(4D) Hybrid EnVar Data Assimilation: Initialization, variable transforms, outer loops

Daryl Kleist

University of Maryland, College Park
Department of Atmospheric and Oceanic Science

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

National Central University, Taiwan – 4 November 2014

Outline

- Data Assimilation Introduction
  - Hybrid
- Initialization and 4DIAU
- Perturbation variables in EnVar
- Outer Loops
- Scale-dependent weighting

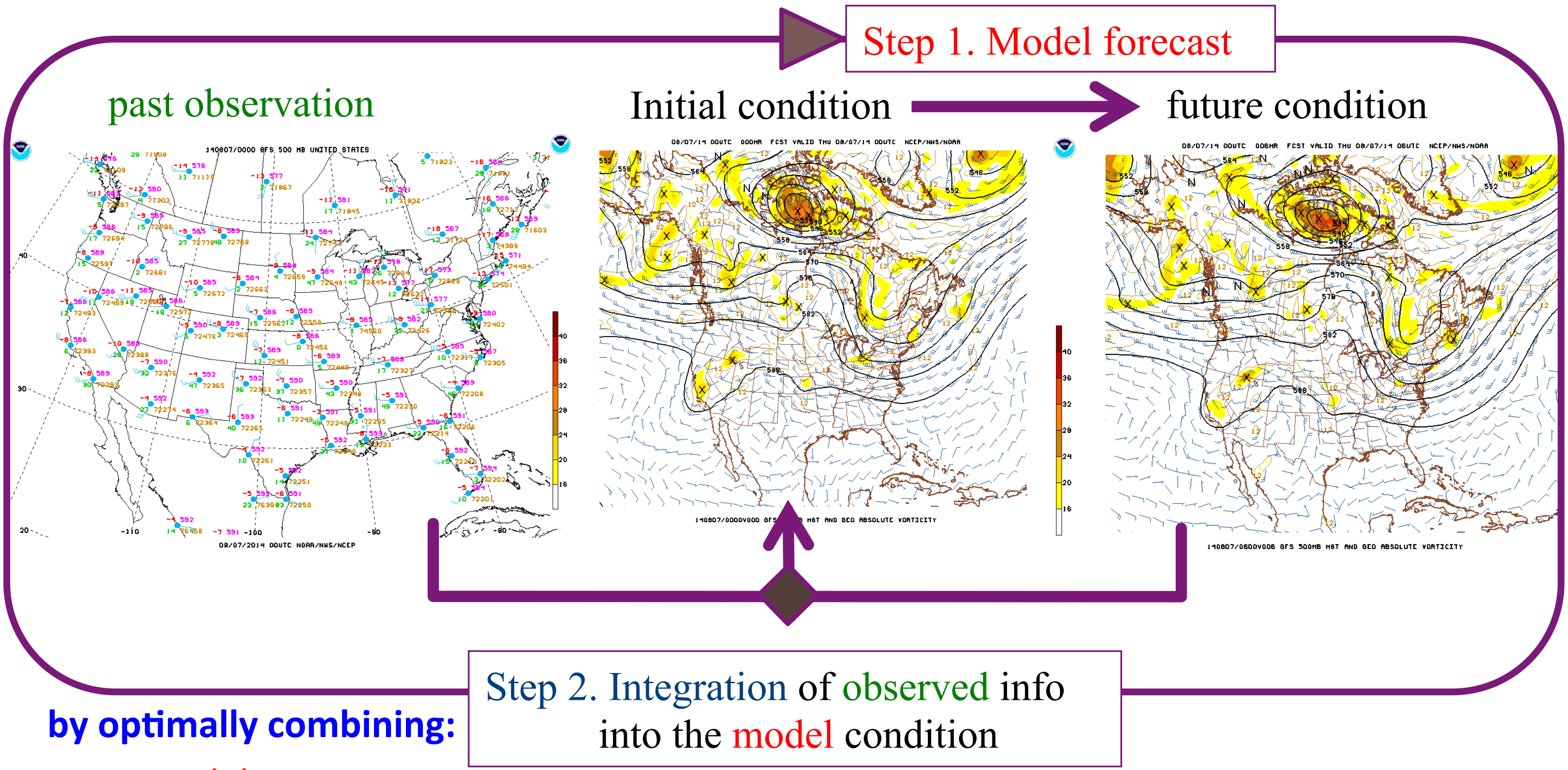
Data Assimilation

- 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**)



by optimally combining:

- Model
- Observations

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).

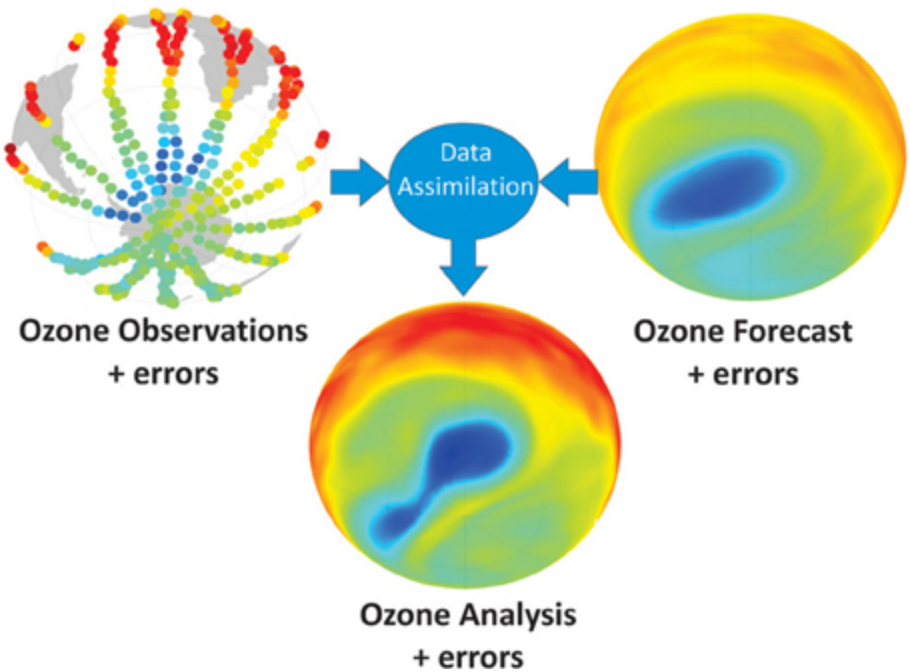


Image courtesy of: AUTHOR=Lahoz William Albert, Schneider Philipp, Data Assimilation: Making Sense of Earth Observation , Frontiers in Environmental Science (DOI=10.3389/fenvs.2014.00016)

DA Algorithms

- Most operational NWP centers use variational (maximum likelihood) algorithms such as 3DVAR, 4DVAR
  - Computationally efficient (for now)
  - Error covariance full rank, static estimate (with typically poor multivariate correlations)

- Move toward ensemble-based estimates of error covariances
  - Time evolving, flow dependent estimate
  - Use small ensemble to represent high dimensional state
  - Relies on well calibrated model to prescribe full multivariate covariances

Hybrid and EnVar Algorithms

- HYBRID
  – Used to describe an algorithm that uses a combined error covariance estimate (“static” plus ensemble)

- ENVAR
  – Used to describe variational-based method (maximum likelihood) that incorporates ensemble perturbations directly into the solver

- 4DVAR versus 4DENVAR
  – 4DVAR uses dynamic model to propagate (TL and AD) information forward and backward within a window
  – 4DEnVar uses 4D ensemble perturbations to prescribe piece-wise trajectory

Hybrid Variational-Ensemble

- Incorporate ensemble perturbations directly into variational cost function through extended control variable
  - Lorenc (2003), Buehner (2005), Wang (2010), etc.

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

$$\mathbf{x}'_\mathrm{t} = \mathbf{x}'_\mathrm{f} + \sum_{n=1}^{N} \left(\boldsymbol{\alpha}^n \circ \mathbf{x}_\mathrm{e}^{\ n} \right)$$

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

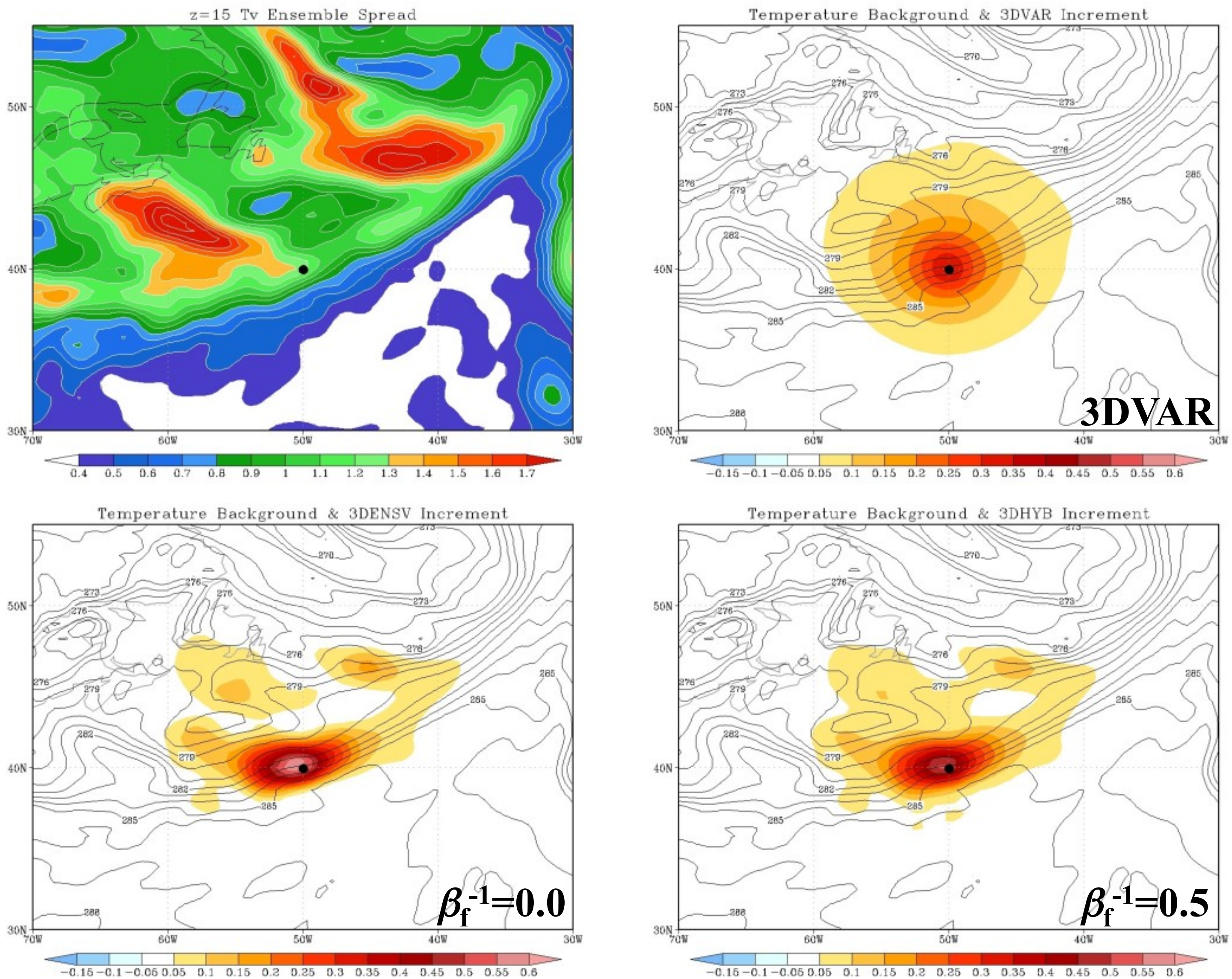
$\mathbf{x}_\mathrm{t}'$: (total increment) sum of increment from fixed/static **B** ($\mathbf{x}_\mathrm{f}'$) and ensemble **B**

