Background Error Covariance Modelling

Mike Fisher

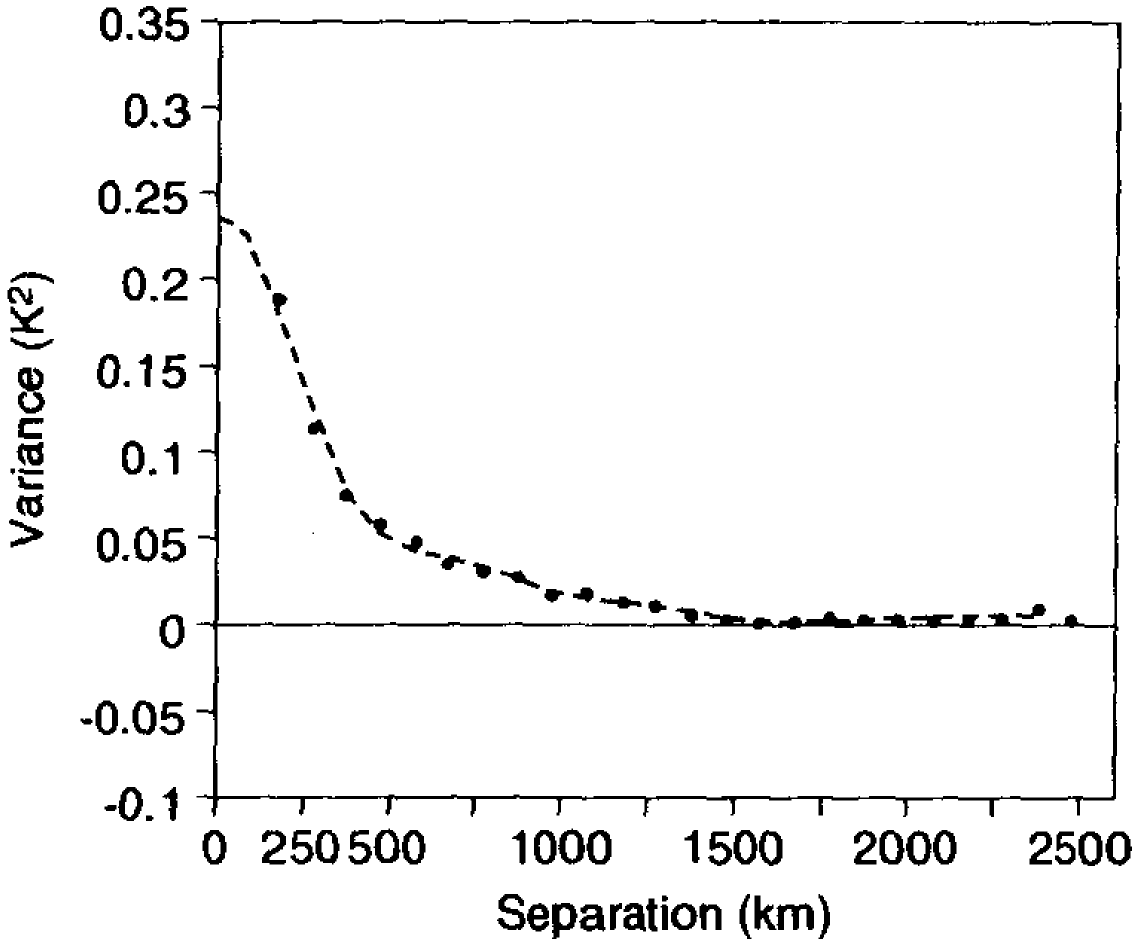
Outline

- Diagnosing the Statistics of Background Error using Ensembles of Analyses
- Modelling the Statistics in Spectral Space
  - Relaxing constraints of isotropy and homogeneity
- Incorporating Balance

Diagnosing Background Error Statistics

- Problem: We cannot produce samples of background error because we don’t know the true state.
- Instead, we must <u>either</u>:
  - Disentangle background errors from the information we do have: innovation statistics.
  - (Hollingsworth and Lönnberg 1986, Tellus 111-136).
- <u>Or</u>:
  - Use a surrogate quantity whose error statistics are similar to those of background error.
    - Forecast differences (Parrish and Derber 1992, MWR 1747-1763)
    - Differences between background fields from an ensemble of analyses.

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.

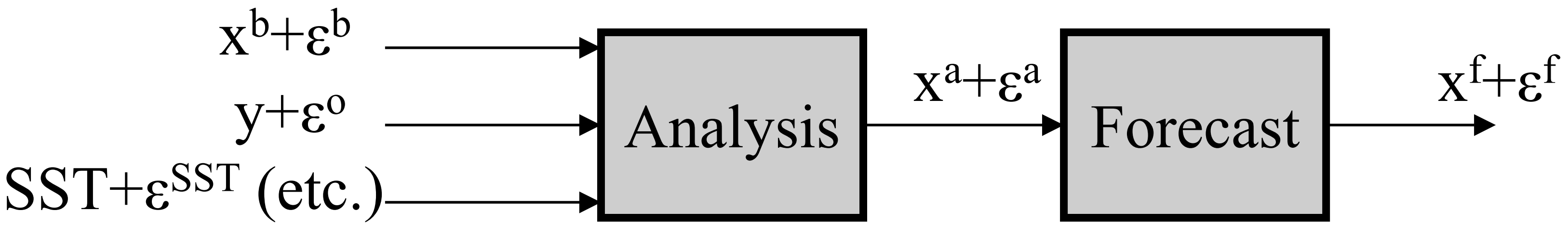
(from Järvinen, 2001)

Estimating Background Error Statistics from Forecast Differences

- Differences are calculated between forecasts and analyses, or between pairs of forecasts, verifying at the same date/time, but with different initial times.
- Parrish and Derber (1992) describe the method as "a very crude first step"!
- The method is widely used.

Estimating Background Error Statistics from Ensembles of Analyses

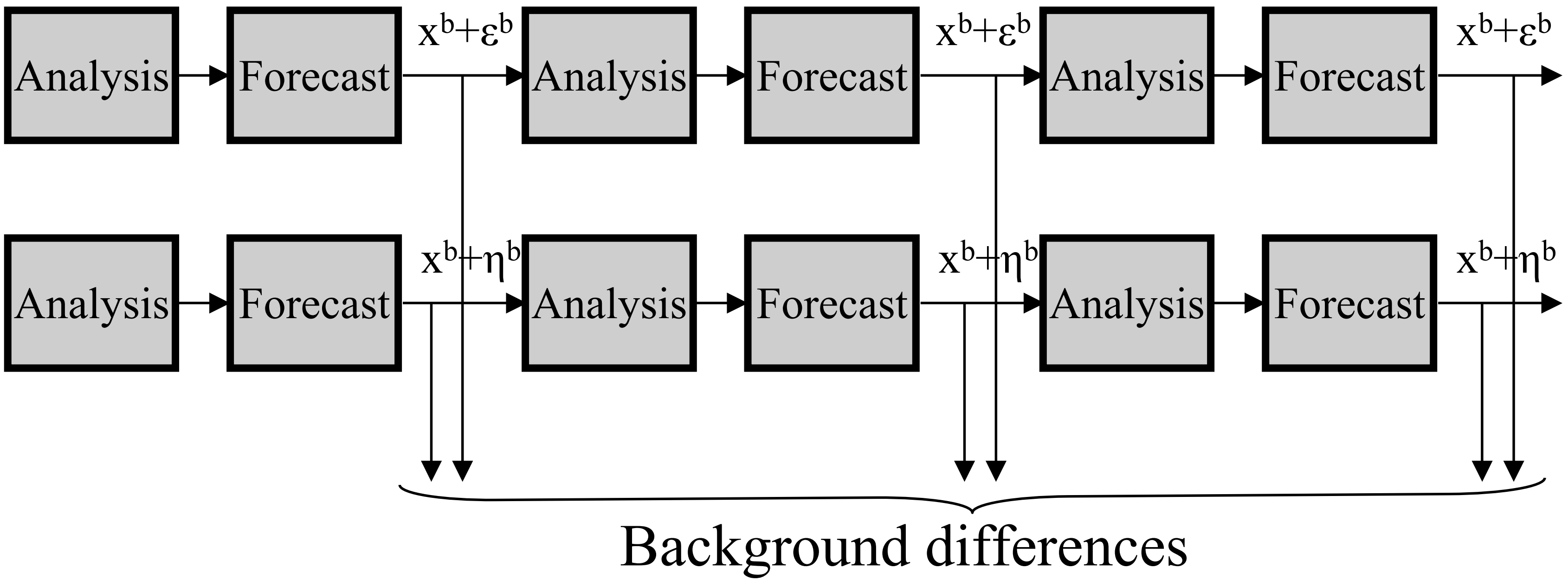
- **Suppose we perturb all the inputs to the analysis/forecast system with random perturbations, drawn from the relevant distributions:**



- **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 several cycles, the system will have forgotten the original initial background perturbations.**

Estimating Background Error Statistics from Ensembles of Analyses

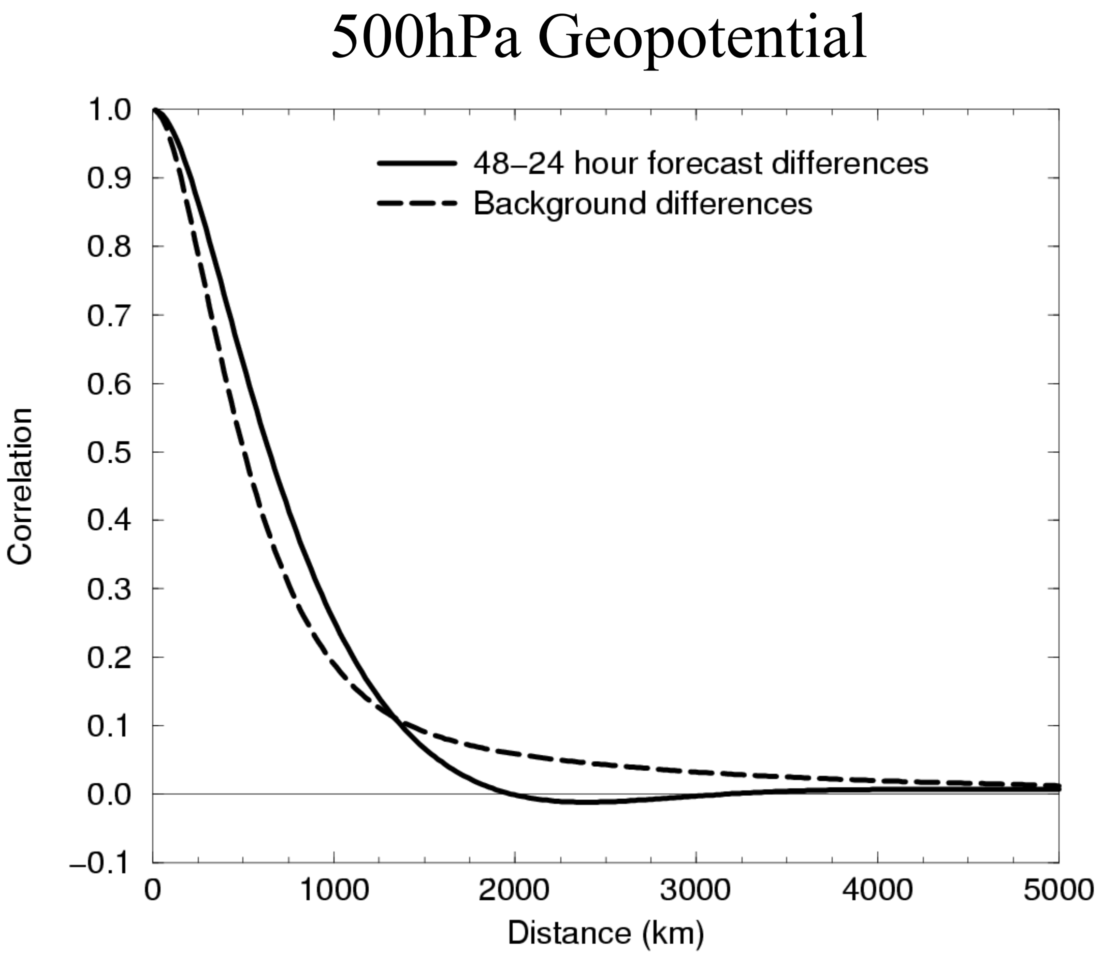
- **Run the analysis system several times with different perturbations, and form differences between pairs of background fields.**
- **These differences will have the statistical characteristics of background error (but twice the variance).**



