GSI Tutorial 2011

Background and Observation Errors: Estimation and Tuning

Daryl Kleist
NCEP/EMC

Background Errors

1. Background error covariance
2. Multivariate relationships
3. Estimating/tuning background errors
4. Balance
5. Flow dependence

3DVAR Cost Function

$$J_{\mathrm{Var}}(\mathbf{x}') = \frac{1}{2}(\mathbf{x}')^{\mathrm{T}} \mathbf{B}_{\mathrm{Var}}^{-1}(\mathbf{x}') + \frac{1}{2}(\mathbf{H}\mathbf{x}' - \mathbf{y}_{\mathrm{o}}')^{\mathrm{T}} \mathbf{R}^{-1}(\mathbf{H}\mathbf{x}' - \mathbf{y}_{\mathrm{o}}') + J_{\mathrm{c}}$$

- J : Penalty (Fit to background + Fit to observations + Constraints)
- **x**' : Analysis increment ($\mathbf{x}_a - \mathbf{x}_b$) ; where $\mathbf{x}_b$ is a background
- $\mathbf{B}_{\mathbf{Var}}$: Background error covariance
- **H** : Observations (forward) operator
- **R** : Observation error covariance (Instrument + Representativeness)
- $\mathbf{y}_o$' : Observation innovations/residuals ($\mathbf{y}_o$-$\boldsymbol{H}\mathbf{x}_b$)
- J_c : Constraints (physical quantities, balance/noise, etc.)

Background Error Covariance

- Vital for controlling amplitude and structure for correction to model first guess (background)

- Covariance matrix
  - Controls influence distance
  - Contains multivariate information
  - Controls amplitude of correction to background

- For NWP (WRF, GFS, etc.), matrix is prohibitively large
  - Many components are modeled or ignored

- Typically estimated a-priori, offline

Analysis (control) variables

- Analysis is often performed using non-model variables
  - Background errors defined for analysis/control (not model) variables
- Control variables for GSI (NCEP GFS application):
  - **Streamfunction (Ψ)**
  - Unbalanced Velocity Potential ($\chi_{unbalanced}$)
  - Unbalanced Virtual Temperature ($T_{unbalanced}$)
  - Unbalanced Surface Pressure ($Ps_{unbalanced}$)
  - Relative Humidity
    - Two options
  - Ozone mixing ratio
  - Cloud water mixing ratio
  - Skin temperature
    - Analyzed, but not passed onto GFS model

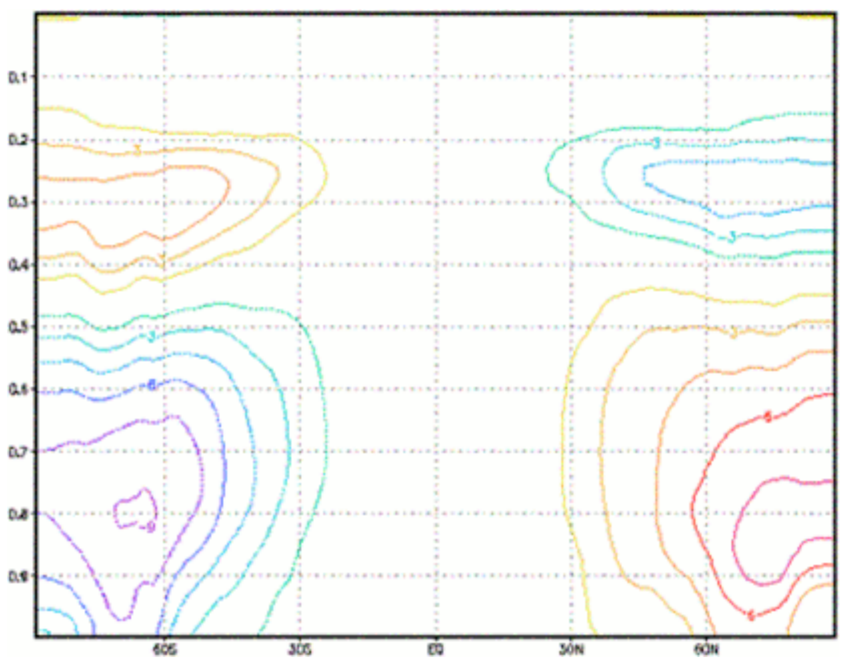
Multivariate Definition

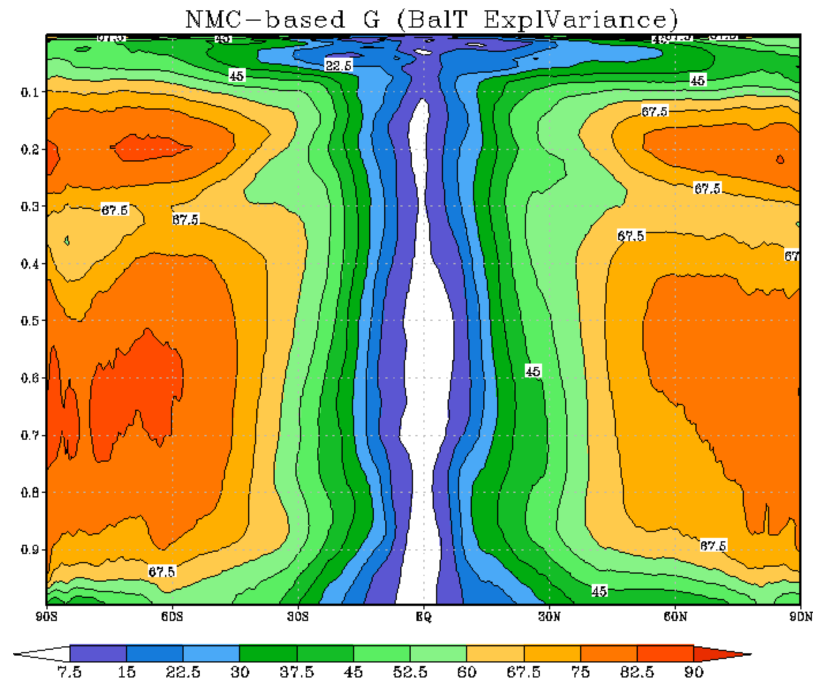
- $\chi = \chi_{\text{unbalanced}} + \mathbf{c}\,\Psi$
- $T = T_{\text{unbalanced}} + \mathbf{G}\,\Psi$
- $Ps = Ps_{\text{unbalanced}} + \mathbf{W}\,\Psi$

- Streamfunction is a key variable
  - defines a large percentage of temperature, velocity potential and surface pressure increment
- **G**, **W**, **c** are empirical matrices (estimated with linear regression) to project stream function increment onto balanced component of other variables

Multivariate Variable Definition

$$T_b = \mathbf{G}\psi$$



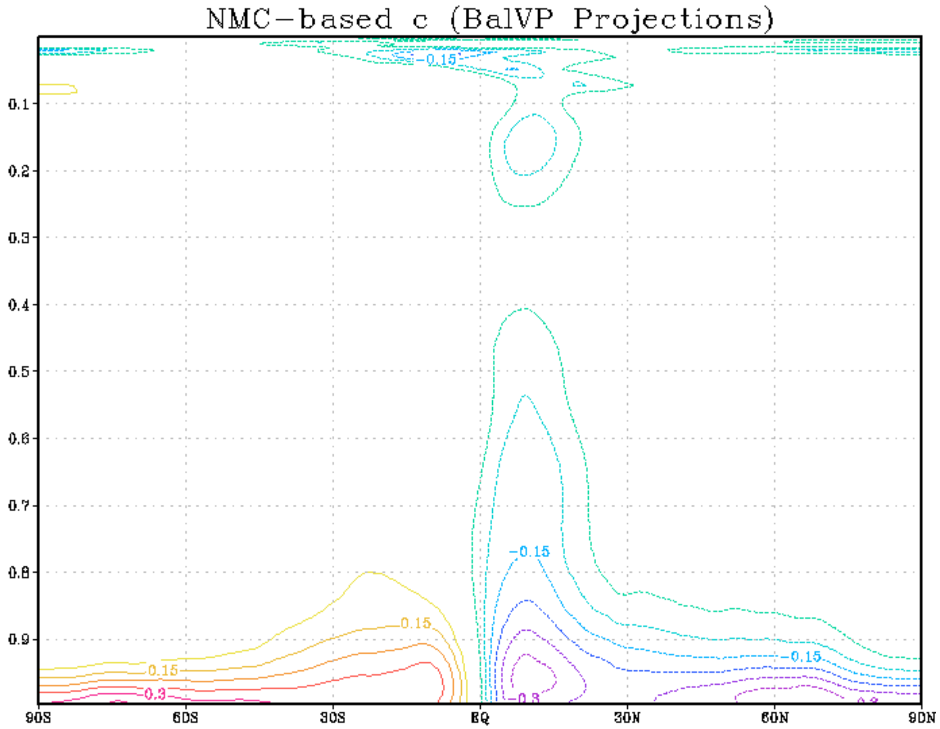


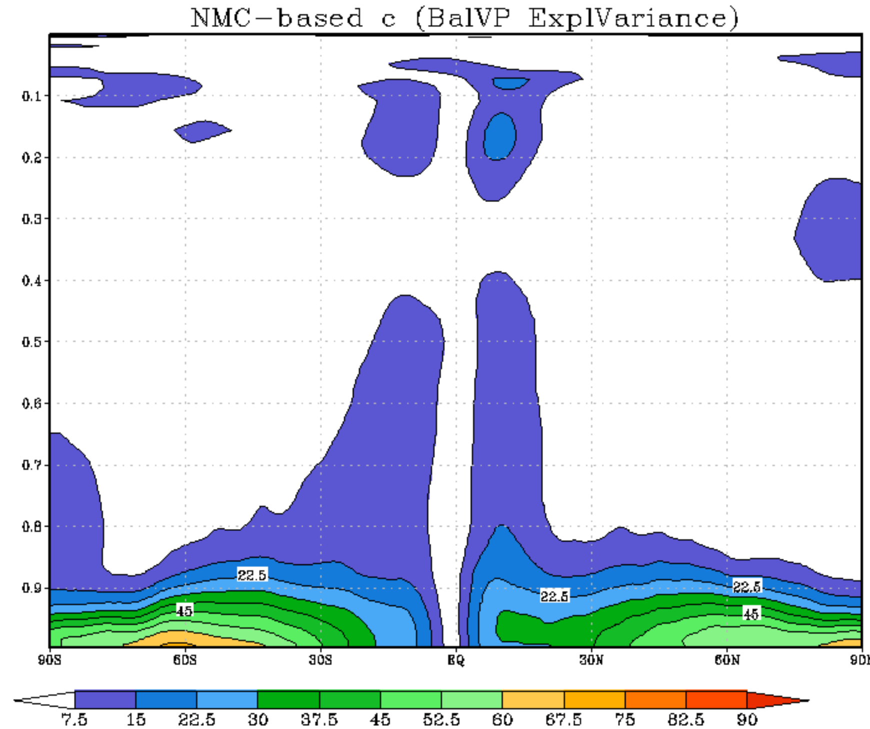
Projection of ψ at vertical level 25 onto vertical profile of balanced temperature ($\mathbf{G}_{25}$)

Percentage of full temperature variance explained by the balance projection

Multivariate Variable Definition

$$\chi_b = \mathbf{c}\psi$$



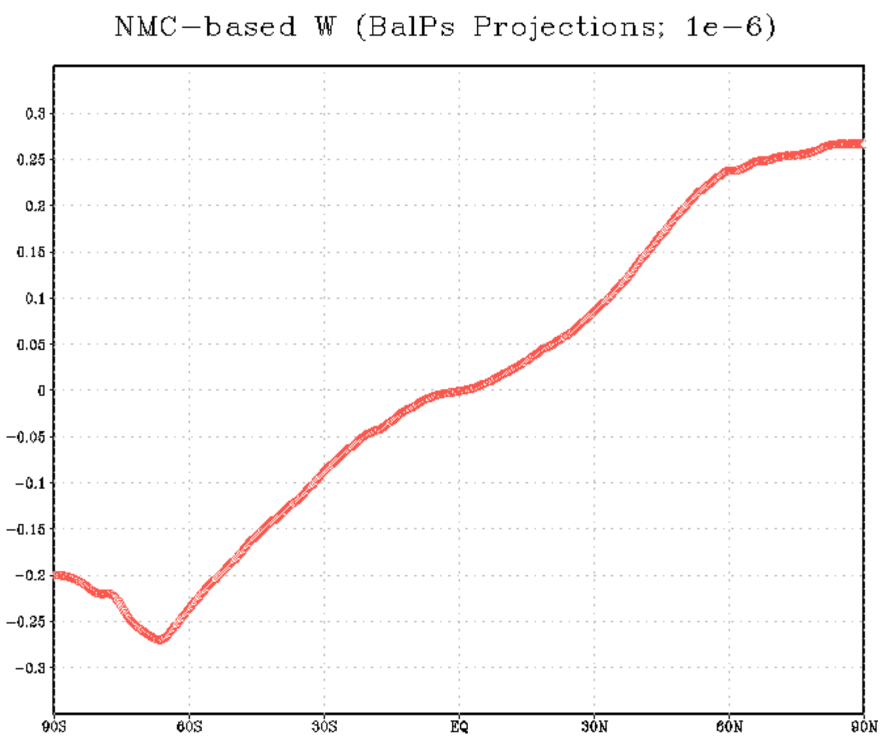


Projection of ψ onto balanced velocity potential (**c**)

