www.ec.gc.ca

Modelling of Ensemble Covariances

Meteorological Research Division
Environment Canada

Mark Buehner
BIRS Data Assimilation Workshop
February 2008

Environment Canada Environnement Canada

Canada

Background

- Canadian NWP centre currently has both a global 4D-Var (for deterministic forecasts) and EnKF (for probabilistic forecasts)
- Provides good opportunity to compare two approaches and to evaluate use of flow-dependent ensemble background-error covariances in a variational system
- Current approaches for modelling background-error covariances in 4D-Var and EnKF represent two extreme cases:
  - 4D-Var: horizontally homogeneous, nearly temporally static
  - EnKF: independently estimated at each grid-point and analysis time
- Unlikely that either approach is optimal, best approach probably somewhere in between
- Therefore, both systems could be improved with a more general approach to covariance modelling (focus is on correlations in following)

Environment Canada Environnement Canada

Canada

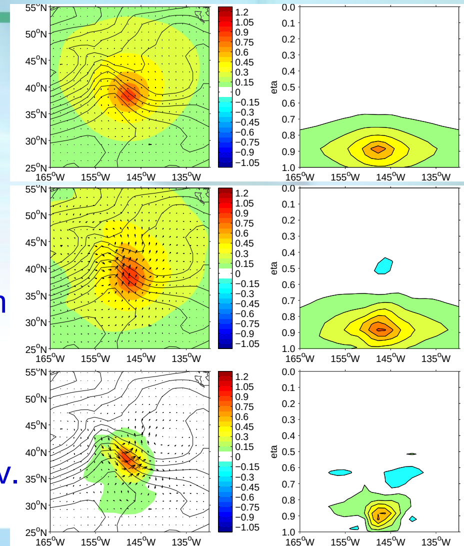
Comparison of Covariances used in 3/4D-Var and EnKF

- Corrections to T and UV in response to a single T obs near the surface
- Black contours show background T
- EnKF error covariances from 128 ensemble members

3D-Var

4D-Var (obs at end of 6h window)

EnKF error cov.



Outline

- Demonstrate complementary effects of spatial and spectral localization applied to ensemble-based error correlations
- Implementation issues in realistic NWP variational assimilation systems
- Simpler approach to incorporate limited amount of heterogeneity

Sampling error in ensemble-based error covariances

- Test ability of ensemble-based covariances to reproduce "true" covariances as function of ensemble size and spatial localization
- Spatial localization:

$$B_{grid\_loc} = B_{samp} \circ L_{grid}$$

where L is a simple "correlation" matrix with monotonically decreasing values as a function of separation distance used to do the localization

- Use operational B matrix as "truth" (homog/isotr. correlations for main analysis variables), generate ensemble members:

$$e^k = B^{1/2} \varepsilon^k \qquad \text{where } \varepsilon^k = N(0,\mathbf{I})$$

- Final value of J_o (all operational data) used as simple measure of accuracy of ensemble-based covariances: ability to fit to observations

Environment Canada Environnement Canada

Canada

Effect of ensemble size and spatial localization on sampling error

Final value of Jo (normalized by value from using "true" **B**) as a function of ensemble size and localization radii:

Localization radii		**Ensemble size**			
Horizontal (km)	**Vertical (ln(P))**	**32**	**128**	**512**	∞
∞	∞	3.15	3.10	2.98	1.00
10 000	∞	2.72	2.30	1.77	0.96
2 800	∞	2.09	1.46	1.12	0.84
10 000	**2**	2.23	1.73	1.31	0.94
2 800	**2**	1.47	1.11	0.97	0.82

Environment Canada Environnement Canada

Canada

Spectral correlation localization

- What happens if same type of localization is applied to the correlation matrix in spectral space?
  - A diagonal correlation matrix in spectral space corresponds with globally homogeneous correlations
  - Represents an extreme case of spectral correlation localization
  - More moderate spectral localization should result in correlations with an intermediate amount of heterogeneity
- Was shown that localization of correlations in spectral space (multiplication) is equivalent with spatial averaging of correlations in grid-point space (convolution)
- Averaging of correlations over a **local** region should be better than either globally homogeneous or independent for each grid point:
  - reduced sampling error through averaging, but
  - still maintain most of spatial/flow dependence of correlations

Environment Canada / Environnement Canada — Canada

Spectral correlation localization

- Localization of correlations in spectral space (multiplication):

$$S\,B_{spec\_loc}\,S^T = \left(S\,B_{samp}\,S^T\right)\circ L_{spec}$$

where

S is spectral transform,

L_{spec} is a "correlation" matrix with monotonically decreasing values as a function of the absolute difference in wavenumber

