

WRFDA

Background Error Estimation

Syed RH Rizvi

National Center For Atmospheric Research

NCAR/ESSL/MMM, Boulder, CO-80307, USA

rizvi@ucar.edu

Talk overview

- What is Background Error (BE) ?
- Some properties of BE
- Role of BE in WRFDA
- Various components of BE
- Impact of BE on minimization and forecasts
- How to compute (“gen_be” utility)?
- Single Observation Test
- Upcoming new features
- Introduction to Practice Session

What is BE?

- The BE covariance matrix describes the probability distribution function (PDF) of forecast errors, assumed Gaussian
- BE is the covariance of (**forecast - truth**) in analysis control variable space

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

- Since **truth** (x^t) is not known, it needs to be estimated
- Common methods for estimating BE
 - **Innovation Method**
 - **NMC Method:** $(x - x^t) \approx (x^{t1} - x^{t2})$
(Forecast differences valid for the same time)
 - **Ensemble Method:** $(x - x^t) \approx (x^{\text{ens}} - \langle x^{\text{ens}} \rangle)$
= (Ensemble - Ensemble mean)

Some properties of **B**

- **B** matrix is square and symmetric. Thus, its eigenvalues are all real and eigenvectors are mutually orthogonal
- It is positive semi-definite. Thus, its eigenvalues are all non-negative. It is a very important property because without this the minimum of the cost function may not exist
- It consists of correlation (**C**) and variance (Σ) parts, $\mathbf{B} = \Sigma \mathbf{C} \Sigma$
- If **V** is an orthogonal matrix ($\mathbf{V}^T \mathbf{V} = \mathbf{I}$) transforming vector **X** to **U** ($\mathbf{U} = \mathbf{V} \mathbf{X}$), then the background error for **X** (**B**) and of **U** (**B^u**) will be related as $\mathbf{B}^u = \mathbf{V}^T \mathbf{B} \mathbf{V}$
- A special representation of **B** is the eigen-representation, where **B^u** is diagonalized. Eigenvectors of **B** form the columns of **V** and the eigenvalues of **B** are the diagonal elements of **B^u**

Role of BE

- B** spreads information, both vertically & horizontally with proper weights to observations and FG. This effect may be understood by introducing a single observation of one (**k**th) element of **x** in the analysis equation

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

In this case **H** or **H** is a row vector with all elements zero except the **k**th, which is = **1** and **y**^o = **y** ; **R** = **σ**² . Thus the analysis equation gives

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

Thus non-zero off-diagonal terms for **B** lead to analysis increment for **l**th element

- In data assimilation, this is not the only mechanism of spreading information. Observation operators (**H** & **H**) also do this job
- If **σ**² << **B**_{kk} ; **x**_k^a ≈ **y** and if **σ**² >> **B**_{kk} ; **x**_k^a ≈ **x**_k^b

Thus if BE is very large compared to observation error, the analysis is closer to the observation. Otherwise it is closer to FG

Role of BE

Contd.

- **B** matrix spreads information between variables and imposes balance
- Since **B** is the last operator in the analysis equation, the analysis increments lies in the subspace of **B**.
- **B** provides a means by which observations can act in synergy. **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.

How BE is represented in WRFDA?

- It is represented with a suitable choice of **U** as follows

$$\mathbf{B} = \mathbf{U}^T \mathbf{U} \quad \text{with} \quad \mathbf{U} = \mathbf{U}_p \mathbf{U}_v \mathbf{U}_h$$

U_h Horizontal Transform

U_v Vertical Transform

U_p Physical Transform

- Horizontal transformation (**U_h**) is via
 - Regional** ----- Recursive filters
 - Global** ----- Power spectrum
- Vertical transformation (**U_v**) is via EOF' s
- Physical transformation (**U_p**) depends upon the choice of the analysis control variable

How BE is represented?

Contd.

- Size of **B** is typically of the order of $10^7 \times 10^7$
- It is reduced by designing the **analysis control variables** in such a way that cross covariance between these variables are minimum
- Currently, **analysis control variables** for WRFDA are the amplitudes of EOF' s of

stream function (ψ)

Unbalanced part of velocity potential (χ_u)

Unbalanced part of temperature (T_u)

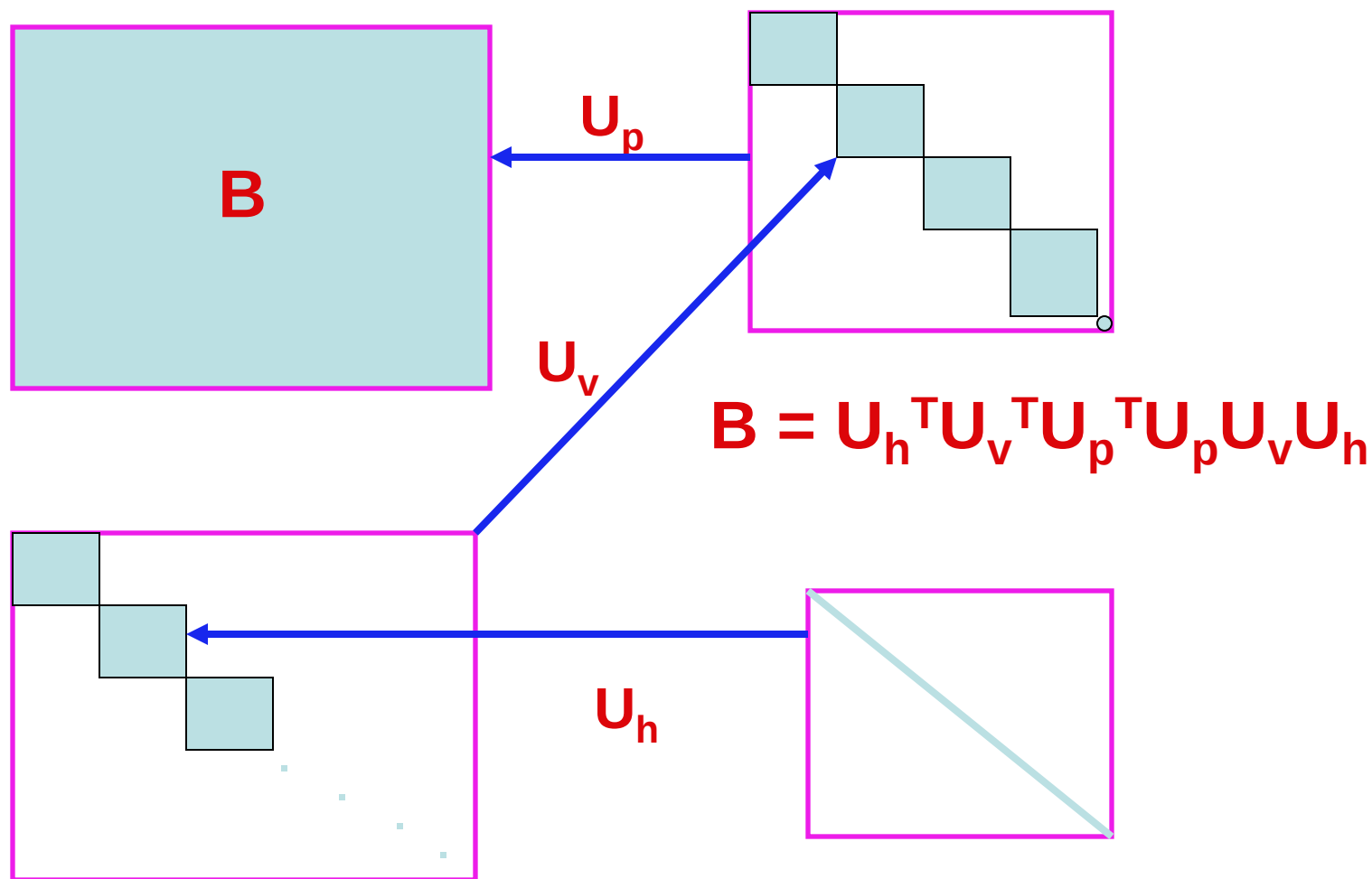
Relative Humidity (q)

