Computational issues for a variational system

~3d-var with situation-dependent background error~

Yoshiaki Sato[1]

JMA visiting NCEP/EMC[1]

Manuel S. F. V. De Pondeca[2], R. James Purser[2],

David F. Parrish[3] and John C. Derber[3]

SAIC at NCEP/EMC[2], NCEP/EMC[3]

WWRP/THORPEX Workshop on 4D-var and Ensemble Kalman Filter inter-comparisons

10-13 November 2008, Buenos Aires, Argentina

Contents

- Introduction
  - NCEP GSI
- Background error in the GSI
  - Recursive filter
- Anisotropic (situation dependent) background error
  - Formulation and sample results
- Summary

NCEP GSI

- GSI, The Gridpoint Statistical Interpolation (Wu et al. 2002)
  - NCEP's regional and global data assimilation system
  - A variational system formulated in the model grid point space
    - Currently 3D-var
      - FOTO (First Order Time-extrapolation to Observations) is available as "simplified 4D-var"
    - 4D-var is being developed by NASA/GMAO and NCEP/EMC
  - The background error is generated by recursive filters
    - Efficient numerical technique with considerable flexibility
      - Horizontally isotropic background error is used in operational version.
        - » It varies with latitude and height for each analysis variable.
      - Situation-dependent background error is available with anisotropic recursive filters.

Basic Formulation

- The cost function and its gradient for the variational method are:

$$2J(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x} + (\mathbf{HMx} - \mathbf{y_o})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HMx} - \mathbf{y_o})$$

$$\nabla_{\mathbf{x}} J(\mathbf{x}) = \mathbf{B}^{-1}\mathbf{x} + \mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HMx} - \mathbf{y_o})$$

  – where,
    - $\mathbf{x}$: Increment, $\mathbf{y_o}$: Observation departure from the first guess
    - $\mathbf{B}$: Background error covariance matrix
    - $\mathbf{R}$: Observation error covariance matrix
    - $\mathbf{H}$: Observation operator
    - $\mathbf{M}$: Time integration operator (=$\mathbf{I}$ for 3Dvar)

    * Penalty terms are omitted

Computational issues:

i) $\mathbf{B}$ matrix is too large to be defined directly.

ii) In 4Dvar, the operators $\mathbf{M}$, $\mathbf{M}^{\mathrm{T}}$ each require at least the cost of 1 model integration.

→ How can we achieve 4D-var benefit w/o $\mathbf{M}$?

Idea to avoid $\mathbf{M}$ operation (I)

- FOTO (First Order Time-extrapolation to Observations, Rancic et al. 2008)
  - Use the (filtered) first order time tendencies of the guess to estimate (extrapolate) the first guess at the given observation time.

$$2J(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x} + (\mathbf{HMx} - \mathbf{y_o})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HMx} - \mathbf{y_o})$$

$$\approx \mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x} + \left(\mathbf{H}\left(\mathbf{x} + \frac{\partial \mathbf{x}}{\partial t}\delta t\right) - \mathbf{y_o}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{H}\left(\mathbf{x} + \frac{\partial \mathbf{x}}{\partial t}\delta t\right) - \mathbf{y_o}\right)$$

$$\because \mathbf{Mx} \approx \mathbf{x} + \frac{\partial \mathbf{x}}{\partial t}\delta t, \quad \frac{\partial \mathbf{x}}{\partial t}:$$ Simplified tangent linear model, gravity oscillations are filtered out (Kleist et al, 2008)

- Since it does not need time integration of the model (and its adjoint), the computational cost is not expensive.

Idea to avoid $\mathbf{M}$ operation (II)

- Situation-dependent background error
  - The analytical solution of the cost function minimization

$$\mathbf{x} = \mathbf{B}\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{M}\mathbf{B}\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{y_o} \quad \text{:general form}$$

$$\mathbf{x} = \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{y_o} \quad \text{:for 3Dvar / 4Dvar at } t=t_{start}$$

$$\mathbf{M}\mathbf{x} = \mathbf{M}\mathbf{B}\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{M}\mathbf{B}\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{y_o} \quad \text{:for 4Dvar at } t=t_{end}$$

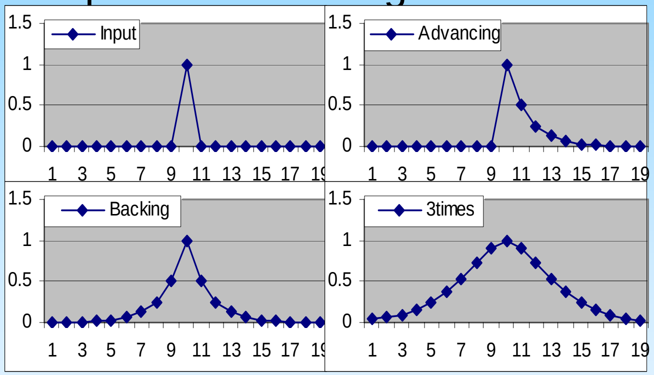
where t_{start}: start of the assimilation window, t_{end}: end of the window

- The analytical solution of the 4Dvar at $t=t_{end}$ is equivalent to the solution of 3Dvar with its $\mathbf{B}$ replaced by $\mathbf{MBM}^{\mathrm{T}}$.
  - The $\mathbf{MBM}^{\mathrm{T}}$ is the $\mathbf{B}$ modified to take into account the forward and adjoint model integration.
- Here, if we could estimate the $\mathbf{MBM}^{\mathrm{T}}$ appropriately without using the $\mathbf{M}$, we might then be able to accomplish the benefit of the flow dependent background errors in 4Dvar and EnKF with the less expensive 3Dvar.

Thus we are focusing on the optimal estimation of the matrix $\mathbf{B}$
And its efficient computation

Background Error in GSI

- The full **B** is too large to be defined directly.
  - GSI uses recursive filters for **Bx** (*) operation.
    - The recursive filter is an efficient numerical technique for simulating diffusion along a grid line (Purser et al. 2003a).



$$q_i = \beta p_i + \sum_{j=1}^{n} \alpha_j q_{i-j}$$ advancing step

$$s_i = \beta q_i + \sum_{j=1}^{n} \alpha_j s_{i+j}$$ backing step

  - p: input, α, β: coefficients, q: intermediate s: output

    - 3D isotropic covariance is simulated by applying the recursive filters along three grid coordinate lines.