- Spatial averaging of correlations in grid-point space (convolution):

$$B_{spec\_loc}(x_1,x_2) = \int B_{samp}(x_1+s,x_2+s)\,\mathcal{L}_{spec}(s)\,ds$$

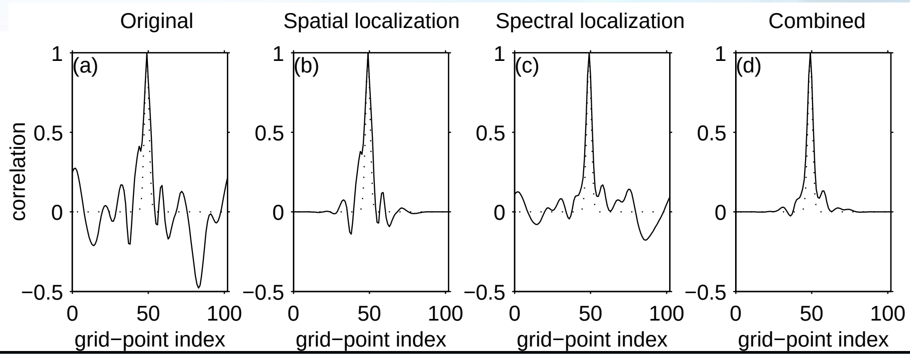
where

$\mathcal{L}_{spec} = (S^{-1}\,L_{spec}\,S^{-T})$ assuming L_{spec} is homogeneous and isotropic in wavenumber space

Environment Canada / Environnement Canada — Canada

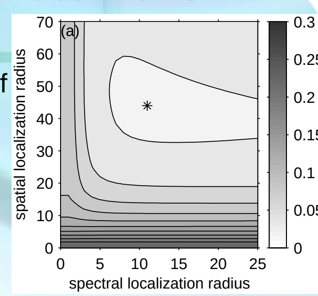
Spatial and spectral correlation localization

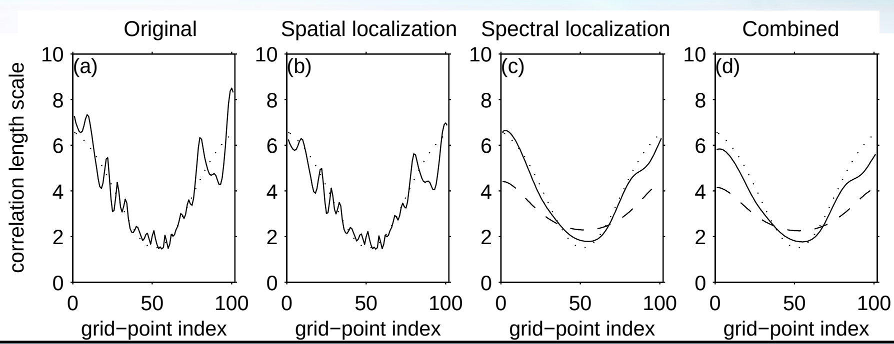
- Idealized 1-D example using prescribed "true" heterogeneous correlations and estimated correlations from 30 realizations
- Spatial localization cannot improve short-range correlations
- Spectral localization cannot remove long-range spurious correlations
- Combination seems to give best result



Spatial and spectral correlation localization

- For this example, a unique optimal combination of spatial and spectral localization exists (minimum rms error of correlations)
- Spectral localization dramatically improves local estimate of correlation length scale: $(-d^2C/dx^2)^{-1/2}$
- With too much spectral localization, start to loose heterogeneity (dashed)





Ensemble-based error covariances in 3D-Var

- Implementation in preconditioned variational analysis:
  - no localization: elements of control vector determine **global** contribution of each ensemble member to the analysis increment:

    $\Delta x = \sum (e^k - \langle e \rangle)\, \xi^k$ (ξ^k is a scalar)

  - spatial localization: elements of control vector determine **local** contribution of each ensemble member to the analysis increment:

    $\Delta x = \sum (e^k - \langle e \rangle) \circ (L_{grid}^{1/2}\, \xi^k)$ (ξ^k is a vector)

  - in each case, J_b is Euclidean inner product:

    $J_b = 1/2\ \xi^T \xi$

  - can also combine with standard **B** matrix:

