Balance Operator in Ensemble Data Assimilation

Kayo Ide, University of Maryland

Cathy Thomas, University of Maryland

Zhijin-Li, NASA JPL

James C. McWilliams, UCLA

With contributions from others

(Special thanks to Jim Purser, Daryl Kleist & Fred Kucharski)

Objectives

- Study effects of balance operator in data assimilation schemes:
  - Hybrid 4DEnVar
  - Local ensemble transform Kalman filter (LETKF)

- Demonstrate how different localization schemes impact the effectiveness of the balance operator.
  - Model space (**B** localization)
  - Observation space (**R** localization)

- Apply for case studies
  - Atmosphere: SPEEDY
  - Ocean: ROMS

Balance Operator: Basics

- Imbalances in initial conditions can degrade forecast skill through the production of fast moving gravity waves

- Conventionally present in variational schemes

- Represents the physical relationship between the variables

- Relationships are defined by regression coefficients, derived from climatological information (Wu etal 2012)

Total = Balanced + Unbalanced

$\Delta\psi = \Delta\psi$

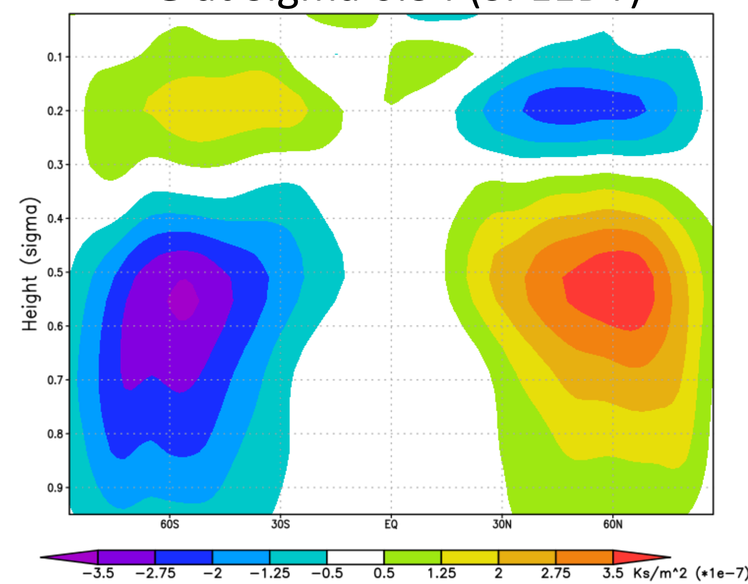
$\Delta\chi = \mathbf{c}\ \Delta\psi + \Delta\chi^{u}$

$\Delta T = \mathbf{G}\ \Delta\psi + \Delta T^{u}$

$\Delta p = \mathbf{\Omega}\ \Delta\psi + \Delta p^{u}$

$\Delta q = \Delta q$

G at Sigma 0.34 (SPEEDY)



Balance Operator: Variational Application

- Transformation of control variables
  - Basic form

$$\Delta \mathbf{x} = \Gamma \Delta \mathbf{z}$$

$$\begin{pmatrix} \Delta\psi \\ \Delta\chi \\ \Delta\mathbf{T} \\ \Delta\mathbf{p} \\ \Delta\mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{c} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{G} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Omega} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \Delta\psi \\ \Delta\chi^u \\ \Delta\mathbf{T}^u \\ \Delta\mathbf{p}^u \\ \Delta\mathbf{q} \end{pmatrix}$$

Γ is invertible:

$$\Delta \mathbf{z} = \Gamma^{-1} \Delta \mathbf{x}$$

  - Preconditioned control
    - Climatological covariance **P** = **B** = **UU**$^{\mathrm{T}}$ with control **v**

$$\Delta \mathbf{x} = \Gamma \mathbf{U} \mathbf{v}$$

    - Ensemble covariance **P** = **C** ∘ **ZZ**$^{\mathrm{T}}$ with control **α** & localization **C**

$$\Delta \mathbf{x} = \Gamma \sum_{m=1}^{M} \mathbf{F} \alpha^m \circ (\mathbf{z}_e^{\mathrm{b}})_k^m$$

→ recursive filter **F**

Clayton etal (2013), Purser et al (2002)

Balance Operator: Variational Application

- Hybrid 4DEnVar Cost function formulation with Balance

Clayton et al (2013),Lorenc et al (2015)

$$J(\mathbf{v},\alpha) = \beta^c \frac{1}{2}\mathbf{v}^T\mathbf{v} + \beta^e \frac{1}{2}\alpha^T\alpha + \sum_{k=1}^{K}\frac{1}{2}(\mathbf{d}_k - \mathbf{H}_k\Delta\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{d}_k - \mathbf{H}_k\Delta\mathbf{x}_k)$$

$$\Delta\mathbf{x}_k = \Gamma(\beta^c\mathbf{U}\mathbf{v} + \beta^e\sum_{m=1}^{M}\mathbf{F}\alpha^m \circ (\mathbf{z}_e^{\mathrm{b}})_k^m)$$

- 4D increment: linear combinations of
  - 4D ensemble perturbations
  - static contribution
- Climatological component contribution is time invariant here
- Balance is enforced by **Γ** operator

Balance Operator: LETKF Application

- Cost function formulation: Hunt et al (2007)

$$J(\mathbf{w}) = \frac{M-1}{2}\mathbf{w}^T\mathbf{w} + \sum_{k=1}^{K}\frac{1}{2}(\mathbf{d}_k - \mathbf{Y}_k^{\mathrm{b}}\mathbf{w})^T\mathbf{R}_k^{-1}(\mathbf{d}_k - \mathbf{Y}_k^{\mathrm{b}}\mathbf{w})$$

$$\mathbf{Y}_k^{\mathrm{b}} = \mathbf{H}\,\Gamma\,\mathbf{Z}_k^{\mathrm{b}} = \mathbf{H}\mathbf{X}_k^{\mathrm{b}}$$

Γ is invertible:

$$\Delta\mathbf{z} = \Gamma^{-1}\Delta\mathbf{x}$$

- LETKF analysis in transformed control

$$\Delta\mathbf{z}_k^{\mathrm{a}} = \mathbf{Z}_k^{\mathrm{b}}\bar{\mathbf{w}}^{\mathrm{a}} \qquad \bar{\mathbf{w}}^{\mathrm{a}} = (\mathbf{W}^{\mathrm{a}})^2\mathbf{Y}_k^{\mathrm{b}}\mathbf{R}_k^{-1}\mathbf{d}_k$$

$$\mathbf{Z}_k^{\mathrm{a}} = \mathbf{Z}_k^{\mathrm{b}}\mathbf{W}^{\mathrm{a}} \qquad (\mathbf{W}^{\mathrm{a}})^{-2} = \nabla_{\mathbf{w}}^2 J(\mathbf{w})$$

Locally, **Γ** has no effect on $\bar{\mathbf{w}}^{\mathrm{a}}$ and $\mathbf{W}^{\mathrm{a}}$

- LETKF analysis: after obtaining increment in **z** globally then apply **Γ**

$$\Delta\mathbf{x}_k^{\mathrm{a}} = \Gamma\Delta\mathbf{z}_k^{\mathrm{a}} \qquad \mathbf{z}_k^{\mathrm{a}} = \mathbf{Z}_k^{\mathrm{b}}\bar{\mathbf{w}}^{\mathrm{a}}$$

$$\mathbf{X}_k^{\mathrm{a}} = \Gamma\mathbf{Z}_k^{\mathrm{a}} \qquad \mathbf{Z}_k^{\mathrm{a}} = \mathbf{Z}_k^{\mathrm{b}}\mathbf{W}^{\mathrm{a}}$$

Effect of Localization

- Motivation
  - For grid points that are at a large distance from each other where the true correlation is small, sampling error likely dominates (Hamill et al 2001).
  - Localization eliminates spurious, long-distance correlations that are likely unphysical.

- Types of localization
  - **B** localization

    Houtekamer & Mitchel 2001
    - Eliminates correlation between grid points that are distant
    - Used in EnVar

$$\mathbf{U}^e = \begin{pmatrix} \mathbf{F} & & & \mathbf{0} \\ & \ddots & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{F} \end{pmatrix}$$

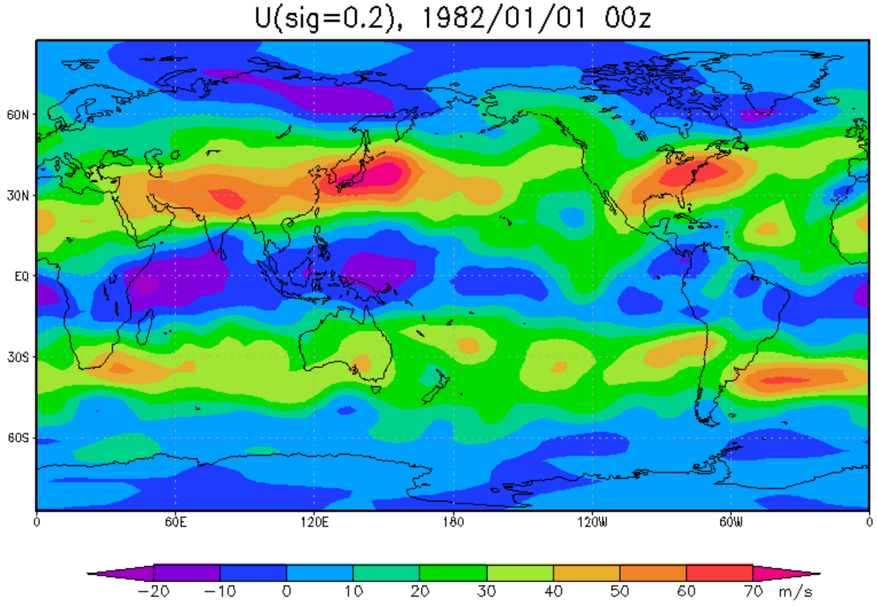
  - **R** localization

    Hunt et al 2007
    - Reduces the influence of observations that are distant
    - Used in LETKF

$$\mathbf{R}^{-1} \Rightarrow \rho_{loc} \circ \mathbf{R}^{-1}$$

Application to SPEEDY (Molteni 2003)

