Balance Operators and Localization in Ensemble Data Assimilation

Kayo Ide, UMD
Catherine Thomas, UMD & NOAA EMC

PSU-UMD DA Workshop
June 2017

Motivation

- Observations are not perfect fit to model or real atmosphere, due to
  - Instrument error
  - Representativeness
  - Observation operator

↓

Imbalance in analysis

= imbalance in initial conditions (dynamical inconsistency)

↓

- Production of fast moving gravity waves which degrades the forecast.

Balances and the relationship between variables need to be properly represented in order to produce a skillful forecast.

Incorporating Balance

- Var approach
  - Strong constraints
  - Weak constrains
    - Additional penalty in the cost function (Courtier and Talagrand, 1990; Gauthier and Thepaut, 2001; Wee and Kuo, 2004)
    - Application of balance operator **Γ** and apply variable localization (Lorenc et al., 2003; Parrish and Derber, 1992; Wu et al., 2002)

- Ensemble approach
  - Underlying assumption: dynamical estimation of the correlation maintains balance.
  - Challenge: Spatial localization to suppress spurious sampling errors
    - using the same length scale for all variables
    - balanced and umblanced variables may have different correlation length scale.

Objectives

- Formulate ensemble data assimilation with balance operator using variable localization as in 3D/4D-Var

- Demonstrate the interference of balance operator with
  - Model-space spatial localization
  - Observation-space spatial localization

- Evaluate the impact through the OSSEs using SPEEDY model

Practical Data Assimilation

- Representation of uncertainty by error covariance matrix
  - Background given by forecast): $\mathbf{x}^b$ with $\mathbf{B}$
  - Observation with known $\mathbf{y}=\mathbf{h}(\mathbf{x})$: $\mathbf{y}^o$ with $\mathbf{R}$

- Optimization (in incremental form: $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$ using $\mathbf{d}=\mathbf{y}^o - \mathbf{h}(\mathbf{x}^b)$)

$$J(\Delta\mathbf{x}) = \underbrace{\frac{1}{2}\Delta\mathbf{x}^T\mathbf{B}^{-1}\Delta\mathbf{x}}_{\text{Background (Forecast)}} + \underbrace{\frac{1}{2}(\mathbf{d}-\mathbf{H}\Delta\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{d}-\mathbf{H}\Delta\mathbf{x})}_{\text{Observations}}$$

$\mathbf{H}$: Linearized obs operator $\mathbf{h}(\mathbf{x})$

  - Variational methods (3DVar, 4DVar) = numerically solve for $\Delta\mathbf{x}$
  - Sequential methods (OI, EnKF) = use analytical form of $\Delta\mathbf{x}$

- Analysis
  - State: $\Delta\mathbf{x}^a = \mathbf{K}\mathbf{d}$
  - Error covariance: $\mathbf{A}=(\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{A}$ with $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-1}$

    or $\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$ with $\mathbf{A} = (\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$

Balance Operator

- Provides multivariate correlation in **B** [control variable localization] based on physical relationship

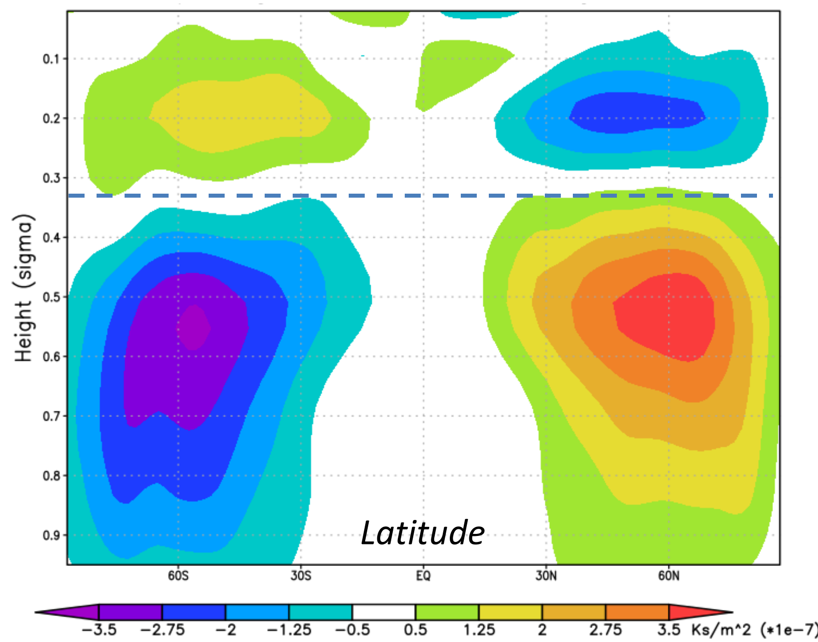
- Conventionally present in Var schemes through the empirical regression coefficients

Following Wu et al (2002)

$$\Delta \mathbf{x} = \Delta \mathbf{x}^{u} + \Delta \mathbf{x}^{balance}$$

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

$\mathbf{G}$ for T at sigma=0.34



Balance Operator

- Control variable transformation

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

$$\Delta\mathbf{x} = \begin{pmatrix} \Delta\mathbf{x}_{\psi} \\ \Delta\mathbf{x}_{\chi} \\ \Delta\mathbf{x}_{T} \\ \Delta\mathbf{x}_{p} \\ \Delta\mathbf{x}_{q} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \Delta\mathbf{x}_{\chi}^{u} \\ \Delta\mathbf{x}_{T}^{u} \\ \Delta\mathbf{x}_{p}^{u} \\ \Delta\mathbf{x}_{q} \end{pmatrix} + \begin{pmatrix} \Delta\mathbf{x}_{\psi} \\ c\Delta\mathbf{x}_{\psi} \\ \mathbf{G}\Delta\mathbf{x}_{\psi} \\ \Omega\Delta\mathbf{x}_{\psi} \\ \mathbf{0} \end{pmatrix} \qquad \Gamma = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ c & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{G} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \Omega & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \qquad \mathbf{z} = \begin{pmatrix} \Delta\mathbf{x}_{\psi} \\ \Delta\mathbf{x}_{\chi}^{u} \\ \Delta\mathbf{x}_{T}^{u} \\ \Delta\mathbf{x}_{p}^{u} \\ \Delta\mathbf{x}_{q} \end{pmatrix}$$

- Role of **Γ**
  - **Γ**: $\Delta\mathbf{x}_{\psi} \rightarrow \Delta\mathbf{x}_{\chi},\ \Delta\mathbf{x}_{T},\ \Delta\mathbf{x}_{p}$
  - **Γ**^{T}: $\Delta\mathbf{x}_{\chi},\ \Delta\mathbf{x}_{T},\ \Delta\mathbf{x}_{p} \rightarrow \Delta\mathbf{x}_{\psi}$

Conventional Hybrid

- Hybrid $\mathbf{B} = \gamma^{clim}\,\mathbf{B}^{clim} + \gamma^{en}\,\mathbf{B}^{en}$

  - Using the controls $\mathbf{z}^{clim}$ and $\mathbf{z}^{en}$ (Wang etal 2009)

$$J(\mathbf{z}^{clim},\mathbf{z}^{en}) = \frac{1}{2}(\mathbf{z}^{clim})^T\mathbf{z}^{clim} + \frac{1}{2}(\mathbf{z}^{en})^T\mathbf{z}^{en} + \frac{1}{2}\sum_k(\mathbf{d}_k - \mathbf{H}_k\Delta\mathbf{x}_k)^T(\mathbf{R}_k)^{-1}(\mathbf{d}_k - \mathbf{H}_k\Delta\mathbf{x}_k)$$

(Kleist and Ide 2015)

$$\Delta\mathbf{x}_k = \underbrace{\beta^{clim}\Gamma\mathbf{U}^{clim}\mathbf{z}^{clim}}_{\text{climatology}} + \underbrace{\beta^{en}\sum_{m=1}^{M}\mathbf{F}^{en}\mathbf{z}^{en}_{(m)}\circ(\mathbf{X}^{en}_{(m)})_k}_{\text{ensemble}}$$

climatology: $\mathbf{B}^{clim} = \mathbf{U}^{clim}\,(\mathbf{U}^{clim})^T$

ensemble: $\mathbf{B}^{en} = \mathbf{X}^{en}\mathbf{U}^{en}\,(\mathbf{U}^{en})^T(\mathbf{X}^{en})^T$