Unbalanced part of surface pressure (p_{s_u})

- With this choice of **analysis control variables** off-diagonal elements of BE is very small and thus its size typically reduces to the order of 10^7

How BE is represented?

Contd.

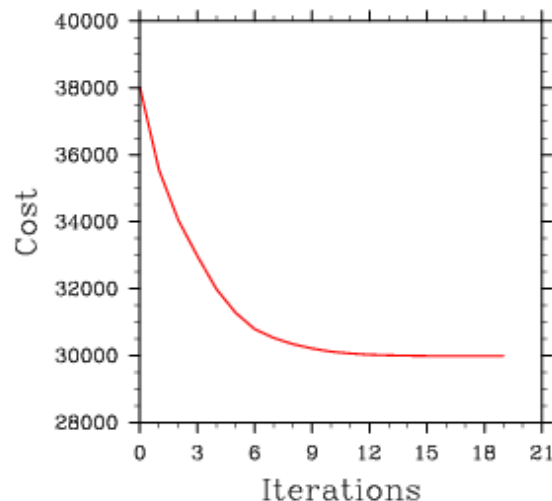


Components of BE

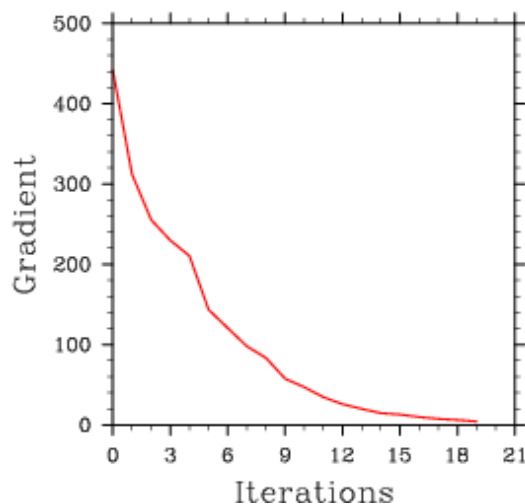
- Regression Coefficient for balanced part of Velocity potential, Temperature and Surface pressure
- Eigen vectors and Eigen values for stream function, unbalanced velocity potential, unbalanced temperature and moisture field
- Horizontal length-scales of control variables for regional option
- Power spectrum of control variables for global option