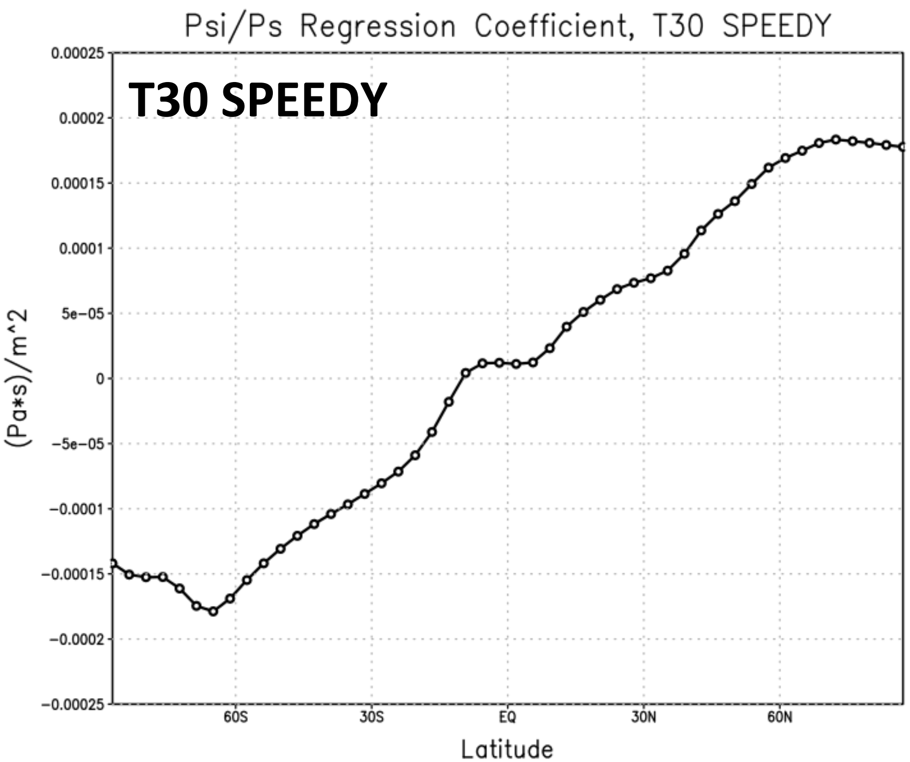
- Model Description
  - **S**implified **P**arameterizations, primitiv**E**-**E**quation **DY**namics
  - Global atmospheric GCM
    of intermediate complexity
- Version 41
  - Provided by Fred Kucharski (ICTP)
  - 3 horizontal resolution options:
    T30, T47, T63
  - 8 vertical levels

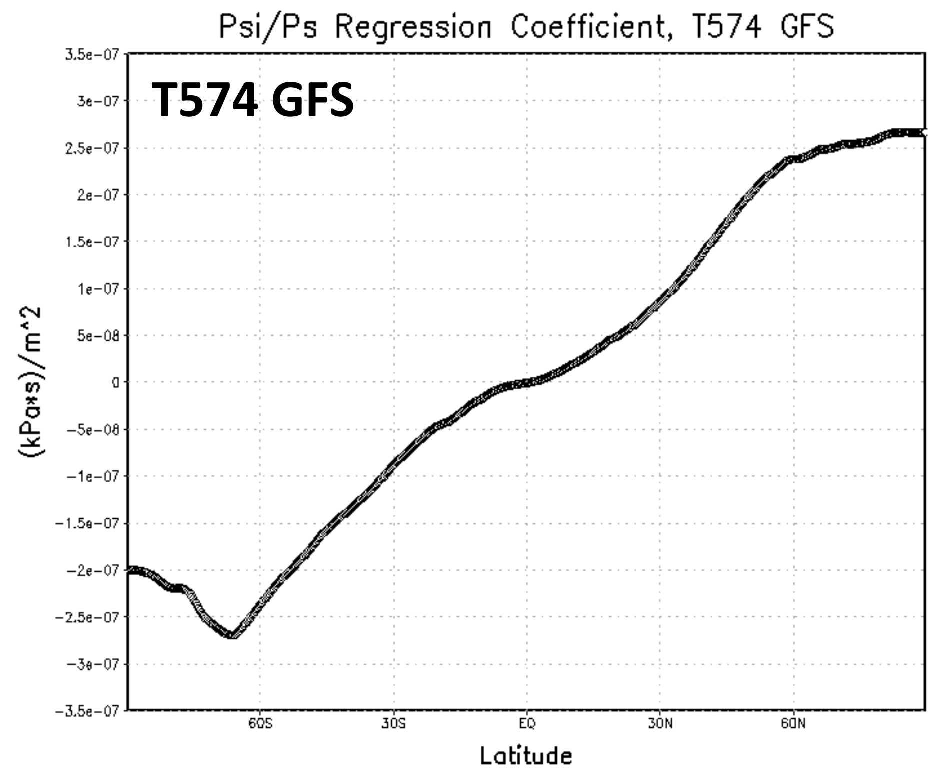


- Output every hour (addition by Miyoshi and Greybush)

Realism of SPEEDY for Balance Experiments

- Regression coefficient **Ω**: $\Delta p = \mathbf{\Omega}\, \Delta\psi + \Delta p^{u}$



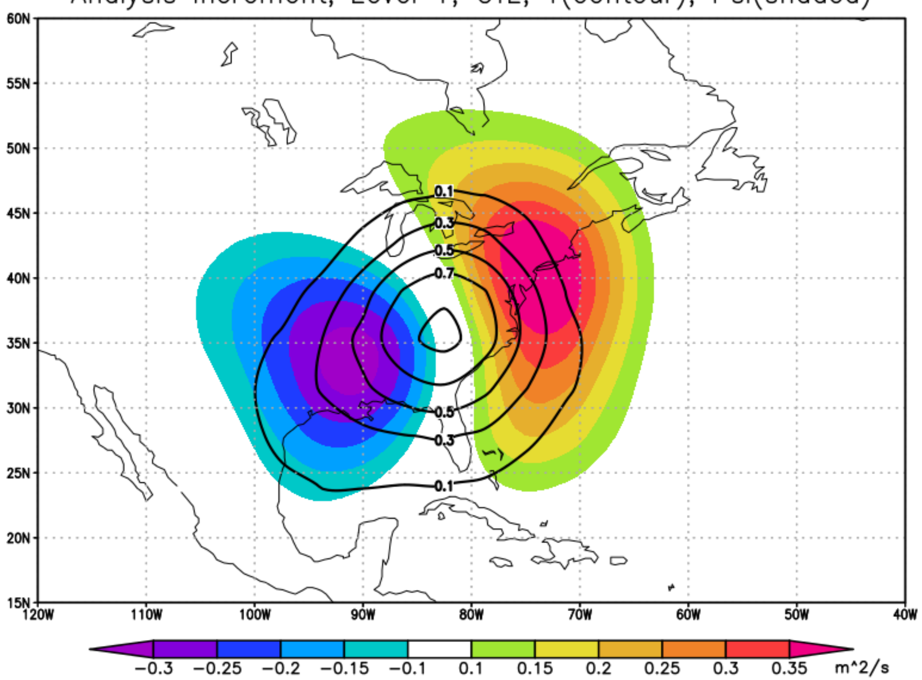


Balance in SPEEDY is reasonable to higher resolution global model based on balance operator

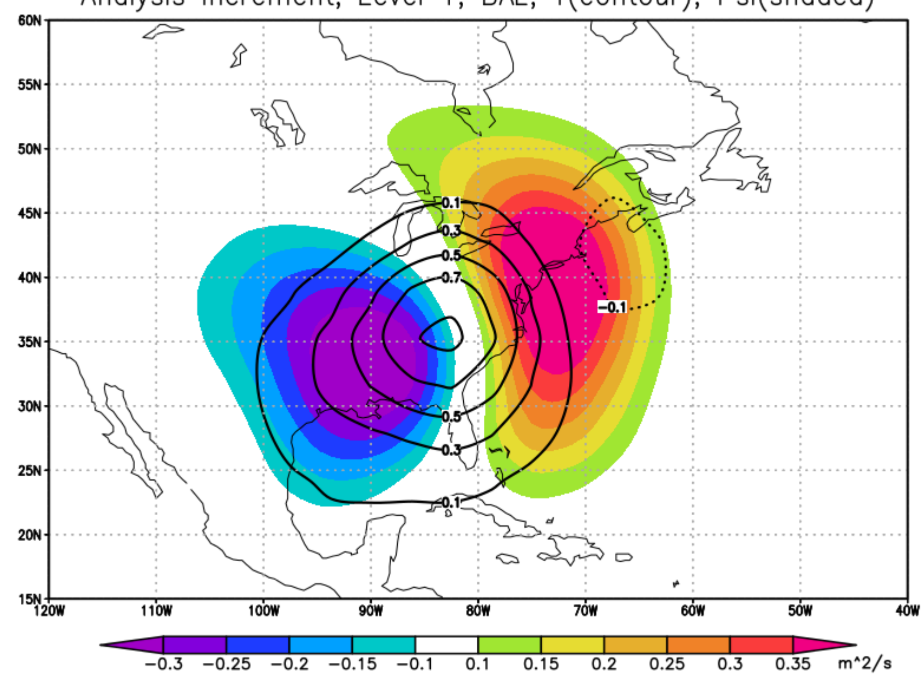
Single Observation (*T*): 4DEnVar (β^e=1). CTL vs BAL

