Algorithm (2): Background Error Modeling and Estimation

Jake Liu[1], **Dale Barker[2]**

[1]National Center for Atmospheric Research

[2]Met Office, United Kingdom

WRFDA Tutorial, July 2019

Outline

- WRFDA-3DVAR: Incremental formulation
- Background error covariance ($\mathbf{B}$) modeling within WRFDA
- Background error covariance ($\mathbf{B}$) estimation: GEN_BE
- Visualizing $\mathbf{B}$: Single Observation Test

WRFDA-3DVar Equation

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}[H(\mathbf{x})-\mathbf{y}]^{\mathrm{T}}\mathbf{R}^{-1}[H(\mathbf{x})-\mathbf{y}]$$

$J(\mathrm{x})$: Scalar cost function

x: The analysis: what we're trying to find!

x^b: Background field (previous forecast)

B: Background error covariance matrix

y: Observations

H: Observation operator: computes model-simulated obs

R: Observation error covariance matrix

However, this cost function is not really what WRFDA uses!

Incremental formulation of 3DVAR and Outer Loop

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[H(\mathbf{x}) - \mathbf{y}]^{\mathrm{T}}\mathbf{R}^{-1}[H(\mathbf{x}) - \mathbf{y}]$$

- Wish to linearise cost function to simplify minimization and reduce cost

- Define first guess x^g and analysis increment $\delta\mathrm{x}$: $\mathrm{x} = \mathrm{x}^g + \delta\mathrm{x}$

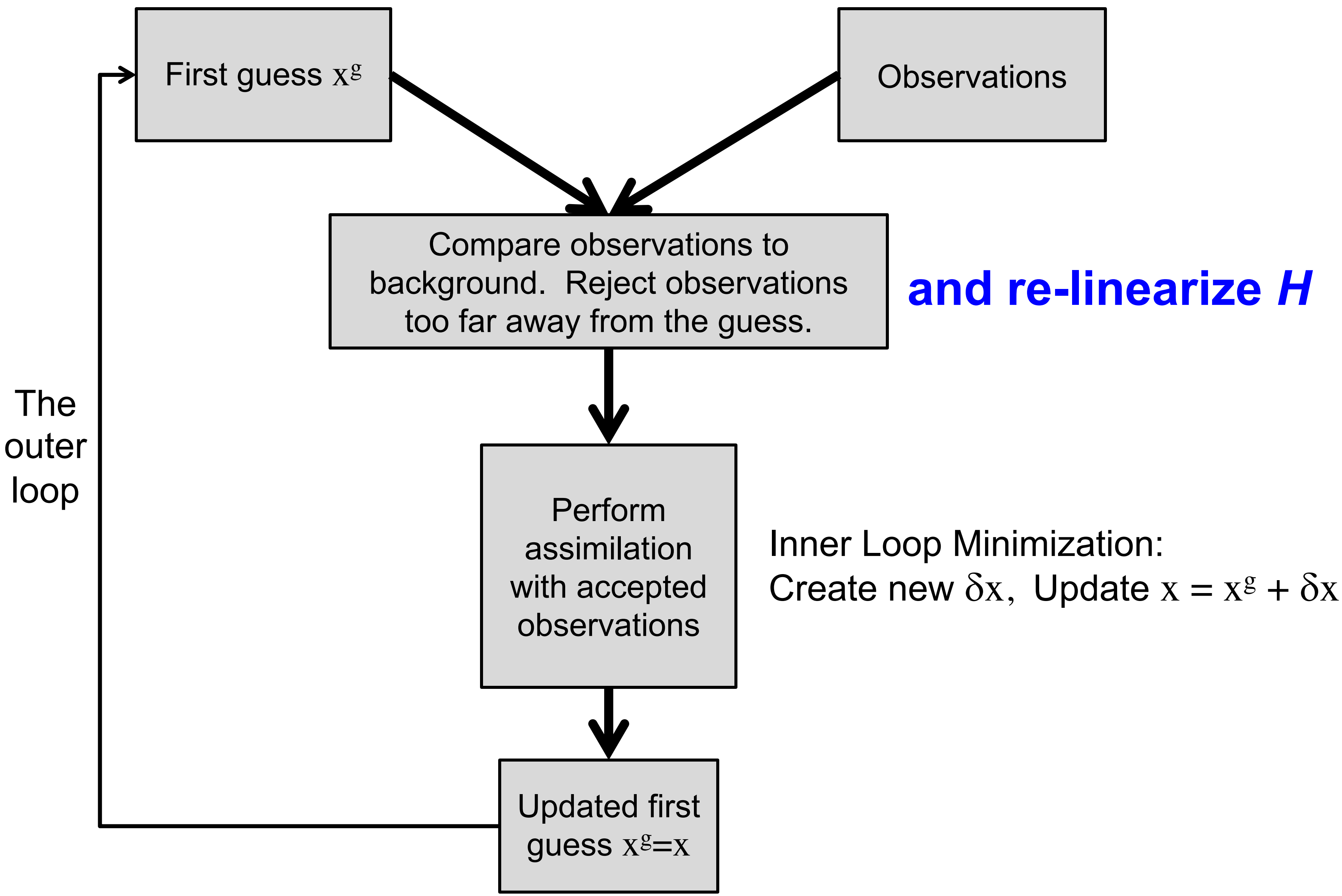
- Define innovation vector $\mathrm{d} = \mathrm{y} - H(\mathrm{x}^g)$ then cost function can be written

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x} - \delta\mathbf{x}^g)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x} - \delta\mathbf{x}^g) + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})$$

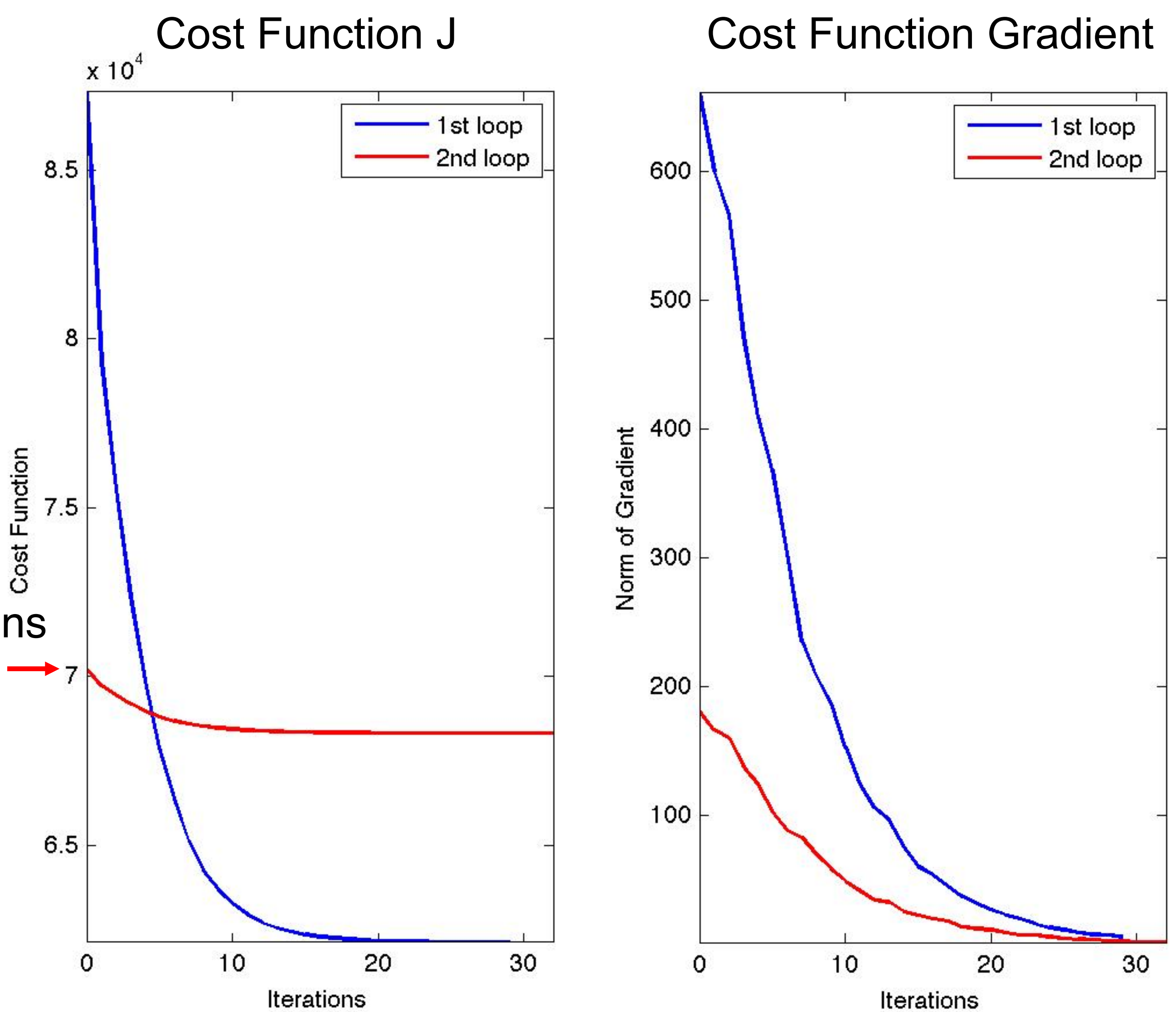
where $\delta\mathrm{x}^g = \mathrm{x}^b - \mathrm{x}^g$ and H is linearization of nonlinear obs. operator H

