

ECMWF

**Data Assimilation
Training Course**

-

**Background Error Covariance
Modelling**

**Elias Holm – slides courtesy Mike Fisher,
Massimo Bonavita**

Importance of Background Covariances

- The formulation of the J_b term of the cost function is crucial to the performance of current analysis systems.
- To see why, suppose we have a single observation of the value of a model field at one gridpoint.
- For this simple case, the observation operator is:

$$H = (0, \dots, 0, 1, 0, \dots, 0).$$

- The gradient of the 3dVar cost function is:

$$\nabla J = B^{-1}(x - x_b) + H^T R^{-1}(Hx - y) = 0$$

- Multiply through by B and rearrange a bit:

$$x - x_b = B H^T R^{-1}(y - Hx)$$

- But, for this simple case, $R^{-1}(y - Hx)$ is a scalar

Importance of Background Covariances

- So, we have: $\mathbf{x} - \mathbf{x}_b \propto \mathbf{B}\mathbf{H}^T$
- But, $\mathbf{H} = (0, \dots, 0, 1, 0, \dots, 0)$
- **=> The analysis increment is proportional to a column of B.**

- The role of B is:
 1. To spread out the information from the observations.
 2. To provide statistically consistent increments at the neighbouring gridpoints and levels of the model.
 3. To ensure that observations of one model variable (e.g. temperature) produce dynamically consistent increments in the other model variables (e.g. vorticity and divergence).

Main Issues in Covariance Modelling

- There are 2 problems to be addressed in specifying B:
 1. We want to describe the statistics of the errors in the background.
 - However, we don't know what the errors in the background are, since we don't know the true state of the atmosphere.
 2. The B matrix is enormous ($\sim 10^7 \times 10^7$).
 - We are forced to simplify it just to fit it into the computer.
 - Even if we could fit it into the computer, we don't have enough statistical information to determine all its elements.

Diagnosing Background Error Statistics

- **Problem:**

- We cannot produce samples of background error. (We don't know the true state.)

- **Instead, we must either:**

- Disentangle background errors from the information we do have: innovation (observation-minus-background) statistics.

- **Or:**

- Use a surrogate quantity whose error statistics are similar to those of background error. Two possibilities are:
 - Differences between forecasts that verify at the same time.
 - differences between background fields from an ensemble of analyses.

Diagnosing Background Error Statistics

- **Three approaches to estimating J_b statistics:**

- 1. The Hollingsworth and Lönnberg (1986) method**

- Differences between observations and the background are a combination of background and observation error.
- The method tries to partition this error into background errors and observation errors by assuming that the observation errors are spatially uncorrelated.

- 2. The NMC method (Parrish and Derber, 1992)**

- This method assumes that the spatial correlations of background error are similar to the correlations of differences between 48h and 24h forecasts verifying at the same time.

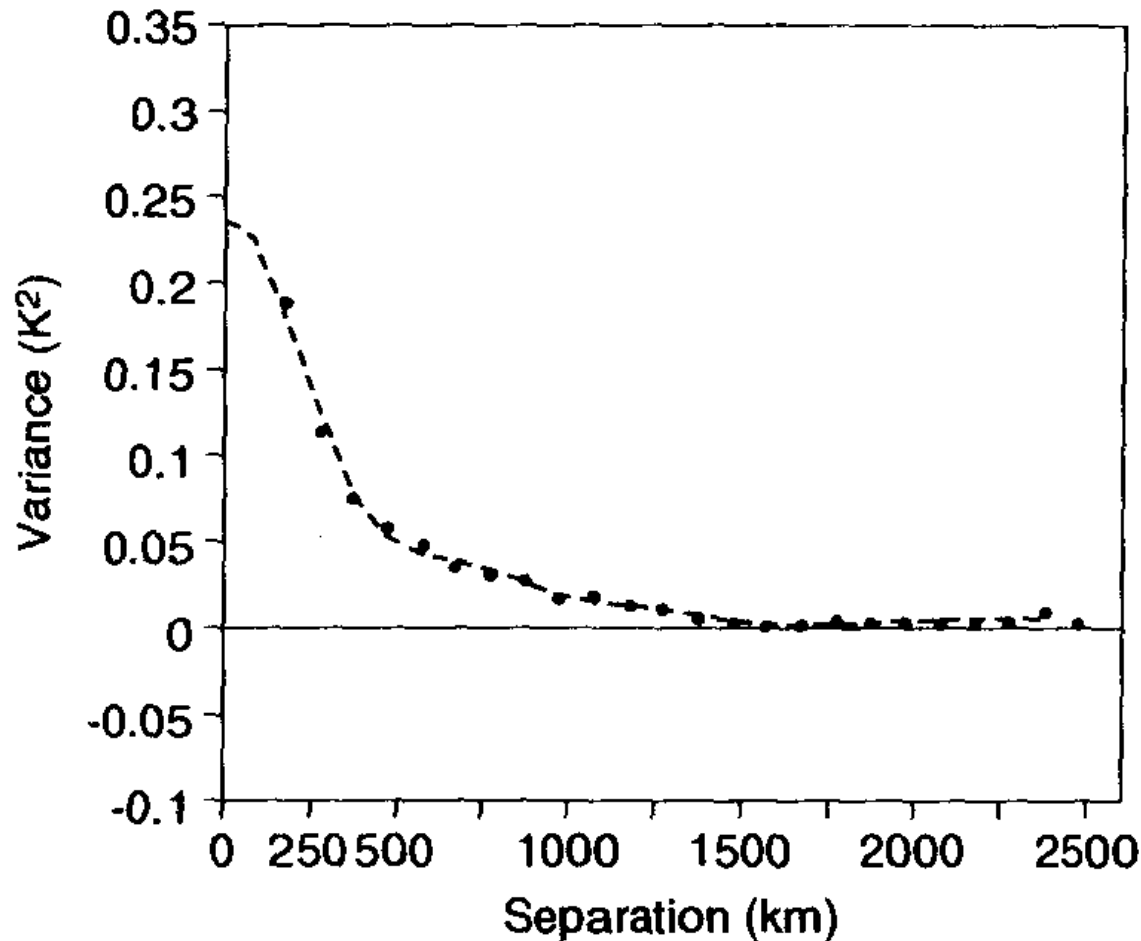
- 3. The Analysis-Ensemble method (Fisher, 2003)**

- This method runs the analysis system several times for the same period with randomly-perturbed observations. Differences between background fields for different runs provide a surrogate for a sample of background error.

Estimating Background Error Statistics from Innovation Statistics

- Assume:
 1. **Background errors are independent of observation errors.**
 2. **Observations have spatially uncorrelated errors (for some observation types).**
- Let $d_i = y_i - H_i(x_b)$ be the innovation (obs-bg) for the i^{th} observation.
- Then, denoting background error by ε , observation error by η , and neglecting representativeness error, we have $d_i = \eta_i - H_i(\varepsilon)$.
 1. $\Rightarrow \text{Var}(d_i) = \text{Var}(\eta_i) + \text{Var}(H_i(\varepsilon))$
 2. $\Rightarrow \text{Cov}(d_i, d_k) = \text{Cov}(H_i(\varepsilon), H_k(\varepsilon))$ (i and k not co-located)
- We can extract a lot of useful information by plotting $\text{Cov}(d_i, d_k)$ as a function of the distance between pairs of observations.

Estimating Background Error Statistics from Innovation Statistics

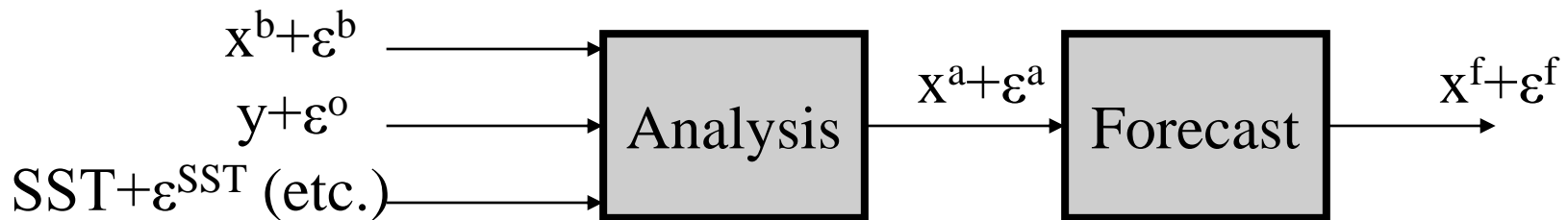


Covariance of $d=y-H(x_b)$ for AIREP temperatures over USA, binned as a function of observation separation.