Impact of BE on Minimization

Cost function minimization for CONUS 200 Km domain

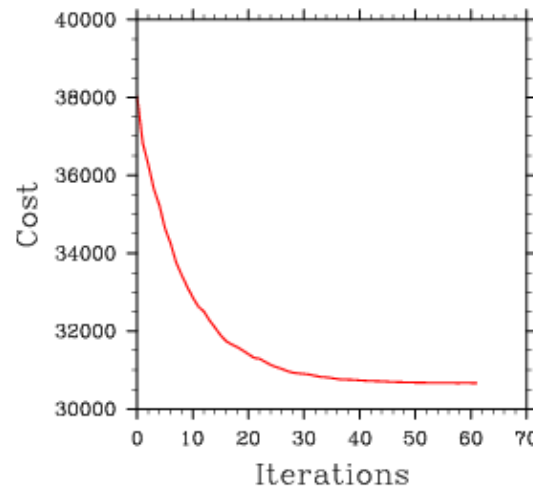


Gradient function for CONUS 200 Km domain

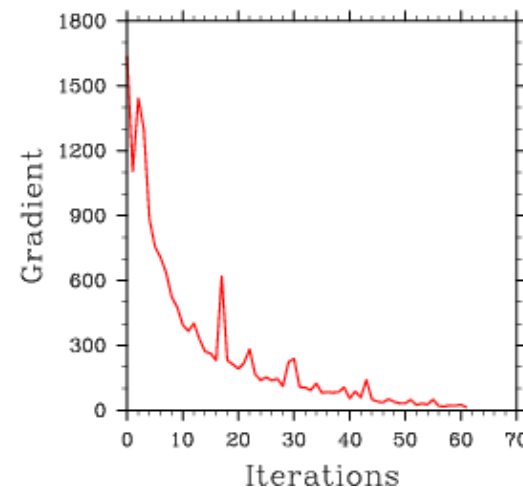


Good BE

Cost function minimization for CONUS 200 Km domain



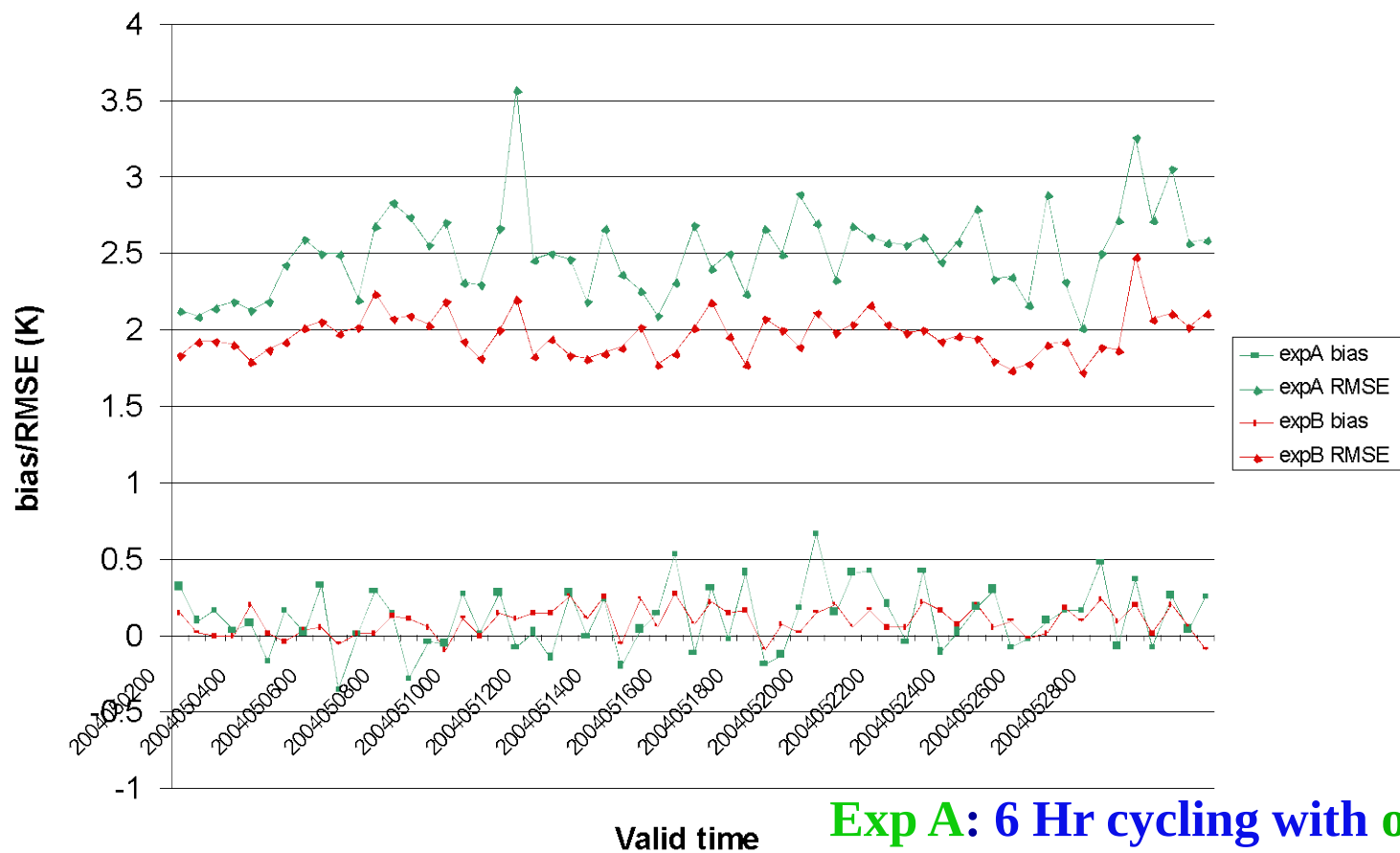
Gradient function for CONUS 200 Km domain



