Random Musings: Update on 4D Hybrid, Toward Multi-Scale, TC initialization in the GFS, and GSI MPI Scalability Issues

Daryl Kleist

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

Central Weather Bureau Seminar
Taipei, Taiwan
1 November 2016

Thanks to Deng-shun Chen (CWB and NCU), Kayo Ide (UMD), Lili Lei (NOAA/ESRL), Rahul Mahajan (NOAA/NCEP), David Parrish (NOAA/NCEP), Cathy Thomas (NOAA/NCEP), Chien-Han Tseng (CWB), Jeff Whitaker (NOAA/ESRL), Wan-shu Wu (NOAA/NCEP)

&

many more for inspiration, collaboration, and several slides.

Before I Begin…

- I want to express my sincere thanks to CWB, and especially Chien-Han Tseng, for the collaboration, hospitality, and support over the past several years
- It is my honor and distinct pleasure to have opportunities to visit with CWB and consult on many exciting developments
- As I strongly believe in its benefits, I will continue to push for further engagement and collaboration between EMC and CWB

Hybrid 4DEnVar

$$J(\mathbf{x}'_c,\mathbf{a}) = b_c \frac{1}{2}(\mathbf{x}'_c)^T \mathbf{B}_c^{-1}(\mathbf{x}'_c) + b_e \frac{1}{2}\mathbf{a}^T\mathbf{L}^{-1}\mathbf{a} + \underbrace{\frac{1}{2}\sum_{k=1}^{K}(\mathbf{H}_k\mathbf{x}'_{(t)k} - \mathbf{y}'_k)^T\mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{x}'_{(t)k} - \mathbf{y}')}_{\text{Jo term divided into observation “bins” as in 4DVAR}}$$

$$\mathbf{z} = \mathbf{B}^{-1}\mathbf{x}'_c \qquad \mathbf{v} = \mathbf{L}^{-1}\mathbf{a}$$

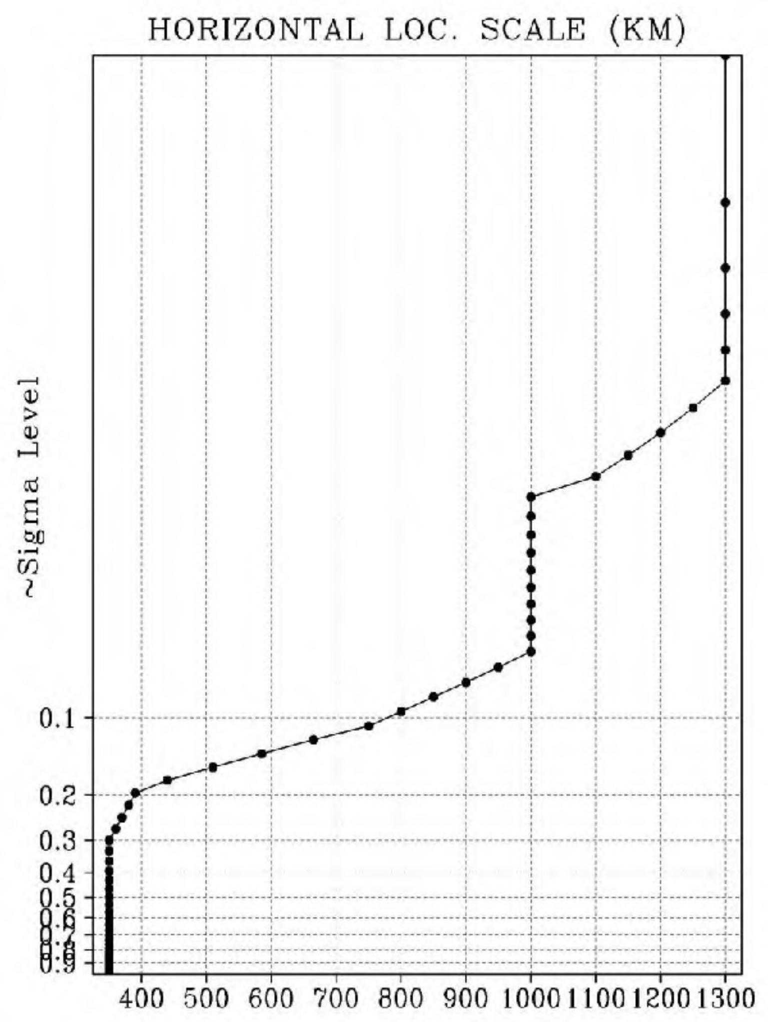
Where the 4D increment is prescribed through linear combinations of the 4D ensemble perturbations plus static contribution, i.e. it is not itself a model trajectory

$$\mathbf{x}'_{(t)k} = \mathbf{C}_k\left[\mathbf{x}'_c + \sum_{m=1}^{M}(\alpha^m \circ (\mathbf{x}_e)_k^m)\right]$$

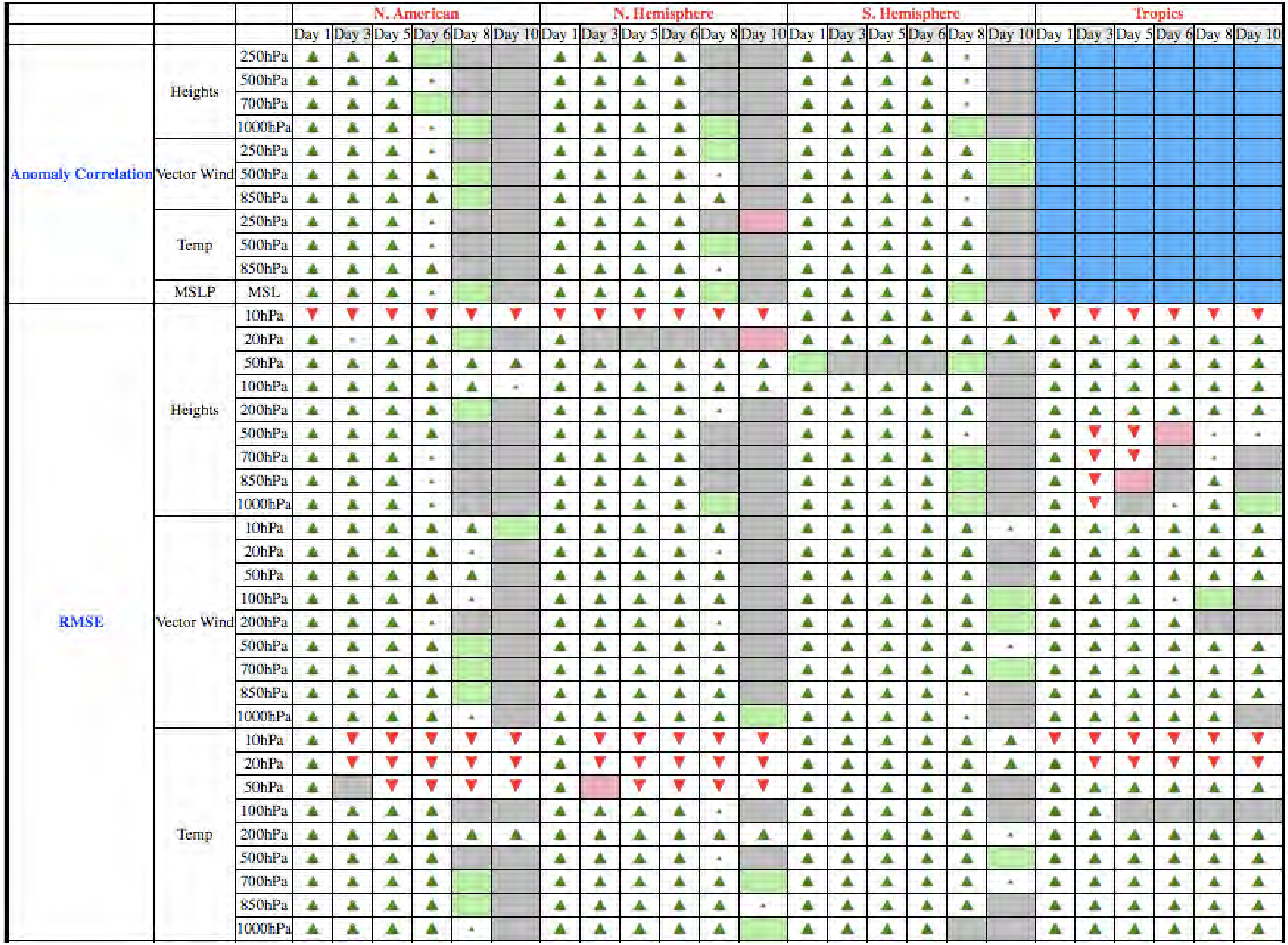
Here, ***static contribution is time invariant***. **C** represents TLNMC balance operator. ***No TL/AD in Jo term*** (**M** and $\mathbf{M}^T$). Linear **H** used in cost function.

Hybrid 4DEnVar Operational Configuration Implemented: May 2016

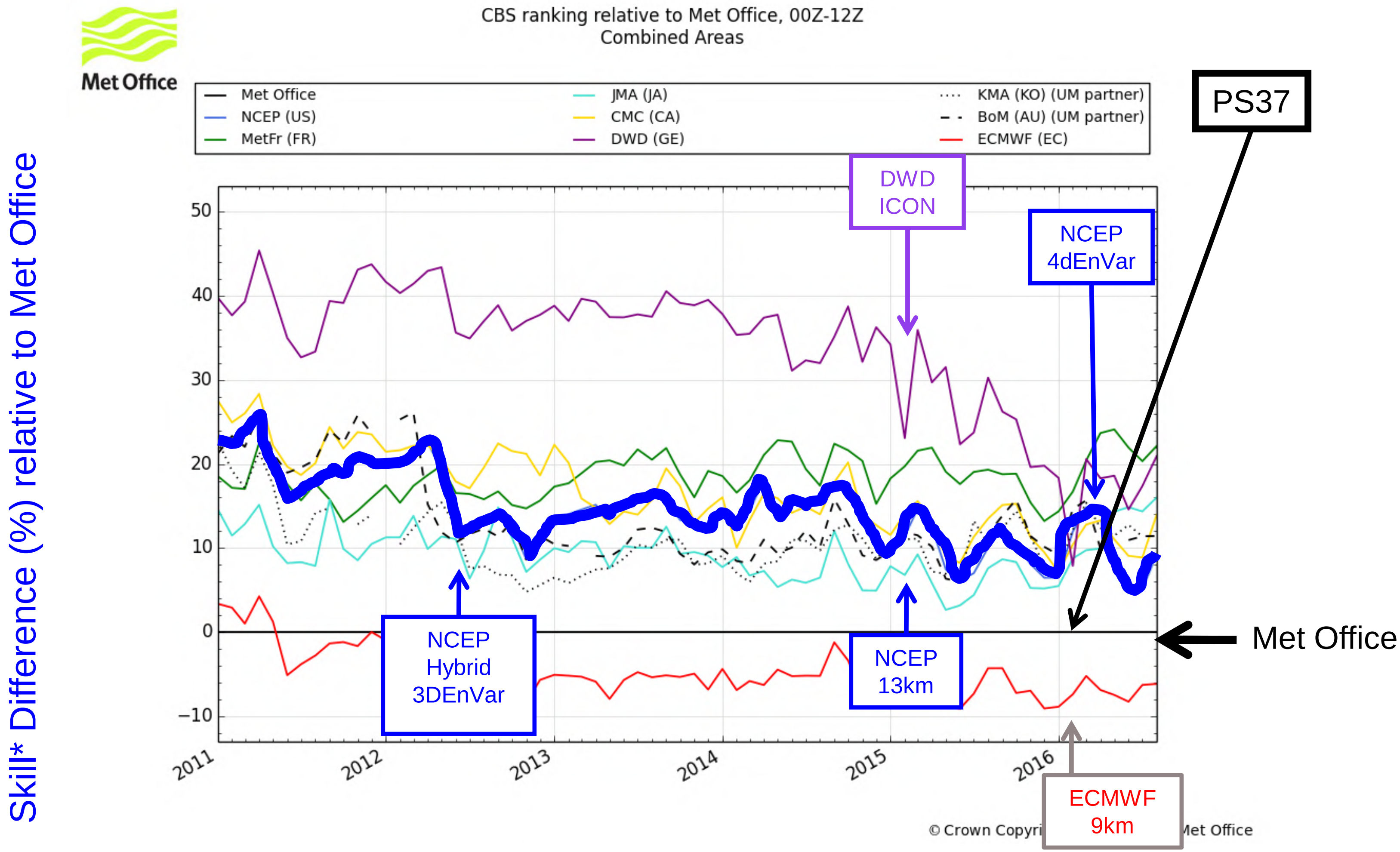
- T1534L64 Deterministic (SL dynamics)
  - T574L64 EnSRF, 80 members, Stoch. Physics, Hourly Output
- 87.5% ensemble, 12.5% climatological for hybrid increment
- Level dependent horizontal localization (divide by 0.38 to convert to GC zero distance)
  - 0.5 scale heights in vertical



