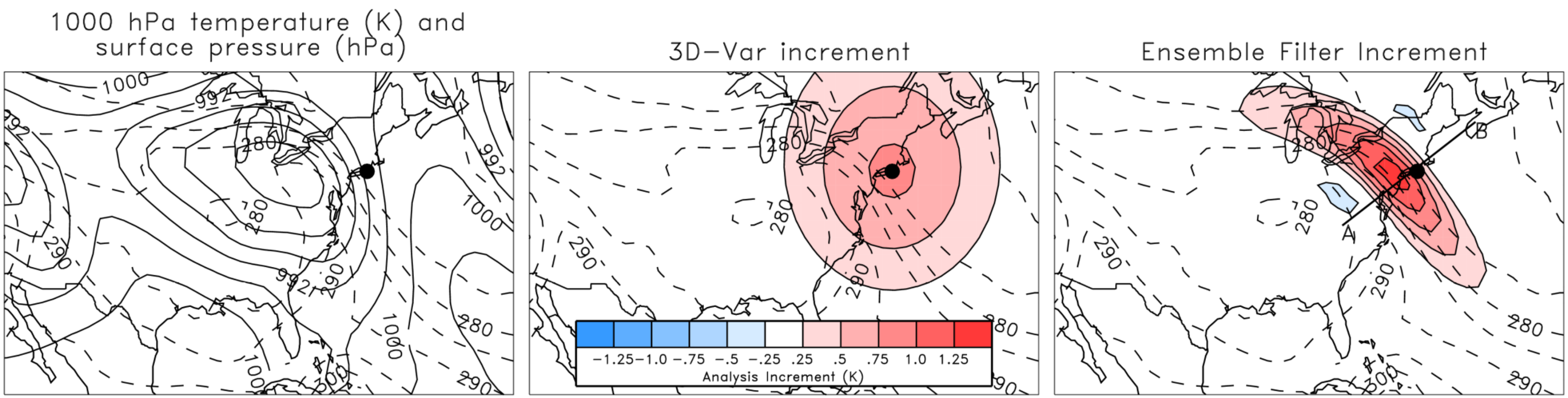
} Ensemble Perturbations

What does $\mathbf{B}_e$ gain us?

- Allows for flow-dependence/errors of the day
- Multivariate correlations from dynamic model
  - Quite difficult to incorporate into fixed error covariance models
- Evolves with system, can capture changes in the observing network
- More information extracted from the observations => better analysis => better forecasts

What does $\mathbf{B}_e$ gain us?

Temperature observation near warm front

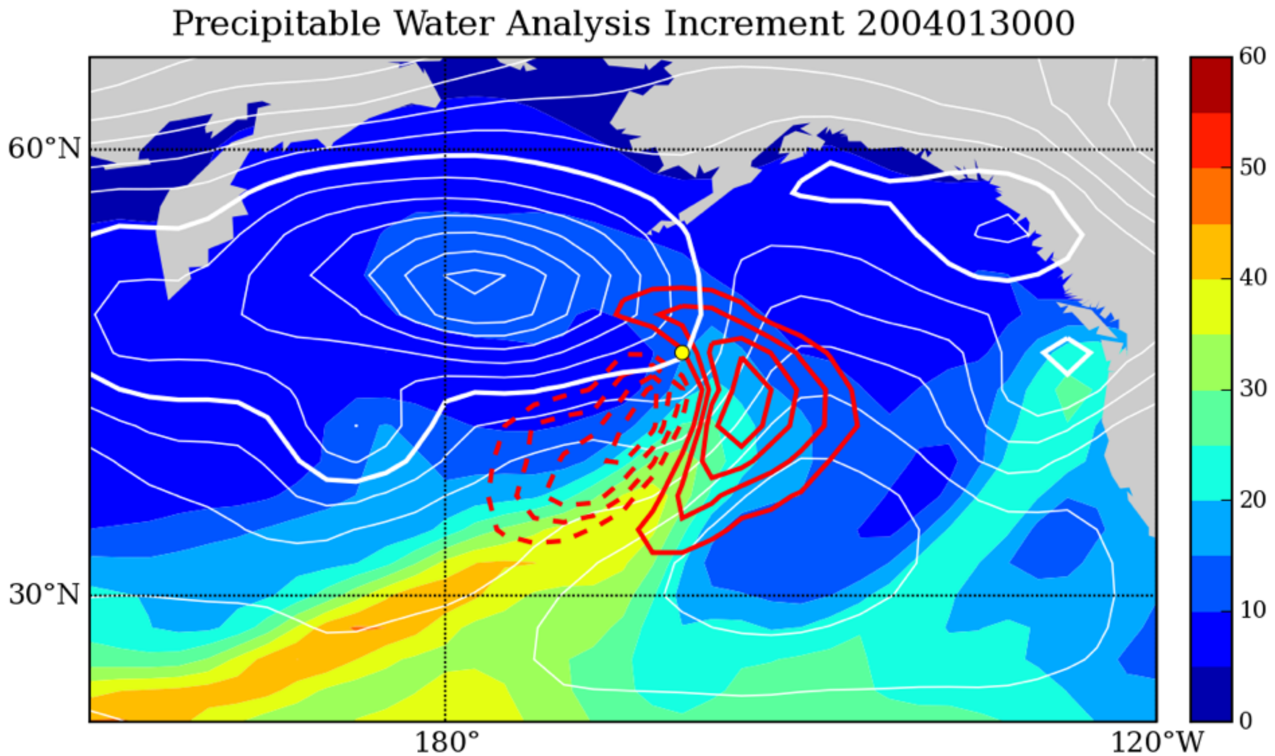


$\mathbf{B}_f$

$\mathbf{B}_e$

What does $\mathbf{B}_e$ gain us?

Surface pressure observation near "atmospheric river"



First guess surface pressure (white contours) and precipitable water increment (A-G, red contours) after assimilating a single surface pressure observation (yellow dot) using $\mathbf{B}_e$.

What is "hybrid DA"?

$$J(\mathbf{x}') = \beta_f \frac{1}{2}(\mathbf{x}')^T \mathbf{B}_f^{-1}(\mathbf{x}') + \beta_e \frac{1}{2}(\mathbf{x}')^T \mathbf{B}_e^{-1}(\mathbf{x}') + \frac{1}{2}(\mathbf{Hx}' - \mathbf{y}')^T \mathbf{R}^{-1}(\mathbf{Hx}' - \mathbf{y}')$$

Simply put, linear combination of fixed and ensemble based **B:**

$\mathbf{B}_f$: Fixed background error covariance

$\mathbf{B}_e$: Ensemble estimated background error covariance

β_f: Weighting factor for fixed contribution (0.25 means 25% fixed)

β_e: Weighting factor for ensemble contribution (typically 1- β_f)

Experiments with toy model

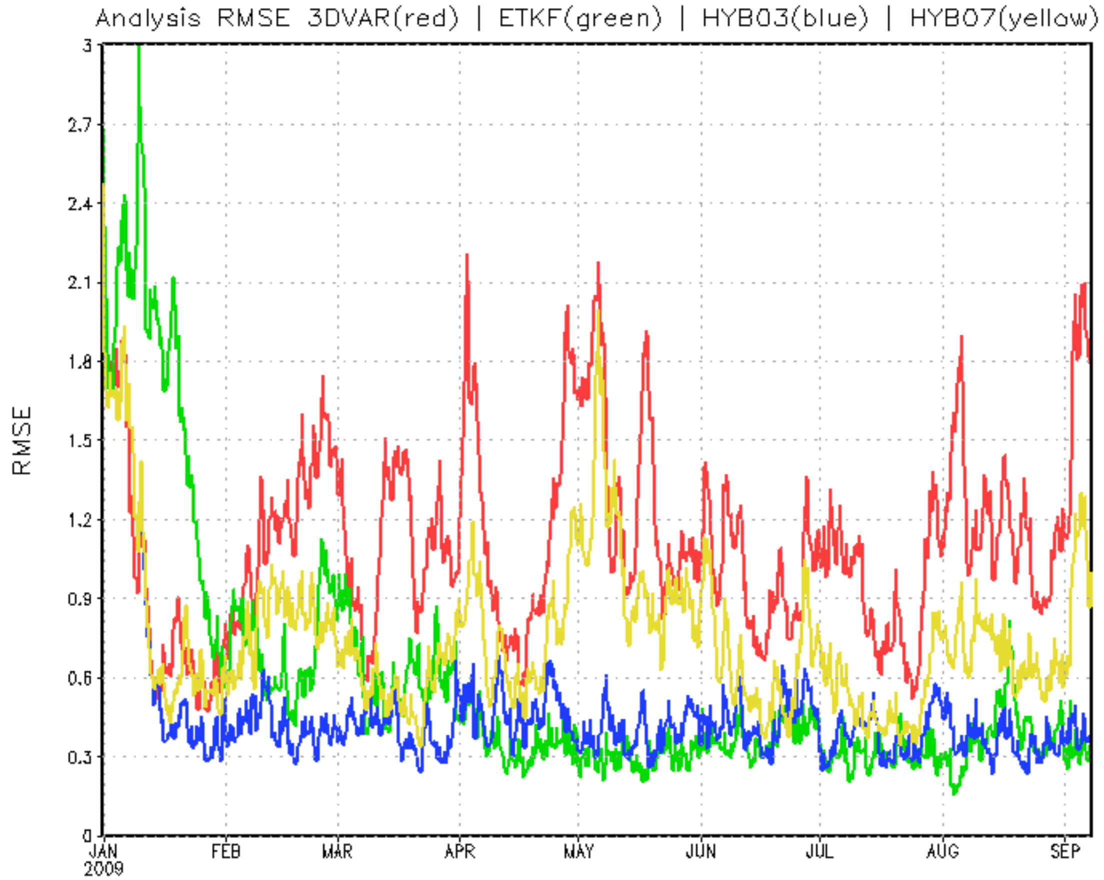
$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F$$

- Lorenz '96
  - 40 variable model, F=8.0, dt=0.025 ("3 hours")
  - 4th order Runge-Kutta

- OSSE: observations generated from truth run every 2*dt ("6 hours")
  - [N(0,1)]

- Experimental design
  - Assimilate single time level observations every 6 hours, at appropriate time, R=1.0
  - F=7.8 (imperfect model) for DA runs

Analysis Error (50% observation coverage)



3DVAR

$\beta_f = 0.7$

$\beta_f = 0.3$

ETKF

- M (ensemble size) = 20, ρ (inflation factor) = 1.1
  - Hybrid (small alpha) as good as/better than ETKF (faster spinup)
  - Hybrid (larger alpha) in between 3DVAR and ETKF

Sensitivity to β

Analysis RMSE (x10) over 1800 cases

β_f	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3DVAR	12.08								
Hybrid	***3.321***	***3.764***	4.074	4.633	5.060	5.770	7.044	8.218	9.595
ETKF	3.871								

- 50% observation coverage (M = 20, $\rho = 1.1$)
  - Improvement a near linear function of weighting parameter
- Small enough weighting (on static error estimate) improves upon ETKF

GSI Hybrid [3D] EnVar

(ignoring preconditioning for simplicity)

- Incorporate ensemble perturbations ***directly*** into variational cost function through extended control variable
  - Lorenc (2003), Buehner (2005), Wang et. al. (2007), etc.

$$J(\mathbf{x}_f', \boldsymbol{\alpha}) = \beta_f \frac{1}{2}(\mathbf{x}_f')^T \mathbf{B}_f^{-1}(\mathbf{x}_f') + \beta_e \frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{\alpha}^n)^T \mathbf{L}^{-1}(\boldsymbol{\alpha}^n) + \frac{1}{2}(\mathbf{H}\mathbf{x}_t' - \mathbf{y}')^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_t' - \mathbf{y}')$$

$$\mathbf{x}_t' = \mathbf{x}_f' + \sum_{n=1}^{N}\left(\boldsymbol{\alpha}^n \circ \mathbf{x}_e{}^n\right)$$