(from Järvinen, 2001)

Estimating Background Error Statistics from Ensembles of Analyses

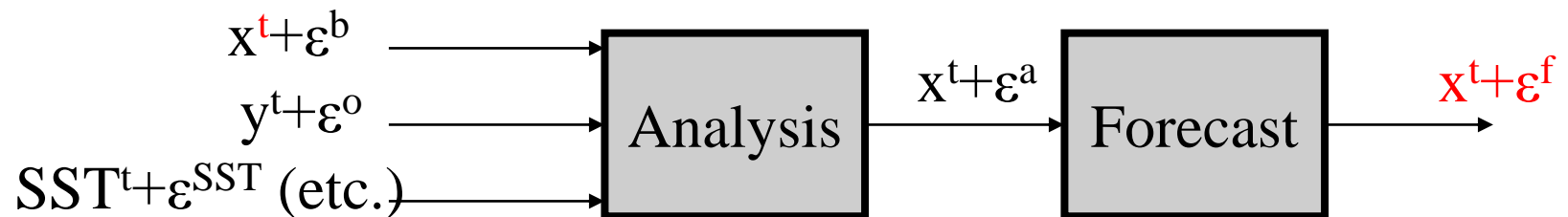
- Suppose we perturb all the inputs to the analysis/forecast system with random perturbations, drawn from the relevant distributions:



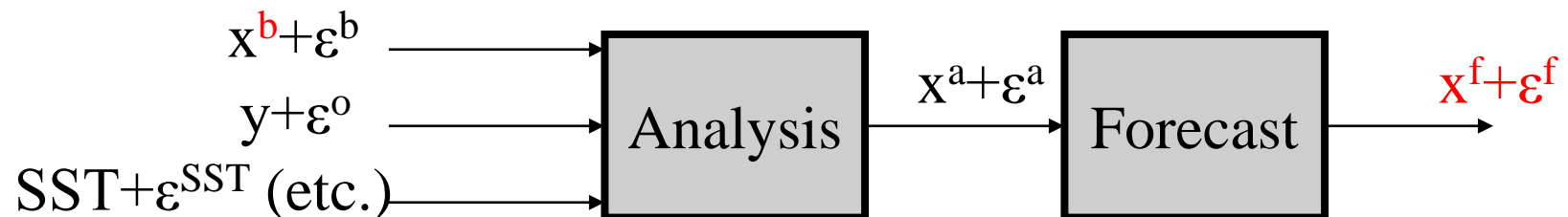
- The result will be a perturbed analysis and forecast, with perturbations characteristic of analysis and forecast error.
- The perturbed forecast may be used as the background for the next (perturbed) cycle.
- After a few cycles, the system will have forgotten the original initial background perturbations.

Estimating Background Error Statistics from Ensembles of Analyses

Normal Analysis

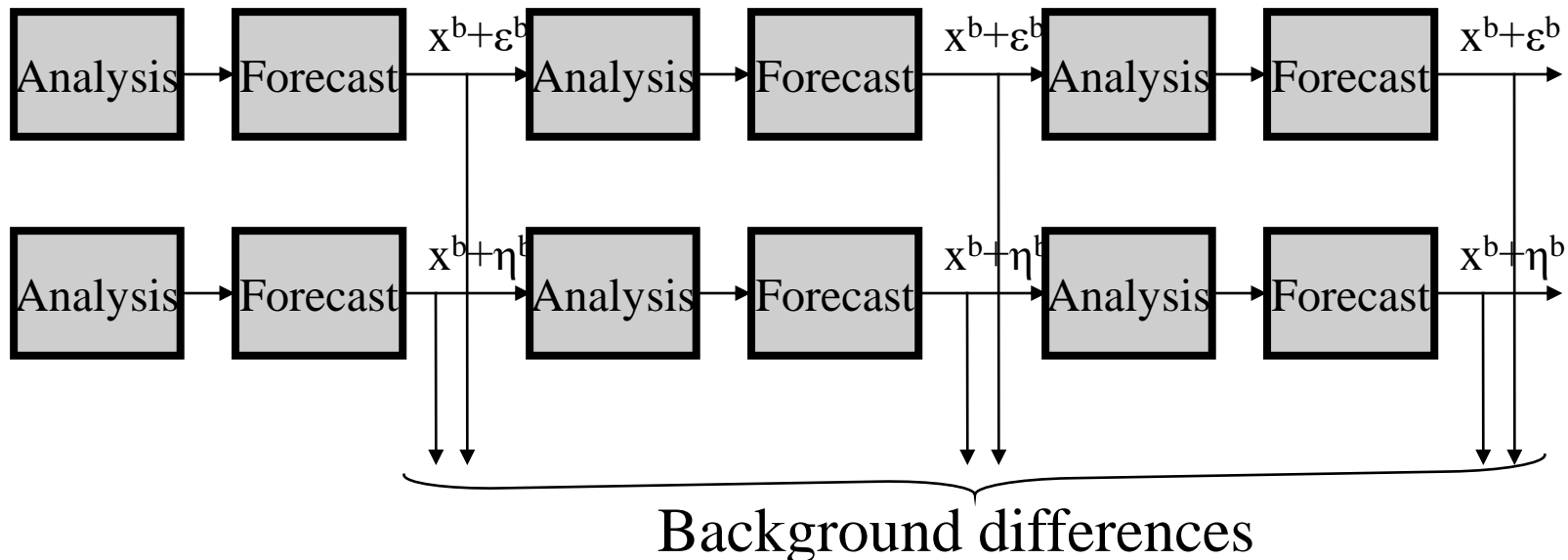


Perturbed Analysis



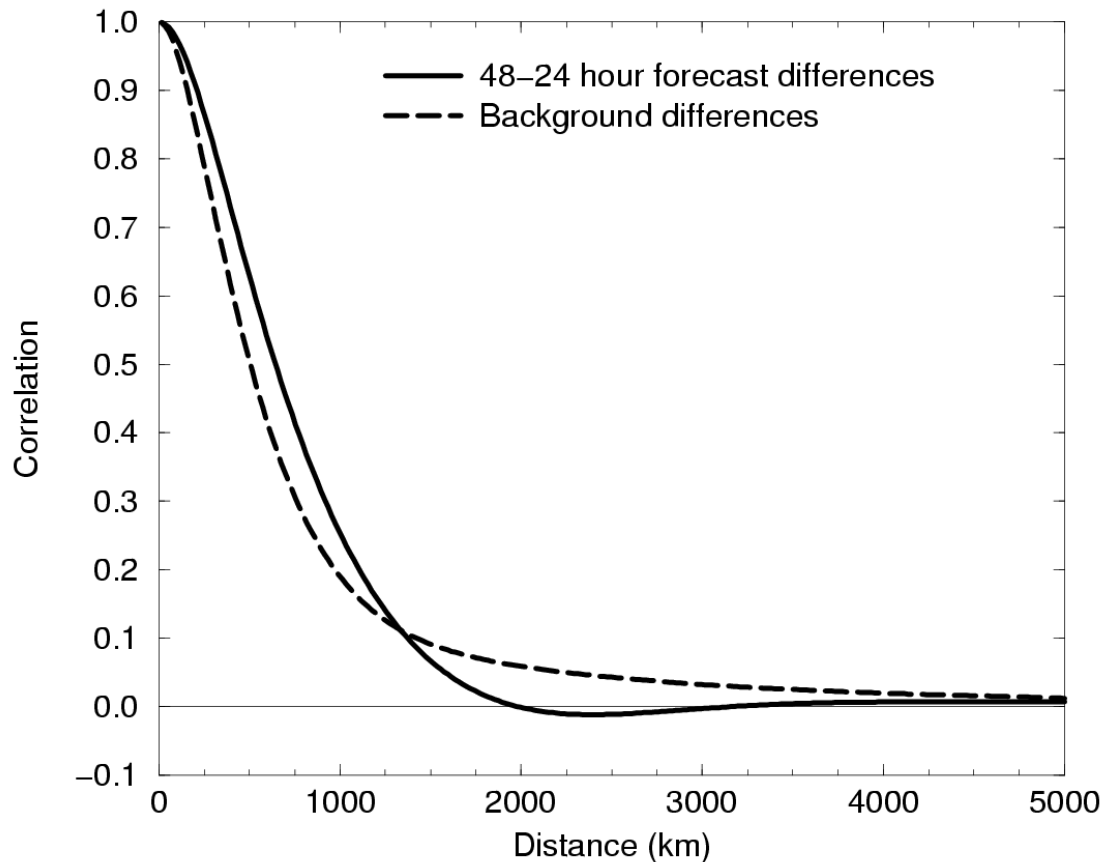
Estimating Background Error Statistics from Ensembles of Analyses

- Run the analysis system **several times with different perturbations**, and form differences between pairs of background fields.
- These differences will have the statistical characteristics of background error (but **twice the variance**).