Summary Scorecard (20130501-20160228)
4DEnVar Package versus Operational GFS



UKMET NWP Index Time Series
Courtesy Gilbert Brunet & Dale Barker

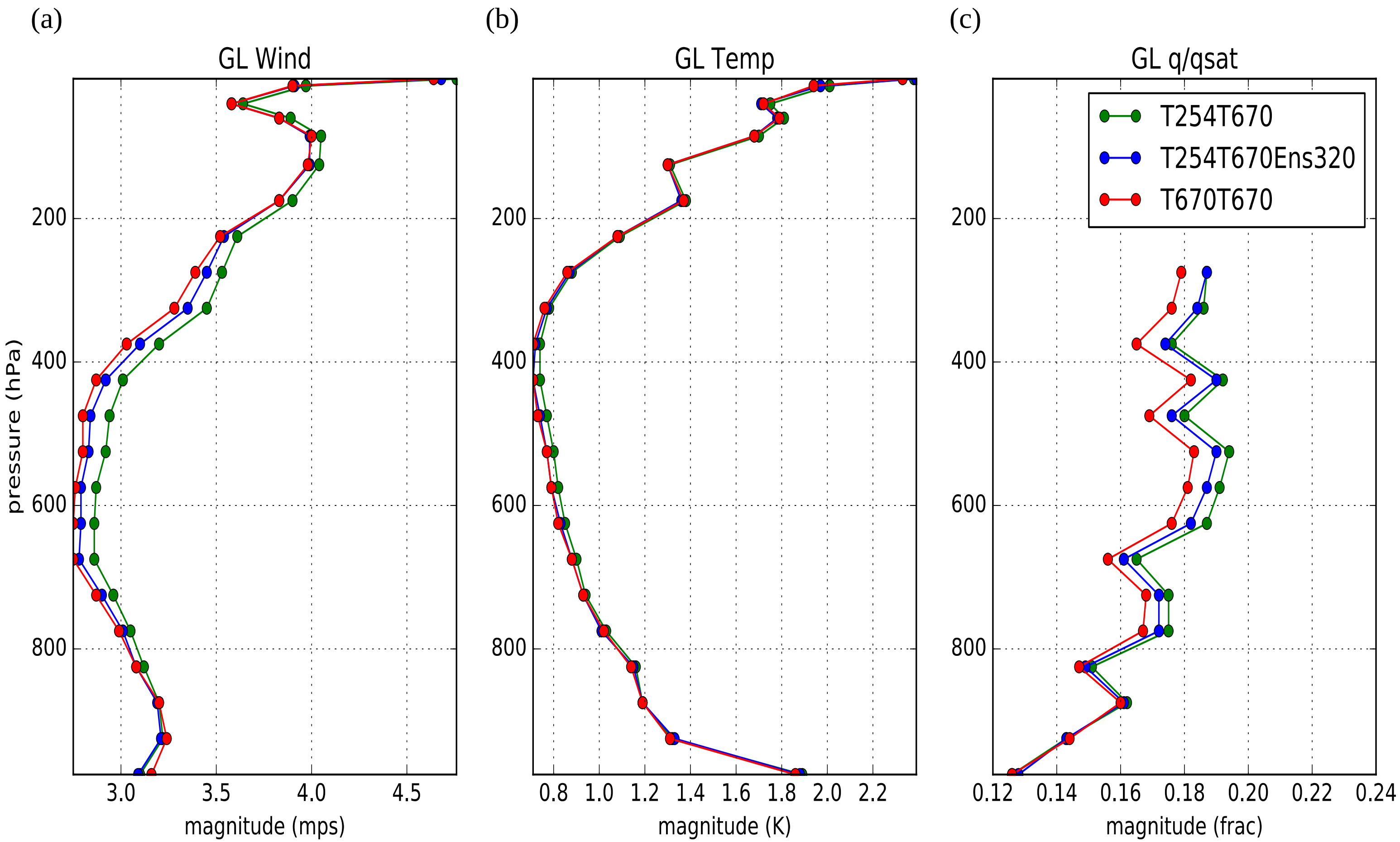


* Parameters: Surface pressure, 500hPa geopotential height, 250hPa/850hPa Winds; Forecast ranges from T+24h to T+120h

Current works in progress on hybrid 4DEnVar

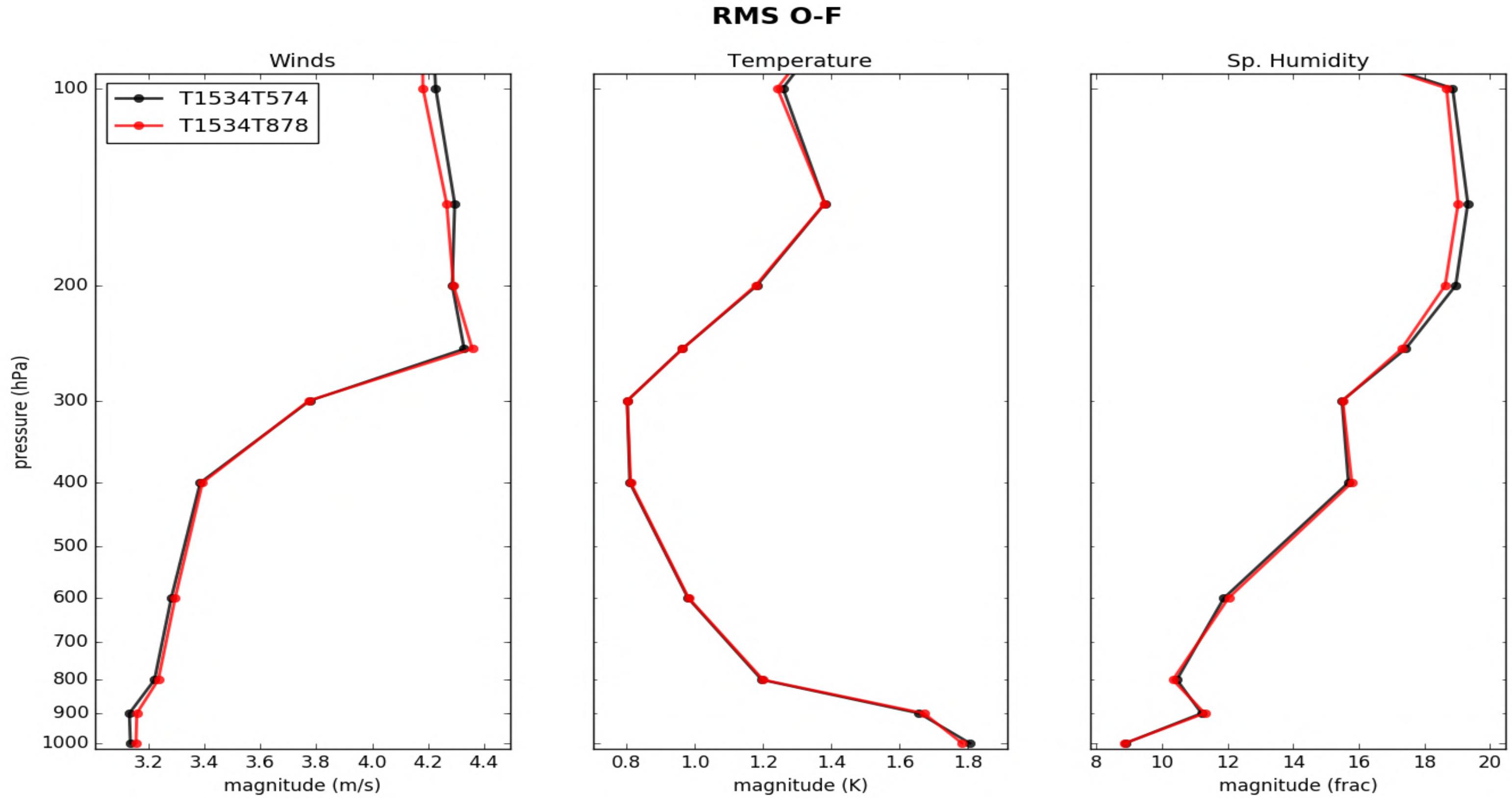
- Replace full field digital filter with 4D incremental analysis update similar to UKMO and CMC systems [already tested]
- ***Explore spectral (waveband) and/or scale-dependent localization***
  - ***Optimistic that this can help with moisture and clouds***
- ***More generalized hybrid weights and/or scale-(waveband-) dependent weights***
- Use of outer loop within 4D EnVar ***or weak constraint***?
- Time evolving static covariance and localization without full TL/PF

Larger Ensemble Size or Higher Resolution?
Lei and Whitaker 2016



Caveat: Larger ensemble size may be especially beneficial if we can rid of vertical localization (especially for satellite data).

Experiments with higher ensemble resolution (Rahul Mahajan, 15 day sample)

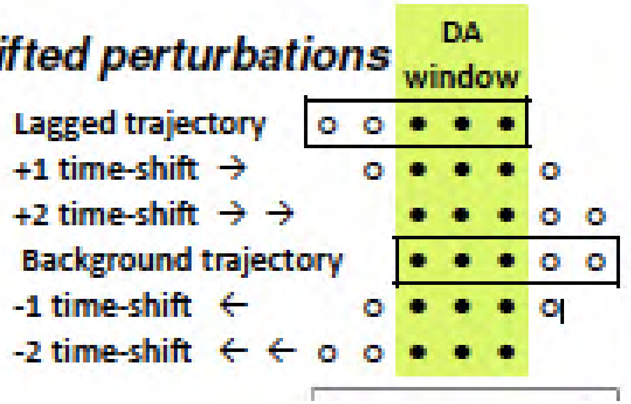


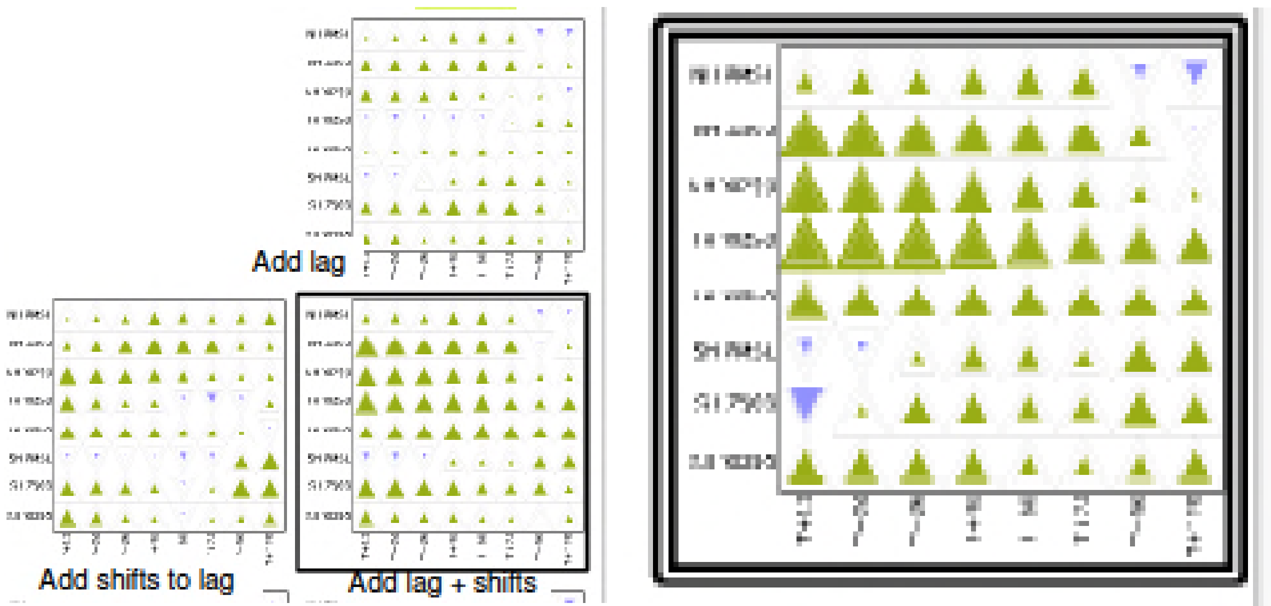
- On new computing asset, test simply increasing resolution of ensemble from T574 to T878. Some model parameters tuned for 878 but everything else identical.
- This also results in higher resolution analysis increment (done at ensemble res.)

Lagging and Shifting
Courtesy Andrew Lorenc (In review)

- ***Use of time-lagged & shifted perturbations***

Lagging uses perturbations valid at the correct time, from longer forecasts from the previous cycle. Shifting uses perturbations valid at slightly different times.





Adding shifting, lagging, and scale-dependent localization gave a 44 member ensemble similar skill as a 200 member ensemble!

http://www.isda2016.net/assets/posters/LorencThebenefit.pdf

Spectral (Waveband) Localization Toward Multi-Scale Analysis

- Currently apply single, level-dependent horizontal localization (removing balance and time (k) index for brevity)

$$\mathbf{x}'_{(t)} = \mathbf{x}'_{c} + \sum_{m=1}^{M} \left(\alpha^{m} \circ (\mathbf{x}_{e})^{m} \right)$$

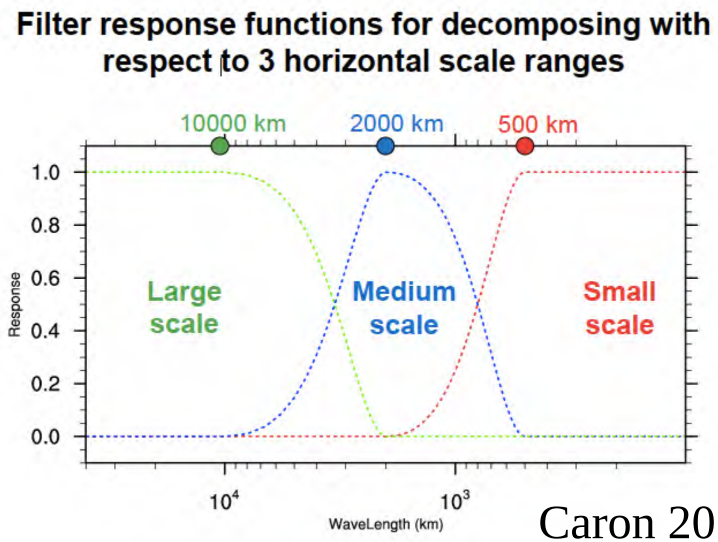
- Buehner (2005, 2012) proposed filtering ensemble perturbations into wavebands* and applying different levels of spatial localization therein

$$\mathbf{v}_{m,w} = \mathbf{L}_{w}^{-1} \mathbf{a}_{m,w} \qquad \mathbf{x}'_{(t)} = \mathbf{x}'_{c} + \sum_{w=1}^{W} \sum_{m=1}^{M} \left(\alpha^{m,w} \circ (\mathbf{x}_{e})^{m,w} \right)$$

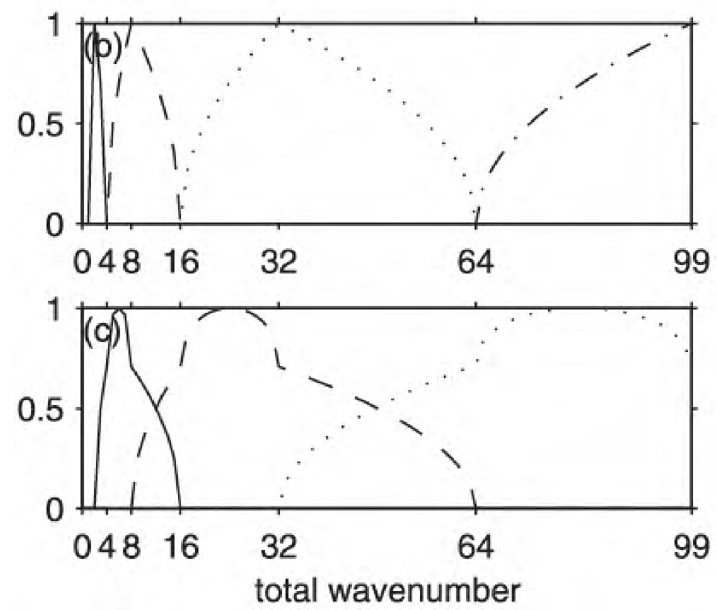
- Ensemble control variable expanded to include wavebands.
  - Localization applied differently for each waveband. No correlations between wavebands
  - Ensemble perturbations filtered into wavebands

Just doing the wavebands (with overlapping weights) is equivalent to spatial smoothing.