T - Contoured
ψ - Shaded

CTL – Base Configuration
without balance operator in the ensemble



BAL – With balance operator
applied in the ensemble

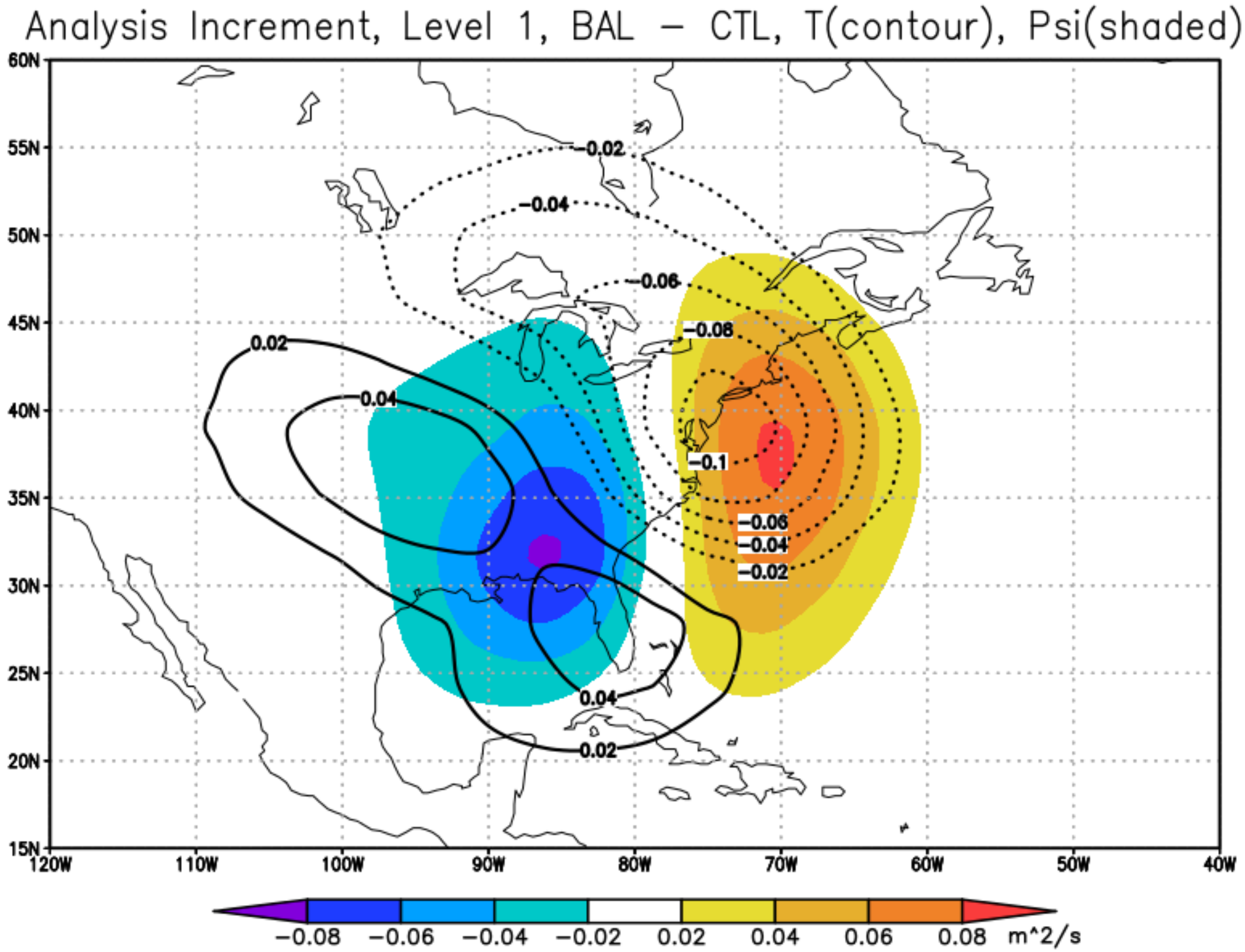


Slight adjustment to both streamfunction and temperature.

Negative regression coefficient at this level.

Single Observation (*T*): 4DEnVar. BAL- CTRL

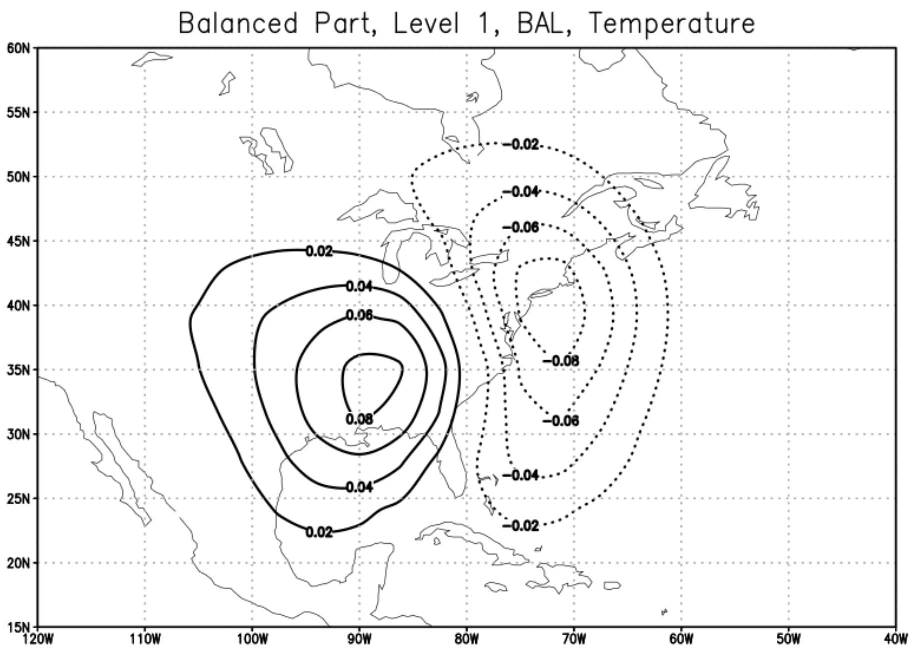
- Difference in both T and ψ



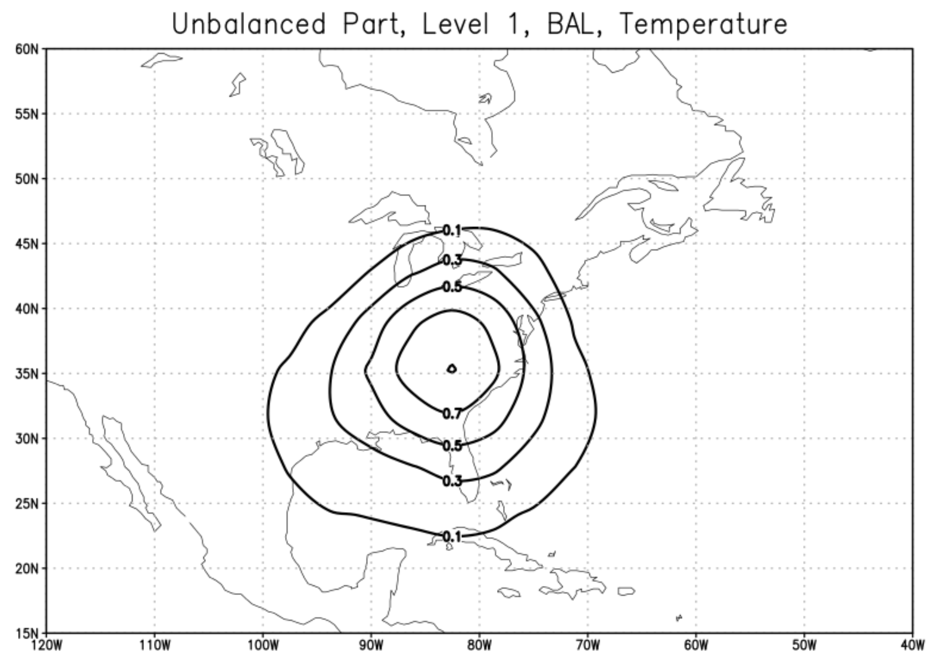
Single Observation (*T*): 4DEnVar. Balanced vs Unbalanced