- To start, $\delta\mathrm{x} = 0$ and $\mathrm{x}^g = \mathrm{x}^b$. At end of minimization $\mathrm{x}^g = \mathrm{x}$ (latest analysis)

Simplistic Outer Loop Schematic



Cost Function/Gradient with 2 outer loops



More observations added in the 2nd outer loop so starting cost function higher

Outline

- WRFDA-3DVAR: Incremental formulation
- Background error covariance (B) modeling within WRFDA
- Background error covariance (B) estimation: GEN_BE
- Visualizing B: Single Observation Test

Control Variable Transform ($\delta x = U v$)

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x} - \delta\mathbf{x}^g)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x} - \delta\mathbf{x}^g) + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})$$

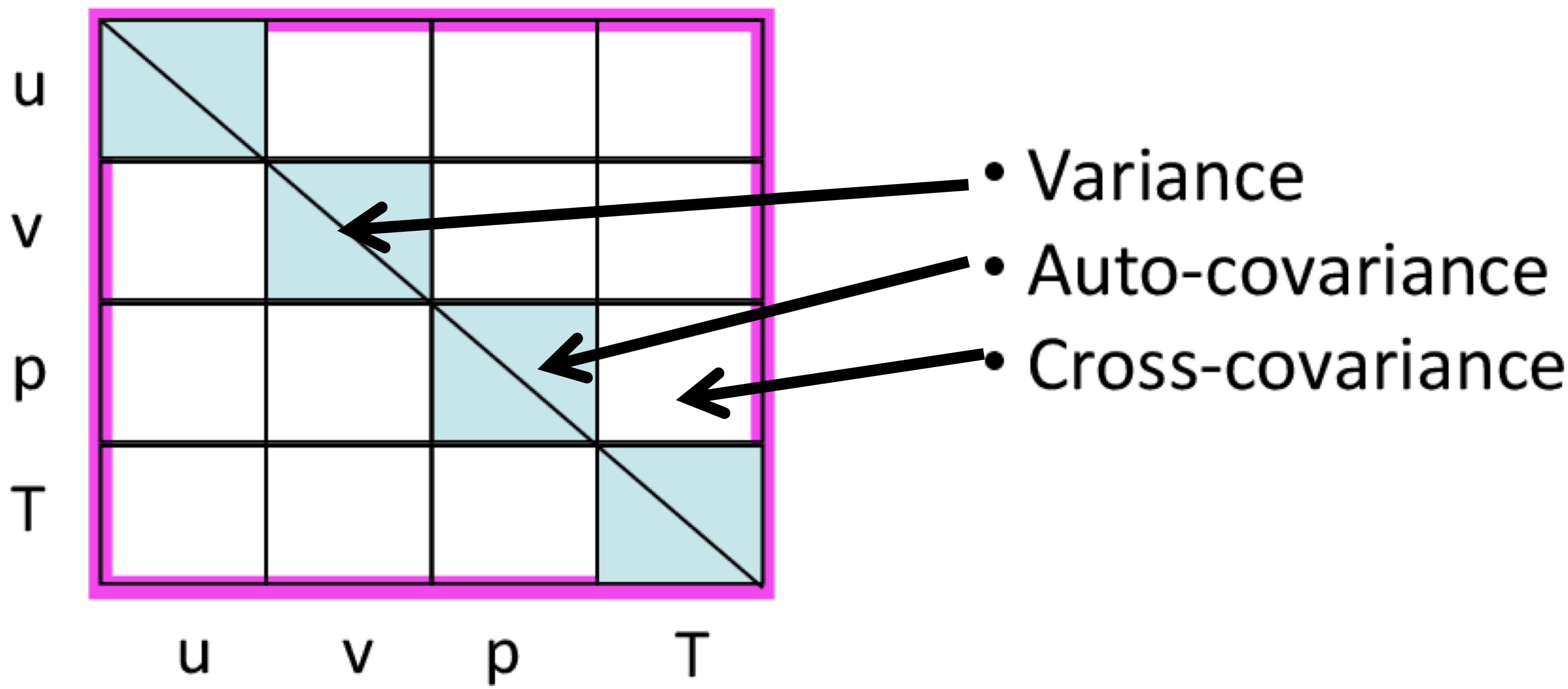
- B is a huge matrix, cannot hope to invert in practice so need to simplify.
- Write $B = UU^T$ and $\delta x = U v$ where v is the **control variable**
- The cost function can then be written:

$$J(\mathbf{v}) = \frac{1}{2}(\mathbf{v} - \mathbf{v}^g)^{\mathrm{T}}(\mathbf{v} - \mathbf{v}^g) + \frac{1}{2}(\mathbf{HUv} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HUv} - \mathbf{d})$$