Estimating Background Error Statistics from Ensembles of Analyses

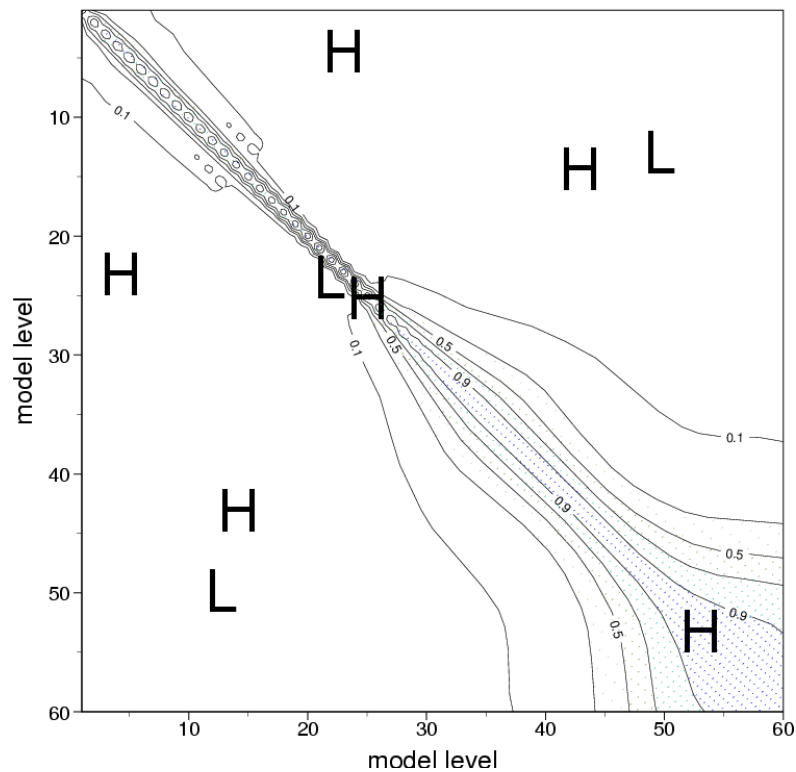
500hPa Geopotential



Estimating Background Error Statistics from Ensembles of Analyses

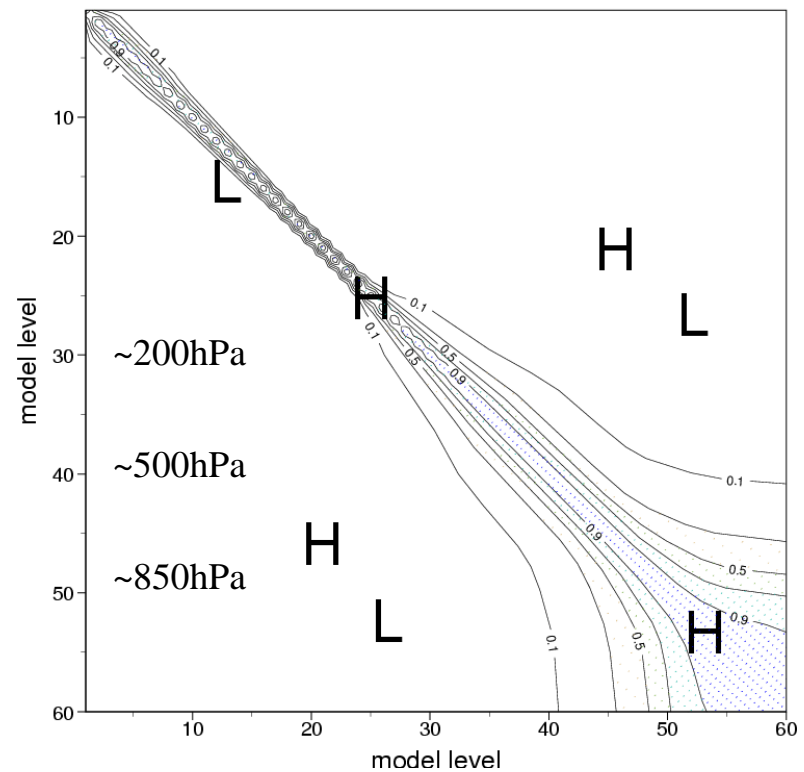
NMC Method

average total vorticity cors



Analysis-Ensemble Method

average total vorticity cors



Estimating Background Error Statistics – Pros and Cons of the Various Methods

● Innovation statistics:

- ☺ The only direct method for diagnosing background error statistics.
- ☹ Provides statistics of background error in observation space.
- ☹ Statistics are not global, and do not cover all model levels.
- ☹ Requires a good uniform observing network.
- ☹ Statistics are biased towards data-dense areas.

● Forecast Differences:

- ☺ Generates global statistics of model variables at all levels.
- ☺ Inexpensive.
- ☹ Statistics are a mixture of analysis and background error.
- ☹ Not good in data-sparse regions.

● Ensembles of Analyses:

- ☹ Assumes statistics of observation error (and SST, etc.) are well known.
- ☺ Diagnoses the statistics of the actual analysis system.
- ☹ Danger of feedback. (Noisy analysis system => noisy stats => noisier system.)

\mathbf{J}_b Formulation – The control variable

- The incremental analysis problem may be rewritten in terms of a new variable, $\boldsymbol{\chi}$, defined by $\mathbf{L}\boldsymbol{\chi} = (\mathbf{x} - \mathbf{x}_b)$, where $\mathbf{L}\mathbf{L}^T = \mathbf{B}$.

- The cost function becomes:

$$J(\boldsymbol{\chi}) = \frac{1}{2} \boldsymbol{\chi}^T \boldsymbol{\chi} + (\mathbf{y} - \mathcal{H}(\mathbf{x}_b) - \mathbf{H}\mathbf{L}\boldsymbol{\chi})^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x}_b) - \mathbf{H}\mathbf{L}\boldsymbol{\chi})$$

- It is not necessary for \mathbf{L} to be invertible (or even square), but it usually is.
- The covariance matrix for $\boldsymbol{\chi}$ is the identity matrix. This is obvious if \mathbf{L} is invertible:

$$\begin{aligned} \overline{\boldsymbol{\chi}\boldsymbol{\chi}^T} &= \overline{\mathbf{L}^{-1}(\mathbf{x} - \mathbf{x}_b)(\mathbf{x} - \mathbf{x}_b)^T \mathbf{L}^{-T}} = \mathbf{L}^{-1} \overline{(\mathbf{x} - \mathbf{x}_b)(\mathbf{x} - \mathbf{x}_b)^T} \mathbf{L}^{-T} \\ &= \mathbf{L}^{-1} \mathbf{B} \mathbf{L}^{-T} \\ &= \mathbf{I} \end{aligned}$$

J_b Formulation – The control variable

- We may interpret L as an operator that takes a control vector χ with covariance matrix I , and introduces correlations to give the background departures, $(x-x_b)$.
- With this interpretation, we may factorize L into a sequence steps, each of which adds some aspect of correlation into the background departures.

The ECMWF J_b Formulation – The balance operator

- The most obvious correlation in the background errors is the balance between mass errors and wind errors in the extra-tropics.
- We therefore define our change of variable as:

$$L = KB_u^{1/2}$$

- where K accounts for all the correlation between variables (e.g. between the mass and wind fields).
- The matrix B_u is a covariance matrix for variables that are uncorrelated with each other.
- $\Rightarrow B_u$ is block diagonal, with one block for each variable.