α^n: extended control variable; $\mathbf{x}_k^\mathrm{e}$:ensemble perturbations

- analogous to the weights in the LETKF formulation

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

Single Temperature Observation



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

The cost function can be expanded

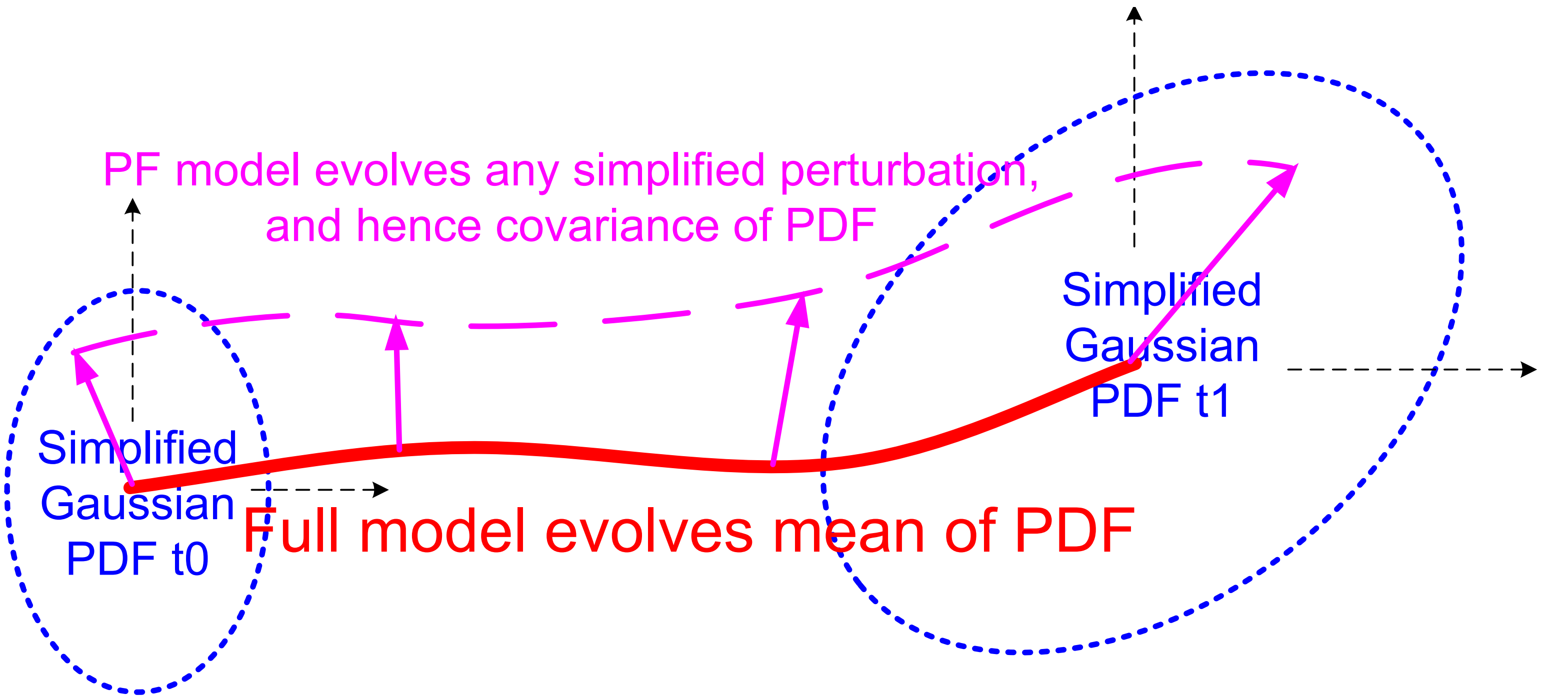
$$J(\mathbf{x}'_\mathrm{f},\boldsymbol{\alpha}) = \beta_\mathrm{f}\frac{1}{2}(\mathbf{x}'_\mathrm{f})^\mathrm{T}\mathbf{B}_\mathrm{f}^{-1}(\mathbf{x}'_\mathrm{f}) + \beta_\mathrm{e}\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{\alpha}^n)^\mathrm{T}\mathbf{L}^{-1}(\boldsymbol{\alpha}^n) + \frac{1}{2}\sum_{k=1}^{K}(\mathbf{H}_k\mathbf{x}'_k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

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

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

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

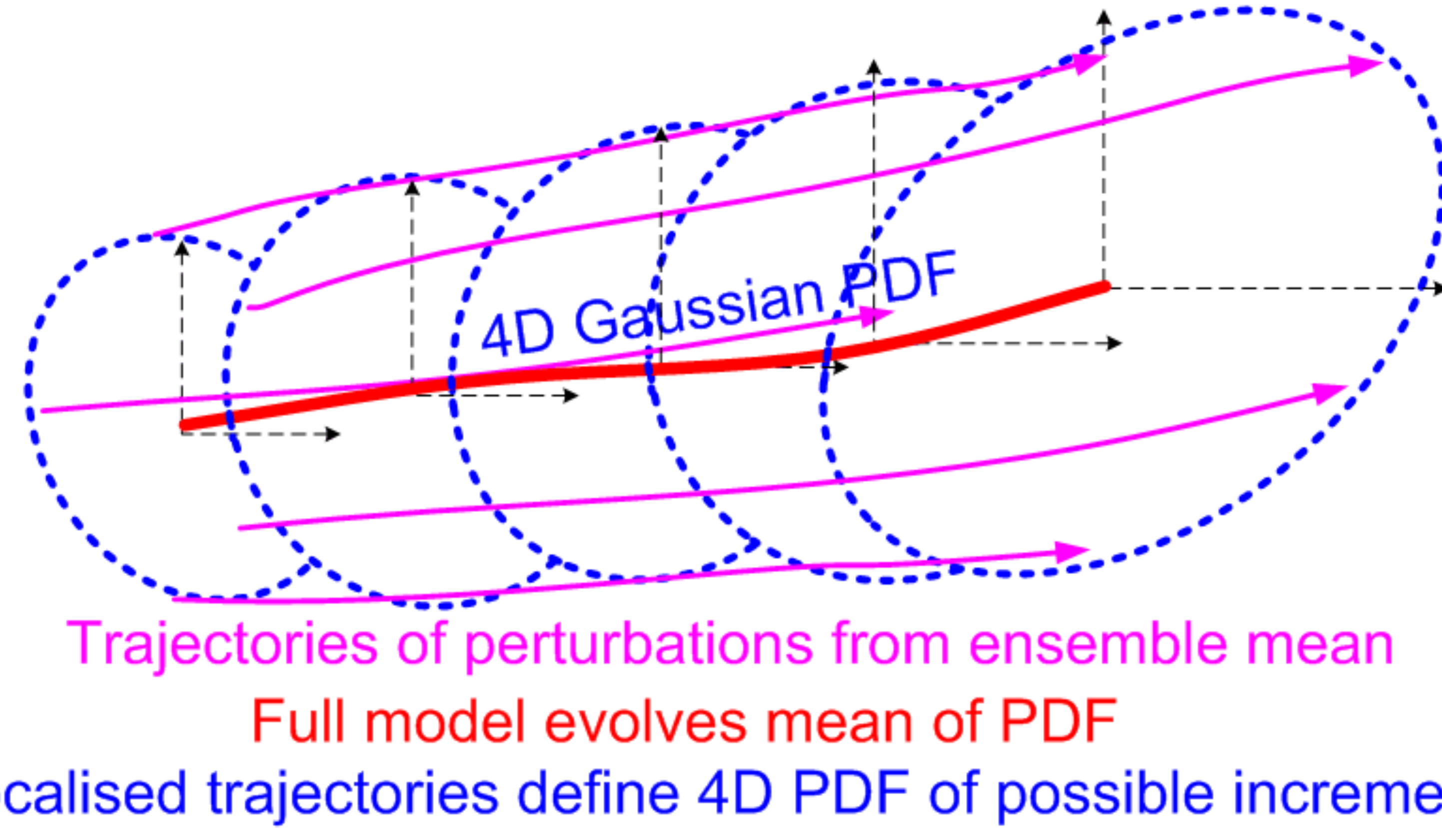
4DVAR



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

Lorenc & Payne 2007

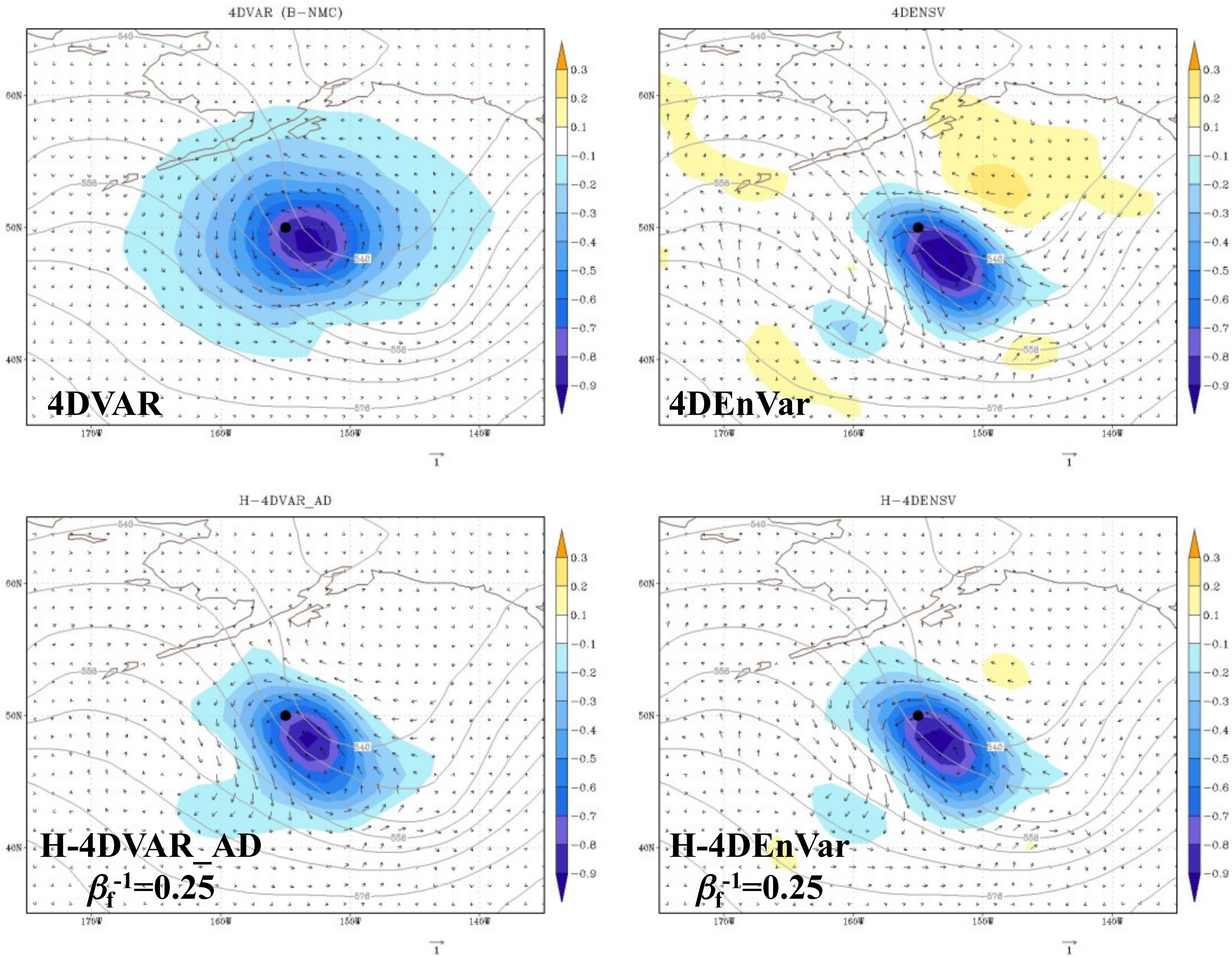
4D EnVar



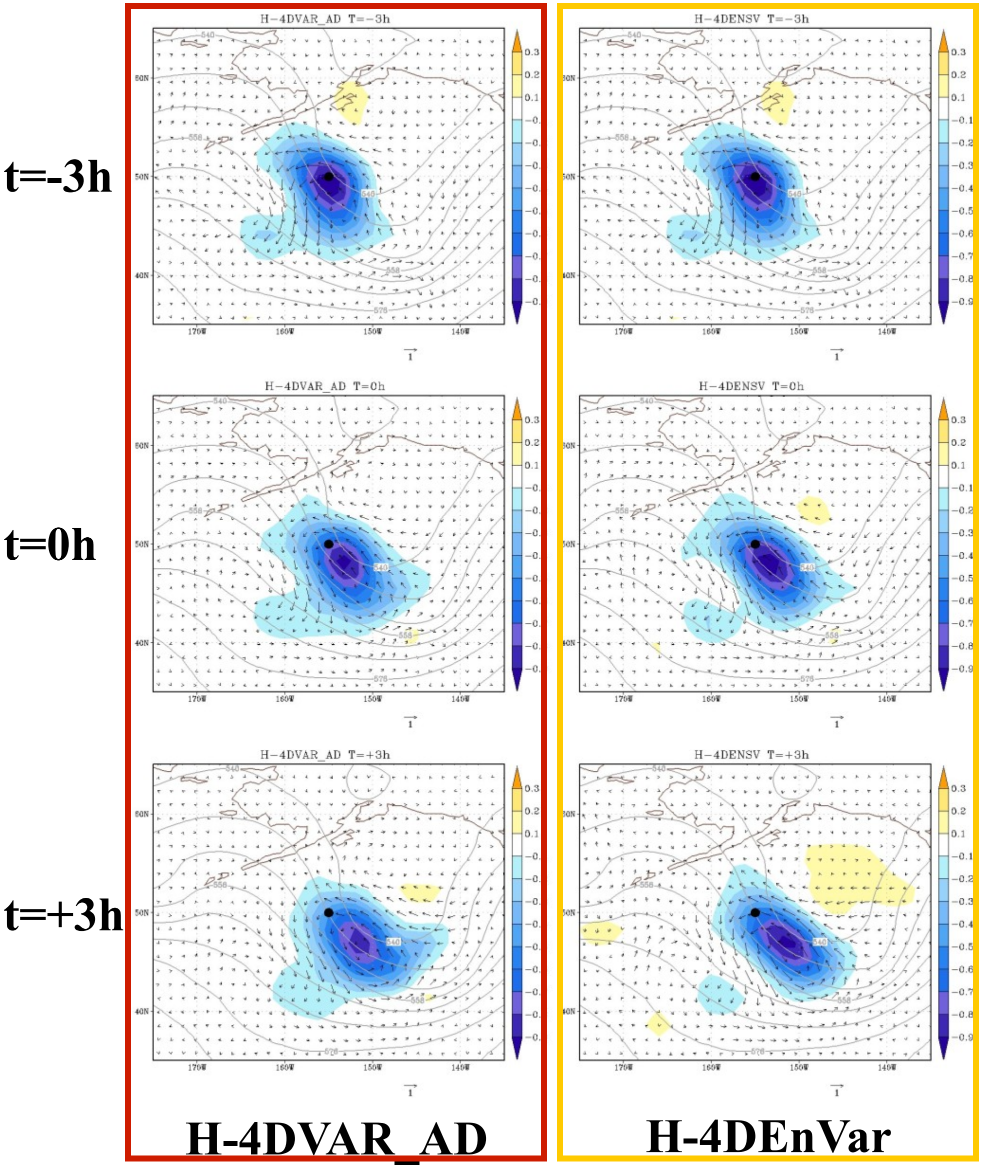
4D analysis is a (localised) linear combination of nonlinear trajectories. It is not itself a trajectory.

