Algorithm (2): Background Error Modeling and Estimation

Zhiquan (Jake) Liu

Mesoscale & Microscale Meteorology Laboratory
National Center for Atmospheric Research

WRFDA Tutorial, July 2017

Outline

- WRFDA-3DVAR: Incremental formulation
- B matrix modeling within WRFDA
- B matrix estimation: GEN_BE package
- Visualize B effect: single observation test

WRFDA-3DVar Equation

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

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

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

$\mathbf{x_b}$**:** Background field

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

1.1 Non-linear 3DVAR Formulation

Non-linear 3DVAR cost function

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

1.2 Incremental 3DVAR Formulation

Linearization, let $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^g$ and $\delta\mathbf{x}^g = \mathbf{x}^b - \mathbf{x}^g$, thus $\mathbf{x} = \delta\mathbf{x} + \mathbf{x}^g$, we have

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

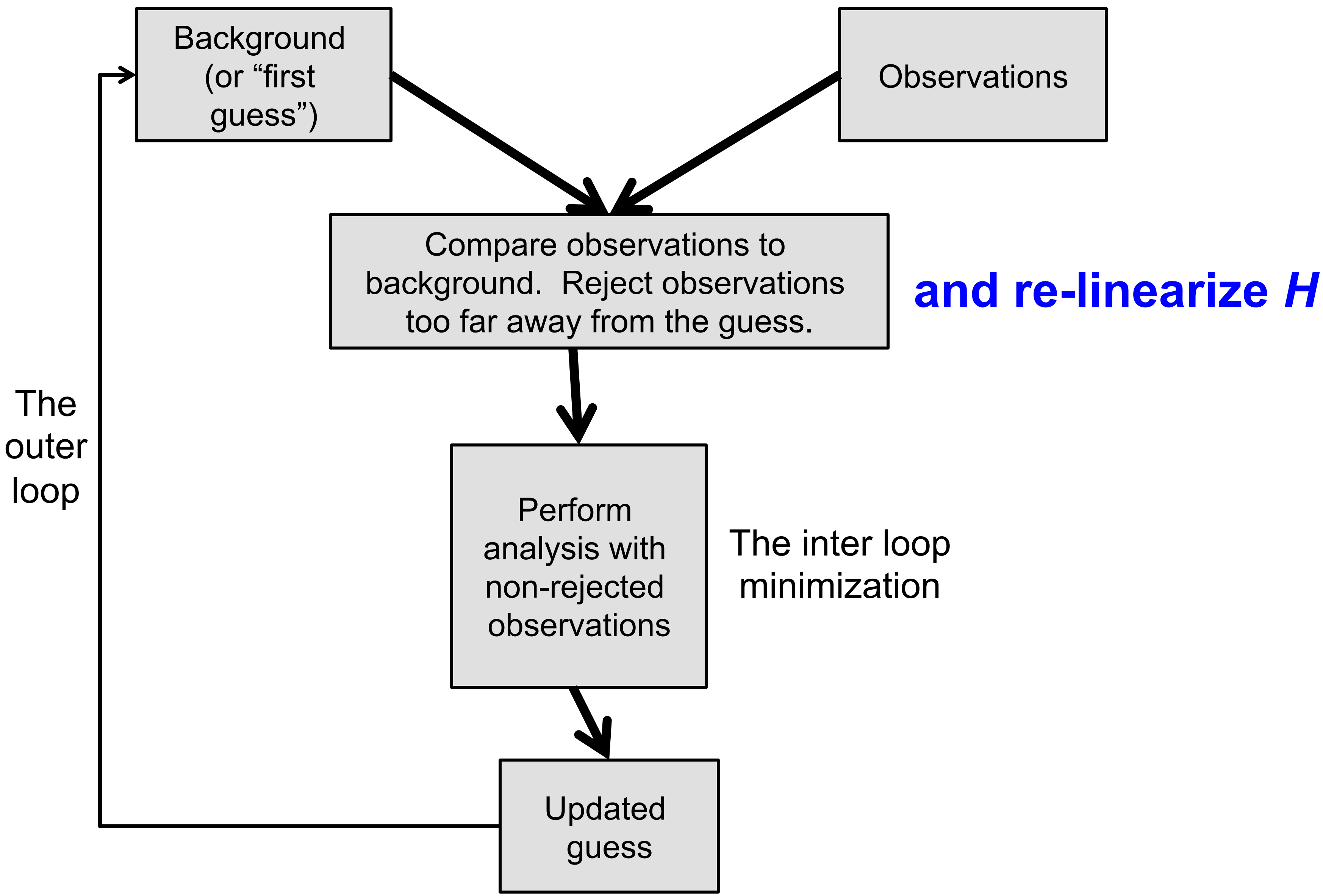
Do Taylor Expansion for observation term

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

where $\mathbf{d} = \mathbf{y} - H(\mathbf{x}^g)$ and $\mathbf{H}$ is the linearized version of H in the vicinity of $\mathbf{x}^g$.

NOTE: Xg is the first guess, not to be confused with the background Xb even though they are the same for the first outer loop. From the 2nd outer loop, Xg is equal to the analysis Xa from previous outer loop.

Simplistic outer loop schematic



1.3 Control Variable Transform (CVT)

To avoid the inverse calculation of large $\mathbf{B}$ matrix, do a change of variable $\delta\mathbf{x} = \mathbf{U}\mathbf{v}$ and $\delta\mathbf{x}^g = \mathbf{U}\mathbf{v}^g$ with $\mathbf{U}$ the square root of $\mathbf{B}$, namely $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{\mathrm{T}/2} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$ or $\mathbf{U} = \mathbf{B}^{1/2}$. Also $\mathbf{B}^{-1} = \mathbf{U}^{-\mathrm{T}}\mathbf{U}^{-1}$. Then the cost function with respect to the control variable $\mathbf{v}$ becomes

$$J(\mathbf{v}) = \frac{1}{2}(\mathbf{v} - \mathbf{v}^g)^{\mathrm{T}}(\mathbf{v} - \mathbf{v}^g) + \frac{1}{2}(\mathbf{H}\mathbf{U}\mathbf{v} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\mathbf{v} - \mathbf{d}) \tag{4}$$

1.4 Solution of Incremental 3DVAR

The minimization of the cost function requires its gradient with respect to $\mathbf{v}$ to be zero, namely

$$\nabla_{\mathbf{v}} J(\mathbf{v}) = (\mathbf{v} - \mathbf{v}^g) + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\mathbf{v} - \mathbf{d}) = 0 \tag{5}$$

After minimization, we get the analysis increment $\mathbf{v}^a$ in control variable space,