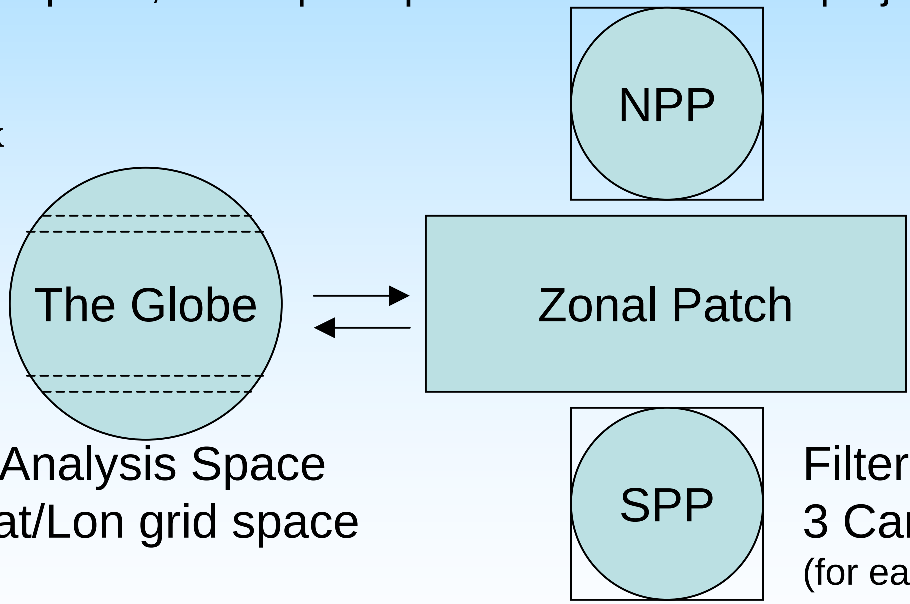
* With the help of the Preconditioned Conjugate Gradient minimization algorithm, the GSI does not need $\mathbf{B}^{-1}\mathbf{x}$ but **Bx.**

Background Error in GSI

- In Global Case
  - The recursive filter is designed for “standard gridded data”.
  - the GSI uses 3 different filter spaces to avoid singularity problem.
    - Zonal patch – Cylindrical projection
    - North polar patch, South polar patch – Polar stereo projection

$$\mathbf{Bx} = \mathbf{P}\begin{pmatrix} \mathbf{B}_{\mathrm{npp}} & & \mathbf{0} \\ & \mathbf{B}_{\mathrm{zp}} & \\ \mathbf{0} & & \mathbf{B}_{\mathrm{spp}} \end{pmatrix}\mathbf{P}^{\mathrm{T}}\mathbf{x}$$

P: map projection operator



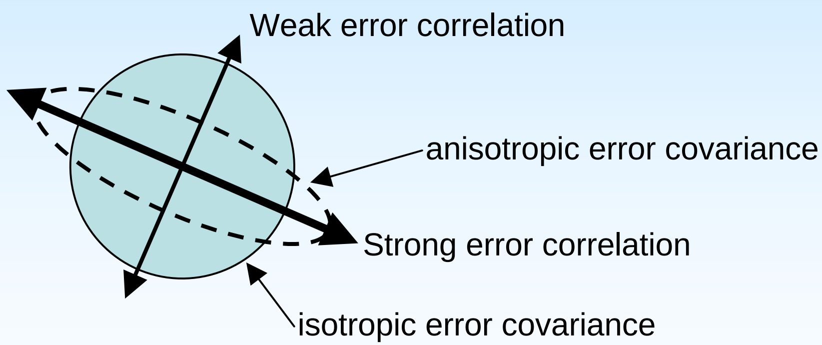
Analysis Space
Lat/Lon grid space

Filter Spaces:
3 Cartesian grid spaces
(for each analysis variables)

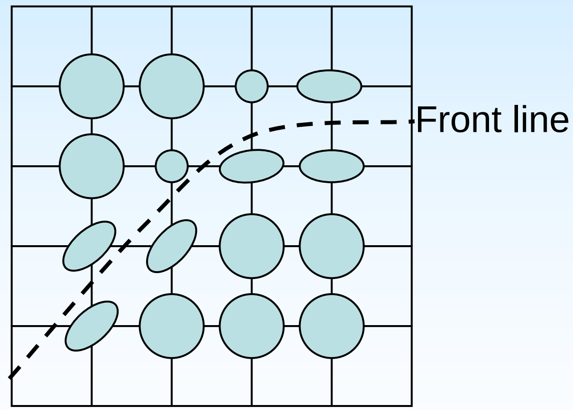
Remaining Issue:
The observational information can not spread beyond the boundary of each patch(I will mention this issue in the later slide)

Anisotropic recursive filter

- For the situation-dependent background error, we need the capability to modify **B** arbitrarily.
  - Anisotropic recursive filter (Purser et al. 2003b)
    - Briefly speaking, the recursive filters are applied to six directions for 3D case (Hexad algorithm) with different diffusion parameters.
    - Anisotropic covariance shape of arbitrary aspect ratio and orientation are represented by the (smooth) centered and normalized second-moment "aspect-tensor" of spatial dispersion.



Schematic view of local aspect tensor



So the next problem is "How do we define the aspect tensor ?"

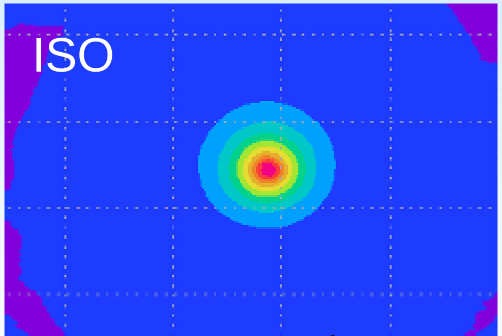
Aspect Tensor Definition

- Isotropic component of the aspect tensor
  - To simulate the isotropic mode with the anisotropic code, we set the aspect tensor $\mathbf{S}$ as:

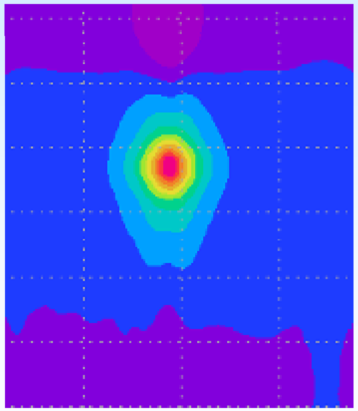
$$\mathbf{S}_{iso}^{-1} = diag\left\{L_h^{-2}, L_h^{-2}, L_v^{-2}\right\}$$

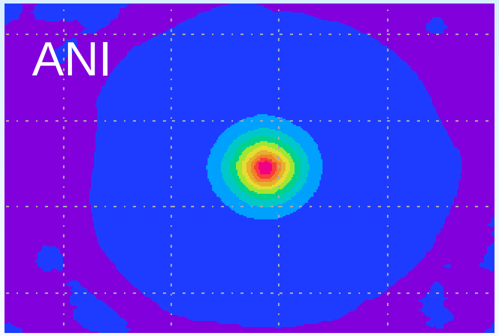
L_h: correlation length scale for Horizontal direction
L_v: correlation length scale for Vertical direction