Courtesy: Andrew Lorenc

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



Time Evolution of Increment

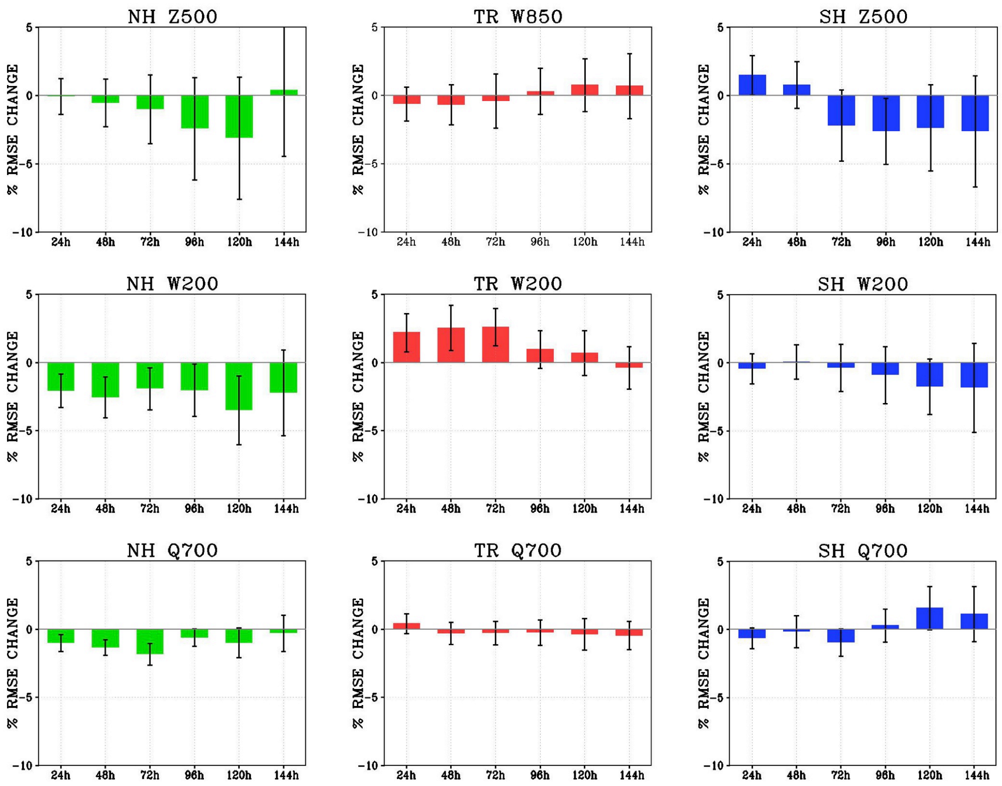


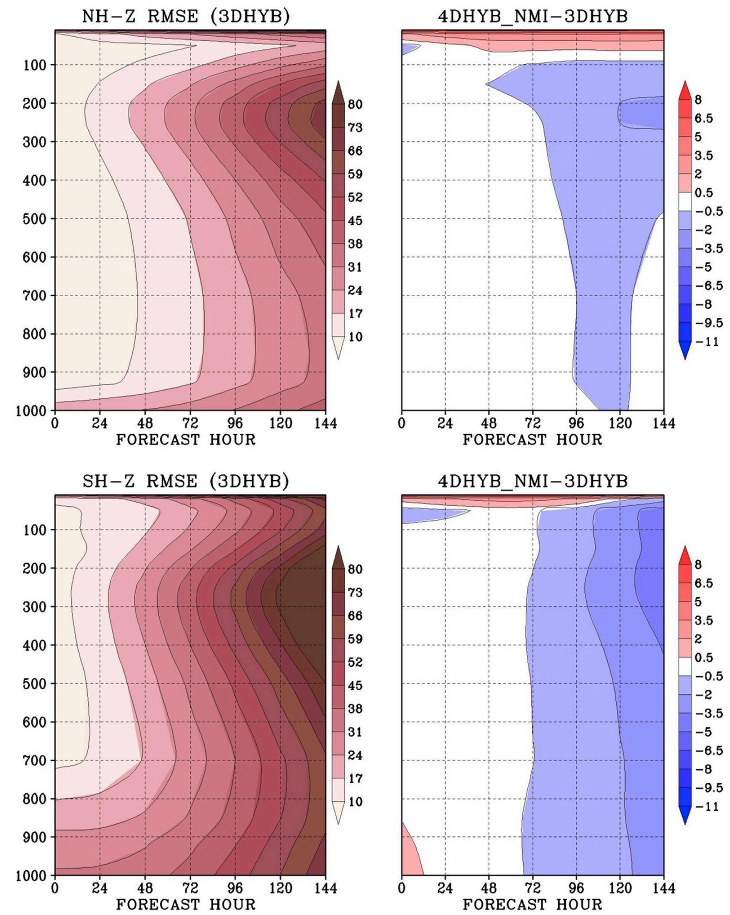
Solution at beginning of window same to within round-off (because observation is taken at that time, and same weighting parameters used)

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)





“Need for initialization”

- Data assimilation process is typically performed intermittently (not continuously)
- Background error specification may be poor
  - Missing correlations in covariance model, inappropriate balances applied
- Model may not be able to properly ingest assimilation-prescribed initial-conditions

Is "noise" important for data assimilation and NWP?

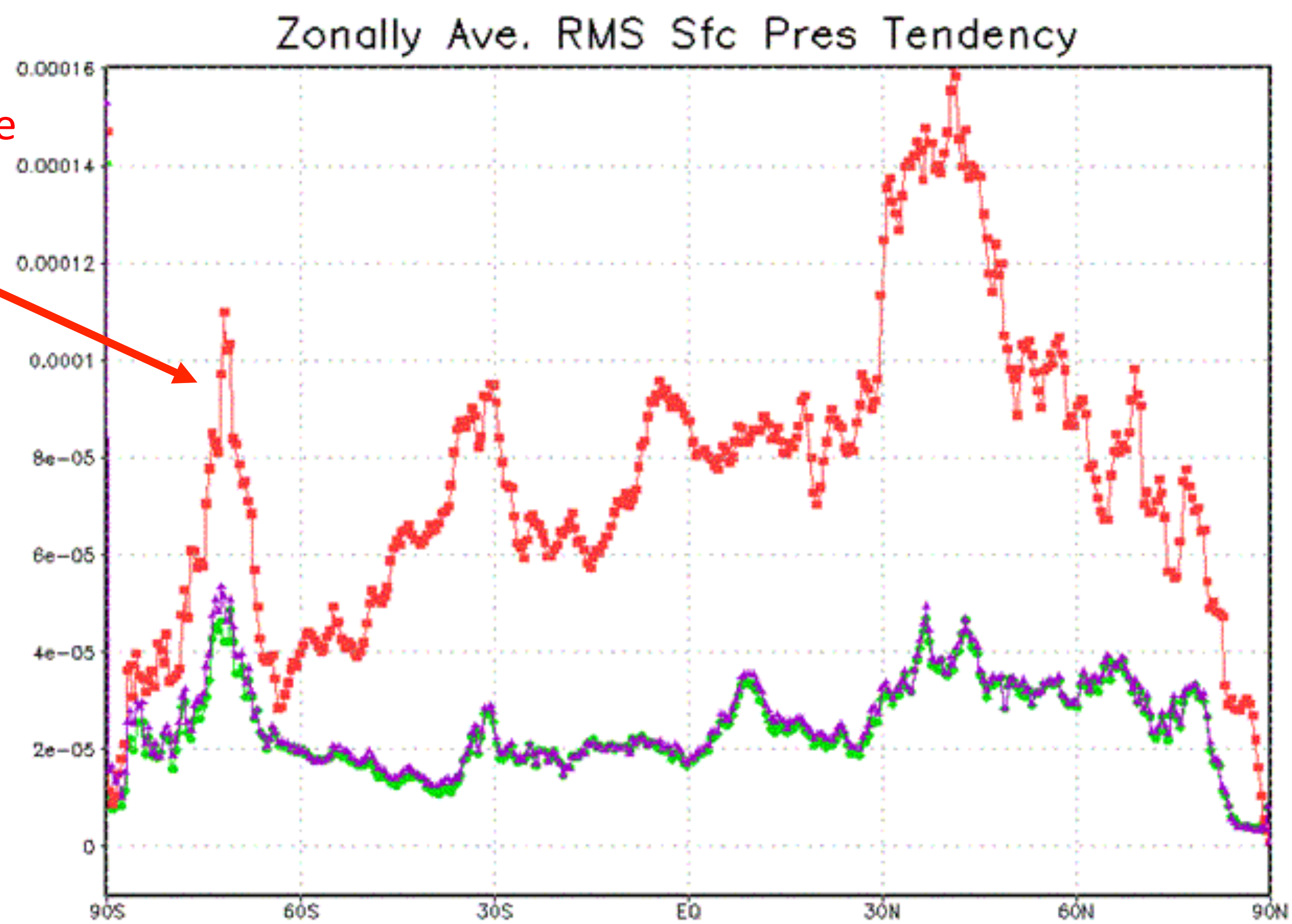
- Fast waves in the NWP system require unnecessary short time steps: inefficient use of computer time

- Gravity waves add high frequency noise to the assimilation system resulting in:
  - rejection of correct observations
  - poor use of observations
    - e.g. deriving wind field properly from satellite radiance observations
  - noisy forecasts with e.g. unrealistic precipitation
    - Spin-up and Spin-down

- Noise in DA system can accumulate through cycling process

Increase in Ps Tendency found in GSI (3DVAR) analyses

Substantial increase without constraint



Zonal-average surface pressure tendency for background (green), unconstrained 3DVAR analysis (red), and 3DVAR analysis with TLNMC (purple).

Potential Corrections for Noise and/or Imbalance

- Noise in the background (first guess/model forecast)
  - Full field digital filters
  - Initialization (Nonlinear Normal Mode Initialization)
    - Analysis draws to data, Initialization pushes away from observations

- Noise in the analysis increment
  - Improved multivariate variable definition
  - Penalty terms
  - *Incremental normal mode initialization*

- Discrepancy in passing increment to model
  - Incremental analysis update

Constraint Options

- Tangent Linear Normal Mode Constraint (Kleist et al. 2009) [3D or 4D]
  - Based on past experience and tests with 3D hybrid, default configuration includes TLNMC over all time levels (quite expensive)

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

- Weak Constraint "Digital Filter" [4D Only]
  - Construct filtered/initialized state as weighted some of 4D states

$$J_{dfi} = \chi \left\langle \mathbf{x}_m - \mathbf{x}_m^\mathrm{i}, \mathbf{x}_m - \mathbf{x}_m^\mathrm{i} \right\rangle$$

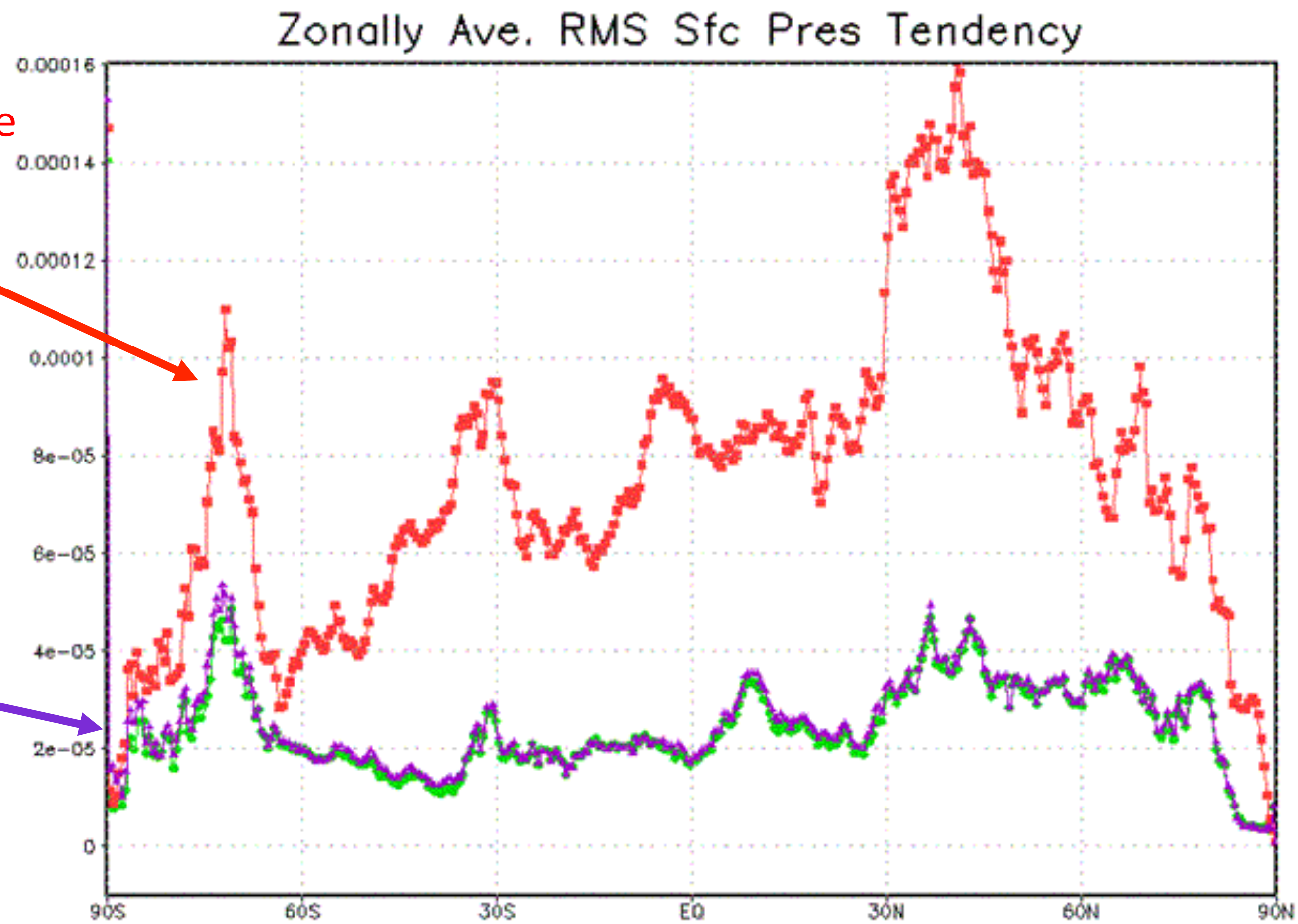
$$\mathbf{x}_m^\mathrm{i} = \sum_{k=1}^{K} \mathrm{h}_{k-m} \mathbf{x}_k^\mathrm{u}$$

- Combination of the two [4D Only]
  - Apply TLNMC to center of assimilation window only in combination with JcDFI (Cost effective alternative?)