Bad BE

Impact of BE on Temperature forecast

12 hr f/c bias/RMSE for Sound T

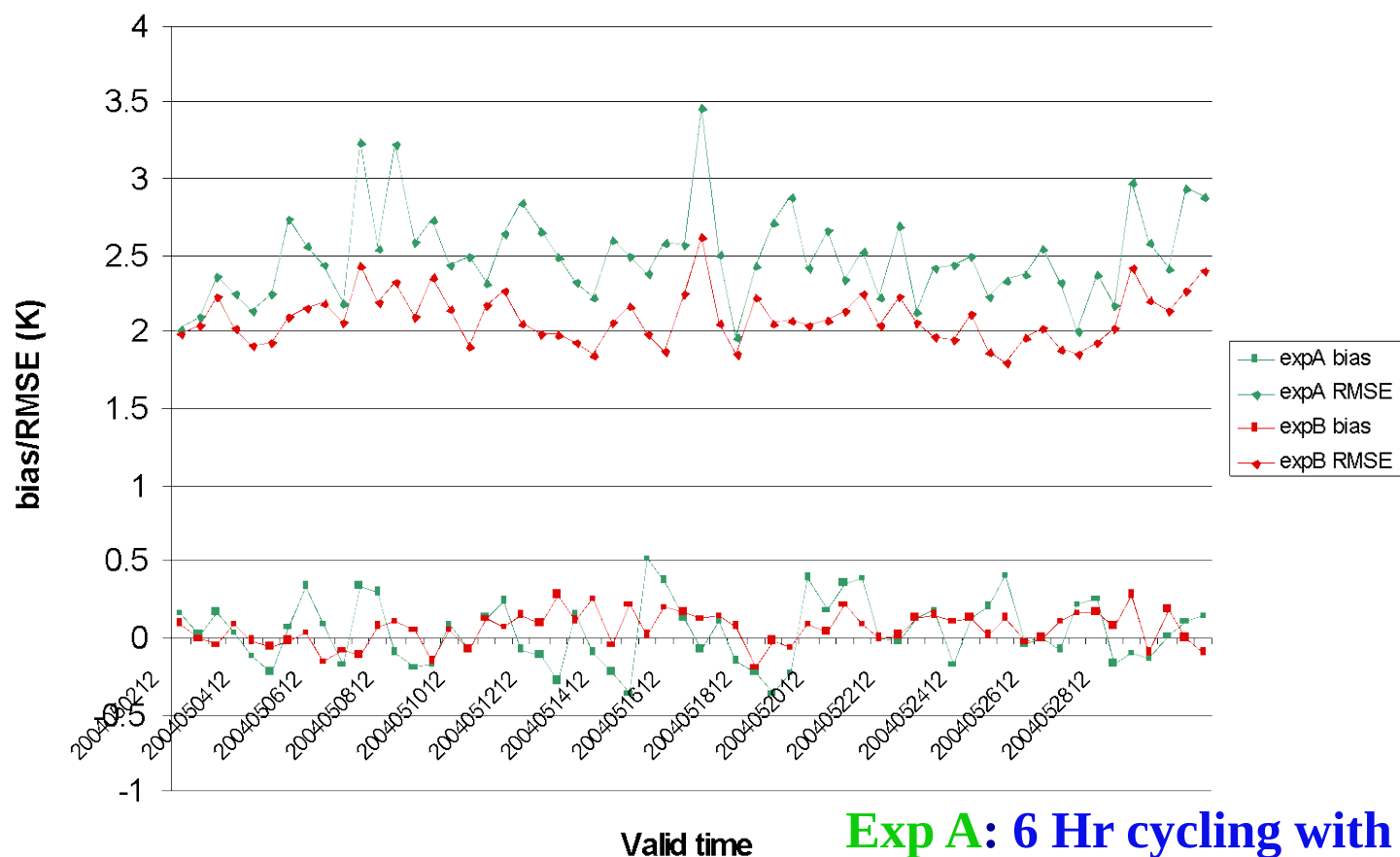


Exp A: 6 Hr cycling with old BE

Exp B: 6 Hr cycling with new BE

Impact of BE on Temperature forecast

24 hr f/c bias/RMSE for Sound T

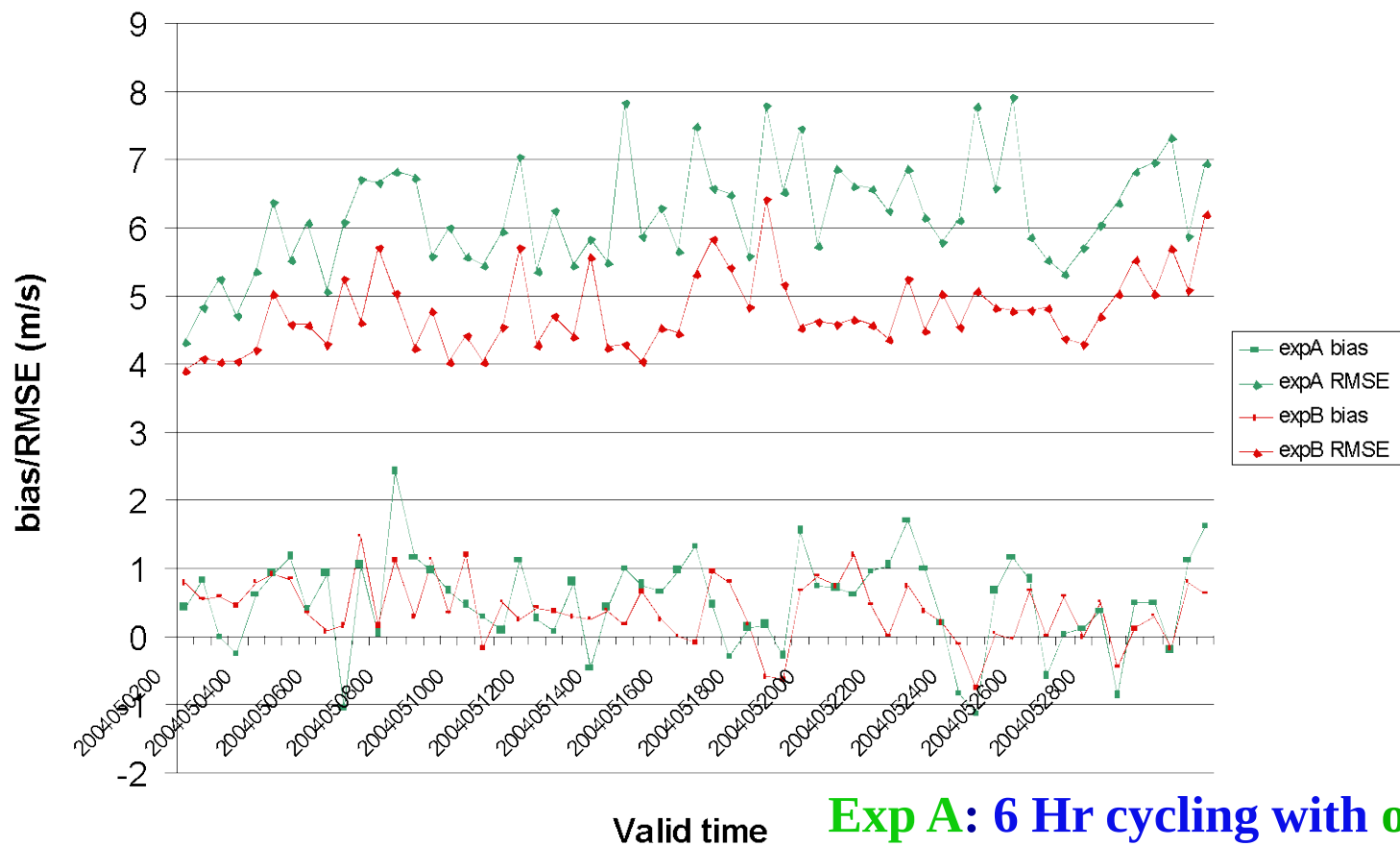


Exp A: 6 Hr cycling with old BE

Exp B: 6 Hr cycling with new BE

Impact of BE on Wind (U Comp.) forecast

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

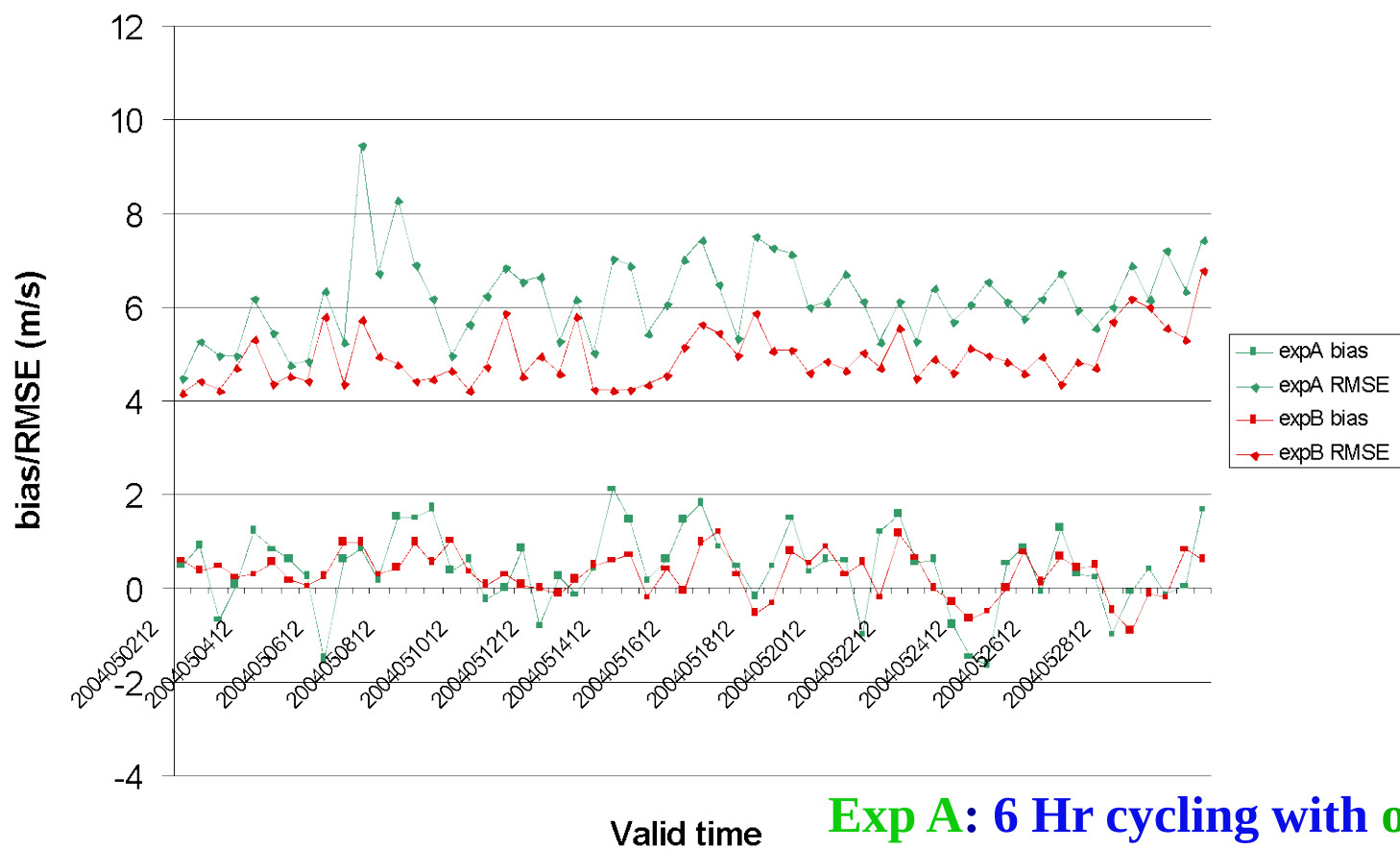


Exp A: 6 Hr cycling with old BE

Exp B: 6 Hr cycling with new BE

Impact of BE on Wind (U Comp.) forecast

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



Exp A: 6 Hr cycling with old BE

Exp B: 6 Hr cycling with new BE

WRFDA “gen_be” utility:

- It resides in WRFDA under the “var” directory
- Computes various components of BE statistics
- Designed both for **NMC** and **Ensemble** methods (“BE_METHOD”)
- It consists of **five** stages
- Basic goal is to estimate the error covariance in analysis control variable space (Coefficients of the EOF’ s for ψ , χ_u , T_u , rh and p_{s_u}) with input from model space (U , V , T , q & P_s)

“gen_be” - Stage0

- Computes (ψ , χ) from (u,v)
- Forms desired differences for the following fields

ψ - Stream Function
 χ - Velocity potential
 T - Temperature
 q - Relative Humidity
 p_s - Surface Pressure

“gen_be” - Stage1

- Reads “gen_be_stage1” namelist
- Fixes “**bins**” for computing BE statistics
- Computes “**mean**” of the differences formed in **stage0**
- Removes respective “**mean**” and forms **perturbations** for

Stream Function	(ψ')
Velocity potential	(χ')
Temperature	(T')
Relative Humidity	(q')
Surface Pressure	(p_s')

“gen_be” - Stage2 & 2a

- Reads “gen_be_stage2” namelist
- Reads field written in stage1 and computes covariance of the respective fields
- Computes regression coefficient & **balanced** part of χ , T & p_s