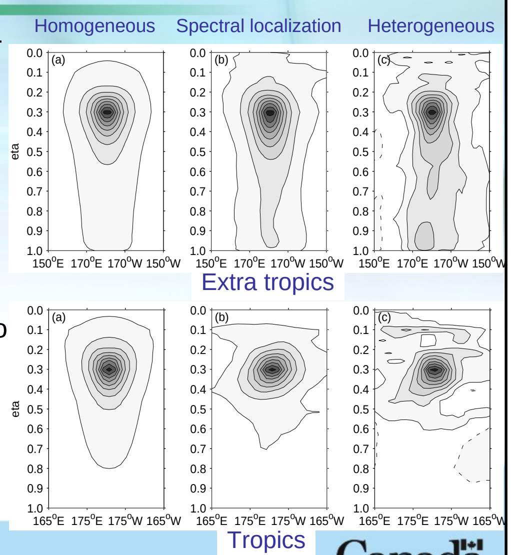
    $\Delta x = \beta^{1/2} \sum (e^k - \langle e \rangle) \circ (L_{grid}^{1/2}\, \xi^k) + (1-\beta)^{1/2} B^{1/2}\, \xi_{HI}$

Spectral correlation localization

- Apply in variational system, similar technique as spatial localization
- Elements of control vector determine local contribution (in spectral space) to analysis increment:

  $\Delta x = S^{-1} \sum (S(e^k - \langle e \rangle)) \circ (L_{spec}^{1/2} \xi^k)$

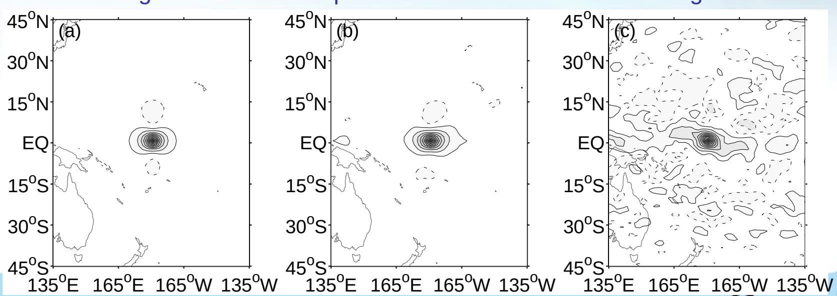
- Spectral correlations forced to zero beyond total wavenumber difference of 10 (Gaussian-like function)



Spectral correlation localization

- Still need to apply spatial localization to damp long-range spurious correlations, however
- Current approach may become prohibitively expensive (memory or time) when combining spatial and spectral localization

Homogeneous | Spectral localization | Heterogeneous



Combining spatial and spectral correlation localization

- transform ensemble members into functions of space and scale:

$$e^k(x) = \sum_n e^k(n)\exp(i2\pi nx) = \sum_n e^k(x,n) \quad \text{(but too big to store)}$$

- both control vector (ξ) and L depend on space and scale, but L could be separable:

$$L_{spec,grid}(x_1,x_2,n_1,n_2) = L_{grid}(x_1,x_2)\, L_{spec}(n_1,n_2)$$

- follow same approach as before:

$$\Delta x = \sum_k \sum_n \left(e^k(x,n) - \langle e(x,n)\rangle\right) \circ \left(L_{spec,grid}^{1/2}\, \xi^k(x,n)\right)$$

- J_b is still the same form: $J_b = 1/2\ \xi^T \xi$
- some similarities with wavelet approach are evident

Simpler approach for resolving limited amount of spatial heterogeneity

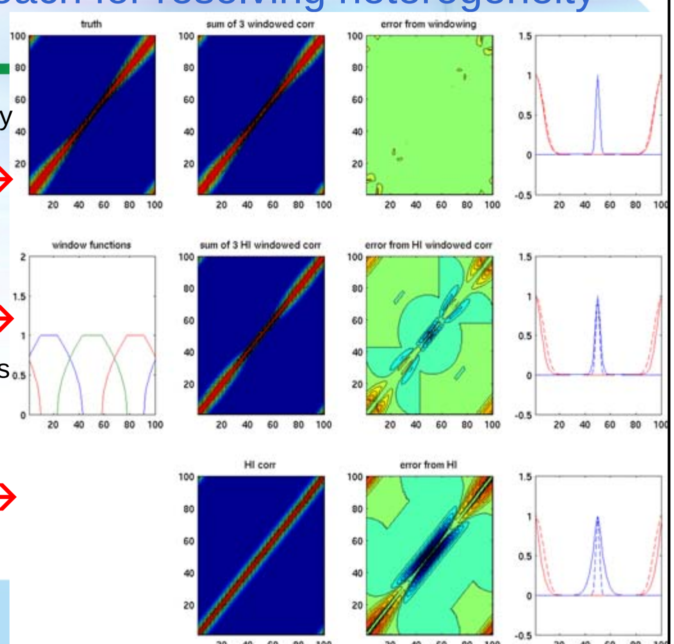
- Tested a simple approach for resolving Tropical vs. Extra-Tropical differences in error correlations in NWP variational system
- Use current approach for modelling homogeneous correlations, but estimate separate statistics for three latitude bands
- Construct increment by combining 3 separate increments:

$$\Delta x = \alpha_{nh}(lat)\, B_{nh}^{1/2}\, \xi_{nh} + \alpha_{tr}(lat)\, B_{tr}^{1/2}\, \xi_{tr} + \alpha_{sh}(lat)\, B_{sh}^{1/2}\, \xi_{sh}$$

- Weighting functions (α) constructed to conserve total variance within transistion zones
- Also results in some limited spatial localization

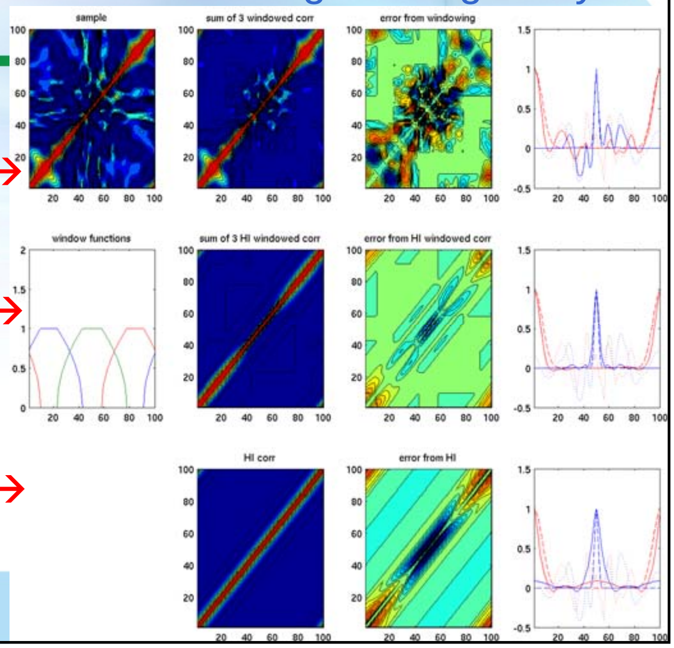
Simpler approach for resolving heterogeneity

- Simple 1D experiment, smoothly varying length scale
- Divide correlations into 3 regions and apply to "truth" →
- Compare with making corr in each region homogeneous →
- Also compare with making correlations homogeneous over entire domain →



Simpler approach for resolving heterogeneity

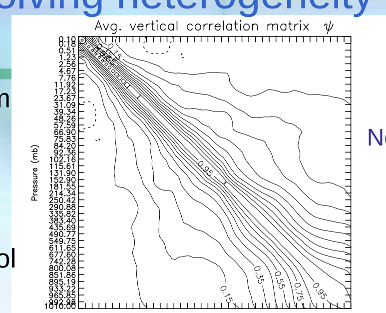
- Apply to sample estimate of correlations (N=20)
- Divide correlations into 3 regions and apply to sample estimate
- Compare with corr homogeneous in each region
- Also compare with making correlations homogeneous over entire domain



Environment Canada Environnement Canada

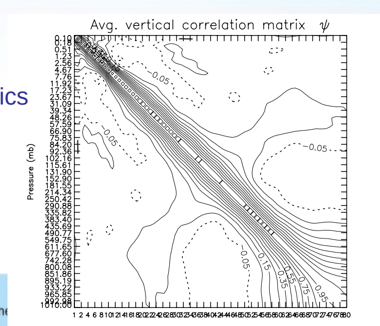
Simpler approach for resolving heterogeneity

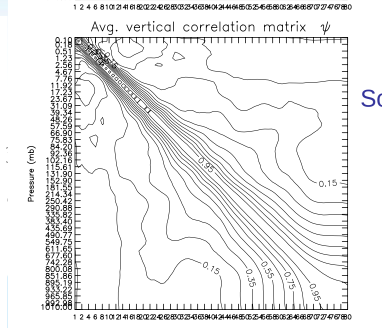
- Tested approach with new NWP system under development
- Vertical correlations of streamfunction shown for 3 latitudinal bands (NMC method)
- Relatively easy to implement, but control vector 3 times larger