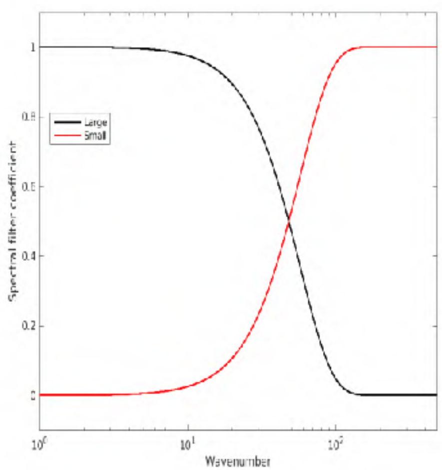
Example Filter Functions



Caron 2016

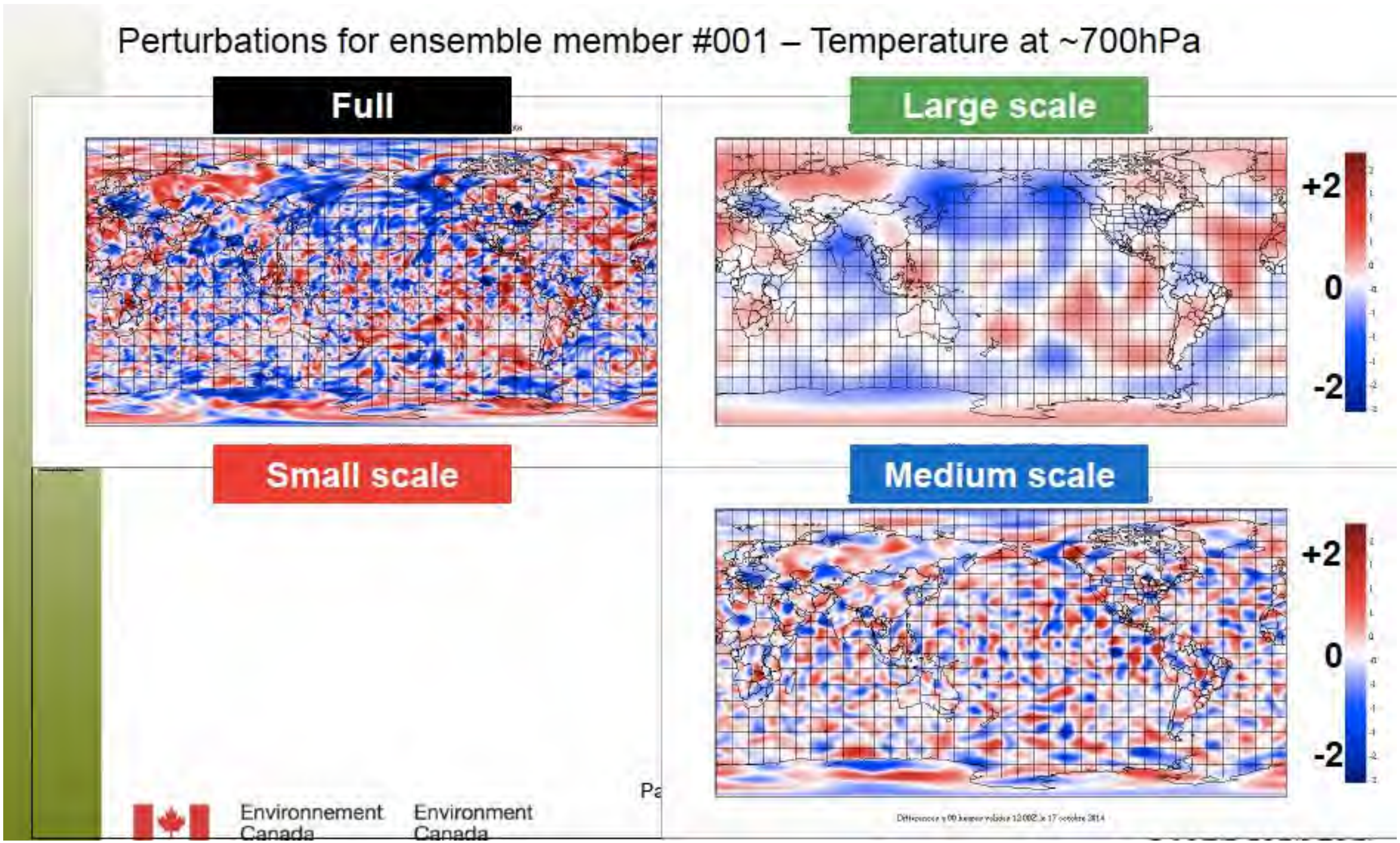


Buehner 2012



Chen 2016

Filtered Perturbations
Courtesy JF Caron (ECCC)



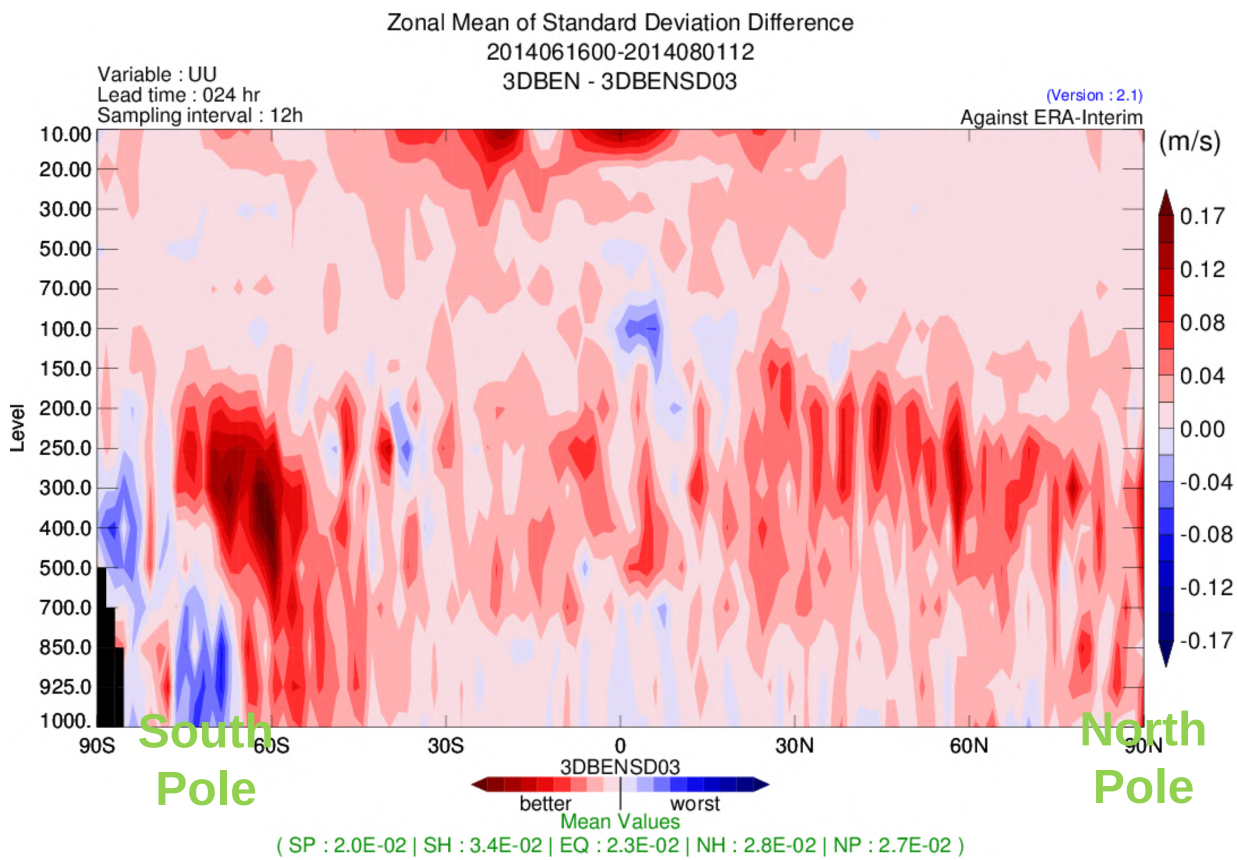
Scale-dependent covariance localization

Forecast impact – Comparison against ERA-Interim

- **Control is better**
- **Scale-Dependent is better**

T+24h
Zonal mean

Std Dev difference for U

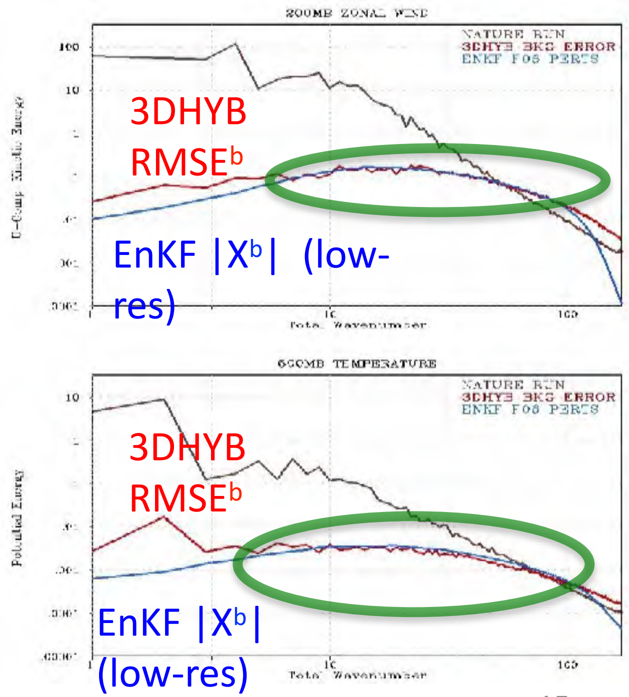


Extension to waveband-dependent weighting

- In addition to waveband-specific localization, expand to beta (hybrid) weights

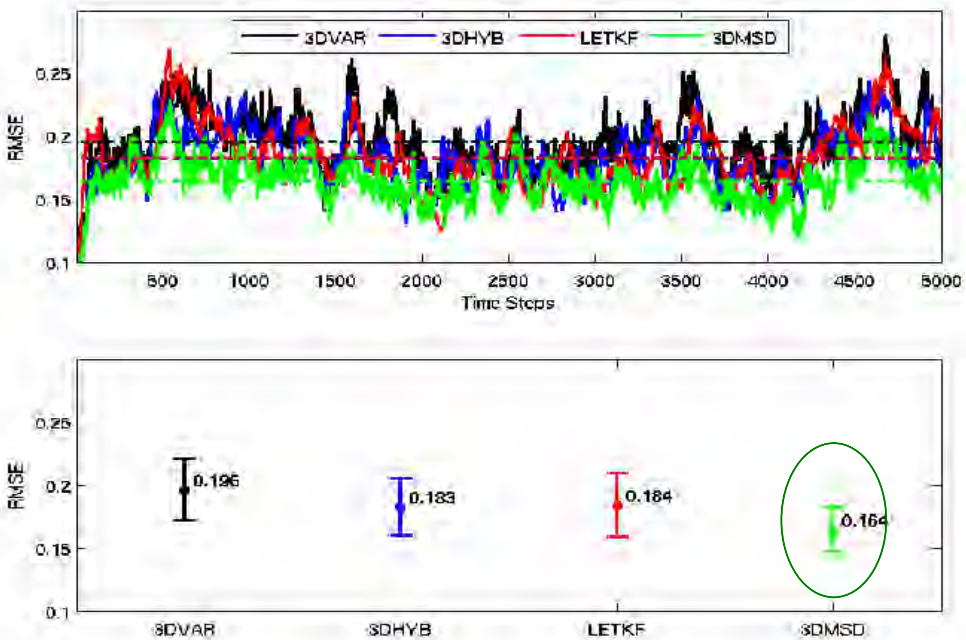
$$J(\mathbf{x}'_c, \mathbf{a}) = b_c \frac{1}{2} (\mathbf{x}'_c)^T \mathbf{B}_c^{-1} (\mathbf{x}'_c) + \sum_{w=1}^{W} b_e^w \frac{1}{2} \mathbf{a}_w^T \mathbf{L}_w^{-1} \mathbf{a}_w + \frac{1}{2} \sum_{k=1}^{K} (\mathbf{H}_k \mathbf{x}'_{(t)k} - \mathbf{y}'_k)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{x}'_{(t)k} - \mathbf{y}')$$

- From OSSE experiments (right) and real data, expectation that ensemble is 'most reliable' for moderate, synoptic-to-meso scales



Example with Toy-Model
Deng-Shun Chen

- Lorenz (2005) multi-scale model. We first demonstrated that two-scale localization can improve hybrid. We then looked at scale-dependent weighting:



- Test with dual resolution configuration (ensemble at half resolution
- Add in simple waveband dependent weighting
  - Large influence from ensemble at larger scales.
  - Larger influence from static B (with small correlation lengths) at small scales

Where we are?