T - Contoured

Balanced



Unbalanced



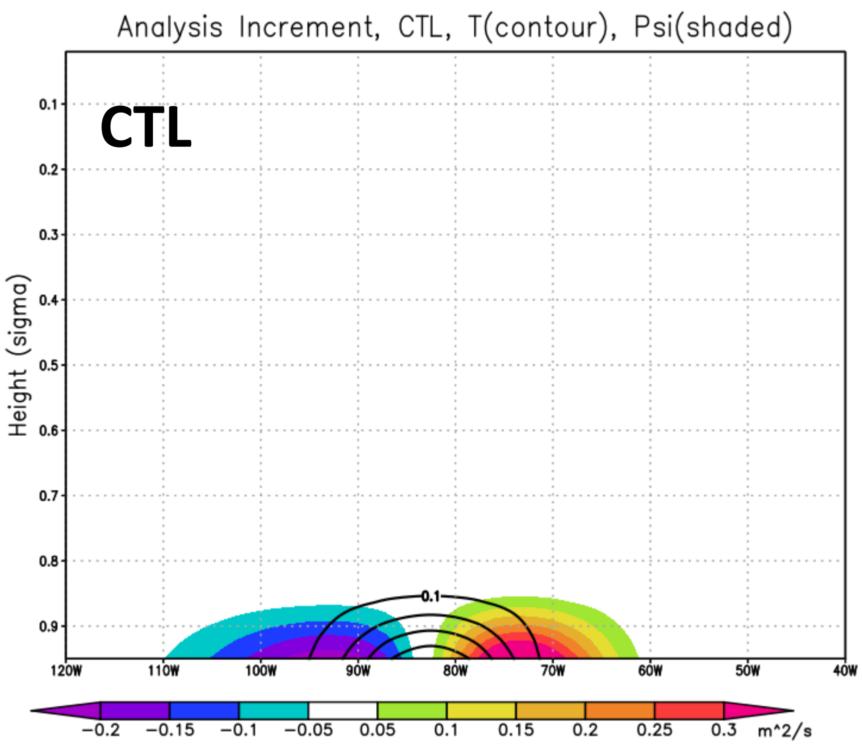
$$\Delta T = \mathbf{G}\,\Delta\psi + \Delta T^u$$

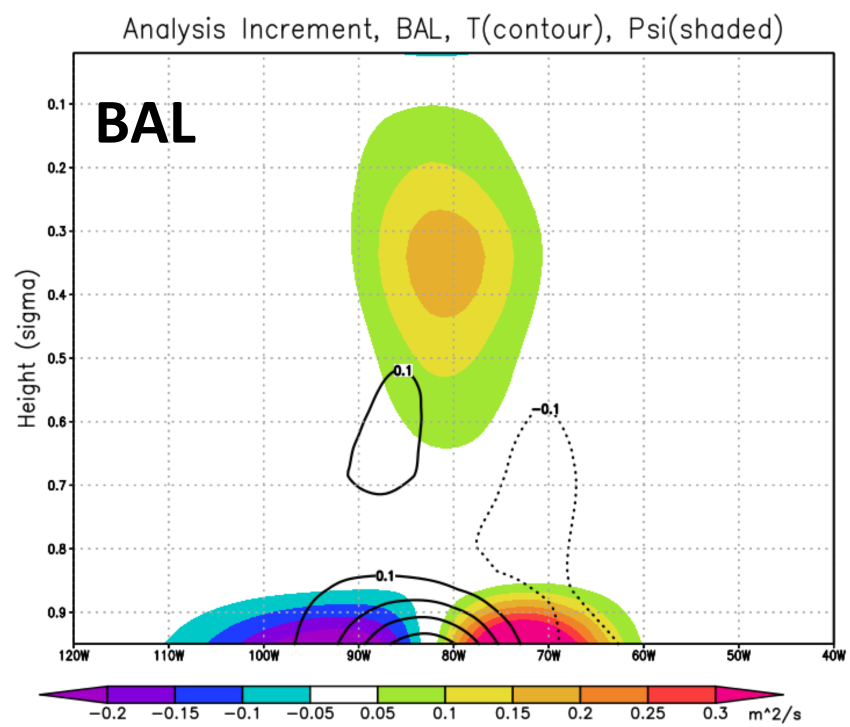
T - Contoured
ψ - Shaded

Single Observation (T): 4DEnVar. Vertical Structure

- Balance operator allows balanced correlations to propagate outside the localization radius.

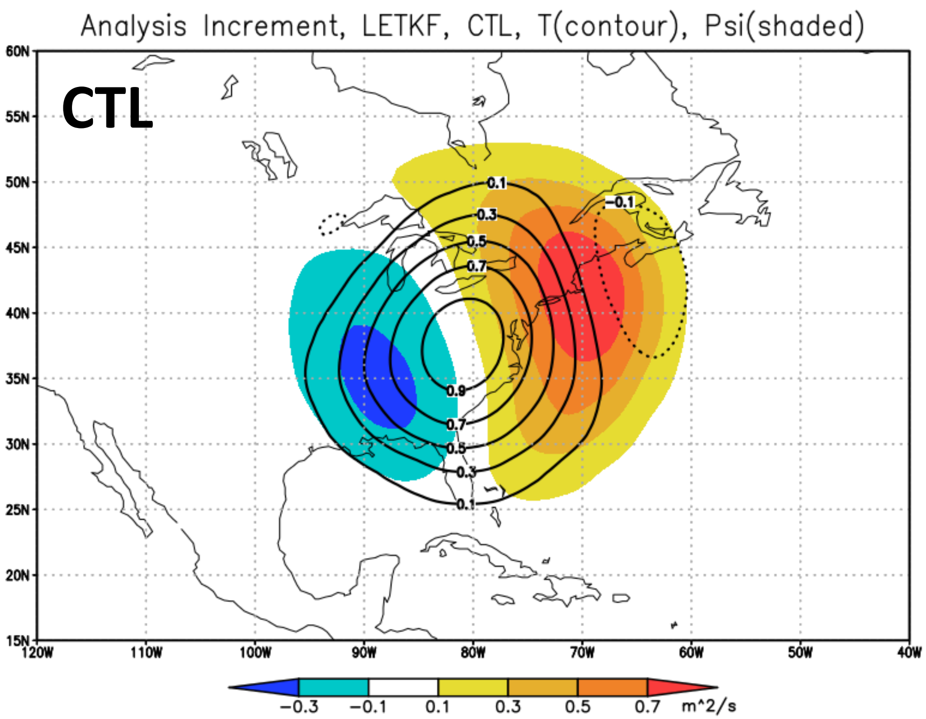
- Two-way propagation of information: $T \leftrightarrows \psi$

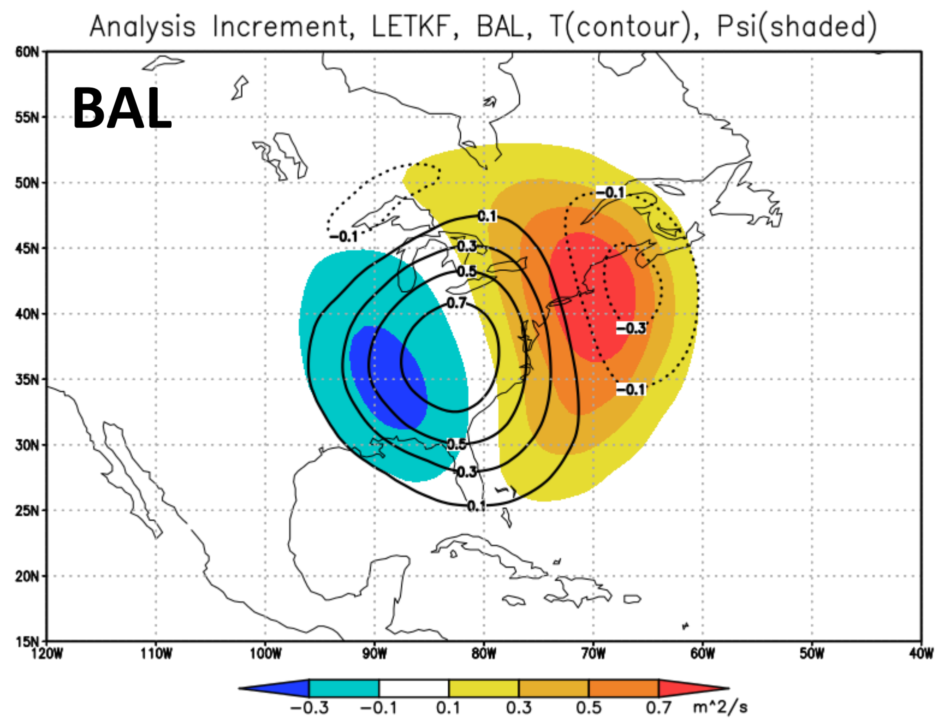




Single Observation (*T*): LETKF. CTL vs BAL

T - Contoured
ψ - Shaded





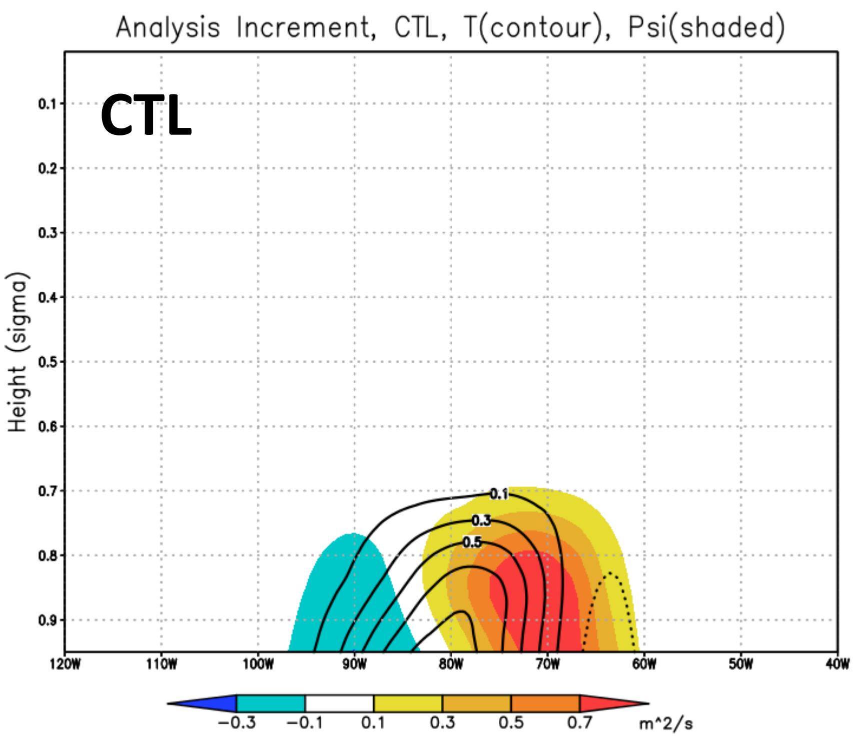
Streamfunction is unaffected.