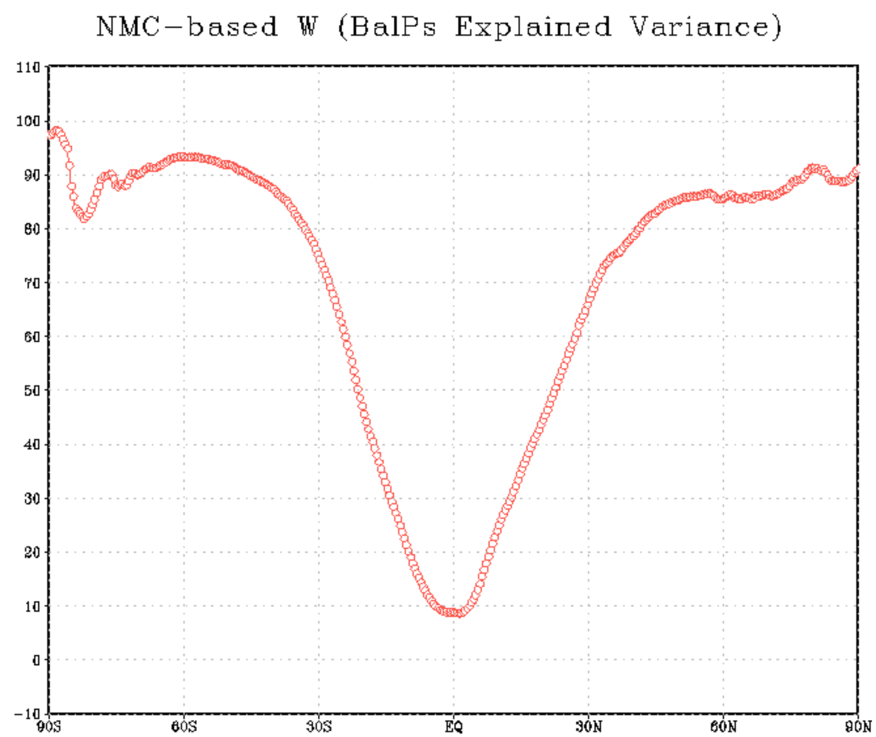
Percentage of full velocity potential variance explained by the balance projection

Multivariate Variable Definition

$$Ps_{\mathrm{b}} = \mathbf{w}\psi$$



Projection of ψ onto balanced surface pressure ($\mathbf{w}$)



Percentage of full surface pressure variance explained by the balance projection

Testing Background Error

- Best way to test background error covariance is through single observation experiments (as shown in some previous plots)

- Easy to run within GSI, namelist options:

  &SETUP

  oneobtest=.true.

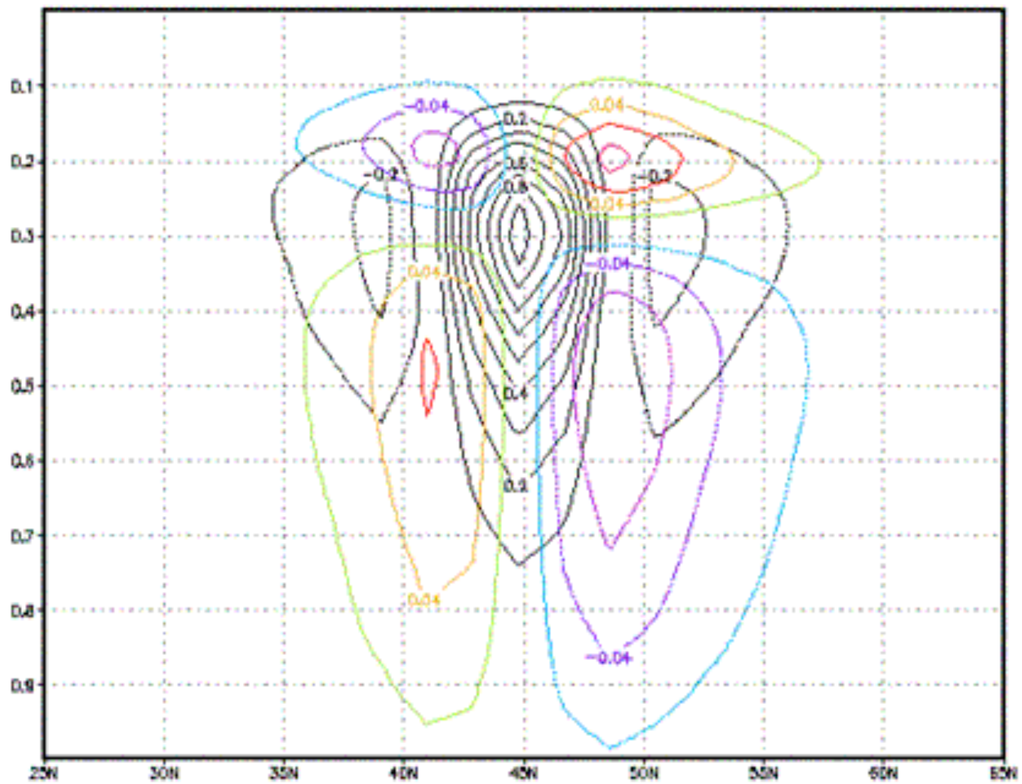
  &SINGLEOB_TEST

  maginnov=1.,magoberr=1.,oneob_type='u',oblat=45.,oblon=180,

  obpres=300.,obdattime= 2010101312,obhourset=0.,

Multivariate Example

Single zonal wind observation (1.0 ms^{-1} O-F and error)



u increment (black, interval 0.1 ms^{-1}) and T increment (color, interval 0.02K) from GSI

Moisture Variable

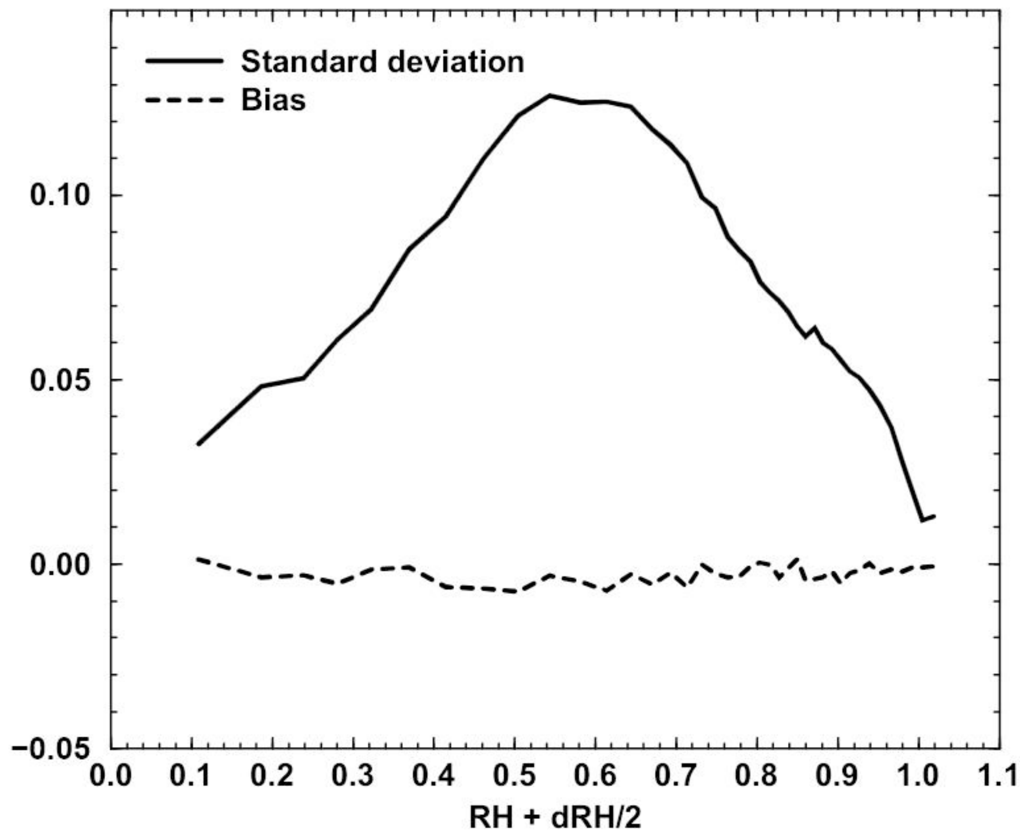
- Option 1
  - Pseudo-RH (univariate within inner loop)
- Option 2*
  - Normalized relative humidity
  - **Multivariate with temperature and pressure**
  - Standard Deviation a function of background relative humidity

$$\frac{\delta \mathrm{RH}}{\sigma\left(\mathrm{RH}^b\right)} = \mathrm{RH}^b\left(\frac{\delta \mathrm{p}}{\mathrm{p}^b} + \frac{\delta \mathrm{q}}{\mathrm{q}^b} - \frac{\delta \mathrm{T}}{\alpha^b}\right)$$

$$\alpha^b = \frac{-1}{\left(\frac{\partial \mathrm{RH}}{\partial \mathrm{T}}\right)}$$

- Holm (2002) ECMWF Tech. Memo

Background Error Variance for RH Option 2



- Figure 23 in Holm et al. (2002); ECMWF Tech Memo

Elements needed for GSI

- For each analysis variable
  - Amplitude (variance)
  - Recursive filter parameters
    - Horizontal length scale (km, for Gaussian)
    - Vertical length scale (grid units, for Gaussian)
      - 3D variables only

- Additionally, balance coefficients
  - **G**, **W**, and **c** from previous slides

Estimating (static) Background Error

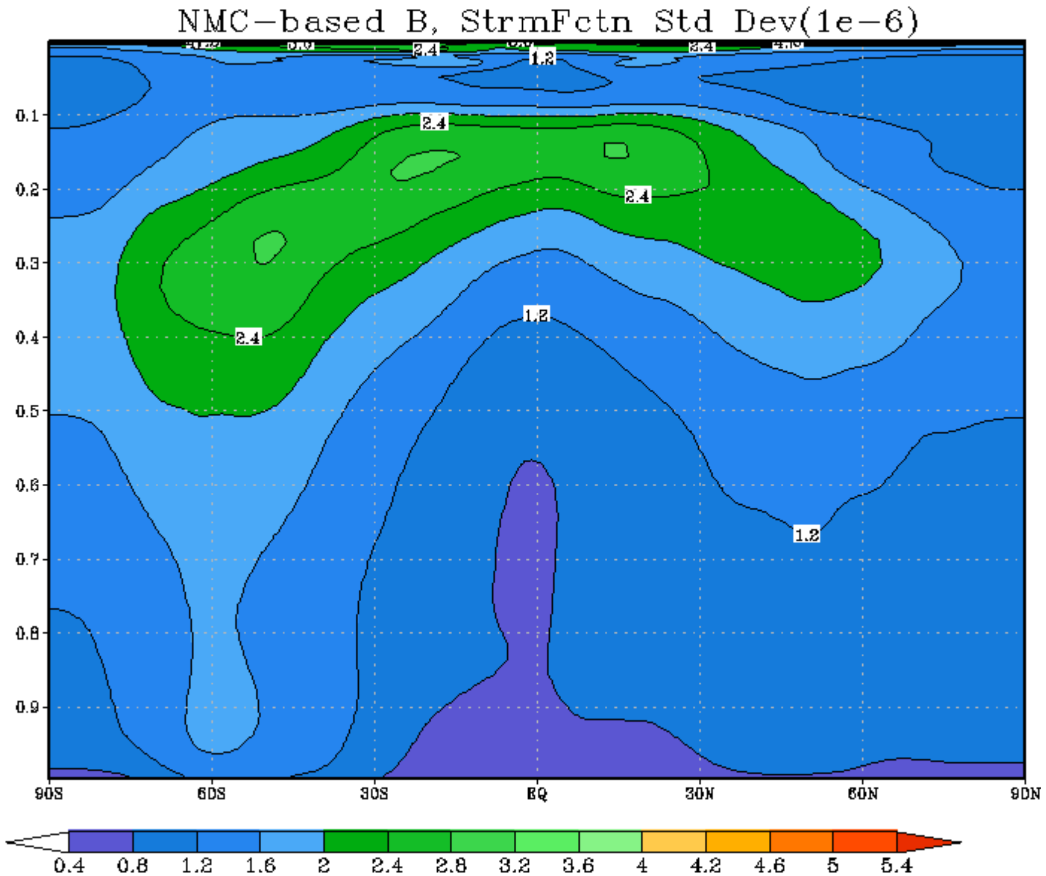
- ***NMC Method****
  - Lagged forecast pairs (i.e. 24/48 hr forecasts valid at same time, 12/24 hr lagged pairs, etc.)
  - Assume: Linear error growth
  - Easy to generate statistics from previously generated (operational) forecast pairs

- Ensemble Method
  - Ensemble differences of forecasts
  - Assume: Ensemble represents actual error

- Observation Method
  - Difference between forecast and observations
  - Difficulties: observation coverage and multivariate components