- Control variable extension installed into GSI
  - Largely untested
- Need to add code to apply waveband filtering of ensemble perturbations
- Extension to add waveband weights trivial
  - But how to estimate what the values should be?
  - Better solution than initial attempts to build full spectral weights
- Extend to true scale-dependent localization
  - Cross-waveband correlations
  - Will require significant code changes

Multi-Scale: Multi-Resolution Ensembles

- Current high-resolution applications use global ensembles within their hybrid schemes (NAM, HWRF, HRRR)
- What if instead we could supplement global (Large scale, Lg) with regional, high resolution (Small scale, Sm) ensemble

$$J(\mathbf{x}'_c, \mathbf{a}^{Lg}, \mathbf{a}^{Sm}) = b_c \frac{1}{2}(\mathbf{x}'_c)^T \mathbf{B}_c^{-1}(\mathbf{x}'_c) + b_e^{Lg}\frac{1}{2}\left(\mathbf{a}^{Lg}\right)^T \mathbf{L}_{Lg}^{-1}\mathbf{a}^{Lg} + b_e^{Sm}\frac{1}{2}\left(\mathbf{a}^{Sm}\right)^T \mathbf{L}_{Sm}^{-1}\mathbf{a}^{Sm} + J_o$$

$$\mathbf{x}'_{(t)} = \mathbf{x}'_c + \sum_{m=1}^{M}(\alpha_{Lg}^m \circ (\mathbf{x}_\mathbf{e}^{Lg})^m) + \sum_{m=1}^{M}(\alpha_{Sm}^m \circ (\mathbf{x}_\mathbf{e}^{Sm})^m)$$

- Filters can be set up just as in spectral/waveband localization to control degree of overlap!
- Analysis increment becomes linear combination of static + large scale (global) + small scale (regional)

Global Relocation Experiments

- During my last two visits, I reported on some recent work on TC recon assimilation OSEs and relocation experiments
- This work has been extended
- Long term goal is to find more general solution to problem within the data assimilation directly (PhD Student Ms. Chu-Chun Chang)
  - Position assimilation in var and EnKF
  - Displacement / Field Alignment

TC Initialization at NCEP

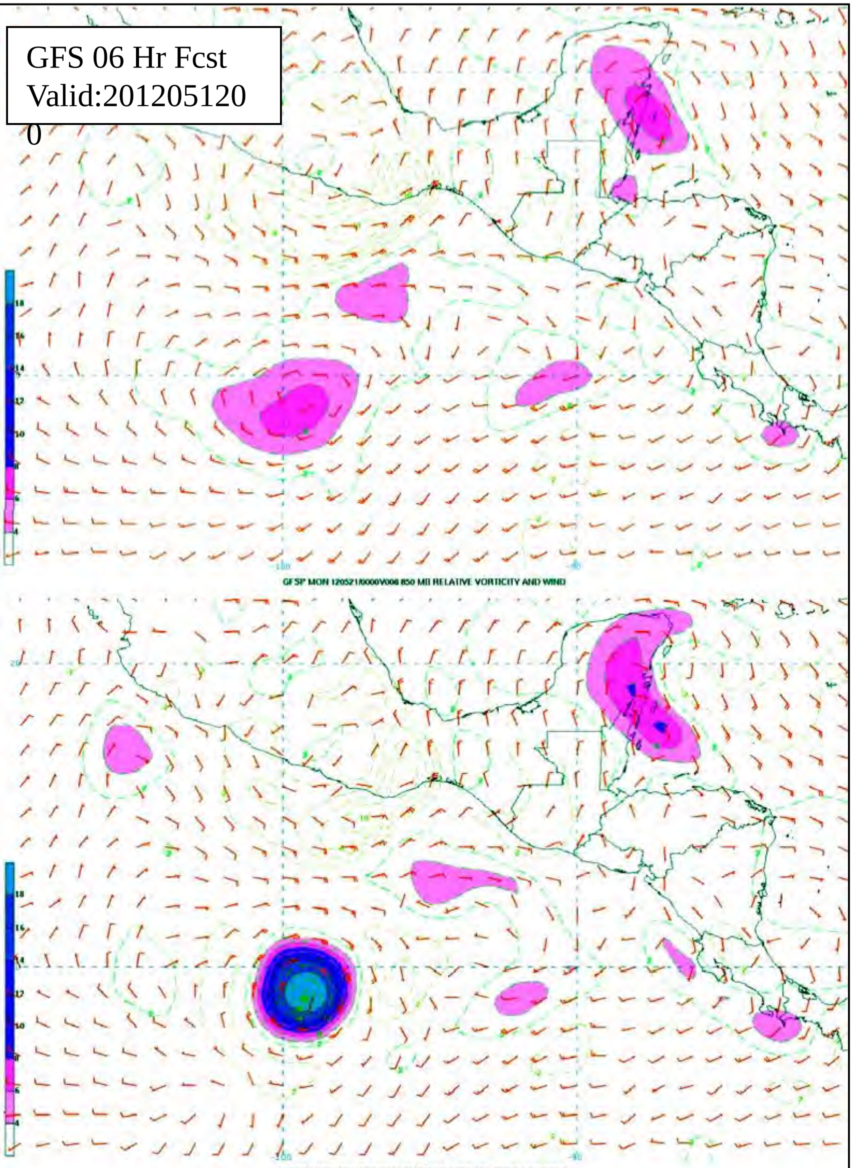
- For the operational GFS / GDAS, there is always some component from outside of the actual assimilation of real observations involved:

1. "Tracker" is run on GDAS forecast
   a. If storm found in forecast/background, ***mechanical relocation*** of vortex
   b. If not found, ***bogus observations*** are generated (winds are assimilated)

2. Advisory minimum sea-level pressure observations are then assimilated with other observations regardless of (1)

Example of Bogus Wind Assimilation

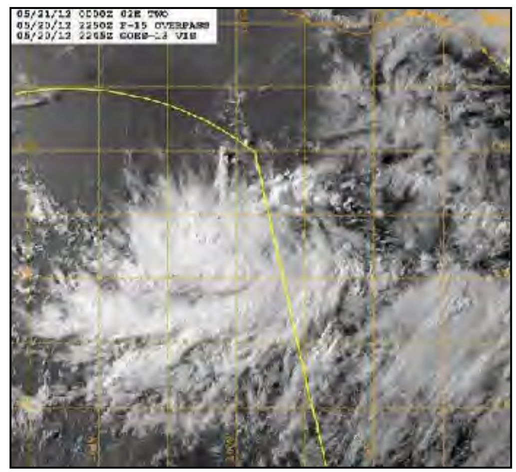
Generally rare in operations, Occurs mainly in genesis situation



Automated tracker “failed” to find coherent vortex to relocat

This can happen because:

- Distance from observation too large
- Too much tilt
- Parameters used to find position misaligned
- Nothing there

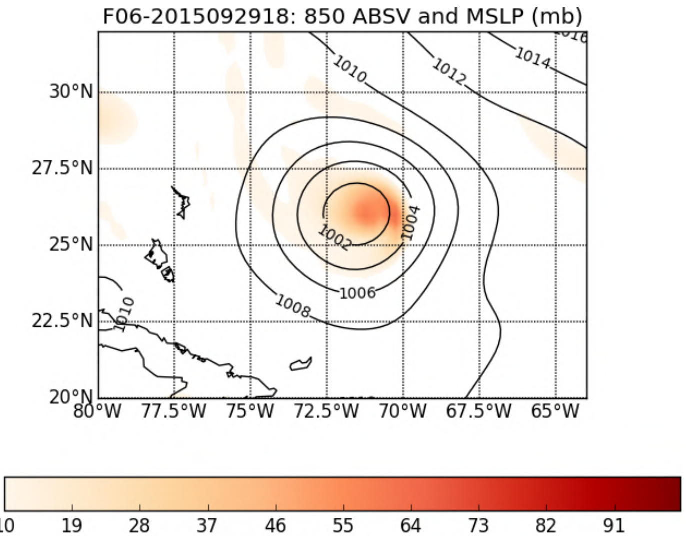


For Bud, tracker “failed” and resultant analysis had radically different vortex due to assimilation of bogus winds (and advisory minSLP)

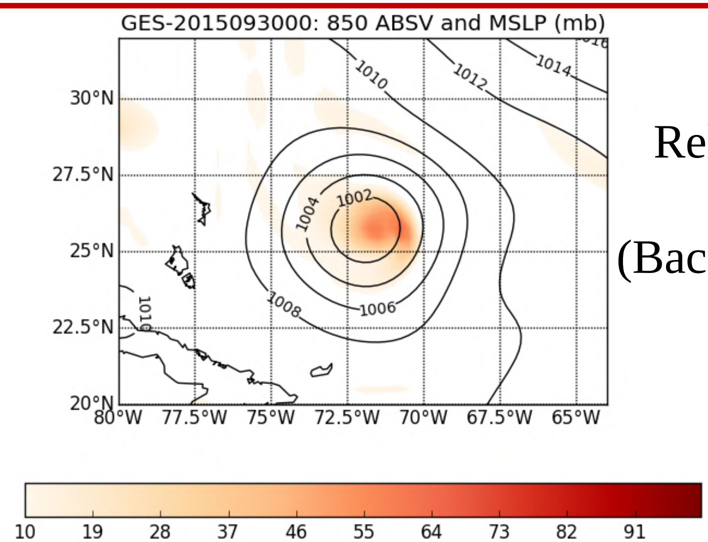
Impact of Relocation on Joaquin (2015093000)

Move Storm SW by ~0.5 degrees

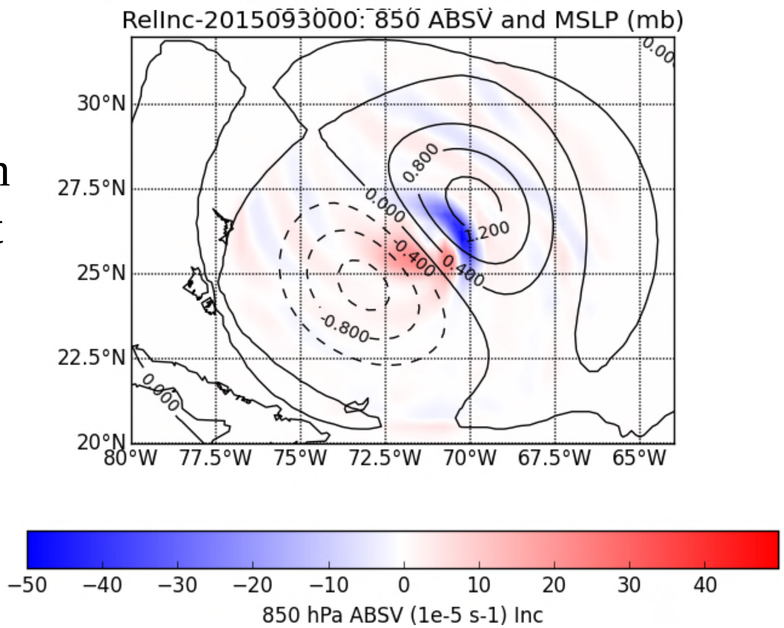
Original F06



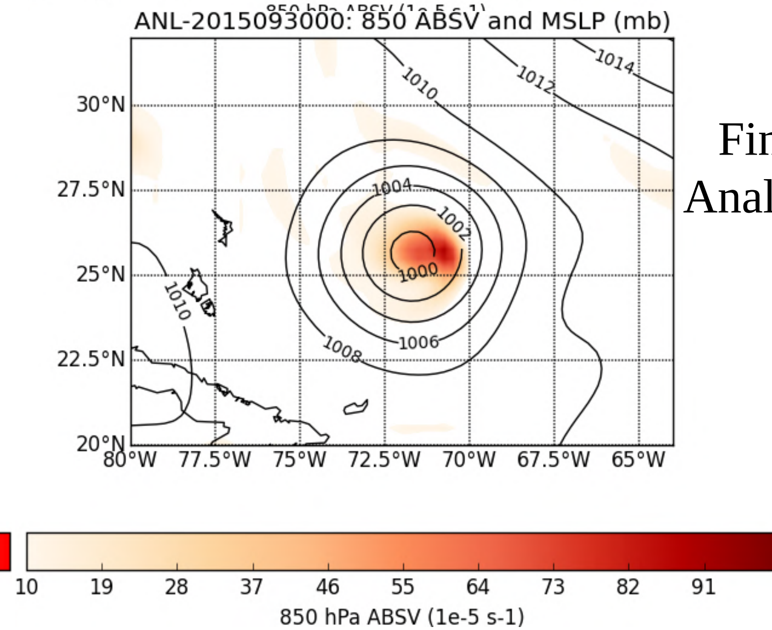
Relocated F06 (Background)