- Balance operator $\mathbf{\Gamma}$: applied to only $\Delta\mathbf{x}^{clim}$ not on $\Delta\mathbf{x}^{en}{}_k$
- Spatial localization: included as recursive filter (Purser and Wu, 2003ab)
  - $\mathbf{U}^{clim}$ for $\Delta\mathbf{x}^{clim}$
  - $\mathbf{U}^{en}$ for $\Delta\mathbf{x}^{en}$ [$\mathbf{F}^{en}$: recursive filter]

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

Ensemble Application: Hybrid EnVar

- Approach

$$\Delta \mathbf{x}_k = \beta^{\mathrm{clim}} \Gamma \mathbf{U}^{\mathrm{clim}} \mathbf{z}^{\mathrm{clim}} + \beta^{\mathrm{en}} \sum_{m=1}^{M} \mathbf{F}^{\mathrm{en}} \mathbf{z}^{\mathrm{en}}_{(m)} \circ (\mathbf{X}^{\mathrm{en}}_{(m)})_k$$

$$\Downarrow$$

$$\Delta \mathbf{x}_k = \Gamma \left[\beta^{\mathrm{clim}} \mathbf{U}^{\mathrm{clim}} \mathbf{z}^{\mathrm{clim}} + \beta^{\mathrm{en}} \sum_{m=1}^{M} \hat{\mathbf{F}}^{\mathrm{en}} \mathbf{z}^{\mathrm{en}}_{(m)} \circ (\mathbf{Z}^{\mathrm{en}}_{(m)})_k \right]$$

$$\mathbf{Z}^{\mathrm{en}}_{(m)} = \begin{Bmatrix} \mathbf{x}_{\psi(m)} - \bar{\mathbf{x}}_{\psi(m)} \\ \mathbf{x}_{\chi(m)}{}^{u} - \bar{\mathbf{x}}_{\chi(m)}{}^{u} \\ \mathbf{x}_{T(m)}{}^{u} - \bar{\mathbf{x}}_{T(m)}{}^{u} \\ \mathbf{x}_{p(m)}{}^{u} - \bar{\mathbf{x}}_{p(m)}{}^{u} \\ \mathbf{x}_{q(m)}{}^{u} - \mathbf{x}_{q(m)}{}^{u} \end{Bmatrix}$$

- Balance operator **Γ** is applied to both $\Delta \mathbf{x}^{\mathrm{clim}}$ to $\Delta \mathbf{x}^{\mathrm{en}}{}_k$
- Spatial localization $\hat{\mathbf{F}}^{\mathrm{en}}$ is applied onto unbalanced part in ensemble

$$\Gamma \hat{\mathbf{F}}^{\mathrm{en}} \mathbf{z}^{\mathrm{en}}_{(m)} \circ (\mathbf{Z}^{\mathrm{en}}_{(m)})_k \neq \mathbf{F}^{\mathrm{en}} \Gamma \mathbf{z}^{\mathrm{en}}_{(m)} \circ (\mathbf{Z}^{\mathrm{en}}_{(m)})_k$$

Ensemble Application: Spatial Localization

- Hybrid EnVar: **B** localization

  Global increment

  $$\Delta\mathbf{x}^{en}{}_k = \Gamma\beta^{en}\sum_{m=1}^{M}\hat{\mathbf{F}}^{en}\mathbf{z}^{en}_{(m)} \circ (\mathbf{Z}^{en}_{(m)})_k$$

  - $\mathbf{Z}^{en}{}_{(m)}$ contain $\mathbf{\Gamma}^{T}$
  - $\hat{\mathbf{F}}^{en}$ propagres increment info spatially regardless of the location of $\mathbf{y}^{o}{}_{k}$

  → Information propagation (2-way)

  – $\mathbf{\Gamma}$: $\Delta\mathbf{x}_\psi$ → $\Delta\mathbf{x}_\chi$, $\Delta\mathbf{x}_T$, $\Delta\mathbf{x}_p$

  – $\mathbf{\Gamma}^{T}$: $\Delta\mathbf{x}_\chi$, $\Delta\mathbf{x}_T$, $\Delta\mathbf{x}_p$ → $\Delta\mathbf{x}_\psi$

- LETKF: **R** localization

  Local increment in **z**

  $$\bar{\mathbf{z}}^{a}_{k} = \sum_{m=1}^{M}\mathbf{w}^{en}_{(m)} \circ (\mathbf{Z}^{en}_{(m)})_k$$

  - $\mathbf{w}^{en}{}_{(m)}$ contains $(\rho\circ\mathbf{R})^{-1}$ after $\mathbf{\Gamma}^{T}$
  - $(\rho\circ\mathbf{R})^{-1}=0$ away from $\mathbf{y}^{o}{}_{k}$ prohibits obs info to propagate regardless of variable types

  → Information propagation (1way)

  – $\mathbf{\Gamma}$: $\Delta\mathbf{x}_\psi$ → $\Delta\mathbf{x}_\chi$, $\Delta\mathbf{x}_T$, $\Delta\mathbf{x}_p$

Summary on Formulation

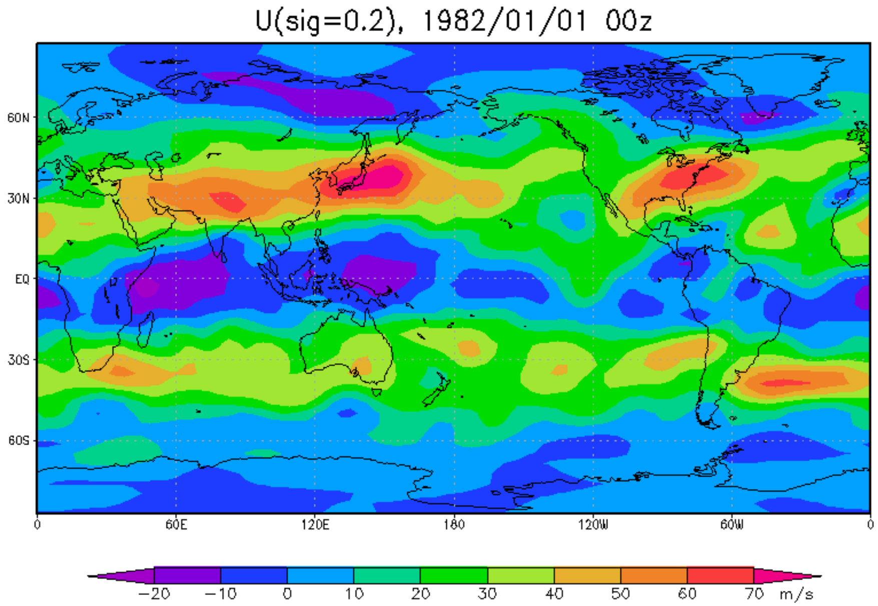
- Balance operators represent the physical relationships between variables.

- Balance operators in ensemble data assimilation allow the localization to be performed on the unbalanced covariances, preserving the balanced covariances.