- Since we don't yet have a method to tune the parameters to simulate the isotropic mode now, we did a subjective calibration in advance.
  - Focusing on just around the signal point, the tail was shorter in the comparison below:

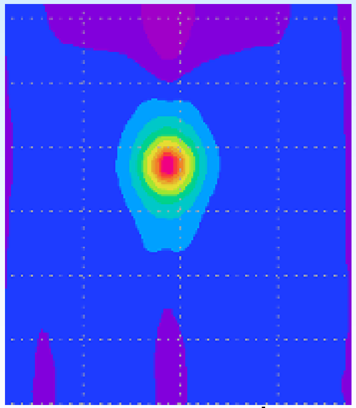


xy-cross section


xz-cross section

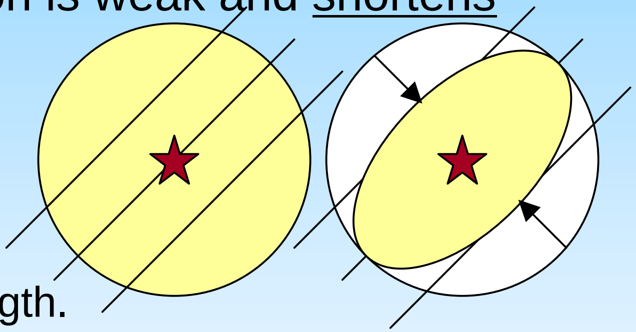


xy-cross section


xz-cross section

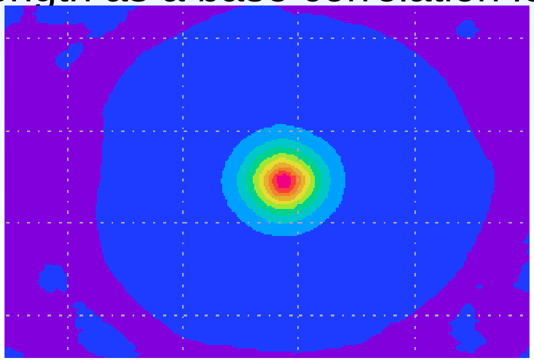
Aspect Tensor Definition (I)

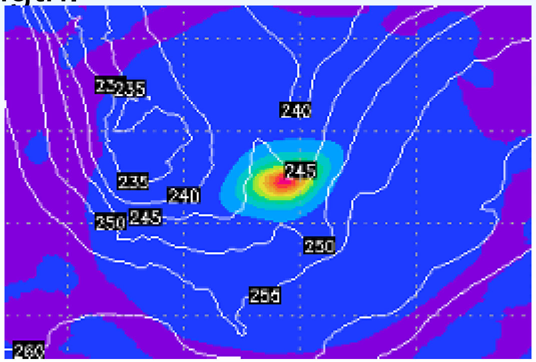
- Anisotropic component of the aspect tensor (Riishøjgaard, 1997)
  - Riishøjgaard method uses some field q as an index of the error correlation.
  - If the gradient is large, it assumes the correlation is weak and <u>shortens</u> the correlation length for the direction.



$$\mathbf{S}_{ani}^{-1} = \alpha \mathbf{S}_{iso}^{-1} + \frac{(\nabla q)(\nabla q)^T}{L_q^2}$$

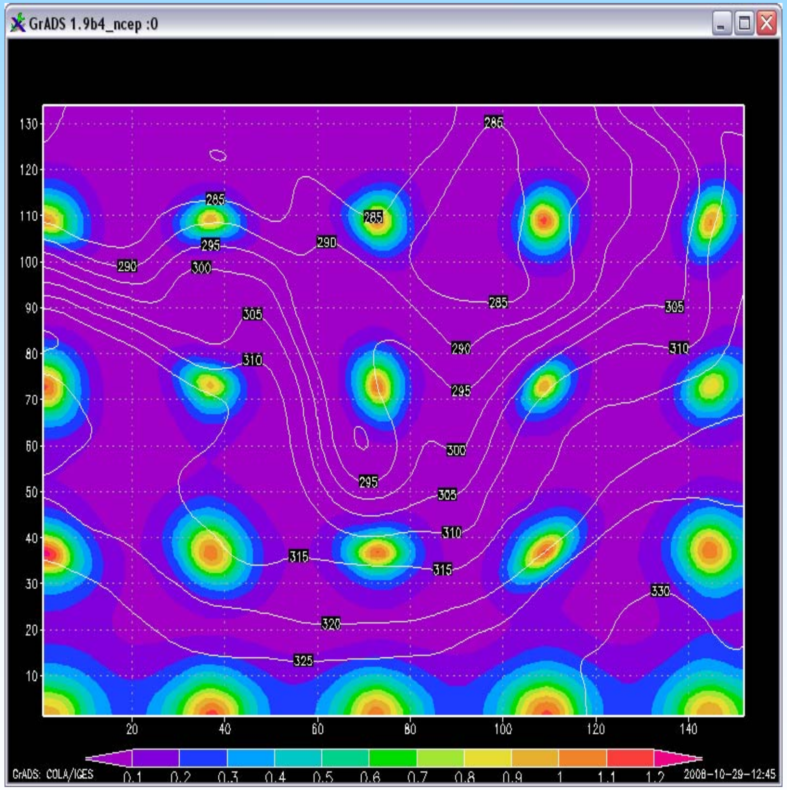
- The 2nd term works to shorten the correlation length.
  - L_q is the "function correlation length"
- So we set an inflation factor α=1/1.2 to stretch the isotropic correlation length as a base correlation length.