β_f & β_e: weighting coefficients for fixed and ensemble covariance respectively

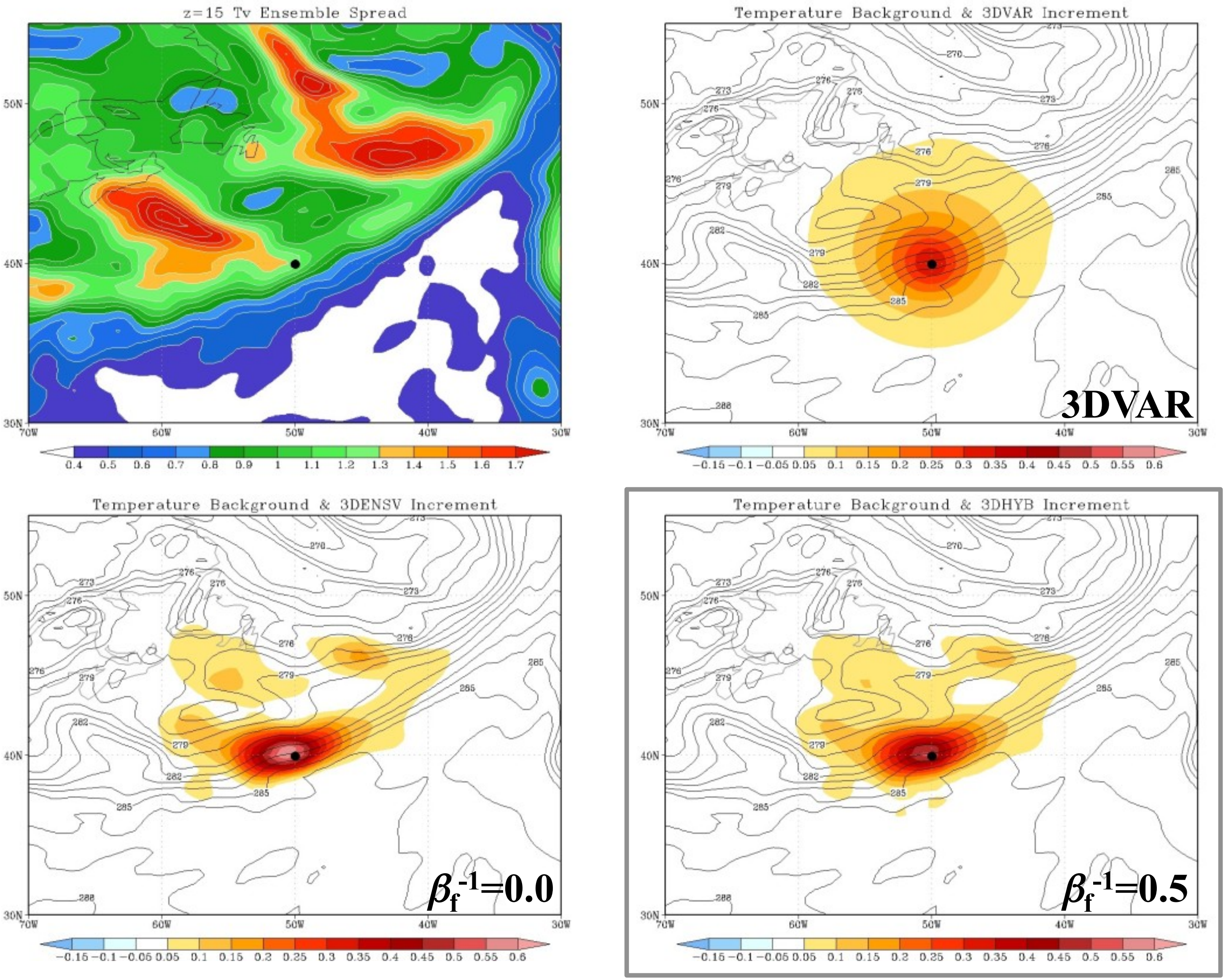
$\mathbf{x}_t'$: (total increment) sum of increment from fixed/static **B** ($\mathbf{x}_f'$) and ensemble **B**

$\boldsymbol{\alpha}_k$: extended control variable; $\mathbf{x}_k^e$:ensemble perturbations

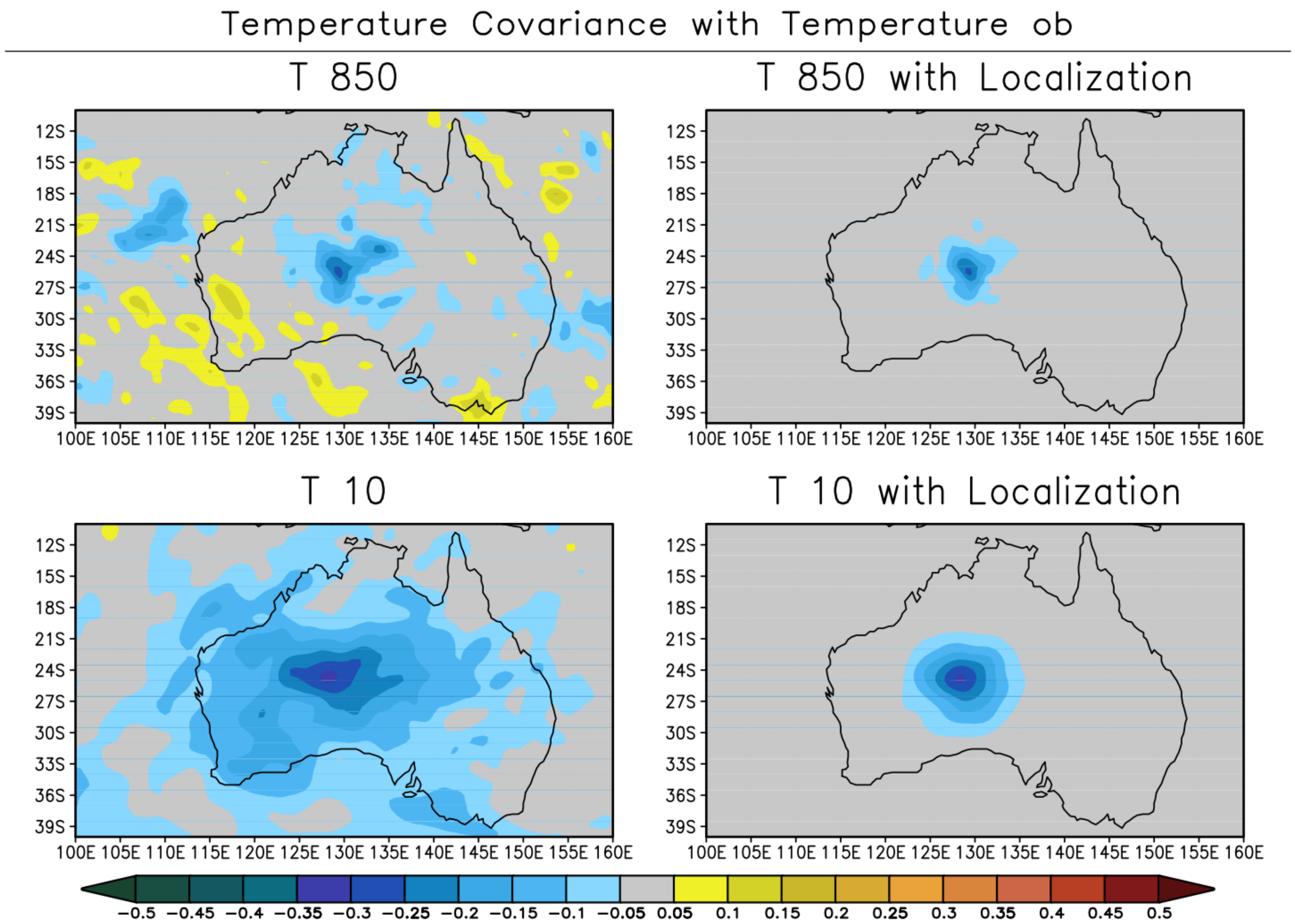
- analogous to the weights in the LETKF formulation

L: correlation matrix [effectively the localization of ensemble perturbations]

Single Temperature Observation



Localization



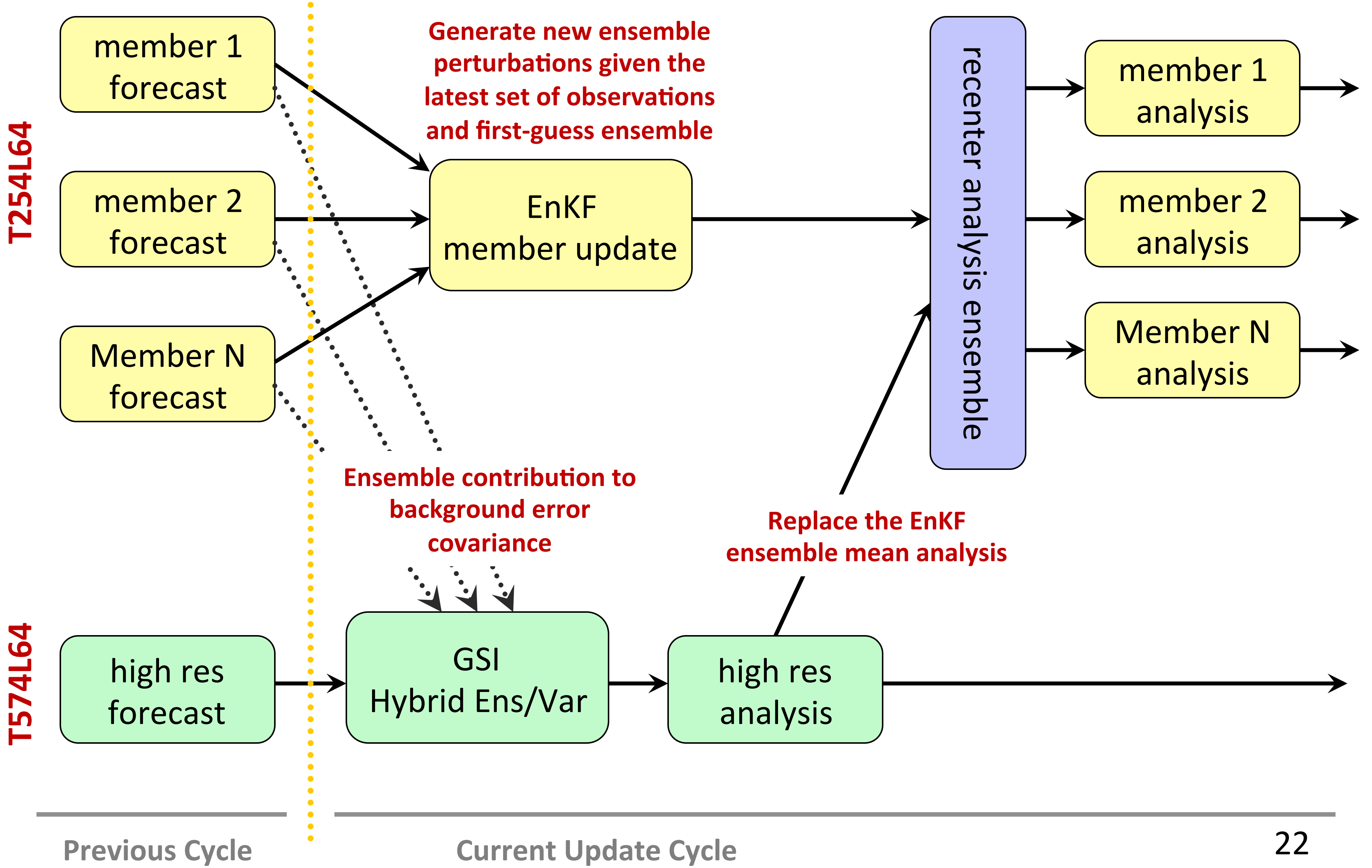
So what's the catch?

- Need an ensemble that represents first guess uncertainty (background error)

- This can mean O(50-100+) for NWP applications
  - Smaller ensembles have larger sampling error (rely more heavily on $\mathbf{B}_f$)
  - Larger ensembles have increased computational expense

- Updating the ensemble: In NCEP operations, we currently utilize an Ensemble Kalman Filter
  - EnKF is a standalone (i.e. separate) DA system that updates every ensemble member with information from observations each analysis time using the prior/posterior ensemble to represent the error covariances. Google "ensemble based atmospheric assimilation" for a good review article by Tom Hamill.

Dual-Res Coupled Hybrid Var/EnKF Cycling



Global Data Assimilation System Upgrade

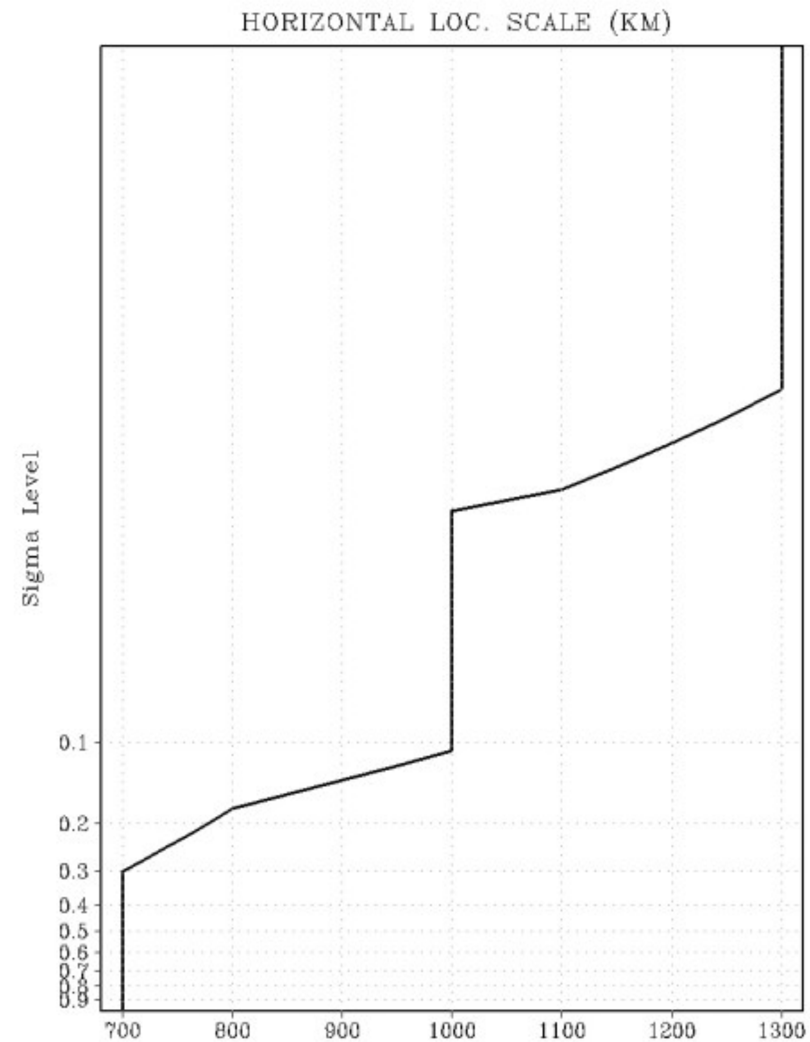
Implemented 22 May 2012

- **Hybrid system**
  - Most of the impact comes from this change
  - Uses ensemble forecasts to help define background error
- **NPP (ATMS) assimilated**
  - Quick use of data after launch
- **Use of GPSRO Bending Angle rather than refractivity**
  - Allows use of more data (especially higher in atmos.)
  - Small positive impacts
- **Satellite radiance monitoring code**
  - Allows quicker awareness of problems (run every cycle)
  - Monitoring software can automatically detect many problems
- **Post changes**
  - Additional fields requested by forecasters (80m variables)
- **Partnership between research and operations**