The ECMWF J_b Formulation – The balance operator

- **K** accounts for the correlations between variables:

$$\begin{pmatrix} \zeta \\ \mathbf{D} \\ (\mathbf{T}, \mathbf{p}_s) \\ \mathbf{q} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{I} & 0 & 0 & 0 \\ \mathbf{M} & \mathbf{I} & 0 & 0 \\ \mathbf{N} & \mathbf{P} & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{pmatrix}}_{\mathbf{K}} \begin{pmatrix} \zeta \\ \mathbf{D}_u \\ (\mathbf{T}, \mathbf{p}_s)_u \\ \mathbf{q} \end{pmatrix}$$

- The inverse is:

$$\mathbf{K}^{-1} = \begin{pmatrix} \mathbf{I} & 0 & 0 & 0 \\ -\mathbf{M} & \mathbf{I} & 0 & 0 \\ (\mathbf{PM} - \mathbf{N}) & -\mathbf{P} & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{pmatrix}$$

The ECMWF J_b Formulation – The balance operator

$$\begin{pmatrix} \zeta \\ \mathbf{D} \\ (\mathbf{T}, \mathbf{p}_s) \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 & 0 & 0 \\ \mathbf{M} & \mathbf{I} & 0 & 0 \\ \mathbf{N} & \mathbf{P} & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} \zeta \\ \mathbf{D}_u \\ (\mathbf{T}, \mathbf{p}_s)_u \\ \mathbf{q} \end{pmatrix}$$

- The most important part of the balance operator is the sub-matrix \mathbf{N} , which calculates a balanced part of $(\mathbf{T}, \mathbf{p}_s)$, determined from the vorticity.
- \mathbf{N} is implemented in 2 parts:
 1. A balanced “geopotential” is calculated from ζ .
 2. Balanced $(\mathbf{T}, \mathbf{p}_s)$ are calculated using statistical regression between $(\mathbf{T}, \mathbf{p}_s)$ and geopotential.
 - (Using regression avoids some numerical problems associated with inverting the hydrostatic equation.)

The ECMWF J_b Formulation – The balance operator

- The original (Derber and Bouttier, 1999) ECMWF balance operator calculated balanced geopotential from vorticity using a statistical regression.
- The regression gave results that were nearly indistinguishable from linear balance.
- We have replaced this part of the balance operator with an analytical balance: nonlinear balance, linearized about the background state.
- This gives a flow-dependent balance operator:

$$\nabla^2 \Phi' = -\nabla \cdot \left(\mathbf{v}_{\psi b} \cdot \nabla \mathbf{v}'_{\psi} + \mathbf{v}'_{\psi} \cdot \nabla \mathbf{v}_{\psi b} + f \mathbf{k} \times \mathbf{v}'_{\psi} \right)$$

- The extra, flow-dependent, terms are particularly important in regions of strong curvature (jet entrances, exits, etc.).

QG Omega Equation

- A similar approach allows us to augment the balance operator with a term that calculates balanced divergence from vorticity and temperature, according to the quasi-geostrophic omega equation:

$$(\sigma \nabla^2 + f_0^2 \frac{\partial^2}{\partial p^2}) \omega' = -2 \nabla \cdot \mathbf{Q}'$$

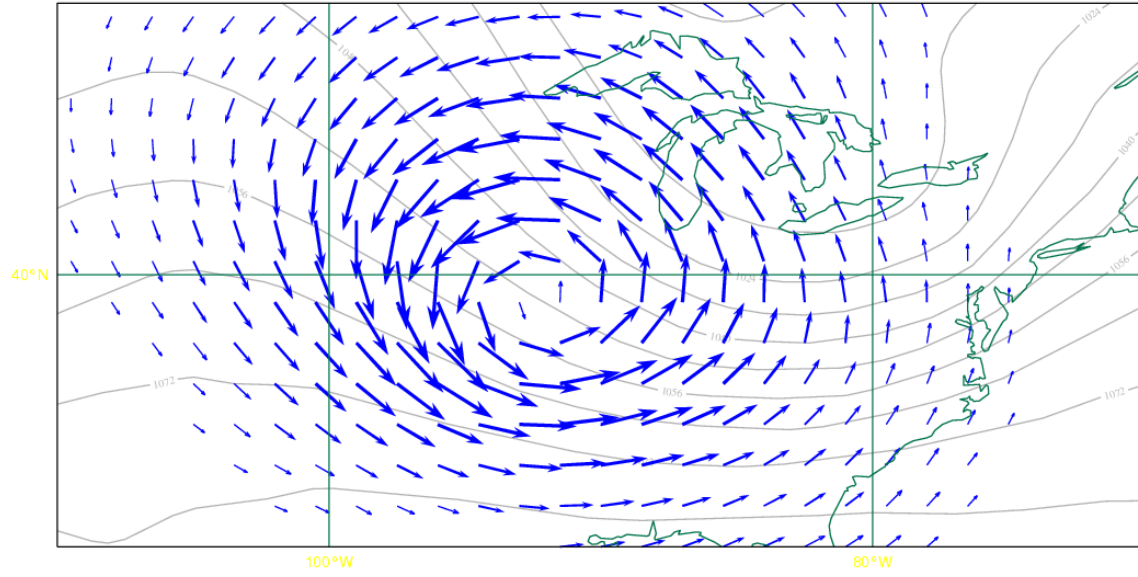
- Linearize Q about the background:

$$\mathbf{Q}' = -\frac{R}{p} \left[\left(\frac{\partial \mathbf{v}'_{\psi}}{\partial x} \cdot \nabla T_b + \frac{\partial \mathbf{v}_{\psi b}}{\partial x} \cdot \nabla T' \right) \mathbf{i} + \left(\frac{\partial \mathbf{v}'_{\psi}}{\partial x} \cdot \nabla T_b + \frac{\partial \mathbf{v}_{\psi b}}{\partial x} \cdot \nabla T' \right) \mathbf{j} \right]$$

Wind increments at level 31 from a single height observation at 300hPa.

Tuesday 14 May 2002 06UTC ECMWF Forecast t+3 VT: Tuesday 14 May 2002 UTC 250hPa geopotential height

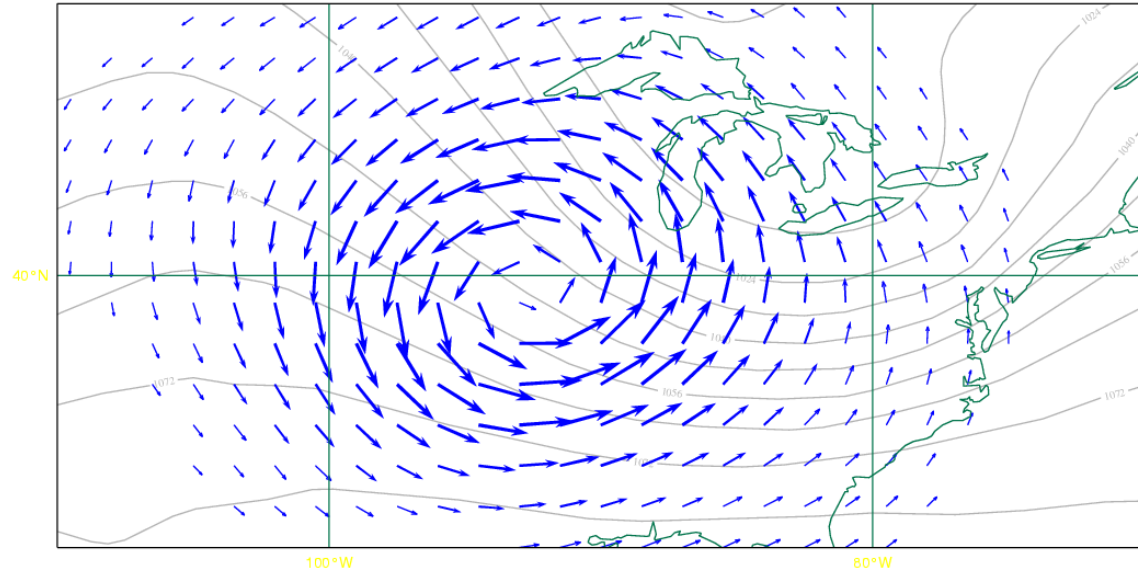
0.0m/s



J_b includes:
Nonlinear balance
equation and omega
equation.