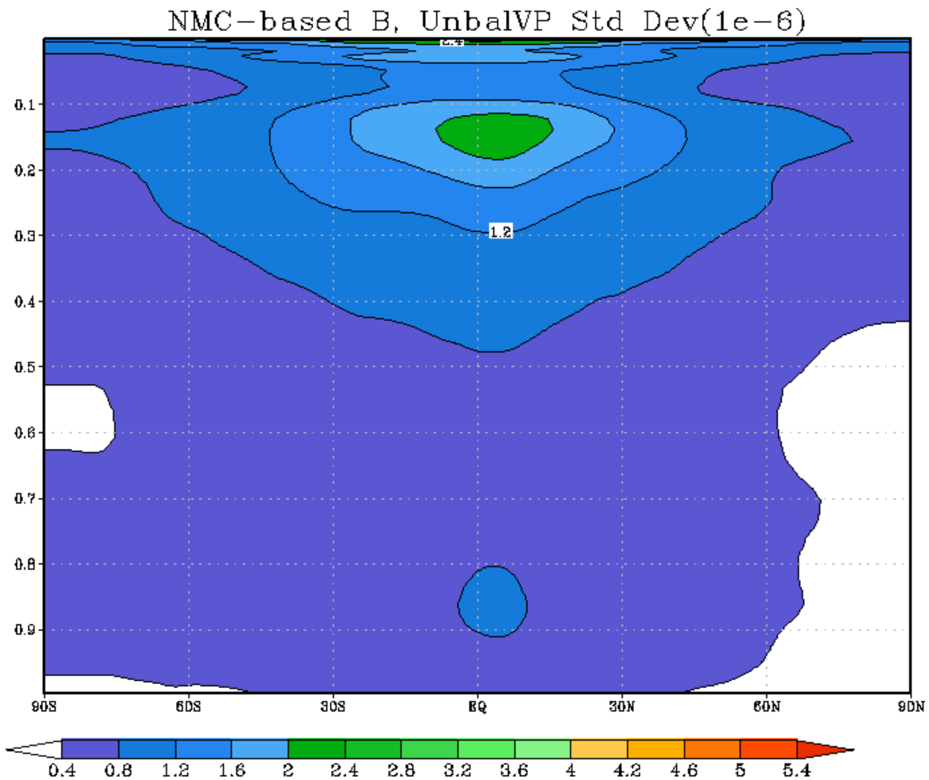
Amplitude (standard deviation)

- Function of latitude and height
- Larger in midlatitudes than in the tropics
- Larger in Southern Hemisphere than Northern Hemisphere

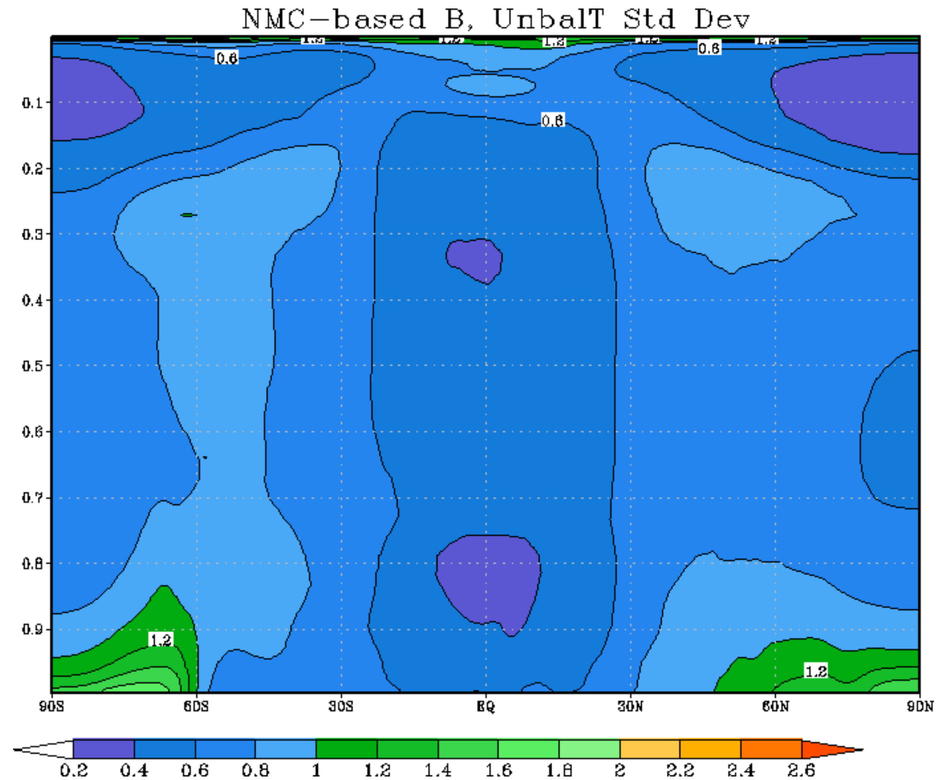


NMC-estimated standard deviation for **streamfunction**, from lagged 24/48hr GFS forecasts

Amplitude (standard deviation)

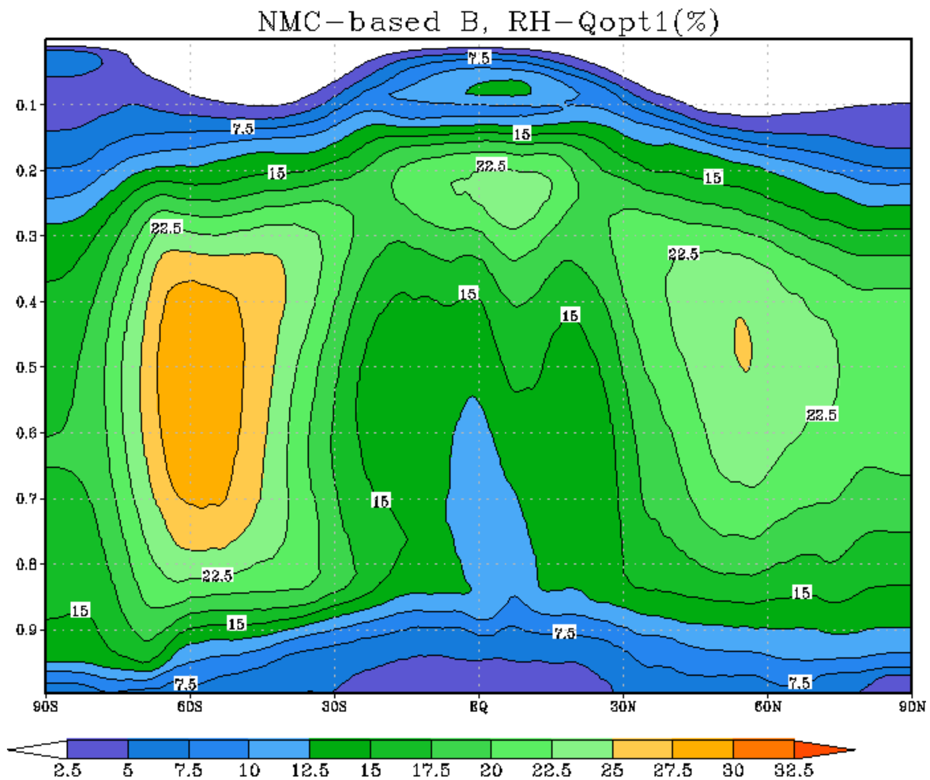


NMC-estimated standard deviation for **unbalanced velocity potential**, from lagged 24/48hr GFS forecasts

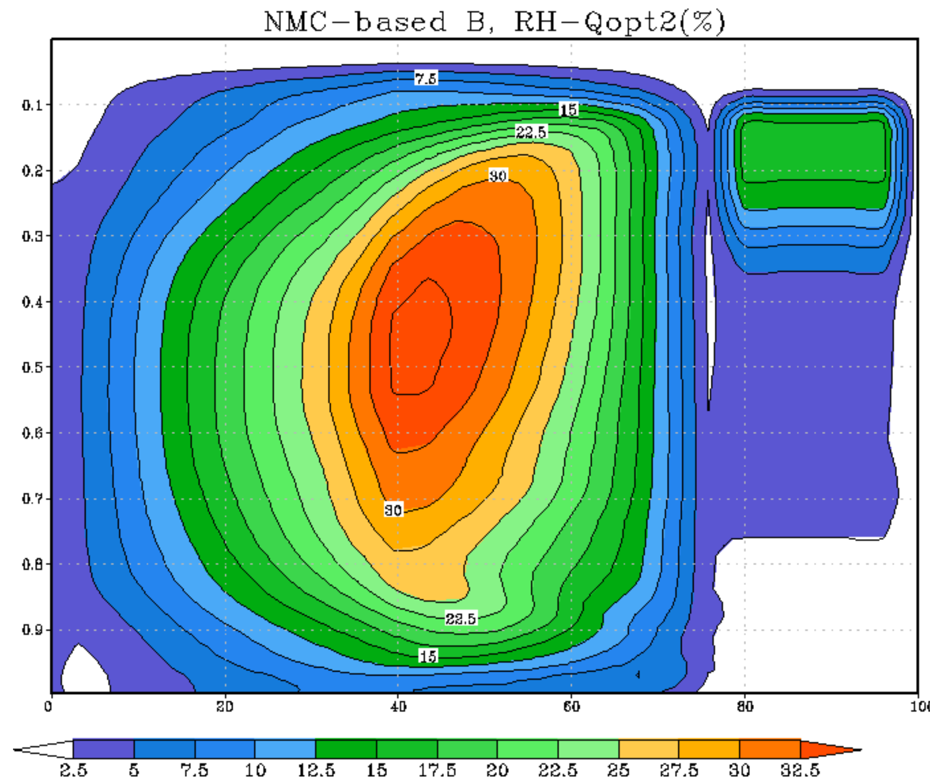


NMC-estimated standard deviation for **unbalanced virtual temperature**, from lagged 24/48hr GFS forecasts

Amplitude (standard deviation)

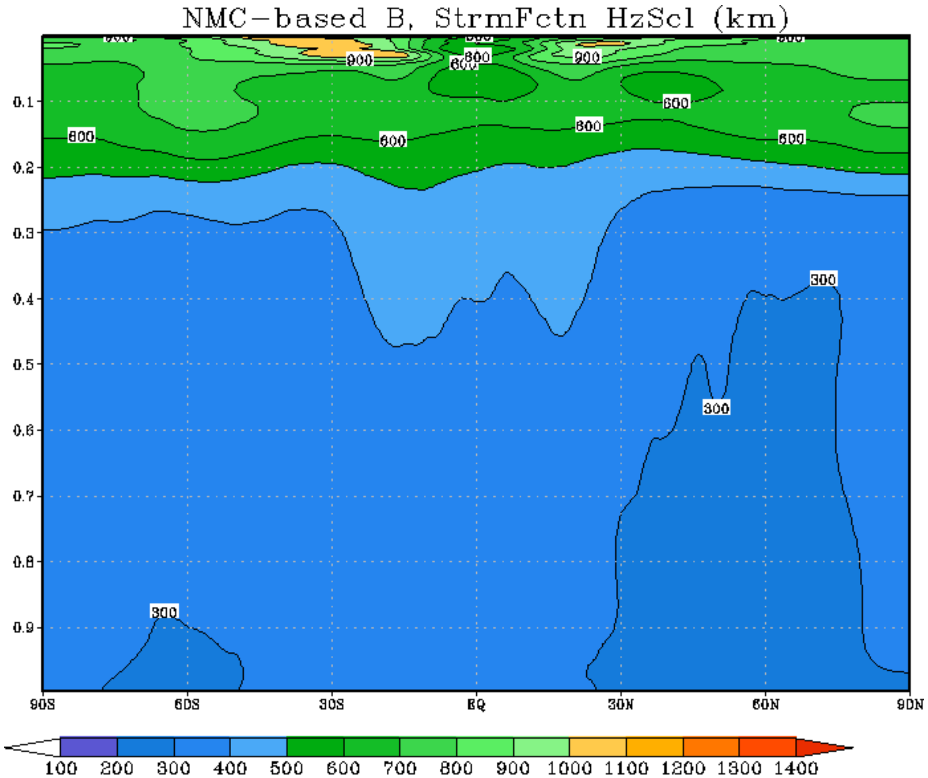


NMC-estimated standard deviation for **pseudo RH (q-option 1)**, from lagged 24/48hr GFS forecasts

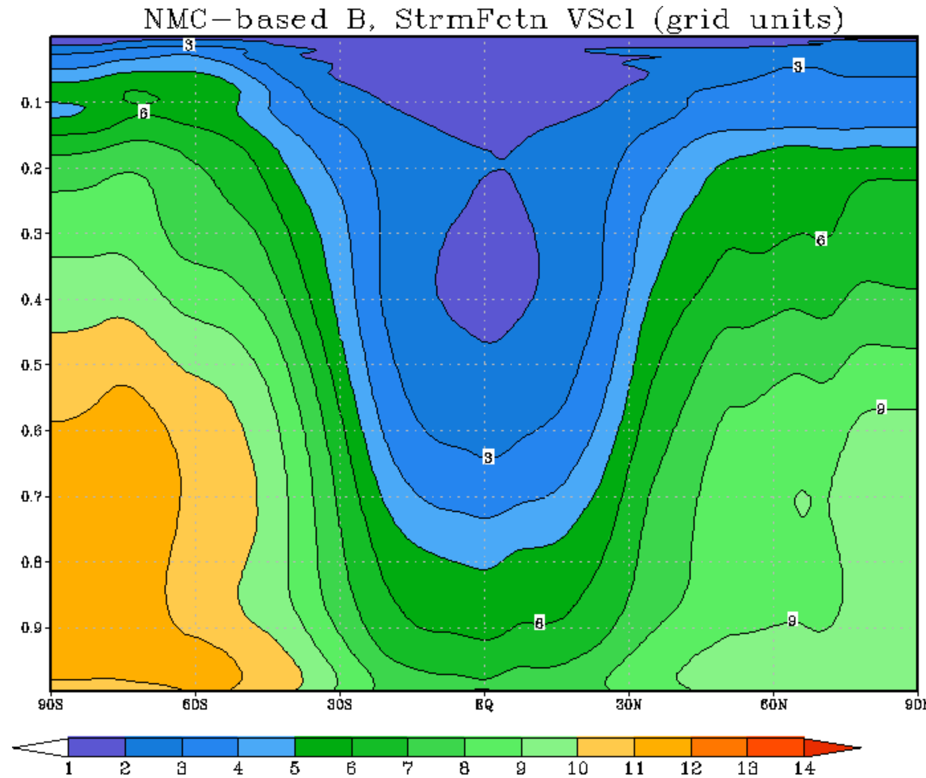


NMC-estimated standard deviation for **normalized pseudo RH (q-option 2)**, from lagged 24/48hr GFS forecasts