Estimating Background Error Statistics from Ensembles of Analyses

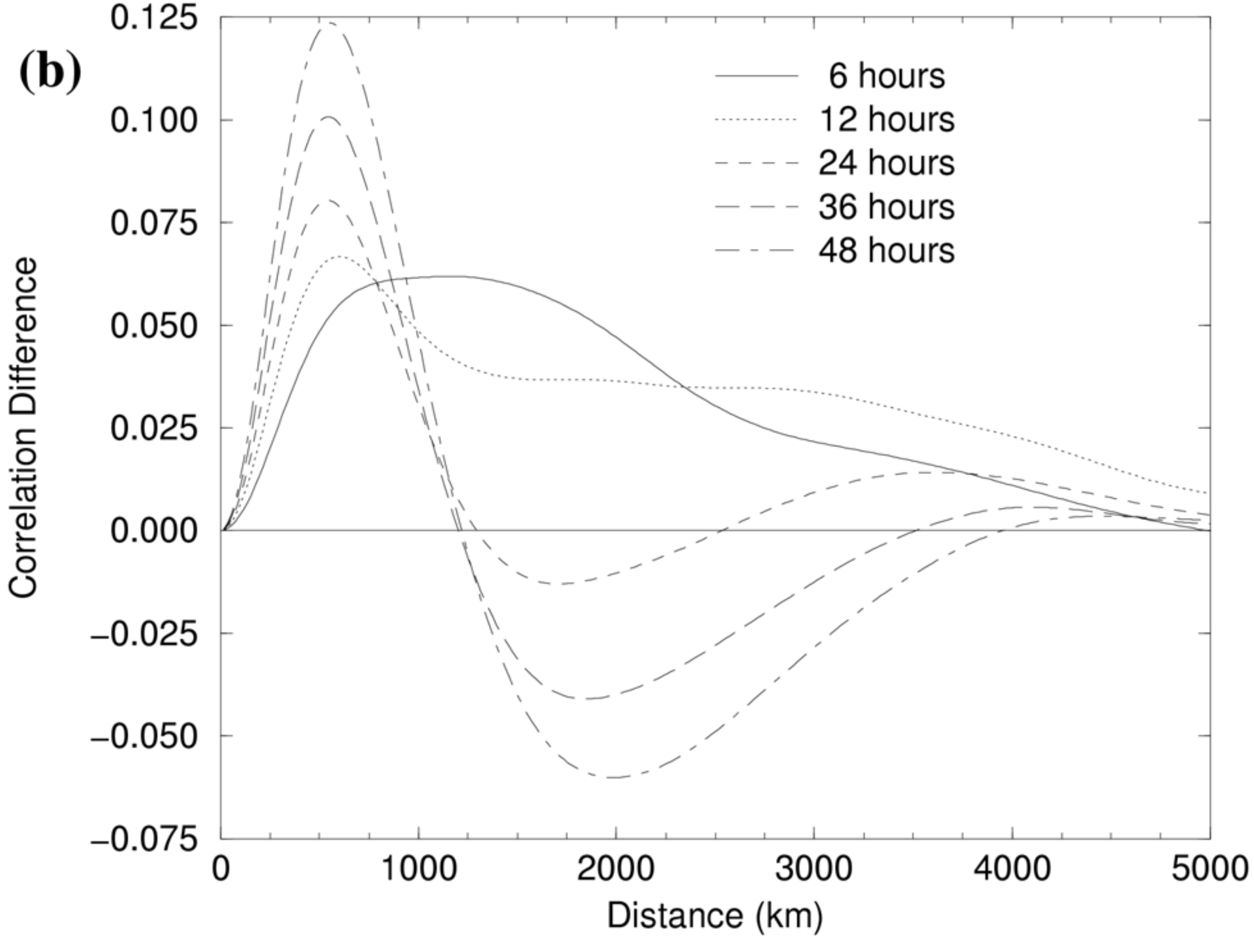
- **Advantages:**
  - The method diagnoses the error characteristics of the actual analysis/forecast system.
  - Analysis error and forecast error at any range can be diagnosed.
  - Does not impose constraints on the observations used, provided their error characteristics are known.
  - Produces global statistics of model variables.
  - Produces good estimates in data-sparse regions.
- **Disadvantages:**
  - Computationally Expensive.
  - Assumes a perfect model. (We could add model error if we knew what it looked like!)
  - Assumes observation (SST, etc.) error characteristics are known.
  - Danger of feedback. Eg: A noisy analysis system => unbalanced stats => Even noisier analysis system.

Estimating Background Error Statistics from Ensembles of Analyses



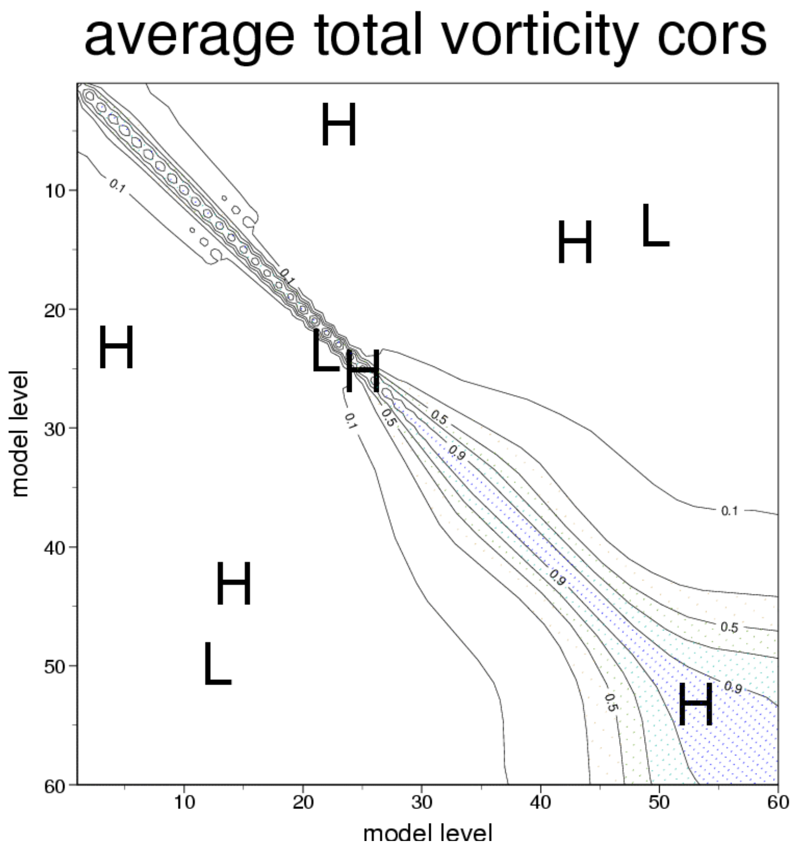
ECMWF

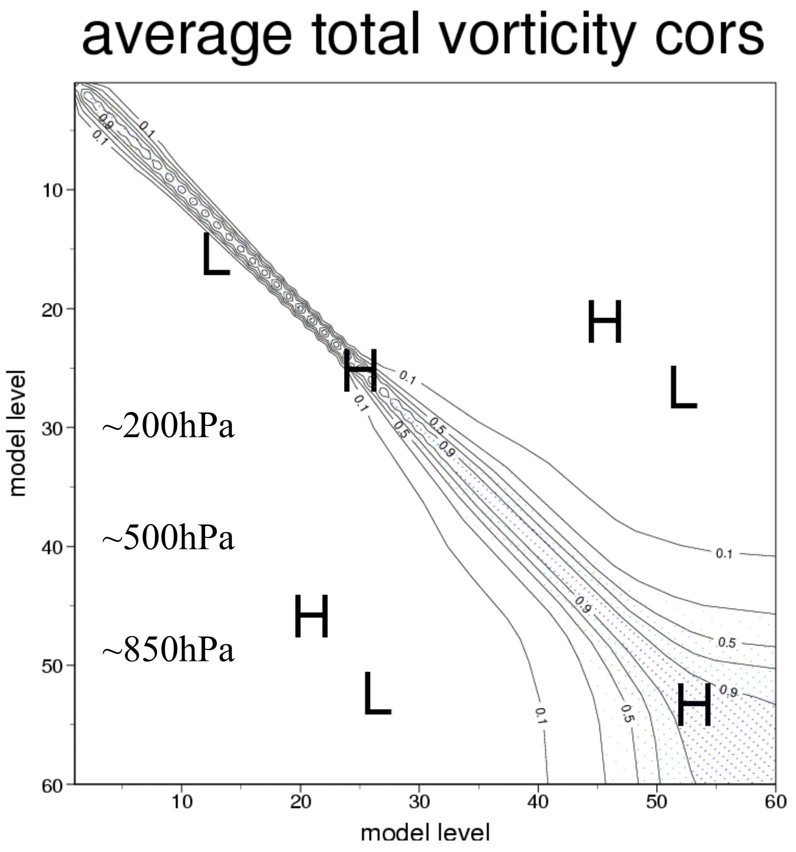
Estimating Background Error Statistics from Ensembles of Analyses



Forecast correlation minus analysis correlation for 500hPa geopotential

Estimating Background Error Statistics from Ensembles of Analyses





Modelling the Statistics

- **Problem:**
  - The background error covariance matrix is big: ~10^7x10^7.
  - We cannot generate enough statistical information to specify this many elements.
  - We cannot store a matrix this big in computer memory.