$$\chi_b = \mathbf{C} \psi'$$

$$T_b(\mathbf{k}) = \sum_l \mathbf{G}(\mathbf{k}, l) \psi'(l)$$

$$p_{s_b} = \sum_l \mathbf{W}(\mathbf{k}) \psi'(k)$$

- Computes unbalanced part

$$\chi_u' = \chi' - \chi_b$$

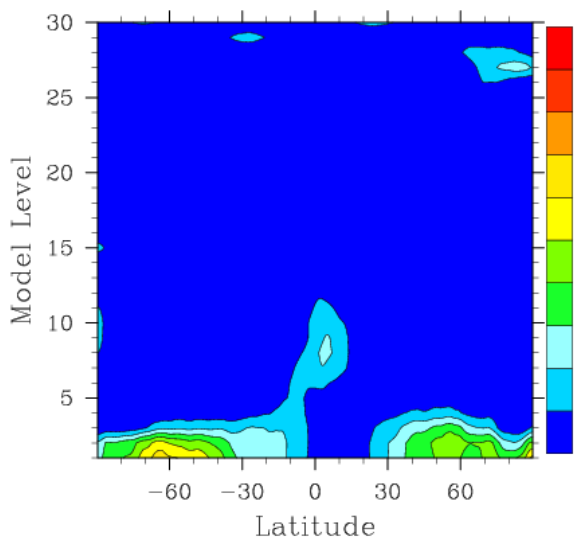
$$T_u' = T' - T_b$$

$$p_{s_u}' = p_s' - p_{s_b}$$

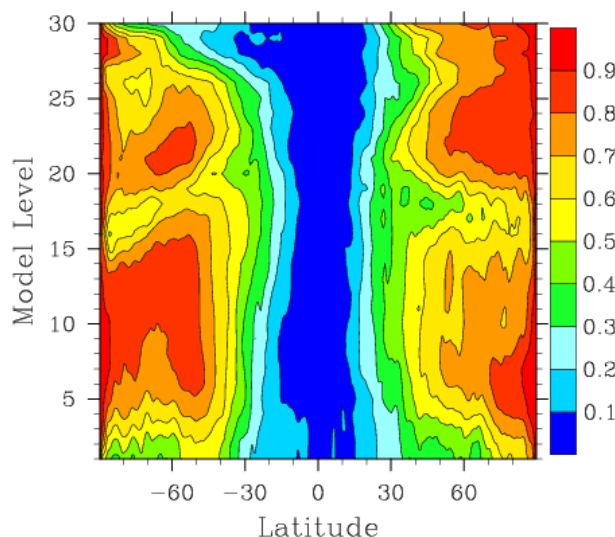
WRFDA Balance constraints

- WRFDA imposes statistical balanced constraints between
 - Stream Function & Velocity potential
 - Stream Function & Temperature
 - Stream Function & Surface Pressure
- How good are these balanced constraints?

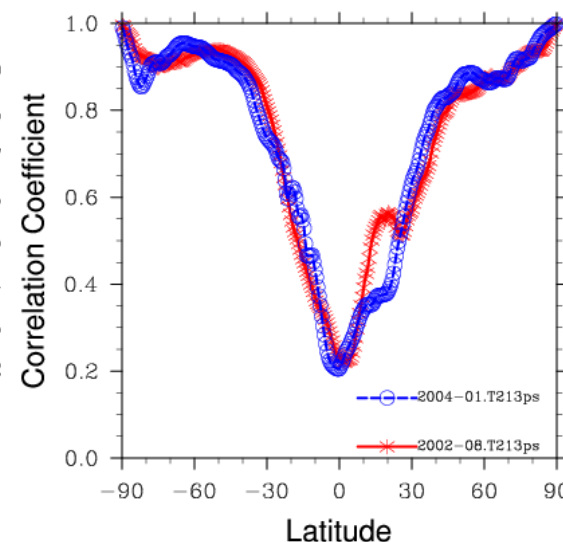
$$\langle \chi_b \bullet \chi \rangle / \langle \chi \bullet \chi \rangle$$



$$\langle T_b \bullet T \rangle / \langle T \bullet T \rangle$$



$$\langle p_{sb} \bullet p_s \rangle / \langle p_s \bullet p_s \rangle$$



Computed based on KMA global model

“gen_be” - Stage3

- Reads “gen_be_stage3” namelist
- Removes mean for χ_u' , T_u' & p_{s_u}'
- Computes eigenvectors and eigen values for vertical error covariance matrix of ψ' , χ_u' , T_u' & q
- Computes variance of p_{s_u}'
- Computes eigen decomposition of ψ' , χ_u' , T_u' & q

“gen_be” - Stage4

- Reads “gen_be_stage4” namelist
- For each variable & each eigen mode, for regional option computes “**lengthscale (s)**”

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

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

- For global option, computes “**power spectrum (D_n)**”

$$D_n = \sum_{m=-n}^n (F_n^m)^2 = (F_n^0)^2 + 2 \sum_{m=1}^n \left[(\text{Re}(F_n^m))^2 + (\text{Im}(F_n^m))^2 \right]$$

Single observation test

- Through single observation test, one can understand
 - structure of BE
 - It identifies the “shortfalls” of BE
 - It gives a broad guidelines for tuning BE

Basic concept:

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

Set single observation (U,V,T etc.) as follows:

$$[y^o - H(x^b)] = 1.0 \quad ; \quad R = I$$

Thus,

$$x^a - x^b = B * \text{constant delta vector}$$

How to activate Single obs test (PSOT)?

“single obs utility” or “psot” 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”

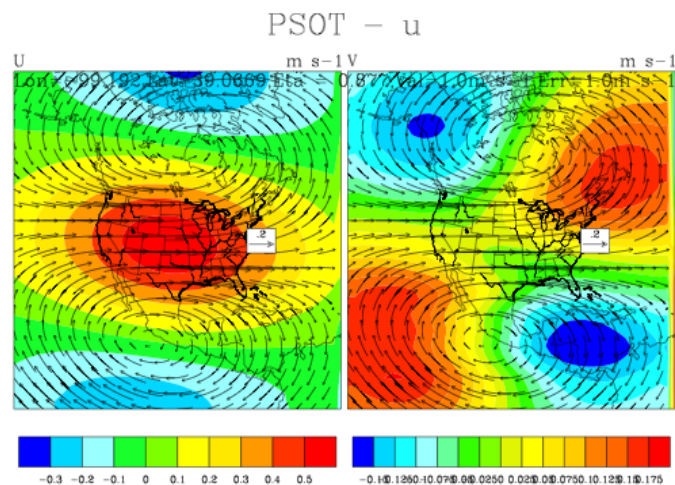
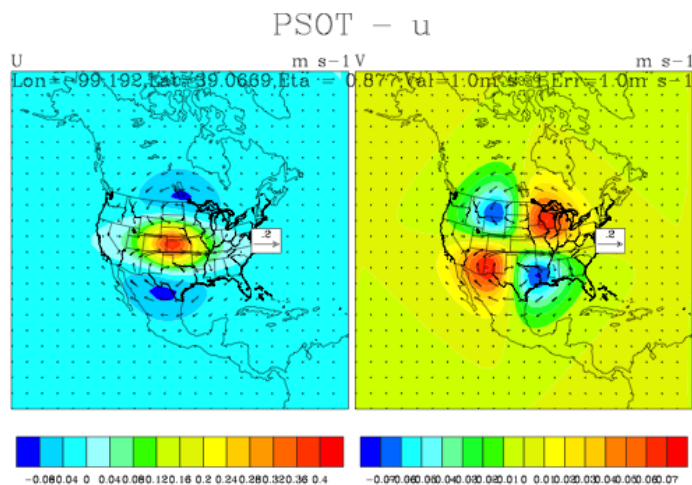
pseudo_y = “Y-coordinate of the observation”

pseudo_z = “Z-coordinate of the observation”

pseudo_val = “Observation innovation”, departure from FG”

pseudo_err = “Observation error”

Single Obs (U) test with different BE



How to perform tuning of BE?

