



WRFDA Background Error (Modeling and Estimation)

Syed RH Rizvi

National Center For Atmospheric Research
NCAR/MMM, Boulder, CO-80307, USA

August 1–August 3, 2016, Boulder, CO



Talk Overview

- Background Error (BE) and its role in DA?
- Modeling of BE
- Estimation of BE (“gen_be” utility)
- Single observation test and tuning of BE
- Impact of BE on analysis and NWP forecast
- Hands on practice session

Background Error (BE)

- If \mathbf{x} is the forecast of the analysis variable and \mathbf{x}^t is the corresponding true state, the BE is defined as the covariance of forecast minus truth ($\mathbf{x} - \mathbf{x}^t$).

$$\mathbf{BE} = \langle (\mathbf{x} - \mathbf{x}^t), (\mathbf{x} - \mathbf{x}^t)^T \rangle$$

- Thus, the BE covariance matrix (\mathbf{B}) describes the probability distribution function (PDF) of the forecast errors ($\mathbf{x} - \mathbf{x}^t$)

Role of \mathbf{B} in DA

- \mathbf{B} appears in the cost function and the analysis equation as,

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

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

- Thus, \mathbf{B} gives proper weight to the background term $(x - x^b)$ in defining the analysis cost function (J)
- Since \mathbf{B} is the last operator in the analysis equation, the analysis increment $(x^a - x^b)$ lies in the subspace of \mathbf{B}
- \mathbf{B} spreads information, both vertically and horizontally with proper weights to observation (y^o) and the background (x^b)

Role of BE in DA

- **B** spreads information between variables and imposes balance across different analysis variables. Thus, a pressure or the temperature observation has the ability to modify the wind analysis and vice-versa
- **B** provides a means by which observations can act in synergy, meaning **B** allows observations to reinforce each other in a way that improves the analysis to a degree that is greater than their individual contributions
- **B** is used for preconditioning the analysis equation

Modeling of BE

Why ?

- **B** is a square, symmetric and positive definite matrix ($x^T \mathbf{B} x > 0$ for all non-zero vectors x) with dimension equal to the number of the analysis variables
- Thus, typically the size of **B** is of the order of $10^7 \times 10^7$ and so, it is not possible to either store or compute its inverse

How ?

- The size of **B** is reduced by designing the actual analysis control variables in such a way that the cross covariance between these variables are minimum (zero)
- Thus assuming all the off-diagonal elements as zero, the size of **B** is typically reduced to the order of 10^7

Modeling of BE

- Let us define a control variable transform (CVT),

$$\delta x = B^{1/2} v$$

or,

$$\delta x = U v$$

Where,

$$\delta x = x - x^b$$

$$U = B^{1/2} \quad \text{and} \quad v v^T = I$$

- Thus $\mathbf{B} = \mathbf{U} \mathbf{U}^T$, and so the modeling of back ground error amounts to approximating the control variable transform (\mathbf{U})
- \mathbf{U} is approximated with a sequence of three linear transforms

$$U = U_p U_v U_h$$

- Thus,

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

Control variable transform (CVT)

$$\mathbf{U} = \mathbf{U}_p \mathbf{U}_v \mathbf{U}_h$$

- $\mathbf{U}_h \longrightarrow$ Horizontal transform. It is applied via recursive filter (Hayden and Purser(1995))
- $\mathbf{U}_v \longrightarrow$ Vertical transform. It is applied through empirical orthogonal functions (EOFs). The EOFs are the eigenvectors of the vertical error covariance matrix (\mathbf{E}). Thus,

$$\mathbf{U}_v = \mathbf{E} \mathbf{\Lambda}^{1/2}$$

Where, $\mathbf{\Lambda}^{1/2}$ is a diagonal matrix holding square root of the eigenvalues of vertical error covariance matrix (\mathbf{E})

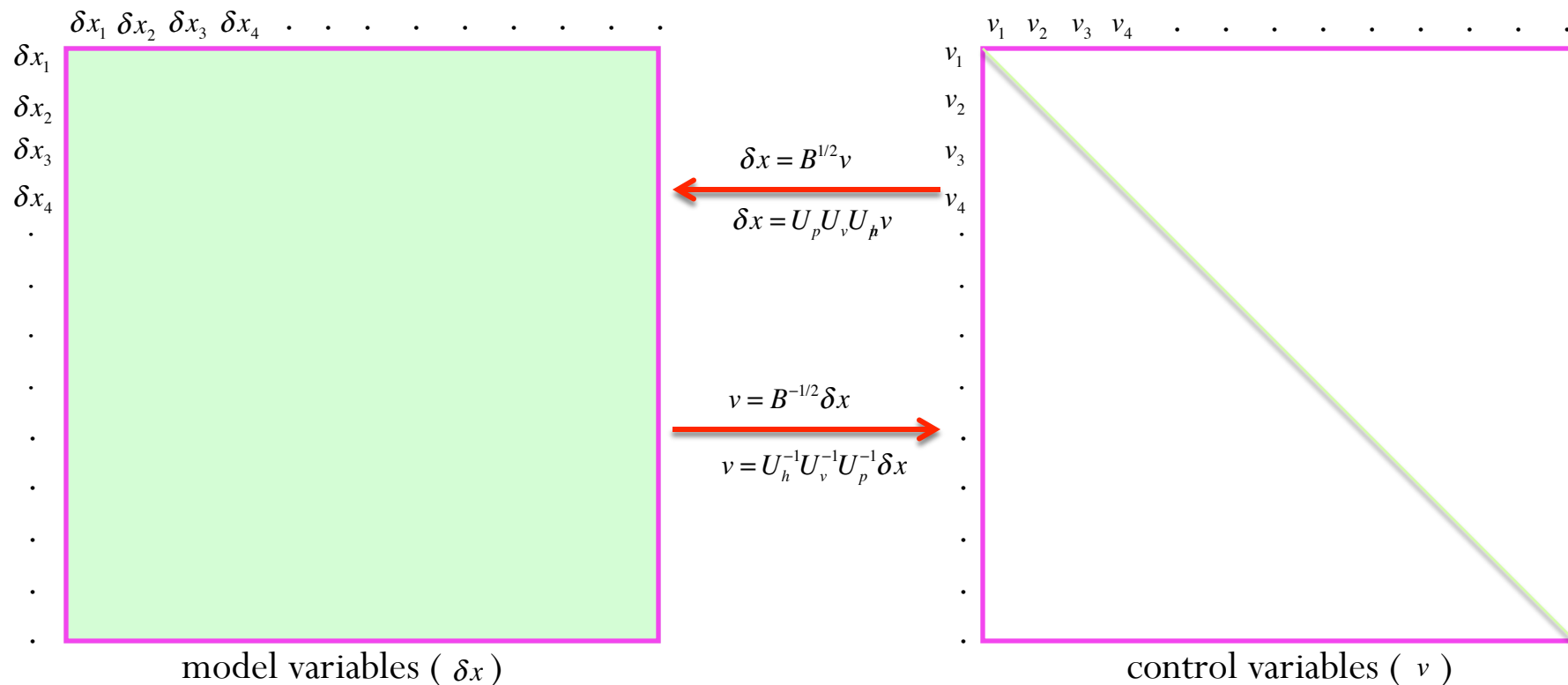
- $\mathbf{U}_p \longrightarrow$ Physical transform. It is applied via statistical balance