Sample covariance field

- Anisotropic covariances sample,
  - estimated by the Riishøjgaard method (index field q=potential temperature)



Aspect Tensor Definition (II)

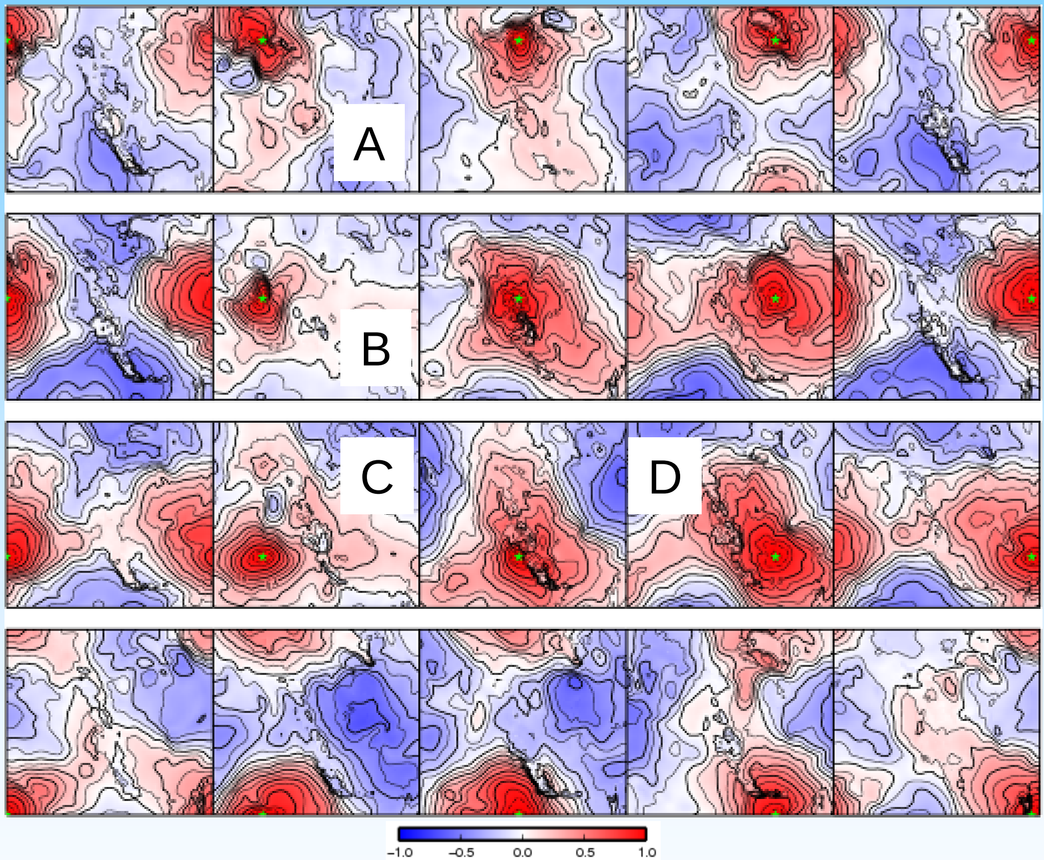
- Anisotropic component of the aspect tensor (Ensemble based)
  - Supposing that "the ensemble perturbation P represents background error structures", we can use the P as q (index field). Taking the summation for the whole ensemble, we get:

$$\mathbf{S}_{ani}^{-1} = \alpha \mathbf{S}_{iso}^{-1} + \beta \frac{\sum_{i=1}^{N} \left(\nabla P_i\right)^T \left(\nabla P_i\right)}{\sum_{i=1}^{N} P_i^2} \qquad \text{where} \qquad \nabla P_i = \left(\frac{\partial P_i}{\partial x}, \frac{\partial P_i}{\partial y}, \frac{\partial P_i}{\partial z} \right)$$

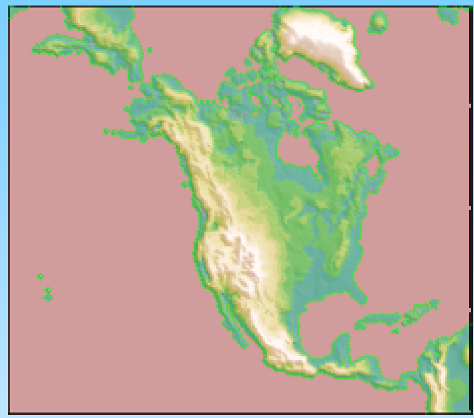
    - The denominator plays the role of a "function correlation length"
    - We might be able to use α=0, provided the determinant of the second term is not 0. However, it would possibly lead to unrealistically long correlation lengths. So we are using α=1/1.2 and β=1 now.

→ To test this formulation,
We compared the estimated covariance with the true perturbation correlation

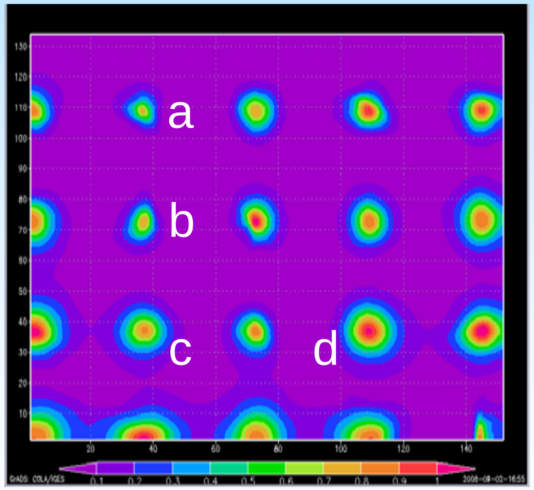
Ensemble perturbation correlation and GSI estimated covariance (NAM)



True ensemble perturbation correlation against some anchor points (denoted by green star). All panels show the full NAM domain