Impact of TLNMC on 3DVAR analysis Effective at removing "noise"

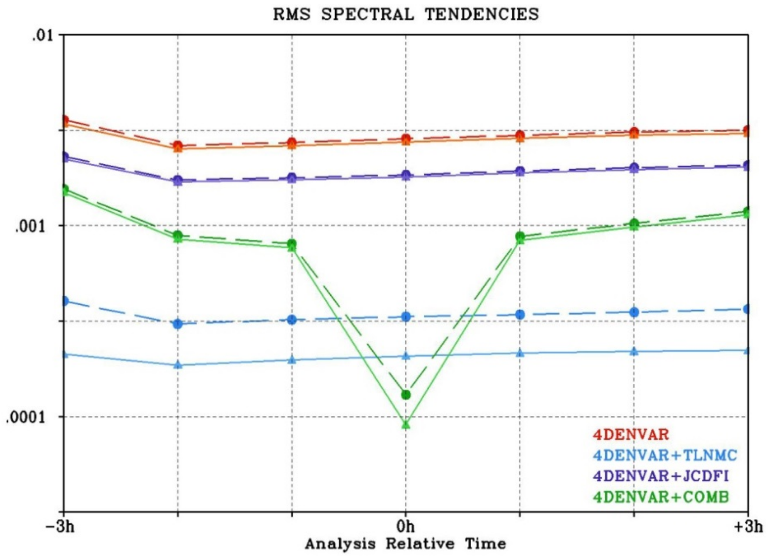
Substantial increase without constraint

Minimal increase with TLNMC

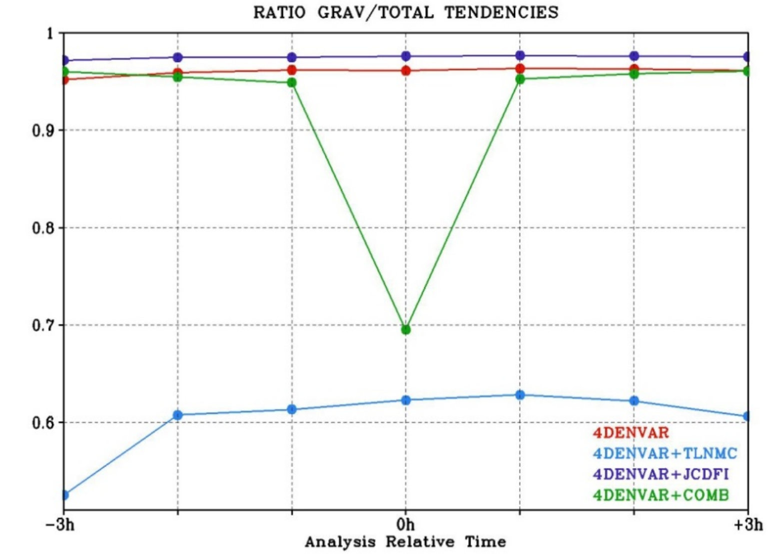


Zonal-average surface pressure tendency for background (green), unconstrained 3DVAR analysis (red), and 3DVAR analysis with TLNMC (purple).

Constraint inter-comparison (single case, Kleist and Ide 2014)

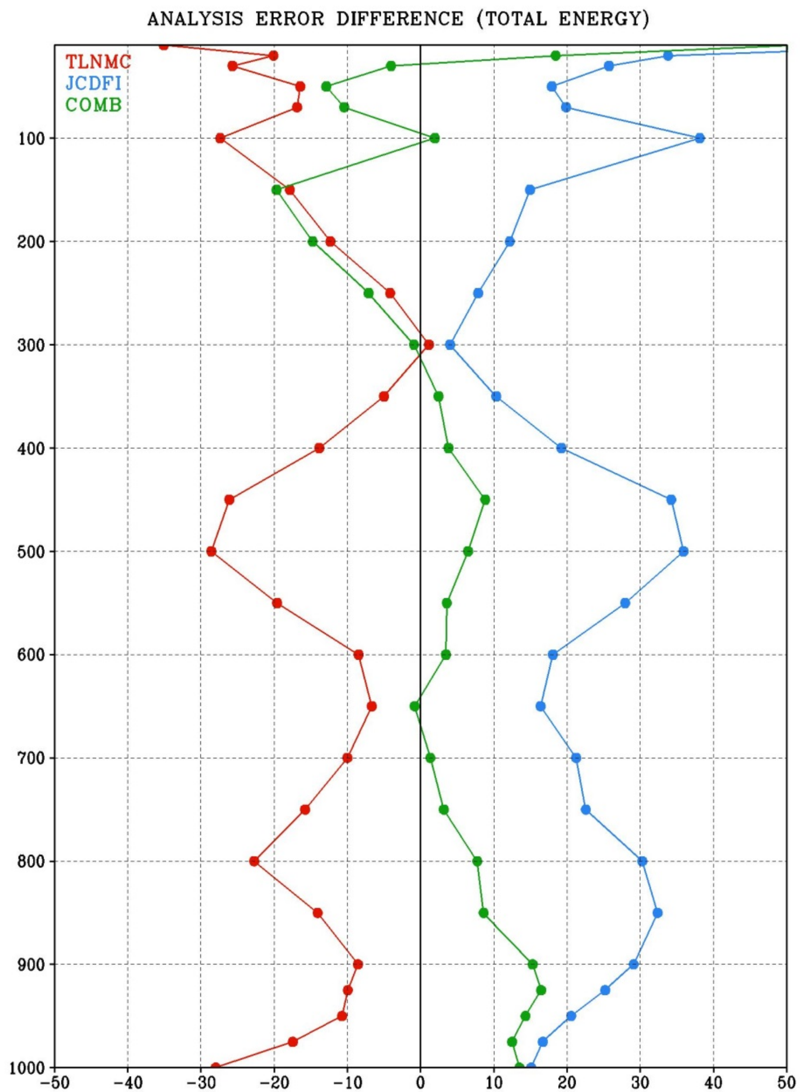


- **Impact on tendencies**
  - Dashed: Total tendencies
  - Solid: Gravity mode tendencies
  - All constraints reduce incremental tendencies



- **Impact on ratio of gravity mode/ total tendencies**
  - JcDFI increases ratio of gravity mode to total tendencies
  - TLNMC most effective (but most expensive)
  - Combined constraint potential (cost effective alternative)

Analysis Error (cycled OSSE)
Hybrid 4D EnVar (Kleist and Ide 2014)



- Time mean (August) change in analysis error (total energy) ***relative*** to 4D hybrid EnVar experiment that utilized no constraints at all
  - **TLNMC universally better**
  - Combined constraint mixed
  - JcDFI increases analysis error

Incremental Constraints Summary

- TLNMC
  - Effective in 3D, 4D, hybrid modes
  - Still room for improvement (physics in tendency model, higher order)
  - Expensive in 4D mode because it has to be applied over all time levels

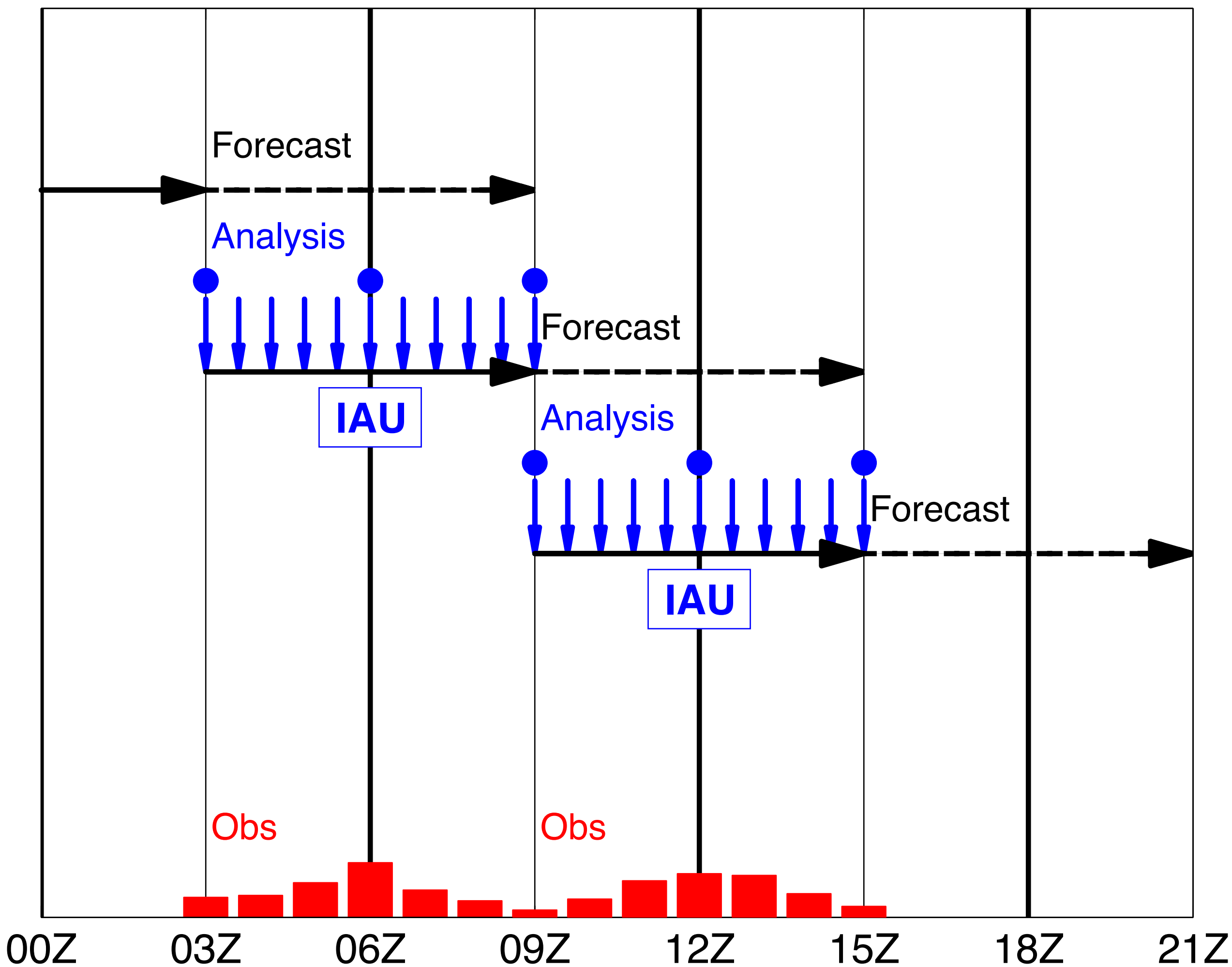
- Weak Constraint Digital Filtering
  - Literature and experience has shown it is effective for 4DVAR
  - However, ineffective for 4D EnVar due to lack of high enough temporal resolution to prescribe high frequency modes to remove (Kleist and Ide 2014)
  - Combination with TLNMC is also not effective

4D Incremental Analysis Update: Motivation

- Analyses can be noisy (deterministic and ensemble-based)
- Imbalances generated by discontinuous nature of analysis, localization & inflation (Greybush, 2011; Kepert, 2009).
- Incremental Analysis Update (Bloom, 1996) helps by using model to distribute a (single) increment over a time window with constant weights (we call this 3DIAU).
  - Propagation of increment neglected, might be significant for fast-moving weather systems.
  - May help spin up unobserved/non-updated state variables
- 4D version of IAU has been proposed by UK Met Office.
- Approximation of "mollified" time-continuous formulation EnKF proposed by Bergemann & Reich (2010).

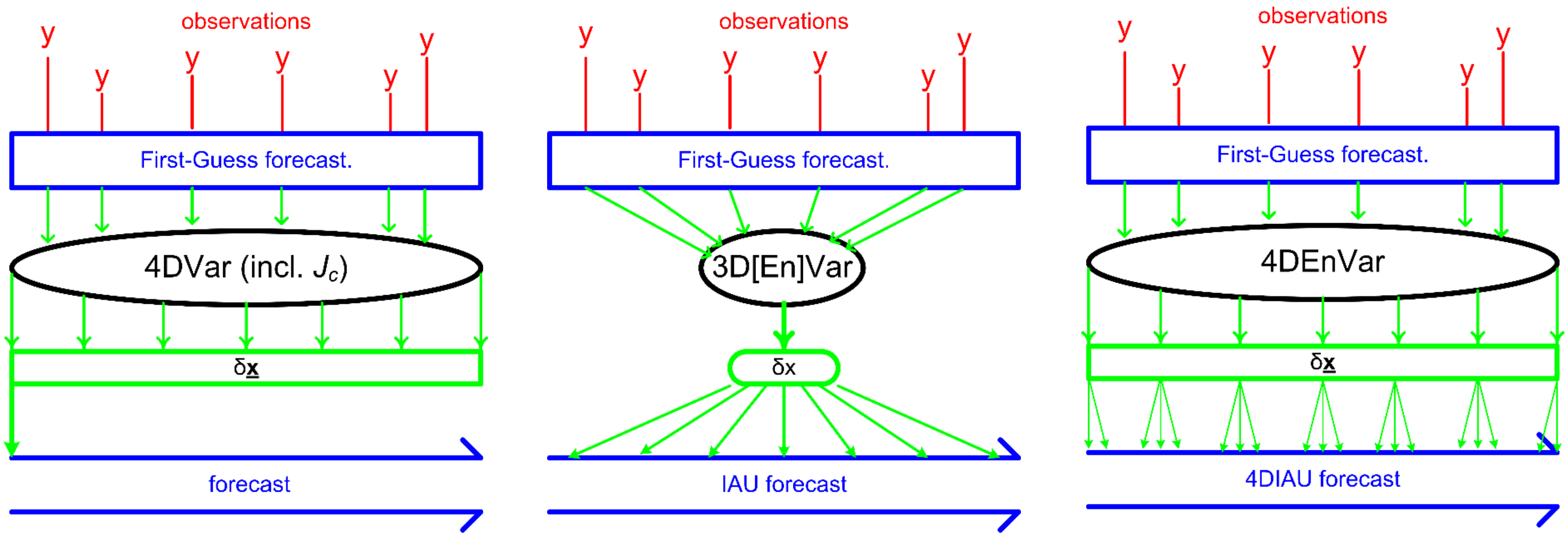
- Here we test the EnKF with a 4DIAU procedure to distribute (time-varying) increments over the assimilation window using the forecast model. (Courtesy Jeff Whitaker and Lili Lei)

Schematic of 4DIAU



JcDFI, 3DIAU, and 4DIAU: From Andrew Lorenc

Initialization



4DVar's
$J_c = \frac{1}{2}\left(\mathbf{F}\delta\underline{\mathbf{x}}\right)^T \mathbf{G}^{-1}\left(\mathbf{F}\delta\underline{\mathbf{x}}\right)$
produces balanced increments by penalizing gravity waves.