Length Scales

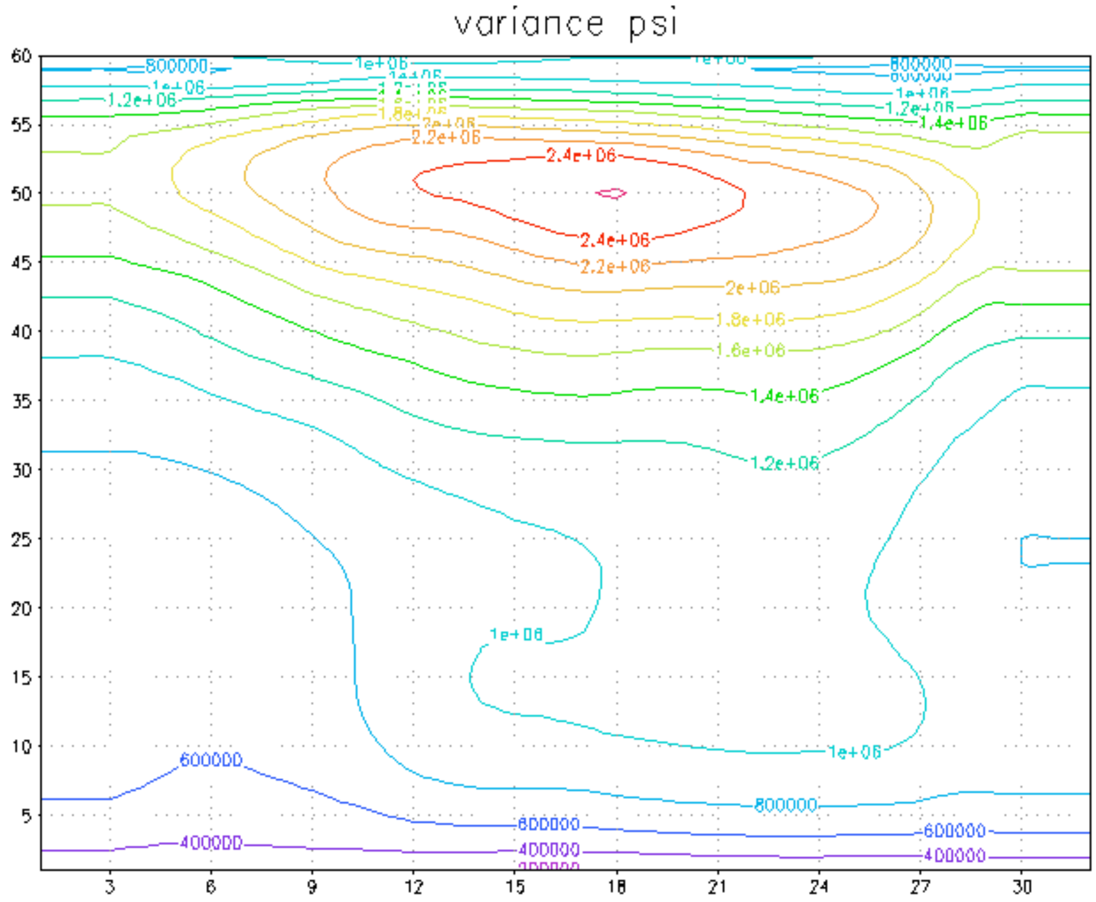


NMC-estimated horizontal length scales (km) for **streamfunction**, from lagged 24/48hr GFS forecasts

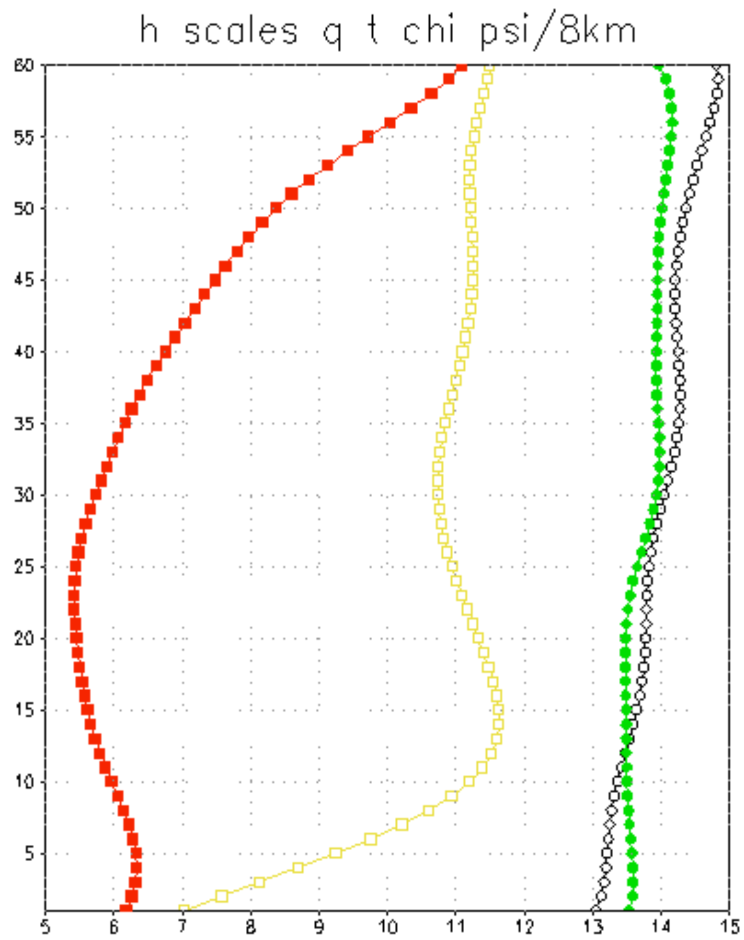


NMC-estimated vertical length scales (grid units) for **streamfunction**, from lagged 24/48hr GFS forecasts