Operational Configuration

- Full **B** preconditioned double conjugate gradient minimization
- Spectral filter for horizontal part of **L**
  - Eventually replace with (anisotropic) recursive filters
- Recursive filter used for vertical
  - 0.5 scale heights
- Same localization used in Hybrid (**L**) and EnSRF
- TLNMC (Kleist et al. 2009) applied to total analysis increment*

$$\mathbf{x}_t' = \mathbf{C}\left[\mathbf{x}_f' + \sum_{n=1}^{N}\left(\boldsymbol{\alpha}^n \circ \mathbf{x}_e{}^n\right)\right]$$



Hybrid Impact in Pre-implementation Tests (to appear in BAMS article)

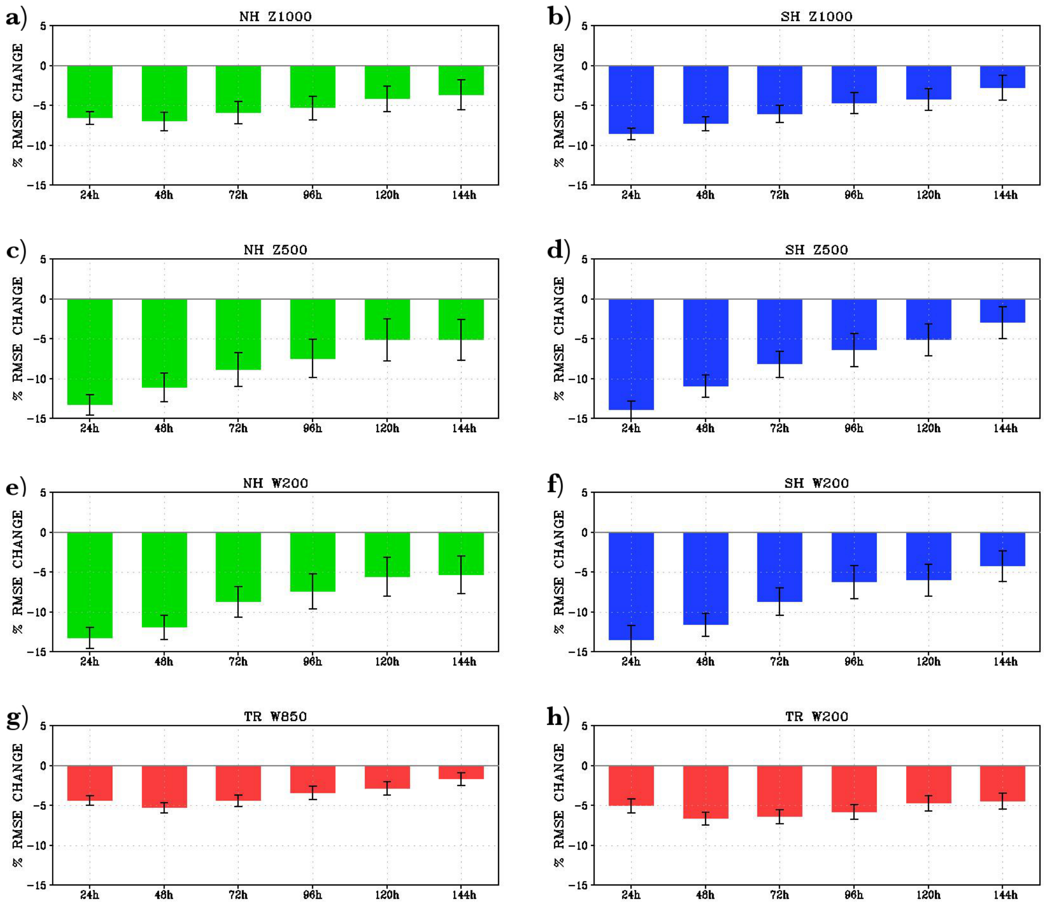


Figure 01: Percent change in root mean square error from the experimental GFS minus the operational GFS for the period covering 01 February 2012 through 15 May 2012 in the northern hemisphere (green), southern hemisphere (blue), and tropics (red) for selected variables as a function of forecast lead time. The forecast variables include 1000 hPa geopotential height (a, b), 500 hPa geopotential eight (c, d), 200 hPa vector wind (e, f, h), and 850 vector wind (g). All verification is performed using self-analysis. The error bars represent the 95% confidence threshold for a significance test.

Hybrid Impact in Pre-implementation Tests (to appear in BAMS article)

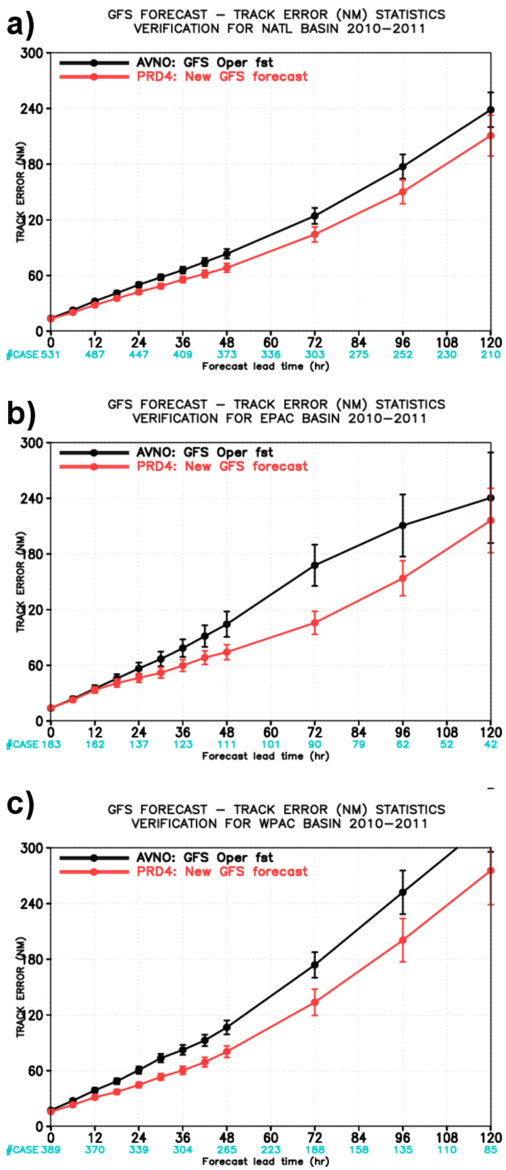
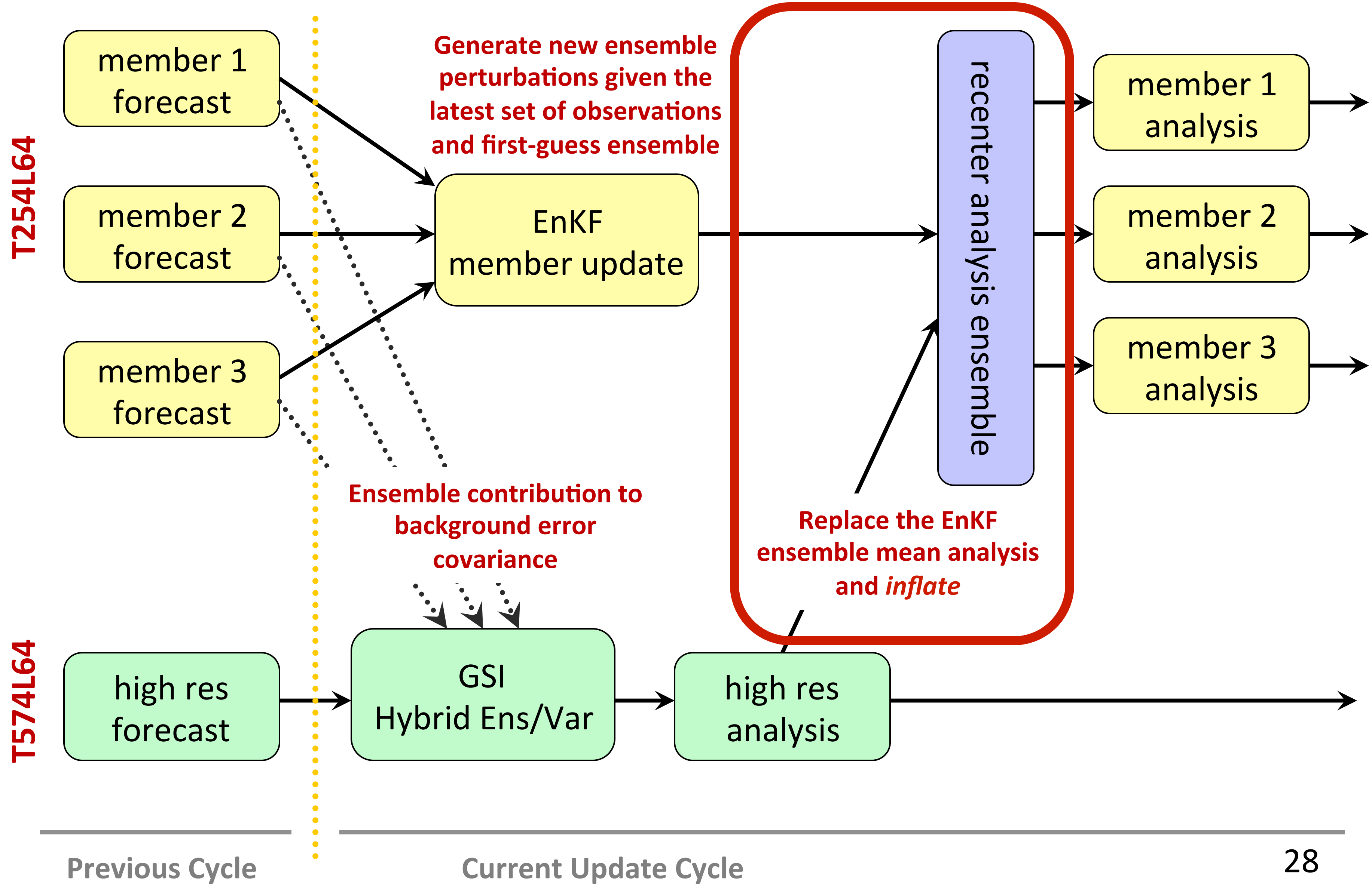


Figure 02: Mean tropical cyclone track errors (nautical miles) covering the 2010 and 2011 hurricane seasons for the operational GFS (black) and experimental GFS including hybrid data assimilation (red) for the a) Atlantic basin, b) eastern Pacific basin, and c) western Pacific basin. The number of cases is specified by the blue numbers along the abscissa. Error bars indicate the 5th and 95th percentiles of a resampled block bootstrap distribution.

NCEP operational 3D ensemble/Var system

- Cycling EnKF (T254) provides an ensemble-based estimate of J_b term in 3DVar.
- 3DVar with ensemble J_b updates a T574 control forecast. The EnKF analysis ensemble is recentered around the high-res analysis.
- A combination of multiplicative and additive inflation is used to represent missing sources of uncertainty in the EnKF ensemble.
- Additive inflation: random draws from a database of 48-24 forecast differences (valid at same time), added to EnKF analysis ensemble.

Dual-Res Coupled Hybrid Var/EnKF Cycling



Post EnKF Inflation (Whitaker and Hamill 2012)

- Multiplicative inflation factor, *ρ,* (function of reduction of spread by assimilation of observations):

$$\rho = \omega\left(\frac{\mathrm{s}^{\mathrm{a}} + \mathrm{s}^{\mathrm{b}}}{\mathrm{s}^{\mathrm{a}}}\right)$$

$$\mathrm{s} = \frac{1}{N-1}\sum_{n=1}^{N}\left(\mathrm{x}_n - \overline{\mathrm{x}}\right)^2$$

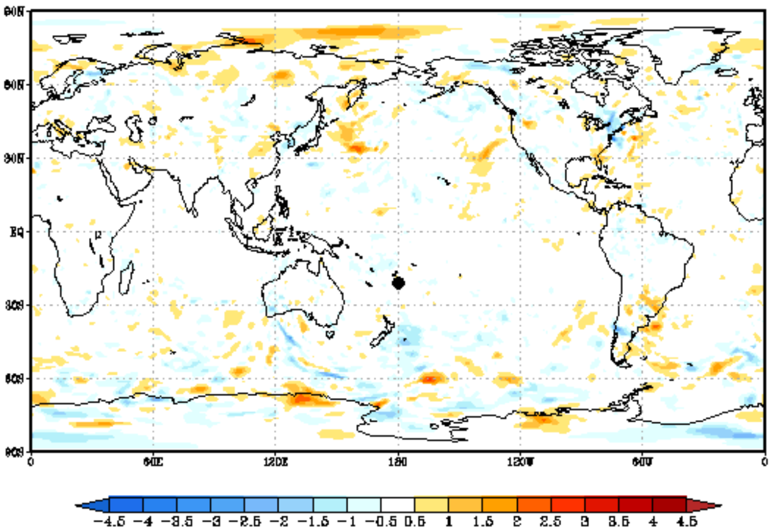
- Additive inflation extracts quasi-balanced pseudo-random perturbations from database of lagged forecast pairs

$$\mathbf{x}_n^{\mathrm{ap}} = \mathbf{x}_n^{\mathrm{a}} + \kappa\left(\mathbf{x}_{\mathrm{r}}'\right).$$

Inflation is an ad-hoc method for overcoming lack of consideration of model/system error within ensemble-filtering systems