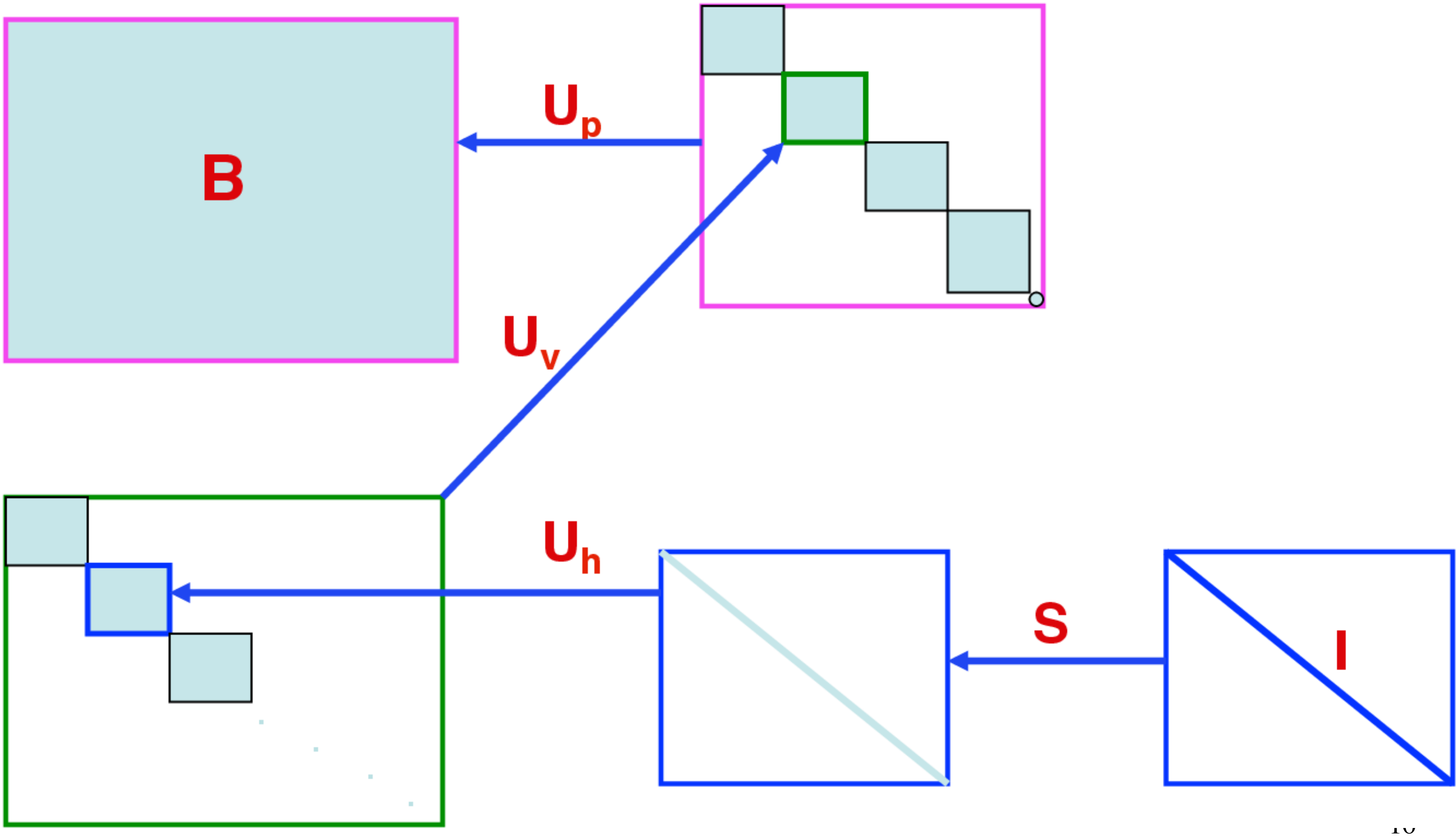
- In WRFDA, $U = U_p U_v U_h$ where:
  - U_p is the physical transform, performed via linear regression.
  - U_v is the vertical transform, performed via EOF decomposition
  - U_h is the horizontal transform, performed via recursive filters.

Properties of B matrix

- **B is square and symmetric**
- **B is positive semi-definite, eigenvalues are positive**



Background Error: Covariance Modeling



WRFDA Control Variables: Up

cv_options	Analysis variables
3	Ψ, unbalanced X, unbalanced t, pseudo rh and unbalanced log (P_s)
5	Ψ, unbalanced X, unbalanced t, pseudo rh and unbalanced P_s
6	Ψ and unbalanced X, unbalanced t, unbalanced pseudo rh and unbalanced P_s
7	u, v, t, pseudo rh, and Ps

- Forecast errors of model space variables (i.e. δx) are typically correlated (e.g. wind, pressure).
- Control variables are designed to have NO cross- (*multivariate*) correlations.

Physical Transform U_p and its inverse

U_p : convert unbalanced field to full field, e.g., for CV5

- Velocity potential/streamfunction regression: $\chi_b(k) = c(k)\psi(k)$;
- Temperature/streamfunction regression: $T_b(k) = \sum_{k1} G(k1,k)\psi(k1)$; and
- Surface pressure/streamfunction regression: $p_{sb} = \sum_{k1} W(k1)\psi(k1)$.

U_p^{-1} : convert full field to unbalanced field

e.g., **for CV6**

$$\chi_u(i,j,k) = \chi(i,j,k) - \alpha_{\psi\chi}(i,j,k)\psi(i,j,k) \tag{7}$$

$$T_u(i,j,k) = T(i,j,k) - \sum_{l=1}^{N_k} \alpha_{\psi T}(i,j,k,l)\psi(i,j,l) - \sum_{l=1}^{N_k} \alpha_{\chi_u T}(i,j,k,l)\chi_u(i,j,l) \tag{8}$$

$$ps_u(i,j) = ps(i,j) - \sum_{l=1}^{N_k} \alpha_{\psi ps}(i,j,l)\psi(i,j,l) - \sum_{l=1}^{N_k} \alpha_{\chi_u ps}(i,j,l)\chi_u(i,j,l) \tag{9}$$

$$rh_u(i,j,k) = rh(i,j,k) - \sum_{l=1}^{N_k} \alpha_{\psi rh}(i,j,k,l)\psi(i,j,l) - \sum_{l=1}^{N_k} \alpha_{\chi_u rh}(i,j,k,l)\chi_u(i,j,l) - \sum_{l=1}^{N_k} \alpha_{T_u rh}(i,j,k,l)T_u(i,j,l) - \alpha_{ps_u rh}(i,j,k)ps_u(i,j)$$

No balance transform for CV7 as all control variables are full fields U, V, T, Q/Qs, Ps

Vertical EOF transform: $\mathbf{U_v}$

- Vertical part of $B = E\Lambda E^T$
  - E: matrix formed by eigenvectors of vertical covariance matrix
  - Λ: diagonal matrix formed by eigenvalues of vertical covariance matrix

- Define $U_v = E\Lambda^{1/2}$

- Inverse transform $U_v^{-1}=\Lambda^{-1/2}E^T$

- EOF can be truncated to save cost:
  - Default setting in WRFDA: 99% of total eigenvalues

Horizontal Transform: U_h

3.1 Recursive Filter

For one-dimensional n grid points, one pass filter is a right-moving recursive filter

$$B_i = \alpha B_{i-1} + (1-\alpha)A_i, i = 1, ..., n \tag{21}$$

followed by a left-moving recursive filter

$$C_i = \alpha C_{i+1} + (1-\alpha)B_i, i = n, ..., 1 \tag{22}$$

Inverse filter is non-recursive

$$A_i = C_i - \frac{\alpha}{(1-\alpha)^2}(C_{i-1} - 2C_i + C_{i+1}) \tag{23}$$

Infinite-pass recursive filter is equivalent to the convolution of a Gaussian covariance function with unfiltered field. U is $\frac{N}{2}$-pass recursive filter with N the total number of passes. Need boundary condition B_0 and C_{n+1}. α is calculated according to the horizontal correlation length-scale.

(1) 2D filter is done by a filter in X-direction, followed by a filter in Y-dir.

(2) namelist rf_passes=6: 3 passes for U, 3 passes for U^T (adjoint of U)

Outline

- WRFDA-3DVAR: Incremental formulation
- Background error covariance ($\mathbb{B}$) modeling within WRFDA
- Background error covariance ($\mathbb{B}$) estimation: GEN_BE
- Visualizing $\mathbb{B}$: Single Observation Test

How to estimate B?