- **To reduce the problem to a manageable size, the matrix is split into a product of very sparse matrices, E.g:**

$$B = L^T \Sigma^T C \Sigma L$$

  - L = balance operator (accounts for inter-variable correlations)
  - Σ = diagonal (in grid space) matrix of standard deviations
  - C = correlation matrix (block diagonal – one block per variable)

Modelling the Correlations

- Three main approaches to modelling the correlations:
  - Spectral convolutions (Courtier et al., 1998, QJRMS pp1783…)
  - Digital filters (Lorenc 1997 J Met Soc Japan pp339… ; Parrish et al. 1997 J Met Soc Japan 359… ; Purser et al. 2001)
  - Diffusion equations (Weaver and Courtier, 2001, QJRMS pp1815…)

Modelling the Correlations – Spectral Method

- **The spectral method uses the equivalence between:**
  - Convolution with a function h of great-circle distance.
  - Multiplication of spectral coefficients by a function of total wavenumber, n.
- **For h a function of great-circle distance:**

$$h \otimes f = \sum_{m,n} h_n f_{m,n} Y_{m,n}(\lambda, \phi)$$

- **Horizontal correlations may be represented by a diagonal matrix, H, of coefficients, h_n:**
  - $C = H^T V H$
  - Horizontal correlations are homogeneous and isotropic
- **An advantage of the spectral method is that the coefficients, h_n, are easily derived:**
  - h_n = (variance for wavenumber n) / (total variance)

Modelling the Correlations – Spectral Method

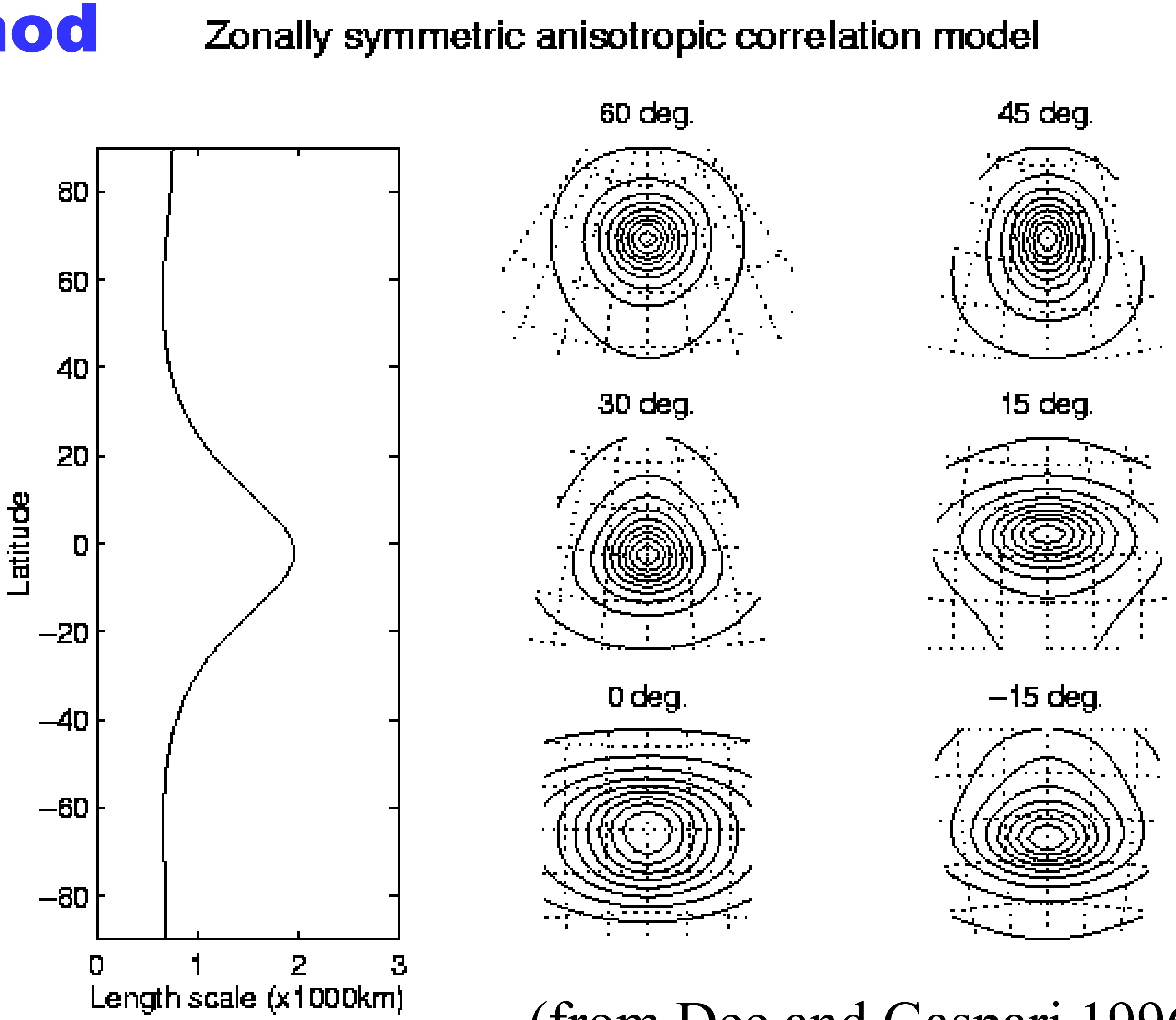
- $C=H^TVH$. V models vertical correlations.
- In the ECMWF system, V is block diagonal, with one block for each n.
  - Non-separable – small horizontal scales have shallow vertical correlation.
  - Non-separability is necessary to get both mass and wind correlations right (Phillips, 1986 Tellus pp314…).
- Other decompositions are possible. E.g. UKMO use:
  - $C = P^TV\ H\ VP$
  - Where P is a projection onto eigenvectors of the global mean vertical correlation matrix
  - V is a diagonal, but its diagonal elements vary with latitude.
  - H is diagonal, but with different coefficients h_n for each vertical eigenvector.

Including Anisotropy in the Spectral Method

- **Dee and Gaspari, 1996:**
  - If $c_0(\underline{x},\underline{y})$ is a correlation function, then so is $c_0(g(\underline{x}),g(\underline{y}))$
  - If $c_0(\underline{x},\underline{y})$ is isotropic, then $c(\underline{x},\underline{y})=c_0(g(\underline{x}),g(\underline{y}))$ is generally anisotropic.
  - We can implement the anisotropic correlation model $c(\underline{x},\underline{y})$ as a two-step process:
    - horizontal coordinate transform: $\underline{X} = g(\underline{x})$
    - Isotropic correlation model: $c_0(\underline{X},\underline{Y})$
- **Dee and Gaspari use a simple function of latitude for $g(\underline{x})$**
- **Desroziers (1997, MWR pp3030…) suggests momentum coordinates:**

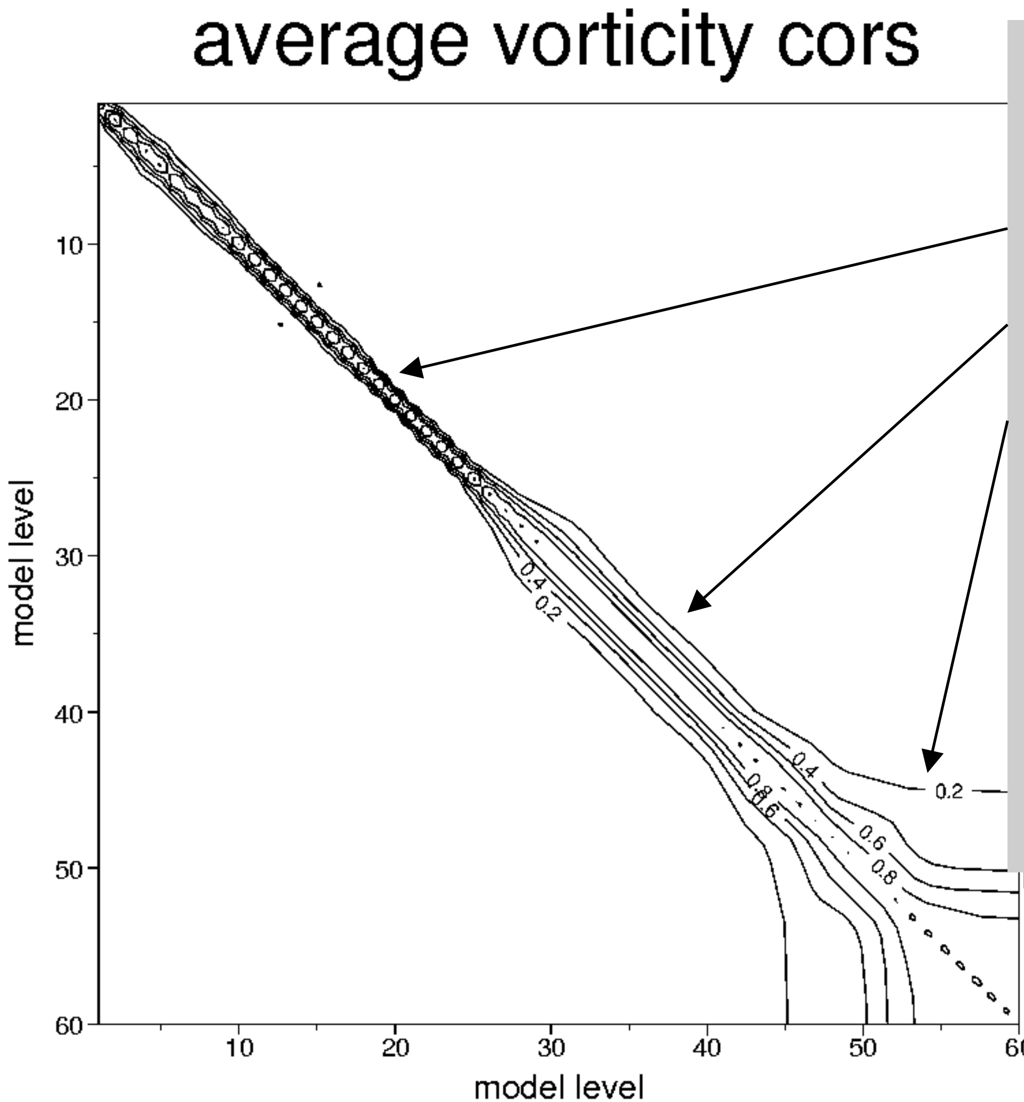
$$\underline{\mathbf{X}} = \underline{\mathbf{x}} + \frac{1}{f}\left(\underline{\mathbf{k}} \times \underline{\mathbf{v}}_g\right)$$

Including Anisotropy in the Spectral Method



(from Dee and Gaspari 1996)

Vertical Coordinate Transforms



Vertical correlations show 3 distinct regions:

Stratosphere

Free Troposphere

Boundary Layer

We could improve the description of vertical correlations by making the boundary-layer top and the tropopause coordinate surfaces.

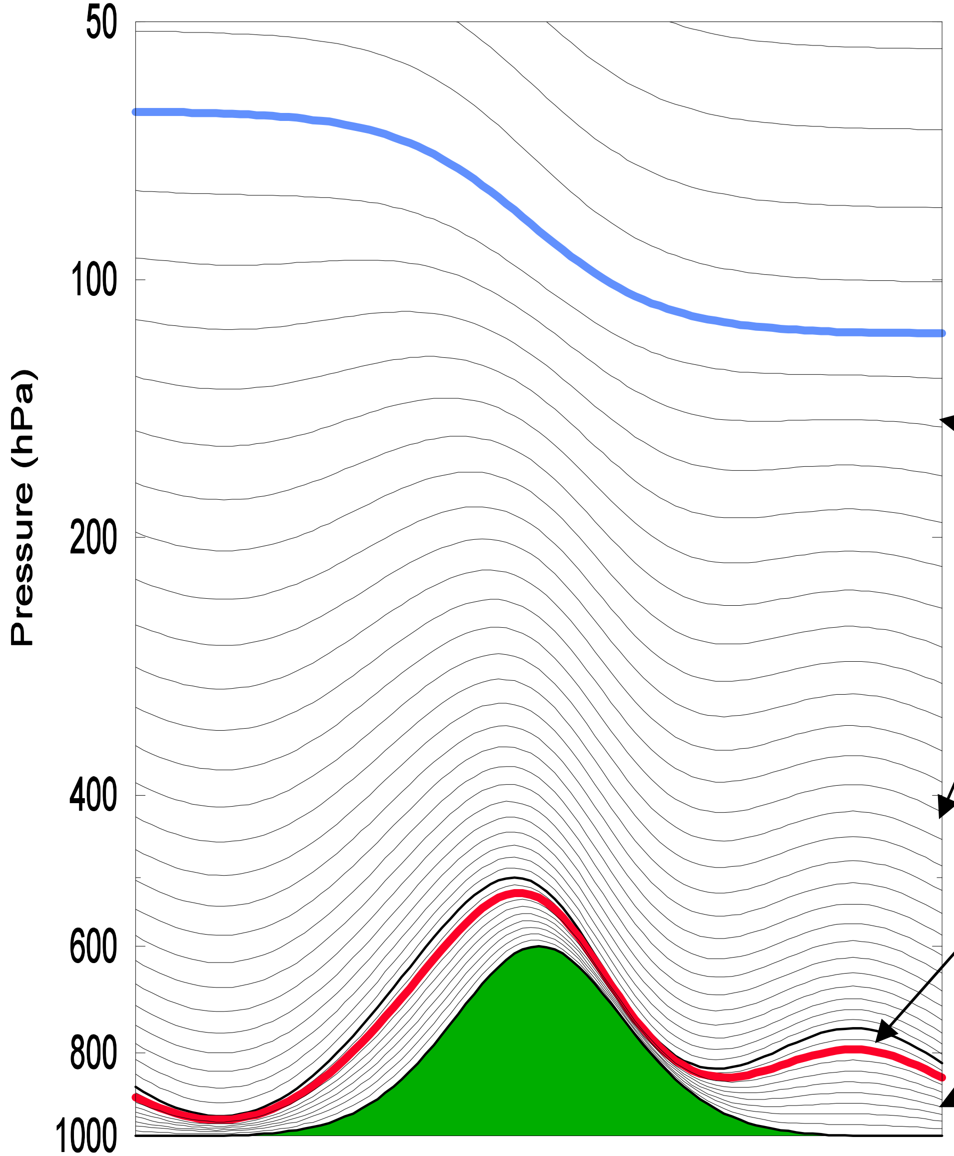
Vertical Coordinate Transforms

- <u>One</u> possibility is to use a boundary-layer following variant of the coordinate introduced by Zhu et al. (1992):