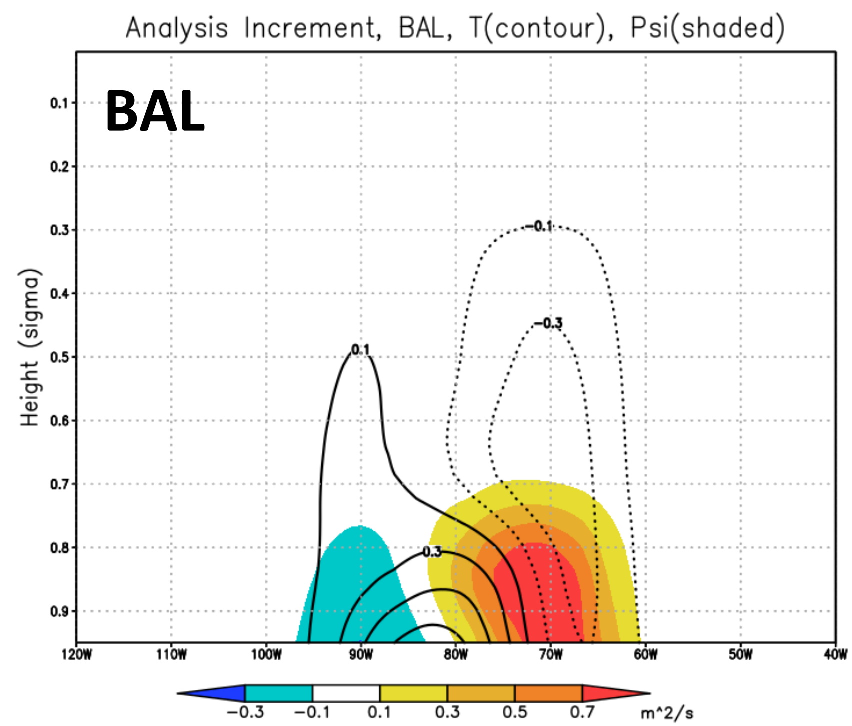
Larger impact on temperature than in the hybrid case.

Single Observation (*T*): LETKF. Vertical Structure

- No propagation of streamfunction information outside of the localization radius.

- Temperature adjusts to the streamfunction only.

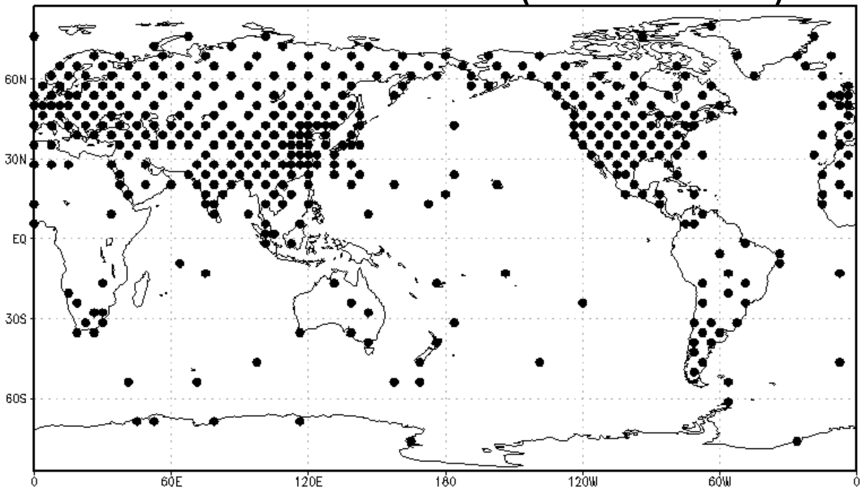




SPEEDY Cycling Experiment: Set-up

- Resolution: T63 Truth with T30 forecasts and analyses
- Beta weighting: 10% Climatological + 90% Ensemble
- Ensemble Size: 20 members
- Inflation: Fixed at 8%
- Experiment length: 2 years (January 1982 – January 1984)
- Observations: simulated radiosonde network and satellite observations

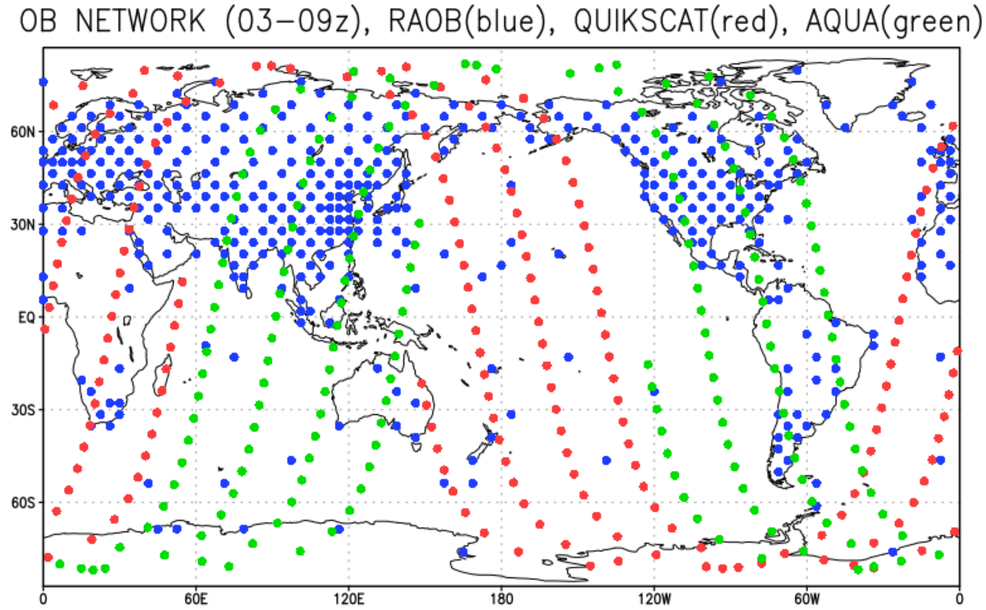
Radiosonde Network (416 Stations)



Observation Type	Observation Error
u	1 m/s
v	1 m/s
T	1 K
P	100 Pa
q	10^{-4} kg/kg

SPEEDY Cycling Experiment: Set-up

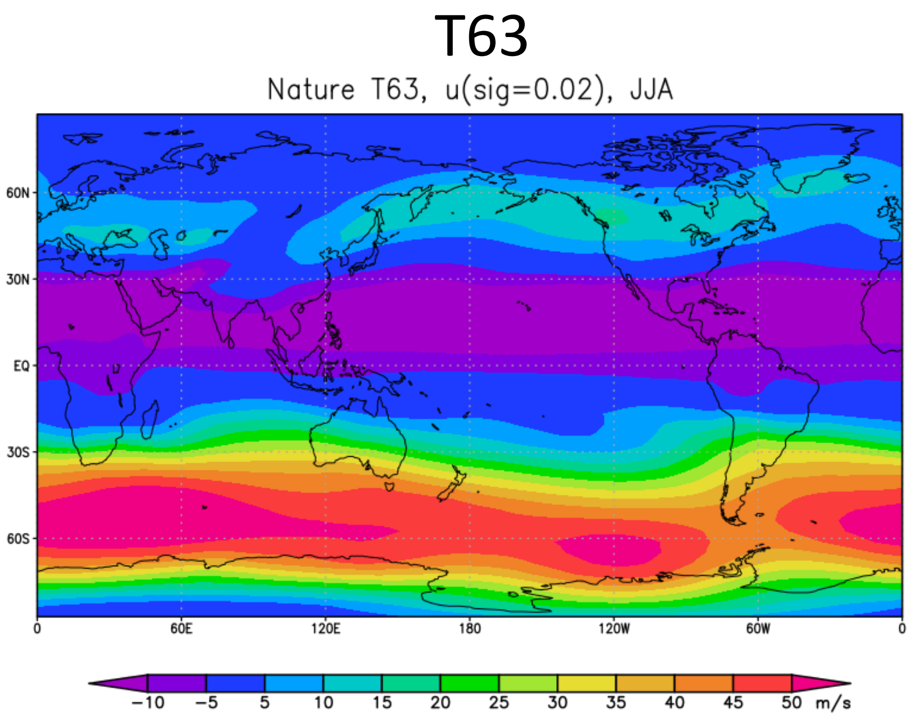
- Satellite
  - AIRS on Aqua and SeaWinds on Quikscat
  - 5 minute intervals with linear time interpolation to an hourly T63 truth