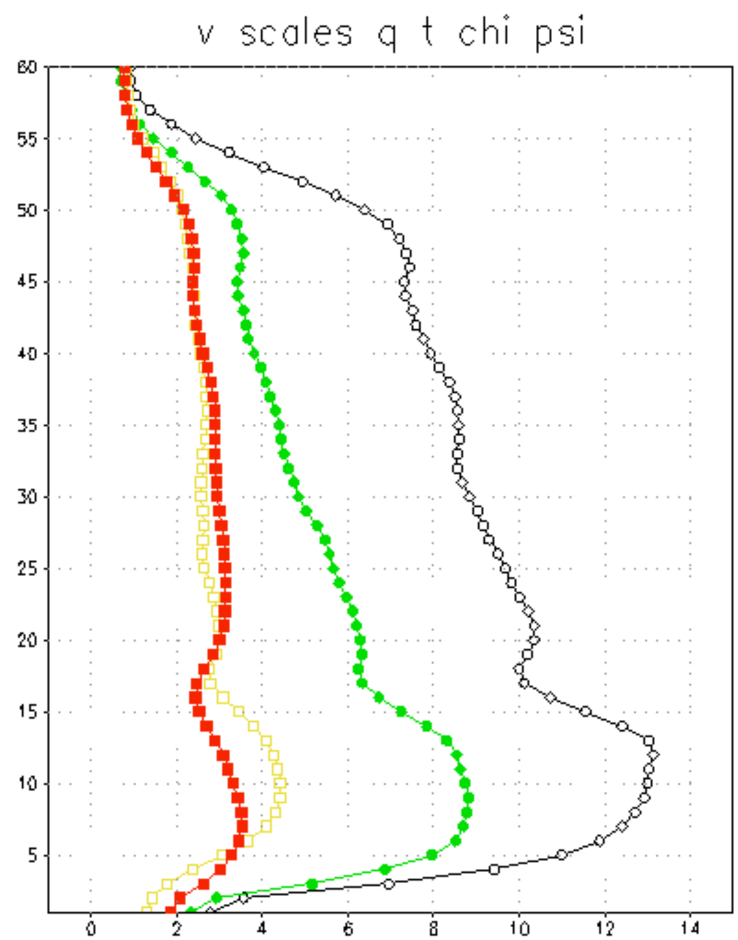
Regional (8km NMM) Estimated NMC Method



Regional Scales

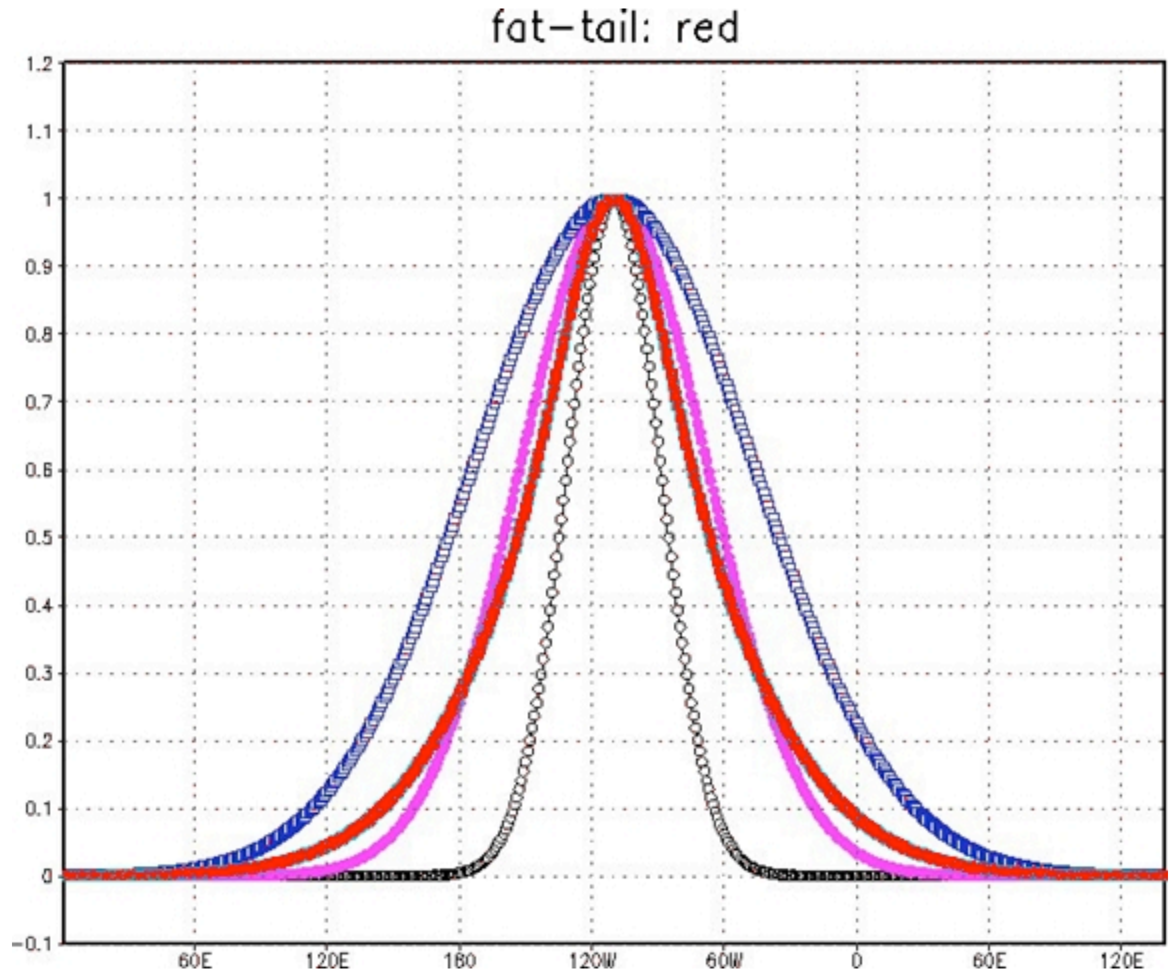


Horizontal Length Scales

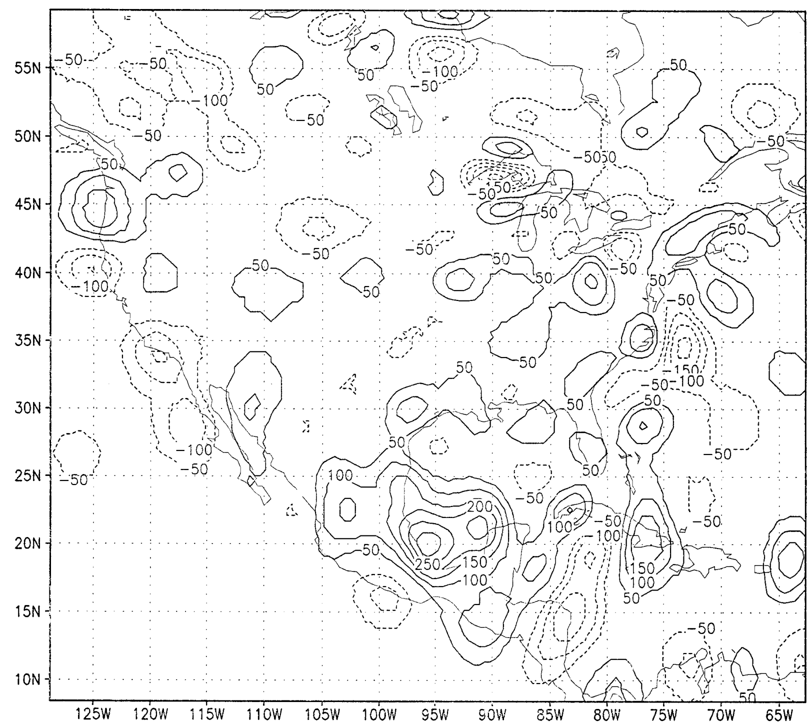


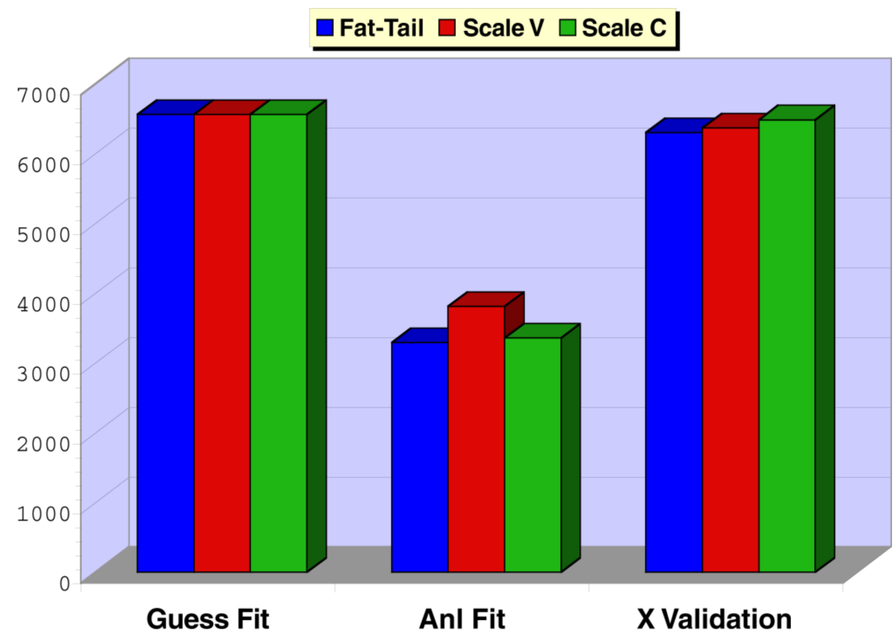
Vertical Length Scales

Fat-tailed power spectrum



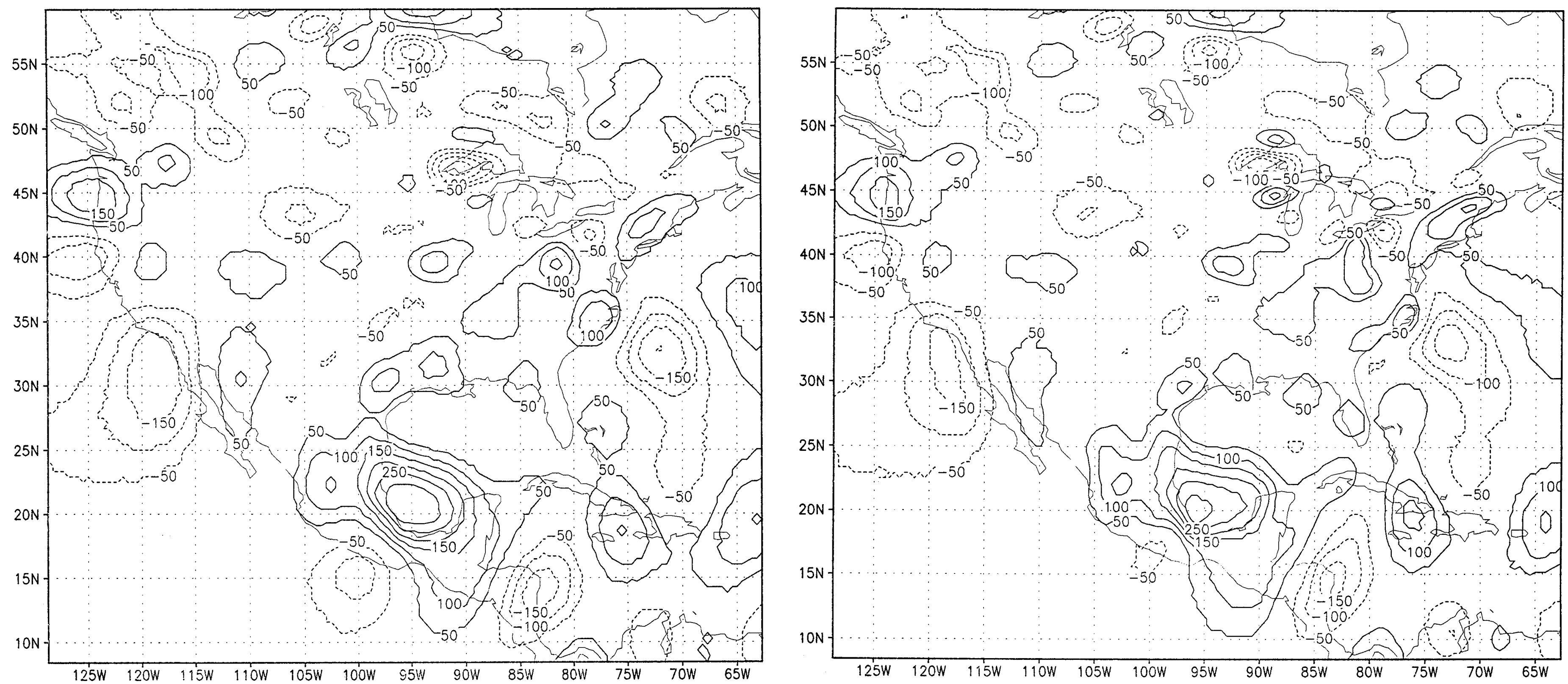
Fat-tailed Spectrum





Surface pressure increment with homogeneous scales using single recursive filter

Fat-tailed Spectrum



Surface pressure increment with inhomogeneous scales using single recursive filter, single scale (left) and multiple recursive filter: fat-tail (right)

Tuning Parameters

- GSI assumes binary fixed file with aforementioned variables
  - Example: berror=$fixdir/global_berror.l64y578.f77
- Anavinfo file contains information about control variables and their background error amplitude tuning weights

control_vector::

!var	level	itracer	**as/tsfc_sdv**	an_amp0	source	funcof
sf	64	0	**0.60**	-1.0	state	u,v
vp	64	0	**0.60**	-1.0	state	u,v
ps	1	0	**0.75**	-1.0	state	p3d
t	64	0	**0.75**	-1.0	state	tv
q	64	1	**0.75**	-1.0	state	q
oz	64	1	**0.75**	-1.0	state	oz
sst	1	0	**1.00**	-1.0	state	sst
cw	64	1	**1.00**	-1.0	state	cw
stl	1	0	**3.00**	-1.0	motley	sst
sti	1	0	**3.00**	-1.0	motley	sst

Tuning Parameters

- Length scale tuning controlled via GSI namelist

  &BKGERR

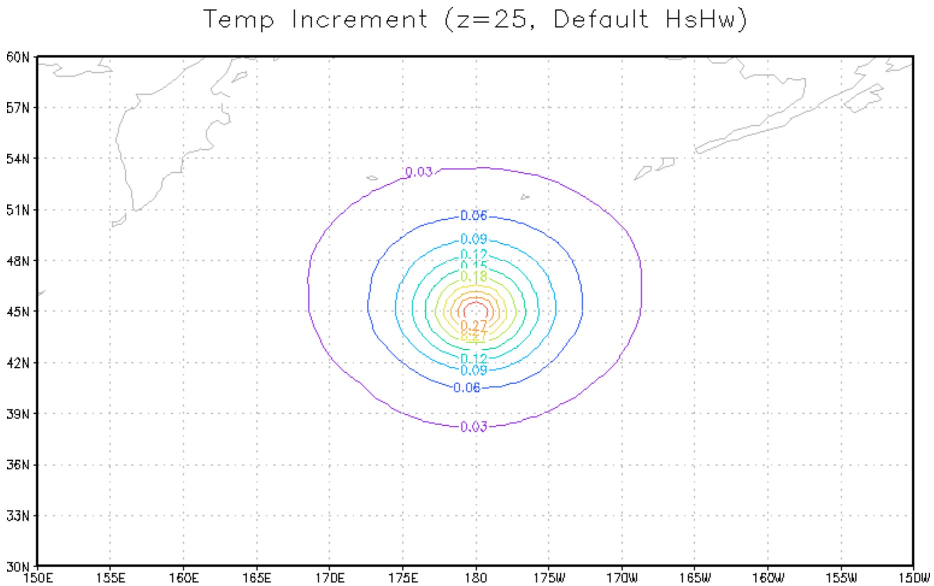
  hzscl = 1.7, 0.8, 0.5

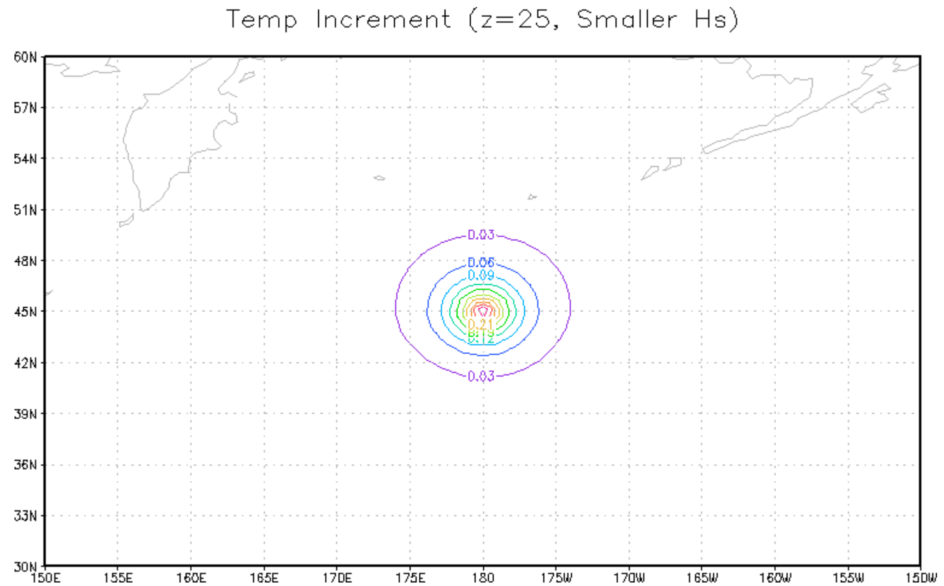
  hswgt = 0.45, 0.3, 0.25

  vs=0.7 [separable from horizontal scales]

- Hzscl/vs/as are all multiplying factors (relative to contents of “berror” fixed file)
- Three scales specified for horizontal (along with corresponding relative weights, hswgt)

Tuning Example (Scales)