$$p_{k+1/2} = \begin{cases} \dfrac{a_{k+1/2} + b_{k+1/2}\, p_B}{c_{k+1/2}\left(T_{k+1/2}\right)^{-1/\kappa} + d_{k+1/2}} & \text{for } k < K_B \\[2ex] p_B \left(\dfrac{p_*}{p_B}\right)^{\frac{k-K_B}{N-K_B}} & \text{for } k \geq K_B \end{cases}$$

- This gives:
  - Level K_B*+1/2* is the boundary-layer top ($p=p_B$).
  - Level *N+1/2* is the surface ($p=p_*$).
  - Levels are evenly spaced in log(*p*) in the boundary layer.
  - Smooth transition between a hybrid pressure coordinate in the lower troposphere, and an isentropic coordinate in the upper troposphere.

Vertical Coordinate Transforms



Fully Isentropic coordinate in the stratosphere.

Hybrid pressure/isentropic coordinate in the free troposphere.

Boundary-layer top is a coordinate surface.

Levels evenly spaced in $\log(p)$ between p_B and p_*.

Including Inhomogeneity in the Spectral Method

- The spectral method gives us full control over the spectral variation of covariance (one matrix for each n), but no control over their spatial variation (homogeneous).
- Specifying covariances in grid space allows us full control over the spatial variation of covariances, but no control over their spectral variation (one matrix for all n).
- By using a wavelet transform, we can compromise between these extremes, and gain partial control over both spectral and spatial variation.

Including Inhomogeneity in the Spectral Method

- A non-orthogonal wavelet transform on the sphere may be defined by a set of functions of great-circle distance:

$$\left\{ \psi_j \left(\left| \mathbf{r} \right| \right) ;\ j = 1 \ldots K \right\}$$

with the property: $\sum_j \hat{\psi}_j^2(n) = 1$

- We then have a "transform pair":

$$f_j = \psi_j \otimes f, \quad f = \sum_j \psi_j \otimes f_j$$

Proof:

$$\left(\widehat{\sum_j \psi_j \otimes f_j} \right)_{mn} = \sum_j \hat{\psi}_j(n) \left(\hat{f}_j \right)_{mn} = \sum_j \hat{\psi}_j^2(n) \hat{f}_{mn} = \hat{f}_{mn}$$

Including Inhomogeneity in the Spectral Method

- **We arrange for $\hat{\psi}_j(n)$ to pick out bands of wavenumbers. For example:**

$$\hat{\psi}_j(n) = \left(\frac{n - N_{j-1}}{N_j - N_{j-1}} \right)^{1/2} \qquad N_{j-1} < n \le N_j$$

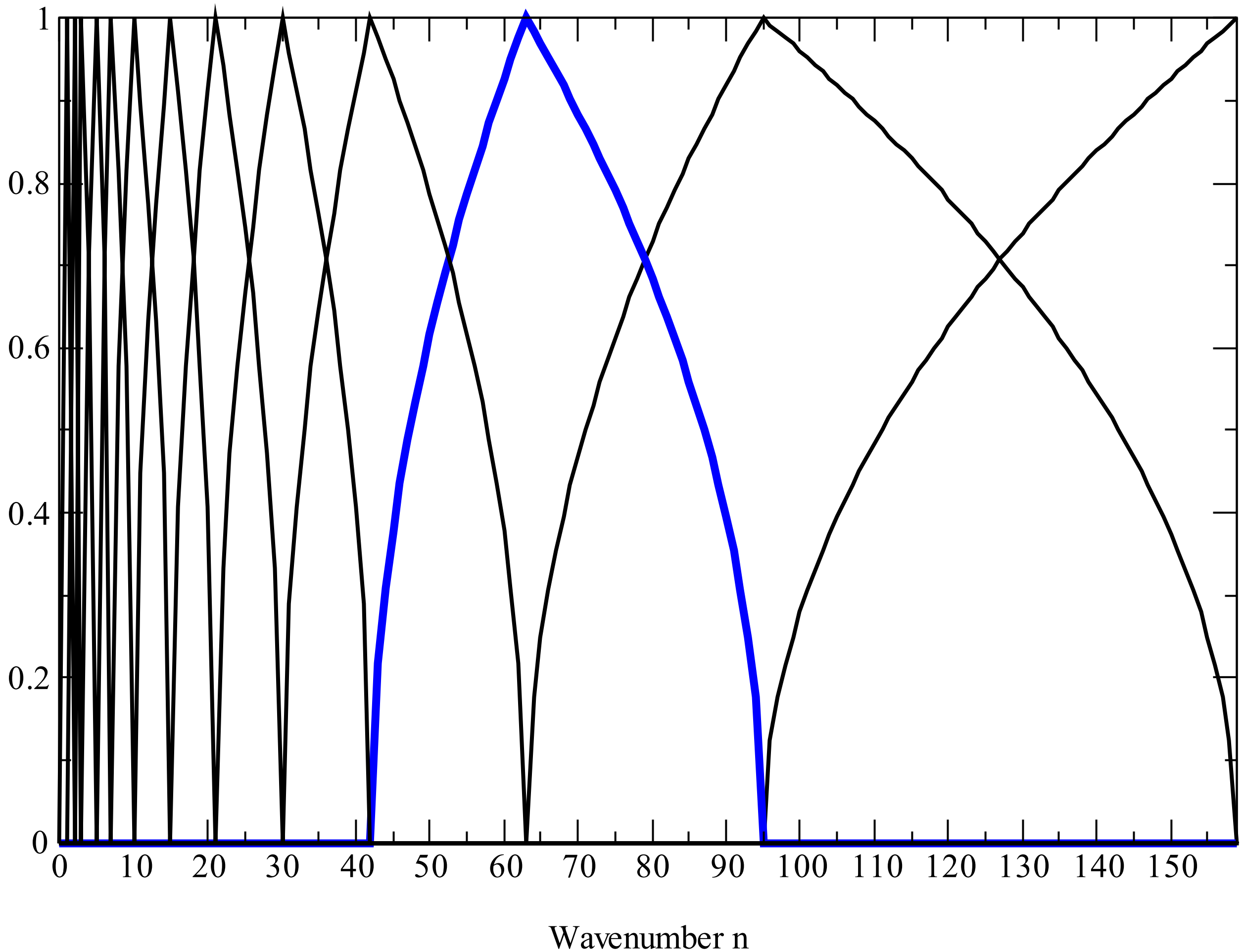
$$\hat{\psi}_j(n) = \left(\frac{N_{j+1} - n}{N_{j+1} - N_j} \right)^{1/2} \qquad N_j < n < N_{j+1}$$

$$\hat{\psi}_j(n) = 0 \qquad \text{otherwise}$$

(square-root of a triangle function)

Including Inhomogeneity

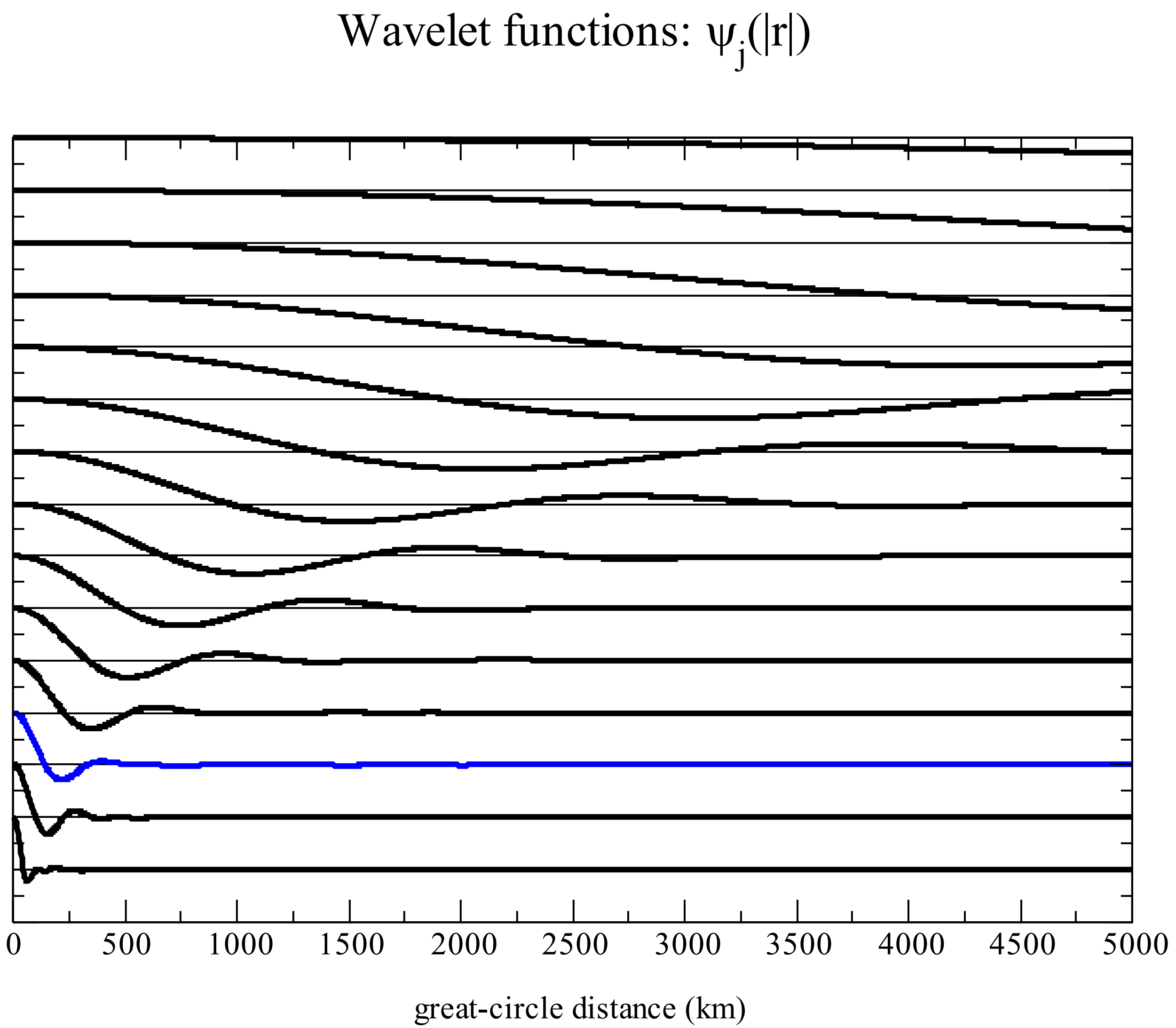
Wavelet functions: $\hat{\psi}_j(n) = \left(\hat{\phi}_j^2(n) - \hat{\phi}_{j-1}^2(n) \right)^{1/2}$



Including Inhomogeneity in the Spectral Method

- In physical space, the functions $\psi_j(|\mathbf{r}|)$ decay with great-circle distance, |r|.