IAU applied a related time-filter (Polavarapu *et al.*, 2004) while adding increments to model.

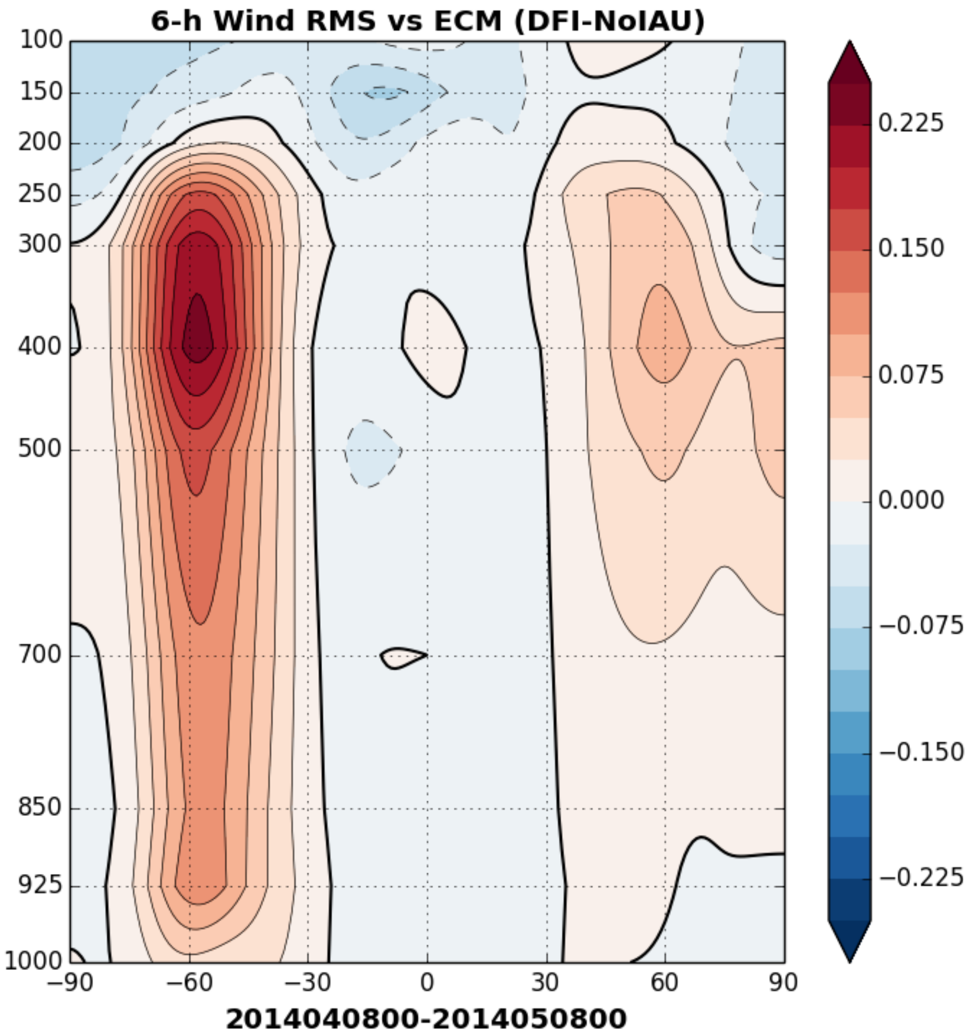
4DIAU has less time-filtering, but is effective at cancelling noise in the increment trajectory.

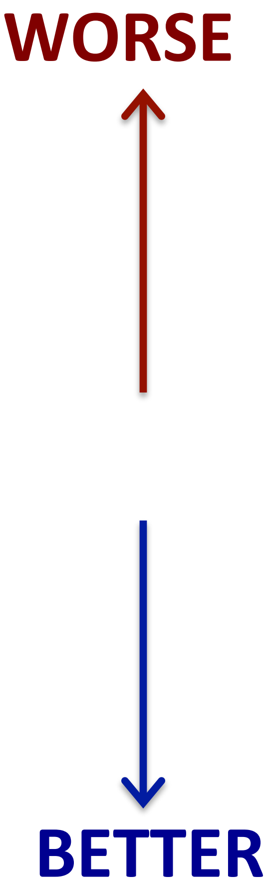
Testing EnKF-4DIAU with NCEP GFS

- T574 ensemble with 80 members
- 1250 km/1.0 scale height localization.
- Stochastic physics and multiplicative inflation (no additive inflation).
- Radiance bias correction comes from a separate EnVar run.
- 6-hour cycling, 3-h forecast output (increments computed at beginning/middle/end of assimilation window for IAU).
- Integration time 2014040100-2014050800; first 7 days are discarded for verification.

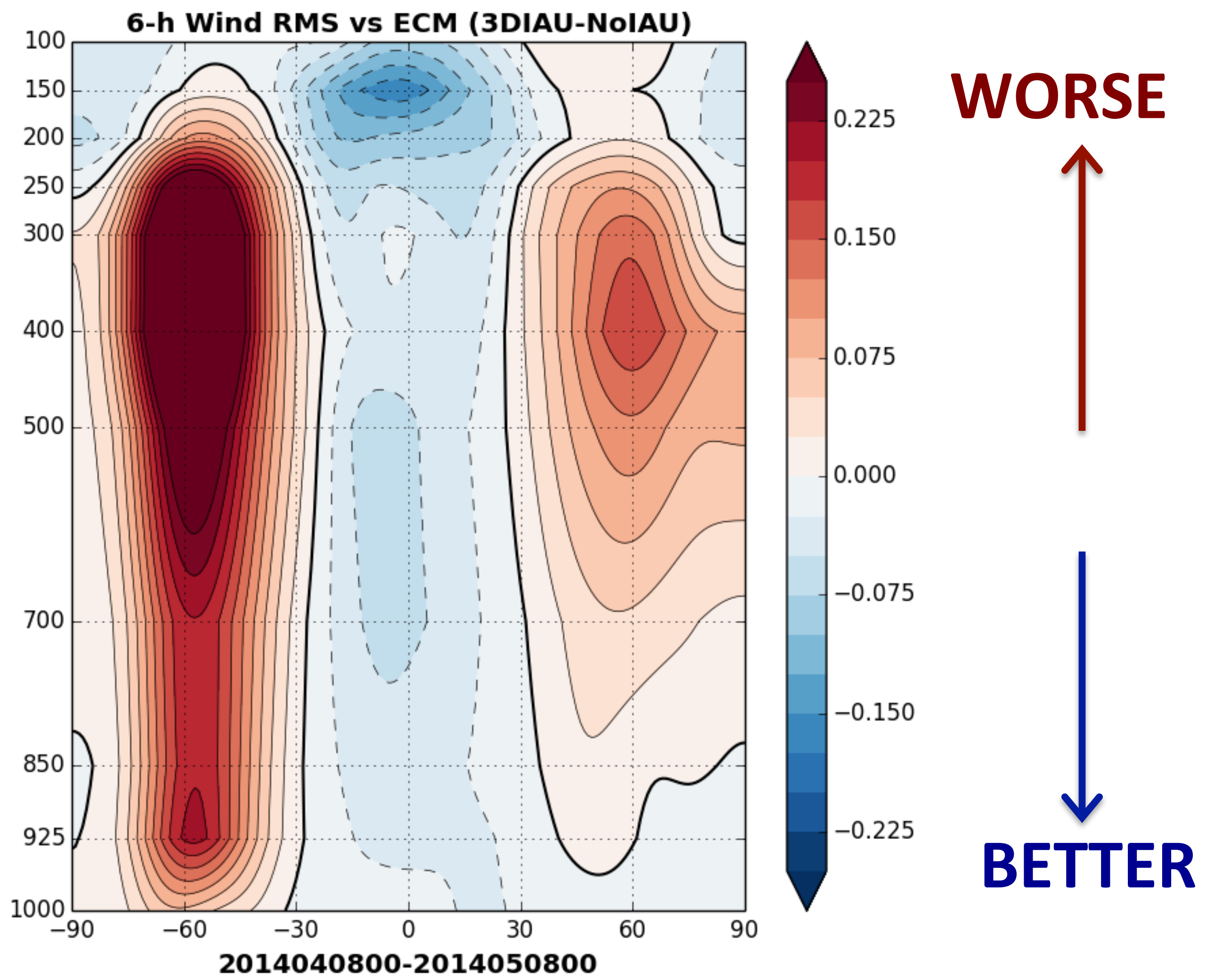
Exp. Name	Exp. Description
EnKF-RAW	Pure EnKF (no DFI or IAU)
EnKF-DFI	EnKF with digital filter (DFI)
EnKF-3DIAU	EnKF with 3DIAU, no DFI
EnKF-4DIAU	EnKF with 4DIAU, no DFI

DFI vs 'raw' EnKF (no IAU or DFI)