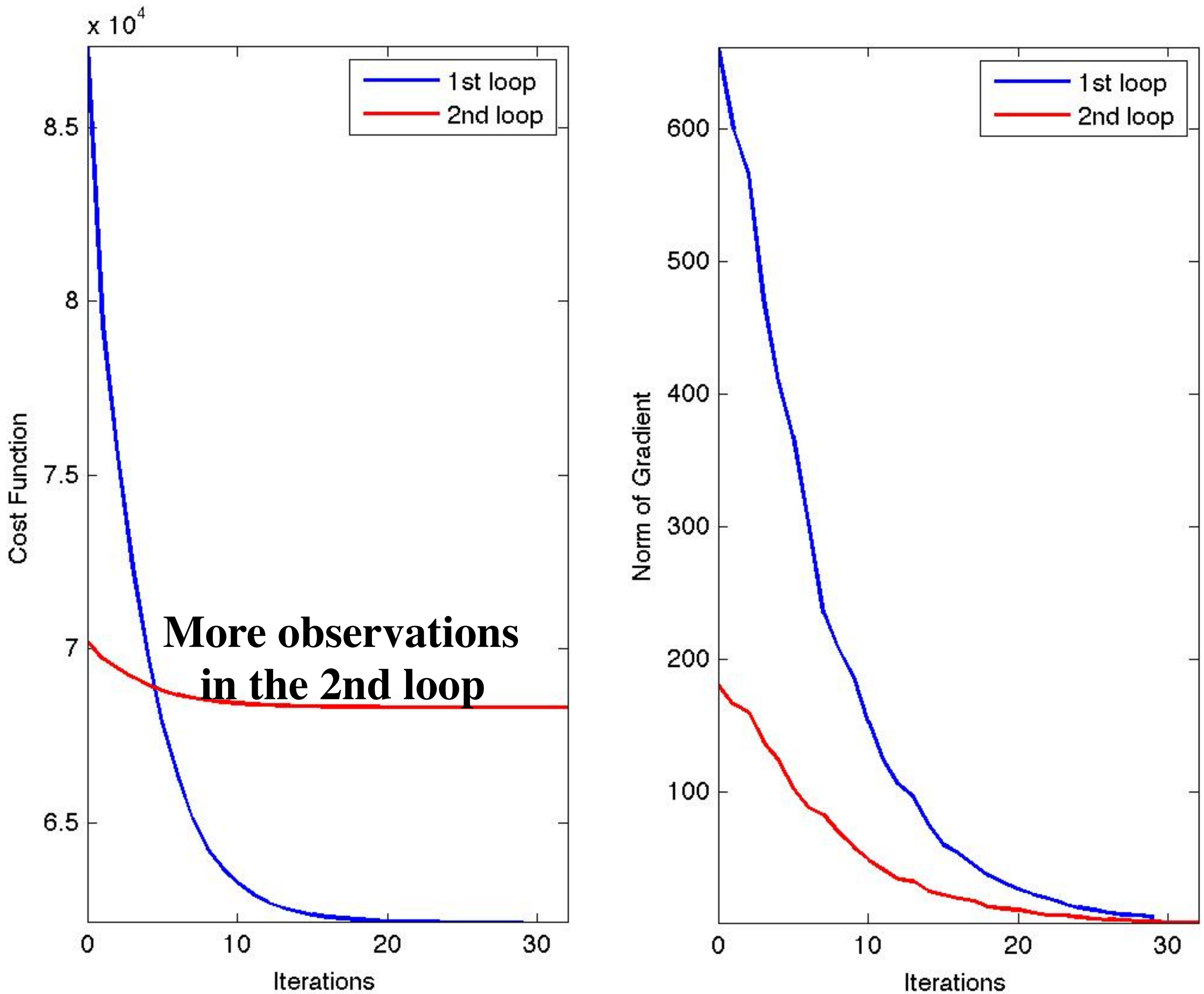
$$\mathbf{v}^a = (\mathrm{I} + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{U})^{-1}(\mathbf{v}^g + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d})$$

The analysis increment and the analysis in model space are

$$\mathbf{x}^a = \mathbf{x}^g + \delta\mathbf{x}^a = \mathbf{x}^g + \mathbf{U}\mathbf{v}^a$$

NOTE: (1) outer loop-1: Xg = Xb; Vg=0; loop-2: Xg = Xa , Vg=Va from previous loop.
(2) For each outer loop, *H* needs to be re-linearized around new Xg ;
(3) d=y-*H*(Xg) is also re-calculated and re-do QC (OMB check).

Cost Function/Gradient with 2 outer loops



Outline

- WRFDA-3DVAR: Incremental formulation
- B matrix modeling within WRFDA
- B matrix estimation: GEN_BE package
- Visualize B effect: single observation test

Role of B (or U) within DA

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

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

$$x^a - x^b = \mathrm{BH}^T(\mathrm{HBH}^T + \mathrm{R})^{-1}[y^o - H(x^b)]$$

$$\mathbf{v}^a = (\mathrm{I} + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{HU})^{-1}(\mathbf{v}^g + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d})$$

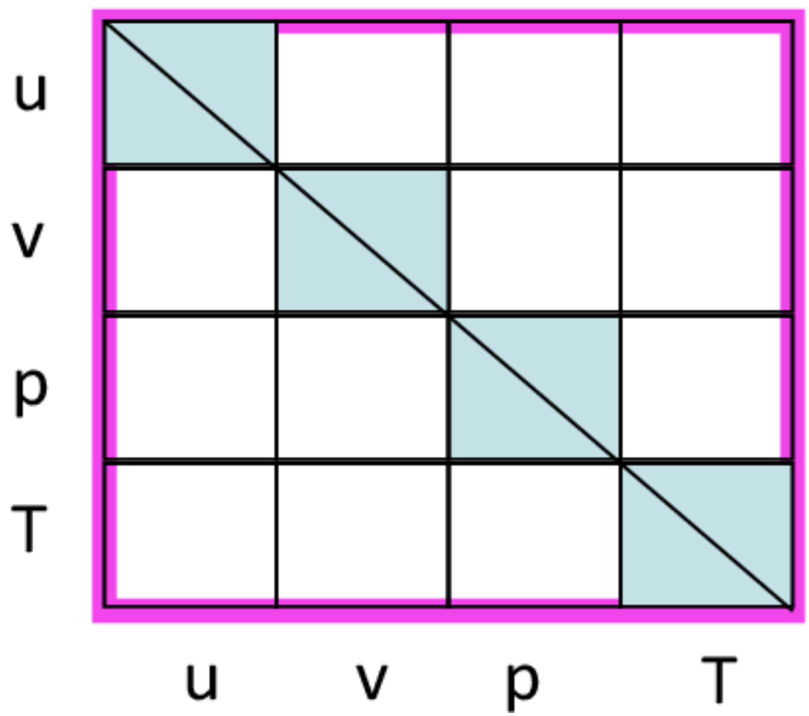
The analysis increment and the analysis in model space are

$$\mathbf{x}^a = \mathbf{x}^g + \delta\mathbf{x}^a = \mathbf{x}^g + \mathbf{U}\mathbf{v}^a$$

- **B/U** gives proper weight to the background term ($\boldsymbol{x} - \boldsymbol{x}^{\mathrm{b}}$)
- **B/U** spreads information spatially (vertical and horizontal) and across different variables