North American Mesoscale (NAM) Model Domain

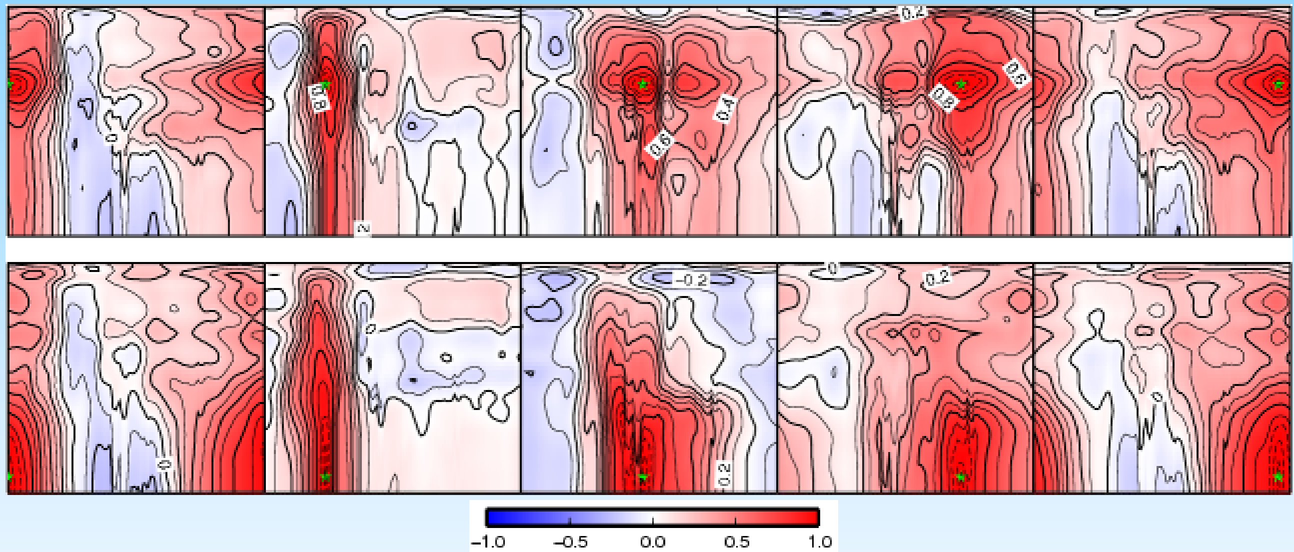


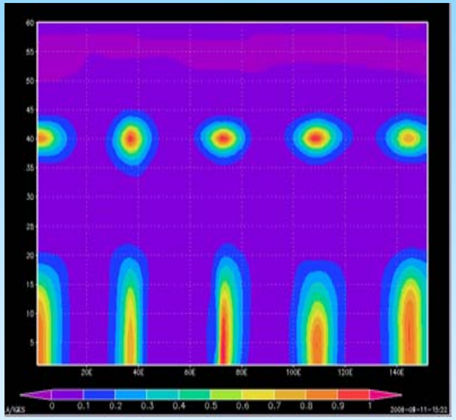
GSI estimated covariances

Roughly speaking, The GSI estimated covariances showed reasonable correspondence with the ensemble perturbation correlation in their extents and shapes.

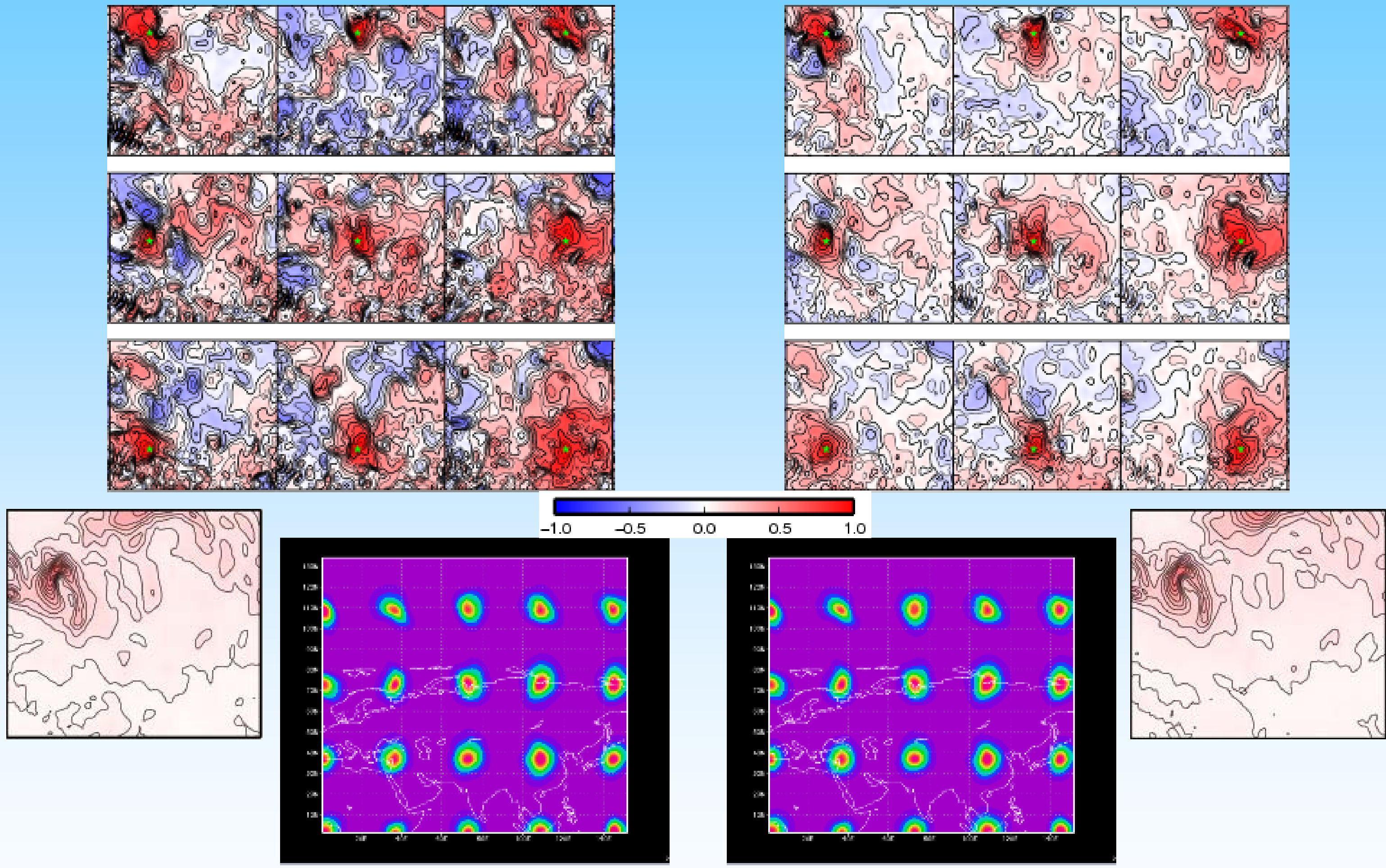
Ensemble perturbation correlation and GSI estimated covariance