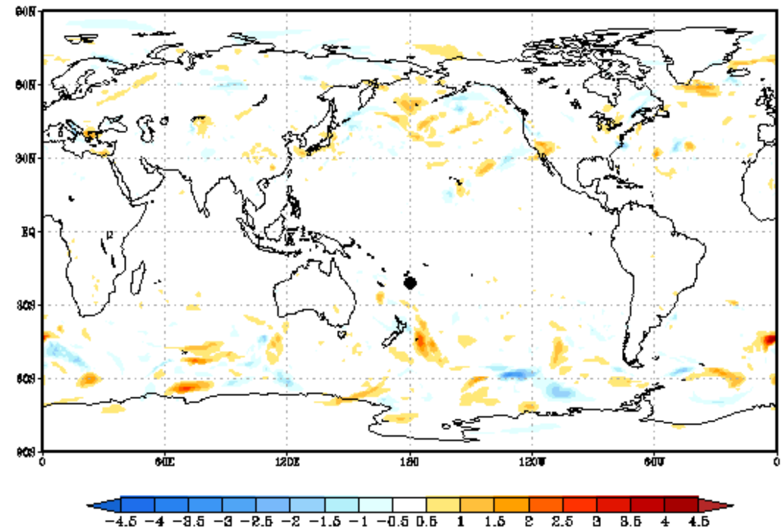
Ensemble Member 01 increments

EnKF Tinc Mem001



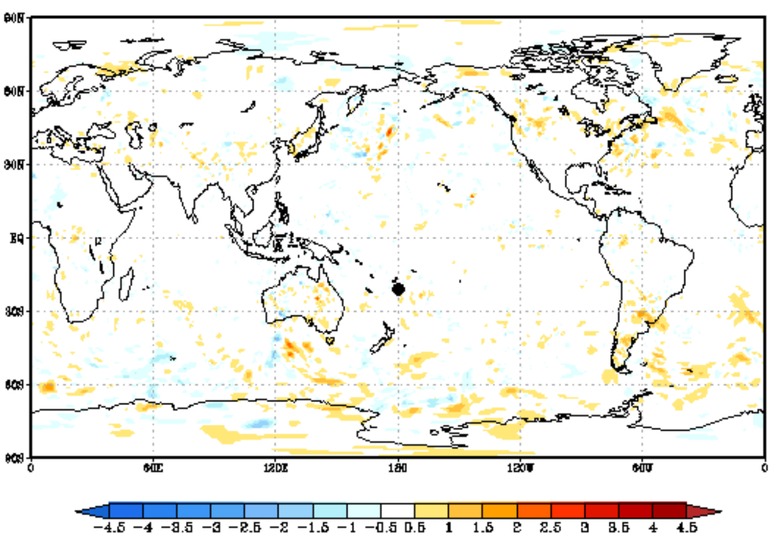
$\delta \mathbf{x}_0^n$

EnKF T Mem001 ADDInflation ONLY



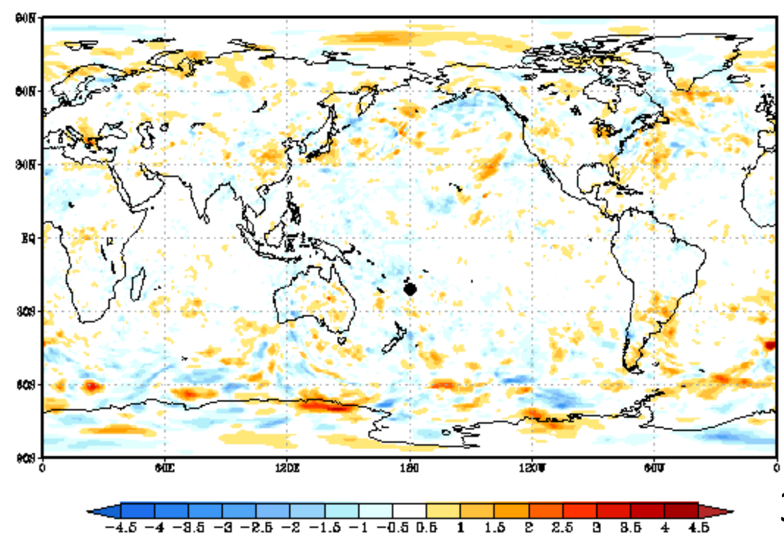
$\delta \mathbf{x}_e^n$

EnKF T Mem001 Recentering ONLY



$\delta \mathbf{x}_r$

EnKF Tinc Mem001 + ADDInflation + Recentering



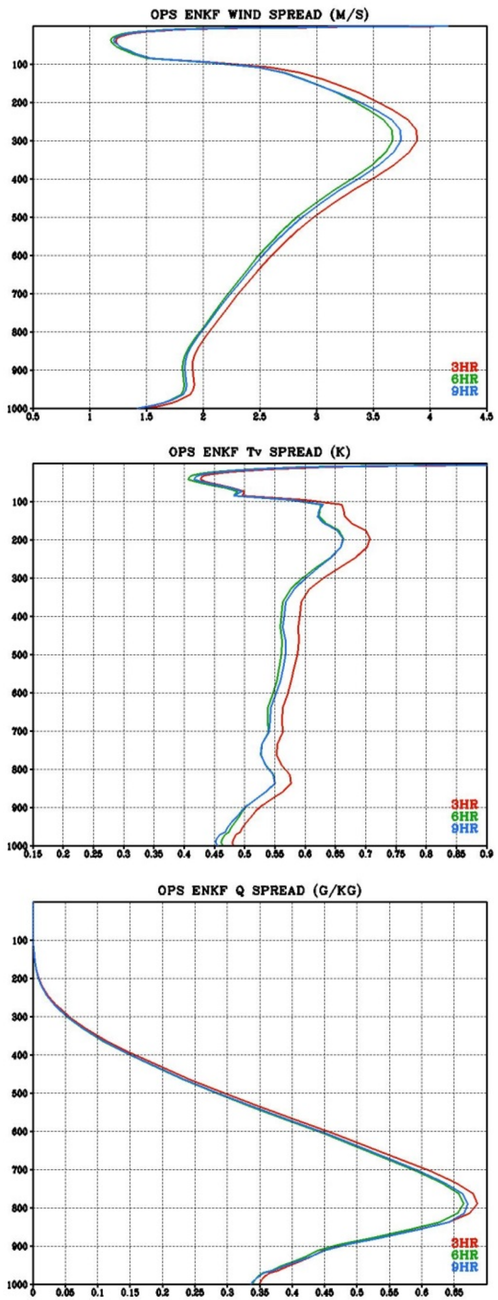
$\delta \mathbf{x}_a^n$

Spread behavior (2014042400)

3HR

6HR

9HR



- Current operations
  - Spread too large
  - Spread decays and recovers

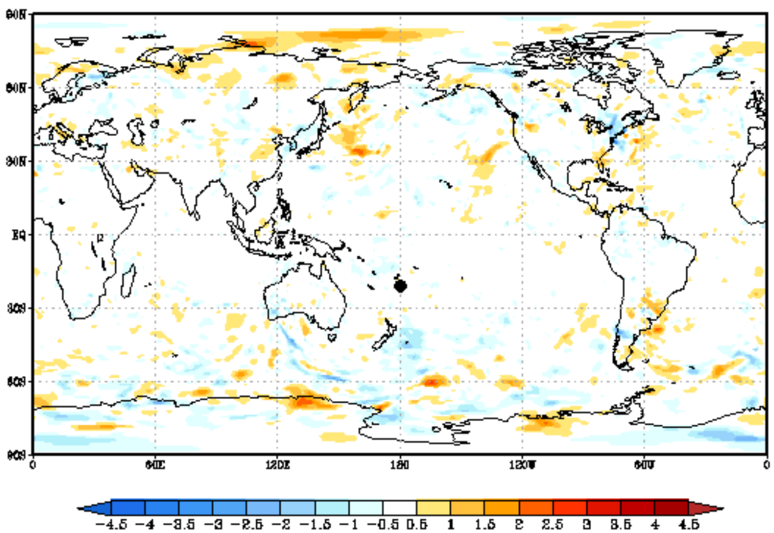
31

Ensemble Member 01 increments

$$\delta \mathbf{x}_a^n = \delta \mathbf{x}_0^n + \delta \mathbf{x}_r + \delta \mathbf{x}_e^n$$