Properties of B matrix

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



- Variance
- Auto-covariance
- Cross-covariance

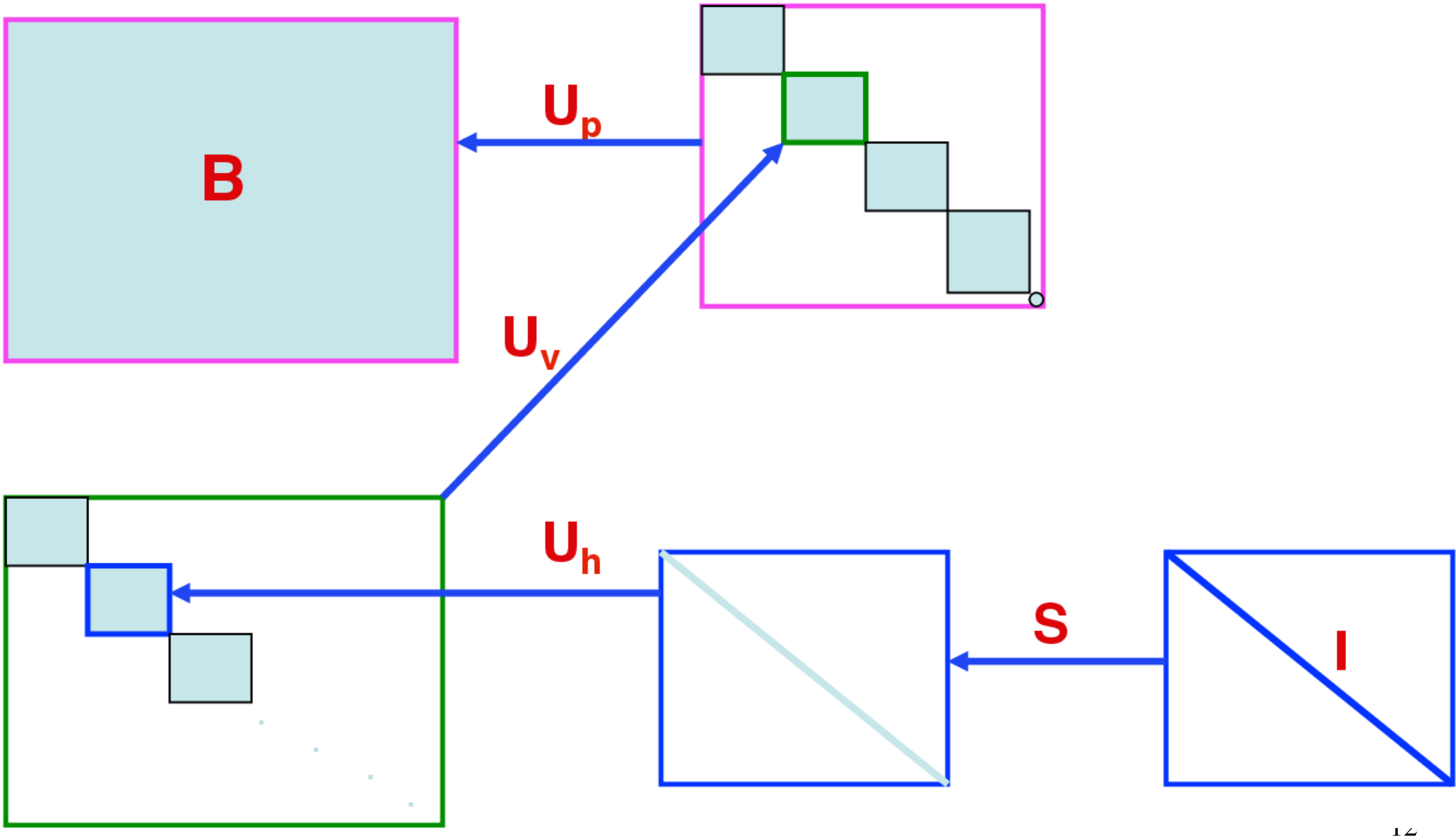
Modeling of $U=B^{1/2}$

$$B = U_p U_v U_h U_h^T U_v^T U_p^T = U_p U_v B_h U_v^T U_p^T$$

- **$U = U_p\ U_v\ U_h$**, 3 sequential transforms
  - Horizontal transform U_h via recursive filter, to model horizontal correlation of individual control variables
  - Vertical transform U_v via EOF decomposition of vertical covariance
  - Balance/Physical transform U_p via linear regression, which determines CV5, 6 or 7

Inverse transform: $U^{-1} = U_h^{-1}\ U_v^{-1}\ U_p^{-1}$

Background Error: Covariance Modeling



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)

More on recursive filter (RF)

- 2-pass RF output approximates a second-order auto-regressive (SOAR) function

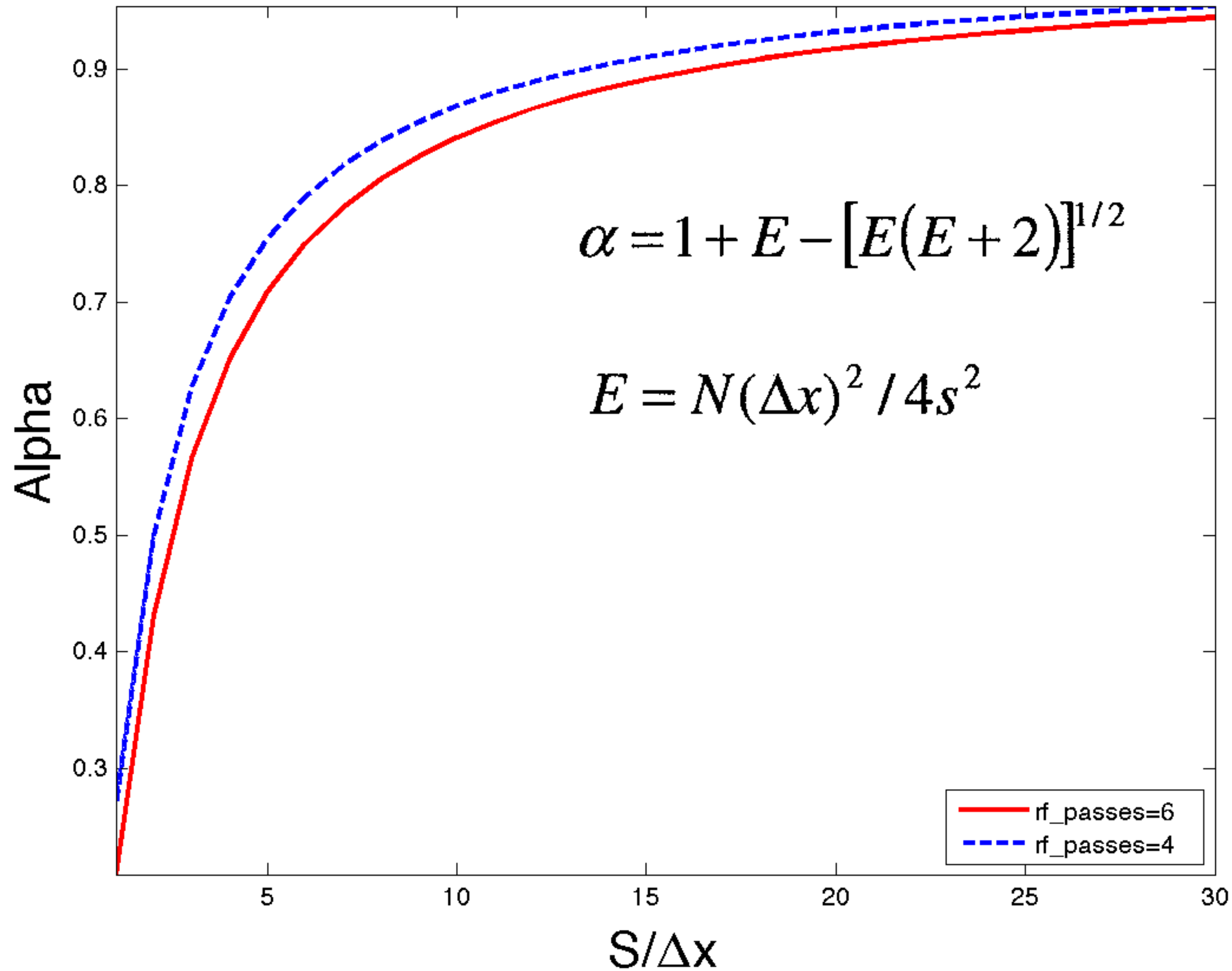
$$\mu_s(r) = \left(1 + \frac{r}{s}\right)\exp\left(-\frac{r}{s}\right)$$