- Vertical Cross Section





Ensemble Size Dependency

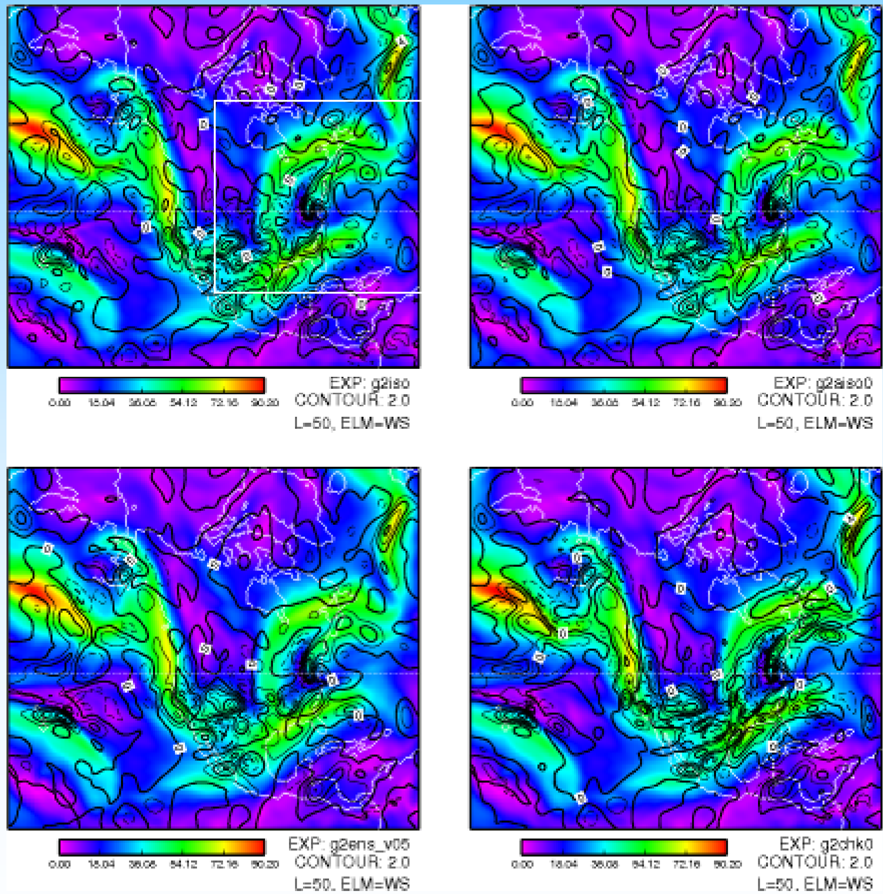


20-member 80-member

Looking at the true ensemble perturbation correlation field, 20-member ensemble shows large sampling error (high correlation far from the anchor point). But the pattern just around the anchor points is not so different from the 80-member ensemble. And our formulation uses the information near the anchor point only. So the difference of the estimated covariance is not large.

Sample innovation field

- Wind Speed at the 50th level (18UTC Dec. 26, 2007), NAM domain



exp1: g2iso
isotropic mode

exp2: g2aiso0
anisofilter w/o anisotropy

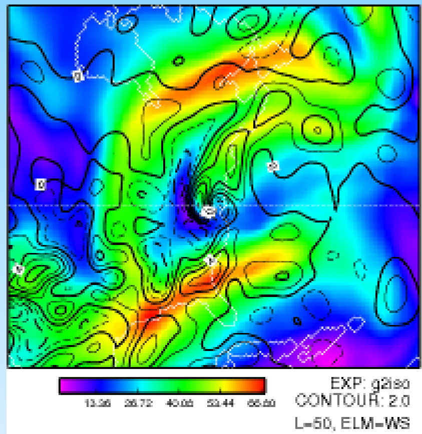
exp3: g2ens_v05
ensemble based anisotropy
for ST and VP
vertical calibration factor=0.5

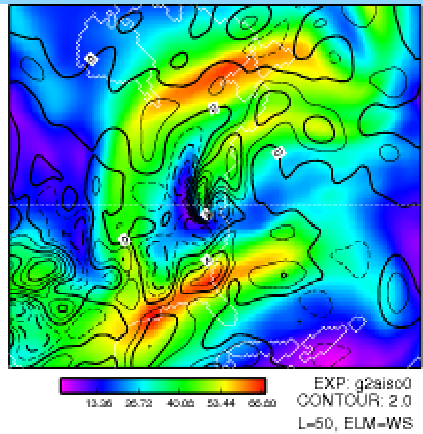
exp4: g2chk0
horizontally surrounding grid
ensemble based anisotropy
for ST and VP

Sample innovation field

- Wind Speed at the 50^{th} level (18UTC Dec. 26, 2007)