Modeling of BE

Thus, for modeling of background error, following is estimated

- Horizontal length-scale for \mathbf{U}_h transform
- Eigenvectors and eigenvalues for \mathbf{U}_v transform
- Regression coefficients for \mathbf{U}_p transform

How CVT ($U=U_p U_v U_h$) works ?



cost function (J): $J(v) = \frac{1}{2} v^T v + \frac{1}{2} (d - HUv)^T R^{-1} (d - HUv)$

gradient of J: $\nabla_v J = v + U^T R^{-1} (d - HUv)$

analysis: $x^a = x^b + Uv$

Choice of analysis variables

| cv_options | Analysis variables |
|------------|--|
| 3 | Ψ , unbalance X , unbalance t , pseudo rh and unbalance $\log(P_s)$ |
| 5 | Ψ , unbalance X , unbalance t , pseudo rh and unbalance P_s |
| 6 | Ψ and unbalance X , unbalance t , unbalance pseudo rh and unbalance P_s |
| 7 | u , v , t , P_s and pseudo rh |

Estimation of Background Error

- For simplicity, the background error distribution is assumed Gaussian
- Since the truth (\mathbf{x}^t) is not known, the forecast error ($\mathbf{x} - \mathbf{x}^t$) needs to be estimated
- There are two common methods for estimating ($\mathbf{x} - \mathbf{x}^t$)

a) NMC method: $(\mathbf{x} - \mathbf{x}^t) \approx (\mathbf{x}^{t1} - \mathbf{x}^{t2})$

The difference of forecasts (with t1 and t2 ICs) valid for the same time

b) Ensemble method: $(\mathbf{x} - \mathbf{x}^t) \approx (\mathbf{x}^{\text{ens}} - \langle \mathbf{x}^{\text{ens}} \rangle)$

Ensemble minus Ensemble mean

“gen_be” utility

“gen_be” utility estimates the different components of the BE

- It is designed to work both for NMC and Ensemble methods with “namelist” option:

BE_METHOD= “NMC” for NMC-method

BE_METHOD =“ENS” for Ensemble method

- It consists of five stages (0 – 4)

Stage0: (forecast error samples)

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

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

- Step 2 – Convert D and ζ to Ψ and χ

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

- Finally, the forecast errors ($\mathbf{x} - \mathbf{x}^t$) are generated for

Ψ - Stream function

χ - Velocity potential

T - Temperature

q - Relative humidity

p_s - Surface pressure

Stage1: (removes 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 (q')
 - Surface pressure (p_s')

Stage2: (Regression coefficients)

- Regression coefficient (α_{xy}) between two variables x and y is estimated as

$$\alpha_{xy} = \frac{\langle x, y \rangle}{\langle x, x \rangle}$$

Where,

$\langle x, y \rangle$ is the covariance between x and y

$\langle x, x \rangle$ is the variance of x

Stage2a: (Input for U_p -transform)

The U_p transform is defined as,

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

Where,

\mathbf{I} - identity matrix, $\mathbf{0}$ – zero matrix and \mathbf{M} , \mathbf{N} , \mathbf{Q} are respectively the regression coefficient matrices for (χ, ψ) , (t, ψ) , and (Ps_u, ψ)

Stage2a: (cv_options=6, the MBE)

The U_p transform is defined as,

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

Where,

P , R , S_1 , S_2 , S_3 and S_4 are respectively the regression coefficient matrices for (t, χ_u) , (Ps_u, χ_u) , (rh, ψ) , (rh, χ_u) , (rh, t_u) and (rh, Ps_u)

Stage3: (Input for U_v -transform)

For all 3-D analysis variables,

- Compute vertical error correlation matrix
- Compute eigenvectors (\mathbf{E}) and eigenvalues ($\mathbf{\Lambda}$) of the vertical error covariance matrix
- Perform $\mathbf{\Lambda}^{-1/2} \mathbf{E}^T$ to compute the amplitude of the corresponding EOFs

Stage4 (Input for U_h -transform)

- a) Computes covariance (z) of the coefficients of the EOF's in distance-wise bins
- b) Assuming the horizontal covariance has exponential decay (Gaussian function) as,

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

- c) Estimate the horizontal length-scale (s) of the covariance using linear curve fitting method as,

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

“gen_be” Bin’s Choice

| 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= $n_j * n_k$ (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
- **Remarks:** Default option is “bin_type=5”

Single observation test (PSOT)

Why?