- Infinite-pass RF output tends to a Gaussian function

$$\mu_g(r) = \exp\left[-\frac{1}{2}\left(\frac{r}{2s}\right)^2\right]$$

Note a factor of 2 in WRFDA, differs from normal Gaussian function

Relation between α and correction lengthscale



Vertical EOF transform: $\mathbf{U_v}$

- $U_v = E\Lambda^{1/2}$
  - E: matrix formed by eigenvectors of vertical covariance matrix
  - Λ: diagonal matrix formed by eigenvalues of vertical covariance matrix

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

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

Balance 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**

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

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

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

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

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

Choices of control variables in WRFDA

cv_options	Analysis variables
3	Ψ, unbalanced X, unbalanced t, pseudo rh and unbalanced log (P_s), Recursive filter in vertical
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, Ps and pseudo rh

- In control variable space (i.e., v), assumes no spatial and multivariate correlation.
- In model variable space (i.e., δx) after apply U transform, we have spatial and multivariate correlation

Control Variable Options in WRFDA

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

• ***What type of background error covariance do I want to use?***

&WRFVAR7

- cv_options: Background error covariance option
  - cv_options = 3 : global, default…see …/var/run/be.dat.cv3
  - cv_options = 5 : regional, generated by "gen_be"
  - cv_options = 6 : regional, generated by "gen_be" with multivariate moisture correlation, new in WRFDA V3.3
  - cv_options = 7 : regional, generated by "gen_be", new in WRFDA V3.7

Outline

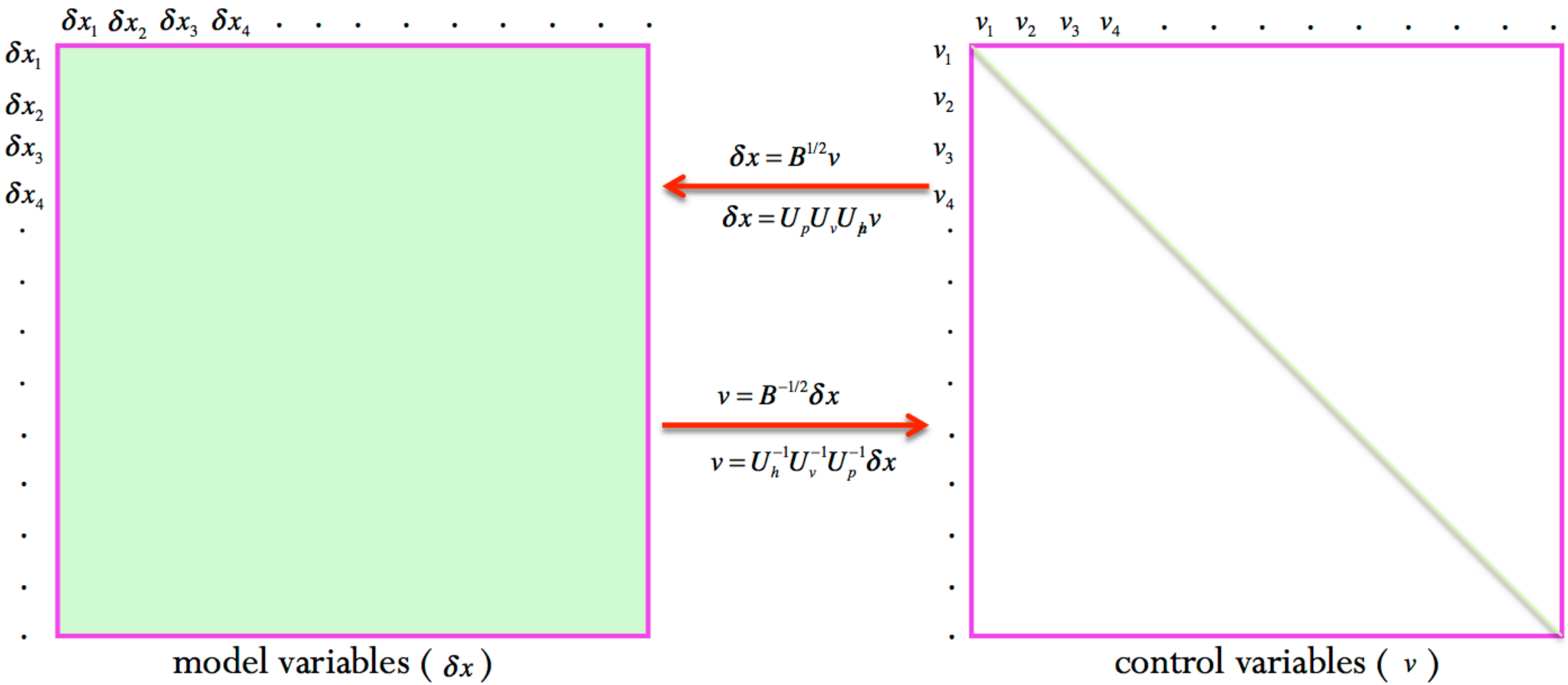
- WRFDA-3DVAR: Incremental formulation
- B matrix modeling within WRFDA
- B matrix estimation: GEN_BE package
- Visualize B effect: single observation test

How to estimate B?

- By definition, $\mathbf{B} = \langle(\mathbf{x} - \mathbf{x}^t), (\mathbf{x} - \mathbf{x}^t)^T\rangle$
  – x^t is the ground truth, which does not exist!

- Alternatively, we can replace $x\text{-}x^t$ by
  – $x^{t1} - x^{t2}$, i.e., NMC method, use the difference of forecasts (for t1 and t2 forecast range) valid at the same time
  – Or ensemble perturbations $x^{ensemble} - \langle x^{ensemble}\rangle$

GEN_BE package: basically performs inverse transform of U



GEN_BE is to compute horizontal correlation length-scales,
Eigenvectors/eigenvalues of vertical covariances,
and balance regression coefficients between control variables
Using large enough sample dataset of forecast difference or ensemble

GEN_BE Stage0: forecast error samples

- Step 1 - (u,v) to horizontal divergence (D) and vorticity (ζ)

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

- Step 2 – Convert D and ζ to Ψ and χ

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

- Finally, the forecast errors are generated using NMC or ensemble method for
  - Ψ - Stream function
  - χ - Velocity potential
  - T - Temperature
  - q/q_s - Relative humidity
  - p_s - Surface pressure

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

  Stream function (ψ')

  Velocity potential (χ')

  Temperature (T')

  Relative humidity (rh')

  Surface pressure (p_s')