Tuesday 14 May 2002 06UTC ECMWF Forecast t+3 VT: Tuesday 14 May 2002 UTC 250hPa geopotential height

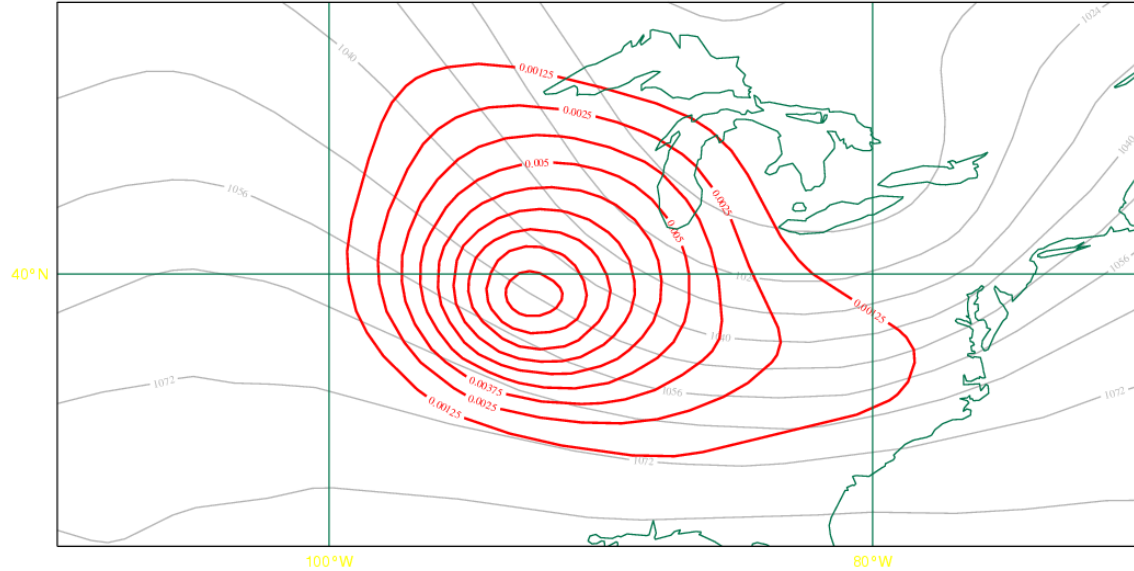
0.0m/s



Linear balance
only.

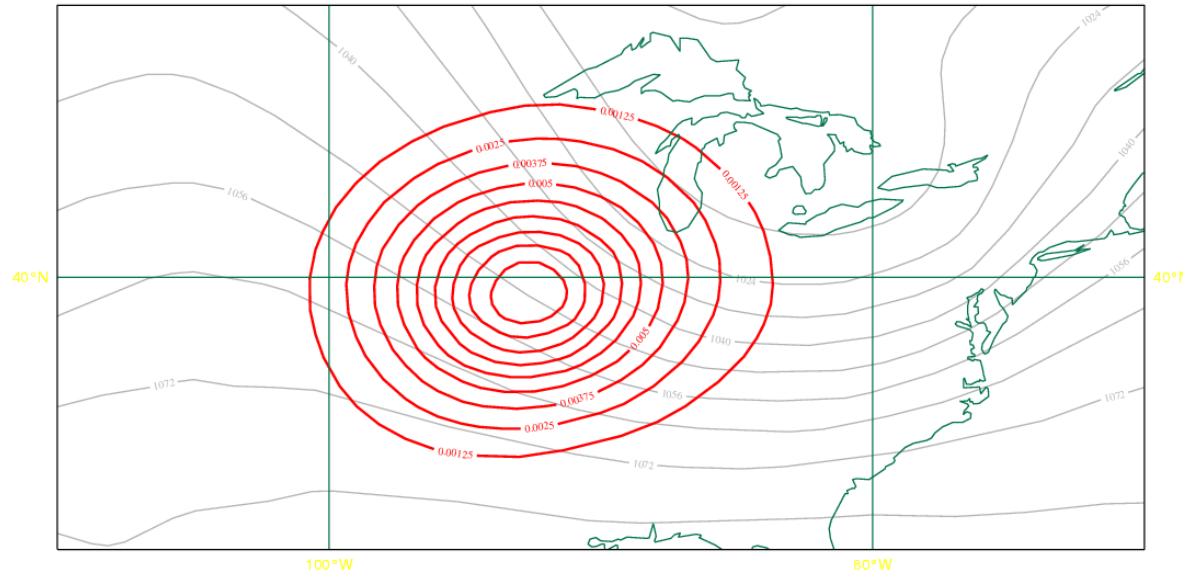
Temperature increments at level 31 from a height observation at 300hPa.

Tuesday 14 May 2002 06UTC ECMWF Forecast t+3 VT: Tuesday 14 May 2002 UTC 250hPa geopotential height



J_b includes:
Nonlinear balance
equation and omega
equation.

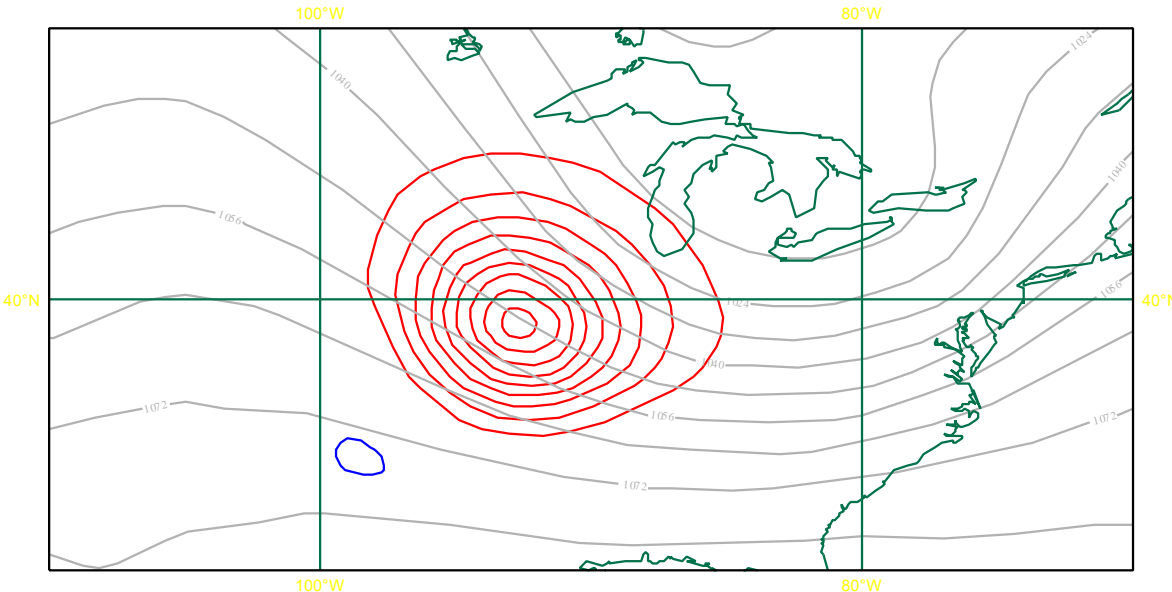
Tuesday 14 May 2002 06UTC ECMWF Forecast t+3 VT: Tuesday 14 May 2002 UTC 250hPa geopotential height



Linear balance
only.

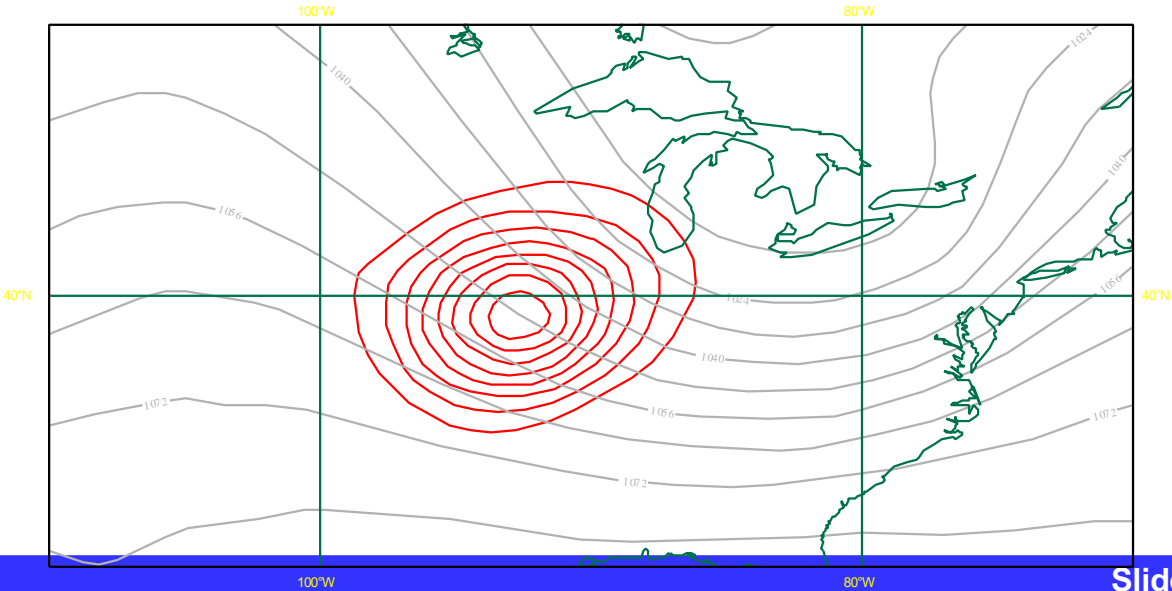
Vorticity increments at level 31 from a height observation at 300hPa.