Including Inhomogeneity



ECMWF

Including Inhomogeneity in the Spectral Method

- **The wavelet functions $\psi_j\left(\left|\mathbf{r}\right|\right)$ are localized in wavenumber and localized spatially.**

- **The transform property:** $f = \sum_j \psi_j \otimes f_j$

  means we can regard $f_j(\lambda, \phi)$ as the coefficient of a spatially- and spectrally localized function:

$$\psi_j\left(\left|\mathbf{r}(\lambda', \varphi') - \mathbf{r}(\lambda, \varphi)\right|\right)$$

Including Inhomogeneity in the Spectral Method

- We construct the Wavelet J_b by providing one vertical covariance matrix $\mathrm{C}_j(\lambda,\phi)$ for each gridpoint and for each waveband, ***j***.
- $\mathrm{C}_j(\lambda,\phi)$ accounts for both horizontal and vertical correlation. It is roughly equivalent to $h_n^2 \mathrm{V}_n$ in the current ECMWF spectral J_b formulation.
- This allows us to provide spatial and spectral variation of vertical and horizontal background error covariances.

Including Inhomogeneity in the Spectral Method

- **The Wavelet J_b defines the control variable to be:**

$$\chi^{\mathrm{T}} = \left(\chi_1^{\mathrm{T}}, \chi_2^{\mathrm{T}}, \ldots, \chi_K^{\mathrm{T}}\right)$$

  - **where** $\chi_j = \mathbf{C}_j^{-1/2}(\lambda,\phi)\left(\psi_j \otimes \boldsymbol{\Sigma}_b^{-1/2}(\mathbf{x}-\mathbf{x}_b)\right)$

- **This gives:** $\mathbf{x}-\mathbf{x}_b = \boldsymbol{\Sigma}_b^{1/2} \sum_j \psi_j \otimes \left[\mathbf{C}_j^{1/2}(\lambda,\phi)\chi_j\right]$

- **The corresponding background error covariance matrix is (schematically):**

$$\mathbf{B} = \boldsymbol{\Sigma}_b^{1/2}\left(\sum_j \psi_j^2 \otimes \left[\mathbf{C}_j(\lambda,\phi)\right]\right)\boldsymbol{\Sigma}_b^{1/2}$$

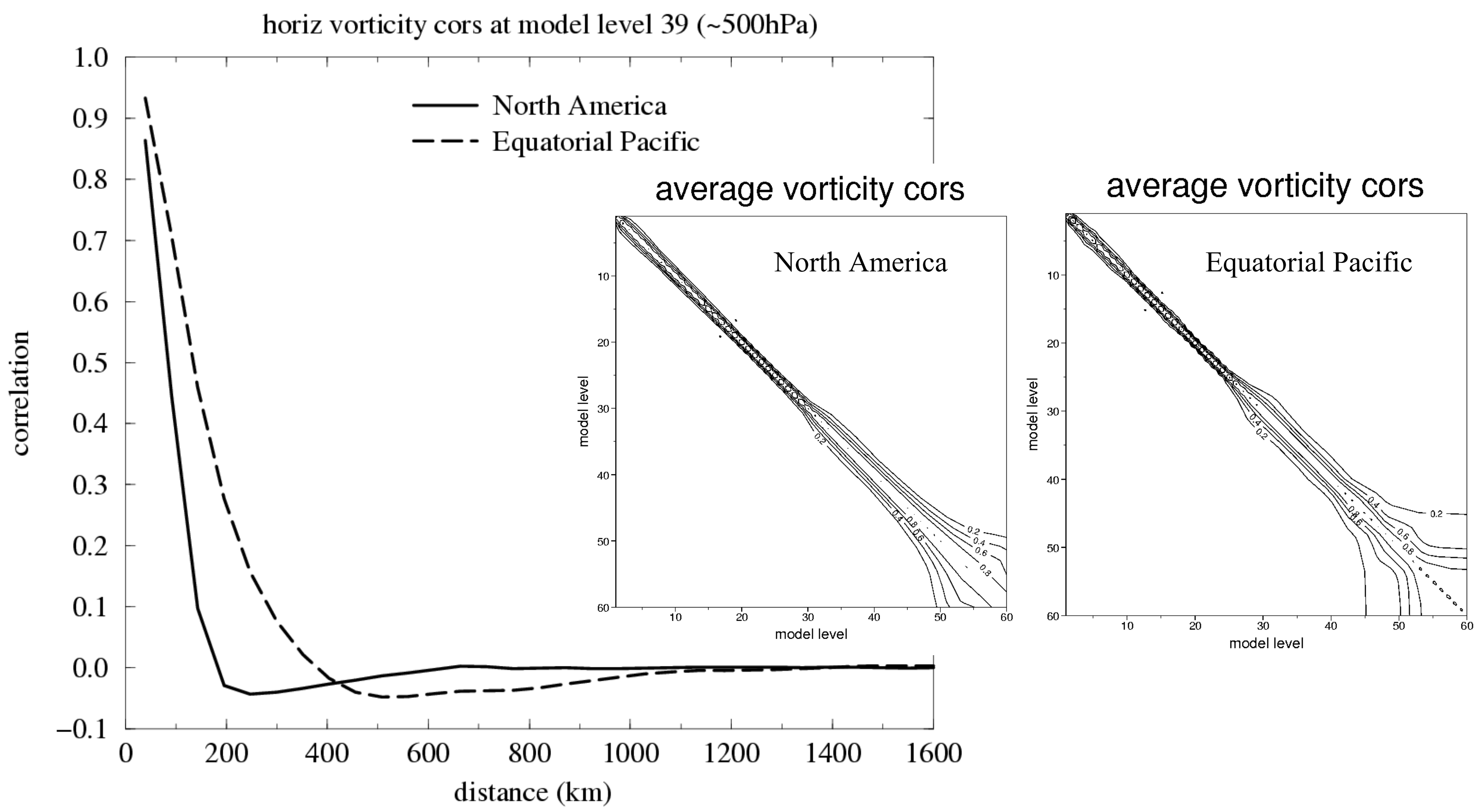
Including Inhomogeneity in the Spectral Method

$$\mathbf{B} = \mathbf{\Sigma}_b^{1/2} \left(\sum_j \psi_j^2 \otimes \left[\mathbf{C}_j(\lambda,\phi) \right] \right) \mathbf{\Sigma}_b^{1/2}$$

- **Remember,** $\sum_j \hat{\psi}_j^2(n) = 1$ **.**
- **So, for a given wavenumber n, B is a weighted average of the matrices** $\mathbf{C}_j(\lambda,\phi)$ **.**
- **At a given gridpoint, B is determined by matrices at neighbouring gridpoints (i.e. gridpoints where** $\psi_j(|\mathbf{r}|)$ **is not close to zero.**

Including Inhomogeneity in the Spectral Method

- **Spherical wavelet transform:**



ECMWF

Incorporating Balance

- One approach is to provide separate J_b cost functions for the balanced and unbalanced components:

$$J_b = (\mathbf{x} - \mathbf{x}_b)_{bal}^T \mathbf{B}_{bal}^{-1} (\mathbf{x} - \mathbf{x}_b)_{bal} + (\mathbf{x} - \mathbf{x}_b)_u^T \mathbf{B}_u^{-1} (\mathbf{x} - \mathbf{x}_b)_u$$

  - With balanced/unbalanced components determined, for example, by projection onto normal modes.

- There are two problems with this:
  - We still need to describe variable-to-variable correlations in B.
  - This J_b cannot be written as $\chi^T \chi$ (unless we double the size of the control vector), making it difficult to choose a control variable for the minimization with good preconditioning properties.

Incorporating Balance

- A better approach was implemented by Derber and Bouttier (1999 Tellus pp195…).
- The use a change of variable:

$$\begin{pmatrix} \zeta_{bal} \\ D_u \\ T_u \\ (p_s)_u \end{pmatrix} = \mathbf{L} \begin{pmatrix} \zeta \\ D \\ T \\ p_s \end{pmatrix}$$

  - Subscripts *bal* and *u* denote “balanced” and “unbalanced”, with for example, $T=T_{bal}+T_u$.

- The transformed variables are treated univariately.
- The balance relationships may be defined analytically, or determined statistically (or a bit of both).

Incorporating Balance

- **Derber and Bouttier's balance operator is:**

$$
\begin{aligned}
\zeta_{bal} &= \zeta \\
D_u &= D - D_{bal}\left(P_{bal}(\zeta)\right) \\
T_u &= T - T_{bal}\left(P_{bal}(\zeta)\right) - T_{div}(D_u) \\
(p_s)_u &= p_s - (p_s)_{bal}\left(P_{bal}(\zeta)\right) - (p_s)_{div}(D_u)
\end{aligned}
$$

- **P_{bal} is a linearized mass variable, determined by statistical regression between spectral coefficients of vorticity and geopotential.**
- **T_{bal} (etc.) is determined by statistical regression between geopotential and temperature (etc.).**
- **T_{div} [and $(p_s)_{div}$] are given by statistical regression between temperature [and p_s] and divergence.**

Nonlinear Balance

- DB99 determine P_{bal} from ζ by regression.
- But, the resulting P_{bal} is nearly indistinguishable from that implied by linear balance.
- A more accurate balance relationship can be achieved using the non-linear balance equation:

$$\nabla^2 P_{bal} = -\nabla . \left(\mathbf{v}_\psi . \nabla \mathbf{v}_\psi + f \mathbf{k} \times \mathbf{v}_\psi \right)$$

- Linearizing about the background state gives a linear, but flow dependent balance operator.
- Nonlinear balance:
  - is important in jet entrance/exit regions
  - describes some tropical modes well

Omega Equation / Richardson's Equation

- **DB99 determine D_{bal} by statistical regression with geopotential.**
- **The regression describes Ekman pumping, but little else.**
- **Augmenting the regression with an analytical equation for balanced divergence should allow the divergence field to be described more accuratey.**
  - **ECMWF now uses a quasi-geostrophic omega equation:**