GEN_BE Stage2 & 2a

- Stage2: Computes regression coefficients (i.e., linear correlation coefficients) between two variables
- Stage2a: Obtain unbalanced fields by removing the balanced part of fields using inverse transform U_p^{-1}
- Balance transform U_p in matrix form

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

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 U_v^{-1}

GEN_BE Stage4

- Calculate correlation as a function of distance between points
- Fit correlation to a Gaussian function with a lengthscale

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

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

- This step is time consuming, there is some trick to speed up
  - Calculate lengthscale for each mode/variable simultaneously, which can be done in script level

GEN_BE: choice of bin_type

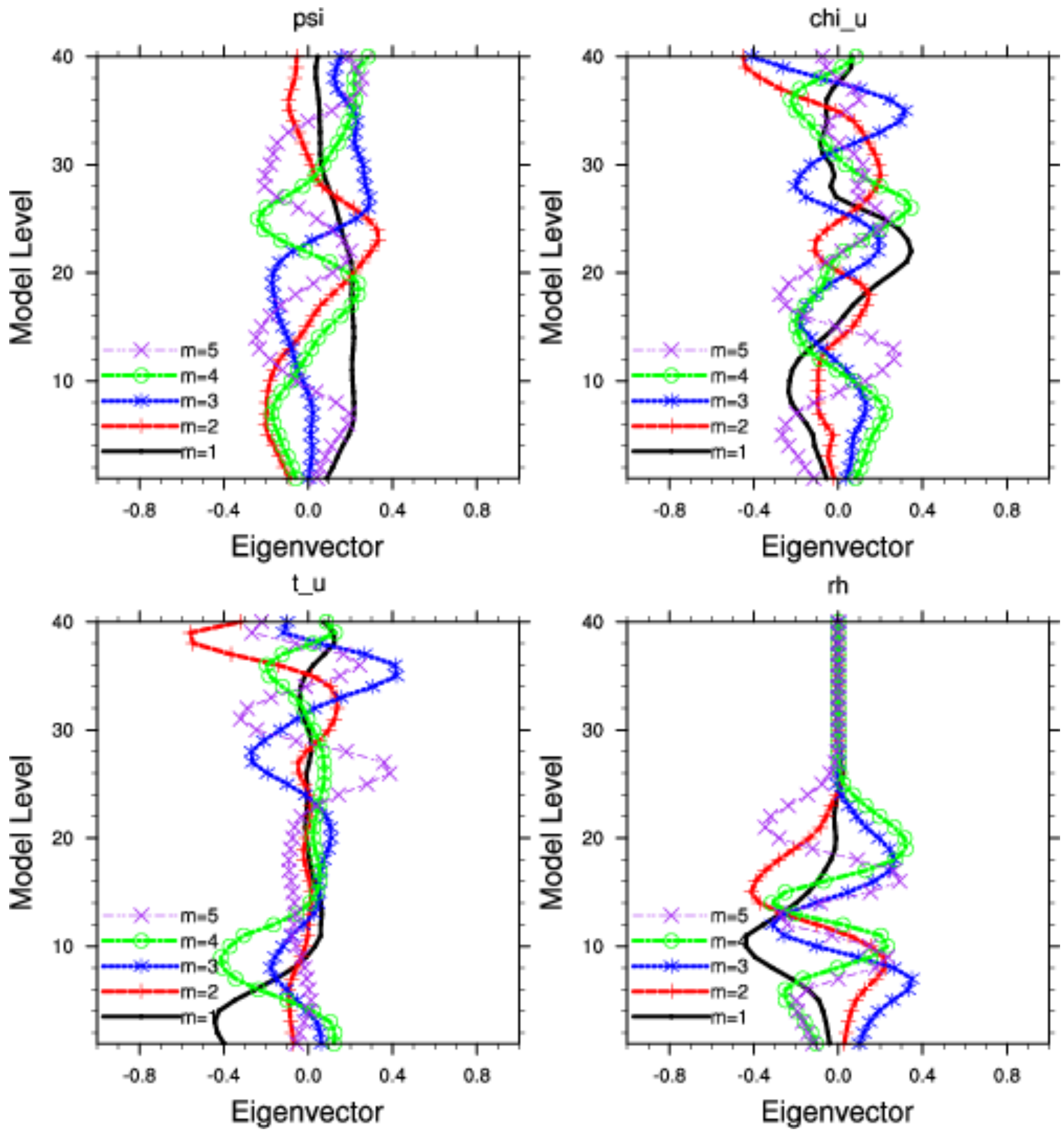
bin_type	Total number of bins (num_bins) and bin's description
0	num_bins= total number of grid points (no binning)
1	num_bins=nj * nk (each latitude is a bin)
2	num_bins= bin_width_lat * bin_width_hgt
3	num_bins=bin_width_lat * n_k (bin_width_lat is defined with lats.)
4	num_bins=bin_width_lat * n_k (bin_width_lat is defined with the number of points in south-north direction)
5	num_bins=n_k (bins with all horizontal points)
6	num_bins=1 (average over all the grid (3D) points)

- n_j – number of points in south-north direction
- n_k – number of points in vertical
- bin_type=5 default option, domain-averaged statistics

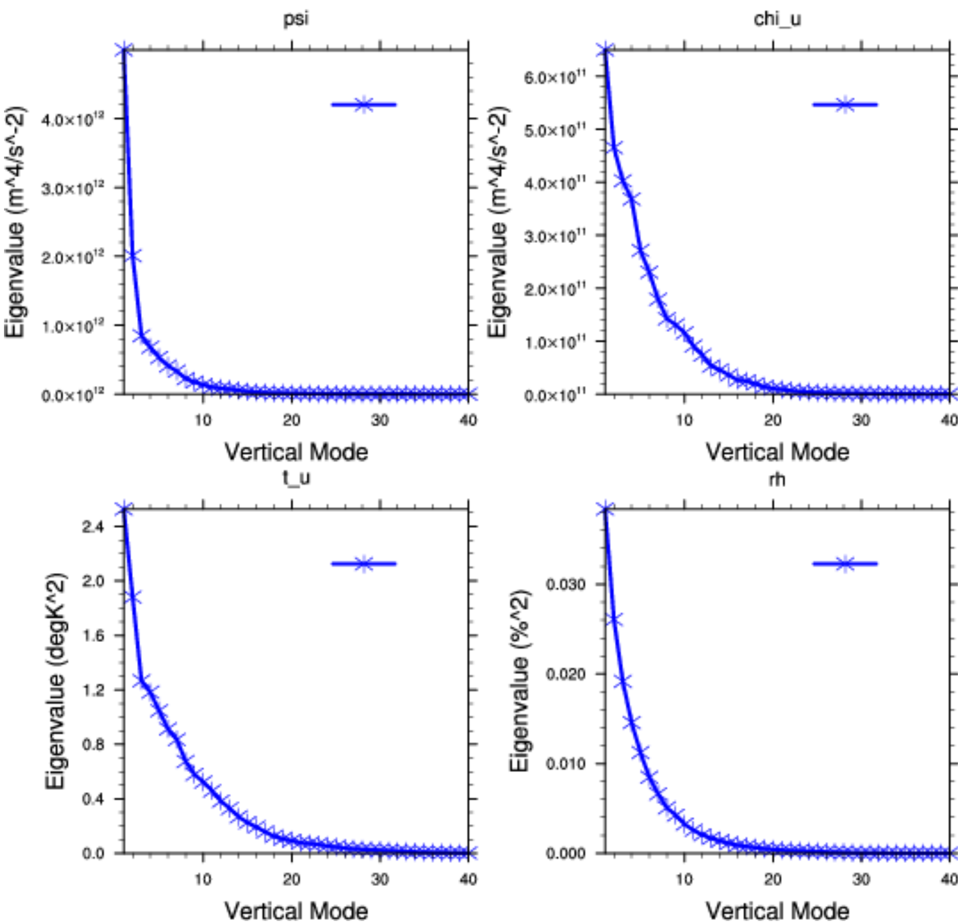
Outline

- WRFDA-3DVAR: Incremental formulation
- B matrix modeling within WRFDA
- B matrix estimation: GEN_BE package
- Visualize B effect: single observation test