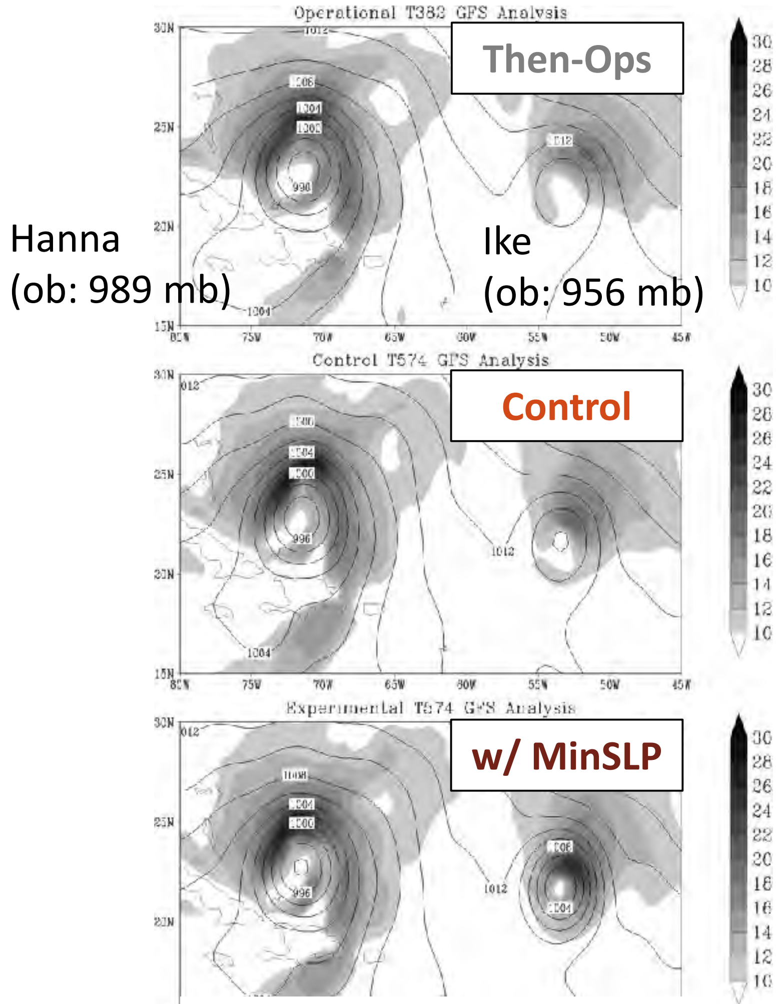
Relocation Increment



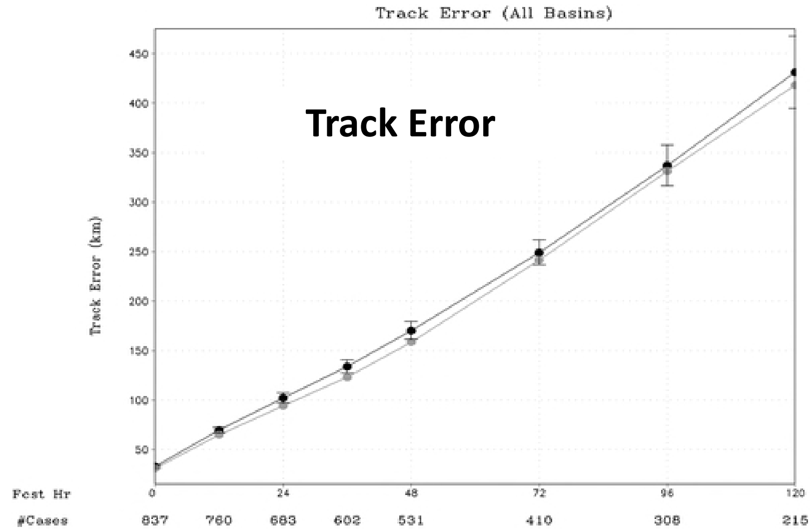
Final Analysis



Advisory MinSLP in GDAS/GFS (Kleist 2011)

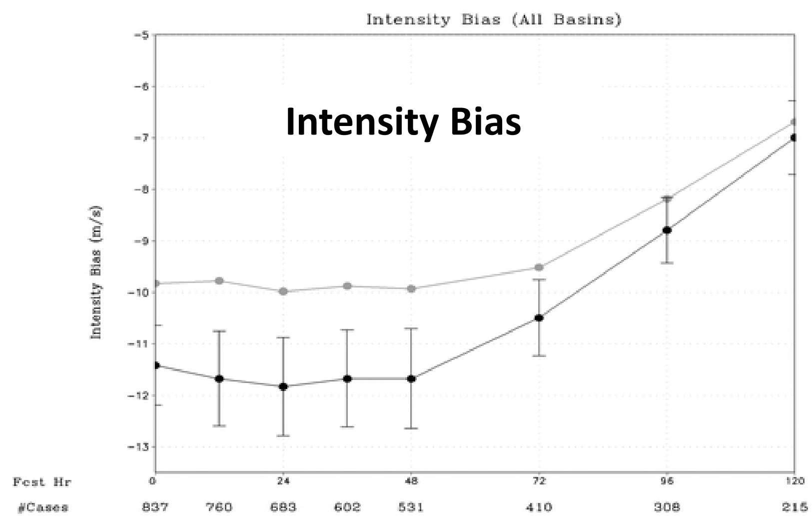


00 UTC 4 September 2008



Control

MinSLP



Control

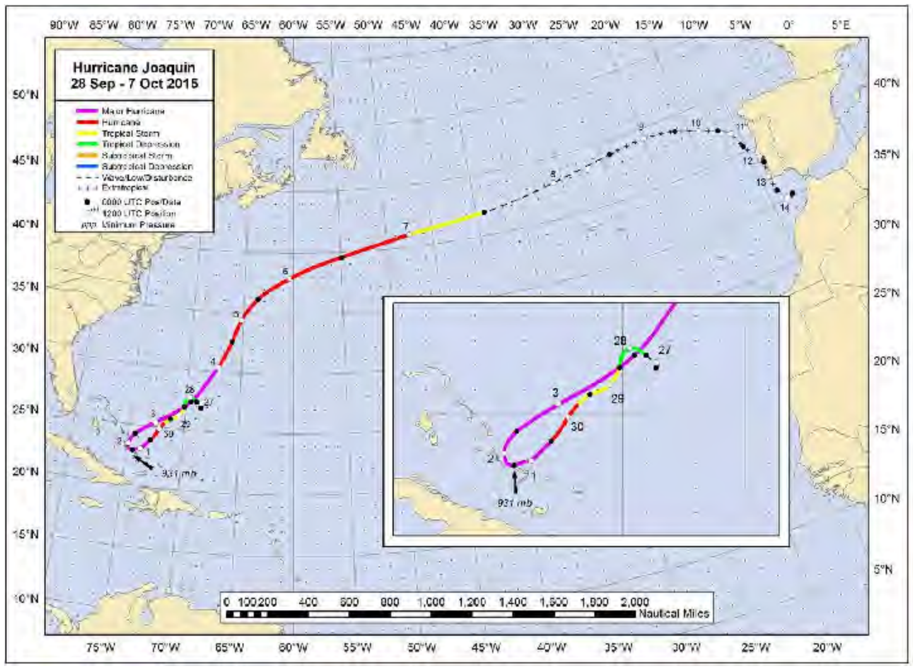
MinSLP

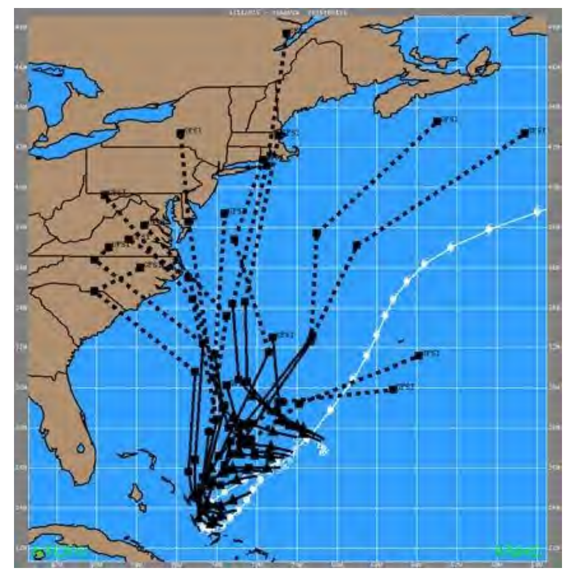
Met Office implemented this recently and have drastically reduced track errors.

Joaquin (2015) Experiment

- Motived by some preliminary experiments in 2012 prior to hybrid assimilation implementation in GDAS, decided to carry out a case study on Joaquin
- Fully-cycled (early and late cut-off) T1534L64 GFS with 80 member EnKF-based ensemble for hybrid data assimilation (3D EnVar)
- Control (with relocation) and Experiment (without) started prior to classification of Joaquin as depression
  - For experiment without relocation the effect is cumulative – we are not evaluating the impact of relocation on any individual operational forecast
- Bogus winds were never generated in operations, control, or experiment
- Advisory MinSLP assimilated into hybrid and EnKF for control and experiment

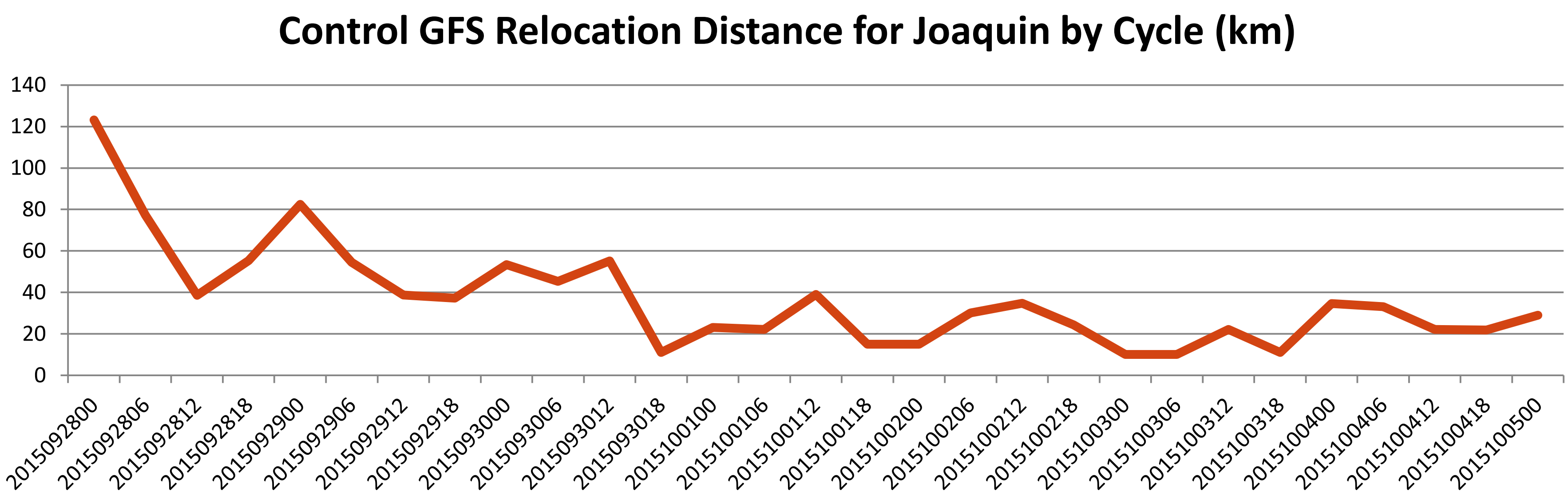
Hurricane Joaquin (2015)





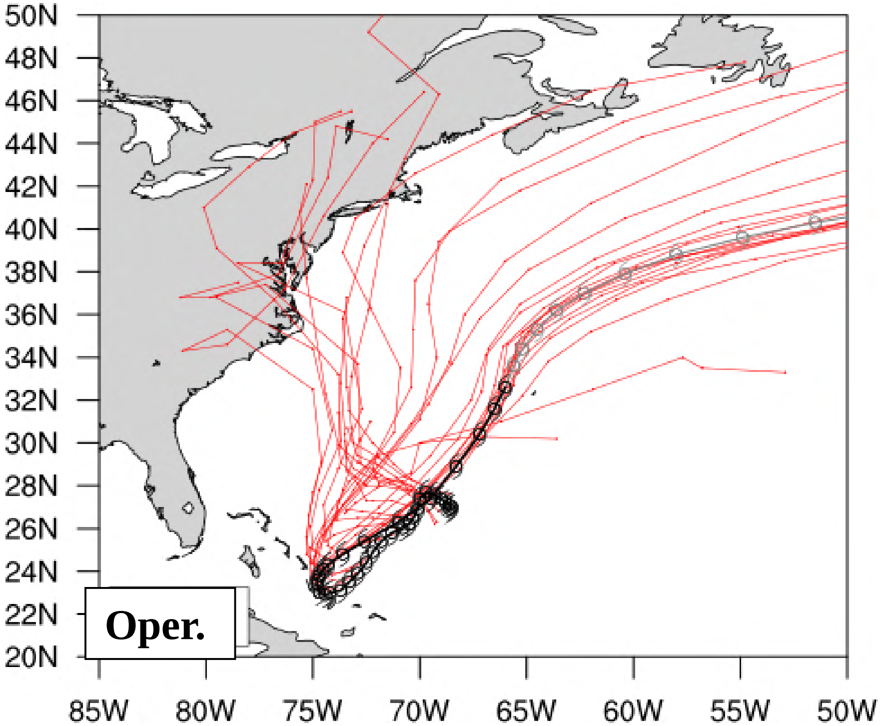
- High Impact in Bahamas
- Some guidance (GFS/HWRF) during early cycles advertised potential U.S. coastal impacts