Assimilation of a single observation helps in understanding the following aspects of the background error

- Its role and the structure
- Identify the “shortfalls”
- Broad guidelines for tuning

PSOT – Basic concept

Let y be the single observation for the k^{th} element of \mathbf{x}^b with standard observation error σ . Then the analysis equation

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

leads to

$$x_l^a - x_l^b = \frac{B_{lk}}{B_{kk} + \sigma^2}(y - x_k^b)$$

Thus,

- If $\sigma^2 \ll B_{kk} \Rightarrow x_k^a = y$
- If $\sigma^2 \gg B_{kk} \Rightarrow x_k^a = x_k^b$
- Thus, if BE is very large compared to observation error, analysis is closer to observation, otherwise it is closer to the first guess (FG) or the background
- A non-zero off-diagonal term B_{lk} of \mathbf{B} leads to non-zero analysis increment for the l^{th} element of \mathbf{x}^a

PSOT – Basic concept

- Set single observation (u, v, t, ps etc.) as,
 unit innovation, $[y^o - H(x^b)] = 1.0$
 unit observation error, $R = 1.0$

The analysis equation

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

gives,

$$x^a - x^b = B\delta$$

Where, δ is a constant delta vector

- Thus, analysis increments with single observation, displays the structure of the background error

How to activate PSOT?

PSOT utility may be activated by setting the following namelist parameters

`num_pseudo = 1`

`pseudo_var` = Variable name like u, t, p, etc.

`pseudo_x` = X-coordinate of the observation

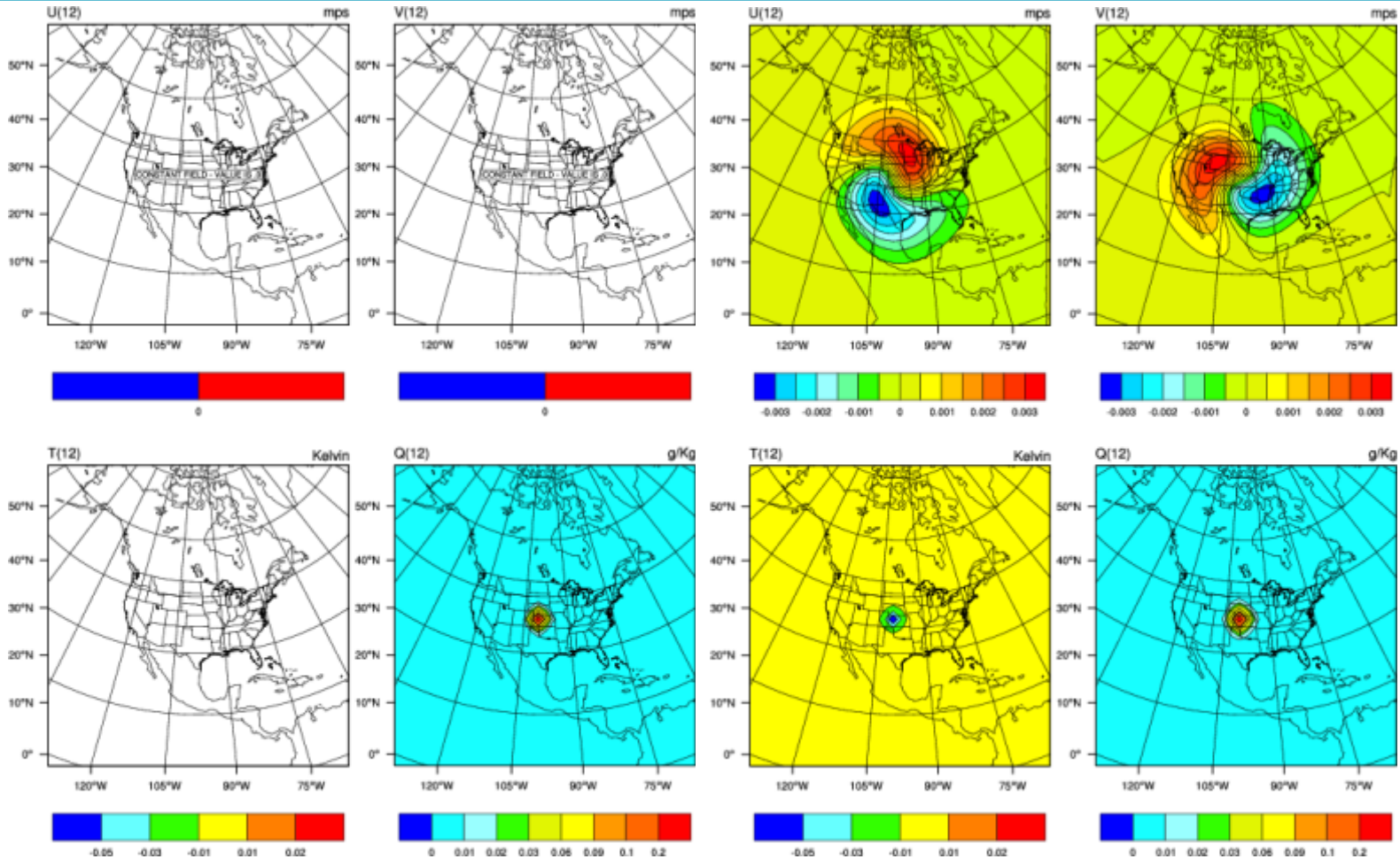
`pseudo_y` = Y-coordinate of the observation