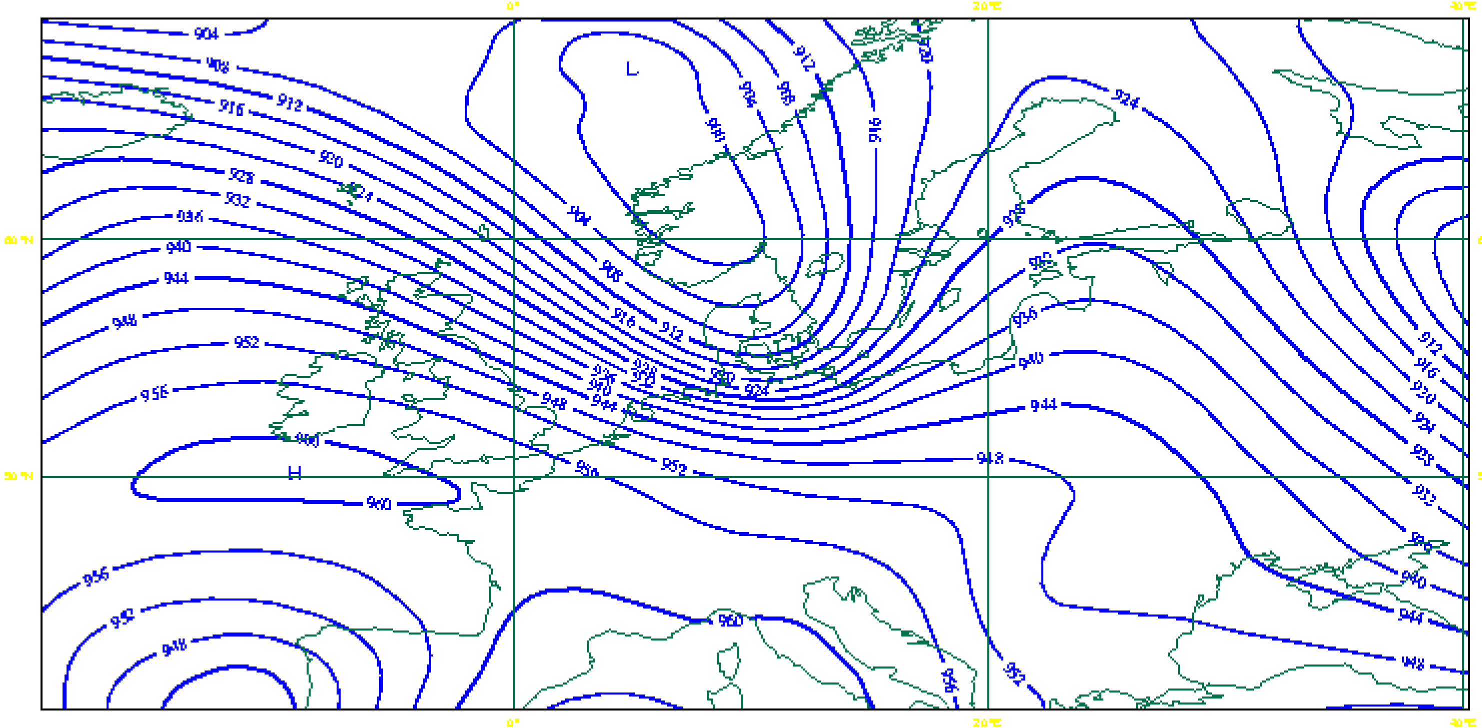
$$(\sigma\nabla^2 + f_0^2 \frac{\partial^2}{\partial p^2})\omega = -2\nabla.\mathbf{Q}$$

  - **UKMO is trying Richardson's equation:**

$$\gamma \frac{\partial}{\partial z}\left\{ p\left[\frac{\partial w}{\partial z} + \nabla_z.\mathbf{v}_{bal} - \frac{Q}{TC_p}\right]\right\} = \frac{\partial p}{\partial z}\nabla_z.\mathbf{v}_{bal} - \frac{\partial \mathbf{v}}{\partial z}.\nabla_z p$$

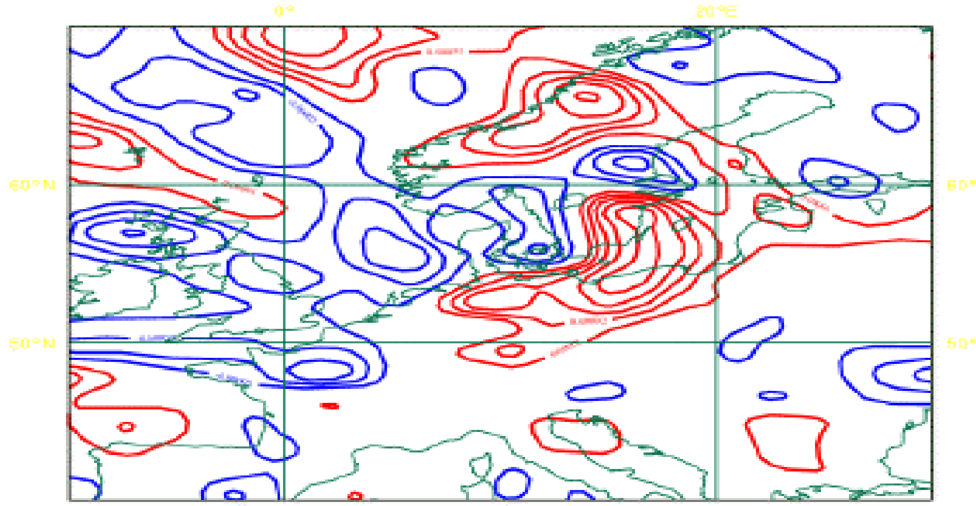
Incorporating Balance

ECMWF Analysis VT:Monday 27 August 2001 12UTC 300hPa geopotential height

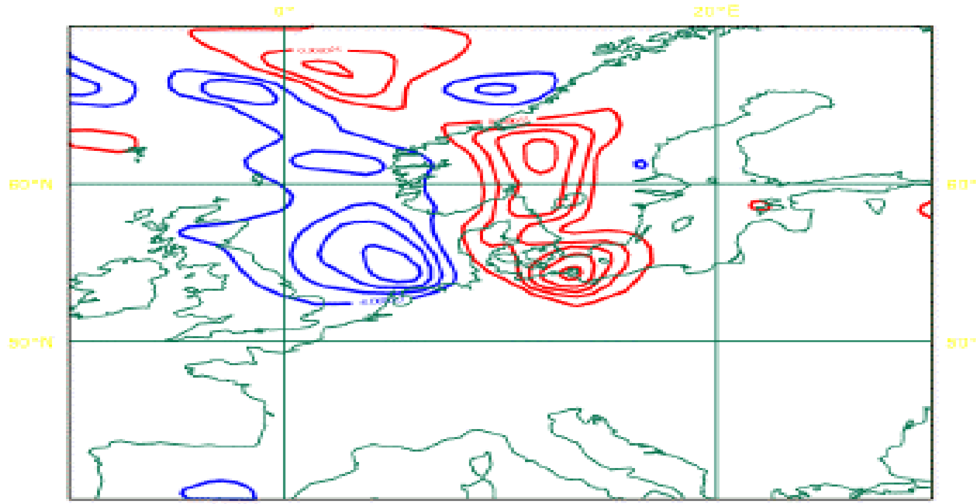


Quasi-geostrophic Balance Operator

Control 3dVar inner-loop analysis 2001/08/27 level 33 divergence.



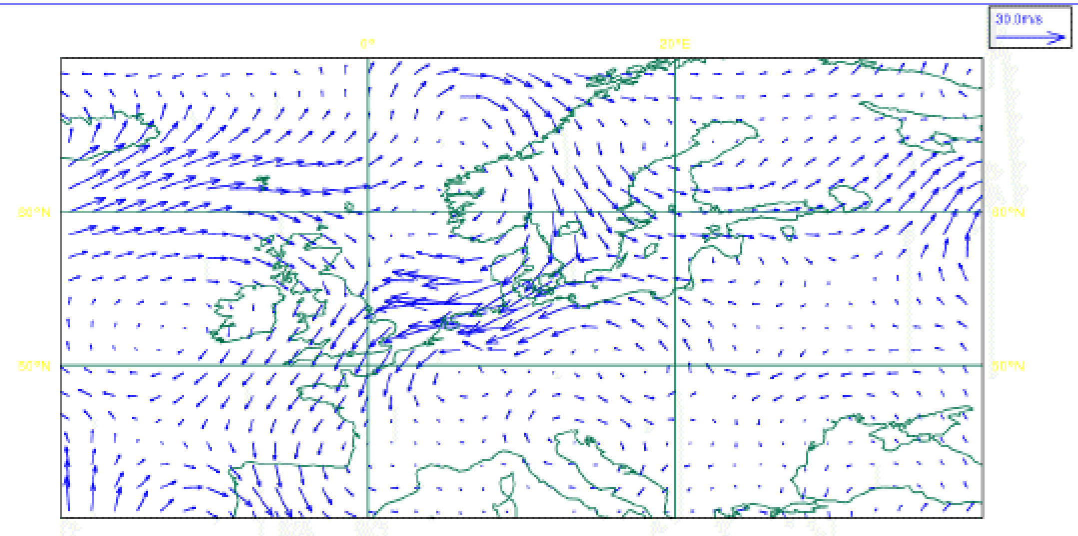
Divergence at level 33 diagnosed from the balanced flow.



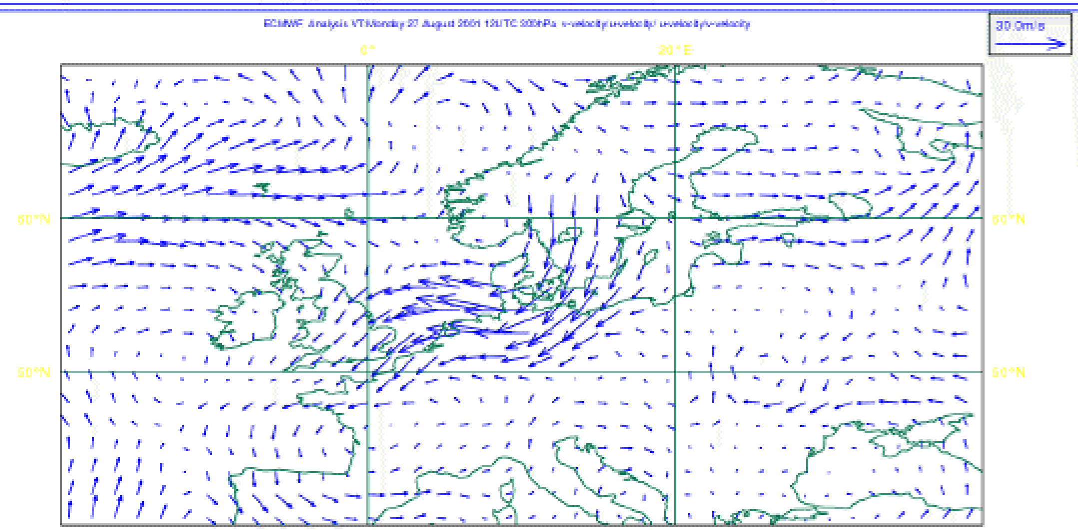
ECMWF

Nonlinear and Quasi-Geostrophic Balance Operator

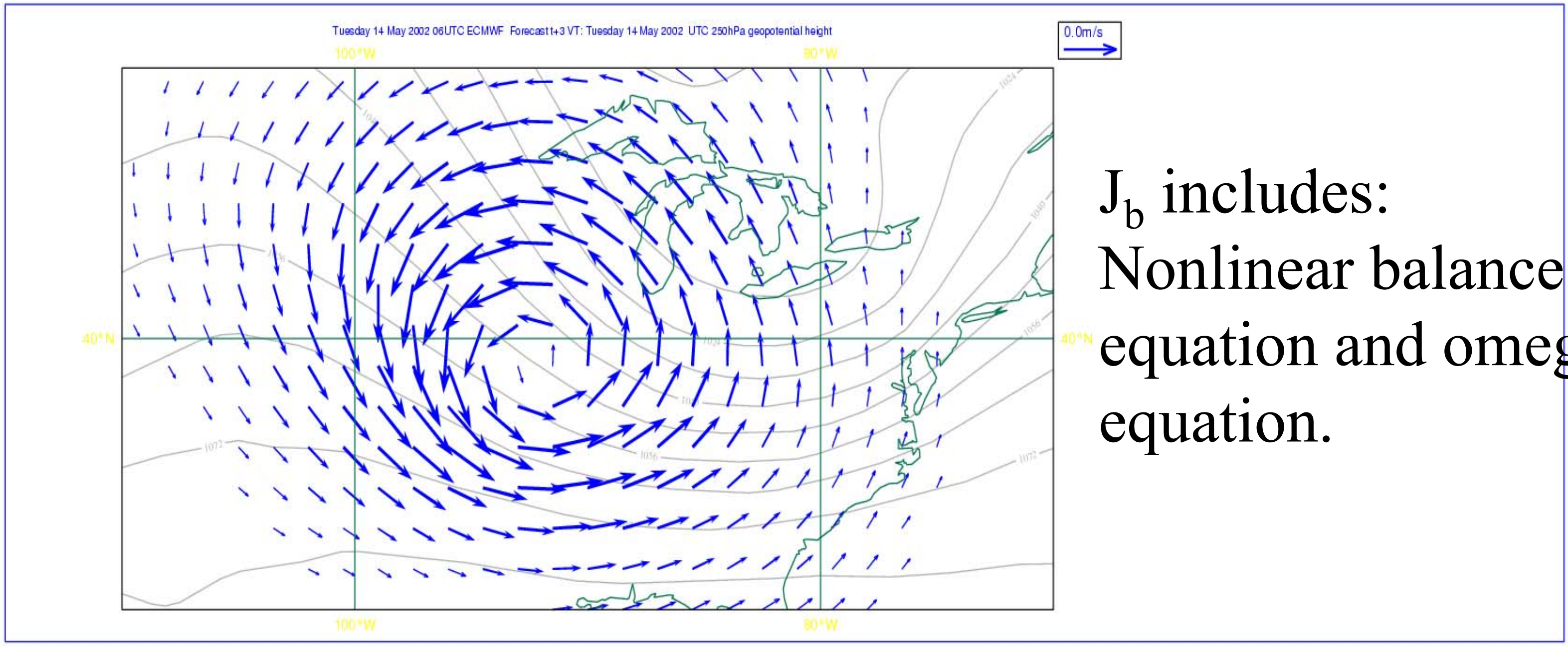
ECMWF analysis 2001/08/27 300hPa ageostrophic wind.