Relocation in Control for Joaquin

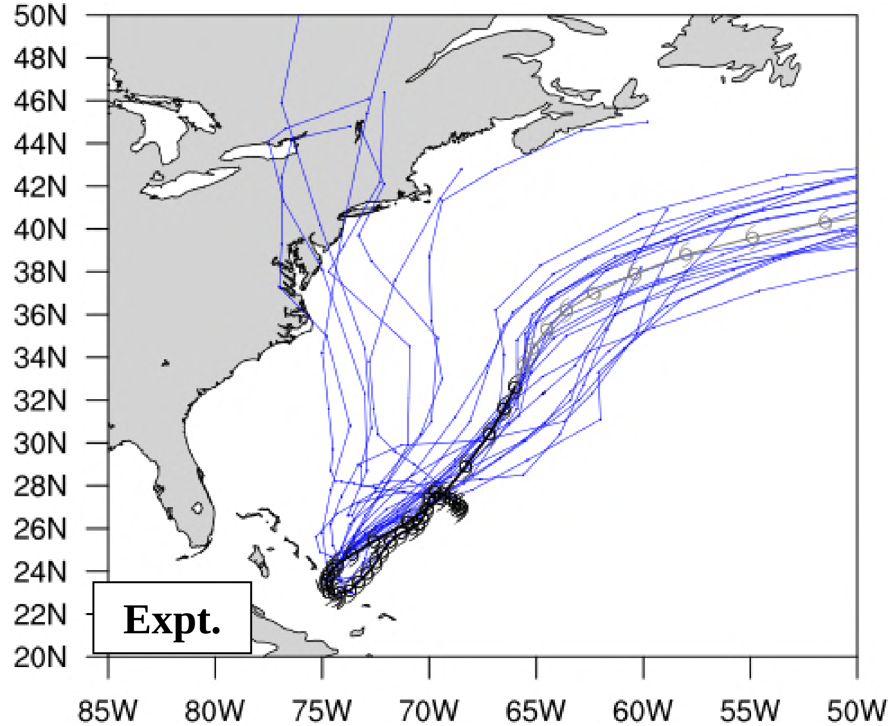


- During depression and TS phase, relocation distance larger than when storm reached hurricane status
- These are approximate – the tracker operates on quarter degree output and relocation is estimated to precision of tenths of degrees
- Also important to keep in mind that NHC analysis position has uncertainty

Track Summary for Experimental Period

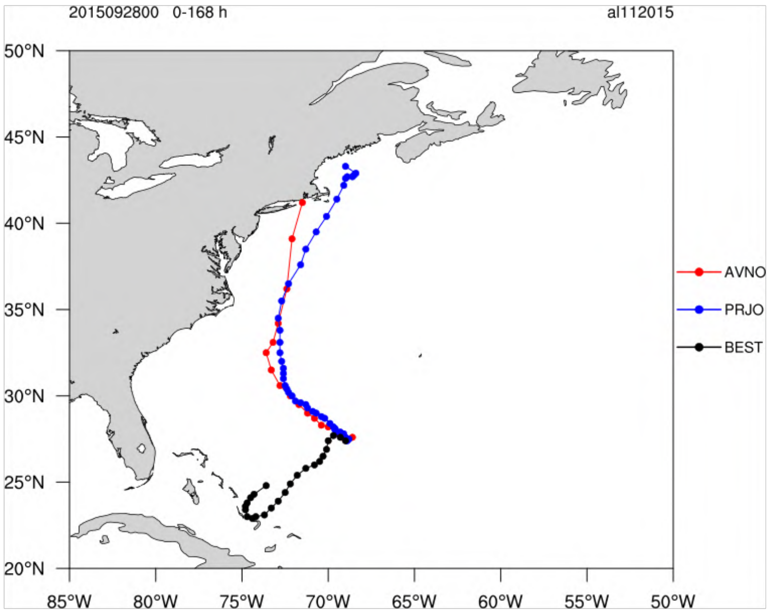


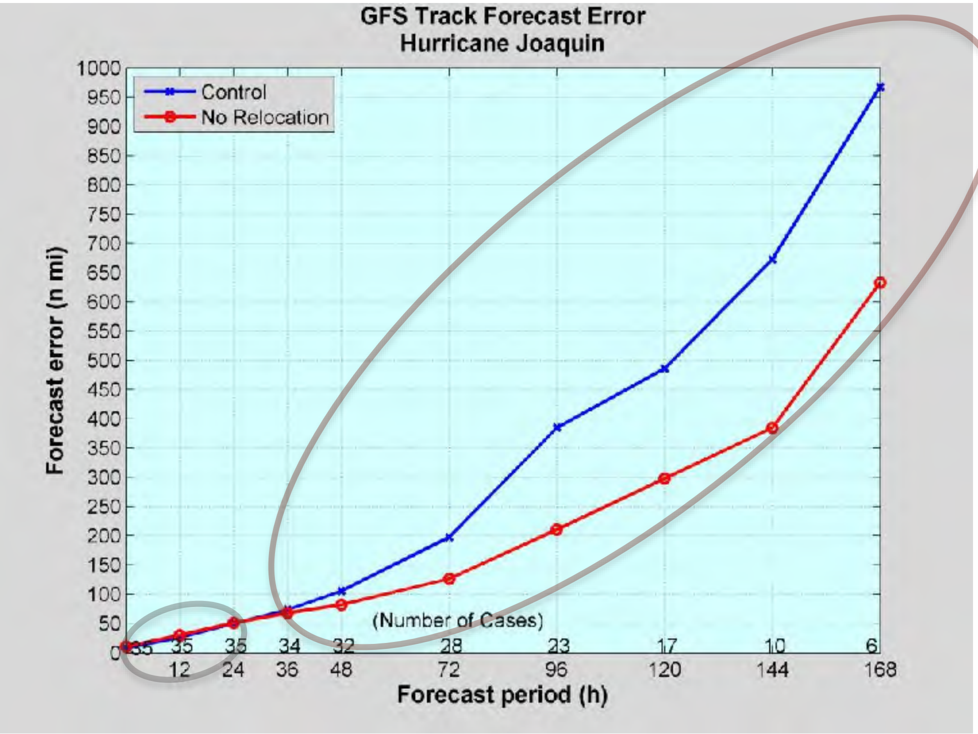
With Relocation



Without Relocation

Joaquin Individual Tracks and Mean Errors



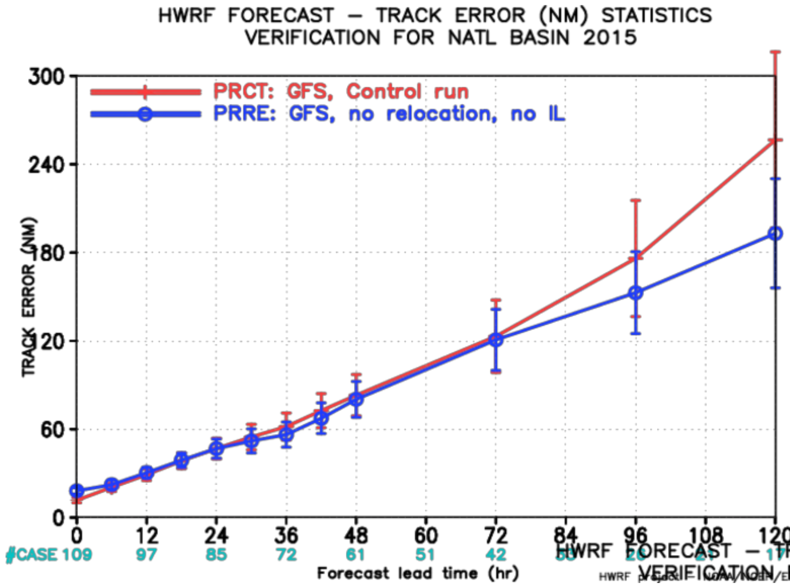


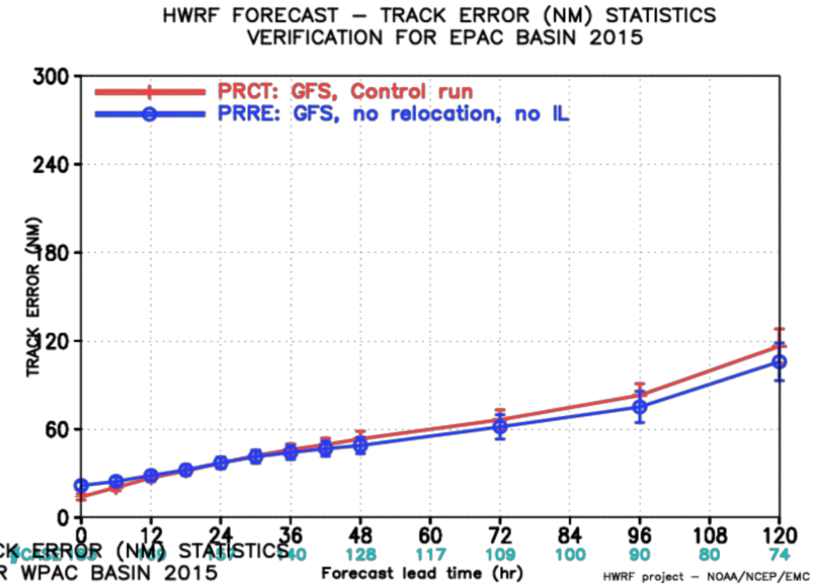
- ***No-relocation runs generally better beyond 24 hours***

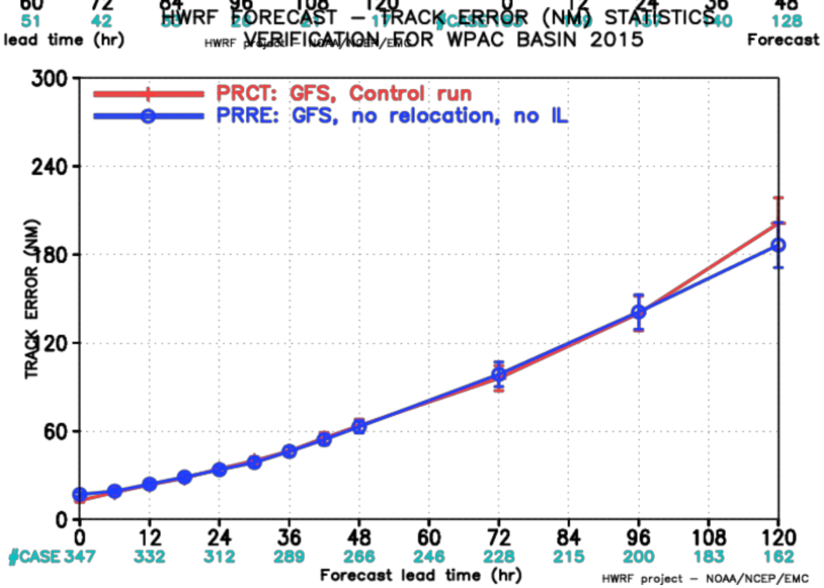
Full season at full resolution

- This sensitivity has prompted a further evaluation with full season cycling
  - Performed on new HPC asset at NOAA
- Fully-cycled (early and late cut-off) T1534L64 GFS with 80 member T574L64 EnKF-based ensemble for hybrid data assimilation (4D EnVar)
- Control (with relocation) and Experiment (without)
  - Experiment is simply turning off mechanical relocation and bogus vortex wind assimilation
- ***Advisory MinSLP assimilated into hybrid and EnKF for control and experiment***

Summary of Results (2 months so far)







NHC Early Track Processed Results % Change (EXP/CTL)

ATLANTIC

VT	NT	PRRI (No Relo)	PRCI (Ctrl)	PRRI-PRCI	Error Change (%)
0	71	5.9	5.9	0	0.00%
12	67	27	26.9	0.1	0.37%
24	61	43.1	45.6	-2.5	-5.48%
36	55	59.8	64.4	-4.6	-7.14%
48	49	75.8	82.9	-7.1	-8.56%
72	37	117.8	122	-4.2	-3.44%
96	23	155.2	187.1	-31.9	-17.05%
120	15	202.2	257.7	-55.5	-21.54%
144	7	189.5	238.5	-49	-20.55%
168	2	285.5	185.9	99.6	53.58%

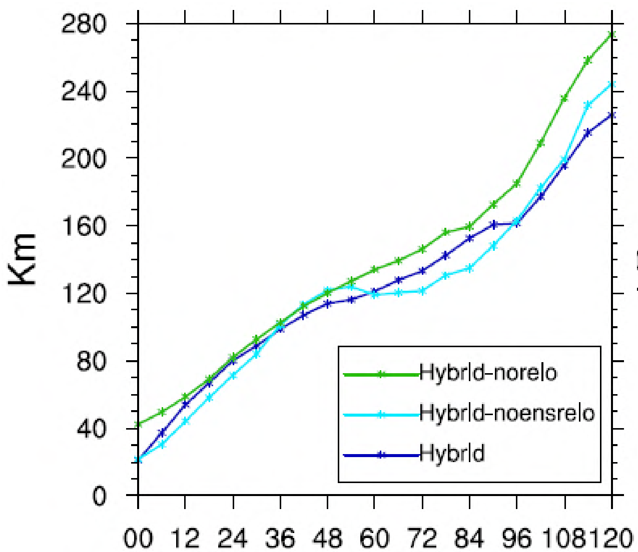
EAST PACIFIC

VT	NT	PRRI (No Relo)	PRCI (Ctrl)	PRRI-PRCI	Error Change (%)
0	62	9.2	9.2	0	0.00%
12	58	29	27.7	1.3	4.69%
24	54	44.2	41.8	2.4	5.74%
36	50	54.1	52.9	1.2	2.27%
48	46	62.3	61	1.3	2.13%
72	42	65.9	70.7	-4.8	-6.79%
96	38	70.3	89.3	-19	-21.28%
120	35	110.9	133.9	-23	-17.18%
144	35	163.7	193.6	-29.9	-15.44%
168	35	216.6	252.8	-36.2	-14.32%

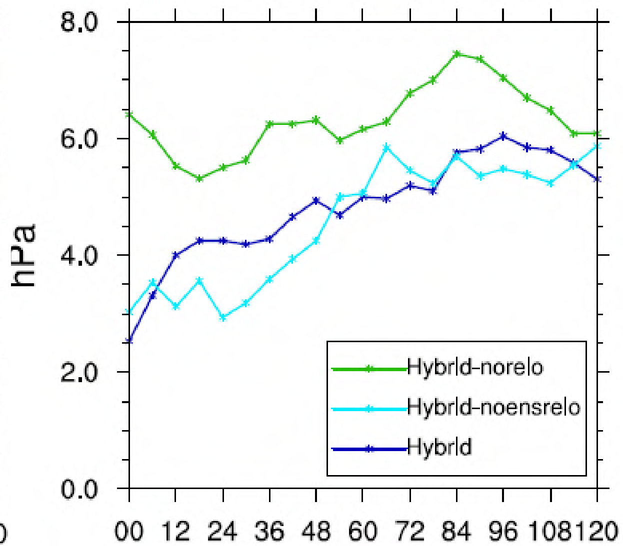
Impacts of vortex initialization and ensemble relocation

mean forecast errors for all cycles (Courtesy Xuguang Wang)