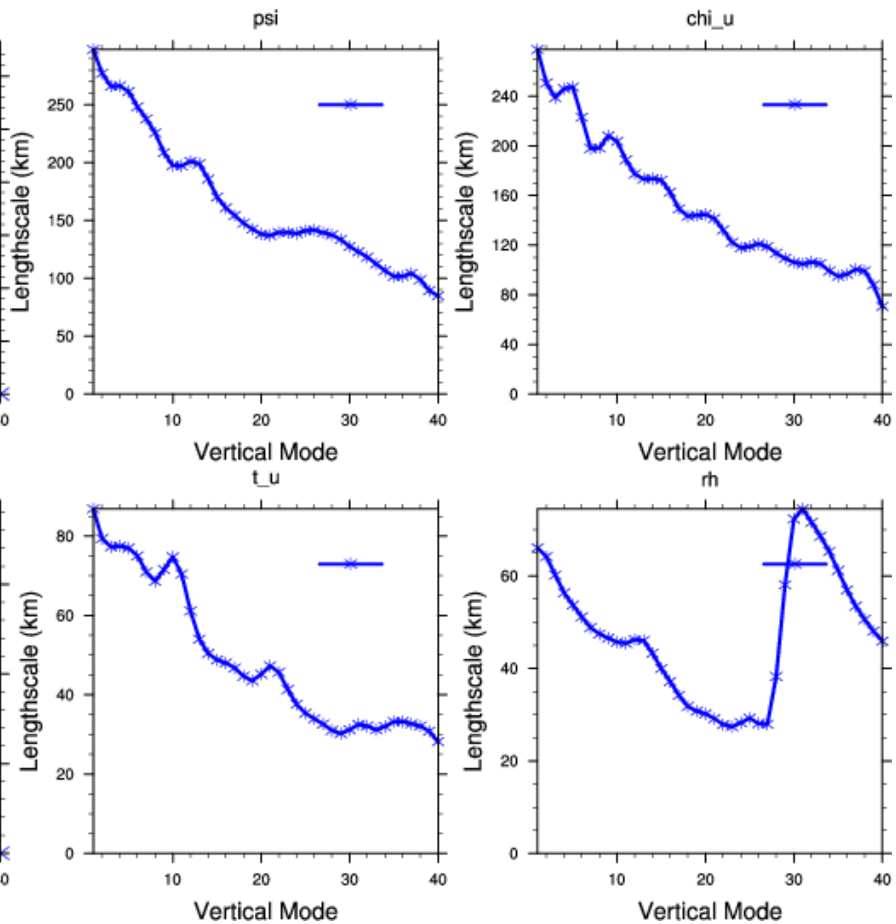
LEADING (FIRST 5) EIGENVECTORS



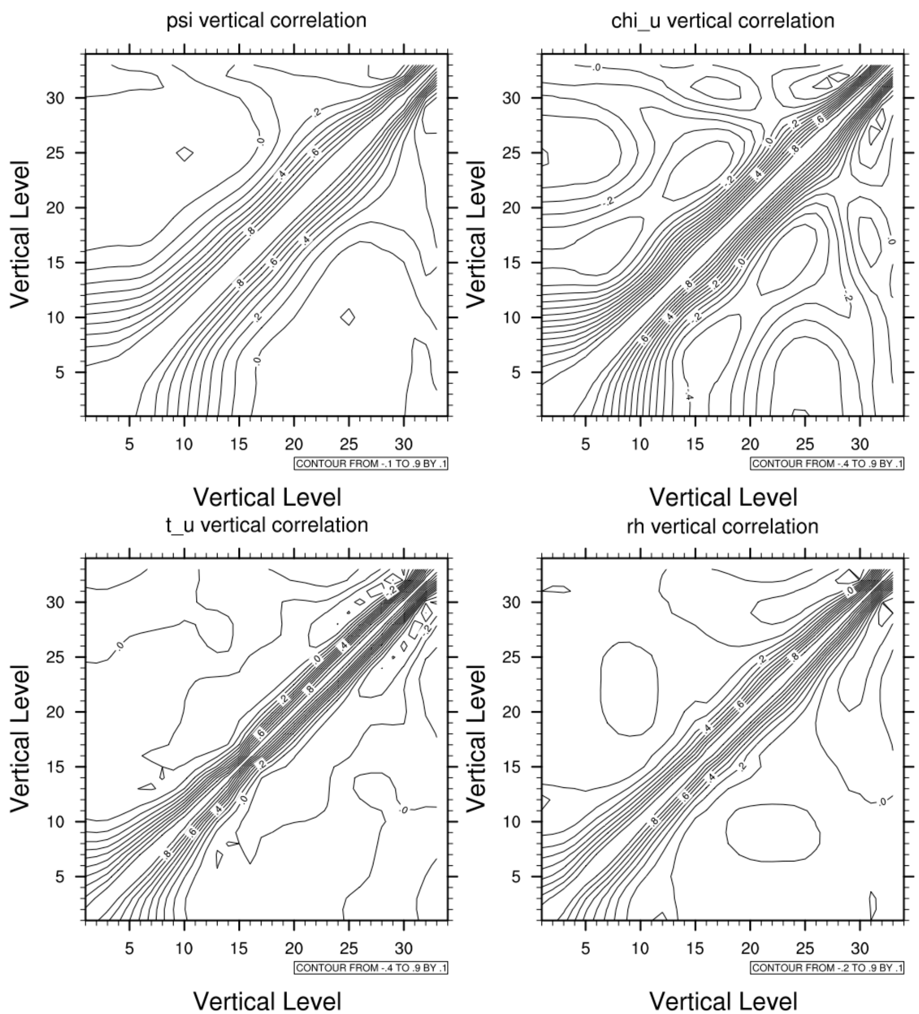
Eigenvalues



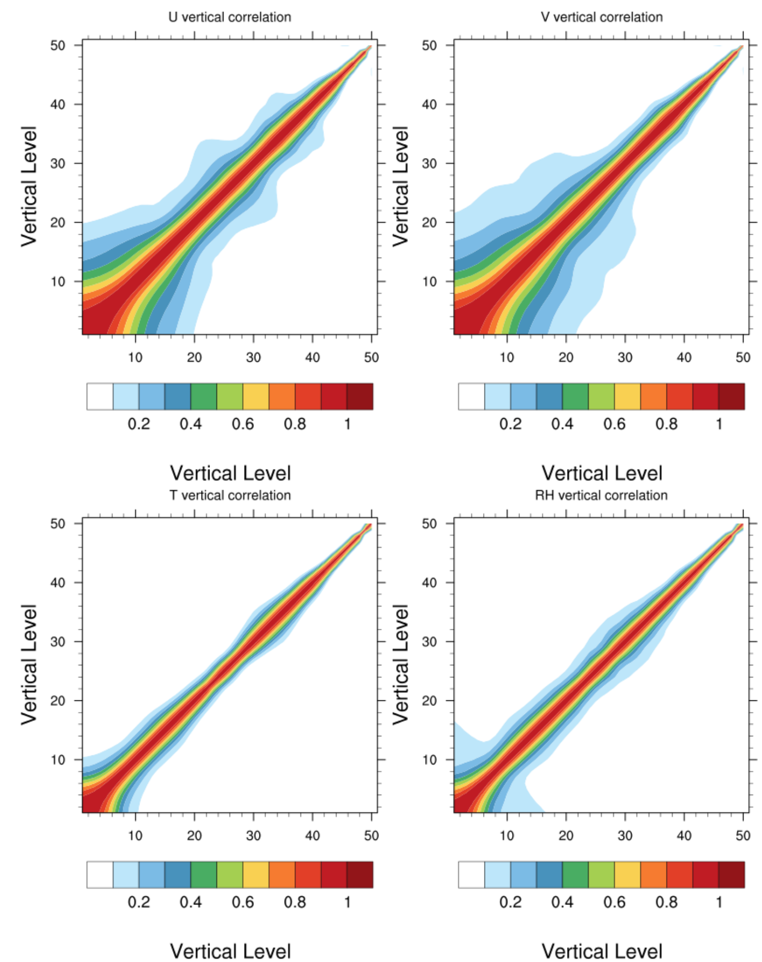
Horizontal lengthscales



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

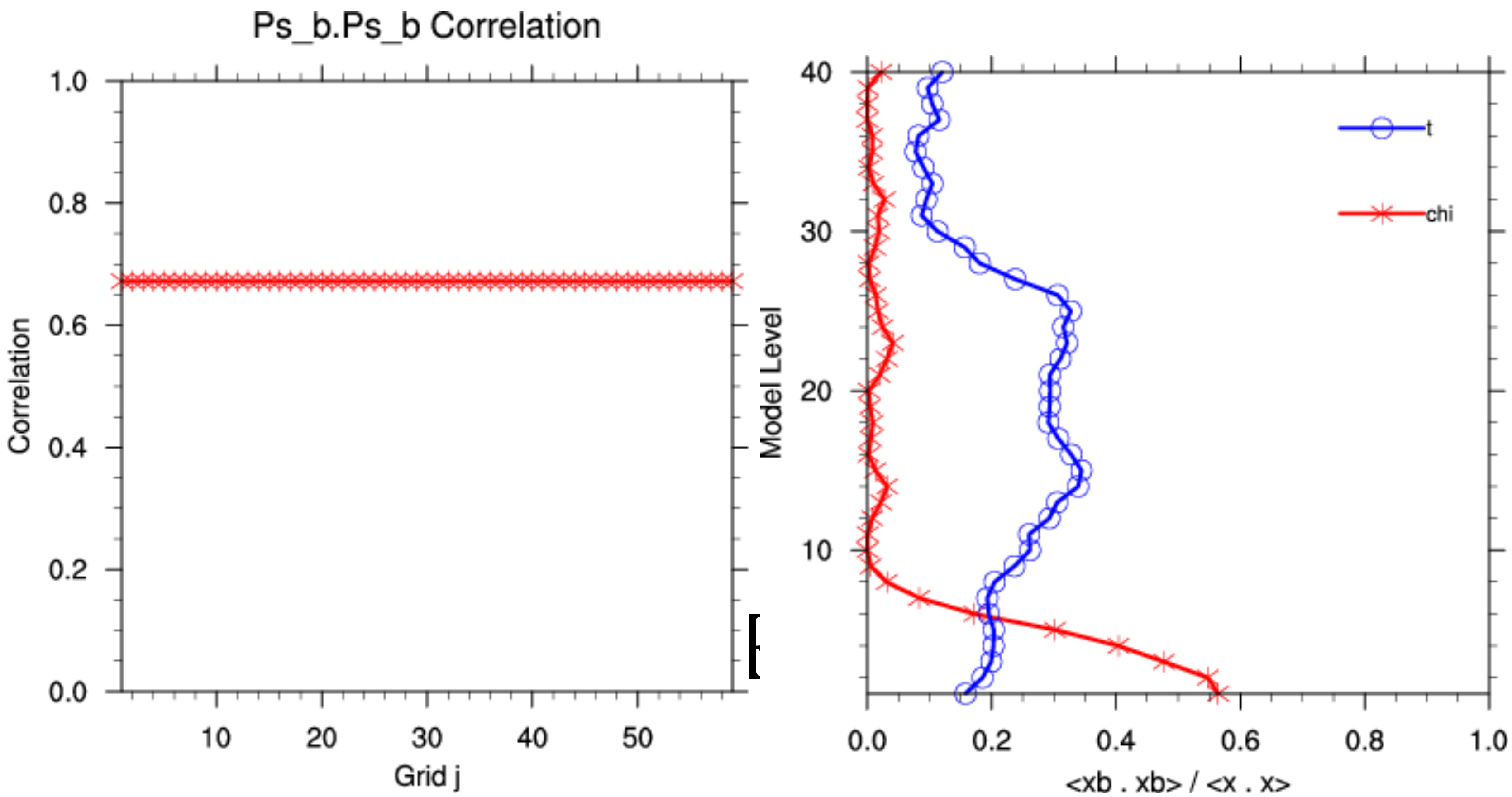


cv_options=5