- The type of localization matters:
  - Model space localization (EnVar) allows a two-way propagation of information.
  - Observation space localization (LETKF) only allows a one-way propagation of information

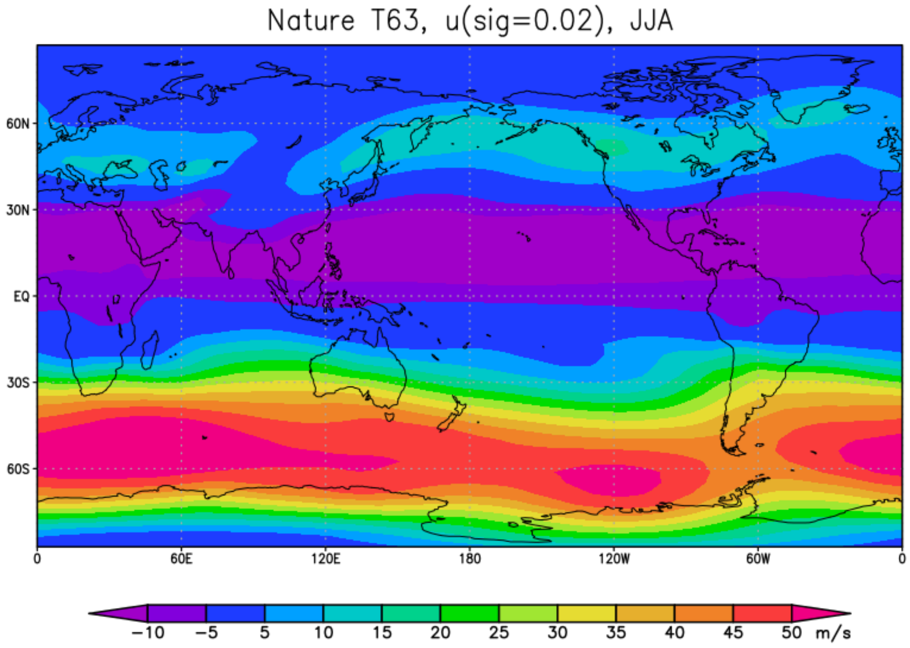
Model Description

- **S**implified **P**arameterizations, primitiv**E**-**E**quation **Dy**namics (SPEEDY) (Moltani, 2003)
  - Global atmospheric general circulation model of intermediate complexity
  - Version 41, provided by F. Kucharski
    - 3 horizontal resolution options: T30, T47, T63
    - 8 vertical levels
- SPEEDY DA System: 3DVar, LETKF, hourly put by T. Miyoshi & S. Greybush

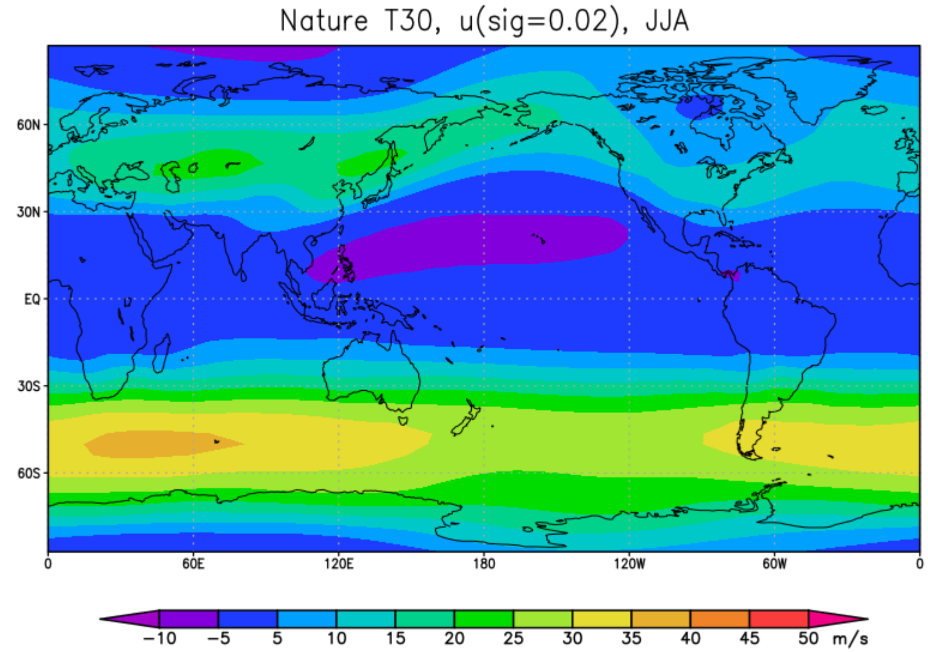


Model Bias in Stratosphere

T63



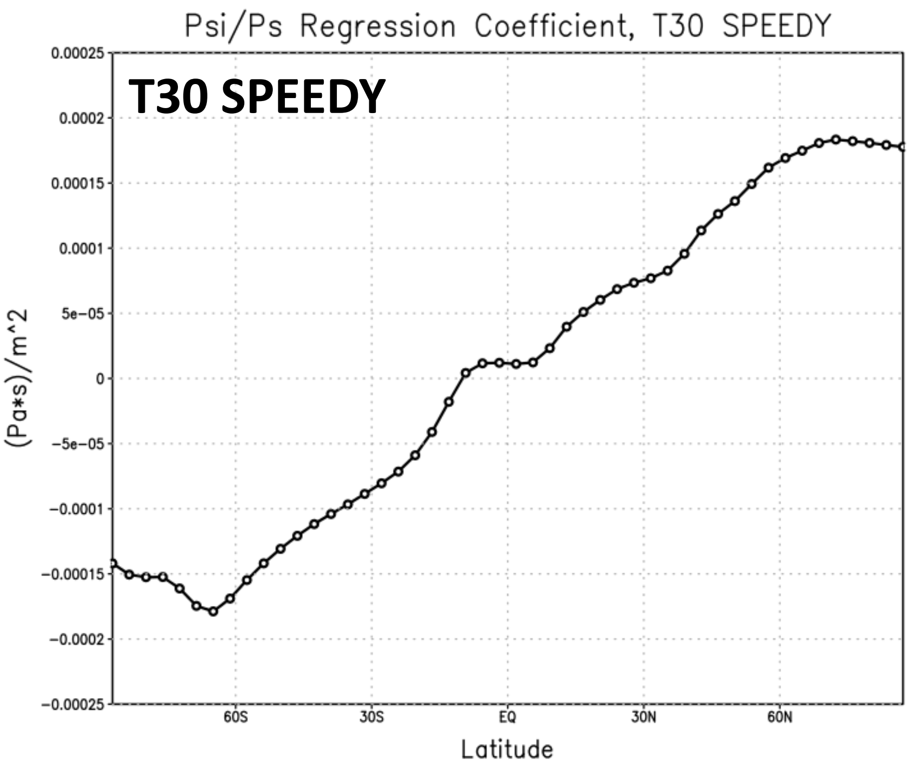
T30

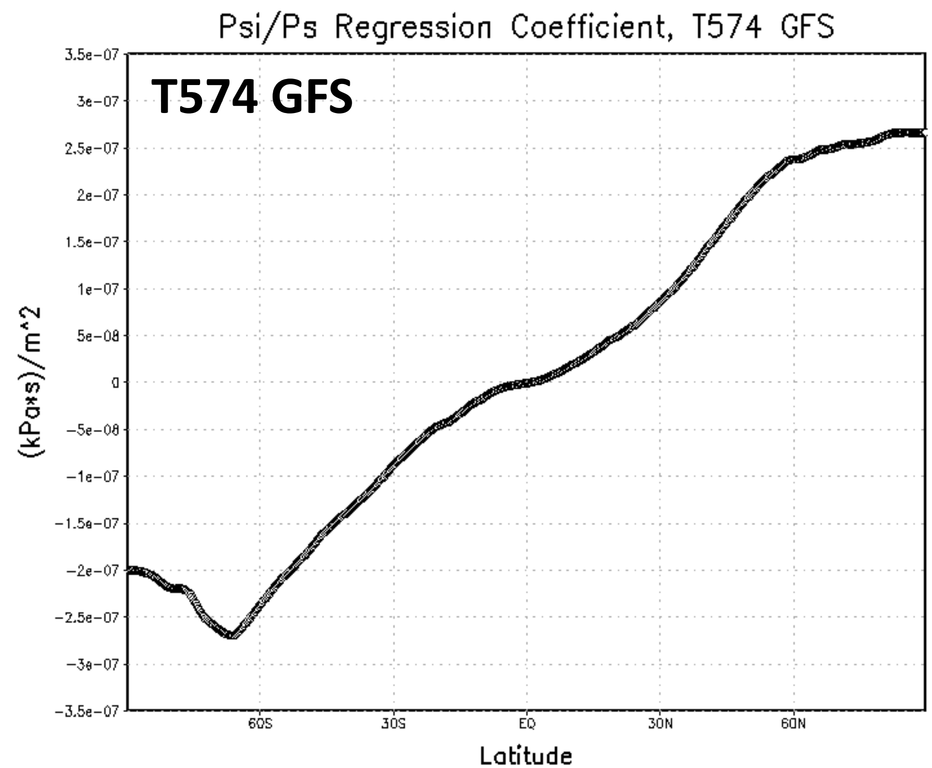


- Stratospheric dynamics highly damped for higher resolution model due to the choice of diffusion coefficient.
- Model biases
  - Stronger in high wavenumber spectral components and cause the model to be unstable.
  - Model bias correction schemes can be helpful (not applied in this work.)