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

LEN_SCALING1 - 5 (Length scaling parameters)

VAR_SCALING1 - 5 (Variance scaling parameters)

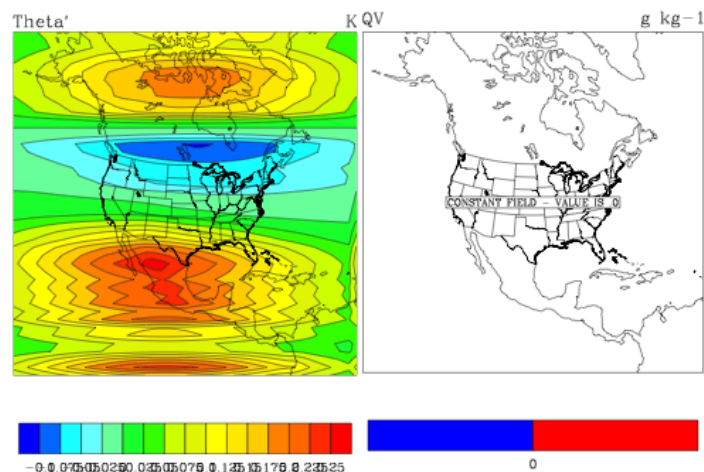
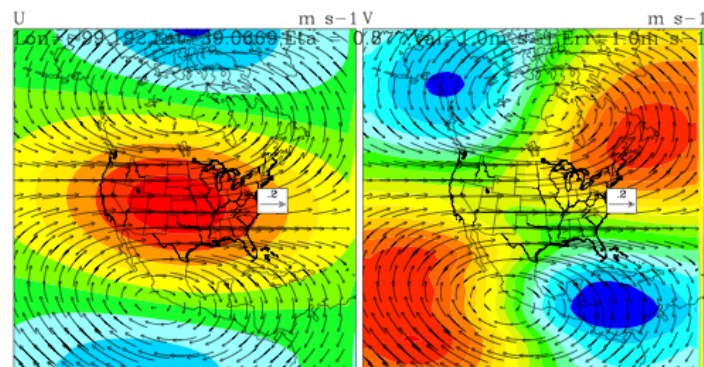
- Vertical component of BE can be tuned with following namelist parameter

MAX_VERT_VAR1 - 5 (Vertical variance parameters)

Results with BE Tuning

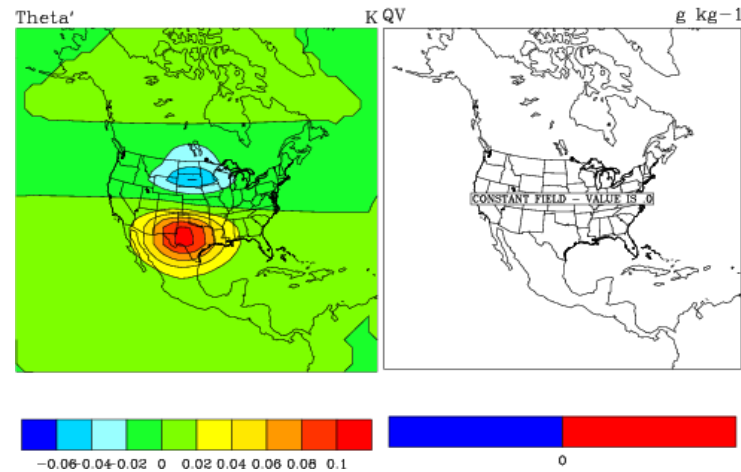
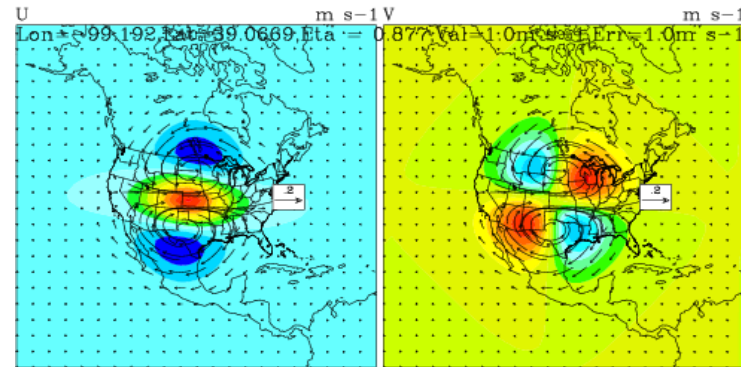
No tuning