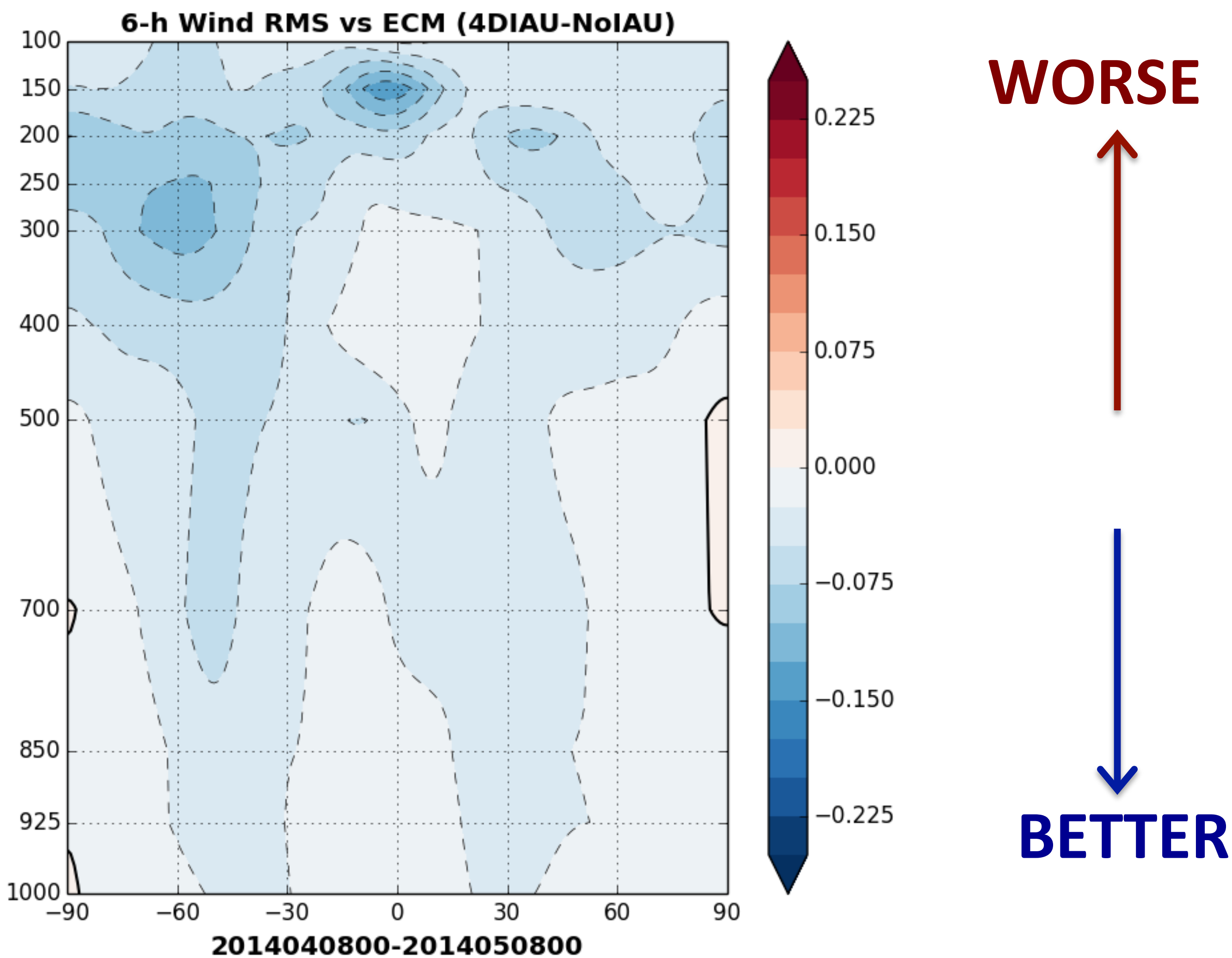




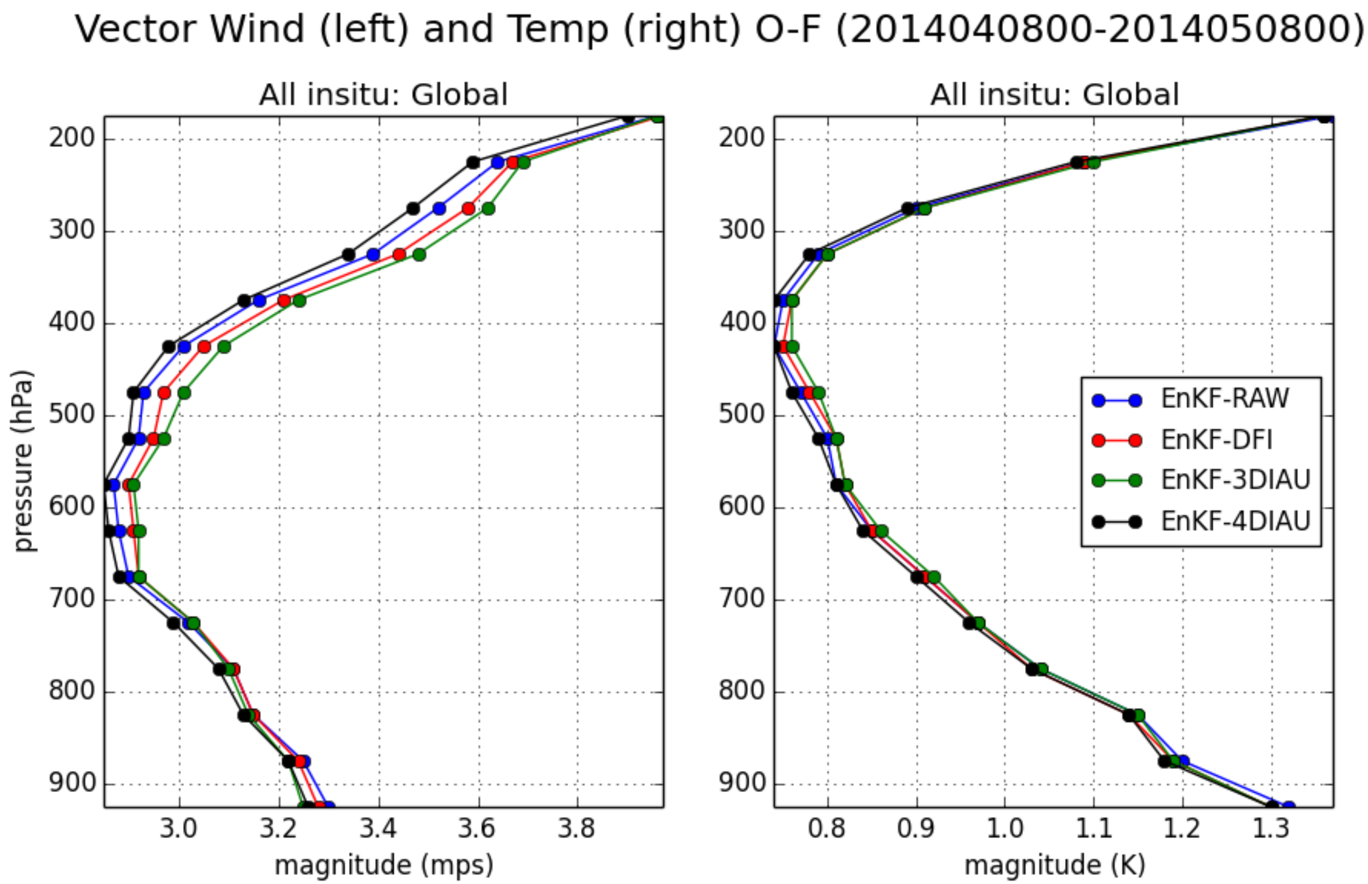
3DIAU vs 'raw' EnKF



4DIAU vs ‘raw’ EnKF



Observation Space Verification (innovation statistics)

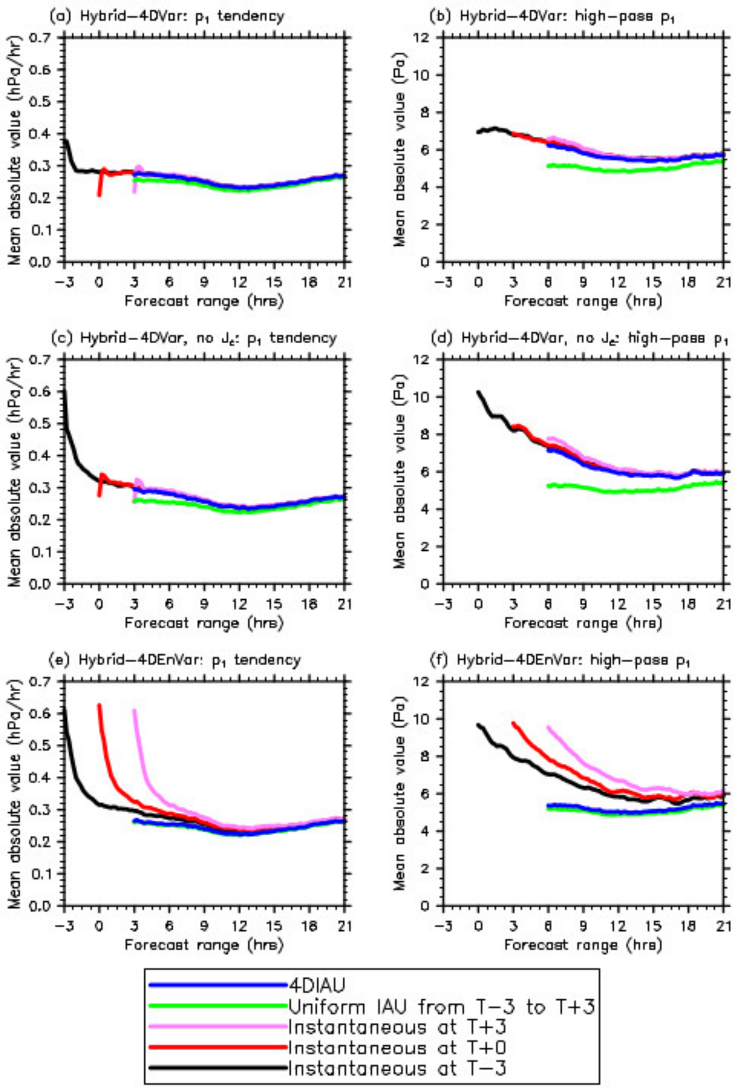


EnKF-DFI has slightly larger errors than the EnKF-RAW.

EnKF-3DIAU produces the largest errors except below 800 hPa.

EnKF-4DIAU slightly better than EnKF-RAW (NoIAU).

Example from MetOffice System Lorenc et al. (2014, Fig. 8)



- Imbalance by 4DEnVar similar to 4DVAR without JC [middle black with bottom black/red/pink]
- 4DEnVar with IAU is as well balanced (or better) than 4DVAR with JC [top black/bottom blue]

IAU Summary

- IAU reduces the imbalances introduced by discontinuous analysis step, localization and inflation.
- 3DIAU can degrade analysis quality, when increments change significantly within the window (extra-tropical storm tracks).
- 4DIAU improves things especially for large ensemble sizes/ long localization scales, when there is strong advection/ propagation of increments and model errors are not large.
  – Requires computing multiple increments.
  – LETKF well suited, since analysis weights need only be computed once.
  – Works well for deterministic as well as ensemble

Initialization Recap

- Analyses can be noise, need to apply initialization
- TLNMC has proven effective, but has deficiencies
- IAU may be good cost-effective alternative, promising results with 4D application in 4D EnVar context
- **NCEP GFS still uses full field digital filter as well, likely to be removed soon (permanently)

Ensemble Variable Choices

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

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

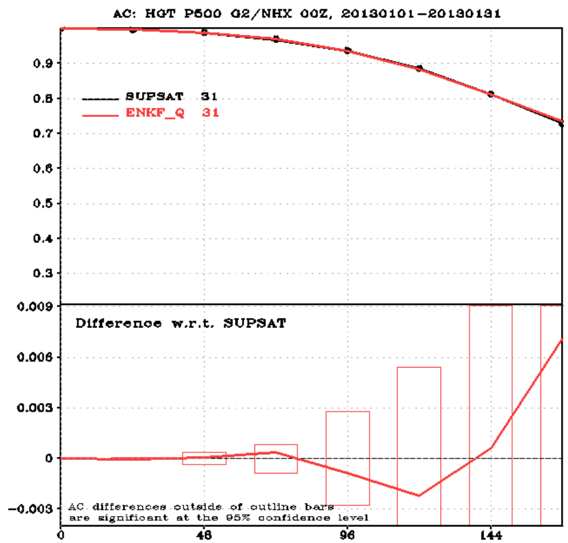
- Original design of GSI (NCEP) EnVar prescribed the ensemble-based increment to be in same space (variables) as the static control variable
  - Streamfunction, velocity potential, pseudo-relative humidity, etc.
- This may not be the best choice
  - For example, using RH perturbations results in accounting for the temperature component twice in the ensemble-based increment (once in T, once in RH)
- Use of alternate choices of ensemble perturbations variables being pursued
  - Those more natural to model state
  - “Balance-aware” for localization
  - Cloud ice & water instead of single cloud condensate variable

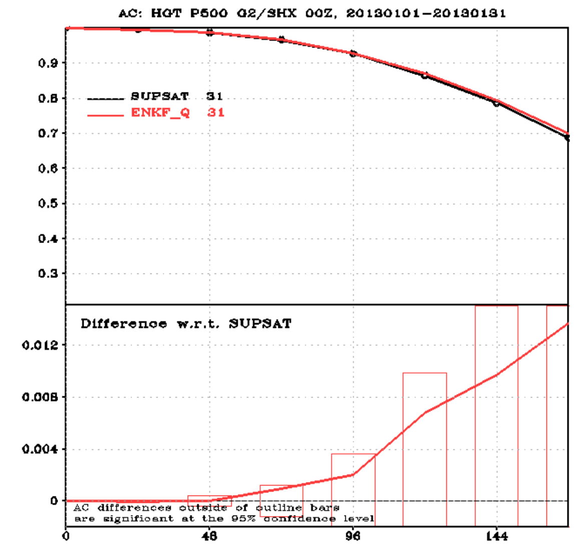
Cycled Tests of Q (ENKF_Q) vs RH (Supsat) in Hybrid 3D EnVar

NH

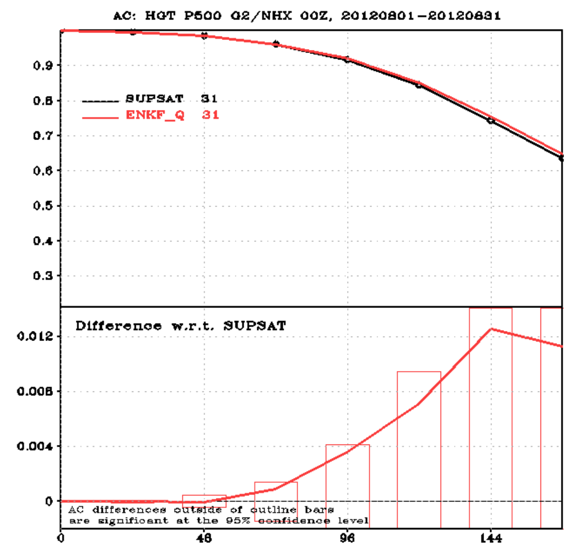
SH

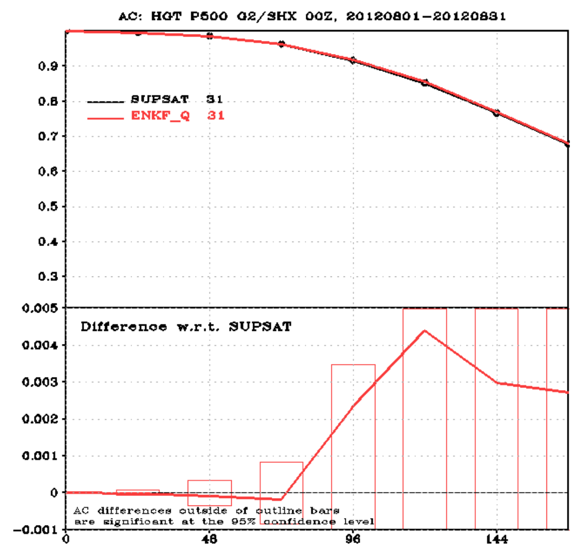
January 2013





August 2013

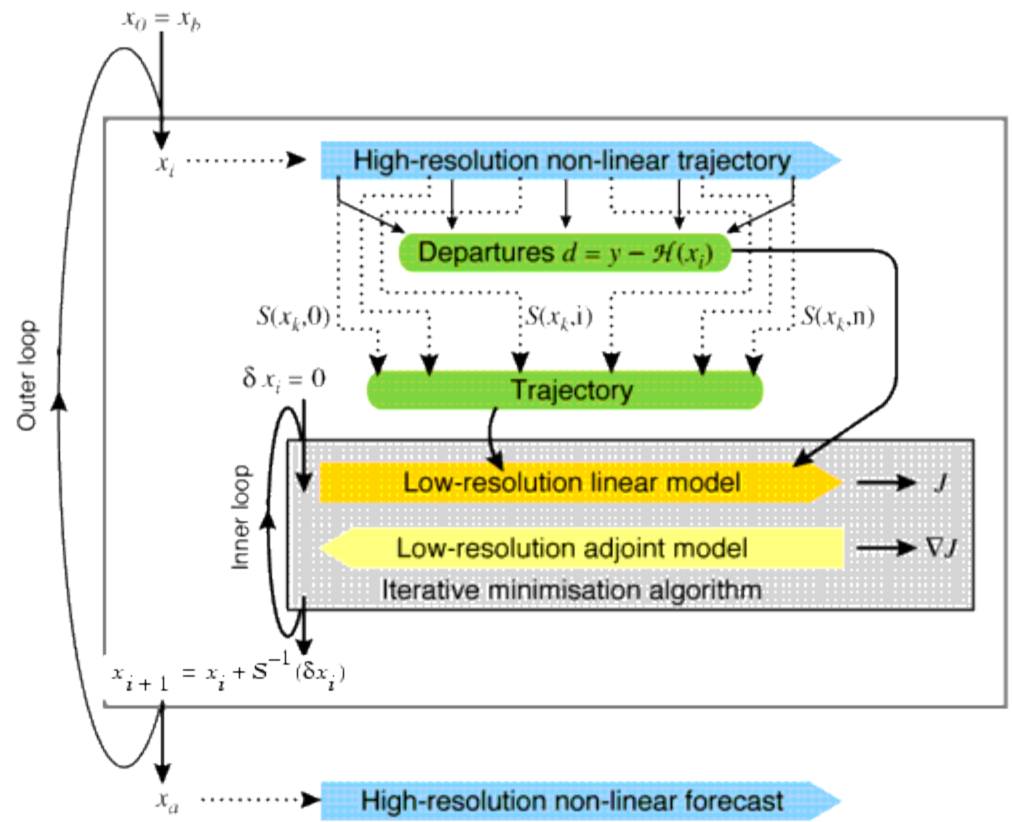




Ensemble Variable Choices

- Q is demonstrably a better choice for ensemble variable for use in EnVar
  - Not double counting for temperature as when using RH
  - Maps onto model state variable directly
- Exploring other choices such as cloud ice & liquid instead of single cloud condensate
  - Particularly for use in cloudy radiance assimilation
- Likely explore variable transforms to help address non-Gaussianity