Tuesday 14 May 2002 06UTC ECMWF Forecast+3 VT: Tuesday 14 May 2002 UTC 250hPa geopotential height



J_b includes:
Nonlinear balance equation and omega equation.

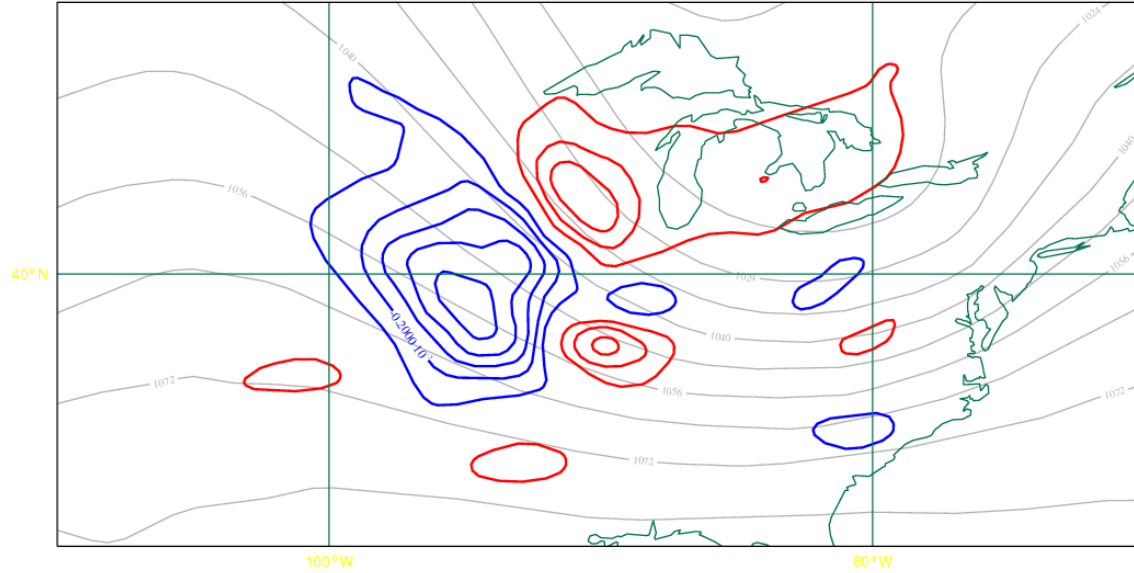
Tuesday 14 May 2002 06UTC ECMWF Forecast+3 VT: Tuesday 14 May 2002 UTC 250hPa geopotential height



Linear balance only.

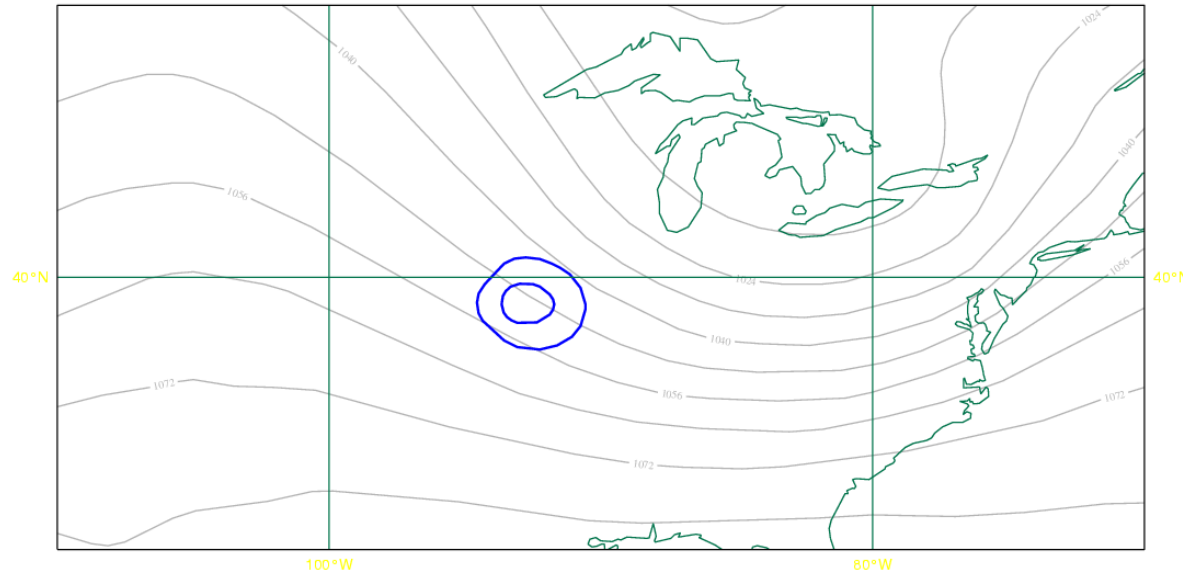
Divergence increments at level 31 from a height observation at 300hPa.

Tuesday 14 May 2002 06UTC ECMWF Forecast+3 VT: Tuesday 14 May 2002 UTC 250hPa geopotential height



J_b includes:
Nonlinear balance
equation and omega
equation.

Tuesday 14 May 2002 06UTC ECMWF Forecast+3 VT: Tuesday 14 May 2002 UTC 250hPa geopotential height



Linear balance
only.

Spectral B model

In variational analysis the B matrix is usually defined implicitly in terms of a **transformation** from the departure $\delta\mathbf{x}$ in state space to a control variable χ :

$$\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b = \mathbf{L}\chi$$

where \mathbf{L} verifies $\mathbf{B} = \mathbf{L}\mathbf{L}^T$

In the **spectral formulation** (Derber and Bouttier, 1999), the change of variable \mathbf{L} has the form:

$$\mathbf{L} = \mathbf{K} \mathbf{B}_u^{1/2}$$

where \mathbf{K} is a **balance** operator going from the set of “unbalanced” variables $[\zeta, \eta_u, (T, ps)_u, q]$ (the “control vector”) to the set of state variables $[\zeta, \eta, (T, ps), q]$

There is a degree of flow-dependence in \mathbf{K} as the balance constraints are linearised about the first-guess trajectory

Spectral B model

$$\delta \mathbf{x} = \mathbf{x} - \mathbf{x}_b = \mathbf{L} \boldsymbol{\chi} \quad \mathbf{L} = \mathbf{K} \mathbf{B}_u^{1/2}$$

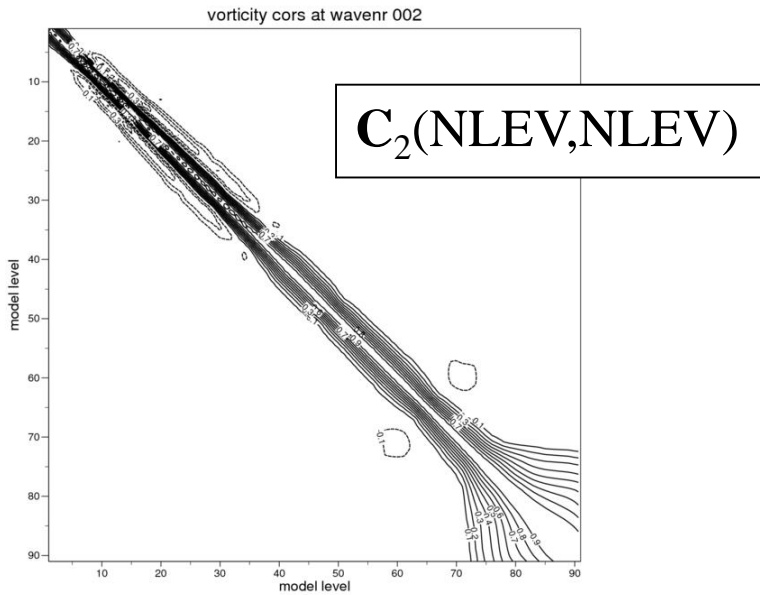
Since we assume that the balance operator accounts for all inter-variable correlations, \mathbf{B}_u is block diagonal

$$\mathbf{B}_u = \begin{pmatrix} \mathbf{B}_\zeta & 0 & 0 & 0 \\ 0 & \mathbf{B}_{D_u} & 0 & 0 \\ 0 & 0 & \mathbf{B}_{(T,p_s)_u} & 0 \\ 0 & 0 & 0 & \mathbf{B}_q \end{pmatrix}$$