PSOT - u



Len_scaling1 & 2 = 0.25

PSOT - u



Multivariate formulation of BE

- New set of analysis control variables (cv_options=6) have been designed

$$\chi_b(i, j, k) = \alpha_{\chi\psi} * \psi(i, j, k)$$

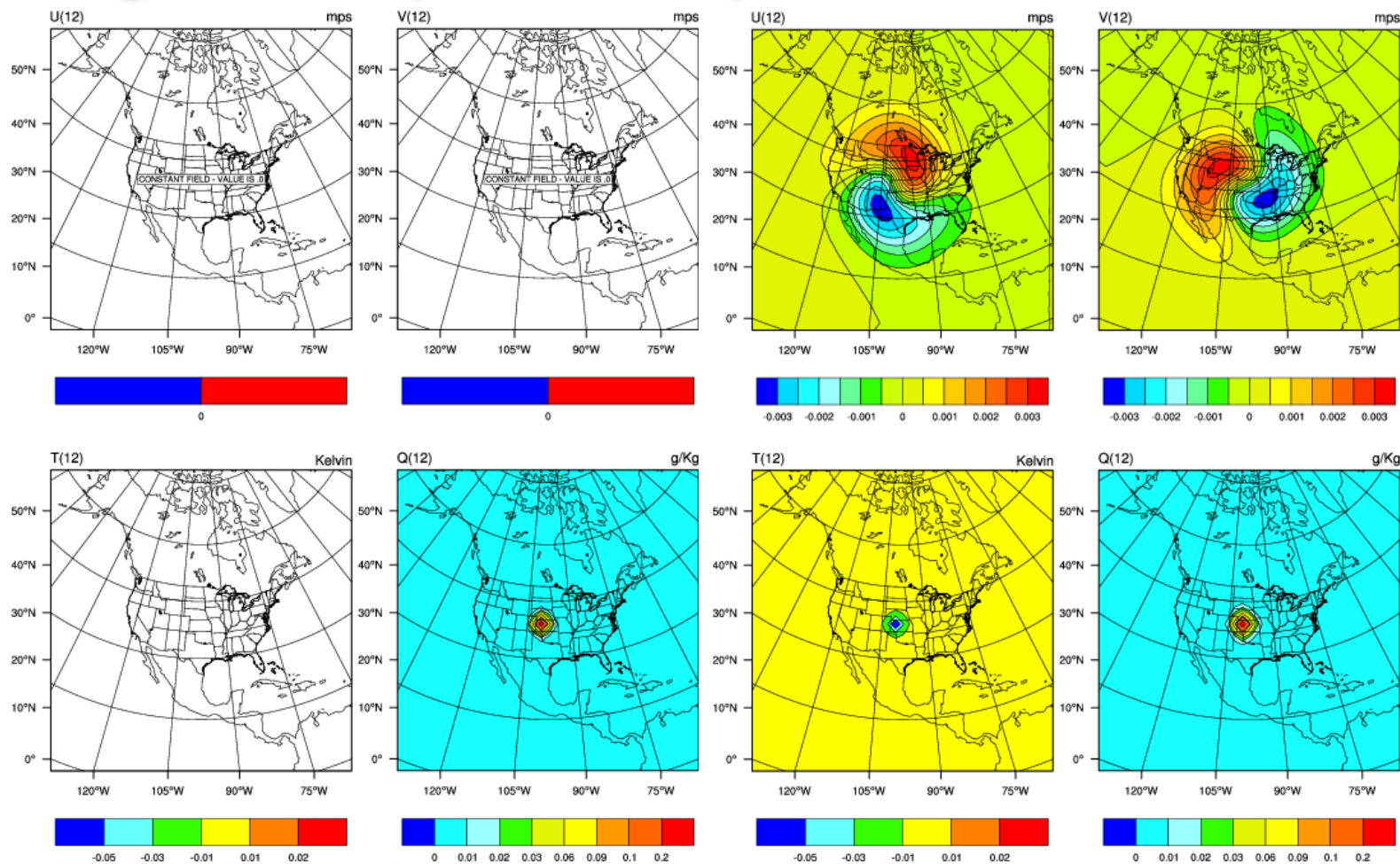
$$T_b(i, j, k) = \sum_{l=1}^{N_k} \alpha_{T\psi}(i, j, k, l) * \psi(i, j, k, l) + \sum_{l=1}^{N_k} \alpha_{T\chi_u}(i, j, k, l) * \chi_u(i, j, l)$$

$$Q_b(i, j, k) = \sum_{l=1}^{N_k} \alpha_{Q\psi}(i, j, k, l) * \psi(i, j, k, l) + \sum_{l=1}^{N_k} \alpha_{Q\chi_u}(i, j, k, l) * \chi_u(i, j, l) + \sum_{l=1}^{N_k} \alpha_{QT_u}(i, j, k, l) * T_u(i, j, l) + \sum_{l=1}^{N_k} \alpha_{ps_uQ}(i, j, l) * ps_u(i, j)$$

$$P_s(i, j) = \sum_{l=1}^{N_k} \alpha_{P_s\psi}(i, j, l) * \psi(i, j, l) + \sum_{l=1}^{N_k} \alpha_{P_s\chi_u}(i, j, l) * \chi_u(i, j, l)$$

Indexes i,j and k are corresponding to West-East, North-South and vertical sigma levels respectively, N_k is the number of sigma levels and α is the regression coefficient between the variables indicated in its subscript.

Single Obs (Moisture) test



`cv_options=5, BE`

`Cv_optins=6, BE`

Upcoming new features

- Some filtering options at various stages
- Background error for cloud hydrometeors like cloud water vapor, ice, snow and rain
- Introduction of new “bin_type=7” for four types of “rain” categories
- Additional diagnostics to study the frequency distribution of background error statistics
- Implementation of Holm (2002) type background error
- Stand alone branch of “gen_be”

Advanced Practice Session – “gen_be”

- Compilation of “gen_be” utility
- Generation of BE statistics
- Familiarization with various graphical utilities to display “gen_be” diagnostics
- Running single observation tests to understand the structure of BE
- BE error tuning

Generation of BE

- “gen_be_wrapper.ksh” script for generating BE for “CONUS” at 200 Km domain with:

Grid Size : 45 x 45 x 28

BE Method : NMC Method

Data Input : January, 2007 forecasts, both from 00 & 12 UTC IC

Basic environment variables that needs to be set are:

- Gen_be executables location (WRFVAR_DIR)
- Forecast input data (FC_DIR)
- Run directory (BE_DIR)
- Data Range (START_DATE, END_DATE)

“gen_be” wrapper script basically executes
“var/scripts/gen_be/gen_be.ksh” script

Gen_be diagnostics

- “gen_be” creates various diagnostic files which may be used to display various components of BE statistics.

- Important files are:

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

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

scalelength: fort.194, fort.179, fort.183, fort.187

Correlation between χ_u & χ_b (chi_u.chi.dat)

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

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

Important Strings that need to be defined in the wrapper script

“var/script/gen_be/gen_be_plot_wrapper.ksh”

BE_DIR --- gen_be Run directory

How to run Single Observation Test ?

- Familiarization with single observation “wrapper” script (“da_run_suite_wrapper_con200.ksh”) to run Single Observation test
- Key parameters 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

BE tuning

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

LEN_SCALING1 - 5 (Length scaling parameters)

VAR_SCALING1 - 5 (Variance scaling parameters)

MAX_VERT_VAR1 - 5 (Vertical variance parameters)

Note: If BE is available for the same domain configuration then its tuning is not required