North

Tropics





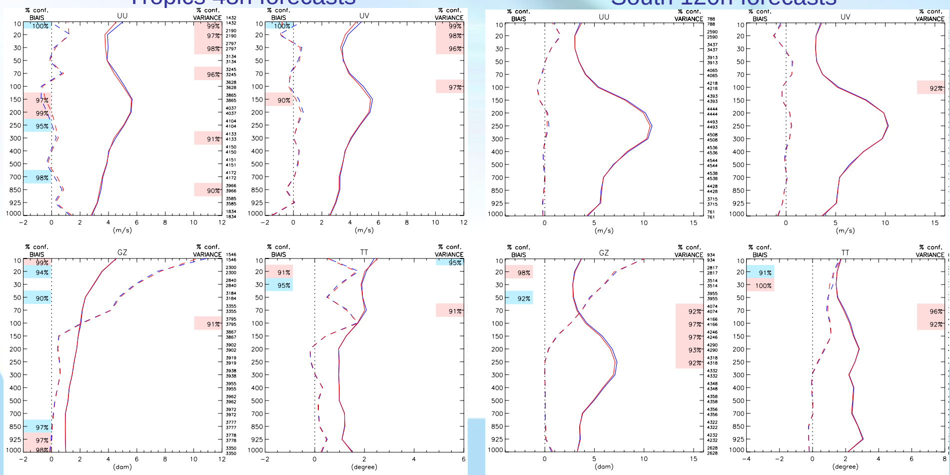
South

Environment Canada

Simpler approach for resolving heterogeneity

- Verifications against radiosondes of forecasts from 2 month 3D-Var experiments with 3 zonal bands vs. globally homogeneous correlations

Tropics 48h forecasts　　　　South 120h forecasts



Extra Slides

Plan for testing EnKF covariances in 4D-Var

- Prompted by workshop planned for November 2008 in Argentina
- Currently, EnKF and 4D-Var are too different to allow useful comparison: horizontal resolution, deterministic vs. probabilistic, etc.
- Design experiments to isolate specific differences:
  1) **standard EnKF**: use ensemble mean for verification (low-res)
  2) **“deterministic” EnKF**: additional member with no perturbations to simulate obs or model error (low-res)
  3) **incremental “deterministic” EnKF**: additional deterministic member at higher horizontal resolution than EnKF ensemble
  4) **incremental 4D-Var with ensemble-based B**: ensemble-based error covariances at beginning of assimilation window with same localization as EnKF
  5) **incremental 4D-Var with static B**: same as operational deterministic analysis system

Environment Canada Environnement Canada

Canada

Plan for testing EnKF covariances in 4D-Var

Specific differences whose impact could be evaluated:

- **smoothing of ensemble mean relative to deterministic forecast**:

  1) standard EnKF vs. 2) “deterministic” EnKF at same resolution

- **different analysis approach with equal covariances at beginning of assimilation window**:

  3) incremental “deterministic” EnKF vs. 4) incremental 4D-Var with ensemble-based B

- ***impact of flow-dependent ensemble-based covariances in 4D-Var**:

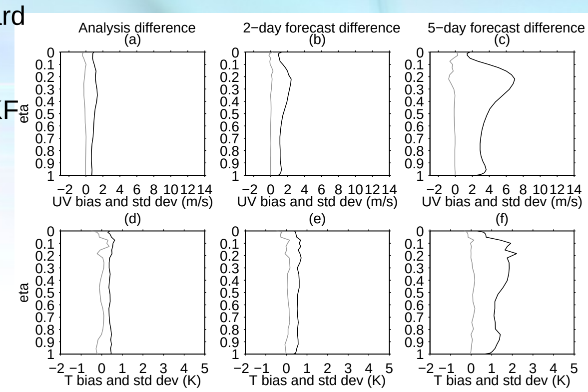
  4) ensemble-based error covariances vs. 5) static covariances

Environment Canada Environnement Canada

Canada

Earlier tests with EnKF error covariances

- Impact of EnKF vs. standard 3D-Var error covariances
- Horizontal and vertical localization applied to EnKF covariances
- Single case of rapidly developing system over Pacific (12 UTC, 27 May 2002)
- Bias (grey curves) and std dev (black curves) of the analysis and forecast differences



Earlier tests with EnKF error covariances

- Forecast error measured vs. analyses from CNTL assimilation experiment
- General improvement from using EnKF error covariances
- Small improvement also seen in scores averaged over 2 week forecast-analysis experiments
- Should revisit, now 4D-Var and EnKF has also been improved