Hzscl = 1.7, 0.8, 0.5

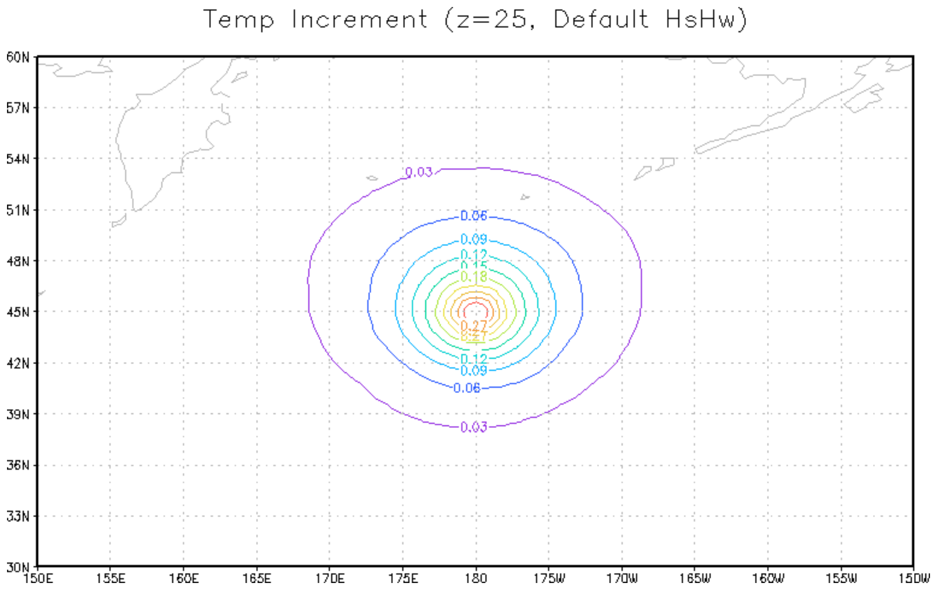
Hswgt = 0.45, 0.3, 0.25

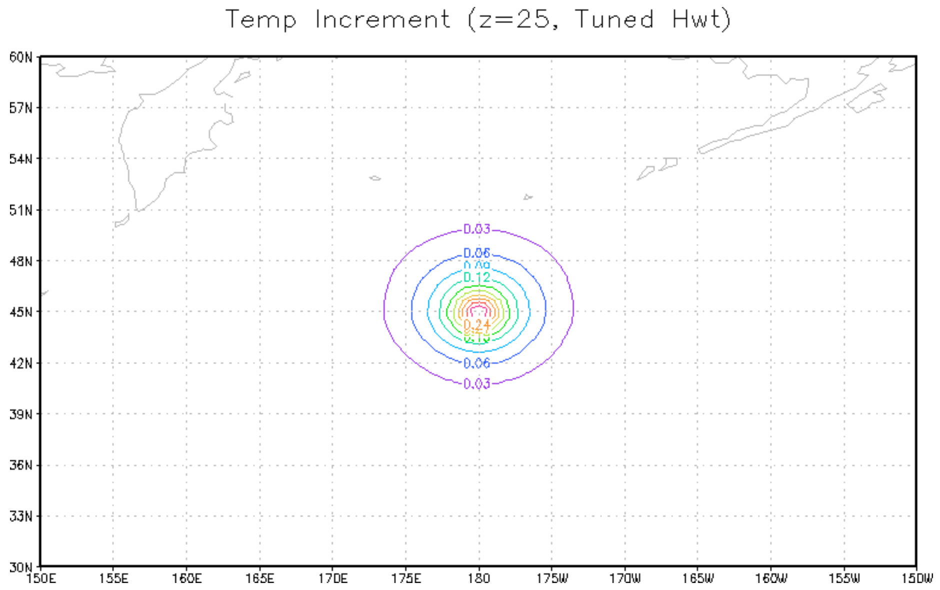
Hzscl = 0.9, 0.4, 025

Hswgt = 0.45, 0.3, 0.25

500 hPa temperature increment (K) from a single temperature observation utilizing GFS default (left) and tuned (smaller scales) error statistics.

Tuning Example (Weights)





Hzscl = 1.7, 0.8, 0.5

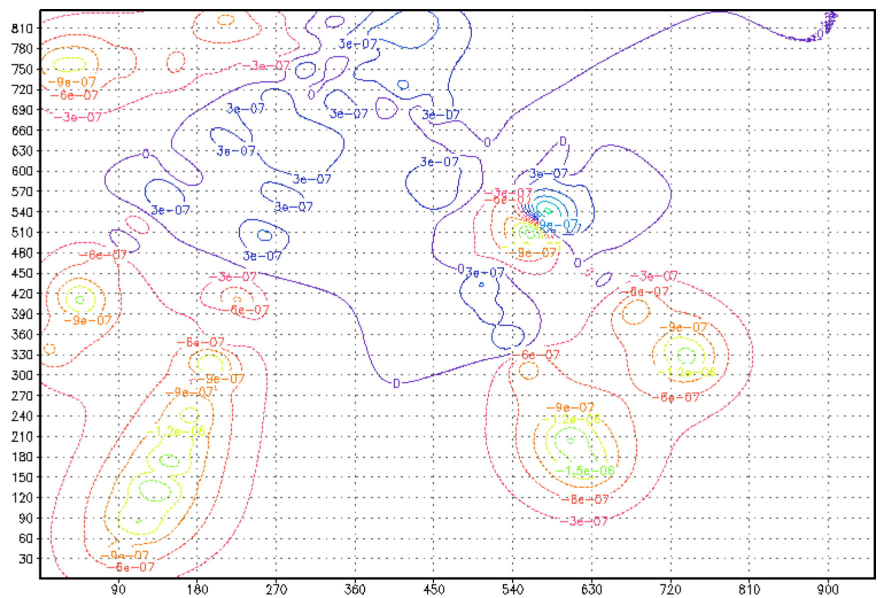
Hswgt = 0.45, 0.3, 0.25

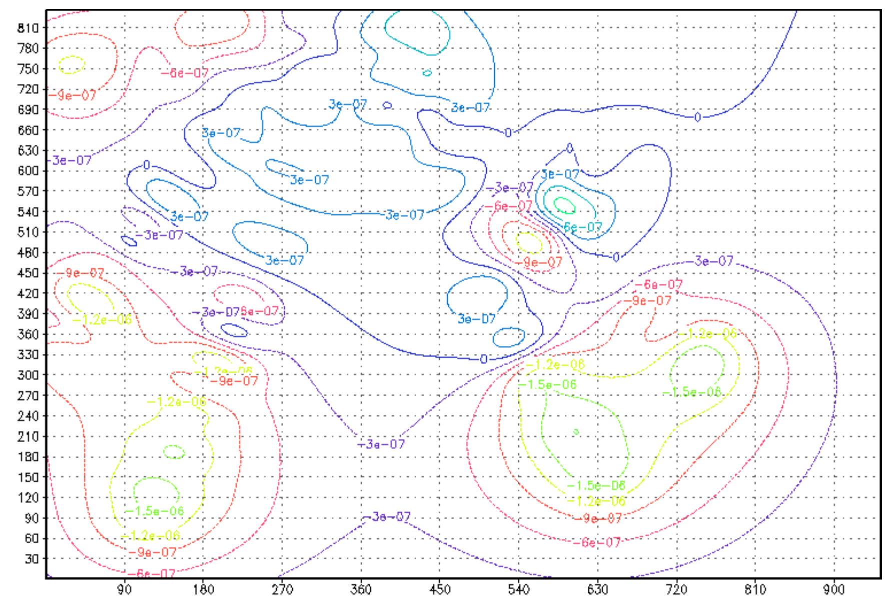
Hzscl = 1.7, 0.8, 0.5

Hswgt = 0.1, 0.3, 0.6

500 hPa temperature increment (K) from a single temperature observation utilizing GFS default (left) and tuned (weights for scales) error statistics.

Tuning Example (ozone)





Ozone analysis increment (mixing ratio) utilizing default (left) and tuned (larger scales) error statistics.

Balance/Noise

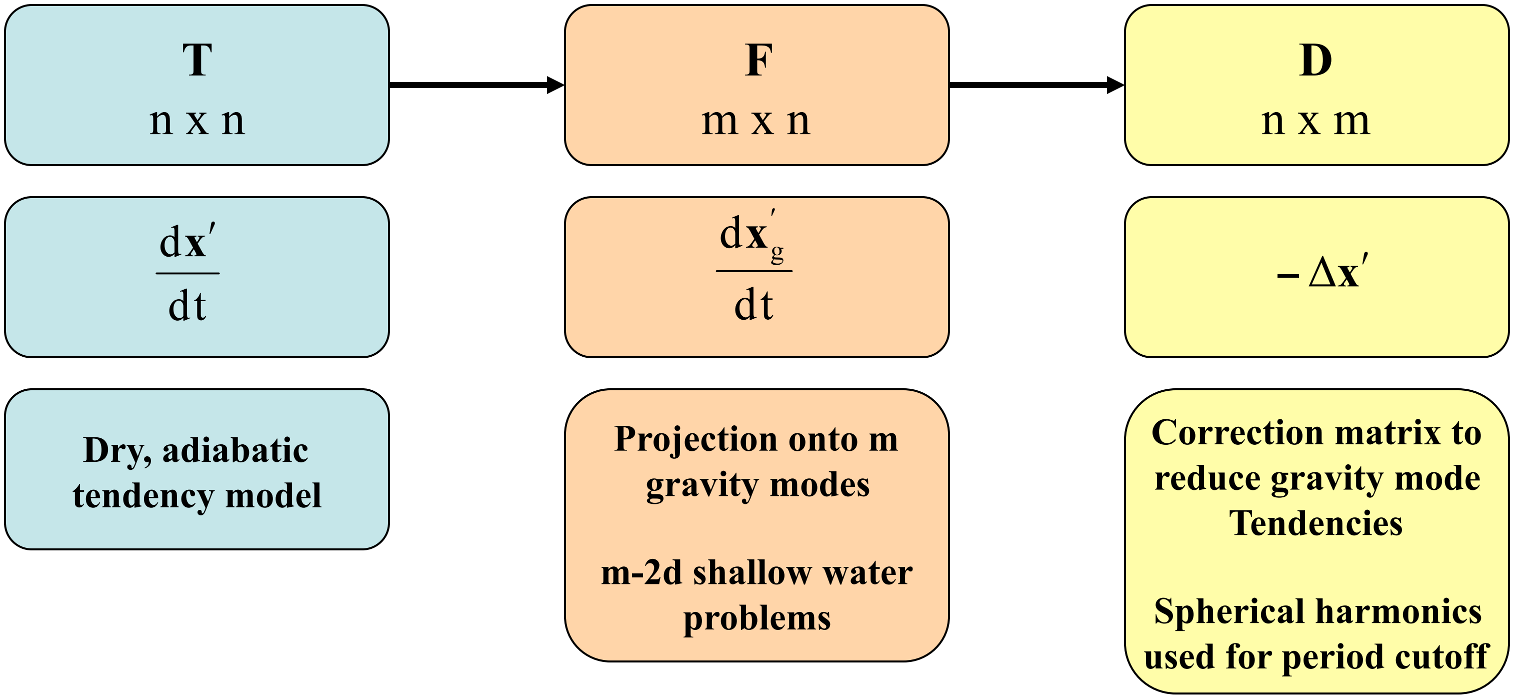
- In addition to statistically derived matrices, an optional (incremental) normal mode operator exists
  - **Not (yet) working well for regional applications**
  - Operational in global application (GFS/GDAS)

$$J\left(\mathbf{x}_c'\right) = \frac{1}{2}\left(\mathbf{x}_c'\right)^T \mathbf{C}^{-T}\mathbf{B}^{-1}\mathbf{C}^{-1}\left(\mathbf{x}_c'\right) + \frac{1}{2}\left(\mathbf{y}_o' - \mathbf{H}\mathbf{x}_c'\right)^T \mathbf{R}^{-1}\left(\mathbf{y}_o' - \mathbf{H}\mathbf{x}_c'\right) + J_c$$

$$\mathbf{x}_c' = \mathbf{C}\mathbf{x}'$$

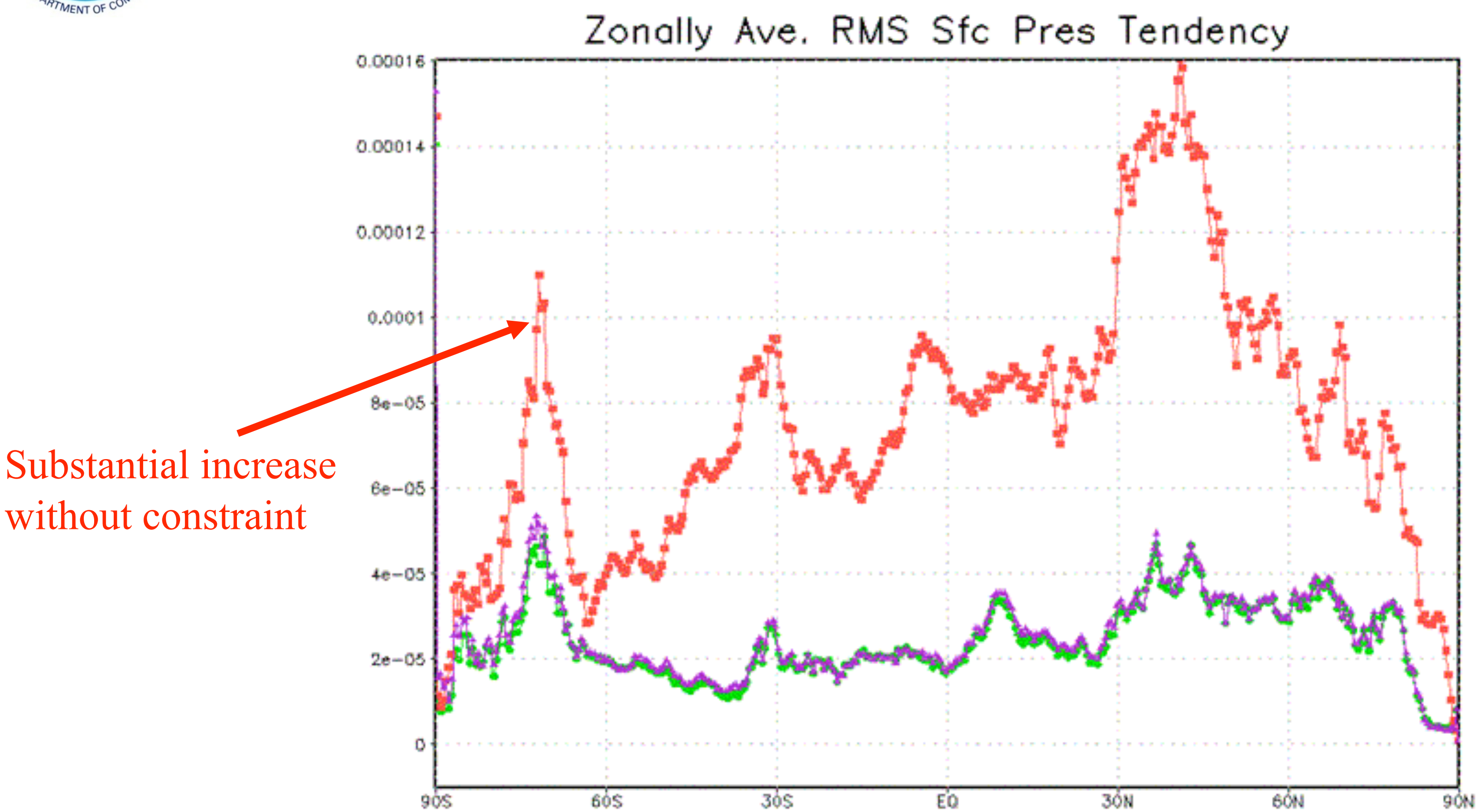
- **C** = Correction from incremental normal mode initialization (NMI)
  - represents correction to analysis increment that filters out the unwanted projection onto fast modes
- No change necessary for **B** in this formulation