`pseudo_z` = Z-coordinate of the observation

`pseudo_val` = Observation innovation, departure from FG

`pseudo_err` = Observation error

Analysis increments with PSOT-q



cv_options=5

cv_options=6

Tuning of BE

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

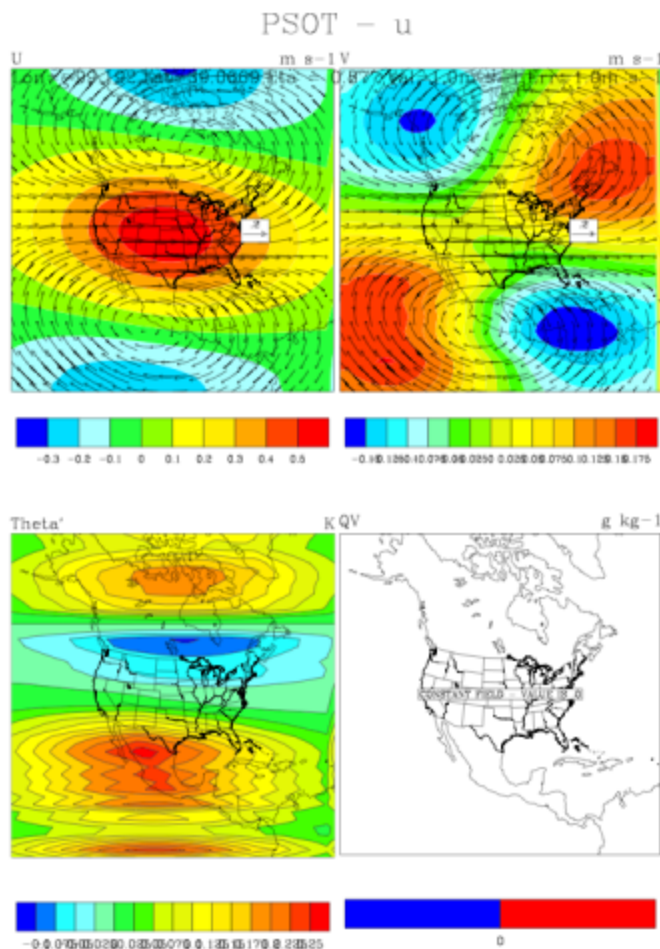
LEN_SCALING1 - 5 (Length scaling parameters)

VAR_SCALING1 - 5 (Variance scaling parameters)

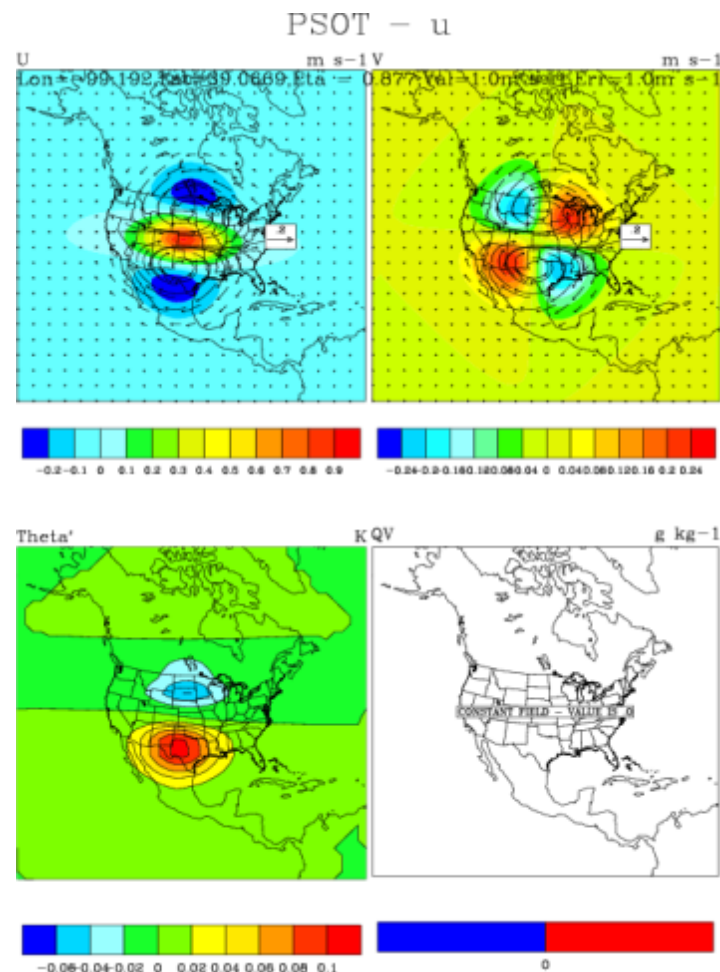
- Vertical component of BE can be tuned with the following five namelist parameters

MAX_VERT_VAR1 - 5 (Vertical variance parameters)

BE tuning (length-scale)



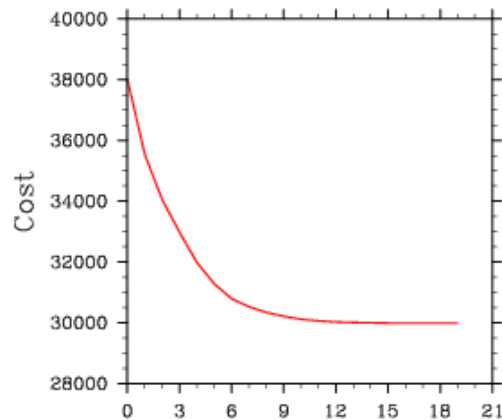
no tuning



tuning (len_scaling1 & 2 = 0.25)

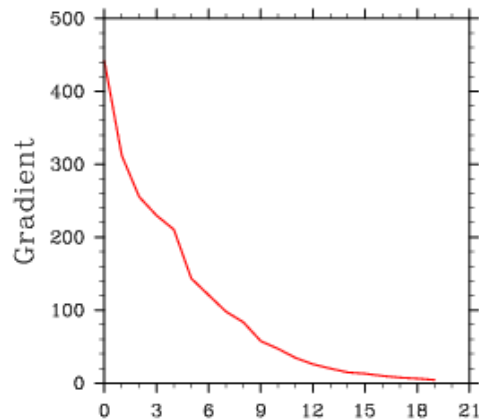
Impact of BE on Minimization

Cost function minimization for CONUS 200 Km domain



Iterations

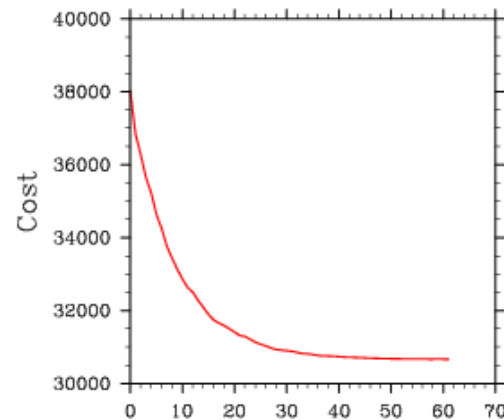
Gradient function for CONUS 200 Km domain