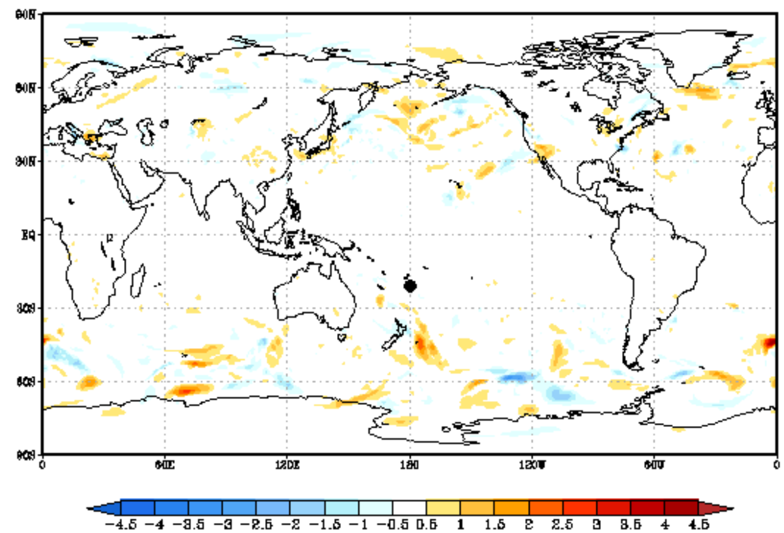
EnKF Tinc Mem001

$\delta \mathbf{x}_0^n$



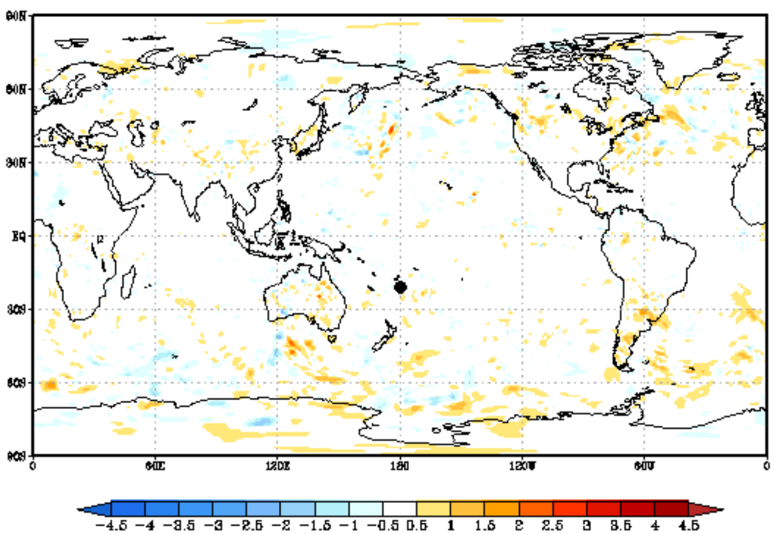
EnKF T Mem001 ADDInflation ONLY

$\delta \mathbf{x}_e^n$



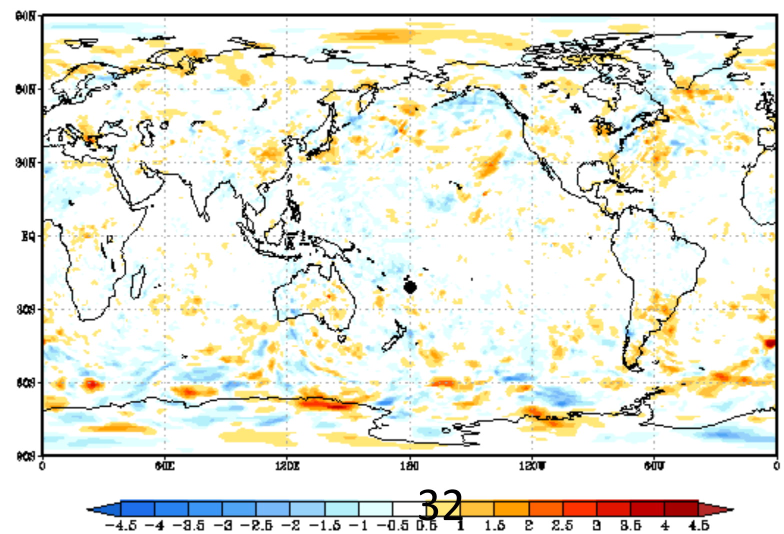
EnKF T Mem001 Recentering ONLY

$\delta \mathbf{x}_r$



EnKF Tinc Mem001 + ADDInflation + Recentering

$\delta \mathbf{x}_a^n$



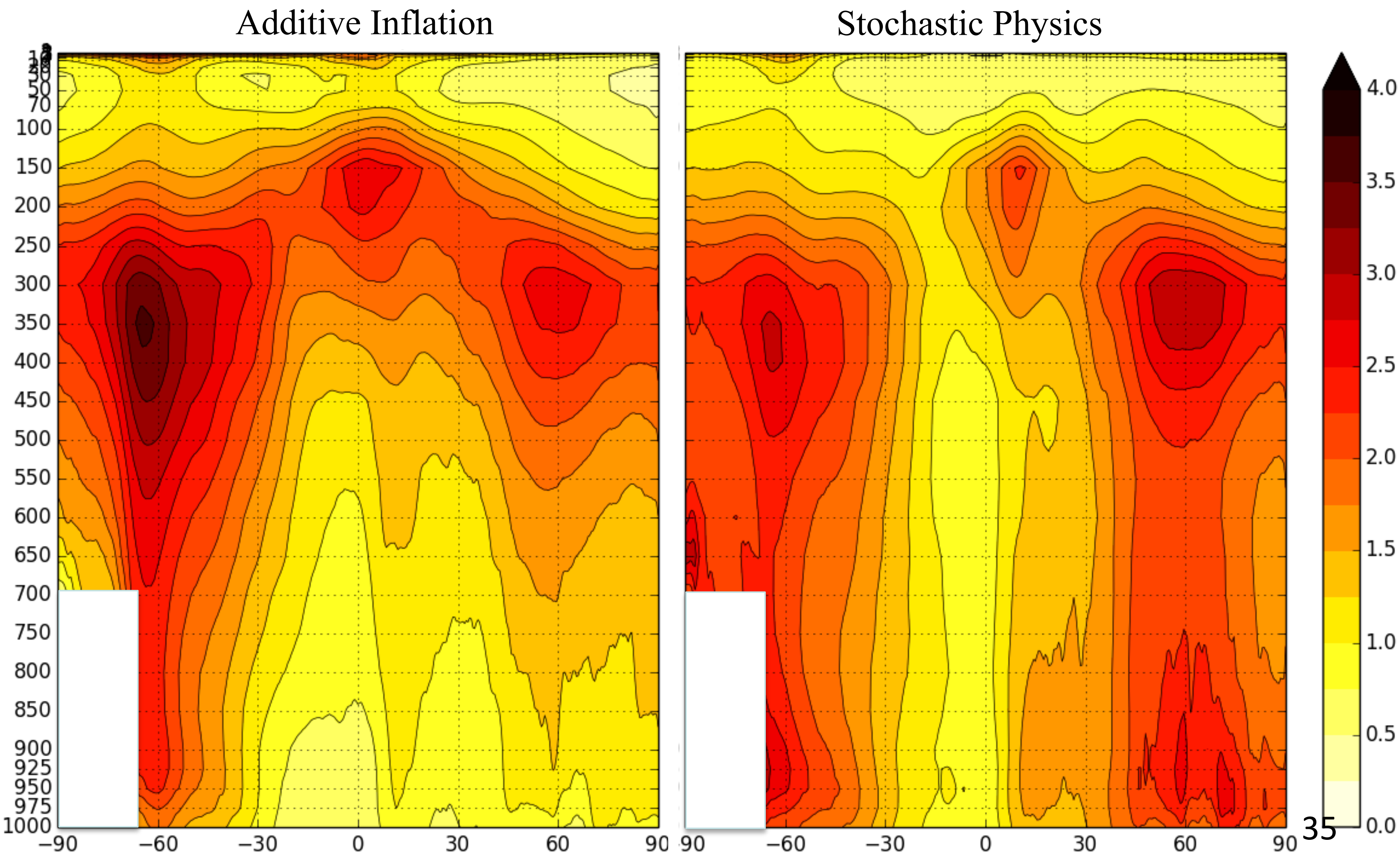
Can we replace the additive inflation by adding stochastic physics to the model?

- Schemes tested:
  - SPPT (stochastically perturbed physics tendencies – ECWMF tech memo 598)
    - *Designed to represent the structural uncertainty of parameterized physics.*
  - SHUM (perturbed boundary layer humidity, based on Tompkins and Berner 2008, *DOI: 10.1029/2007JD009284)*
    - *Designed to represent influence of sub-grid scale humidity variability on the the triggering of convection.*
  - SKEB (stochastic KE backscatter – also see tech memo 598)
  - VC (vorticity confinement, based on Sanchez et al 2012, *DOI: 10.1002/qj.1971).* Can be deterministic and/or stochastic.
    - *Both SKEB and VC aim to represent influence of unresolved or highly damped scales on resolved scales.*
- All use stochastic random pattern generators to generate spatially and temporally correlated noise.

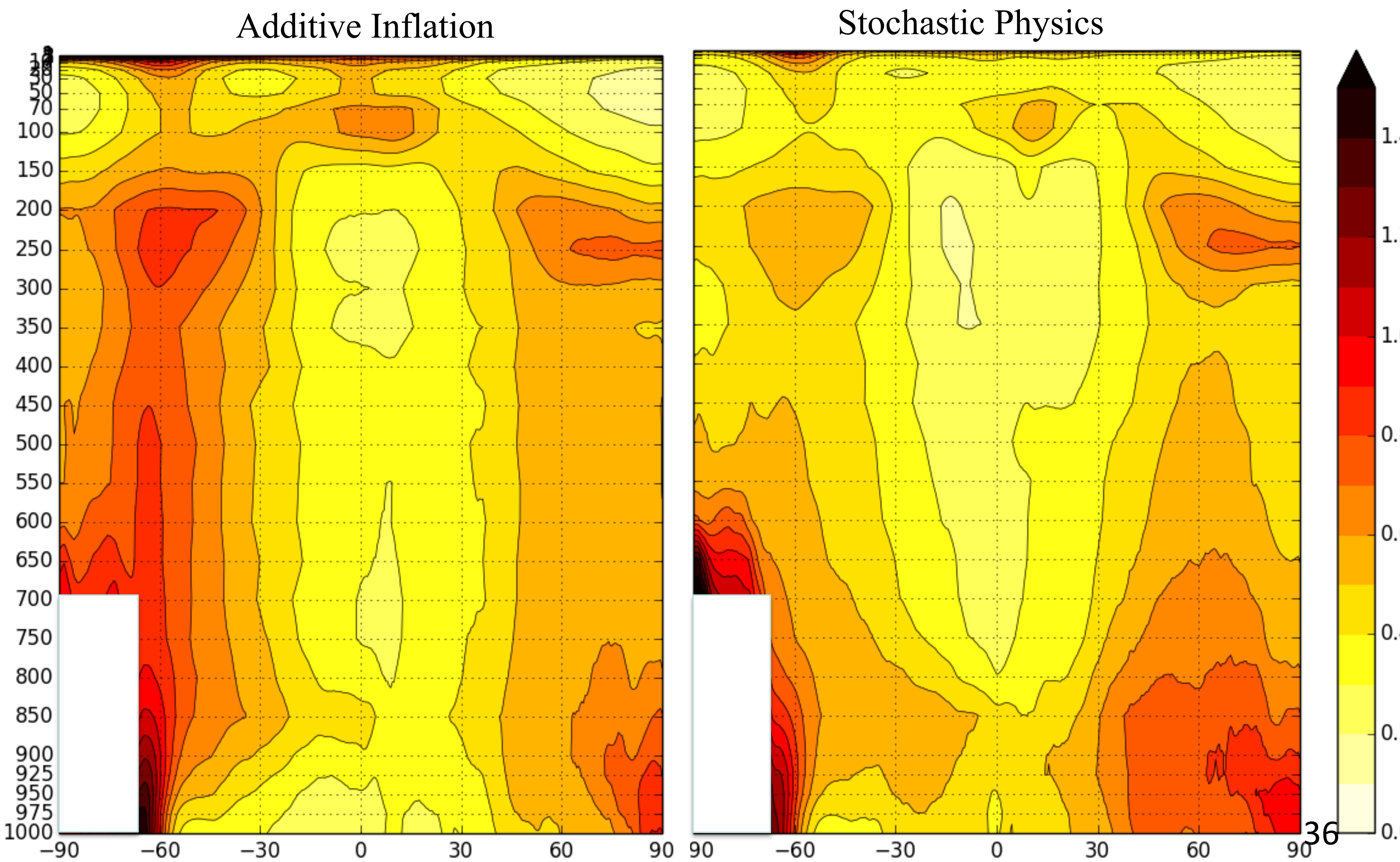
Experiments

- **Control**:
  - As in NCEP ops (using additive inflation), but using semi-lagrangian GFS with T574 ensemble.
- **Expt**:
  - Replace additive inflation with combination of SPPT, SHUM, SKEB and VC. Spatial/temporal scales of 250km/6 hrs for each (except 1000 km/6 hrs for VC). VC purely stochastic. Amplitudes set to roughly match additive inflation spread. Multiplicative inflation as in NCEP ops.
- **Period**: Sept 1 to Oct 15 2013, after 7 day spinup.

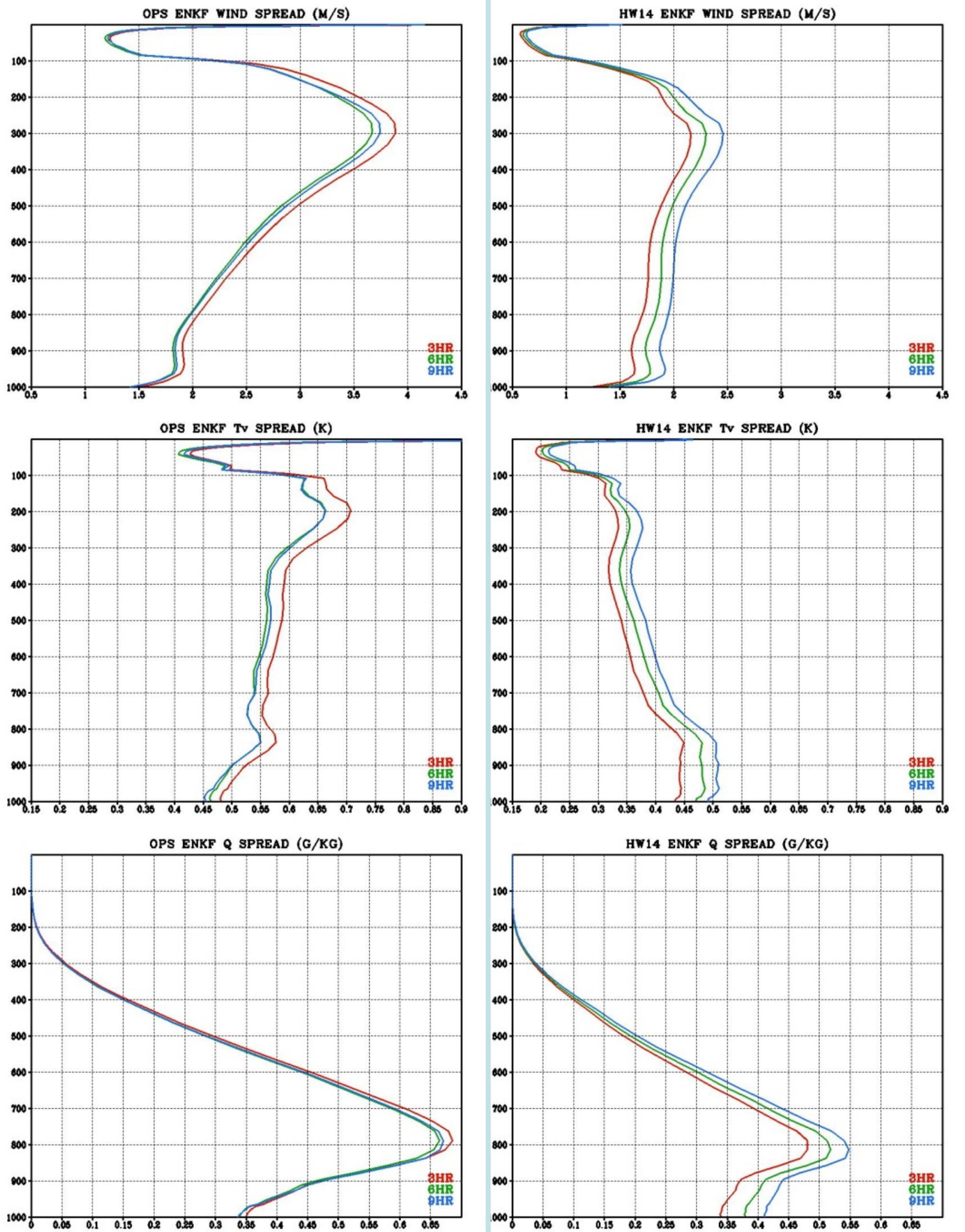
6-hr forecast spread (zonal wind)



6-hr forecast spread (temperature)



Better spread behavior (2014042400)



3HR

6HR

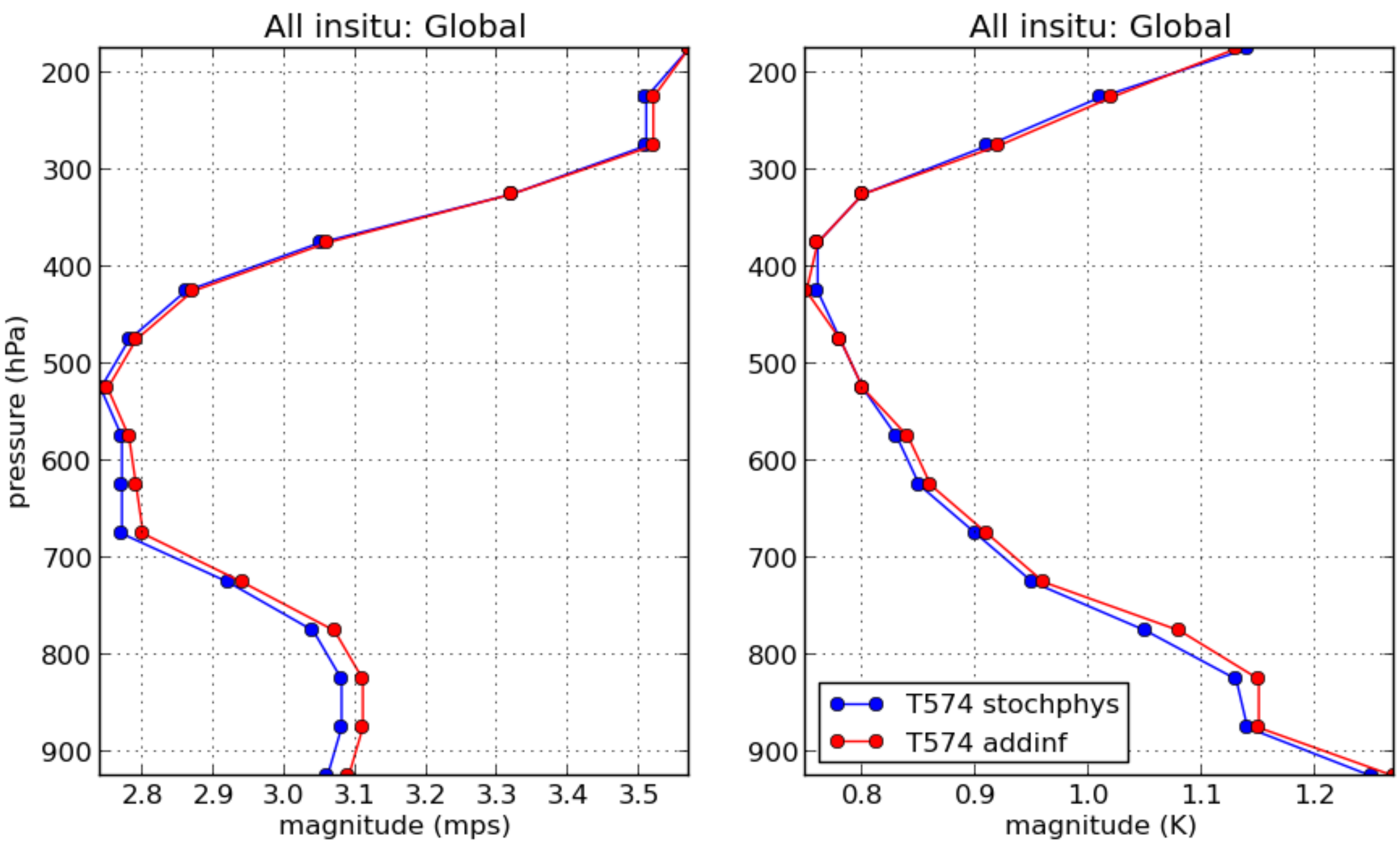
9HR

- Current operations
  - Spread too large
  - Spread decays and recovers
- Stochastic Physics
  - Spread decreased overall (consistent with error estimates)
  - Spread grows through assimilation