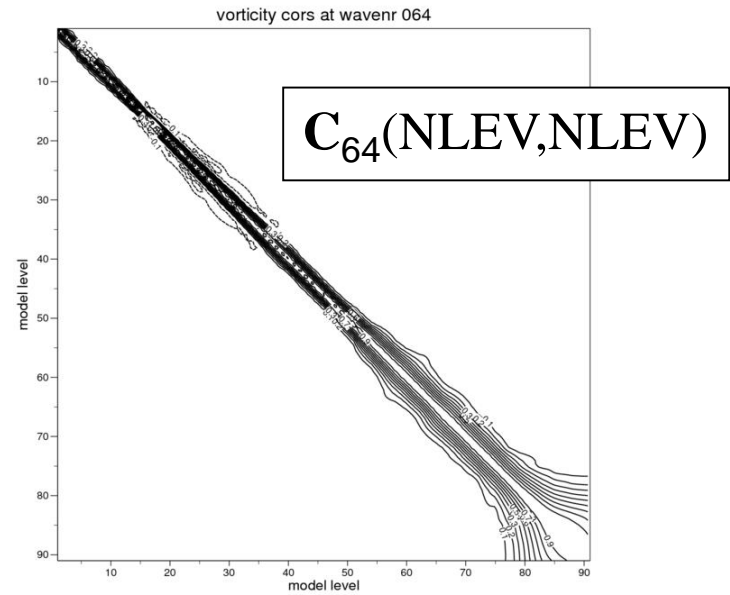
Each block in \mathbf{B}_u is of the form $\boldsymbol{\Sigma}^T \mathbf{C} \boldsymbol{\Sigma}$.

$\boldsymbol{\Sigma}$ is the **gridpoint** standard deviation of background errors.

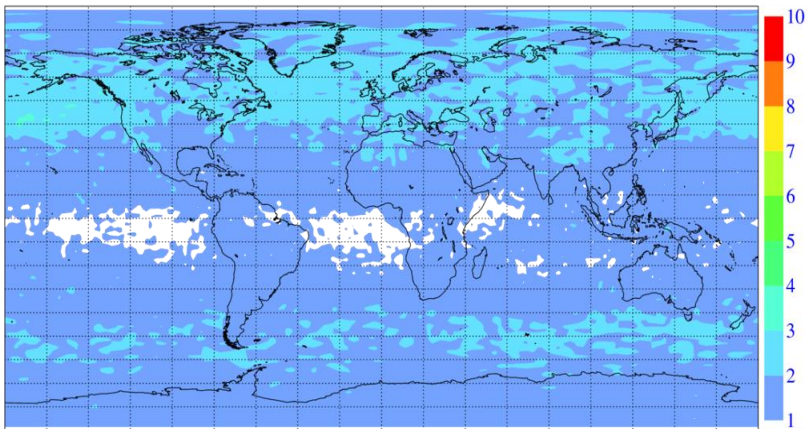
\mathbf{C} models the **autocorrelation** of the control variables. It is block diagonal with one full vertical correlation matrix for each spectral wavenumber, i.e. $C_n(\text{NLEV}, \text{NLEV})$ (non-separable B model)



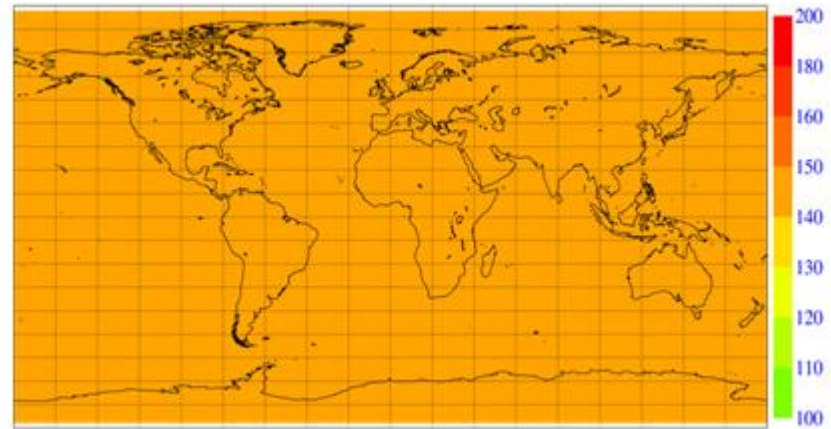
Vorticity correl. wavenum=2



Vorticity correl. wavenum=64



Vorticity bg error stdev, 500hPa



Vorticity bg error corr. Lscale, 500hPa

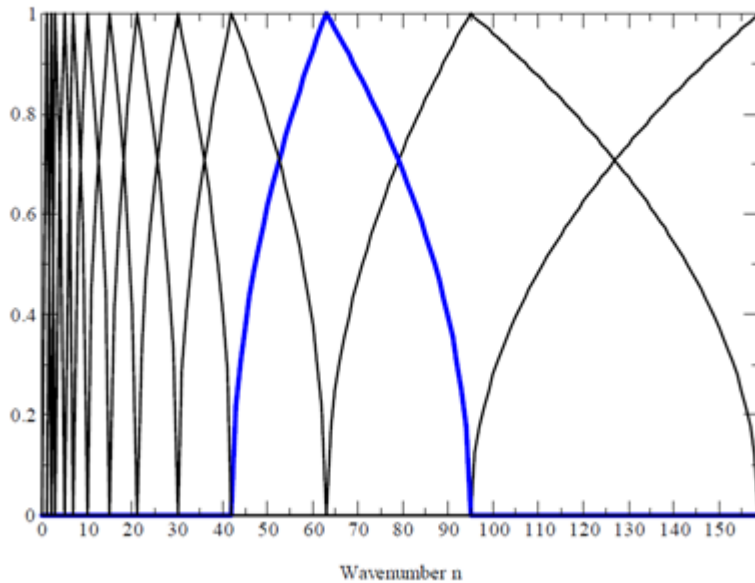
From Spectral to Wavelet B model

- The **spectral B** model is one end of the spectrum: full resolution of the variation of vertical correlation with horizontal scale, but it allows no horizontal variability of the vertical/horizontal correlations
- The other end of the spectrum is represented by the separable formulation which allows full horizontal variation of the correlations (we may specify a different vertical covariance matrix for each horizontal grid point), but has no variation of vertical correlation with horizontal scale
- The **wavelet B** (Fisher, 2003) is a compromise between these two extremes and allows **a degree of variation of correlation with both wavenumber and horizontal location**

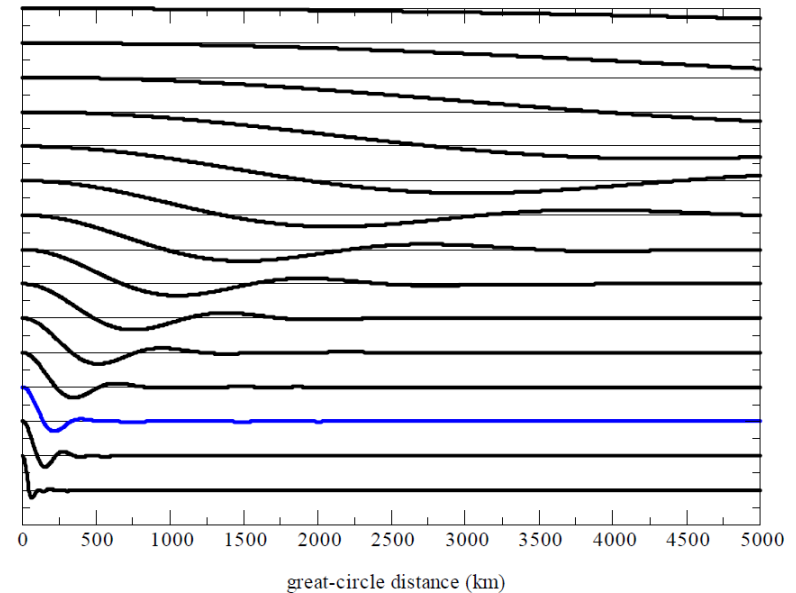
Wavelet B model

- The **wavelet B** is based on a **wavelet expansion** on the sphere.
- The basis functions (wavelets) are chosen to be **band-limited** and, to a good approximation, **spatially localized**

Wavelet functions: $\hat{\psi}_j(n) = (\hat{\phi}_j^2(n) - \hat{\phi}_{j-1}^2(n))^{1/2}$