Balance Operator $\mathbf{\Gamma}$: $\Delta\mathbf{x} = \mathbf{\Gamma}\,\mathbf{z}$

- SPEEDY has realistic balance operator $\mathbf{\Gamma}$ in comparison with GFS
  - Example: $\mathbf{\Omega}$ for $\Delta\mathbf{x}_p = \mathbf{\Omega}\Delta\mathbf{x}_\psi + \Delta\mathbf{x}_p^u$





DA Setup

- Configuration
  - CTL: Control without balance operator **Γ** operated on ensemble
  - BAL: Balance Ensemble DA with balance operator **Γ**

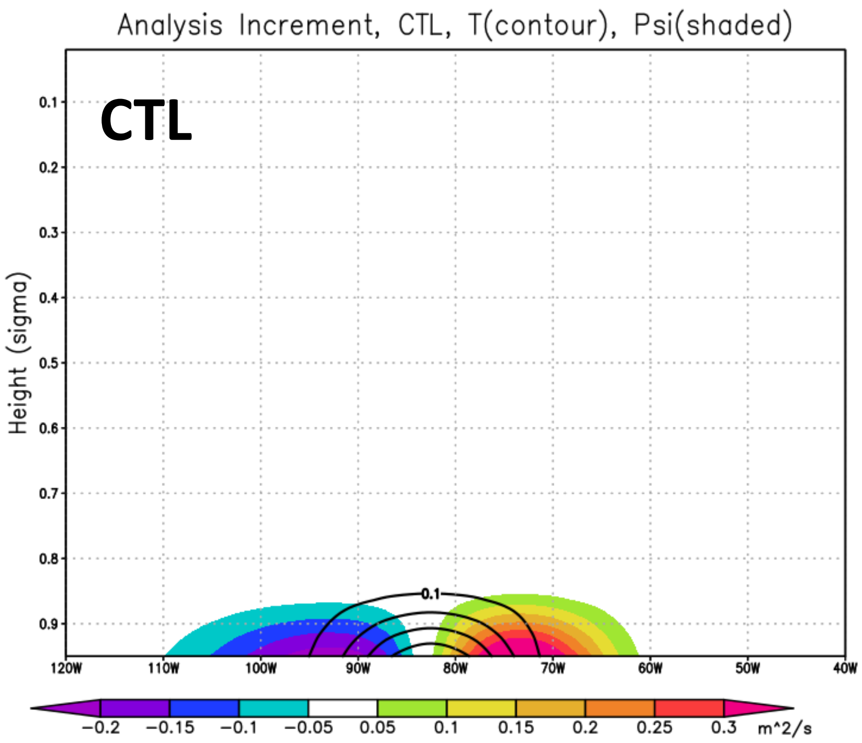
- DA schemes
  - Hybrid 4DEnVar: Model space localization with $\mathbf{\Gamma}^T$
  - 4DLETKF: Obs space localization without $\mathbf{\Gamma}^T$

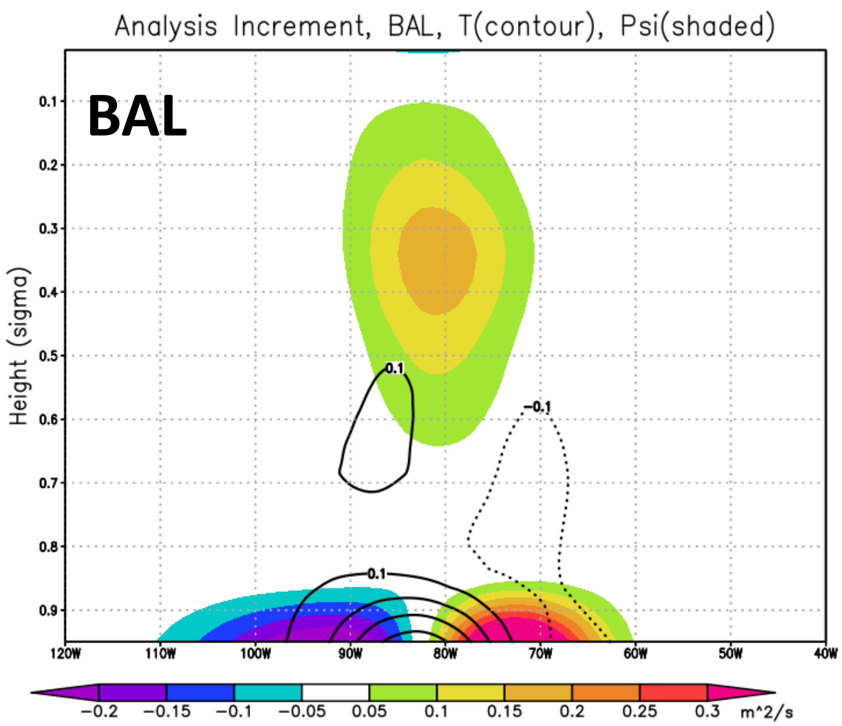
- Experiments
  - Single obs impact tests
    - *T* at the lowest level & at the time of analysis
    - with β^{en} =1 for 4DEnVar direct comparison of space localization schemes
  - Cycling experiments
    - NR: T63
    - DA system: T30