Outer Loop for 4DVAR

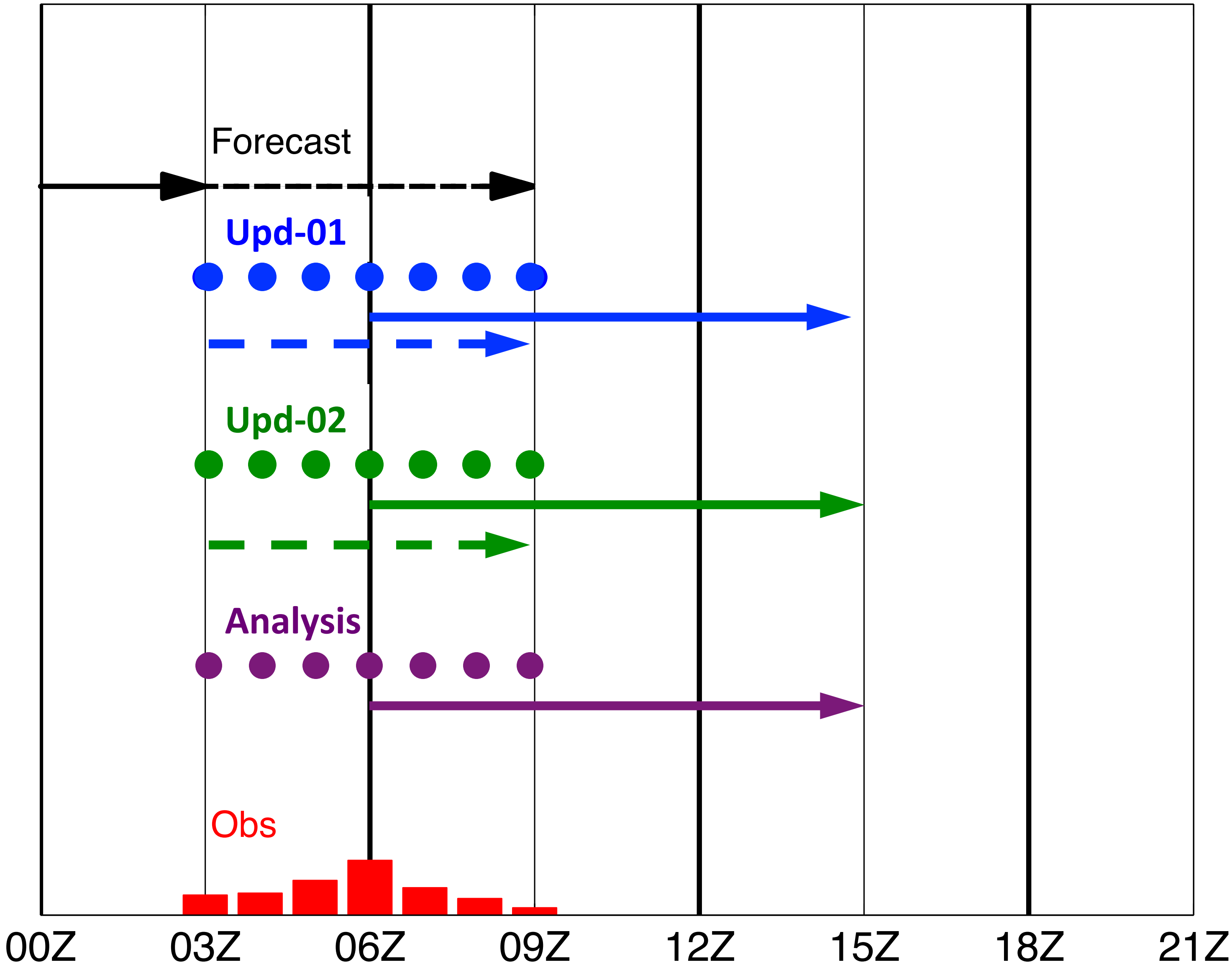


Outer loop utilizing full (high resolution) non-linear model could be adapted to 4D EnVar. This has proven to be beneficial within the context of an EnKF (Quasi Outer Loop, Yang et al. 2012)

Figure Courtesy: ECMWF

39

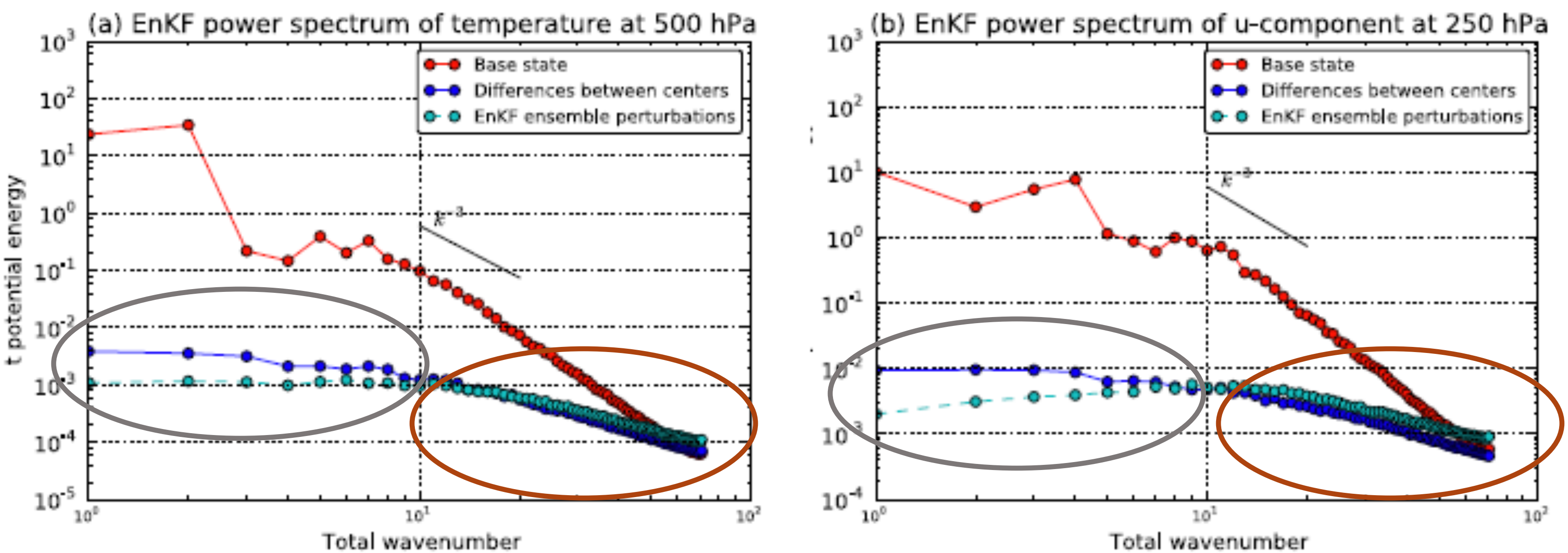
4D EnVar Outer Loop



4D EnVar Outer Loop

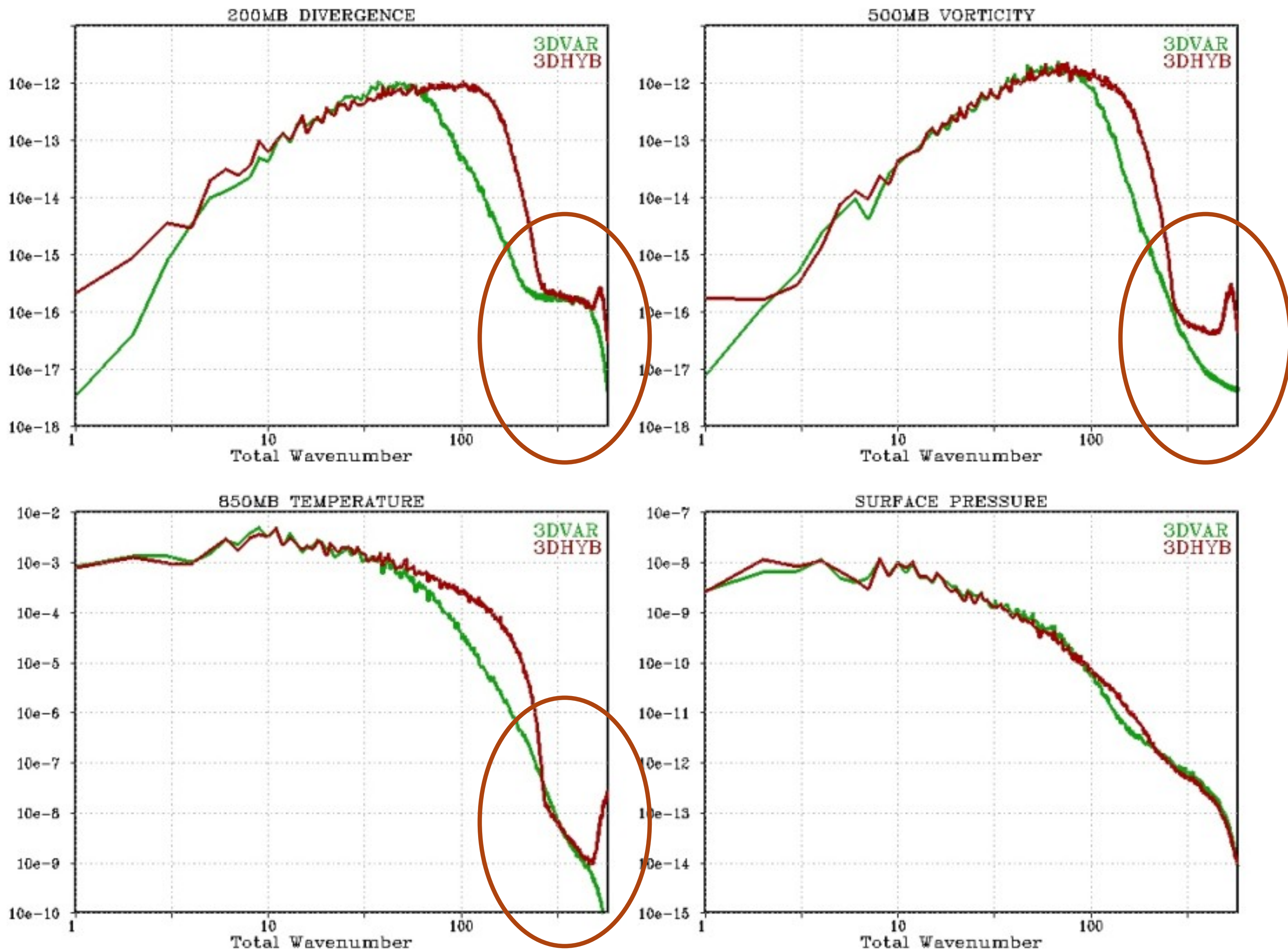
- 4D EnVar solves for increment valid at each of the discretized times
  - In case of developmental NCEP GFS version, hourly in a 6 hour window
- This includes an increment valid at the beginning of the window
  - Currently T-3 from "analysis time"
- Model can be restarted from this state to create new background for solver, just as in 4DVAR outer loop (or QOL)
- Allows for treatment of nonlinearity
  - Testing underway for NCEP GFS 4DEnVar

Scale-Dependence Motivation (Courtesy: Tom Hamill)



(1) Generally more power at all wavenumbers relative to ETR.
(2) Overestimate of power (i.e., amplitude of perturbations) at small scales. Likely this is attributable to inappropriate analysis increments due to the use of smaller-than ideal ensemble size (n=80) in the EnKF, and still-crude methods (covariance localization) for filtering usable signal from sampling noise.