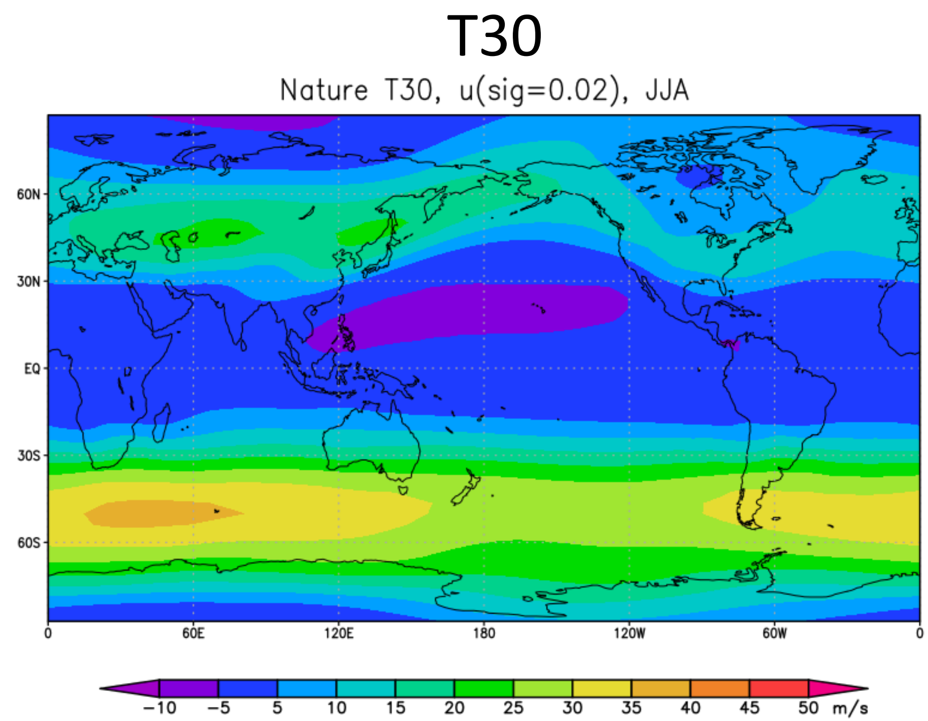


- AIRS:
  - *T* profile: 2 K error
  - *q* profile up to middle model level: $2x10^{-4}$ kg/kg error
- SeaWinds:
  - *u* and *v* at lowest model level: 1.5 m/s error

Model Bias: T63 vs T30

- Top level u, JJA



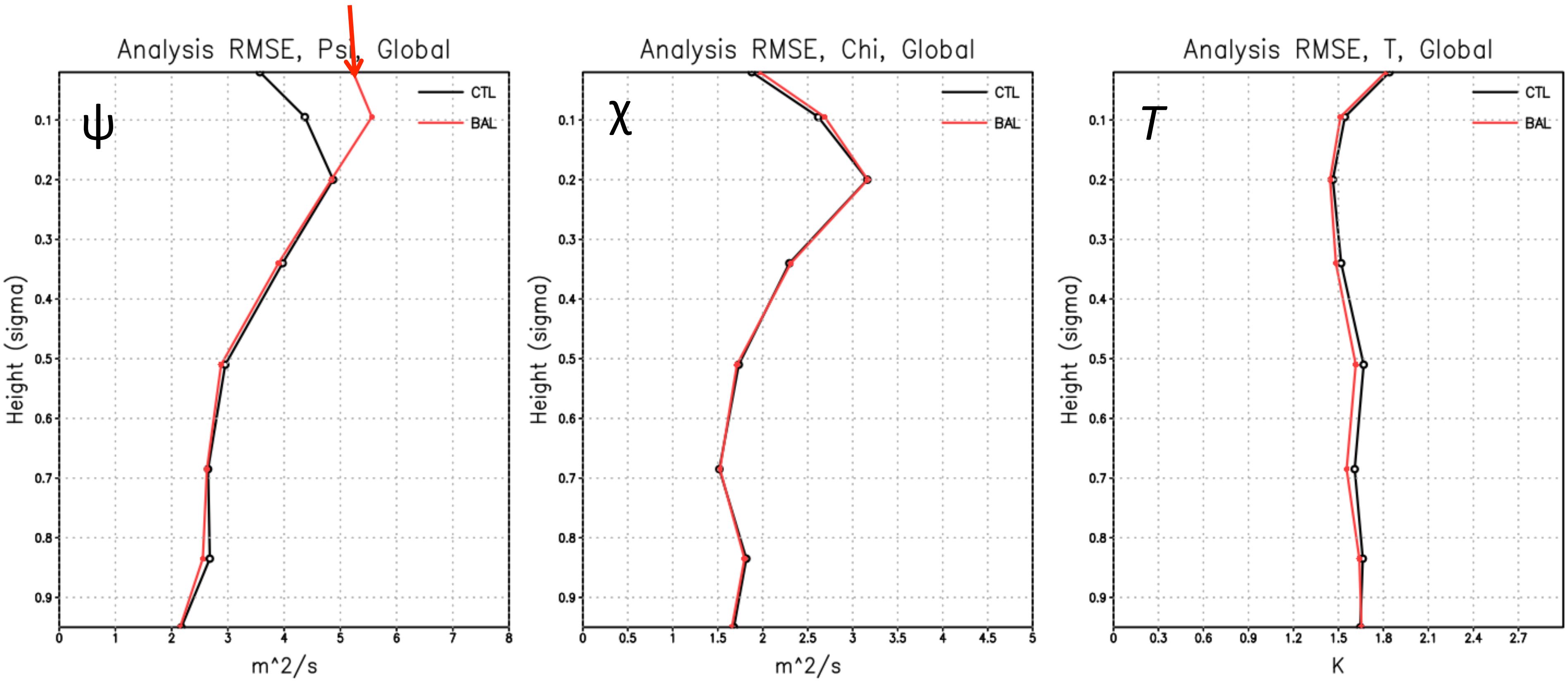


- Stratospheric diffusion coefficients were not adjusted for the higher resolution.
- The stratosphere is more damped in the T30 model than in the T63.
- If severe (regional) bias is untreated, balance operator can degrade the performance.

SPEEDY Cycling Experiment: Hybrid. Analysis RMSE

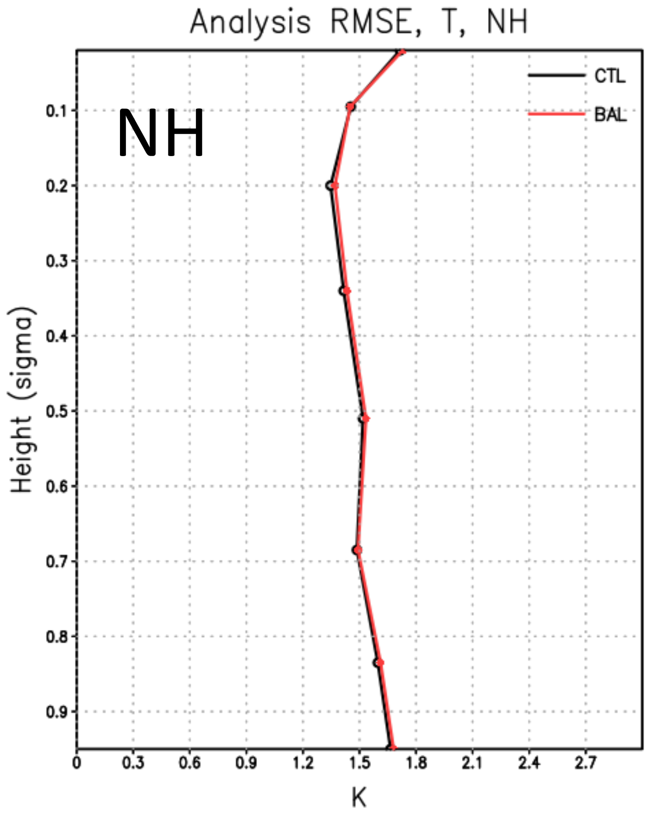
CTL
BAL

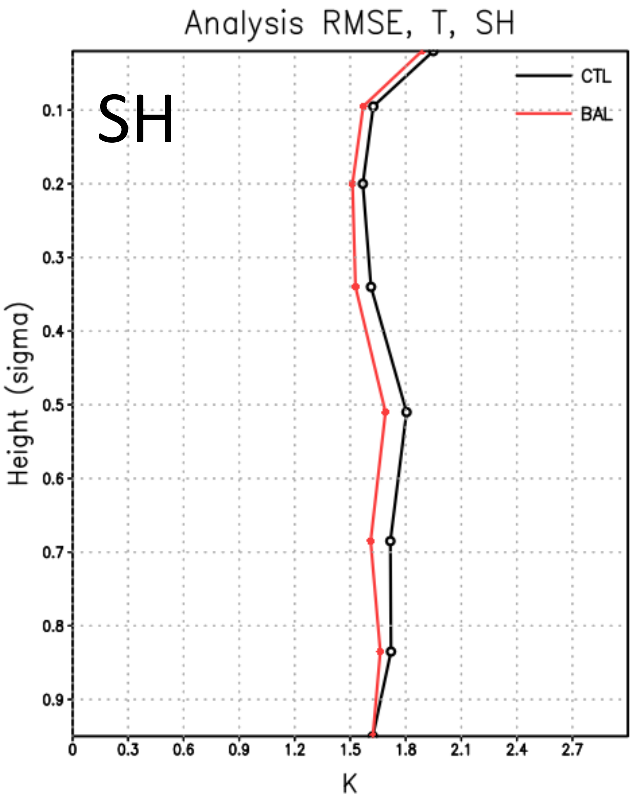
Due to severe model bias



- Greatest positive impact where the balance operator works on the full column: T and χ
- Negative impact where the model bias is prevalent: stratospheric wind fields