Single *T* Obs Impact Tests

- EnVar (β^{clim},β^{en})=(0,1): Model space localization

T - Contoured
ψ - Shaded



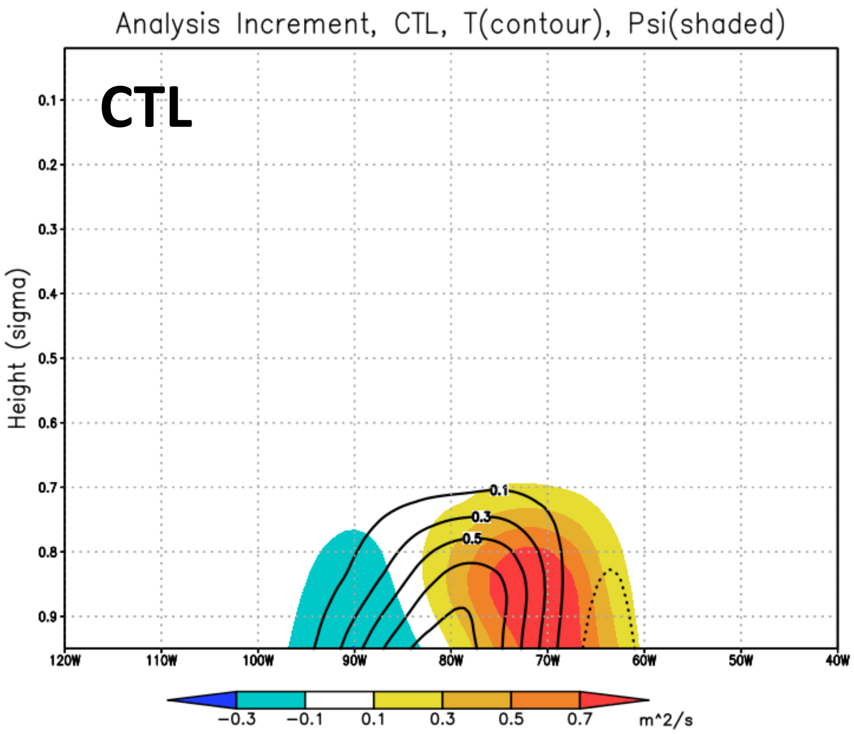


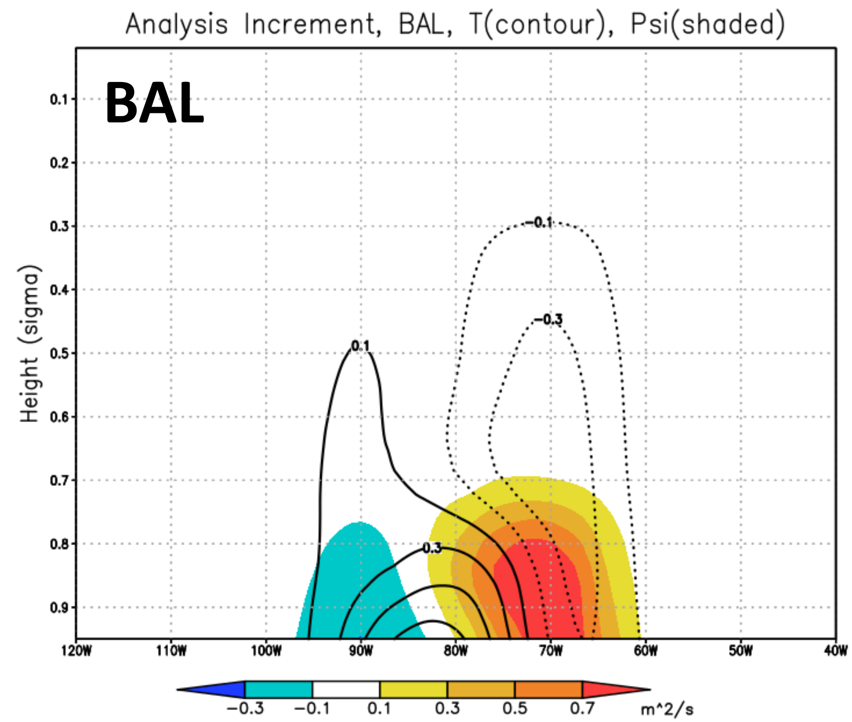
- Two way adjustment
  - *T*→ ψ
  - ψ→ *T*

Single Obs Impact Tests

- LETKF: Obs space localization using the same ensemble

T - Contoured
ψ - Shaded



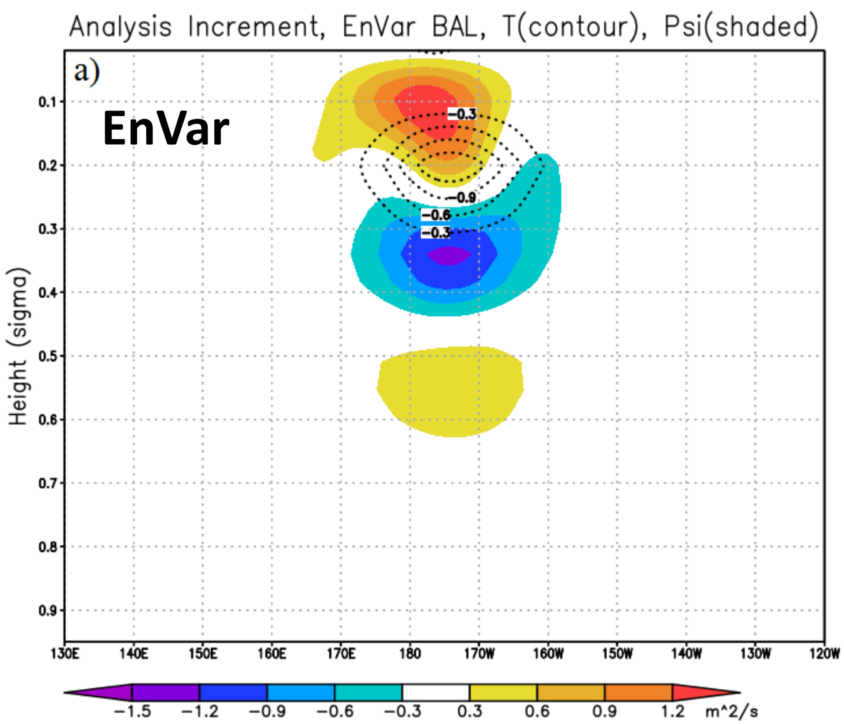


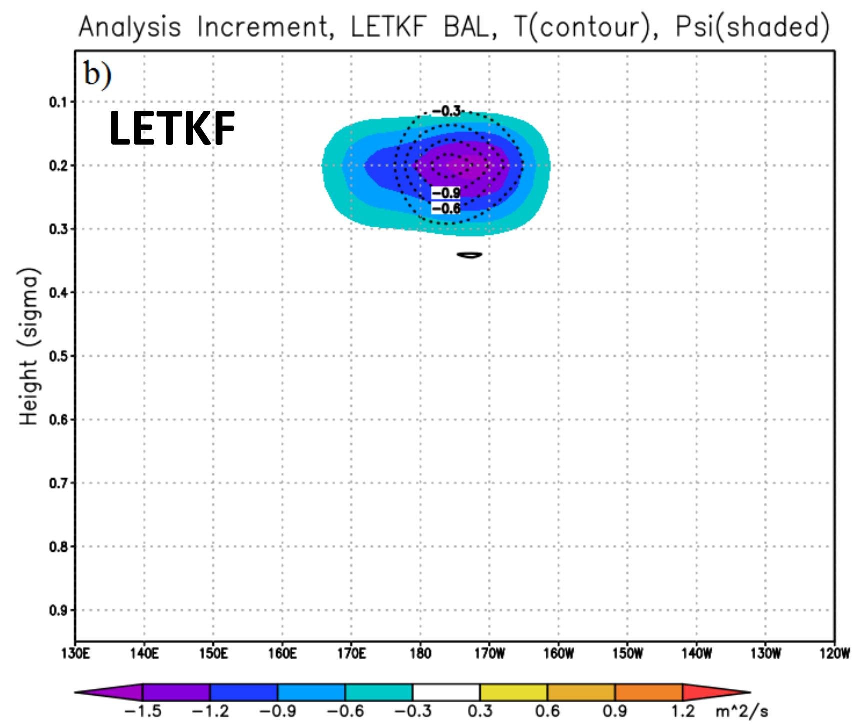
- One way adjustment
  - $T \rightarrow \psi$

Single Observation Impact Tests

- BAL cases for upper air *T* using **Γ**

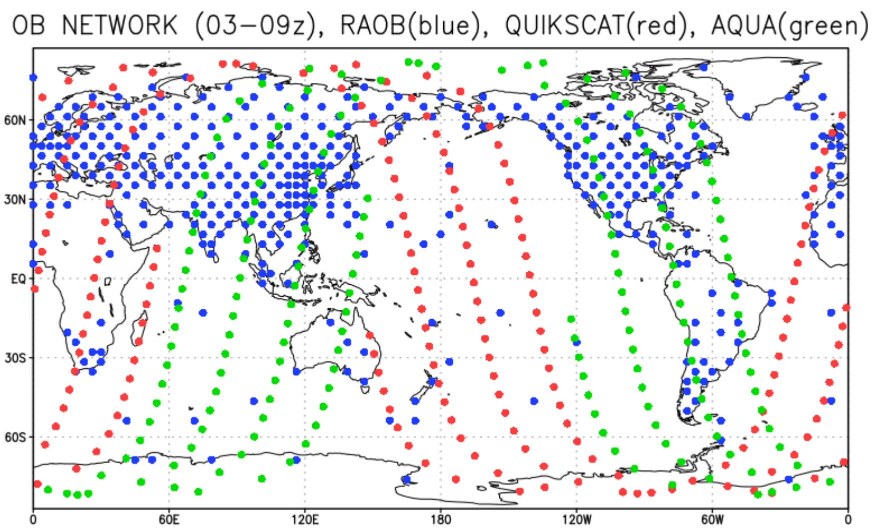
T - Contoured
ψ - Shaded





Cycling Experiments

- Observing system
  - Radiosonde: (*u, v, T*) full profile & *q* bottom four levels
  - SeaWinds: (*u, v*) at lowest level
  - AIRS: *T* full profile & *q* bottom four levels
- DA System
  - Hybrid with (β^{clim},β^{en})=(10%,90%)
  - Ensemble Size: 20 members
  - Inflation: Fixed at 8%
  - Experiment length: 2 years (January 1982 – January 1984)



Observation Type	Observation Error
Radiosonde	
u, v	1 m/s
T	1 K
P	100 Pa
q	10^{-4} kg/kg
Satellite	
u, v	1.5 m/s
T	2 K
q	2×10^{-4} kg/kg