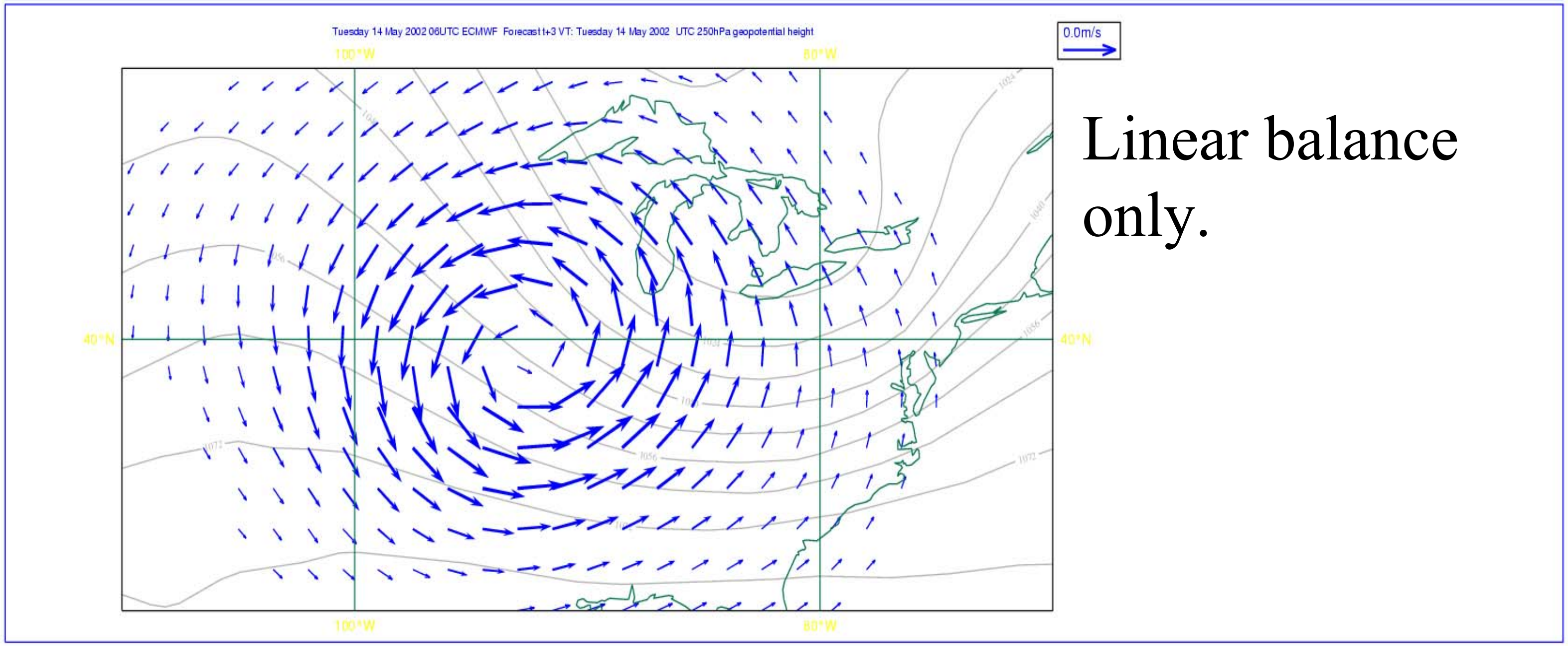
Ageostrophic wind at model level 33 diagnosed from the balanced flow.



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

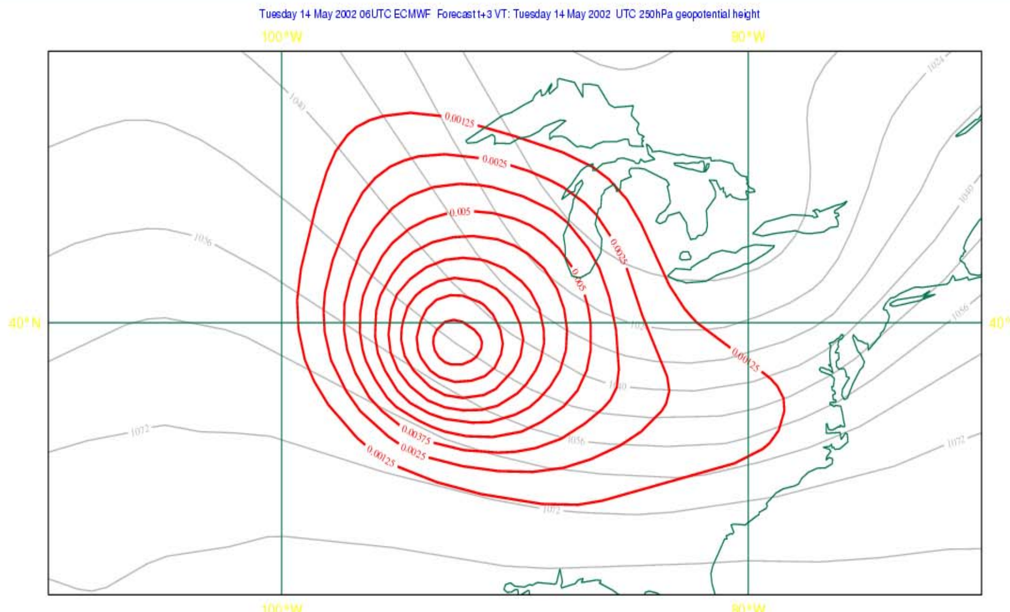


J_b includes: Nonlinear balance equation and omega equation.

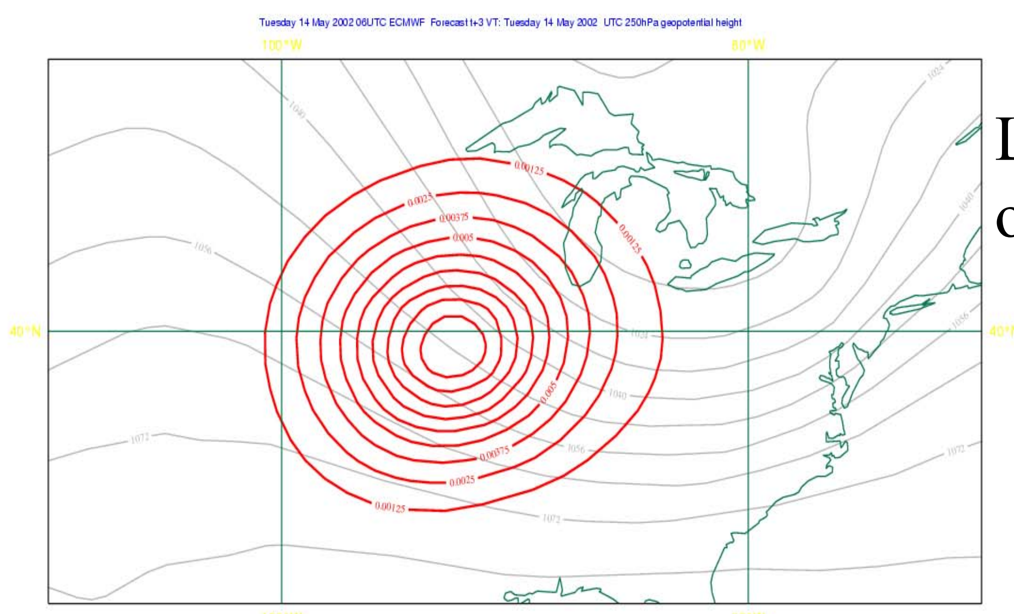


Linear balance only.

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

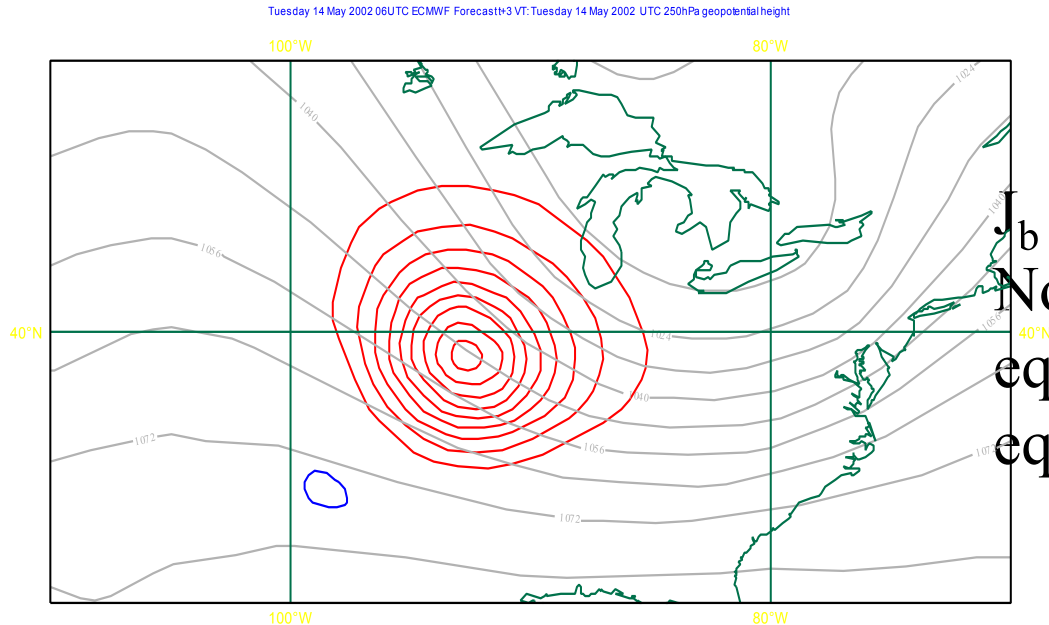


J_b includes: Nonlinear balance equation and omega equation.