- The background error covariance matrix is defined as

$$B = < \varepsilon \, \varepsilon^T > = < (x - x^t) \, (x - x^t)^T >$$

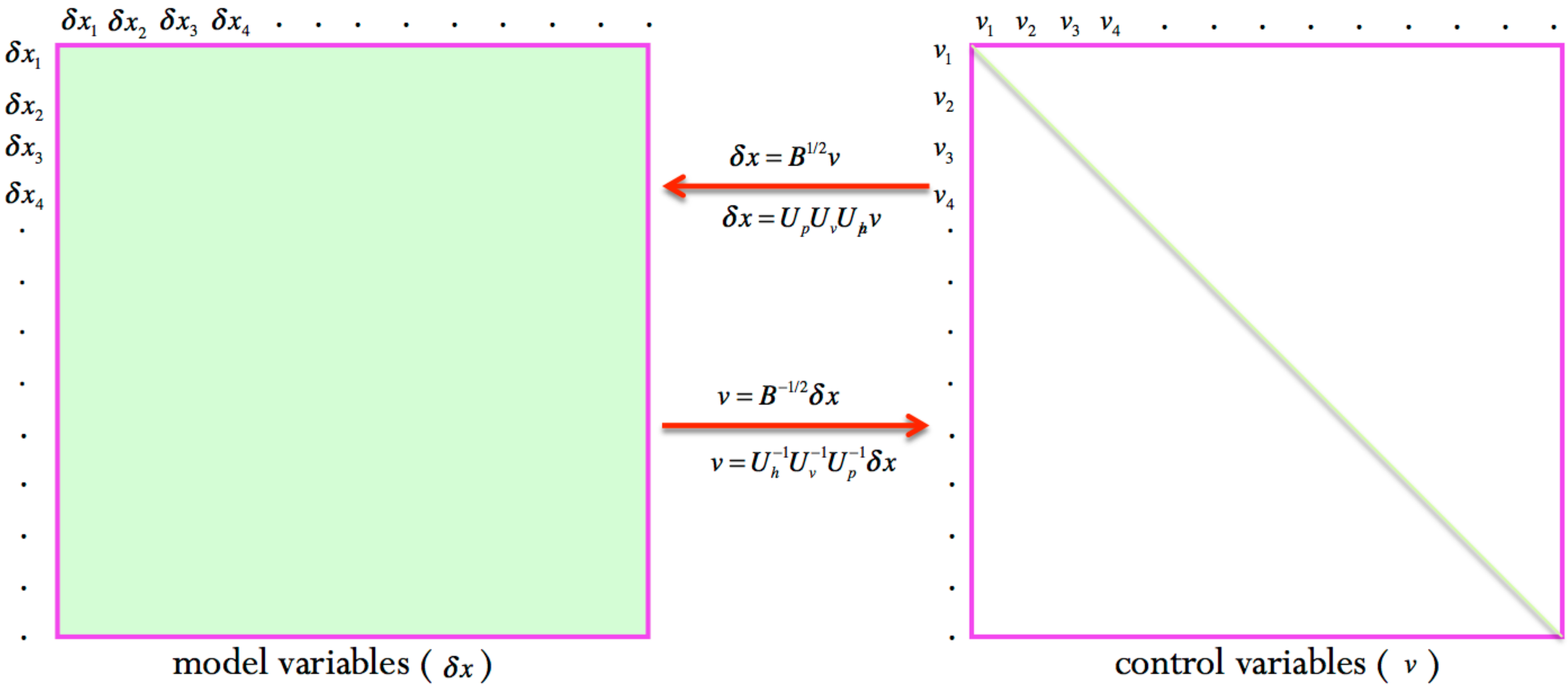
- Where ε is the forecast error and x^t the truth (which we do not know).

- In practice, we approximate $x - x^t$ by one of two methods:

  – NMC-method: $x - x^t = x^{t1} - x^{t2}$ the difference between two forecasts valid at the same time.

  – Ensemble technique: $x - x^t = x^{ens} - <x^{ens}>$ a set of ensemble perturbations from an ensemble forecasting system.

GEN_BE: Perform U inverse transform



GEN_BE computes:

1) Balance regression coefficients between analysis variables,
2) Eigenvectors/eigenvalues of vertical covariances, and
3) horizontal correlation length-scales (as a function of EOF mode),

using large enough sample dataset of forecast difference or ensemble

GEN_BE Stage0: Create 'Standard' Fields

- Estimate forecast errors using NMC or ensemble method for each of the following 'standard' fields:

ψ - Stream function
χ - Velocity potential
T - Temperature
q/q_s - Relative humidity
p_s - Surface pressure

- Convert (u,v) to vorticity (ζ) and horizontal divergence (D)

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \qquad D = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$$

- Convert ζ and D to ψ and χ, via solution of Poisson equations:

$$\nabla^2 \chi = D \qquad \nabla^2 \psi = \zeta$$

GEN_BE Stage1: remove temporal mean

- Computes temporal mean of the forecast error samples generated in stage0

- Removes temporal mean to form the perturbations for standard fields.

ψ - Stream function
χ - Velocity potential
T - Temperature
q/q_s - Relative humidity
p_s - Surface pressure

GEN_BE Stage2 & 2a

- Stage2: Computes regression coefficients (i.e., linear correlation coefficients) between two variables (M, N, P, Q, R, S below)

- Stage2a: Obtain unbalanced fields by removing the balanced part of fields using inverse transform U_p^{-1}

- Balance transform U_p in matrix form: completes $v_p = U_p^{-1}\, dx$

$$
\begin{pmatrix} \Psi \\ \chi \\ t \\ Ps \\ rh \end{pmatrix} = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ M & I & 0 & 0 & 0 \\ N & 0 & I & 0 & 0 \\ Q & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix} \begin{pmatrix} \psi \\ \chi_u \\ t_u \\ Ps_u \\ rh \end{pmatrix}
$$

CV5

$$
\begin{pmatrix} \Psi \\ \chi \\ t \\ Ps \\ rh \end{pmatrix} = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ M & I & 0 & 0 & 0 \\ N & P & I & 0 & 0 \\ Q & R & 0 & I & 0 \\ S_1 & S_2 & S_3 & S_4 & I \end{pmatrix} \begin{pmatrix} \psi \\ \chi_u \\ t_u \\ Ps_u \\ rh_u \end{pmatrix}
$$

CV6

GEN_BE Stage3: EOF of vertical covariances

- Computes vertical covariances for unbalanced 3D fields

- Performs EOF decomposition of vertical covariances to obtain eigenvectors and eigenvalues

- Projects unbalanced fields into vertical modes using inverse transform: calculates $v_v = U_v^{-1} v_p$

GEN_BE Stage4

- Calculate horizontal error correlation as a function of distance between points