Iterations

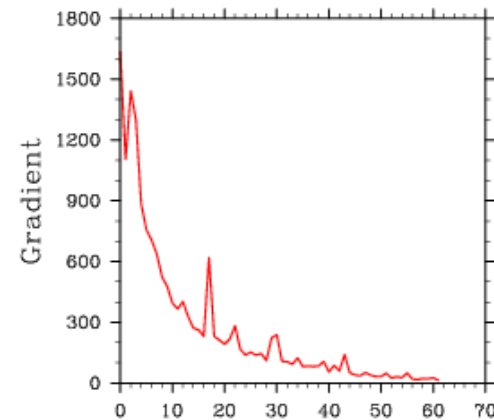
good BE

Cost function minimization for CONUS 200 Km domain



Iterations

Gradient function for CONUS 200 Km domain

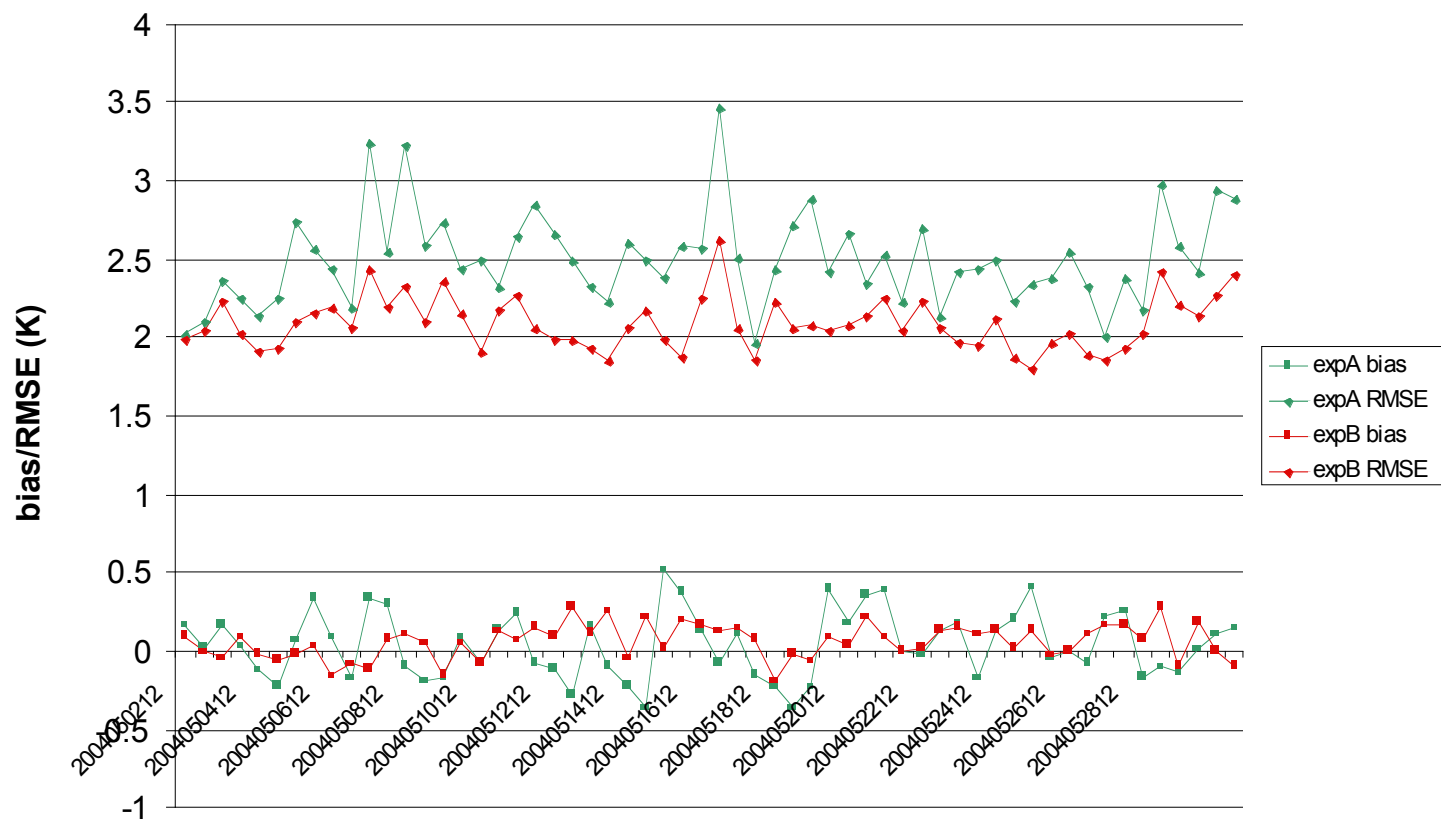


Iterations

bad BE

Impact of BE on Temp. Forecast

24 hr f/c bias/RMSE for Sound T



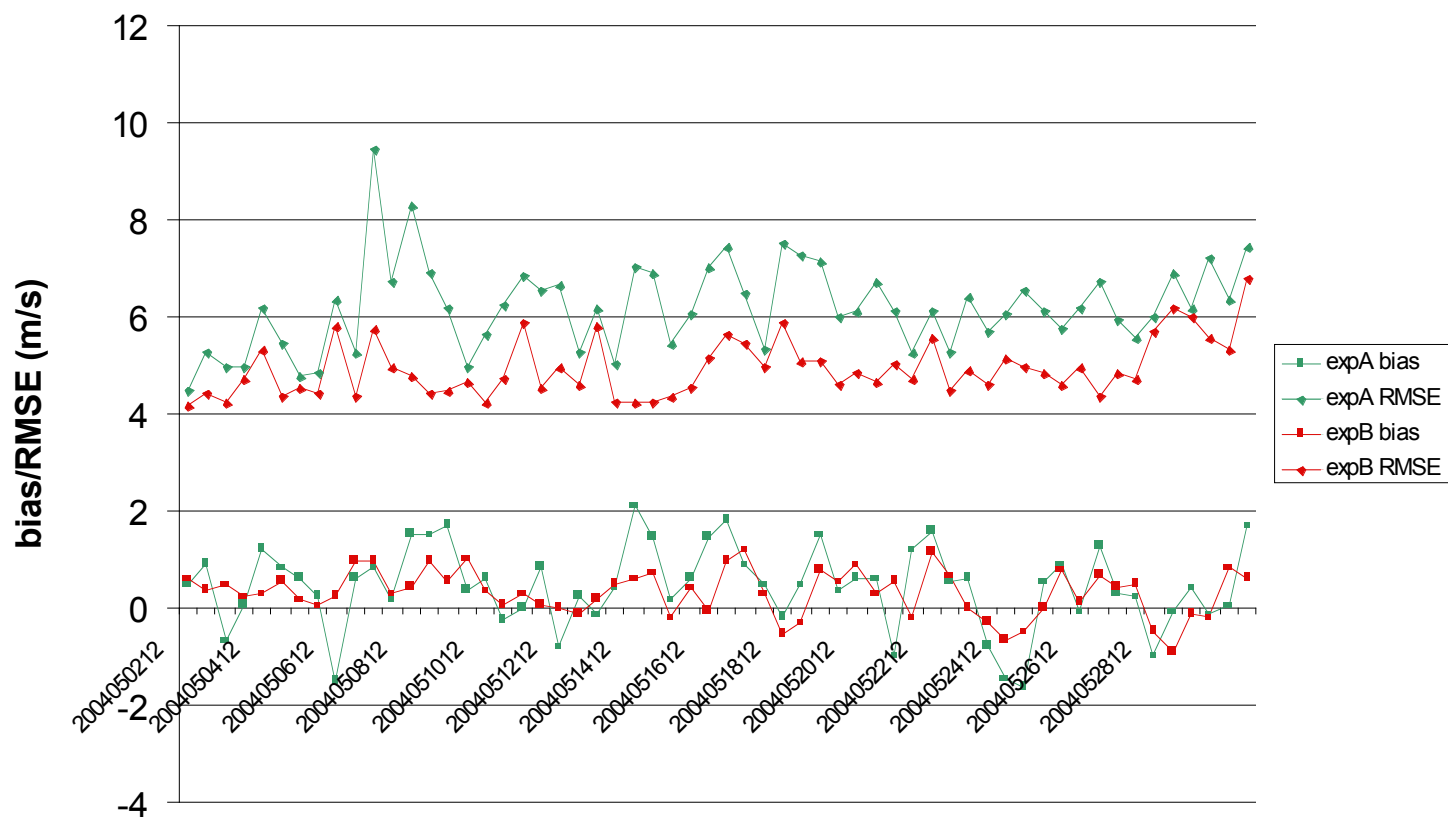
Valid time

expA: 6 Hr DA cycling with **bad BE**

expB: 6 Hr DA cycling with **good BE**

Impact of BE on U-comp Forecast

24 hr f/c bias/RMSE for Sound U-comp



Valid time **expA:** 6 Hr DA cycling with **bad BE**
expB: 6 Hr DA cycling with **good BE**



Advanced practice “gen_be”

- Compilation of “gen_be” utility
- Generation of BE statistics
- Familiarization with “gen_be” diagnostics
- Running PSOT to understand the structure of BE
- BE tuning

Generation of BE

“gen_be_wrapper.ksh” script for generating BE at 60 km “CONUS” domain with:

Grid Size : 90 X 60 X 41 (staggered grid points)

BE Method : NMC Method

Data Input : 12 and 24 hour forecasts (already run)

Basic environment variables to be set in the wrapper script:

WRFVAR_DIR (code location) ; FC_DIR (forecast location)

START_DATE (1st pert time) ; END_DATE (last pert time)

NUM_LEVELS (half sigma levels) ; RUN_DIR (run directory)

gen_be diagnostics

- “gen_be” creates various diagnostic files which may be used to display different components of BE
- Important diagnostics files are:

Eigen vectors: fort.174, fort.178, fort.182, fort.186

Eigen values: fort.175, fort.179, fort.183, fort.187

scale-length: fort.194, fort.195, fort.196, fort.197

Correlation between X_u & X_b (chi_u.chi.dat)