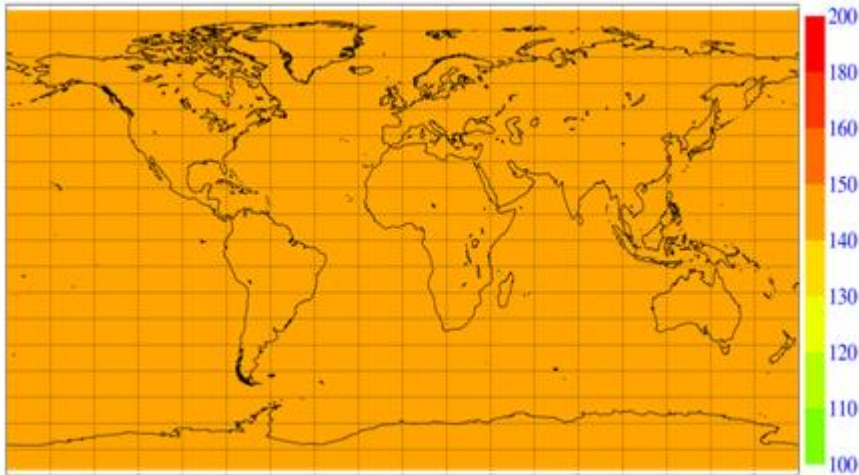


Wavelet functions: $\psi_j(r)$

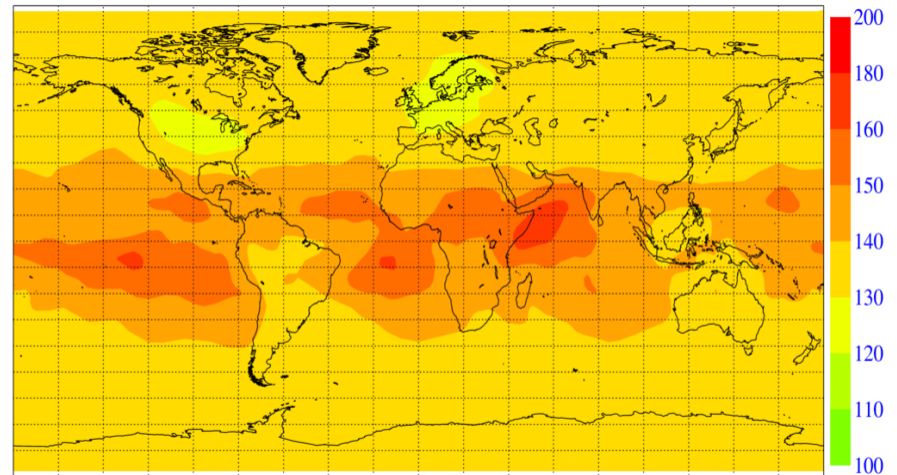


Wavelet B model

- The correlation matrices $\mathbf{C}_n[N_{lev} \times N_{lev}]$ are now of the form $\mathbf{C}_j[N_{lev} \times N_{lev}](\lambda, \varphi)$, where j is now the index of the wavelet component
- The choice of the wavelet bandwidths $[N_j, N_{j+1}]$ determines the **trade-off between spectral and spatial resolution**. If the bands are narrow, the corresponding wavelet functions are not spatially localized, and vice versa



Climat. Spectral B
Vorticity bg error corr. Lscale, 500hPa



Climat. Wavelet B
Vorticity bg error corr. Lscale, 500hPa

Flow-dependent wavelet **B** model

The wavelet **B** formulation:

$$(\mathbf{x} - \mathbf{x}_b) = \mathbf{L}\boldsymbol{\chi} = \mathbf{K}\boldsymbol{\Sigma}_b^{1/2} \sum_j \boldsymbol{\psi}_j \otimes [\mathbf{C}_j^{1/2}(\lambda, \phi)\boldsymbol{\chi}_j]$$

can be made flow-dependent by obtaining flow-dependent estimates of the **background error variances** ($\boldsymbol{\Sigma}_b$) and **correlations** ($\mathbf{C}_j(\lambda, \phi)$) from the EDA background perturbations

Diffusion Operators and Digital Filters

- The spectral approach is efficient and convenient for models with regular (e.g. spherical or rectangular) domains.
- It is difficult to use if the domain is not regular (e.g. ocean models).
- Because the spectral approach is based on convolutions, it is difficult to incorporate inhomogeneity and anisotropy.
- Diffusion operators and digital filters provide alternatives to the spectral approach that address these difficulties.

Diffusion Operators

- The 1-dimensional diffusion equation:

$$\frac{\partial \eta}{\partial t} - \kappa \frac{\partial^2 \eta}{\partial x^2} = 0$$

- Has solution at time T :

$$\eta(x, T) = \frac{1}{\sqrt{4\pi\kappa T}} \int_{-\infty}^{\infty} e^{-(x-x')^2 / 4\kappa T} \eta(x', 0) dx'$$

- That is, $\eta(x, T)$ is the result of convolving $\eta(x, 0)$ with the Gaussian function:

$$\frac{1}{\sqrt{4\pi\kappa T}} \exp\left(-x^2 / 4\kappa T\right)$$

Diffusion Operators

- The one-dimensional result generalizes to more dimensions, and to different geometries (e.g. on the sphere).
- Weaver and Courtier (2001) realized that numerical integration of a diffusion equation could be used to perform convolutions for covariance modelling.
- Irregular boundary conditions (e.g. coastlines) are easily handled.
- More general partial differential equations can be used to generate a large class of correlation functions:

$$\frac{\partial \eta}{\partial t} + \sum_{p=1}^P \kappa_p \left(-\nabla^2 \right)^p \eta = 0$$

Diffusion Operators

- The change of variable needs the square-root of the diffusion operator. Fortunately, because the operator is self-adjoint, the square-root is equivalent to integrating the equation from time 0 to $T/2$.
- Inhomogeneous covariance models can be produced by making the diffusion coefficients vary with location.
- Anisotropic covariances can be produced by using tensor diffusion coefficients.
- Disadvantages:
 - Calculation of the normalization coefficient ($1/\sqrt{4\pi\kappa T}$ in the 1-D example) is expensive in the general case.
 - The relationship between the diffusion coefficients and the shape of the correlation function is complicated. It is difficult to generate suitable coefficients to match the correlations implied by data.

Digital Filters

- In one-dimension, convolution with a Gaussian may be achieved, to good approximation, using a pair of recursive filters:

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

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

- In two dimensions, the Fourier transform of the Gaussian factorizes:

$$\exp\left(-\frac{a^2(k^2 + l^2)}{2}\right) = \exp\left(-\frac{a^2 k^2}{2}\right) \exp\left(-\frac{a^2 l^2}{2}\right)$$

- => 2-D convolution may be achieved by 1-D filtering in the x-direction, and then in the y-direction.

- NB: This factorization only works for Gaussians!

Digital Filters

- **Non-Gaussian covariance functions may be produced as a superposition of Gaussians.**
 - I.e. the filtered field is the weighted sum of convolutions with a set of Gaussians of different widths.
- **Inhomogeneous covariances may be synthesized by allowing the filter coefficients to vary with location.**
- **Simple anisotropic covariances (ellipses), with different north-south and east-west length scales, can be produced by using different filters in the north-south direction.**
- **However, fully general anisotropy (bananas) requires 3 independent filters (north-south, east-west, and SW-NE) in 2 dimensions and 6 filters in 3 dimensions.**

Digital Filters

- **There is a close connection between digital filter methods and diffusion operator methods.**
 - One timestep of integration of a diffusion operator can be viewed as one application of a digital filter.
- **Advantages of Digital Filters:**
 - Computational Efficiency
 - Generality
- **Disadvantages:**
 - Filter coefficients are difficult to determine from data.
 - Grid geometry, polar singularities and boundary conditions must be handled carefully.

Summary

- **A good B matrix is vitally important in any (current) data assimilation system.**
- **In a large-dimension system, covariances must be modelled: The matrix is too big to specify every element.**
- **Innovation Statistics are the only real data we have to diagnose background error statistics, but they are difficult to use.**
- **Analysis ensembles allow us to generate a good surrogate for samples of background error.**
- **Spectral representations work well for simple geometries (spherical or rectangular domains) but anisotropic and/or inhomogeneous covariances are tricky!**
- **Wavelet formulation allows inhomogeneous covariances.**
- **Diffusion operators and digital filters have fewer limitations, but calculating the diffusion/filter coefficients is non-trivial.**