- Fit correlation to a Gaussian function with a lengthscale

$$\boxed{z(r) = z(0)\exp\{-r^2 / 8s^2\}}$$

$$\boxed{y(r) = 2\sqrt{2}[\ln(z(0) / z(r)]^{1/2} = r / s}$$

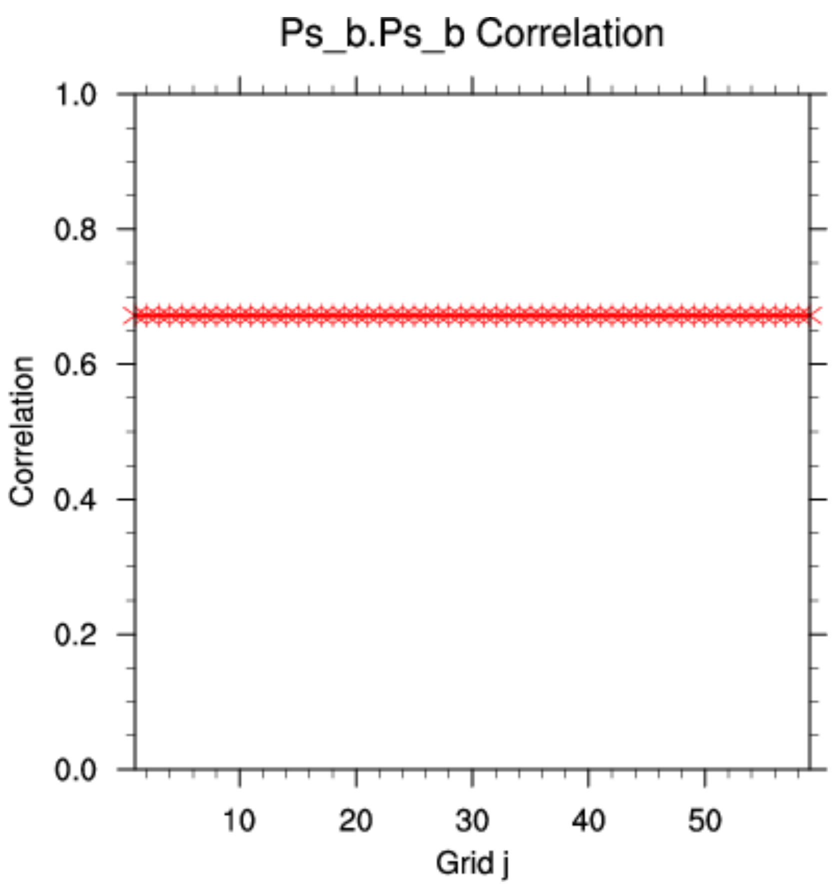
- Calculate horizontal lengthscale for each component of v_v (vertical modes of each control variable) simultaneously – parallelism at the script level.

Outline

- WRFDA-3DVAR: Incremental formulation
- Background error covariance (B) modeling within WRFDA
- Background error covariance (B) estimation: GEN_BE
- Visualizing B: Single Observation Test

Up: Multivariate Correlation Example

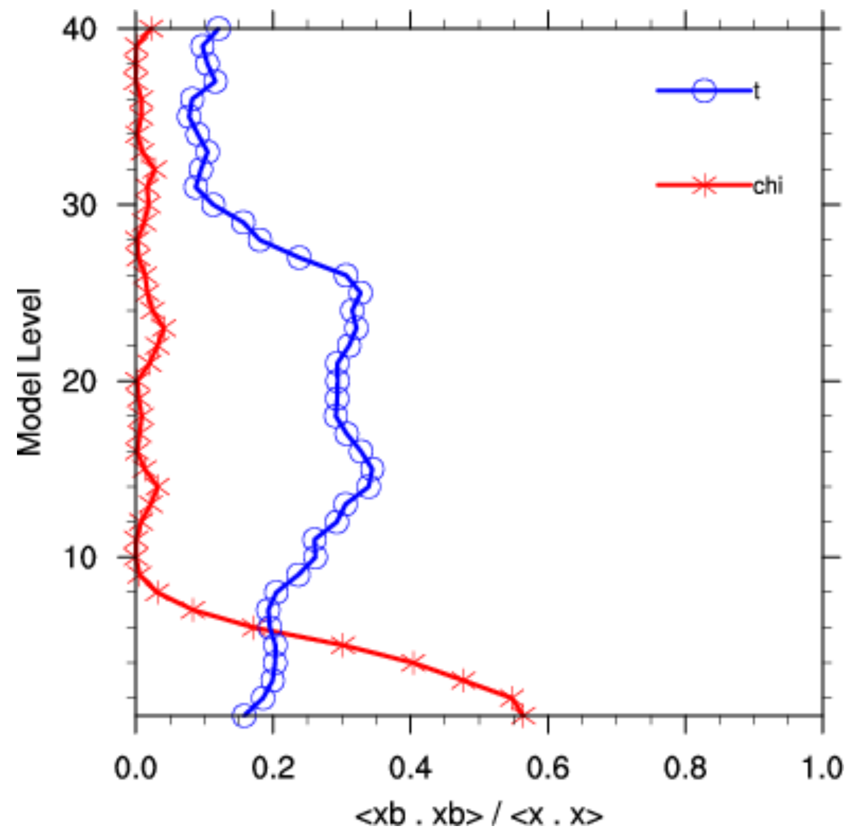
Correlation between surface pressure and balanced pressure $p_{sb} = \sum_{k1} W(k1)\psi(k1).$



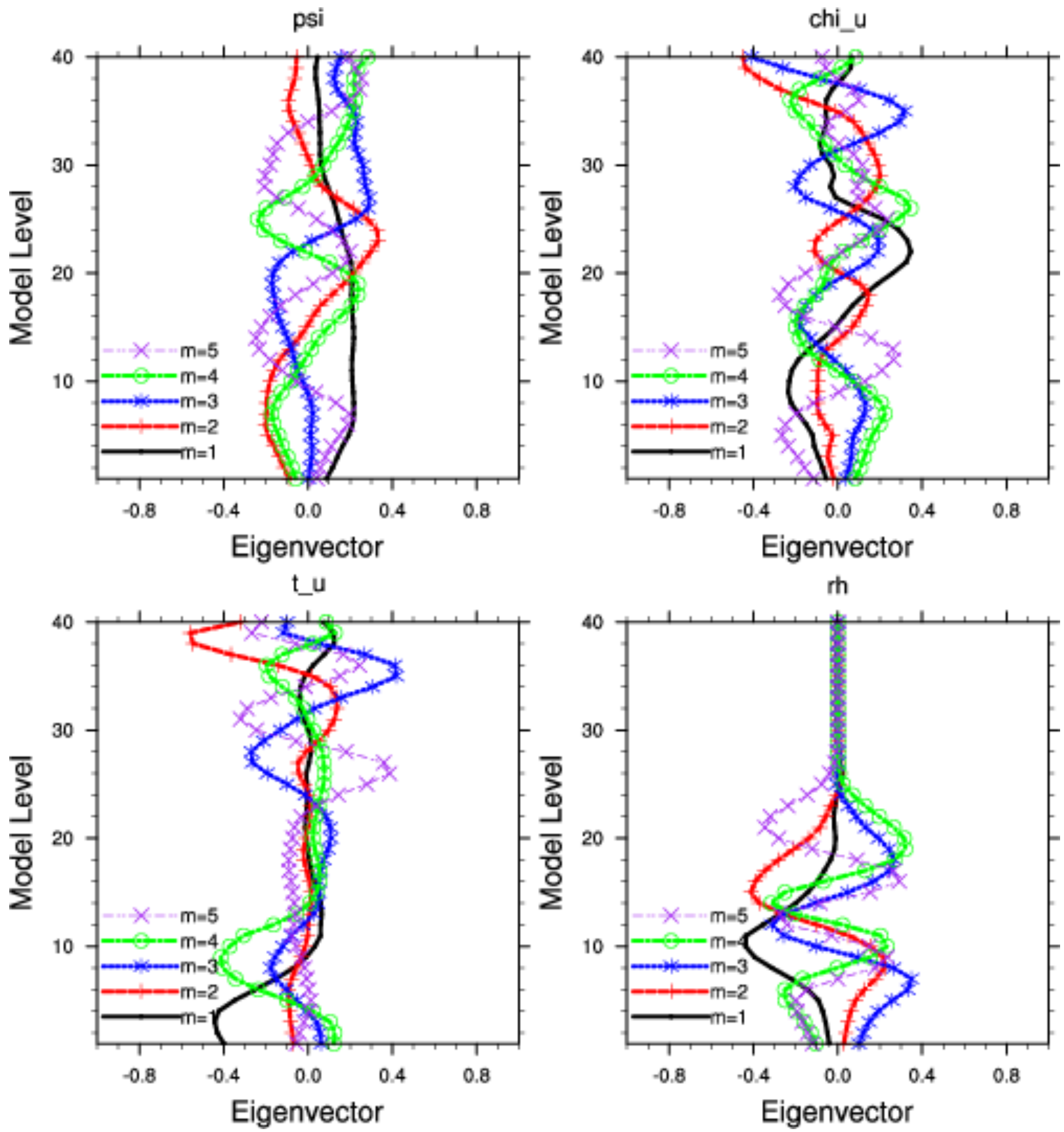
Correlation between temp/ and balanced temp/potential