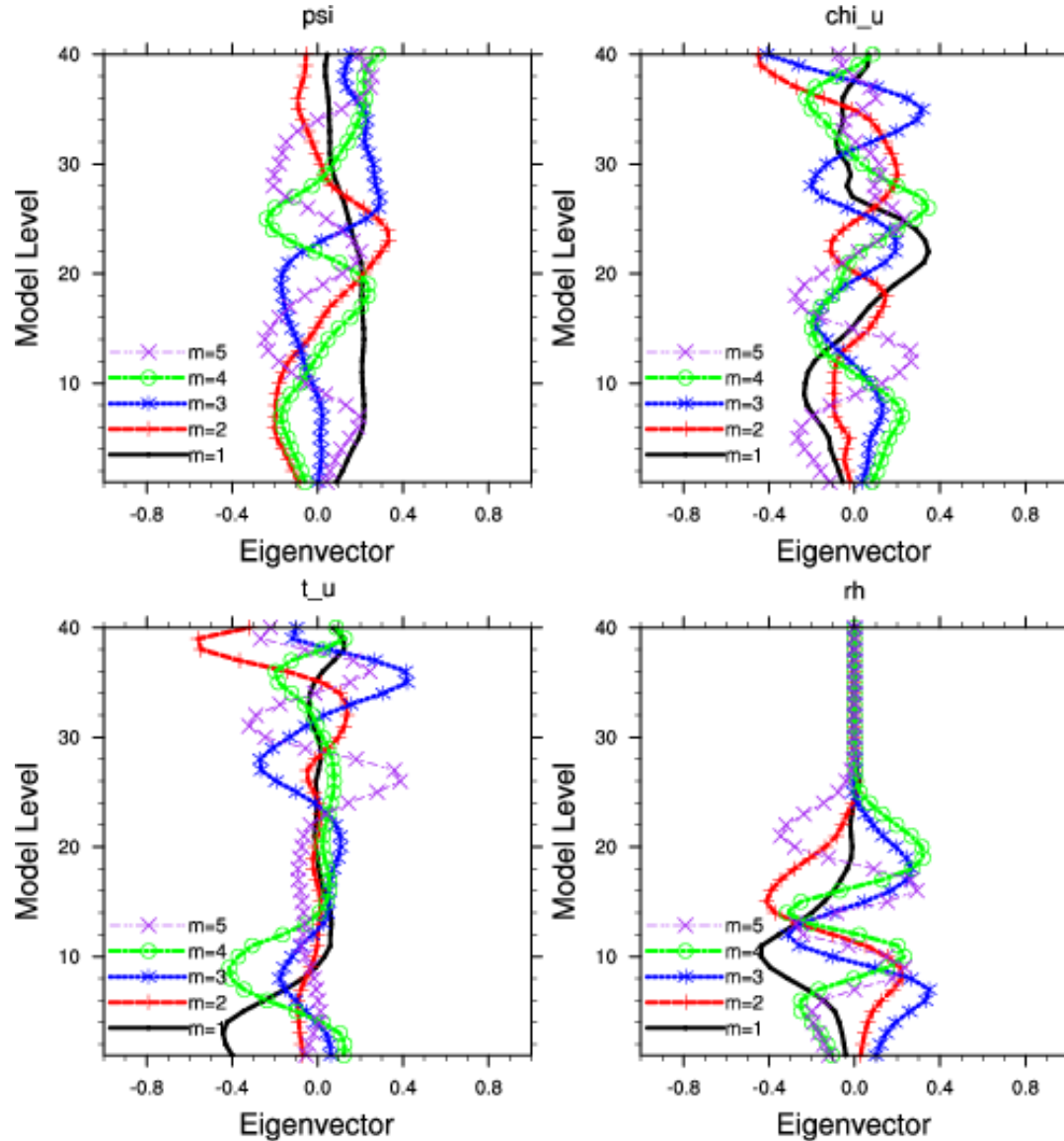
Correlation between T_u & T_b (T_u.T.dat)

Correlation between p_{s_u} & (ps_u.ps.dat)

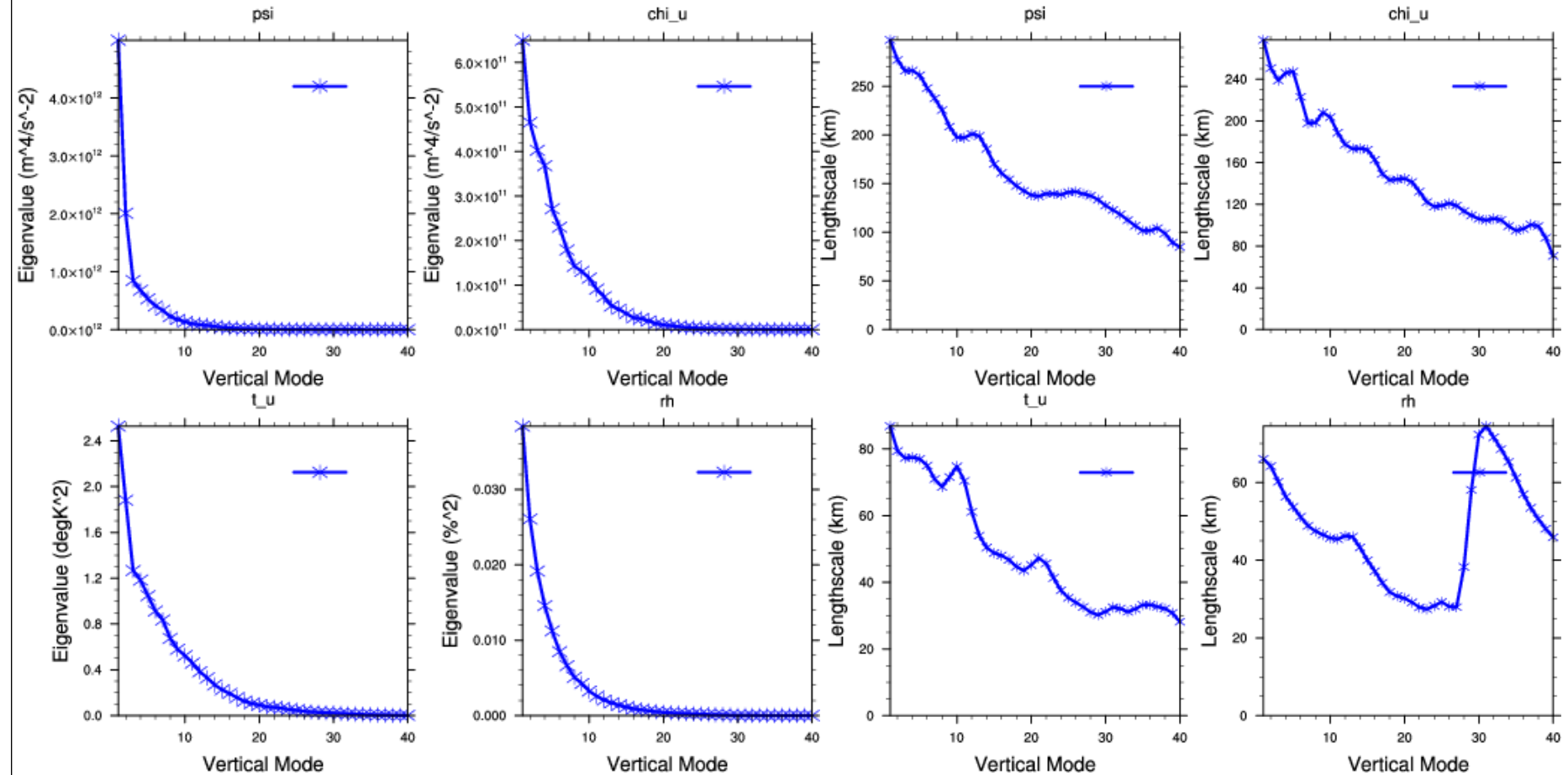
- Sample wrapper script for the display of BE diagnostics
“var/script/gen_be/gen_be_plot_wrapper.ksh”