C=[I-DFT]x’



- Practical Considerations:
  - **C** is operating on **x**’ only, and is the tangent linear of NNMI operator
  - Only need one iteration in practice for good results
  - Adjoint of each procedure needed as part of variational procedure

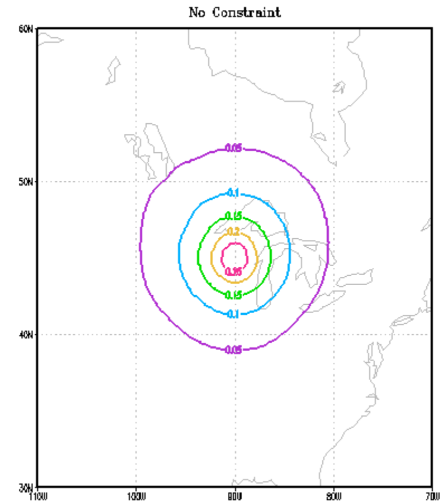
Noise/Balance Control



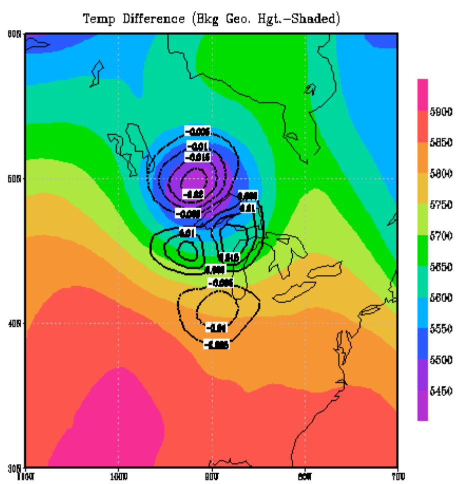
Zonal-average surface pressure tendency for background (green), unconstrained GSI analysis (red), and GSI analysis with TLNMC (purple).

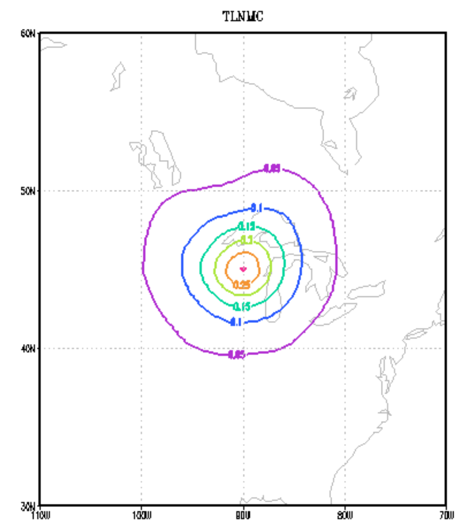
Example: Impact of Constraint

- Magnitude of TLNMC correction is small
- TLNMC adds flow dependence even when using same isotropic **B**



Isotropic response





Flow dependence added

500 hPa temperature increment (right) and analysis difference (left, along with background geopotential height) valid at 12Z 09 October 2007 for a single 500 hPa temperature observation (1K O-F and observation error)

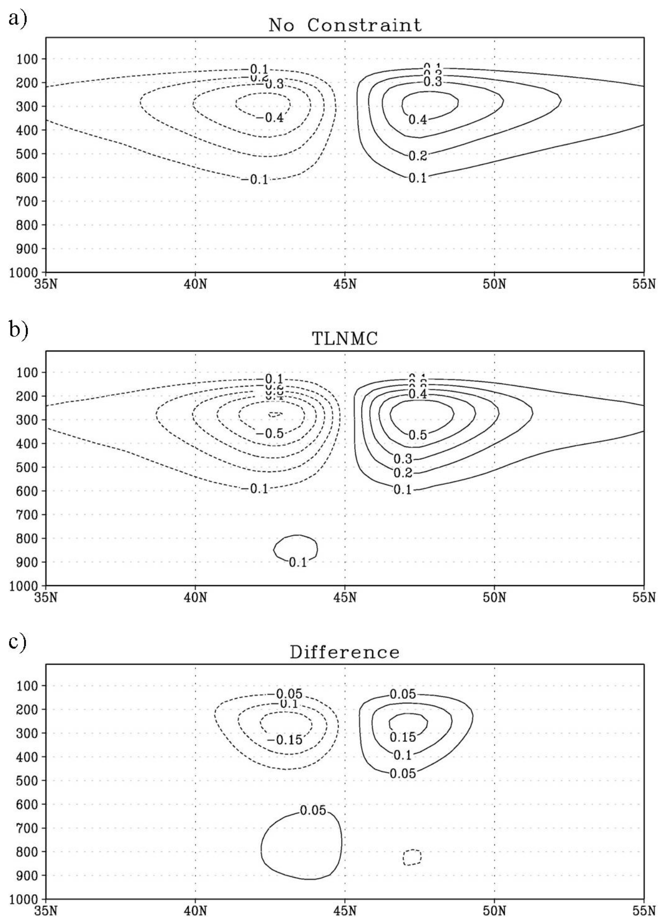
Single observation test (T observation)

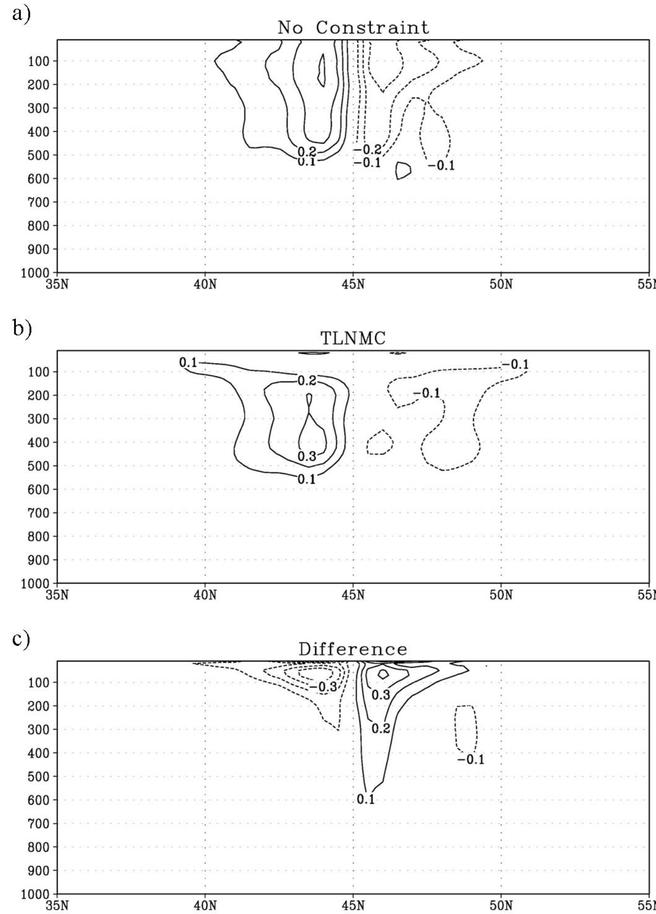
U wind

Ageostrophic U wind

From multivariate B

TLNMC corrects





Smaller ageostrophic component

Cross section of zonal wind increment (and analysis difference) valid at 12Z 09 October 2007 for a single 500 hPa ***temperature*** observation (1K O-F and observation error)

Adding Flow Dependence

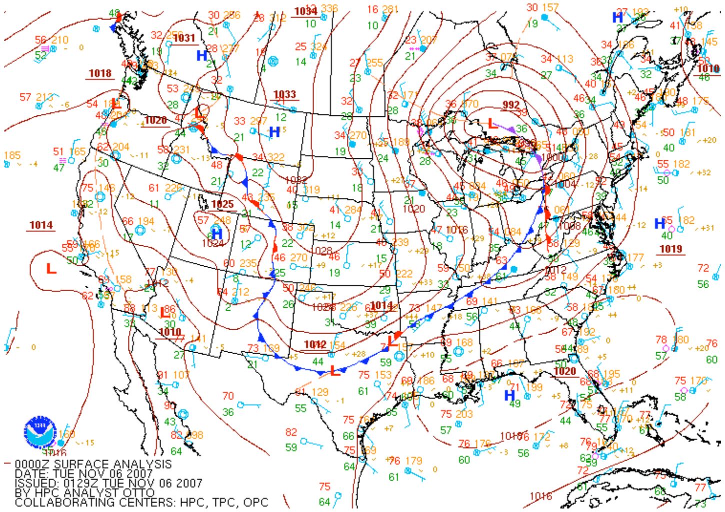
- One motivation for GSI was to permit flow dependent variability in background error

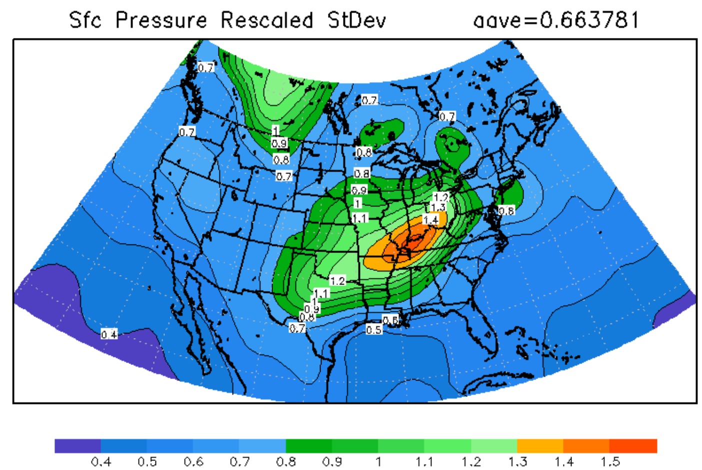
- Take advantage of FGAT (guess at multiple times) to modify variances based on 9h-3h differences
  - Variance increased in regions of large tendency
  - Variance decreased in regions of small tendency
  - Global mean variance ~ preserved

- Perform reweighting on streamfunction, velocity potential, virtual temperature, and surface pressure only

Currently global only, but simple algorithm that could easily be adapted for any application

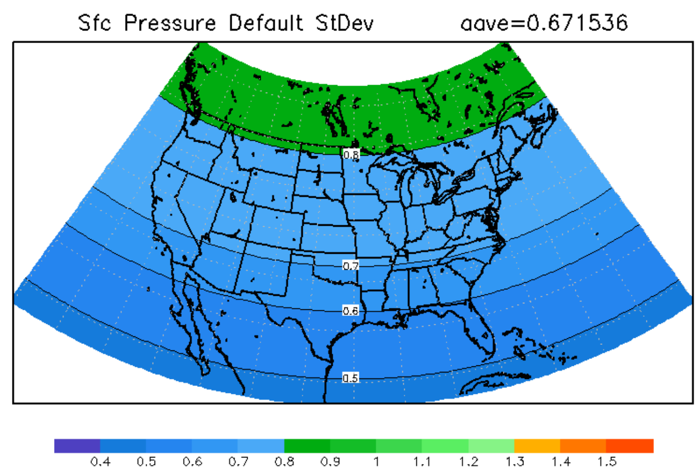
Variance Reweighting





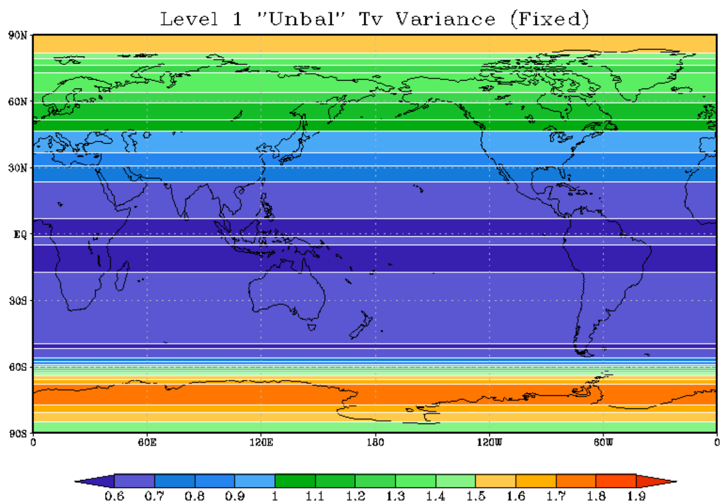
Surface pressure background error standard deviation fields