$$\chi_b(k) = c(k)\psi(k)$$

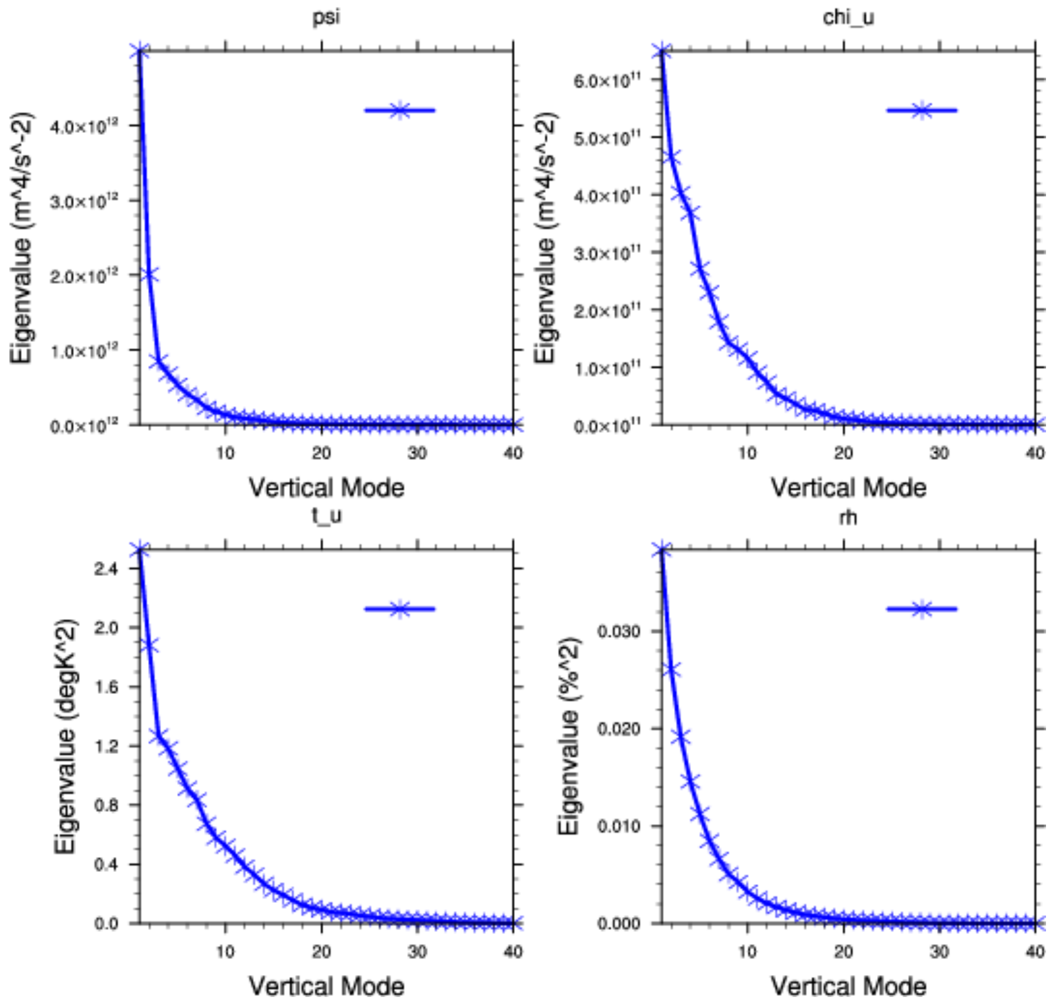
$$T_b(k) = \sum_{k1} G(k1,k)\psi(k1)$$



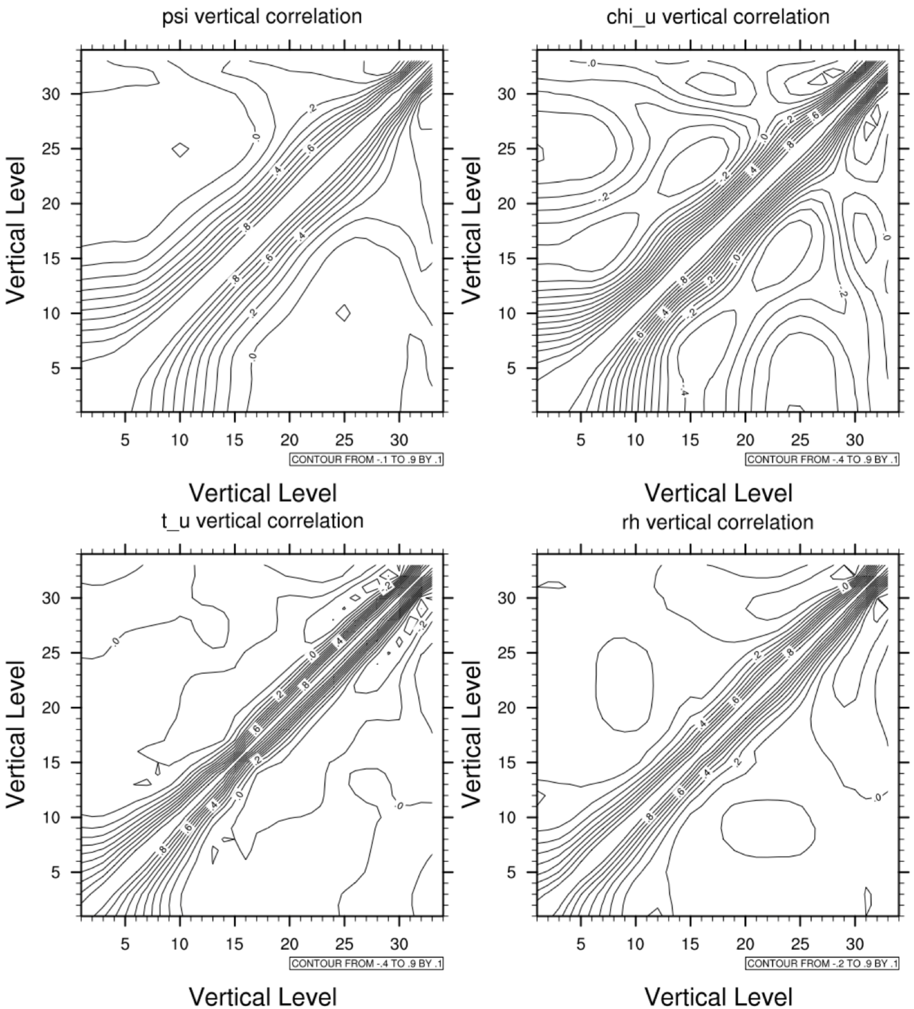
Uv: Leading (First 5) Eigenvectors Example