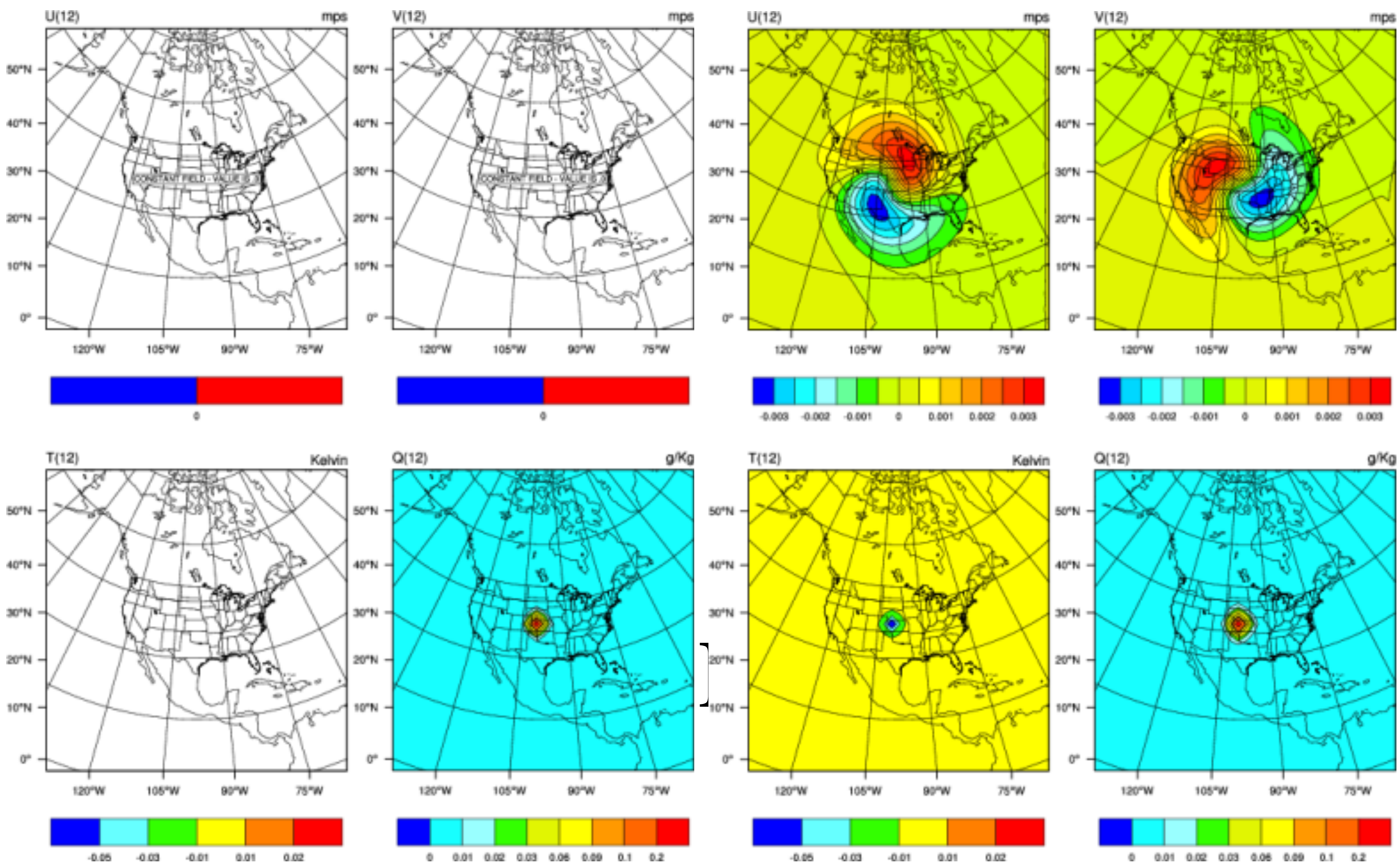


cv_options=7

Correlation coefficients w.r.t. stream function



SINGLE OBSERVATION TEST: Q



cv_options=5 cv_options=6

BE Tuning via namelist parameters

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

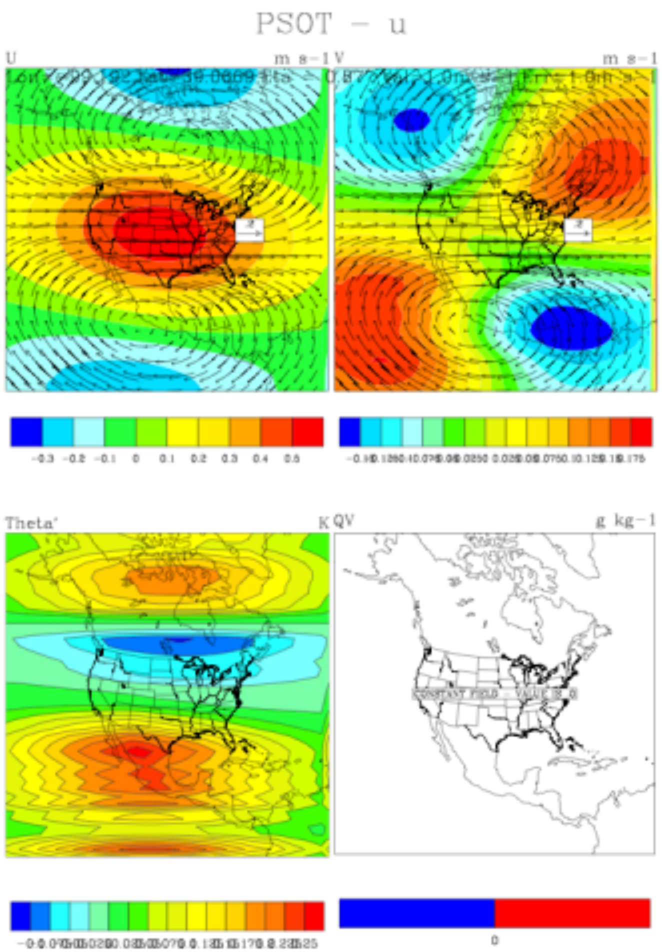
  LEN_SCALING1 - 5 (Length scaling parameters)