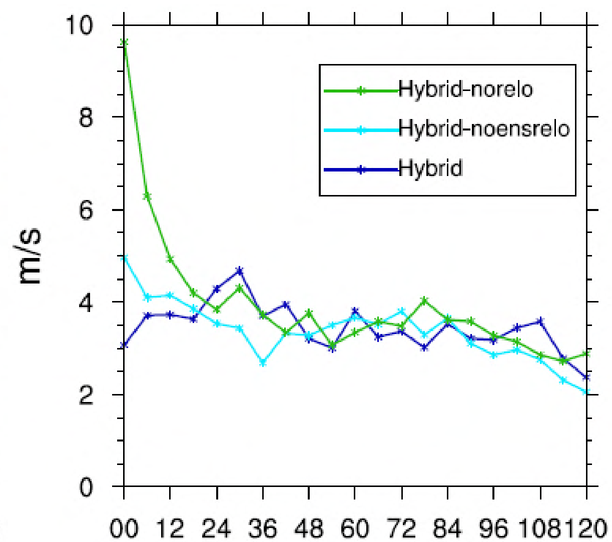
Track (km)



MSLP (hPa)



Vmax (m/s)

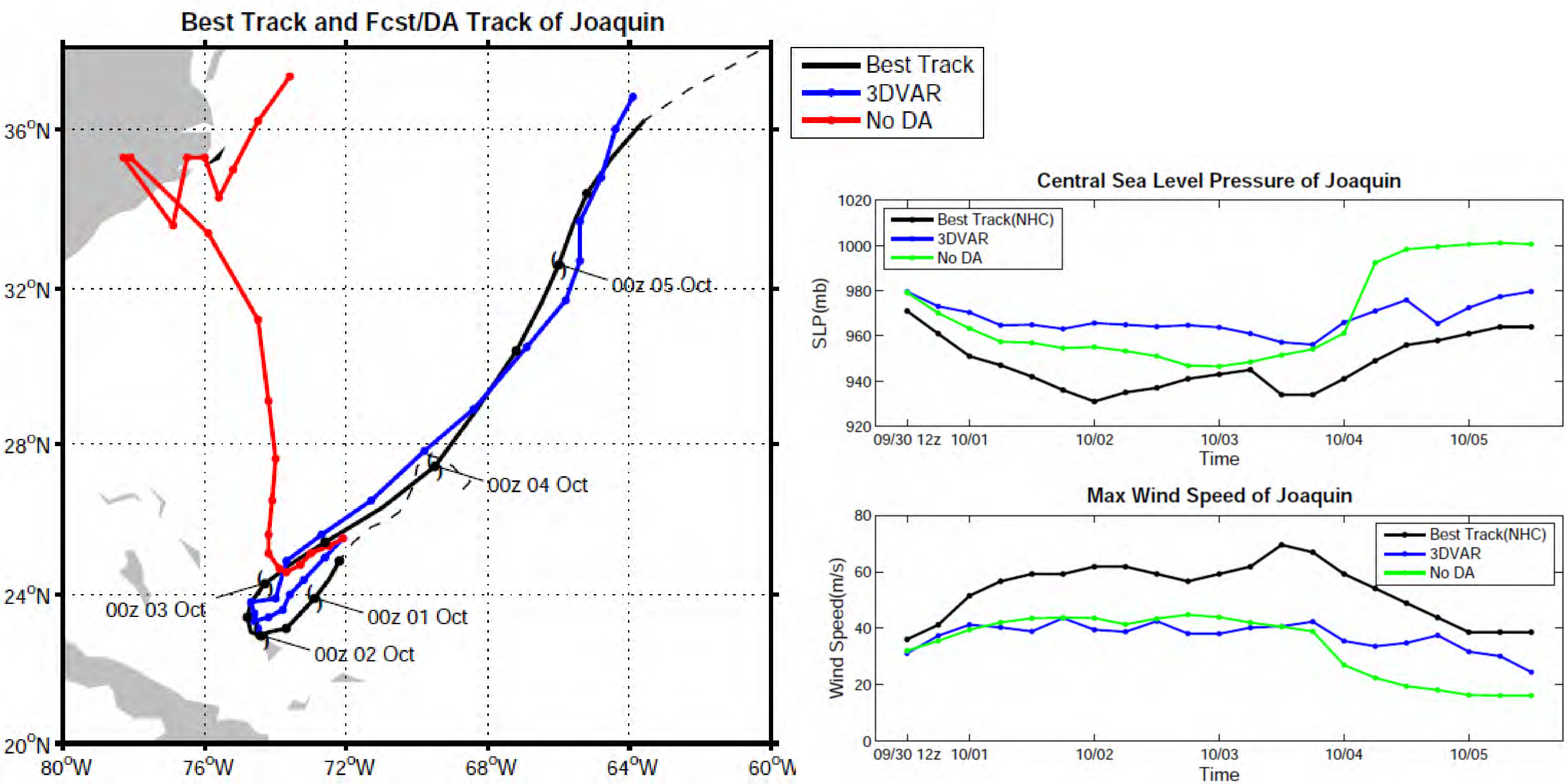


- Hybrid-norelo (green) performs worst.
- Hybrid-noensrelo (light blue) is better than Hybrid-norelo, but still worse than Hybrid (dark blue) especially for Vmax forecasts at earlier lead times.
- Together with the previous structure analyses, Hybrid with both ensemble relocation and control vortex relocation/initialization improves the forecasts compared to both Hybrid-noensrelo and Hybrid-norelo.

Testing Alternatives

- Exploring two alternatives directly in the DA
  - Direct assimilation of storm position (and size)?
    - Very difficult to build observation operators for VAR
  - Field Alignment
    - "DWRF" demonstrated for OSSE
- Hurricane Joaquin (2015) selected as first test case, completing a hierarchy of test simulations
  - Cold start WRF runs with GFS initial conditions
  - Ensemble WRF integrations from GFS initial conditions and pseudo-random initial perturbations
  - WRFVAR 3DVAR cycling

Very preliminary work:
WRF Simulations of Joaquin



DA helps with track, but intensity is not very good. Some of this due to resolution.

Summary

- Hybrid 4DEnVar now operational, many things to work on to improve system
  - 4D IAU tested as suitable replacement for DFI
  - Spatial/Spectral (Waveband) Localization and Weighting
    - Preliminary coding already done
  - Outer Loops
  - Ensemble size versus resolution
    - Lagging/Shifting as alternative

- TC relocation demonstrably worse for global simulations after 24h or so.
  - Trying to understand why
  - Does not seem to be general to regional, high resolution configuration

- NGGPS and FV3.....

Thank you!

- Backup slides on GSI MPI scalability issues.

GSI Scalability: Main MPI Domain Decomposition

- The main domain decomposition in GSI is done on evenly sized subdomains
  - All variables/levels stored
  - Schematic example for 50 MPI tasks

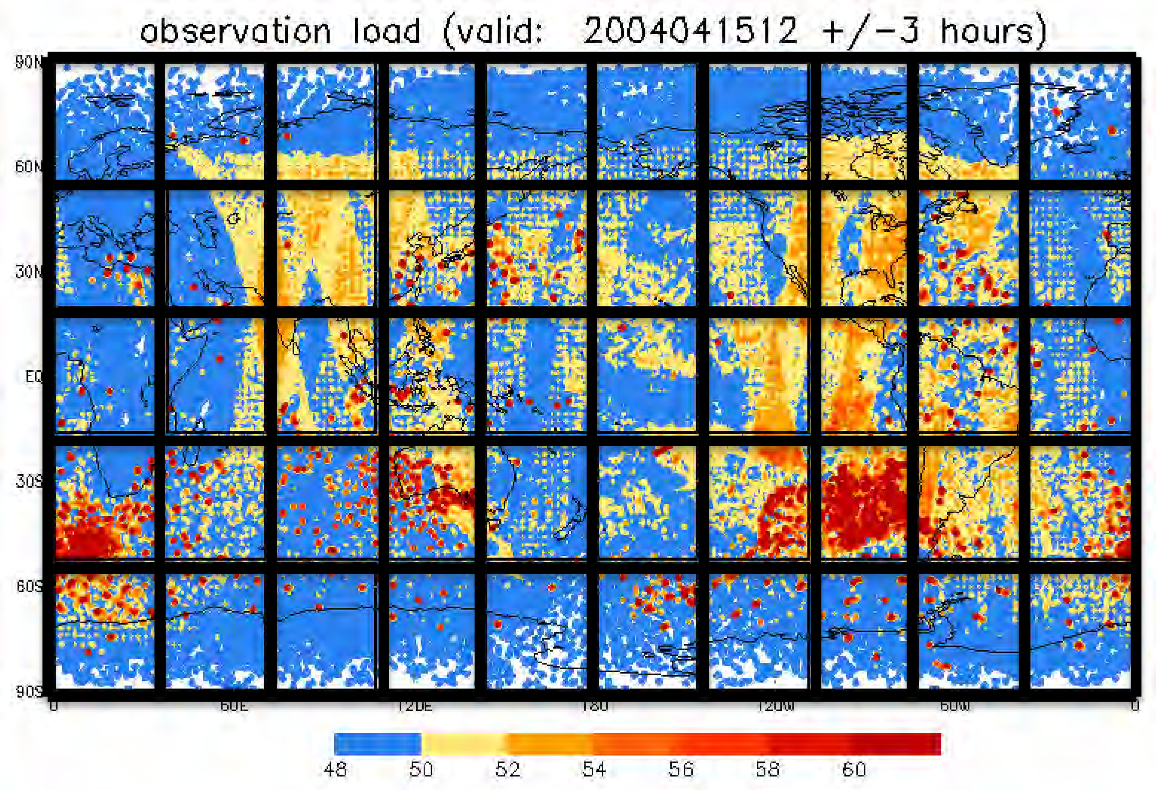
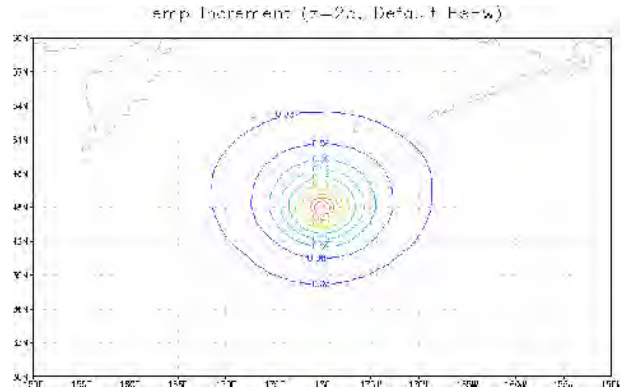


Figure adapted from Treadon 2005, 1st GSI User's Workshop

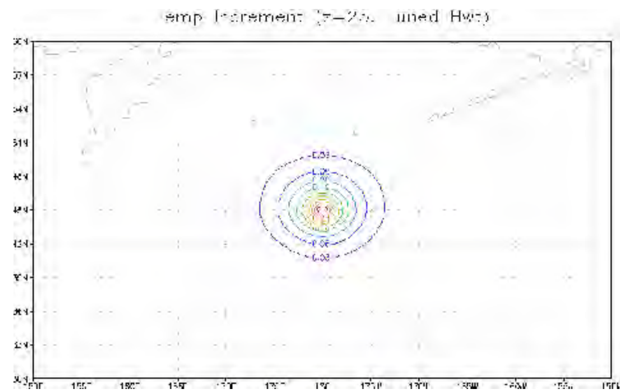
Static Error Covariance in GSI

- Horizontal correlations applied using series of recursive filters
  - Sum of three Gaussians to model fat-tailed spectrum
- For global grids, this requires special treatment to avoid end points at the pole