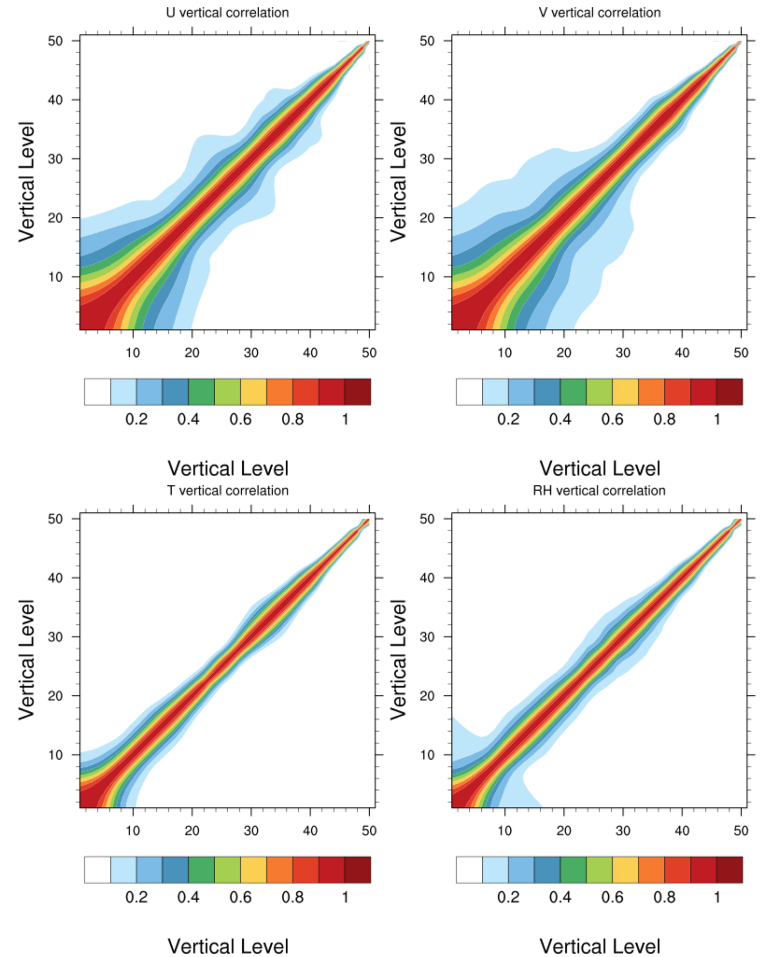
Uv: Eigenvalues Example



Vertical Correlation ($U_v U_v^T = E \Lambda E^T$)