Note: BE_DIR is set to “gen_be” RUN_DIR directory

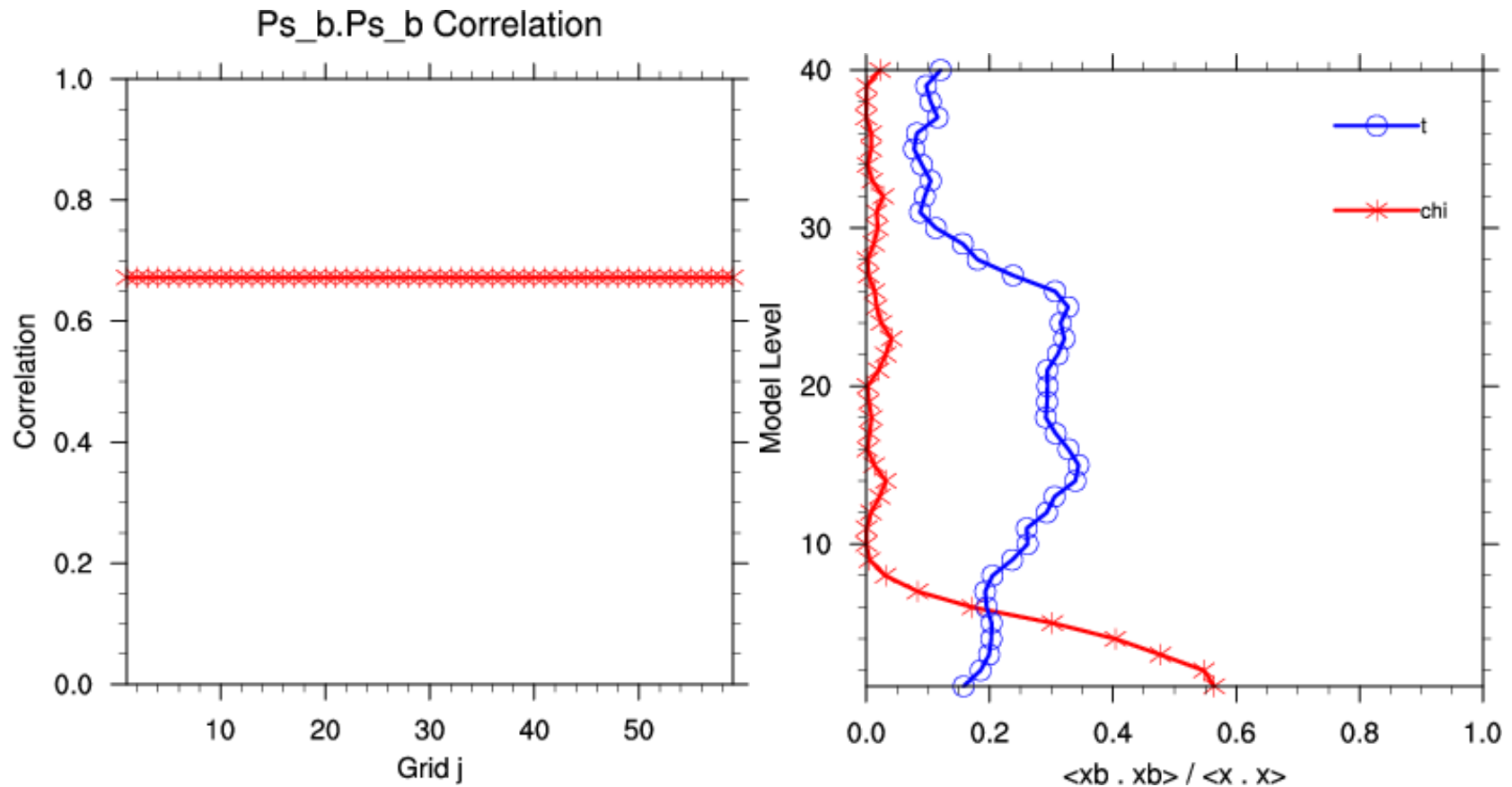
Leading (first 5) Eigenvectors



Eigenvalues and length-scales



Balance Correlations



How to run PSOT?

- Use following script from the WRFDA TOOLS package to build the PSOT wrapper script “var/scripts/wrappers/da_run_suite_wrapper_con200.ksh”
- Key parameters to set are:
 - Type of observation (pseudo_var)
 - Obs co-ordinates (pseudo_x, pseudo_y & pseudo_z)
 - Observation value (pseudo_val)
 - Observation error (pseudo_err)
- Display analysis increments to understand BE structure

Tuning of BE

- Understand the role of BE-tuning parameters through namelist options

LEN_SCALING1 - 5 (Length scale)

VAR_SCALING1 - 5 (Horizontal variance)

MAX_VERT_VAR1 - 5 (Vertical variance)