Cycling Experiments

- Metrics for evaluation of the OSSE Experiments

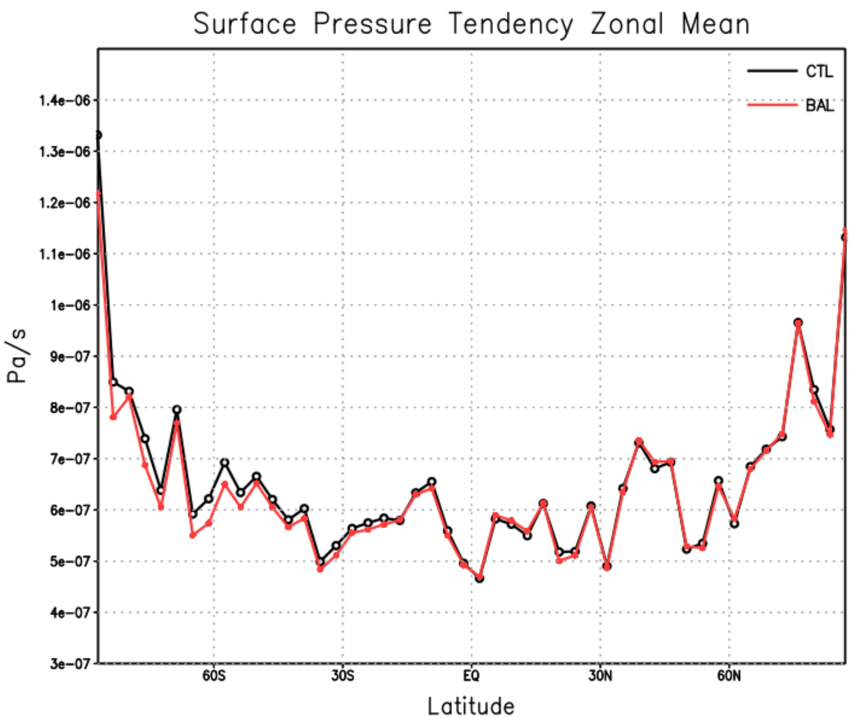
  - Balance (surface pressure tendency)

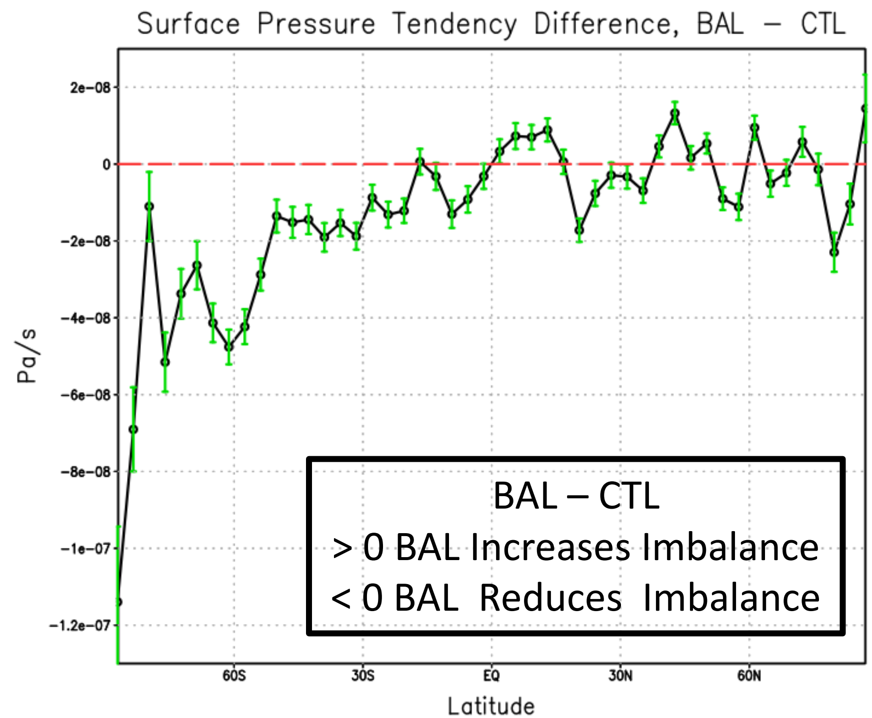
  - Analysis skills (RMSE)

  - Forecast skills (Anomaly Correlations)

Measure of Balance

- Global p_{sfc} tendency

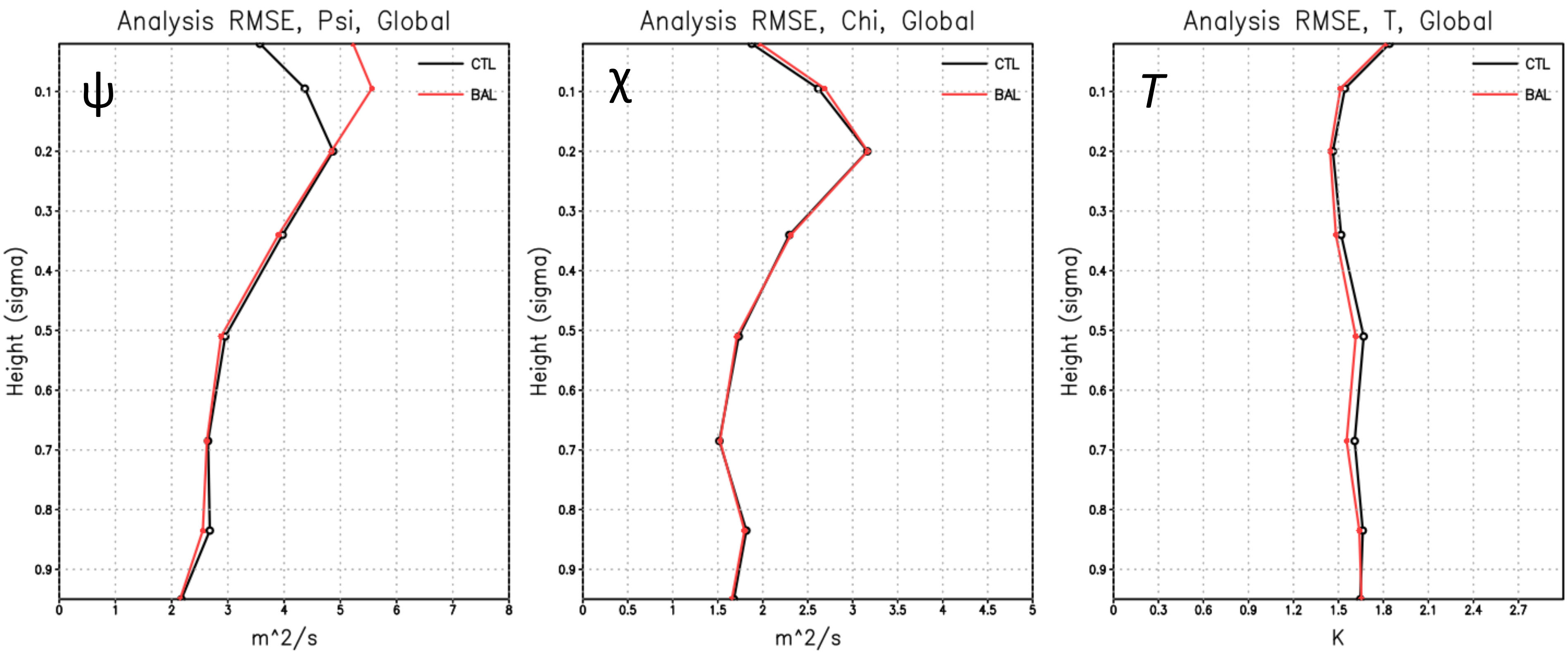




- Significantly reduced in Hybrid case
- Practically unchanged in LETKF case (not shown)