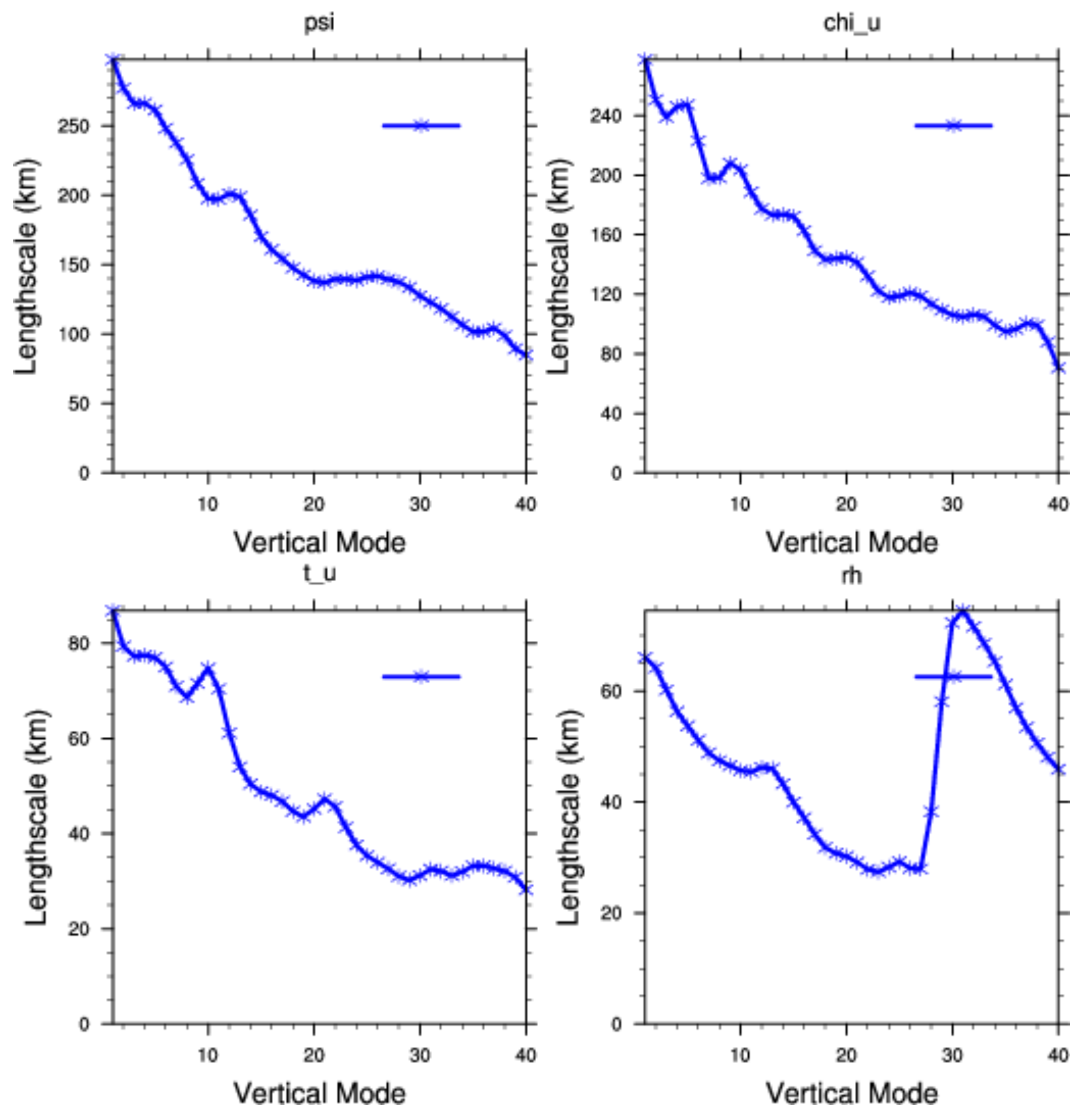


cv_options=5

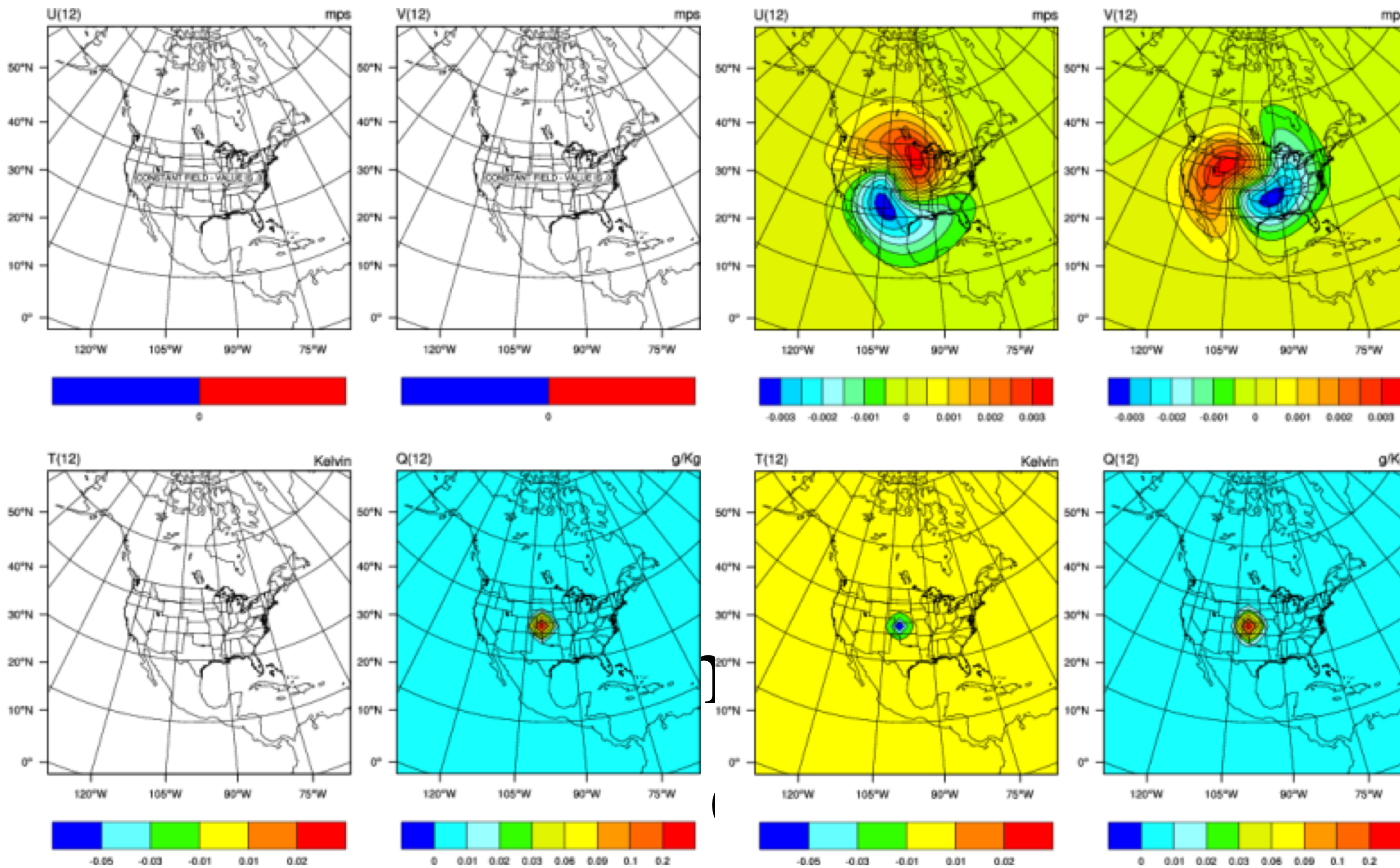


cv_options=7

Uh: Horizontal Lengthscales Example



The End Result: Single q Observation Test



cv_options=5

cv_options=6

Empirical BE Tuning via namelist parameters

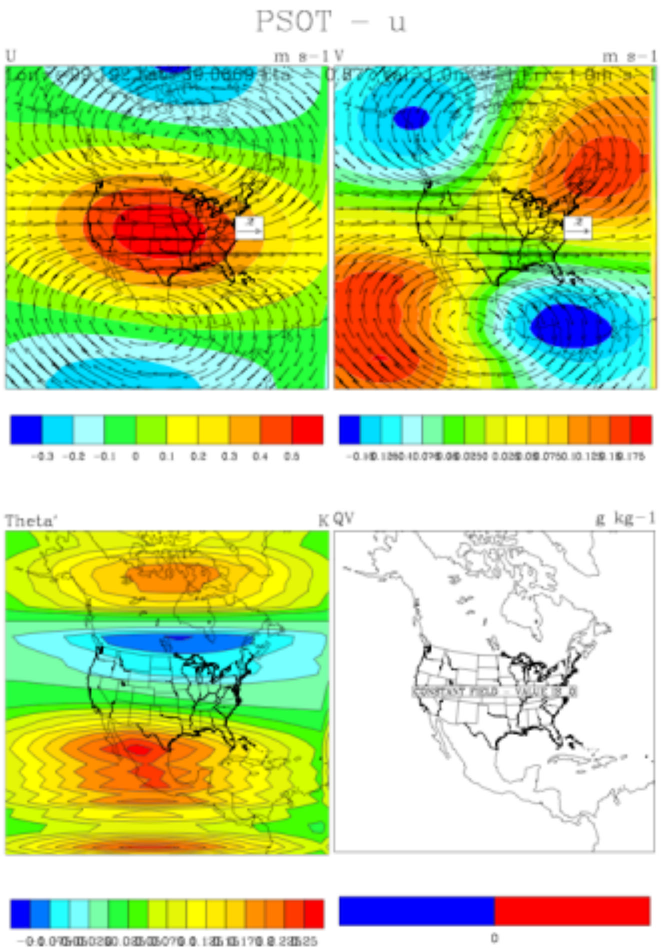
- Horizontal component of BE can be tuned with following namelist parameters

  - LEN_SCALING1 - 5 (Length scaling parameters)
  - VAR_SCALING1 - 5 (Variance scaling parameters)

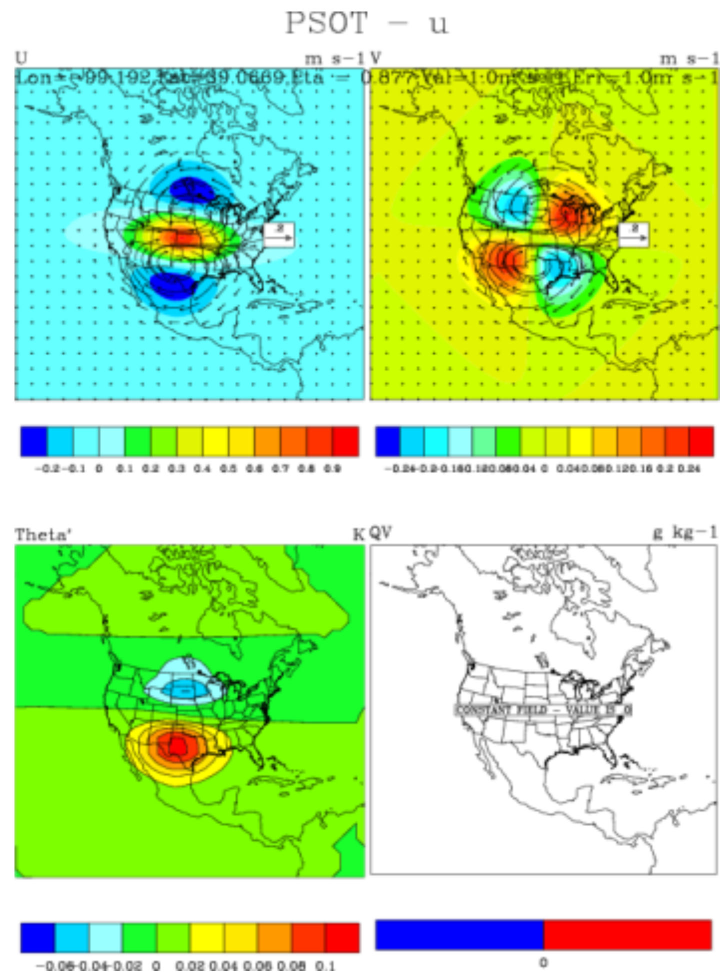
- Vertical component of BE can be tuned with the following namelist parameters

  - MAX_VERT_VAR1 - 5 (Vertical variance parameters)

Impact of Empirical Tuning of B



no tuning



tuning (len_scaling1 & 2 =0.25)