  VAR_SCALING1 - 5 (Variance scaling parameters)

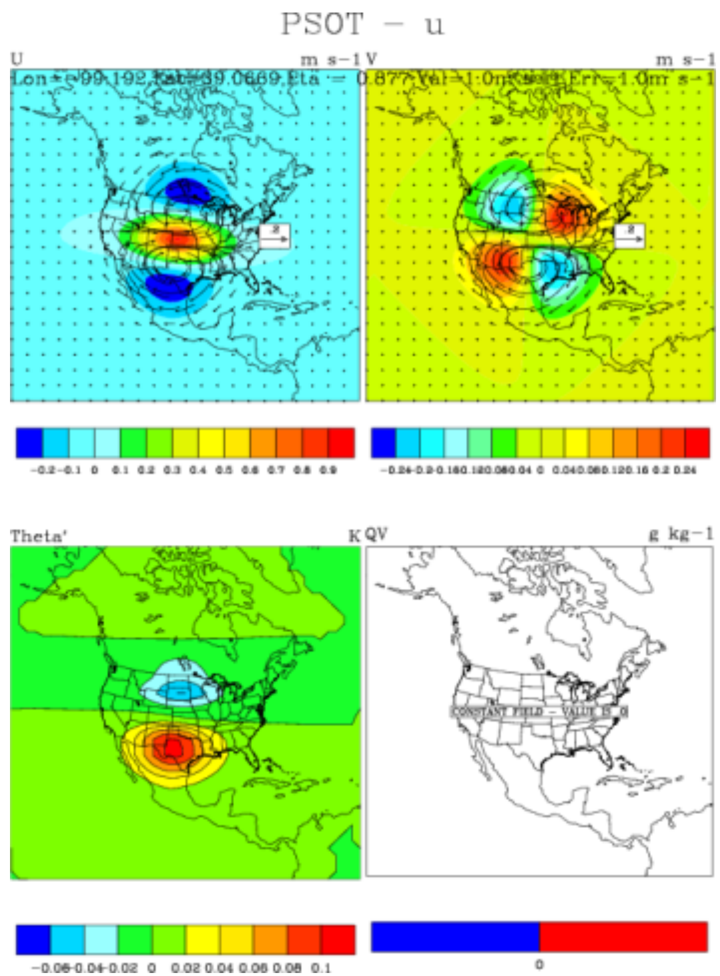
- Vertical component of BE can be tuned with the following five namelist parameters

  MAX_VERT_VAR1 - 5 (Vertical variance parameters)

BE TUNING (LENGTH-SCALE)



no tuning



tuning (len_scaling1 & 2 =0.25)