Hzscl = 1.7, 0.8, 0.5

Hswgt = 0.45, 0.3, 0.25

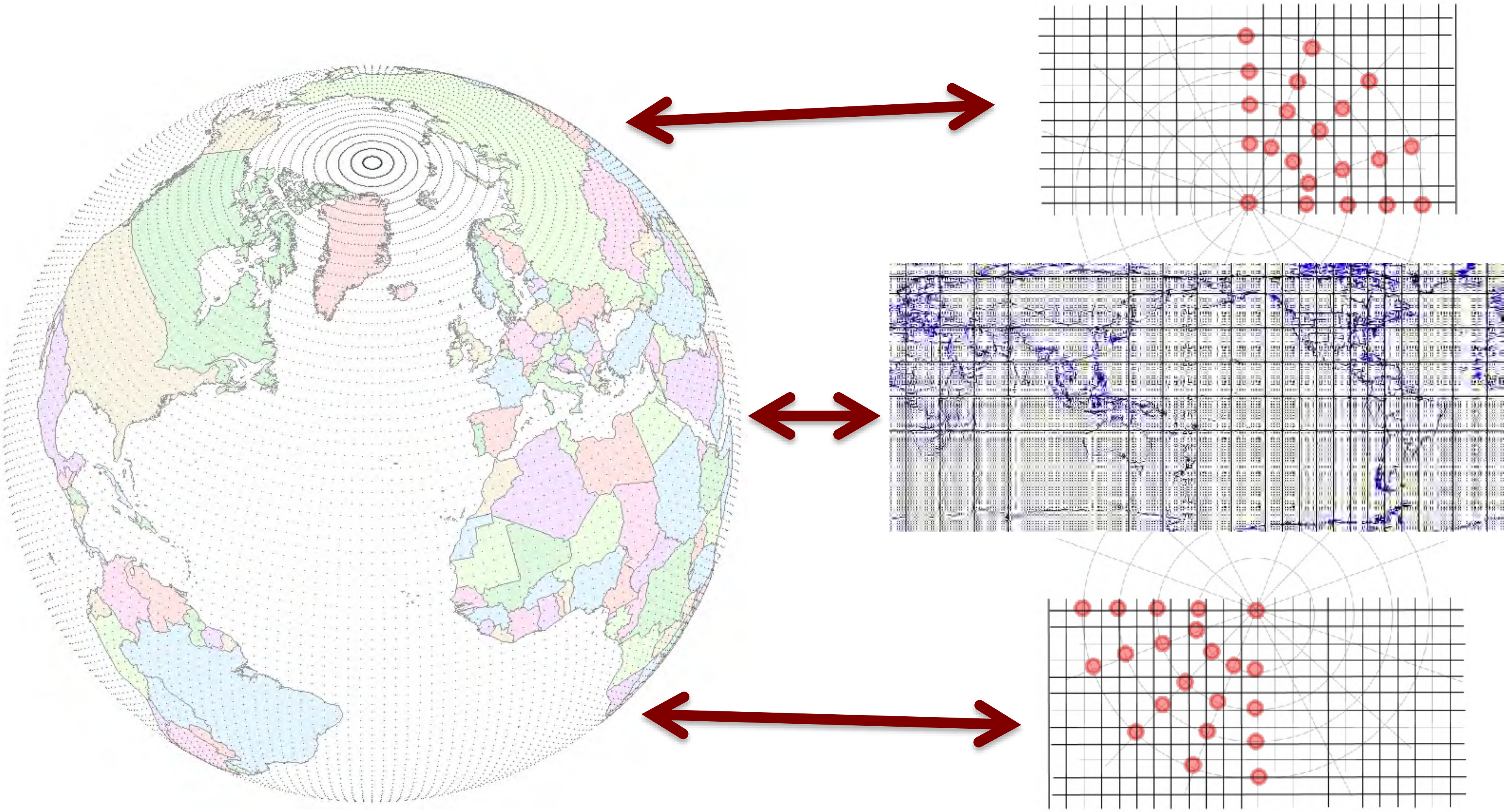


Hzscl = 1.7, 0.8, 0.5

Hswgt = 0.1, 0.3, 0.6

500 hPa temperature increment (K) from a single temperature observation utilizing GFS default (top) and tuned (weights for scales) error statistics.

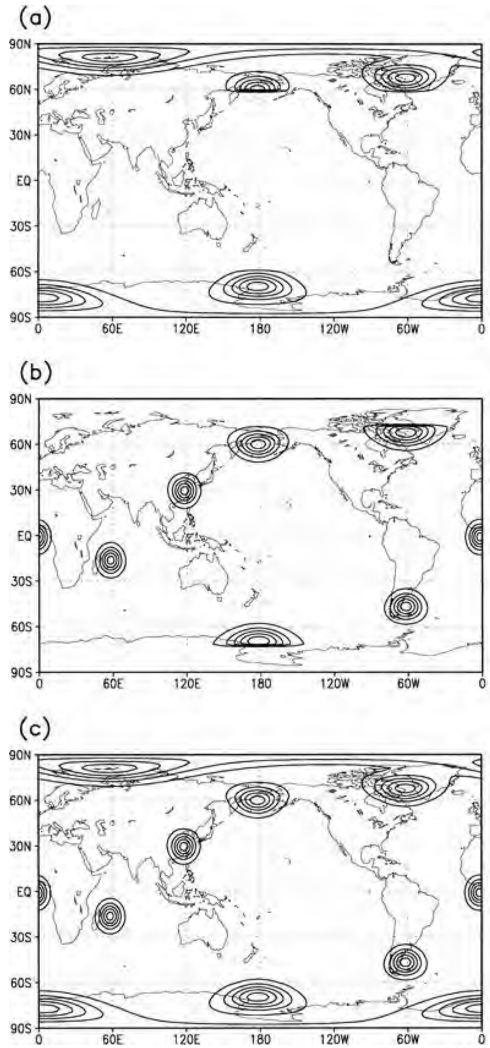
Analysis versus Recursive Filter Grids



Result after Merge of Patches

Results of recursive filter on some randomly located impulses in (a) the polar patches, (b) the zonal band, and (c) merged from the subdomains without blending. The contour levels are 0.1, 0.3, 0.5, 0.7, and 0.9

From Wu et al. (2002)



Impact on communications

- Horizontal recursive filters (***as currently coded***) requires data on ***full 2d horizontal slabs***
  - ***This also has memory implications***

- Transform (Columns/Slabs) requires massive data movement

- The all to all communications limit scalability within the minimization
  - To maximize distribution of work for 2D slabs/horizontal RFs, one could place one variable from one level on each MPI task (and thread therein)
  - Current global gsi: 6 3D control variables, 64 vertical levels, + 4 surface variable (pressure and skin temperature-water/land/ice)…or 388 slabs

- ***Some*** of the requirement to have 2D slabs is a function of the desire to have reproducibility across different task counts

Scalability tests on WCOSS (courtesy Russ Treadon)

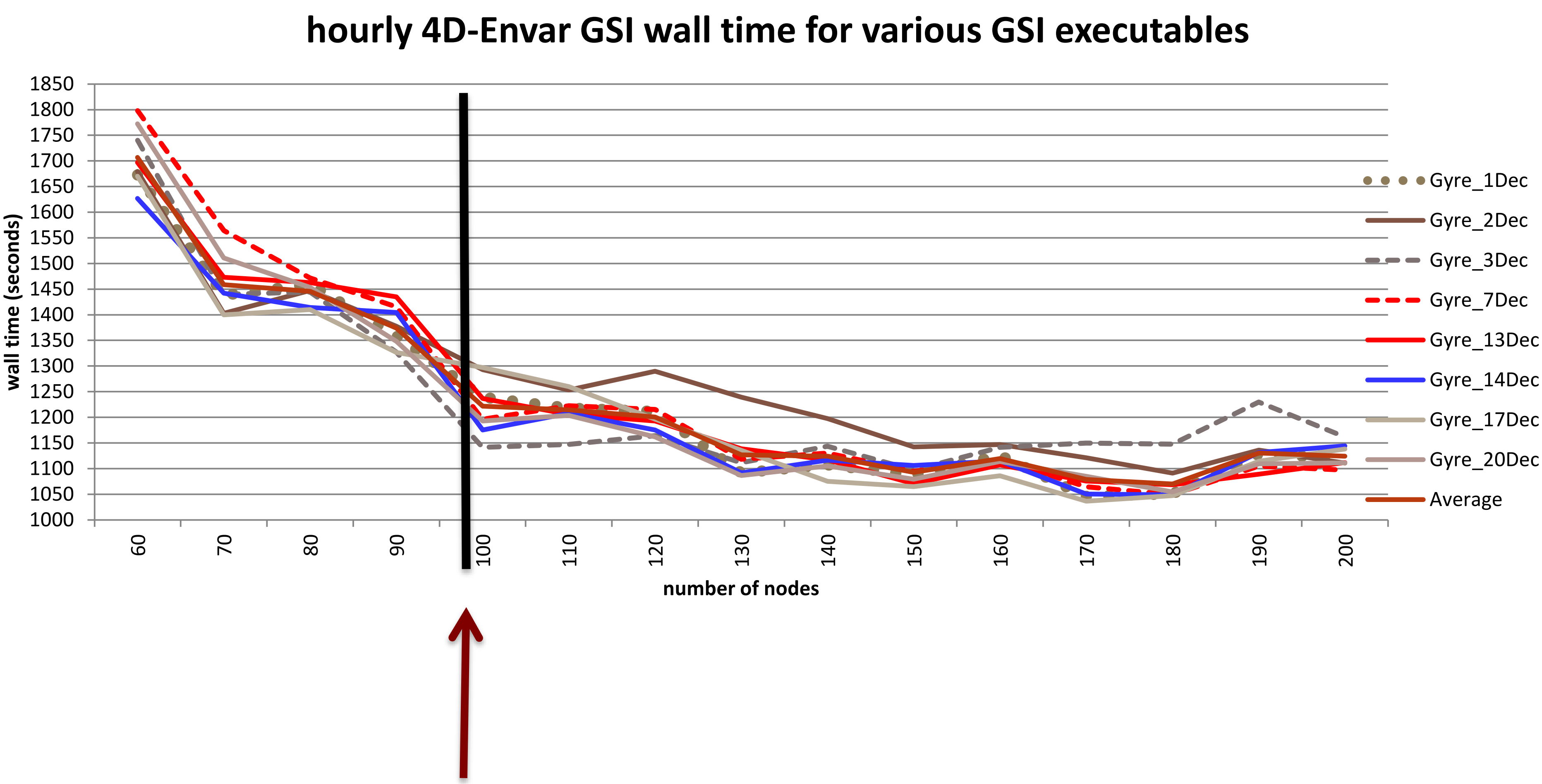
- Series of GSI runs were carried out on WCOSS to test timing in preparation for May implementation of hybrid 4D EnVar

- Configurations utilized hybrid openmp/mpi
  - 4 MPI tasks/node with 4 threads each (*)

- Preoperational configuration:
  - Hybrid 4D EnVar with hourly bins within 6 hour window (7 slots), 2x100 iterations, 12.5% static error contribution, full complement of observations, TLNMC on all time levels (I/O is non-trivial because of hourly states)
  - ***Horizontal localization for ensemble control variable is done using spectral operators which requires substantial communication as well***

- Tests carried out for several days, increasing the node (processor) count in increments of 10 (40)

Scalability tests on WCOSS (courtesy Russ Treadon)



97 nodes = 388 MPI tasks in these runs

What's the point?

- Recursive filters have served us well, but current implementation is already limiting spatial resolution possibilities for the operational window (20 minutes)

- 12.5% weighting. Can/should be build static covariance model that is highly scalable?

- Some thought should be put into using "global" operators given potential implications for MPI bottlenecks
  - 2D recursive filters
  - Spectral operators*

Longer term

- Desire to eliminate the need for these global-type operators

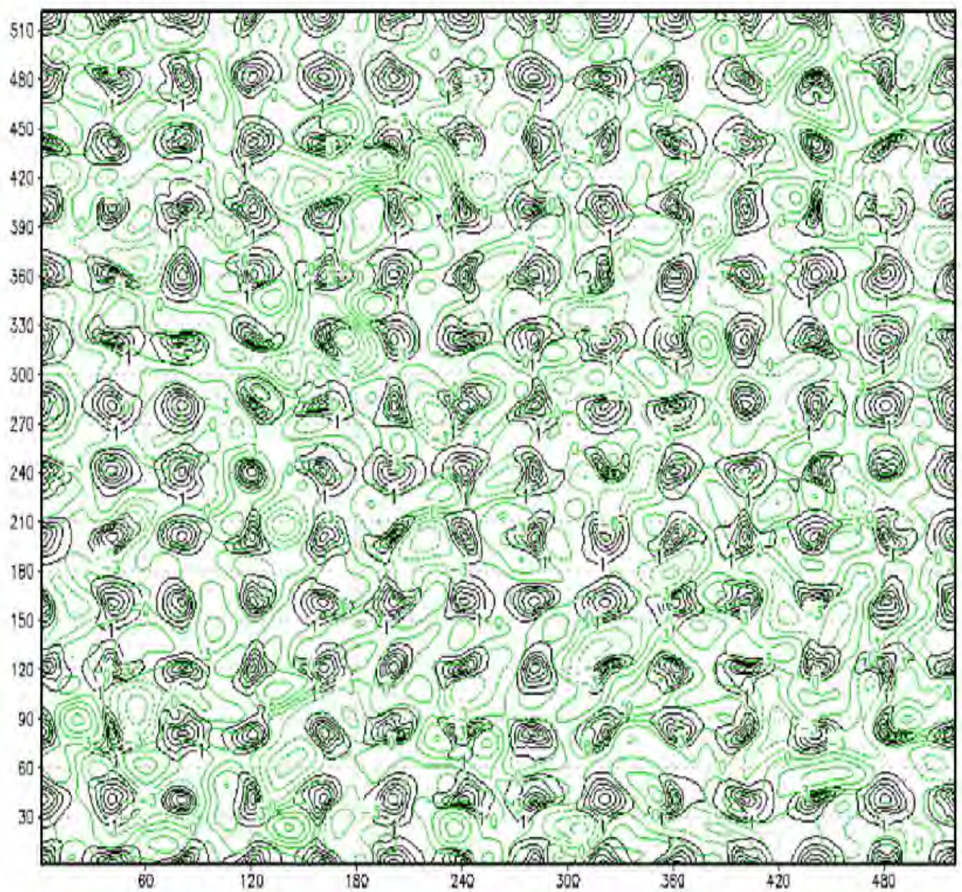
- One potential promising avenue is to use a multigrid algorithm
  - Multigrid correlation algorithm has been tested for both isotropic and anisotropic correlation functions, with a demonstration test using the RTMA 2-D application of GSI.
  - In addition to being applicable to correlation functions, potential exists to use multigrid solvers as an alternative to conjugate gradient
    - Cioaca et al. (2013), Debreu et al. (2013), Gratton et al. (2013), Kang et al. (2014), others

- Certainly other options out there
  - Diffusion operator-based

Example correlation function tested in 2D

$$\mathbf{C}(\mathbf{x}_1,\mathbf{x}_2,f_1,f_2) = \exp\left[- |\mathbf{x}_2 - \mathbf{x}_1|^2 \,/\, (L_{xy}(\mathbf{x}_1)\, L_{xy}(\mathbf{x}_2)) - ((f_2 - f_1)/L_f)^2 \right]$$

inhomogeneous isotropic Riishøjgaard anisotropic



521 x 521 grid Riishøjgaard coupled correlation. Test point every 40 grid points in x and y, 194 samples all together. Green contours are for field that the correlations follow. Black contours are 0.1, 0.3, 0.5, 0.7, 0.9. Correlation computed to accuracy of 0.01 with cubic interpolation grid transfers. Cost to setup information needed for multigrid algorithm: 214 seconds. Cost for v = C*u : 0.07 seconds. Need to work on the setup cost.