Analysis Skills

- Hybrid

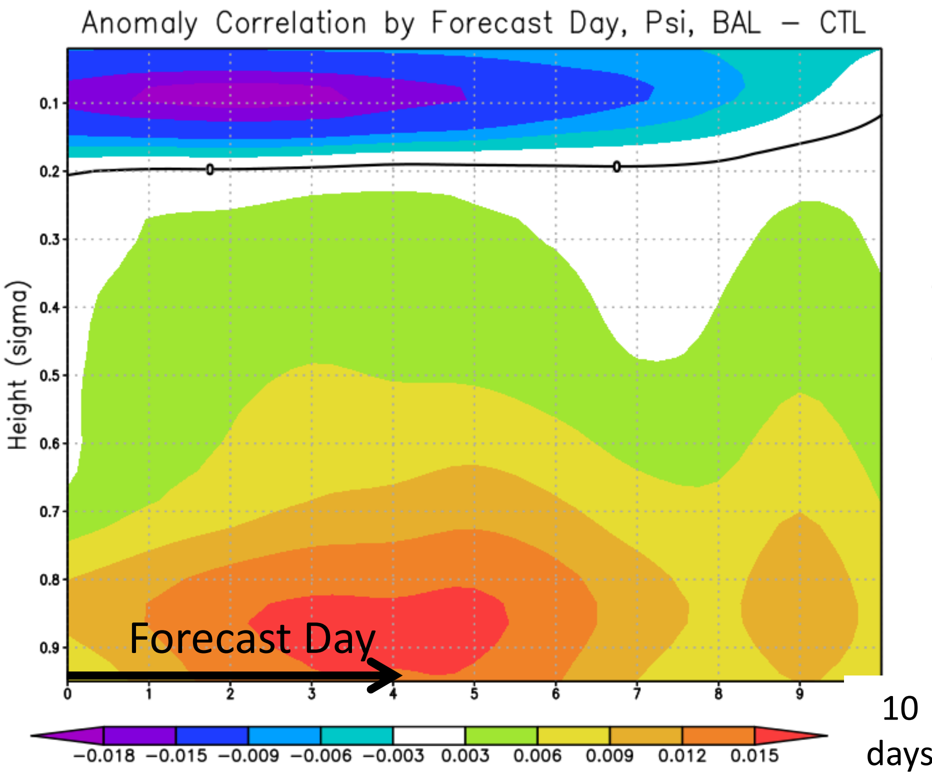


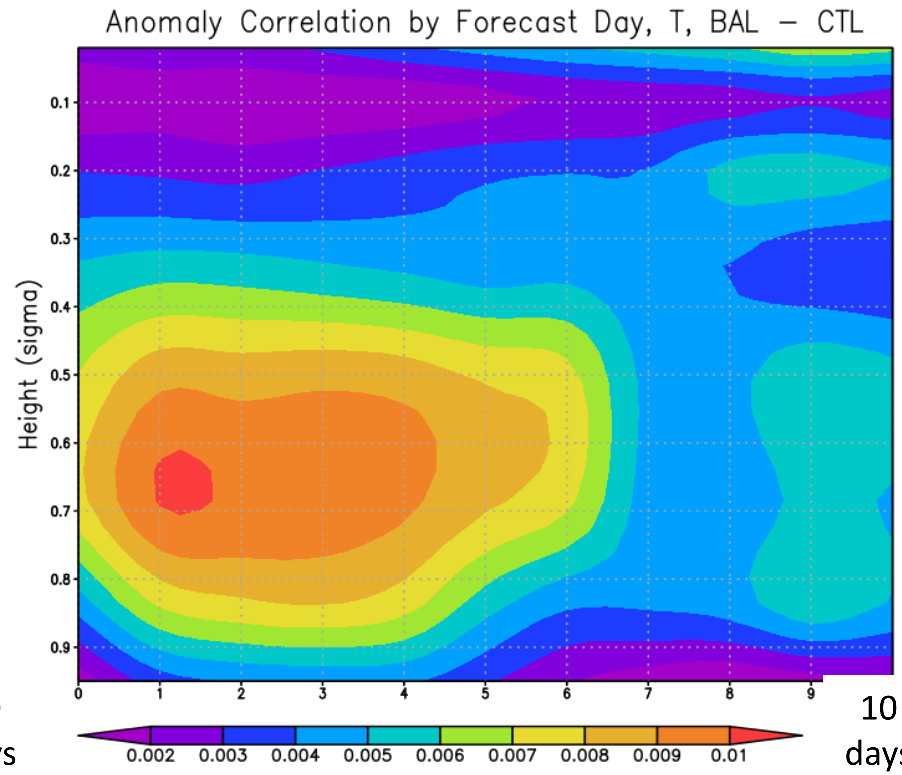
- Significant positive impact where **Γ** works in full column (ψ and *T*)
- Negative impact on ψ in stratosphere where the model bias is prominent

Forecast Skill

- Hybrid ψ

T



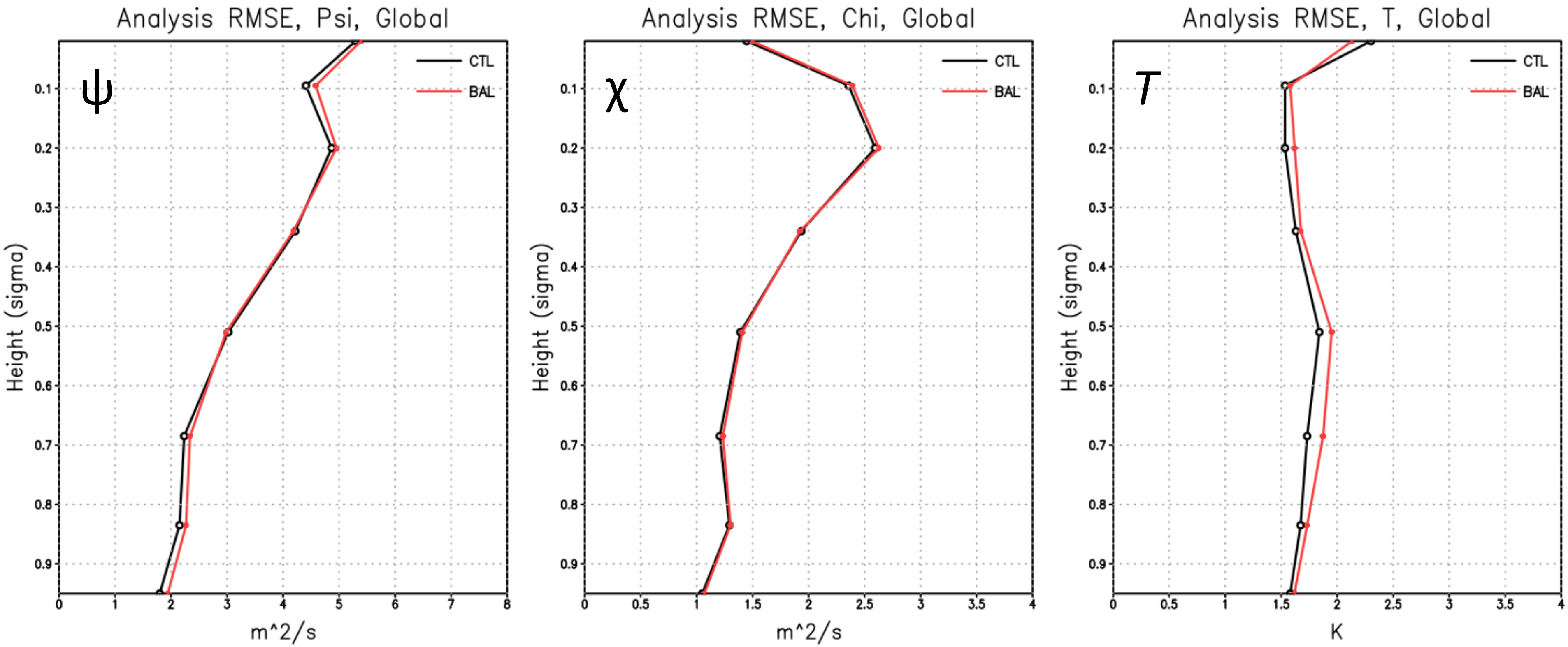


- Forecast skill for T and tropospheric ψ are improved for all forecast lengths.