Impact on O-F (observation innovation std. dev)

sqrt of $\mathbf{d}_b^o(\mathbf{d}_b^o)^T$ where $\mathbf{d}_b^o = \mathbf{y}^o - H(\mathbf{x}^b)$

Vector Wind (left) and Temp (right) O-F (2013091000-2013101412)



Hybrid 4D-Ensemble-Var [H-4DEnVar]

The 4D EnVar cost function can be easily expanded to include a static contribution

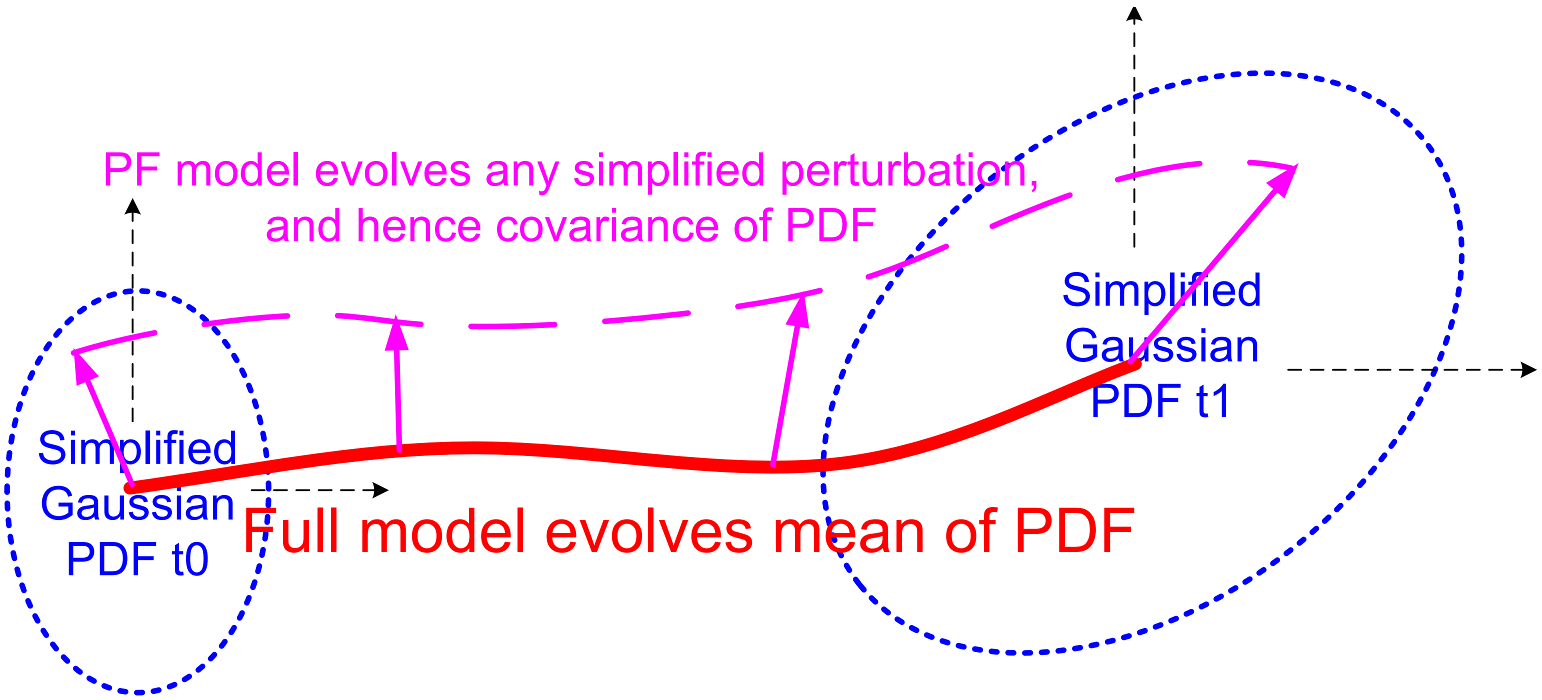
$$J\left(\mathbf{x}_{\mathrm{f}}',\boldsymbol{\alpha}\right)=\beta_{\mathrm{f}}\frac{1}{2}\left(\mathbf{x}_{\mathrm{f}}'\right)^{\mathrm{T}}\mathbf{B}_{\mathrm{f}}^{-1}\left(\mathbf{x}_{\mathrm{f}}'\right)+\beta_{\mathrm{e}}\frac{1}{2}\sum_{n=1}^{N}\left(\boldsymbol{\alpha}^{n}\right)^{\mathrm{T}}\mathbf{L}^{-1}\left(\boldsymbol{\alpha}^{n}\right)+\frac{1}{2}\su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

Where the 4D increment is prescribed exclusively through linear combinations of the 4D ensemble perturbations plus static contribution

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

Here, the static contribution is considered time-invariant (i.e. from 3DVAR-FGAT). Weighting parameters exist just as in the other hybrid variants.

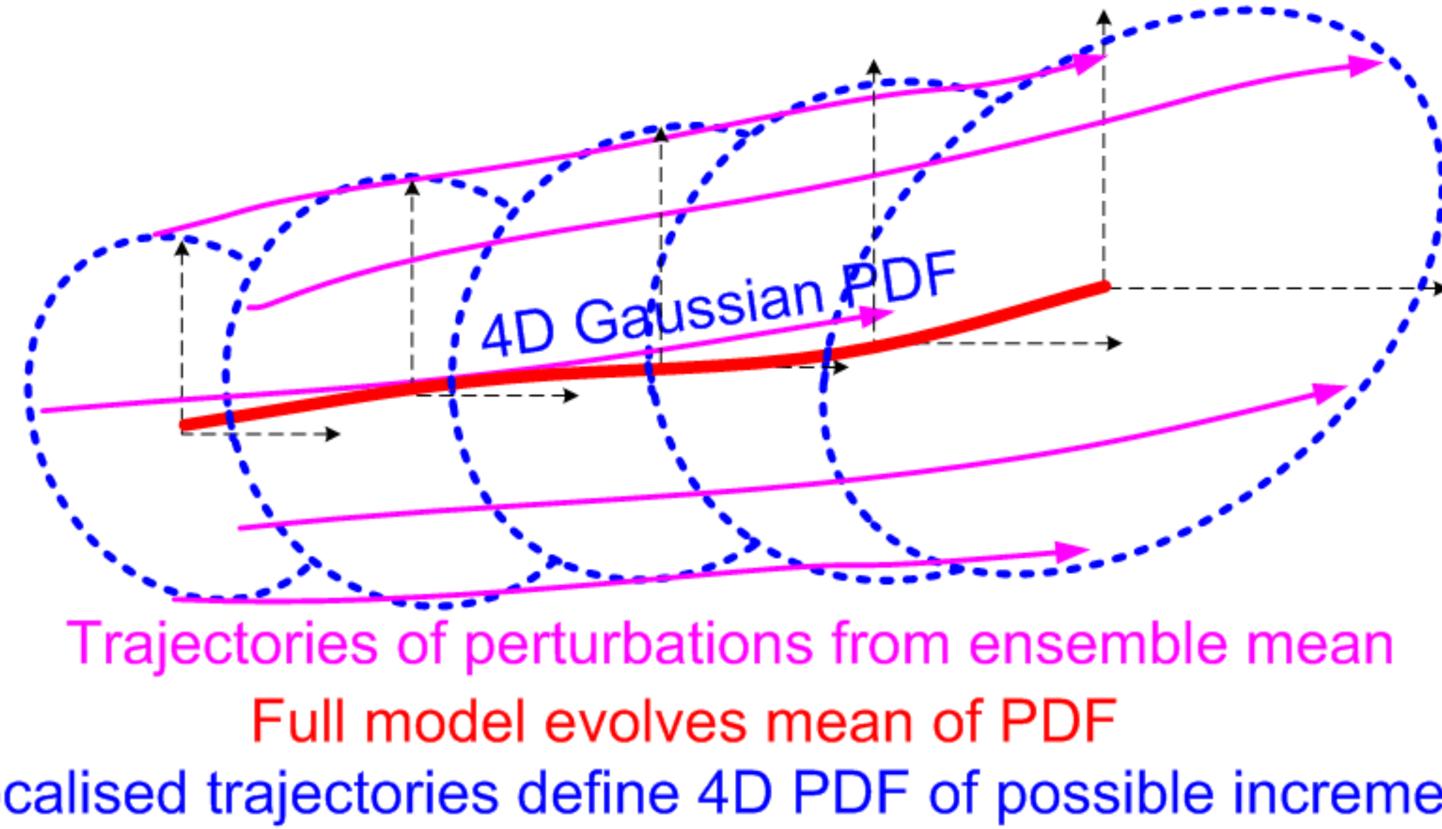
4DVAR



4D analysis increment is a trajectory of the PF model.