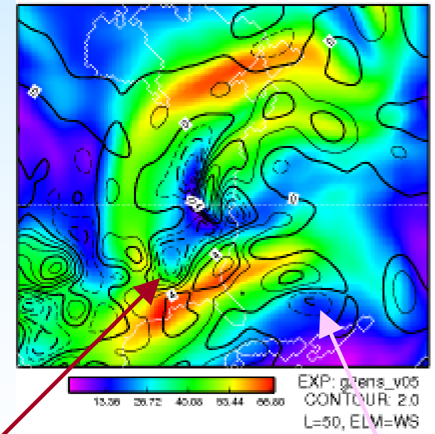
exp1: g2iso
isotropic mode

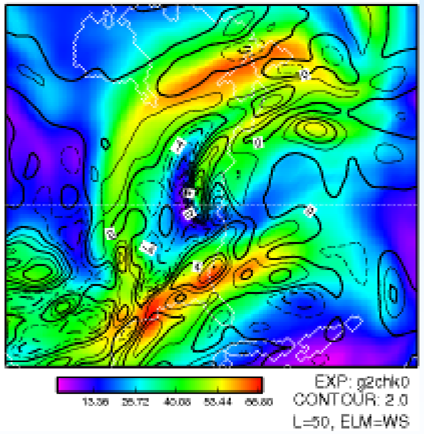




exp2: g2aiso0
anisofilter w/o anisotropy

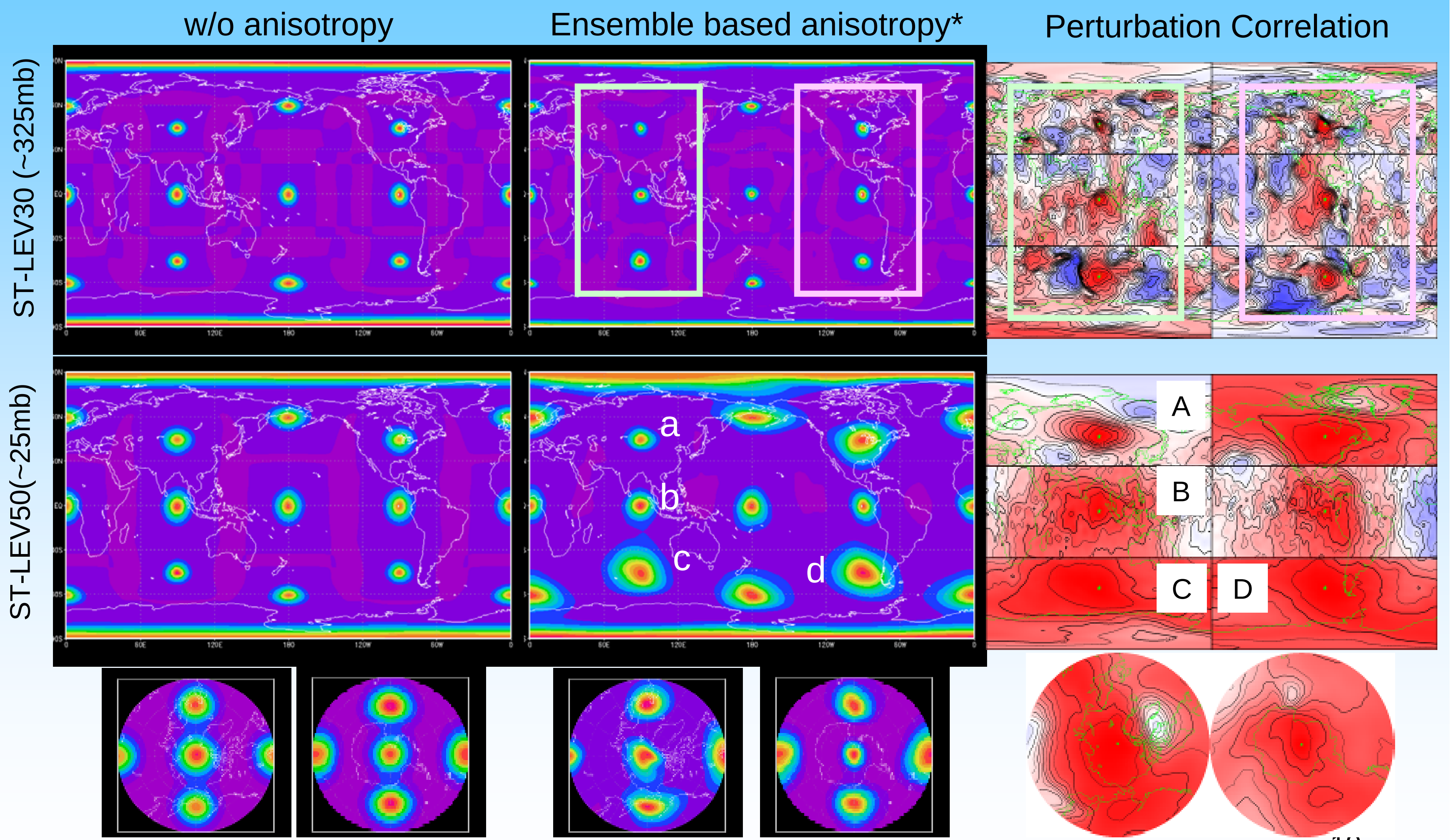
exp3: g2ens_v05
ensemble based anisotropy
for ST and VP
vertical calibration factor=0.5





exp4: g2chk0
horizontally surrounding grid
ensemble based anisotropy
for ST and VP

In Global Mode

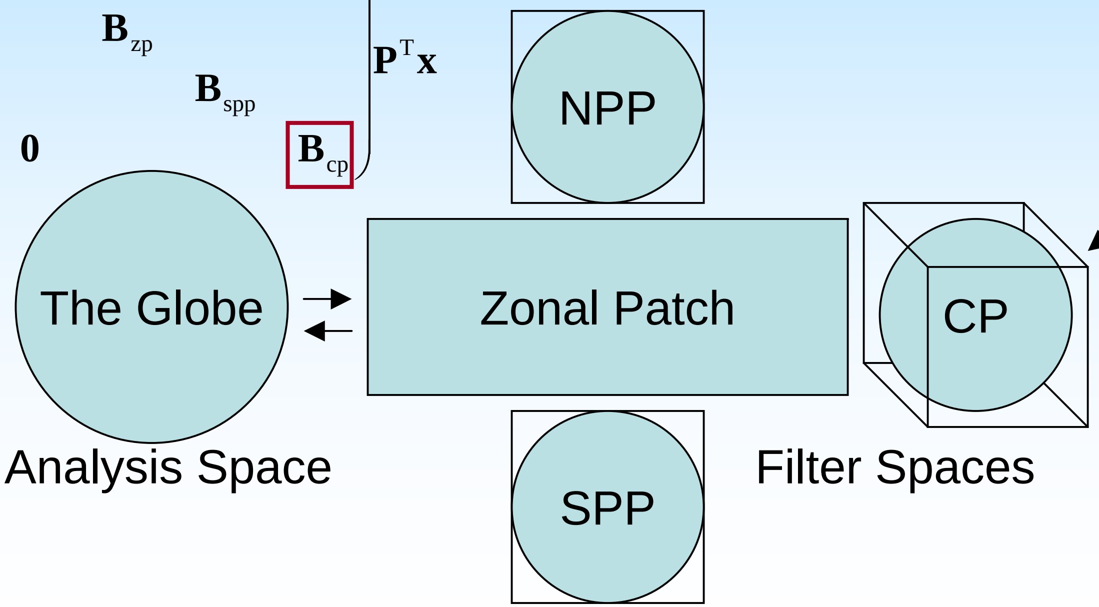


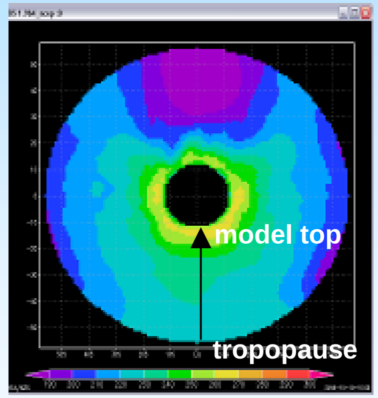
* Note: a modified formulation has been used in this case to allow for very long correlation lengths.

Remaining Issue in the Global Case

- Patch boundary problem
  - The observational information can not spread beyond the boundary of each patch, though there are planetary scale correlation length in the stratosphere.
  - Our idea for this issue: To add Cubic patch for the stratosphere
    - Under development

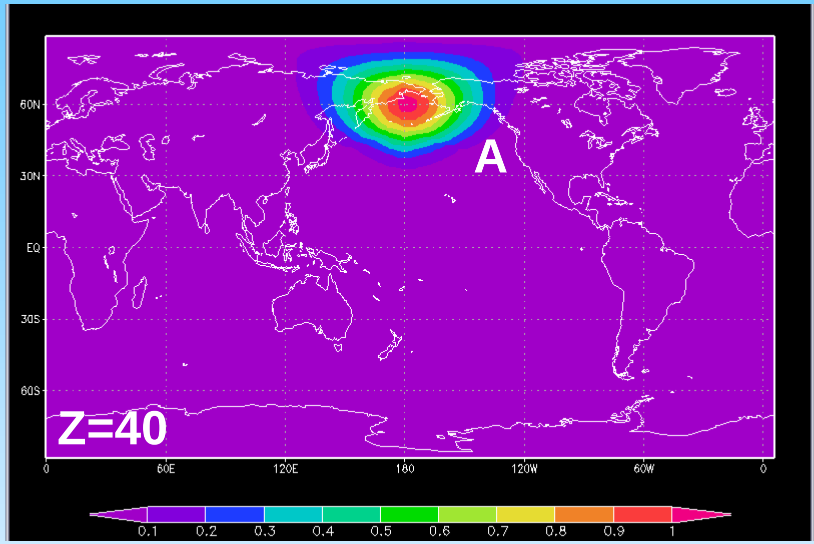
$$\mathbf{Bx} = \mathbf{P}\begin{pmatrix} \mathbf{B}_{npp} & & & \mathbf{0} \\ & \mathbf{B}_{zp} & & \\ & & \mathbf{B}_{spp} & \\ \mathbf{0} & & & \boxed{\mathbf{B}_{cp}} \end{pmatrix}\mathbf{P}^{T}\mathbf{x}$$