BAL – CTL
> 0 BAL Improves
< 0 BAL Degrades

Analysis Skill

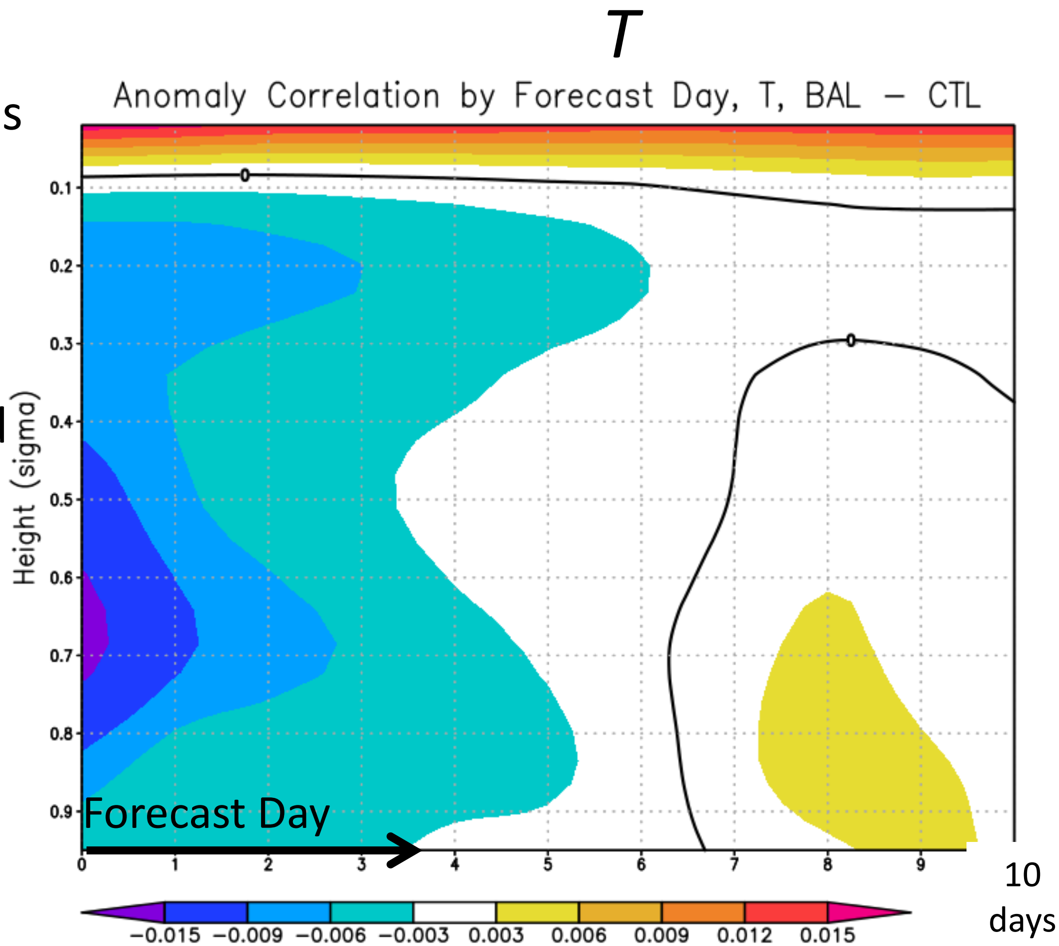
- LETKF



- Negative impact on ψ may arise through model integration
- Negative impact on *T* due to **Γ** the effect of to move analysis away from obs

Forecast Skill

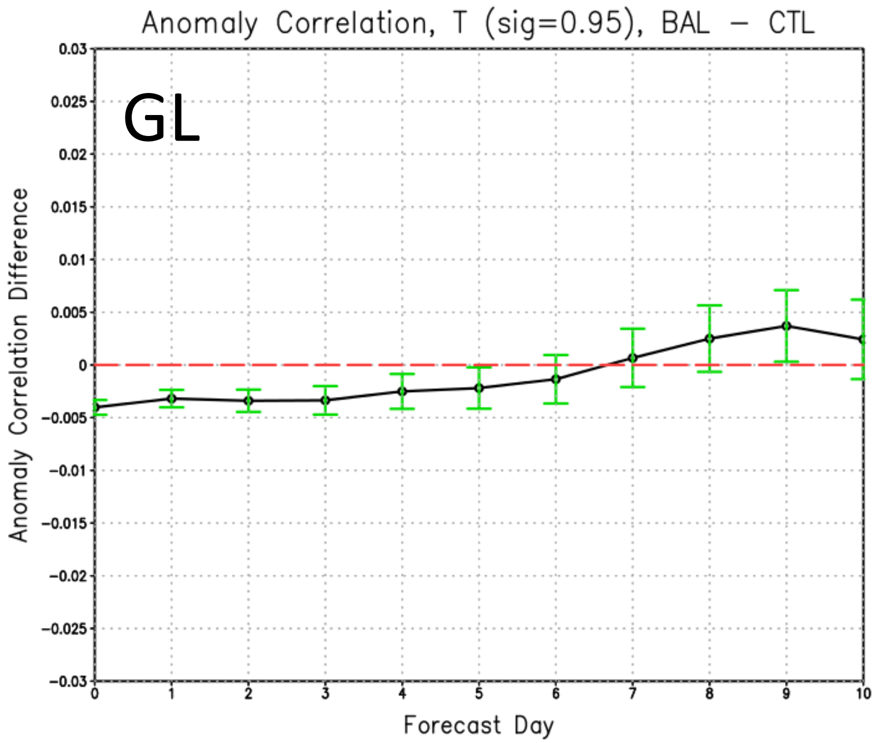
- LETKF
  - Stratospheric improvement is dominated by the southern polar region.
  - The forecast skill is degraded for short forecast lead times because *T* adjusts to ψ.
  - At longer lead times, the forecast skill is improved.
  - *Is the improvement due to the balance operator?*

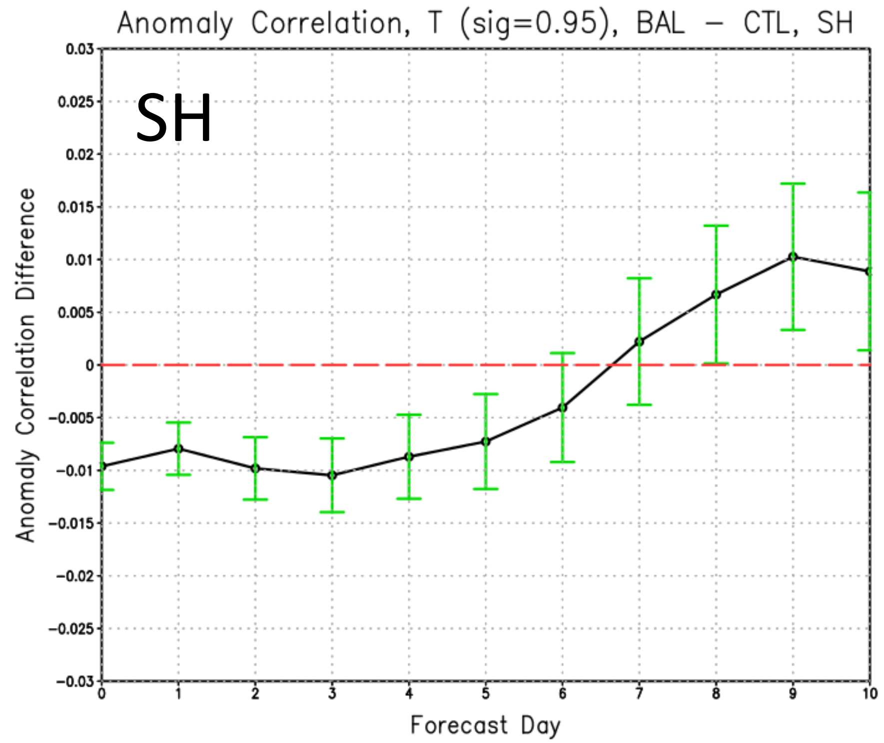


BAL – CTL
> 0 BAL Improves
< 0 BAL Degrades

Forecast Skill

- LETKF





- Transition from negative to positive skill
  - is significant
  - amplified for regions where the balance operator has a greater impact

Summary

- Balance operator
  - A balance operator was applied to two ensemble DA schemes: Hybrid 4DEnVar and LETKF.
  - The type of spatial localization impacts the effectiveness of the balance operator, with the Hybrid 4DEnVar showing greater improvements than the LETKF.
- Variable localization
  - Two forms of variable localization (VM, VO) were formulated within three ensemble DA schemes (EnSRF, LETKF, EnVar).
  - The form of variable localization makes a larger difference in application than the DA scheme.
- Overall
  - Construction of the background error covariance is critical to model performance.
  - The form of localization, either model space or observation space, is significant for many applications.