Lorenc & Payne 2007

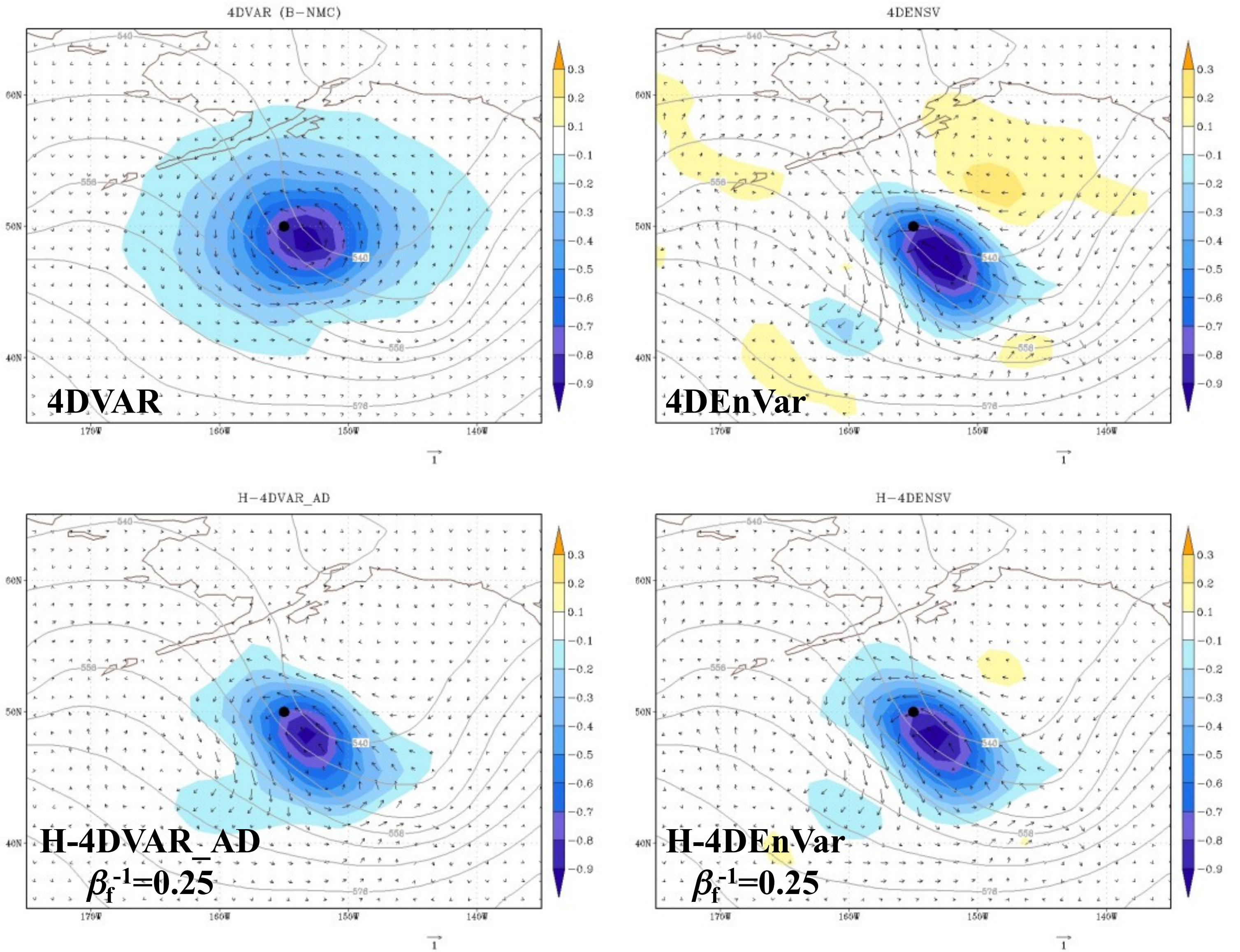
4D EnVar



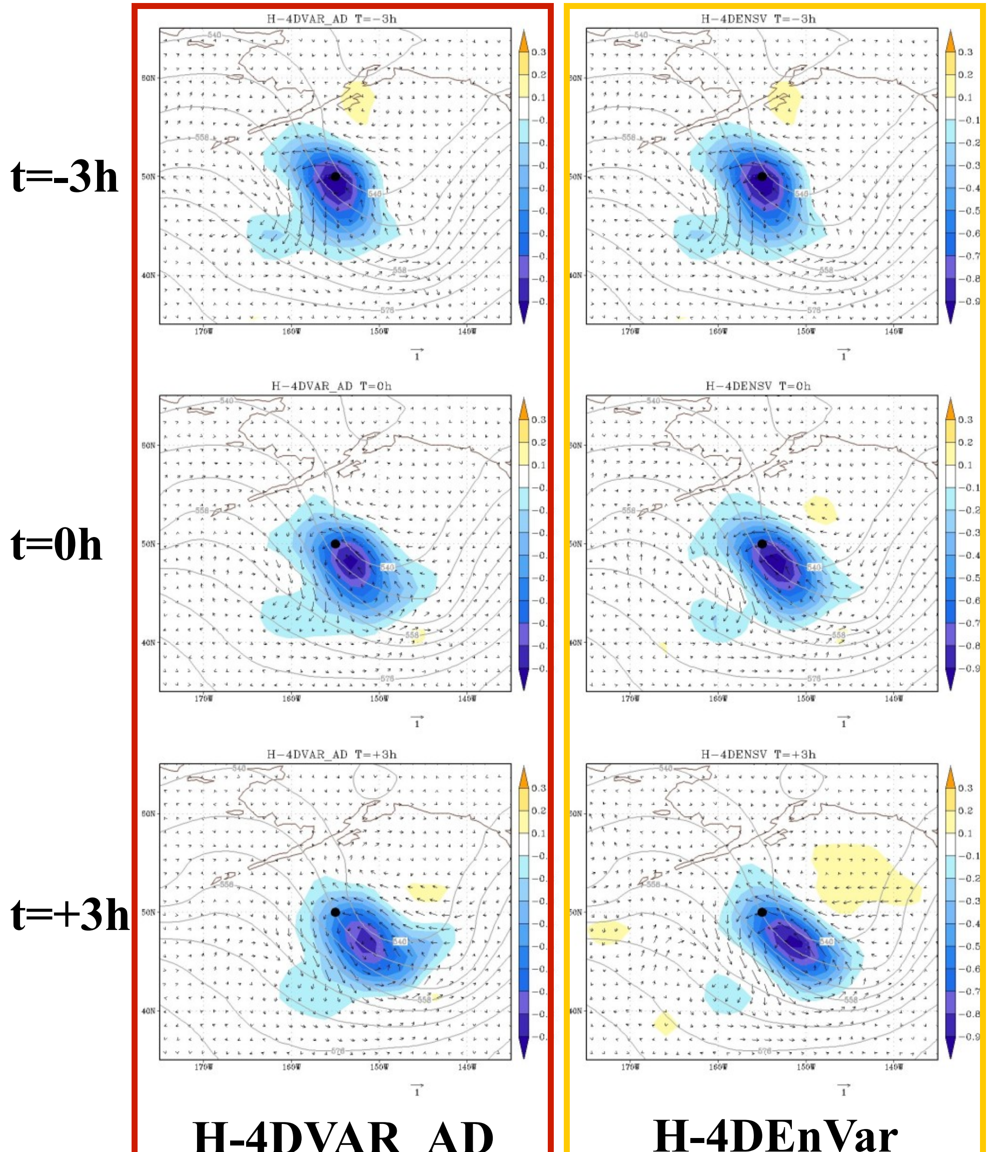
4D analysis is a (localised) linear combination of nonlinear trajectories. It is not itself a trajectory.

Courtesy: Andrew Lorenc

Single Observation (-3h) Example for 4D Variants



Time Evolution of Increment

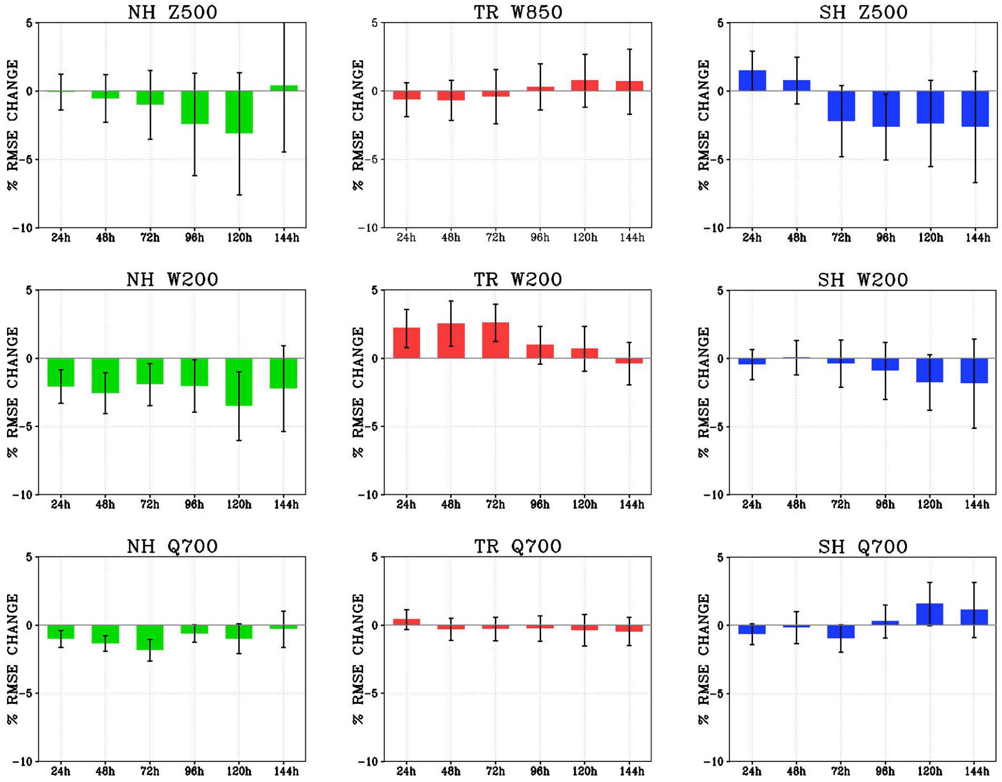


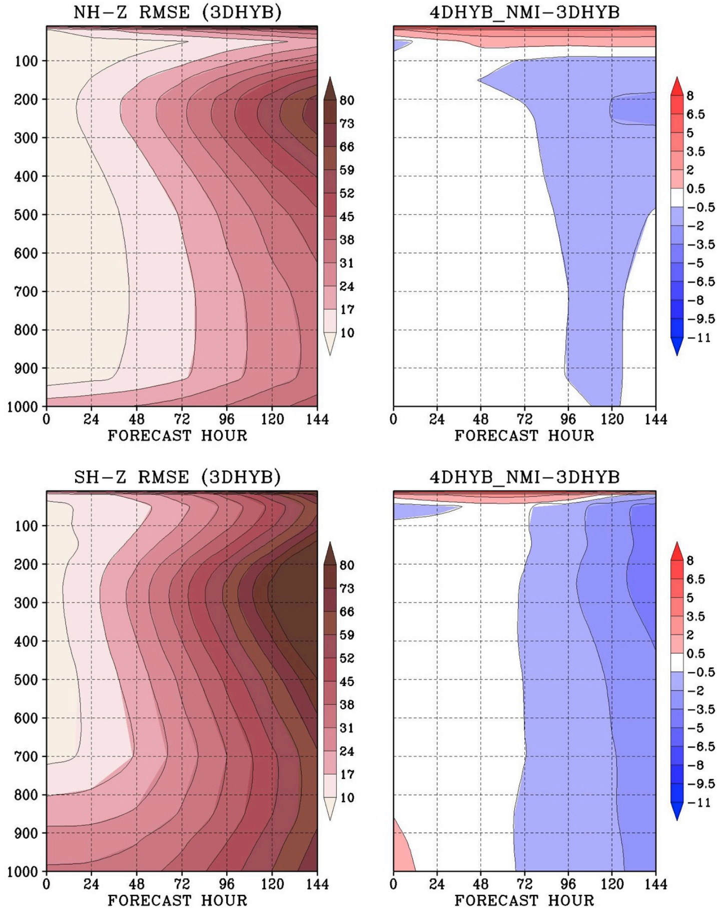
Solution at beginning of window same to within round-off (because observation is taken at that time, and same weighting parameters used)

Evolution of increment qualitatively similar between dynamic and ensemble specification

** Current linear and adjoint models in GSI are computationally unfeasible for use in 4DVAR other than simple single observation testing at low resolution

OSSE Cycling Experiments
Hybrid 4DEnVar relative to 3DEnVar
Kleist and Ide 2014 (MWR)

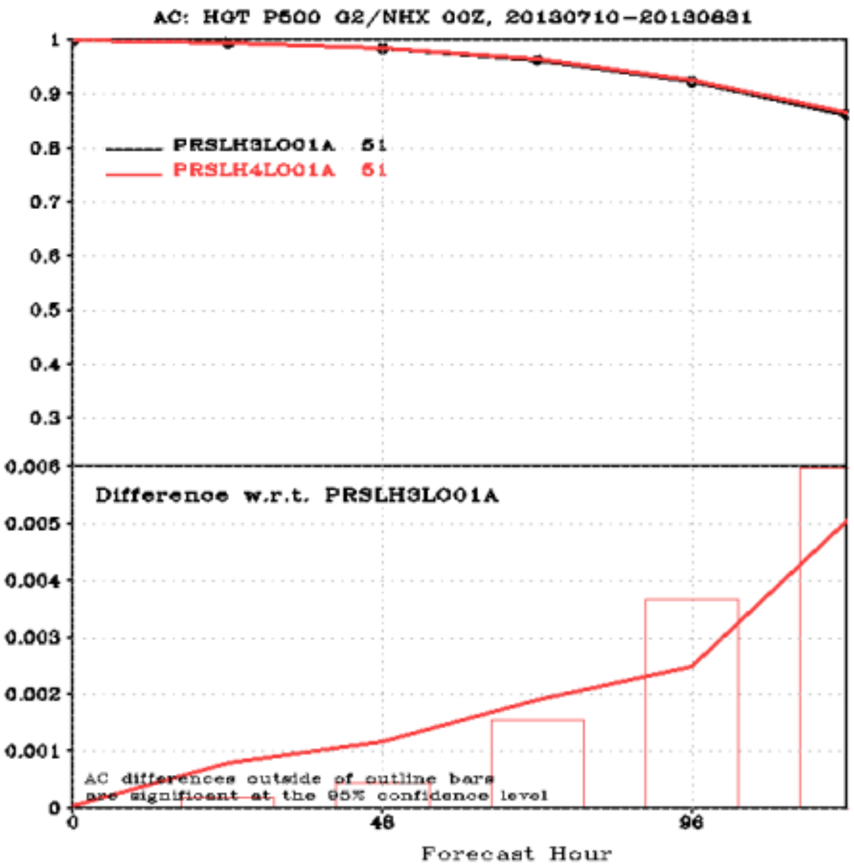




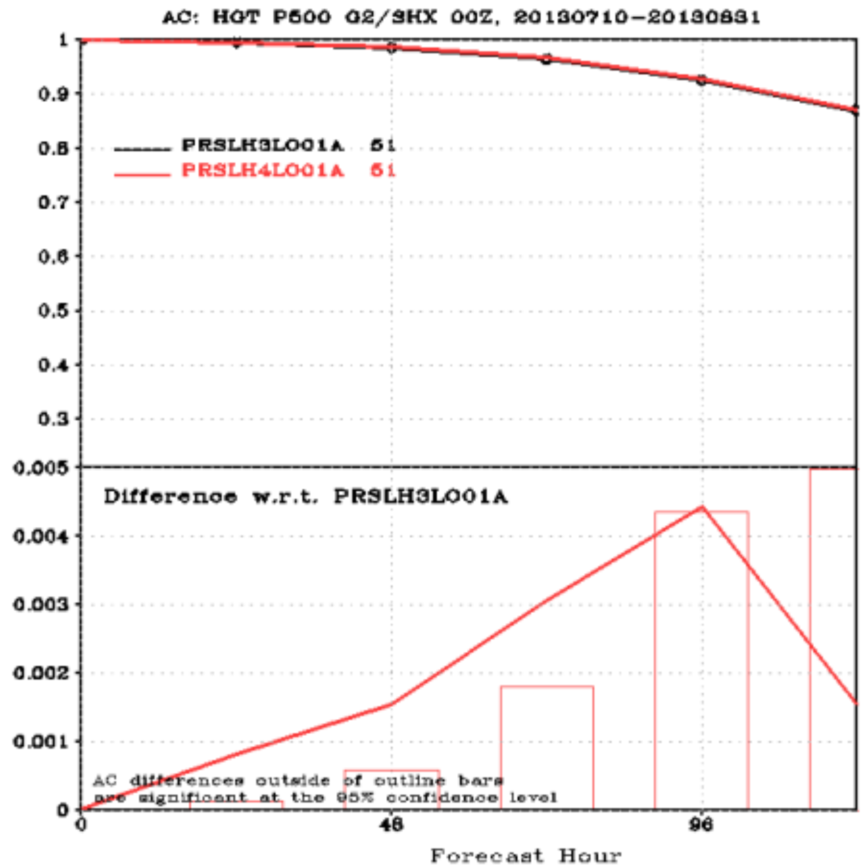
Recent Low Resolution GFS/GDAS Experiments with real observations

- Due to the impending implementation of the new Semi-Lagrangian model, tests needed to be done with new model, configuration, etc.
- T670 Semi-Lagrangian with an 80 member T254 Semi-Lagrangian ensemble
  - Similar ratio to what is to be implemented with T1543/T574 system
  - Experiments with both additive inflation and stochastic physics as replacement
    - Stoch. Physics is resolution sensitive, requires tuning.
- Compare hybrid 3DEnVar to 4DEnVar (minimal additional tuning such as localization, weights, etc.)