Spectrum of Dual-Resolution Hybrid Increment (interp. aliasing)



Scale Dependent Weights

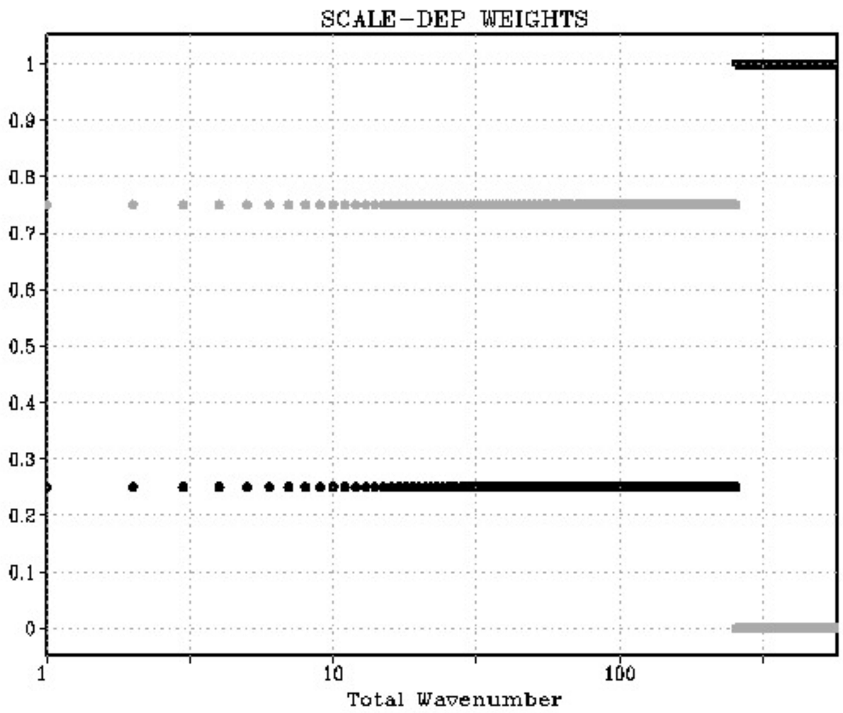
- Can be achieved by reformulating some of the problem in spectral space, such as the introduction of a new control variable for ensemble based increment and spectral operator **S**

$$\boldsymbol{\alpha} = \mathbf{L}\boldsymbol{v}$$

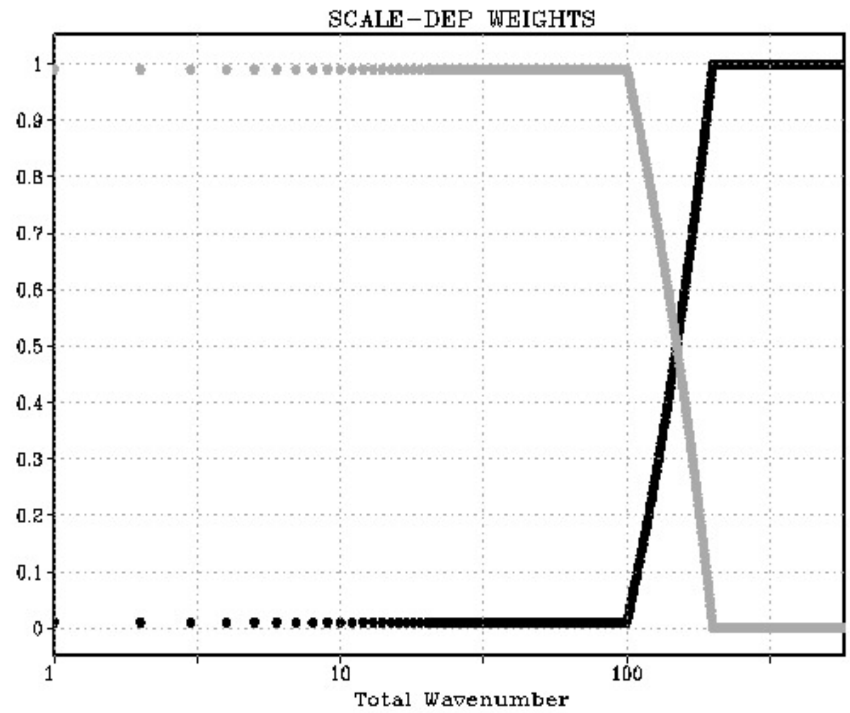
$$\mathbf{x}_{\mathrm{e}}' = \mathbf{S}^{-1}\left[\left(\boldsymbol{\beta}_{\mathrm{e}}^{\mathrm{s}}\right)^{-1}\mathbf{S}\left(\sum_{n=1}^{N}\left(\boldsymbol{\alpha}^{n} \circ \mathbf{x}_{\mathrm{e}}^{\ n}\right)\right)\right]$$

- Allows to rely more heavily on ensemble in part of the spectrum that is not dominated by sampling error (and vice versa). Also, allows one to revert entirely to static B where ensemble has no information at all (below truncation)

Scale-dependent β

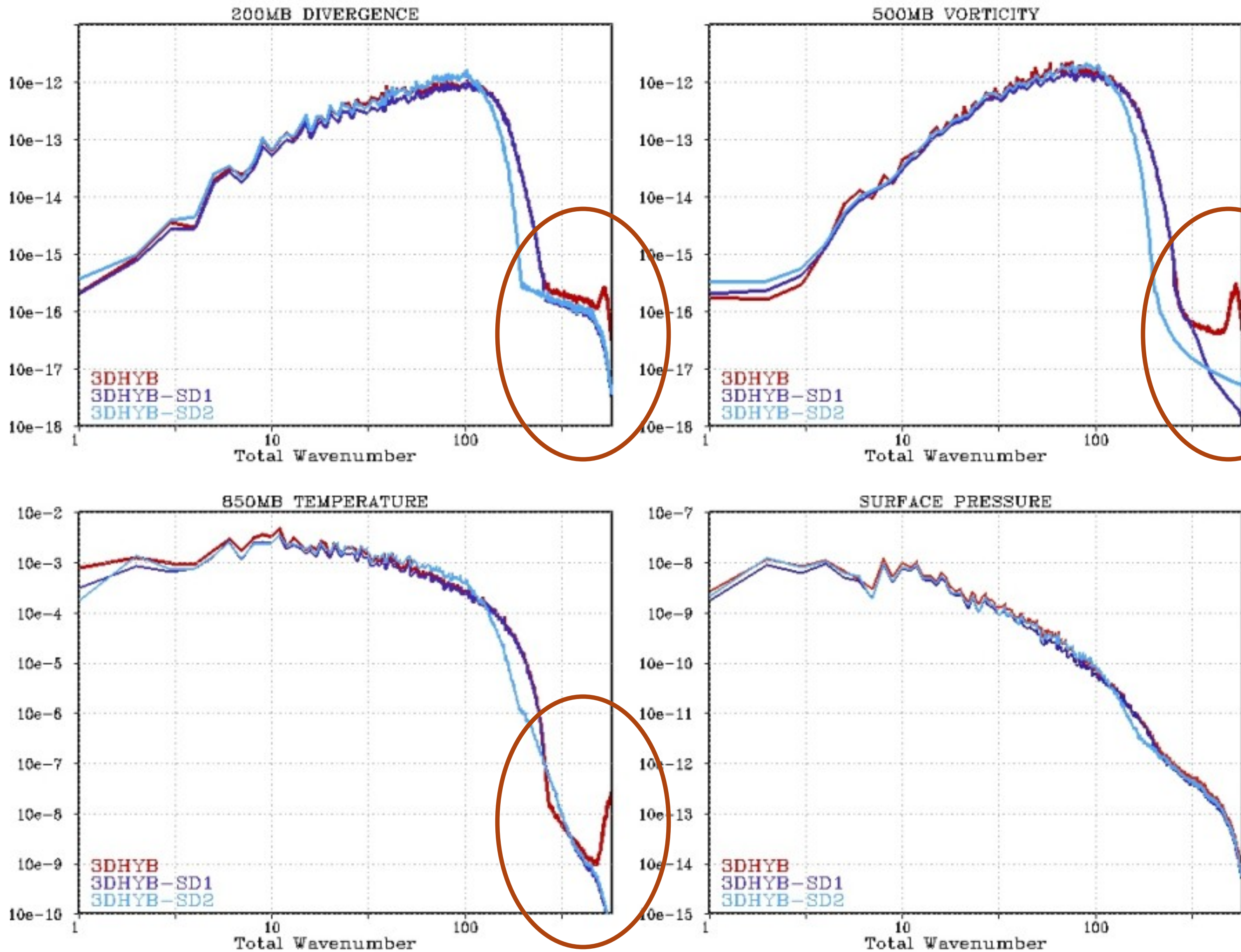


SD1

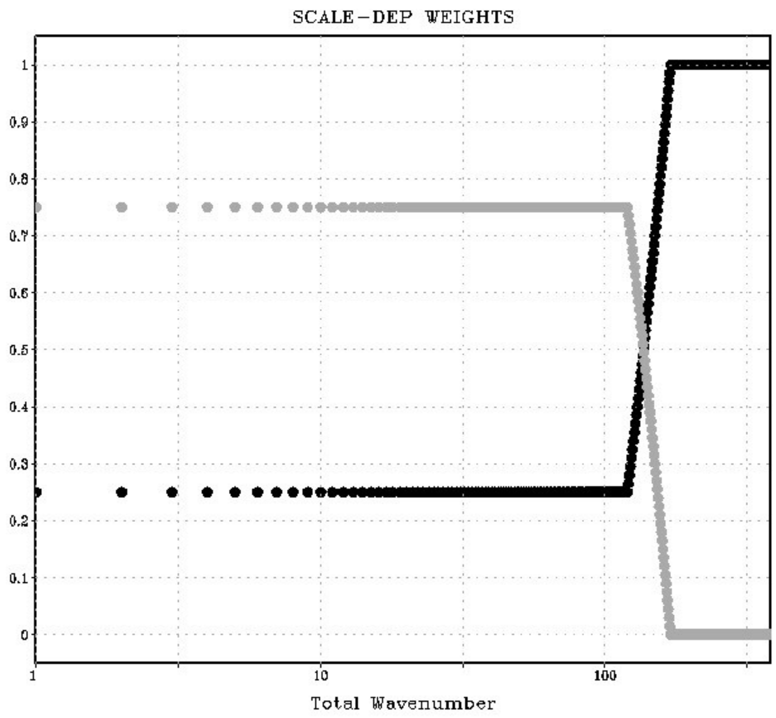


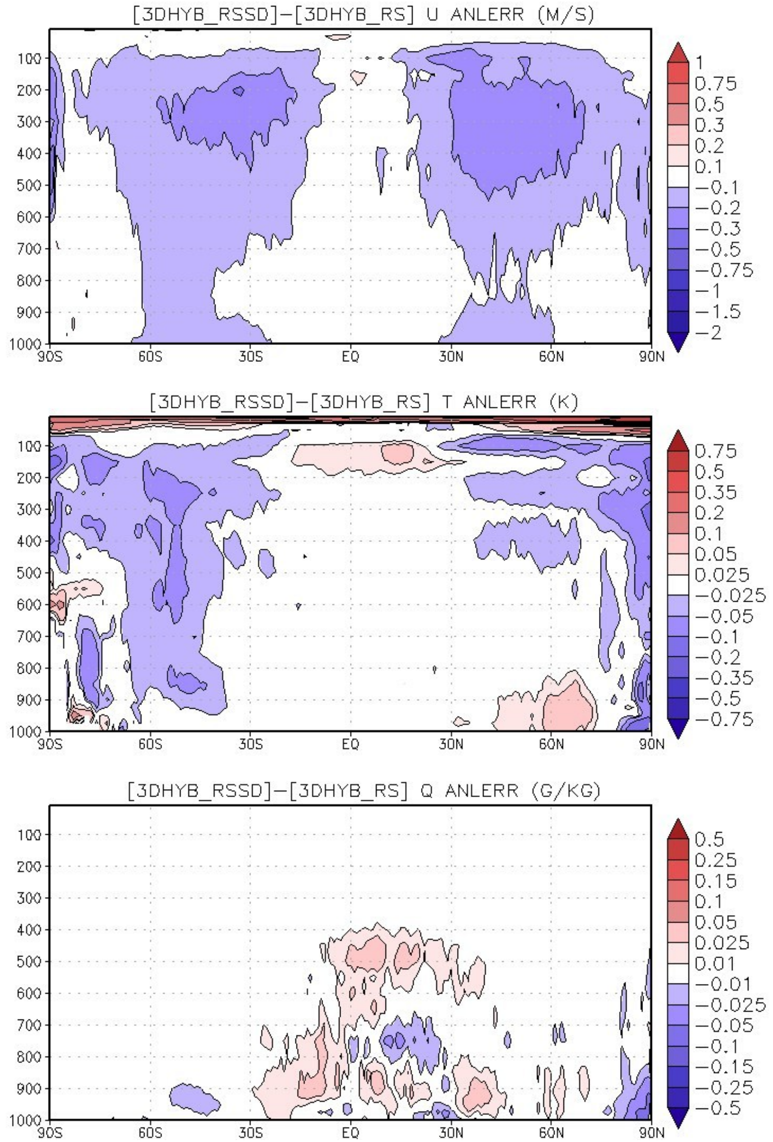
SD2

Spectrum of Dual-Resolution Increment with SD-weighting



Cycling Results
Analysis Error





Scale Dependence

- Not all information from ensemble created equal
  - High frequencies dominated by sampling error

- Preliminary testing shows scale-dependent weighting effective
  - However, more parameters to consider
  - Expensive because of spectral transforms

- Could be used in combination with wave band filtering to perform scale-dependent localization

Summary

- Significant progress has been made on 4D EnVar development and testing for operational NWP at NCEP (see talk tomorrow at CWB)
- Further improvements expected through use of improved initialization
  - Removal of DFI, use of 4D IAU
  - What to do about TLNMC remains open question
  - Handing of ensemble also important
- Choice of variable for use in EnVar may be important
  - Some have thought of this within context of EnKF already
  - Scale dependent weighting shows some promising early results
- Much of the literature shows that hybrid 4DEnVar is not quite as good as hybrid 4DVar
  - Can close some of this gap with initialization (4DIAU) and perhaps outer loop (to be determined)
  - Significant work remains in coming up with more optimal static B

Opportunities

- Collaboration!

- Research Scientist at Univ. of Maryland
  - Hiring to work on hybrid 4D EnVar for aviation industry, 2 year project with likely extension or follow-on opportunities

- Graduate Students at Univ. of Maryland