SPEEDY Cycling Experiment: Hybrid. Analysis RMSE





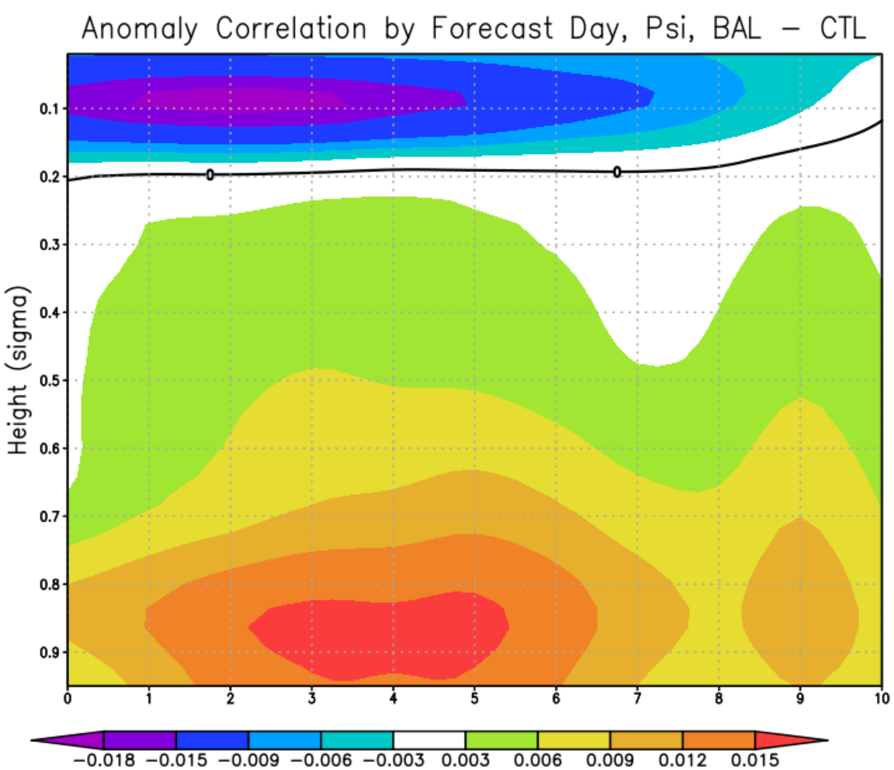
Balance operator has a greater impact in the southern hemisphere

Lower percentage of radiosonde data

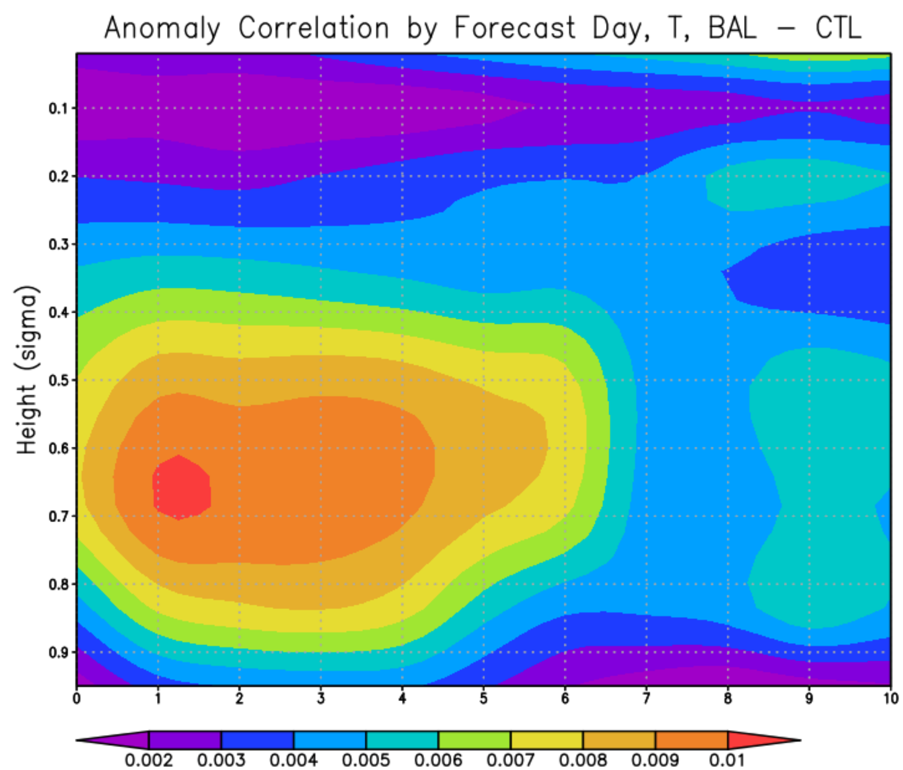
SPEEDY Cycling Experiment: Hybrid. Anomaly Correlation

BAL – CTL
> 0 BAL Improves
< 0 BAL Degrades

ψ



T



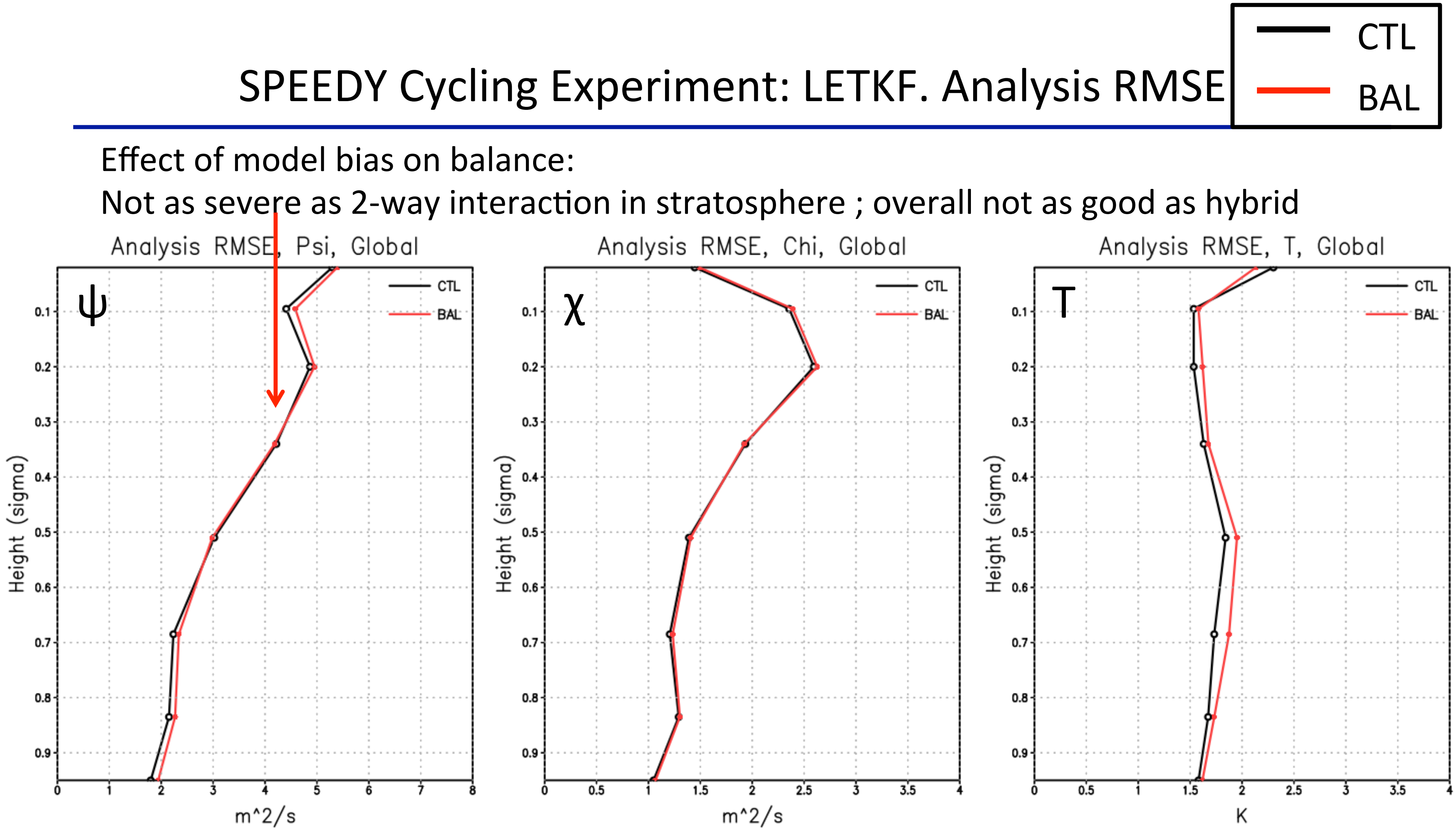
→ Forecast Day

Forecast skill for temperature and tropospheric streamfunction are improved for all forecast lengths.