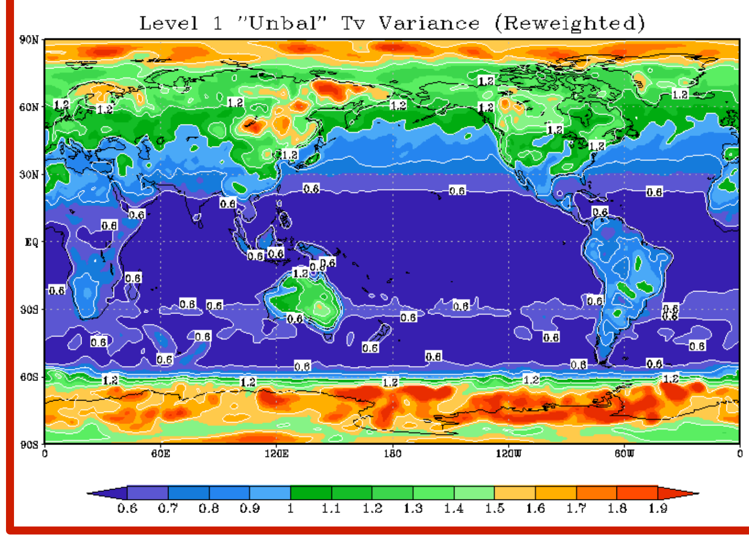
a) with flow dependent re-scaling
b) without re-scaling



Valid: 00 UTC November 2007

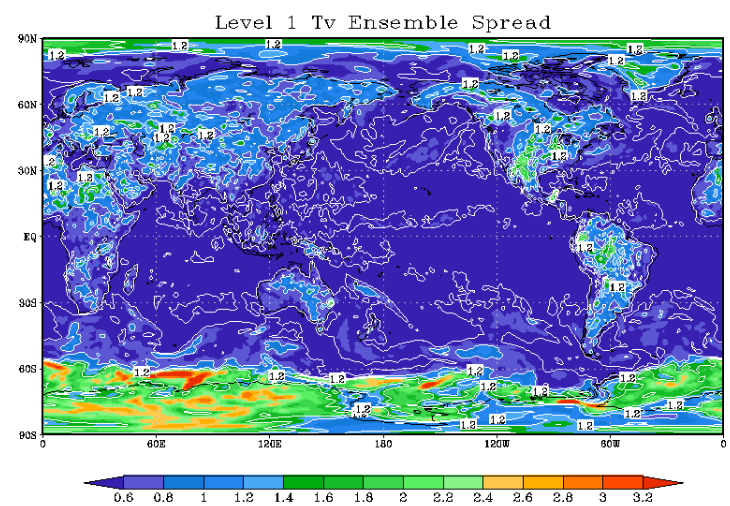
Variance Reweighting





- Although flow-dependent ***variances*** are used, confined to be a rescaling of fixed estimate based on time tendencies
  – No cross-variable or length scale information used
  – Does not necessarily capture 'errors of the day'

- Plots valid 00 UTC 12 September 2008



Hybrid Variational-Ensemble

- Incorporate ensemble perturbations directly into variational cost function through extended control variable
  - Lorenc (2003), Buehner (2005), Wang et. al. (2007), etc.

$$J(\mathbf{x}_f', \alpha) = \beta_f \frac{1}{2} (\mathbf{x}_f')^T \mathbf{B}^{-1} (\mathbf{x}_f') + \beta_e \frac{1}{2} (\alpha)^T \mathbf{L}^{-1} (\alpha) + \frac{1}{2} (\mathbf{y}_o' - \mathbf{H}\mathbf{x}_t')^T \mathbf{R}^{-1} (\mathbf{y}_o' - \mathbf{H}\mathbf{x}_t')$$

$$\mathbf{x}_t' = \mathbf{x}_f' + \sum_{n=1}^{N} (\alpha^n \circ \mathbf{x}_e^n) \qquad \frac{1}{\beta_f} + \frac{1}{\beta_e} = 1$$

β_f & β_e: weighting coefficients for fixed and ensemble covariance respectively

$\mathbf{x}_t$: (total increment) sum of increment from fixed/static **B** ($\mathbf{x}_f$) and ensemble **B**

α^n: extended control variable; $\mathbf{x}_e^n$:ensemble perturbations

L: correlation matrix [localization on ensemble perturbations]

2:30 GSI/ETKF Regional Hybrid Data Assimilation - Arthur Mizzi (MMM/NCAR)

Observation Errors

1. Overview
2. Adaptive Tuning

3DVAR Cost Function

$$J_{\mathrm{Var}}(\mathbf{x}') = \frac{1}{2}(\mathbf{x}')^{\mathrm{T}}\mathbf{B}_{\mathrm{Var}}^{-1}(\mathbf{x}') + \frac{1}{2}(\mathbf{Hx}' - \mathbf{y}_{\mathrm{o}}')^{\mathrm{T}}(\mathbf{R})^{-1}(\mathbf{Hx}' - \mathbf{y}_{\mathrm{o}}') + J_{\mathrm{c}}$$

- J : Penalty (Fit to background + Fit to observations + Constraints)
- **x**’ : Analysis increment ($\mathbf{x}_a - \mathbf{x}_b$) ; where $\mathbf{x}_b$ is a background
- $\mathbf{B}_{\mathbf{Var}}$: Background error covariance
- **H** : Observations (forward) operator
- **R : Observation error covariance (Instrument + Representativeness)**
  - **Almost always assumed to be diagonal**
- $\mathbf{y}_o$’ : Observation innovations/residuals ($\mathbf{y}_o$-$\boldsymbol{H}\mathbf{x}_b$)
- J_c : Constraints (physical quantities, balance/noise, etc.)

Tuning

- Observation errors contain two parts
  - Instrument error
  - Representativeness error

- In general, tune the observation errors so that they are about the same as the background fit to the data

- In practice, observation errors and background errors can not be tuned independently

Adaptive tuning

- Talagrand (1997) on $E[J(\mathbf{x}_a)]$
- Desroziers & Ivanov (2001)
  - $E[J_o] = \frac{1}{2}\,\mathrm{Tr}\,(\mathbf{I} - \mathbf{HK})$
  - $E[J_b] = \frac{1}{2}\,\mathrm{Tr}\,(\mathbf{KH})$
    - **K** is Kalman gain matrix
    - **H** is linearlized observation forward operator
- Chapnik et al.(2004)
  - robust even when **B** is incorrectly specified

Adaptive tuning

Tuning Procedure: $$J(\delta\mathbf{x}) = \frac{1}{\varepsilon_b^2} J_b(\delta\mathbf{x}) + \frac{1}{\varepsilon_o^2} J_o(\delta\mathbf{x})$$

Where ε_b and ε_o are background and observation error weighting parameters

$$\varepsilon_o = \sqrt{\frac{2J_o}{\mathrm{Tr}(\mathbf{I} - \mathbf{HK})}}$$

$$\mathrm{Tr}(\mathbf{I} - \mathbf{HK}) = N_{obs} - \left(\sum \xi \mathbf{R}^{-\frac{1}{2}} \mathbf{H} \delta\mathbf{x}_a \left(\mathbf{y} + \xi \mathbf{R}^{\frac{1}{2}} \right) + \sum \xi \mathbf{R}^{-\frac{1}{2}} \mathbf{H} \delta\mathbf{x}_a (\mathbf{y}) \right)$$

Where ξ is a random number with standard normal distribution (mean:0, variance:1)

Adaptive tuning

1) &SETUP

    oberror_tune=.true.

2) If Global mode:

    &OBSQC

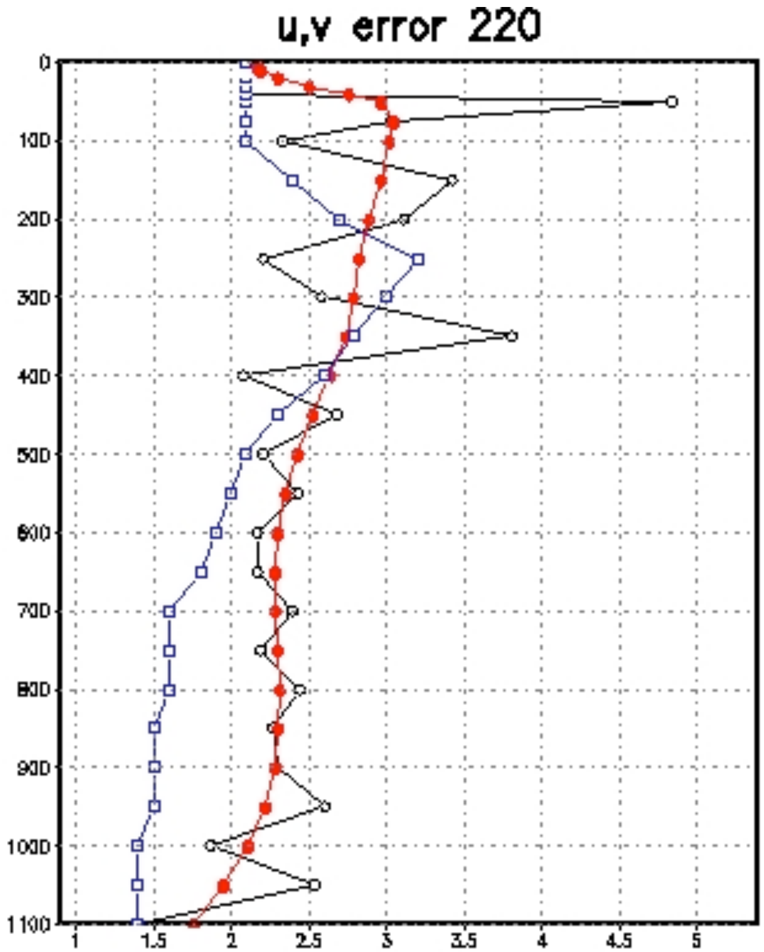
        oberrflg=.true.

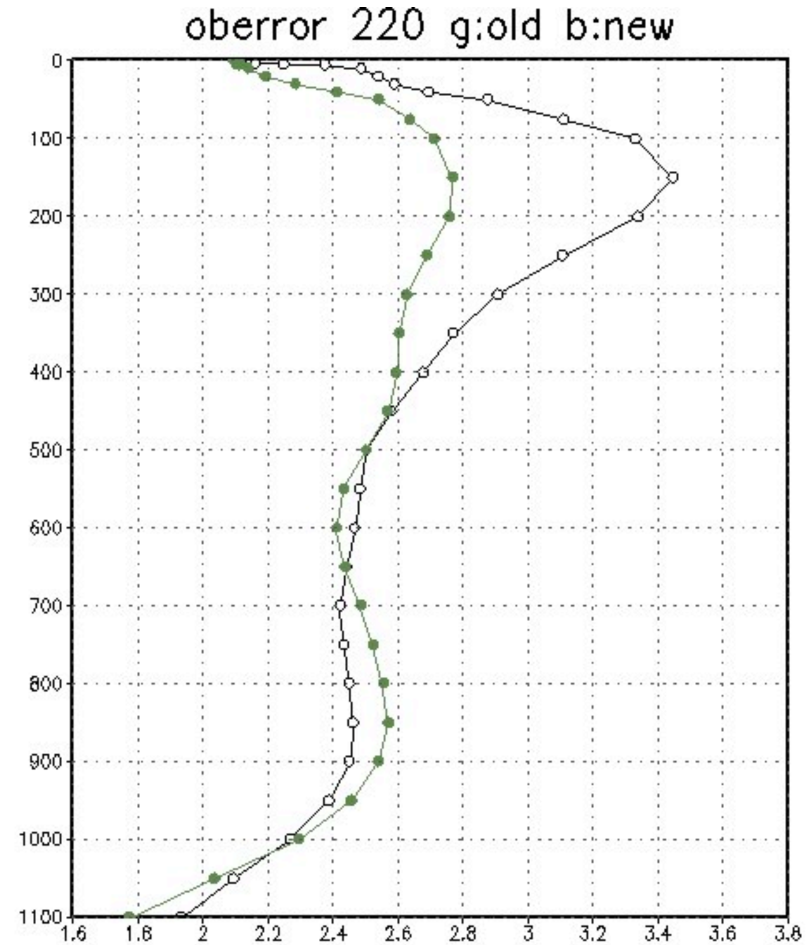
(Regional mode: oberrflg=.true. is default)

Note: GSI does not produce a 'valid analysis' under the setup

Aside: Perturbed observations option can also be used to estimate background error tuning (ensemble generation)!

Adaptive Tuning





Alternative: Monitoring Observations from Cycled Experiment

1. Calculate the covariance of observation minus background (O-B) and observation minus analysis (O-A) in observation space

   (O-B)*(O-B) , (O-A)*(O-A), (O-A)*(O-B), (A-B)*(O-B)

2. Compare the adjusted observation errors in the analysis with original errors

3. Calculate the observation penalty ((o-b)/r)**2)

4. Examine the observation regions

Summary

- Background error covariance
  - Vital to any data assimilation system
  - Computational considerations
  - Recent move toward fully flow-dependent, ensemble based (hybrid) methods

- Observation error covariance
  - Typically assumed to be diagonal
  - Methods for estimating variance are well established in the literature

- Experience has shown that despite all of the nice theory, error estimation and tuning involves a lot of trial and error