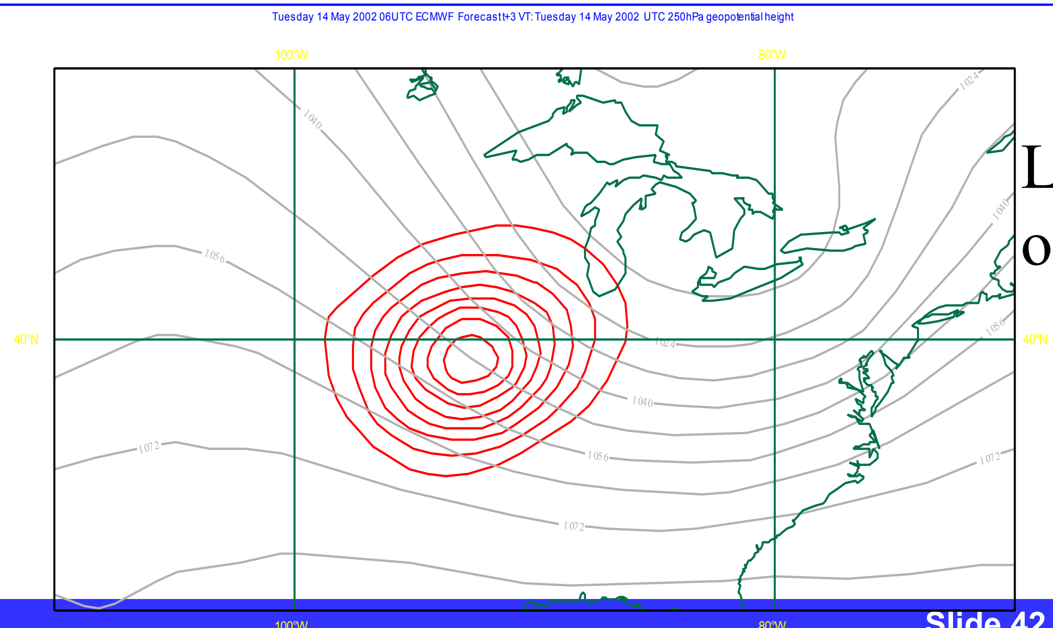


Linear balance only.

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

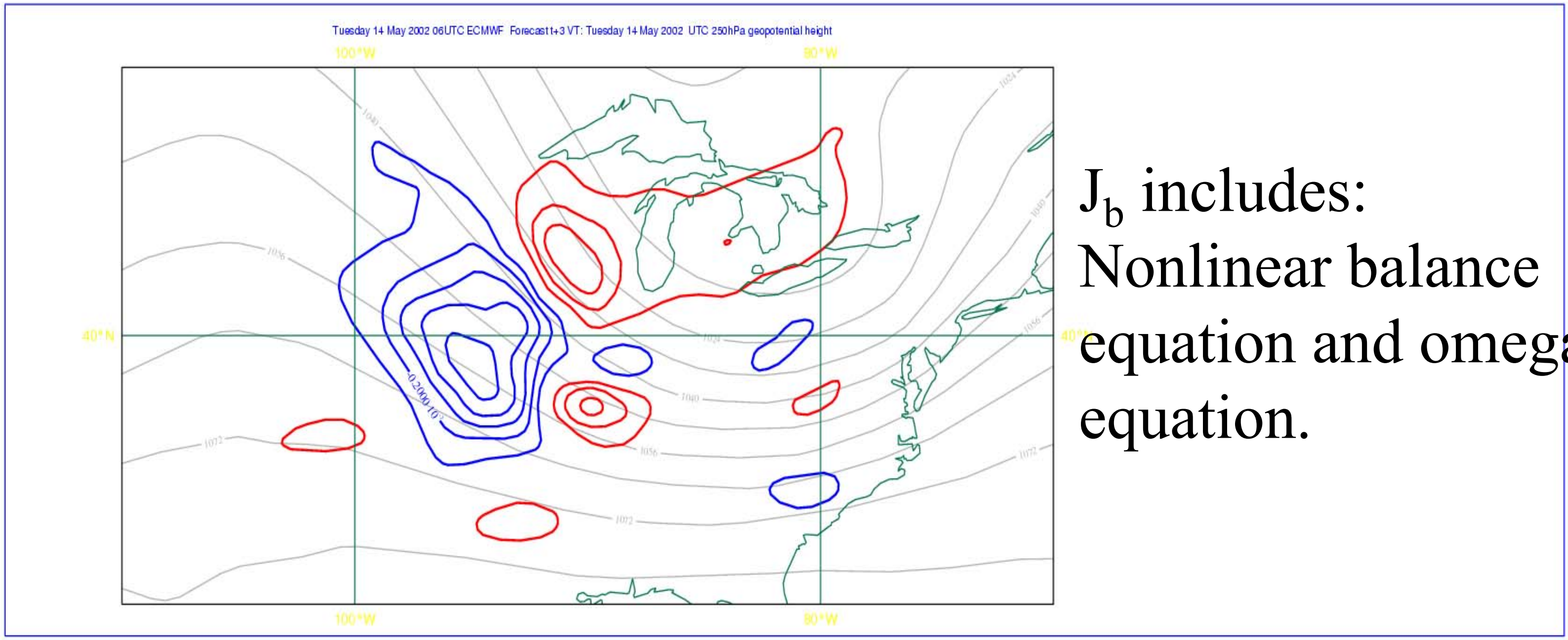


J_b includes: Nonlinear balance equation and omega equation.

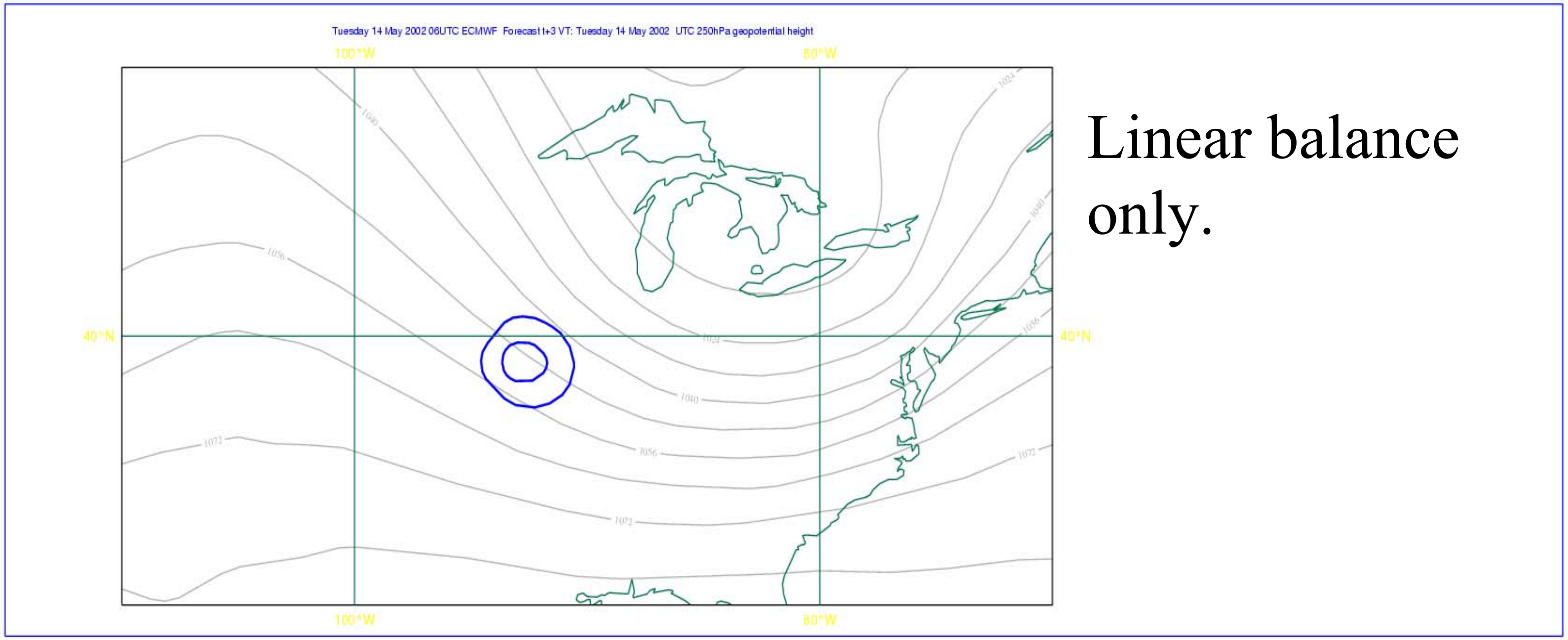


Linear balance only.

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



J_b includes: Nonlinear balance equation and omega equation.



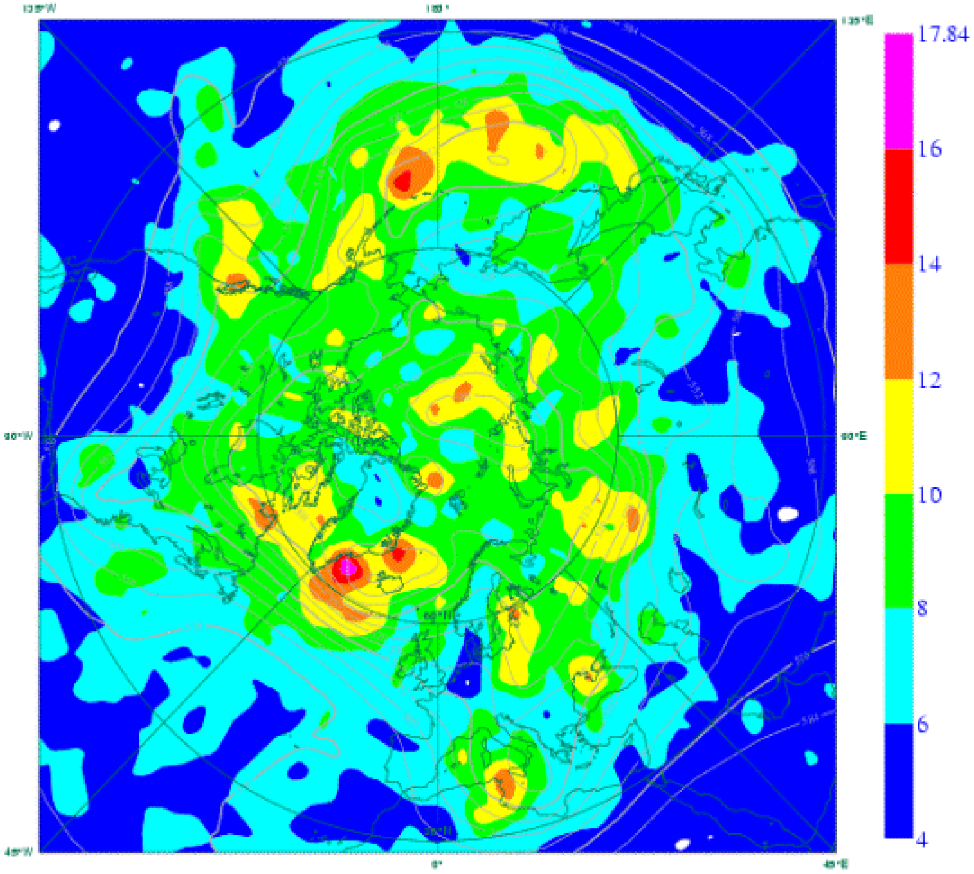
Linear balance only.

An Added Bonus: Flow-dependent σ_b!

Shaded: Diagnosed background error for geopotential on model level 39.

Contoured: 500hPa height.

ECMWF Analysis VT:Saturday 1 February 2003 12UTC 500hPa geopotential height



Conclusions

- The background error covariance matrix is vitally important for any data assimilation method.
- Ensembles of analyses provide an attractive method for diagnosing statistics of background error.
- There is a clear link with ensemble Kalman filtering.
- Anisotropy, Inhomogeneity and flow-dependence can all be incorporated (in a variety of different ways) without necessarily requiring a Kalman filter approach.