SPEEDY Cycling Experiment: LETKF. Analysis RMSE

Effect of model bias on balance:

Not as severe as 2-way interaction in stratosphere ; overall not as good as hybrid



The balance operator moves the temperature away from the observations to be brought into balance with the streamfunction.

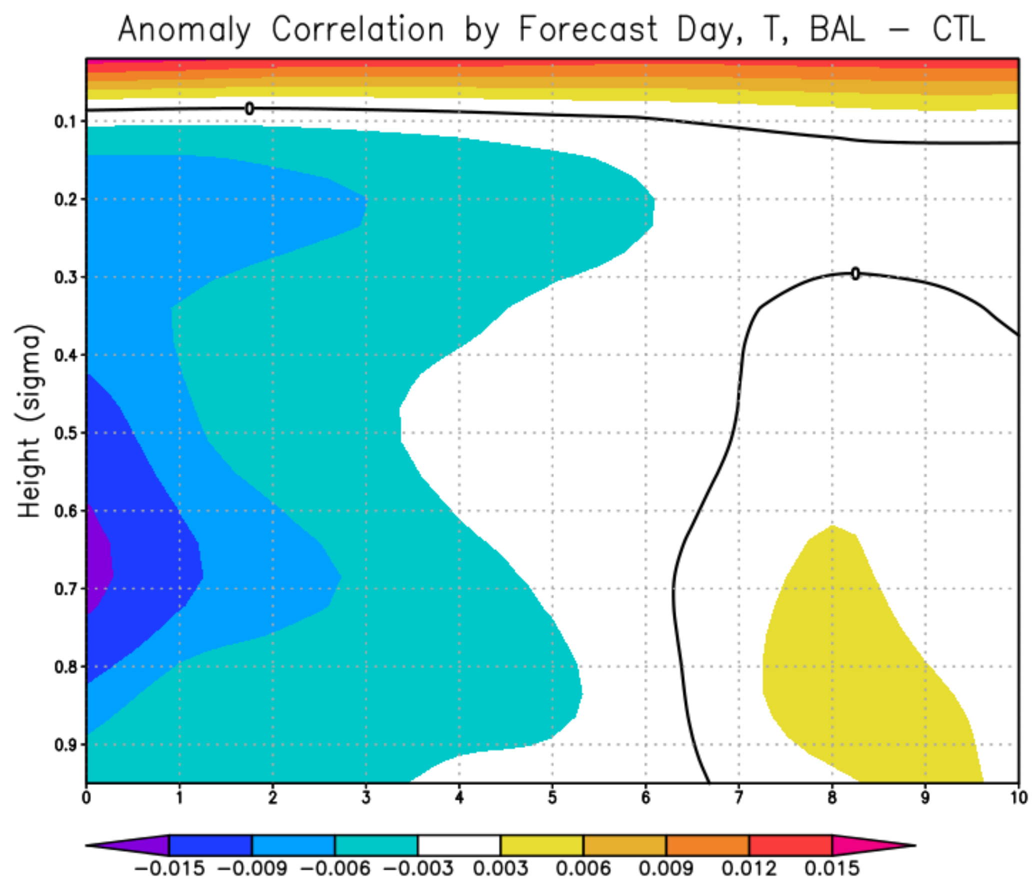
Negative impact on streamfunction is through integration only since the balance operator does not impact it.

SPEEDY Cycling Experiment: LETKF. Anomaly Correlation

BAL – CTL
> 0 BAL Improves
< 0 BAL Degrades

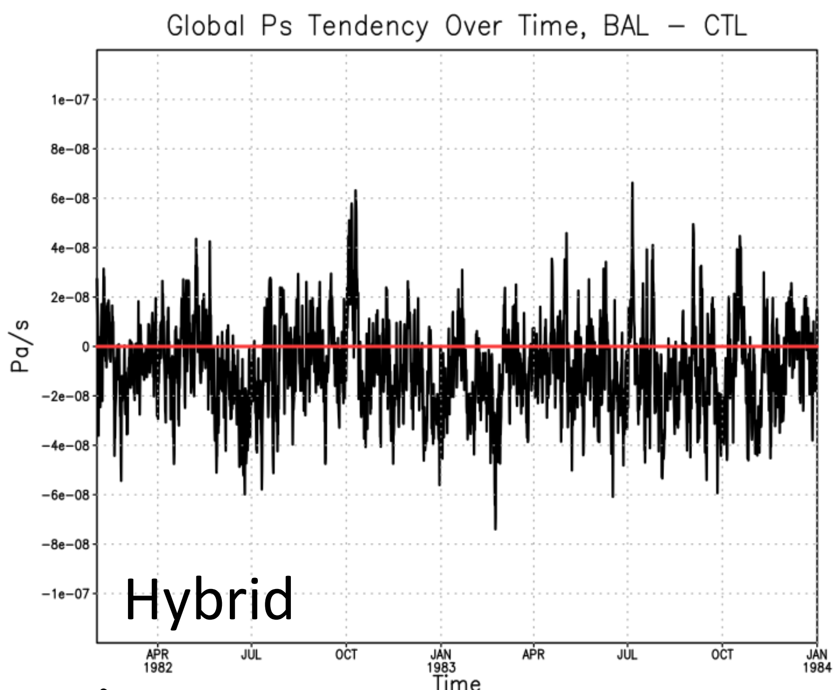
- Stratospheric improvement is dominated by the southern polar region.
- The forecast skill is degraded for short forecast lead times.
  - Temperature adjusts to streamfunction.
- At longer lead times, the forecast skill is improved.
- *Is the improvement due to the balance operator?*

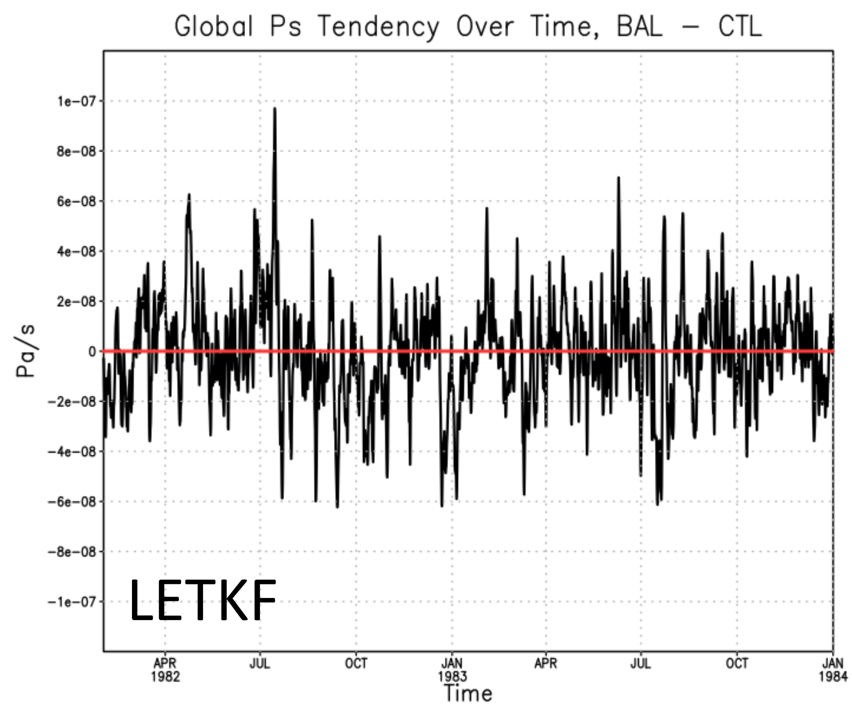
T



Measure of Balance

BAL – CTL
> 0 BAL Increases Imbalance
< 0 BAL Reduces Imbalance





Time

Global surface pressure tendency is significantly reduced in the Hybrid case.

Summary

- Study of balance operator to two ensemble data assimilation schemes:
  - Hybrid 4DEnVar and LETKF
  - Localization in model space and observation space

- Hybrid 4DEnVar provides additional balance by propagating information outside of the traditional localization radius and preserves the balanced information provided by the ensembles.

- The type of localization in ensemble data assimilation methods impacts the effectiveness of applying the balance operator to the ensemble, with the Hybrid 4DEnVar showing greater improvements than the LETKF.

Reference

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.,* **138**, 1550–1566, doi:10.1175/2009MWR3157.1.

——, ——, ——, ——, and ——, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.,* **138**, 1567–1586, doi:10.1175/2009MWR3158.1.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.,* **139**, 1445–1461, doi:10.1002/qj.2054.

Greybush, S. J., E. Kalnay, T. Miyoshi, and K. Ide, 2011: Balance and Ensemble Kalman Filter Localization Techniques. *Mon. Wea. Rev.,* **139**, 511-522.

Gong, Jiandong, 2015: Use ensemble error covariance information in GRAPES 3DVAR -- initial results. College Park, MD, June 17, 2015.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.,* **129**, 2776–2790.

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

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman Filter. *Physica D*, **230**, 112–126.

Miyoshi, T., 2005: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. dissertation, University of Maryland, 226 pp.

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parameterizations. I: model climatology and variability in multi-decadal experiments. *Clim. Dyn.,* **20**, 175-191.

Wu, W.-S., D. F. Parrish, and R. J. Purser, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.,* **130**, 2905-2916.

Thank You