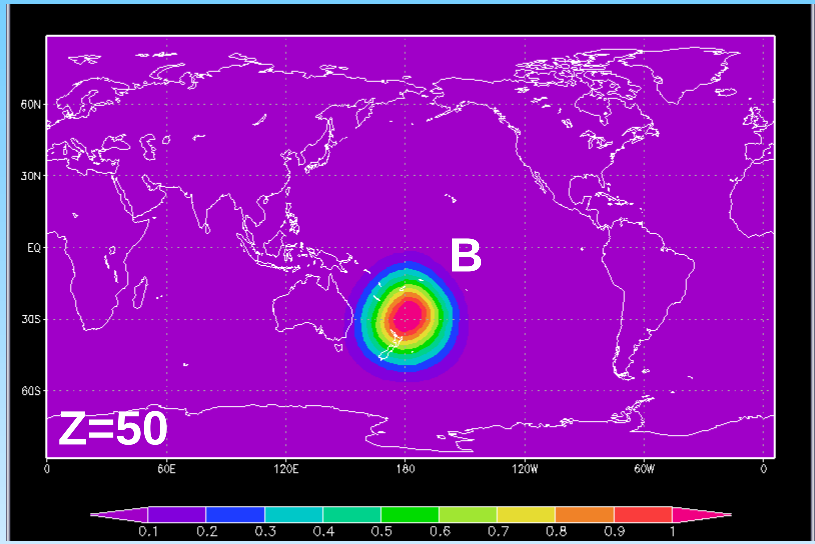


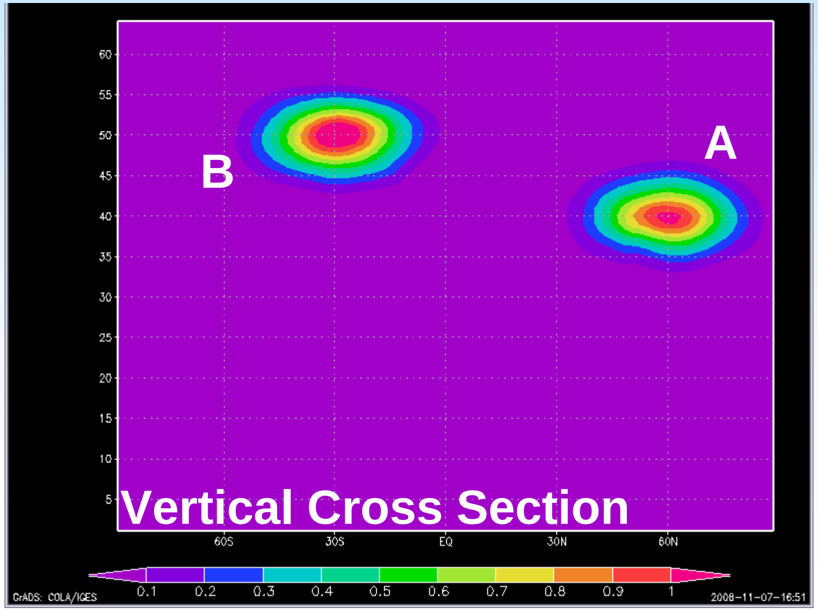


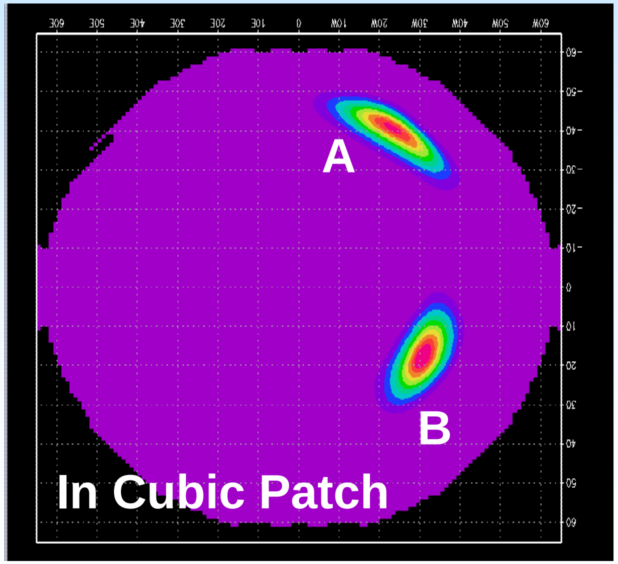
Model bottom:
higher resolution
Model top
Lower resolution

Sample Covariances in Cubic Patch









Horizontally Isotropic Covariance Case

Summary

- For computational efficiency,
  - we are improving the 3Dvar to achieve most of the 4Dvar benefits without the expensive time integration operator.
    - FOTO
    - Flow Dependent Background Error
  - With ensemble data input, we were able to estimate plausible flow dependent background errors without applying a time integration operator.
  - Though the forecast skill is still only comparable to the experiment which used isotropic background error case (not shown), we think that there is room to improve the analysis with the improvement of background error specifications.

To Do

- Calibration issues
  - We are using subjective calibration right now to simulate the covariance.
  - Objective method would be better
- Alternate ensemble data input
  - We used NCEP global ensemble forecast data for this test.
  - But, we could use other ensemble data, such as from EnKF.
- Global planetary scale correlation:
  - We are developing an additional patch for the planetary scales.
- And…
  - The situation dependent background error can be combined with 4D-Var to combine the advantages of both.

References

- Kleist et al 2008: Improving incremental balance in the GSI 3DVAR analysis system, *Mon. Wea. Rev.*, to be published.
- Purser et al. 2003a: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1524-1535.
- Purser et al. 2003b: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.*, **131**, 1524-1535.
- Riishøjgaard 1998: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus*, **50A**, 42-57.
- Rancic et al. 2008: The development of the first-order time extrapolation to the observation (FOTO) method and its application in the NCEP global data assimilation system. Preprints, *12th Conf. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans and Land Surfaces*, New Orleans, LA, Amer. Meteor. Soc., J6-1.
- Wu et al. 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905-2916.