3D v 4D hybrid in SL GFS

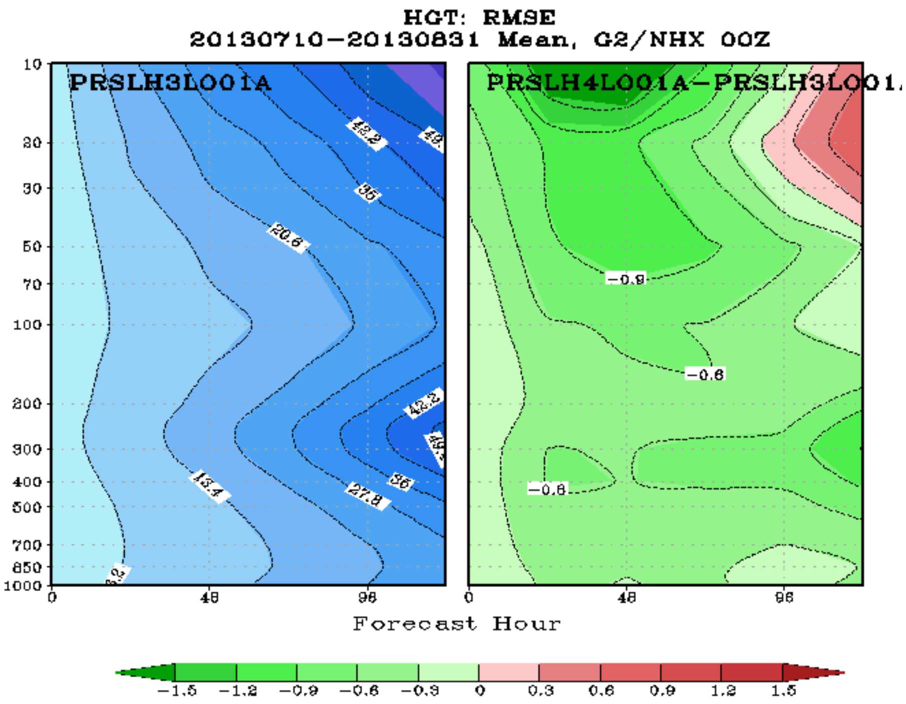


Northern Hemisphere

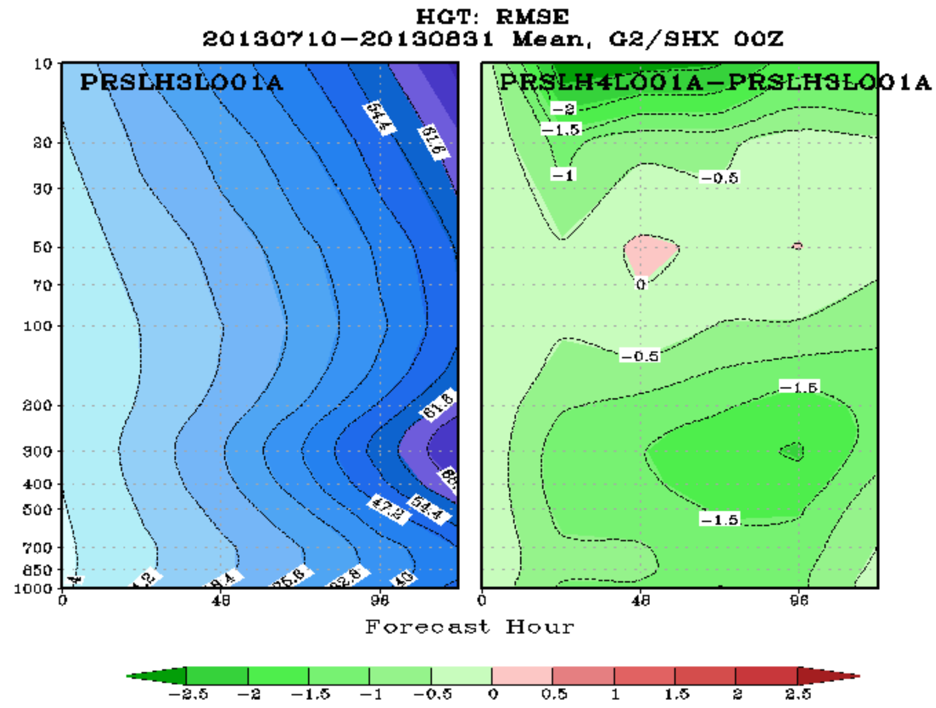


Southern Hemisphere

3D v 4D hybrid in SL GFS



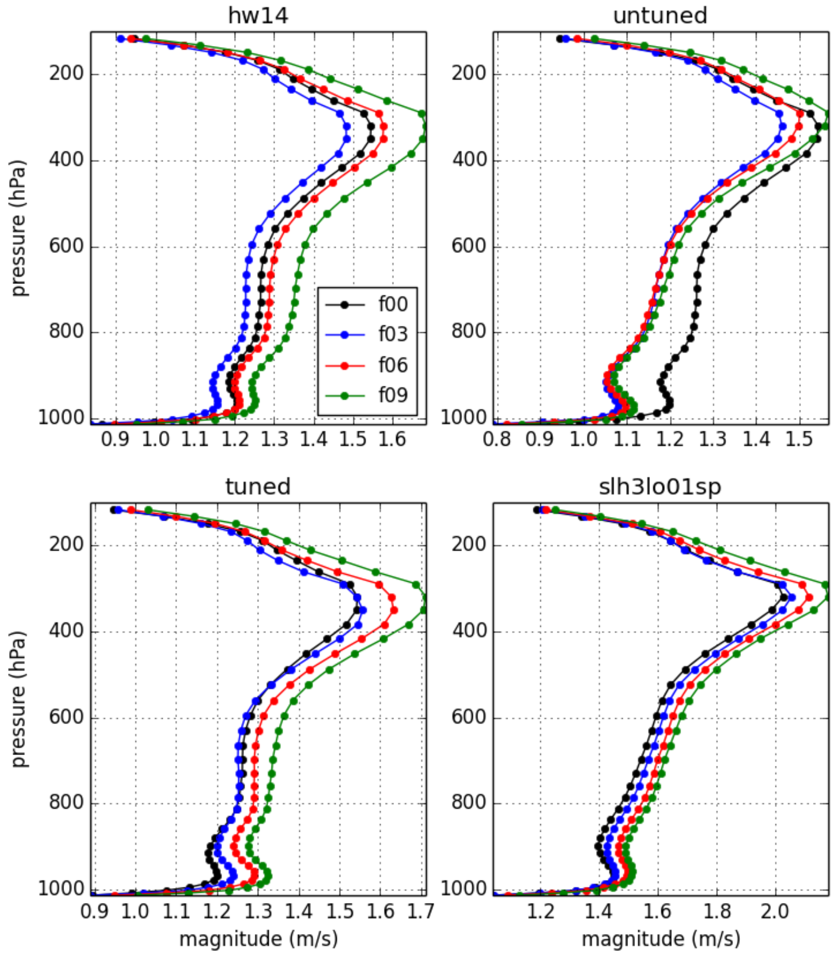
Northern Hemisphere



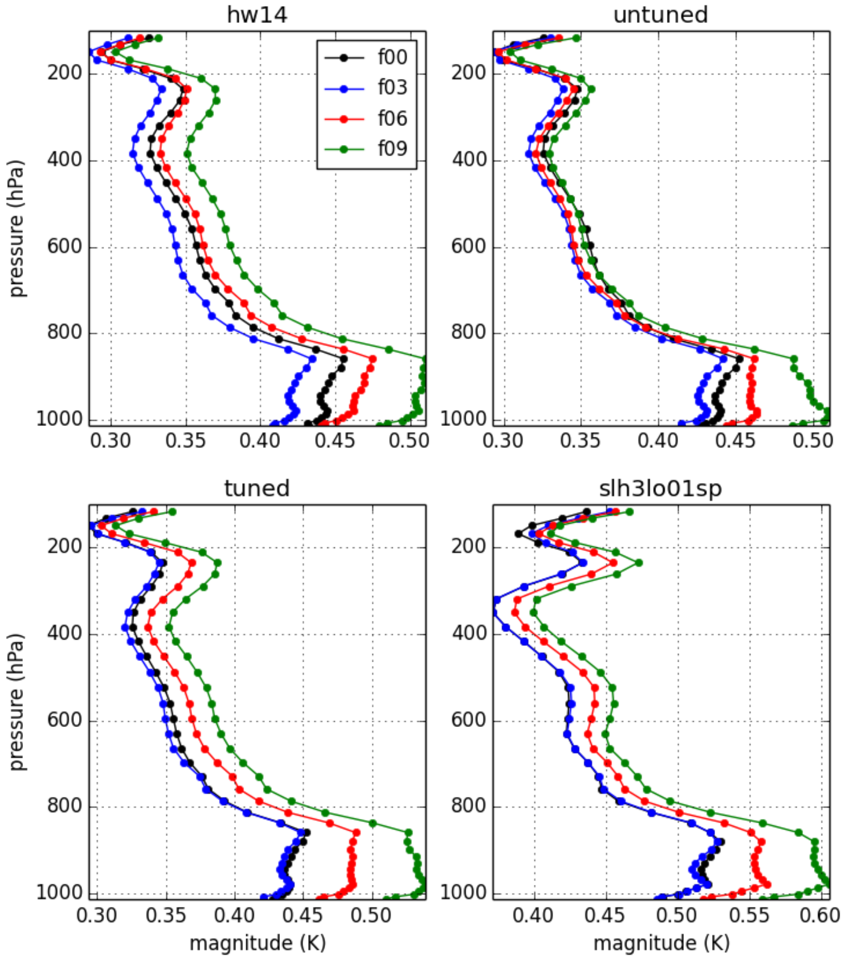
Southern Hemisphere

Stochastic Physics Tuning

Ens. Spread in U Wind
(Global)



Ens. Spread in Temp.
(Global)



Innovation statistics from 3D and 4D runs with tuned stoch. physics (preliminary)

UV, Counts and O-F, 3D (RMSE1) 4D (RMSE2)

lev1-lev2	COUNTS1	COUNTS2	RMSE1	RMSE2
1200-1000	16525	16527	**3.14**	3.16
1000- 900	51483	51485	3.27	3.27
900- 800	47937	47938	**3.15**	3.19
800- 600	93195	93192	3.26	**3.25**
600- 400	96109	96108	3.60	**3.58**
400- 300	75313	75315	3.84	3.84
300- 250	22882	22880	4.25	**4.23**
250- 200	48625	48624	4.18	**4.11**
200- 150	56236	56238	4.25	**4.18**
150- 100	101212	101218	4.31	**4.30**
100- 50	134273	134267	4.48	**4.45**
2000- 0	998436	998445	4.07	**4.04**

T, Counts and O-F, 3D (RMSE1) 4D (RMSE2)

lev1-lev2	COUNTS1	COUNTS2	RMSE1	RMSE2
1200-1000	19237	19242	**1.92**	1.93
1000- 900	48727	48739	1.74	**1.73**
900- 800	46272	46275	**1.19**	1.20
800- 600	90544	90549	0.98	**0.97**
600- 400	111865	111863	0.90	**0.89**
400- 300	78711	78727	0.85	0.85
300- 250	22780	22782	1.07	**1.05**
250- 200	43890	43899	1.31	**1.27**
200- 150	44370	44382	1.20	**1.18**
150- 100	76347	76366	1.30	1.30
100- 50	72433	72448	1.88	**1.86**
2000- 0	789836	789967	1.38	**1.37**

2013070106-2013071418

4D yielding consistent improvements in O-F with stochastic physics.
Experiments and evaluation are still ongoing to assess forecast impact.

O-F comparison (3D/4D) similar in additive inflation context

Summary

- The "hybrid" EnVar option in GSI uses perturbations from an ensemble of short term forecasts to better estimate the background error covariance term.
  - Added expense (mostly IO)
  - Added complexity
    - Running/updating ensemble
    - Additional DA component (EnKF)?

- Although any ensemble can be used, should be representative of forecast/background error
  - Feel effects of observations

Summary (2)

- Implementation of hybrid 3D EnVar lead to significant improvements in deterministic model skill
- Working toward extending 3D hybrid to 4D in NCEP operations (non-adjoint 4D EnVar)
- Within 4D, much work to do:
  - Outer loop, initialization, improve climatological contribution and/or increase ensemble size, computational aspects including I/O
    - Some of the I/O issues can be mitigated by the fact that the analysis uses the ensemble forecasts from previous cycle. Can employ ensemble post-processor.
- **4D EnVar Target: Late 2015/early 2016 (some of this is machine dependent as NCEP gets their next upgrade)**