

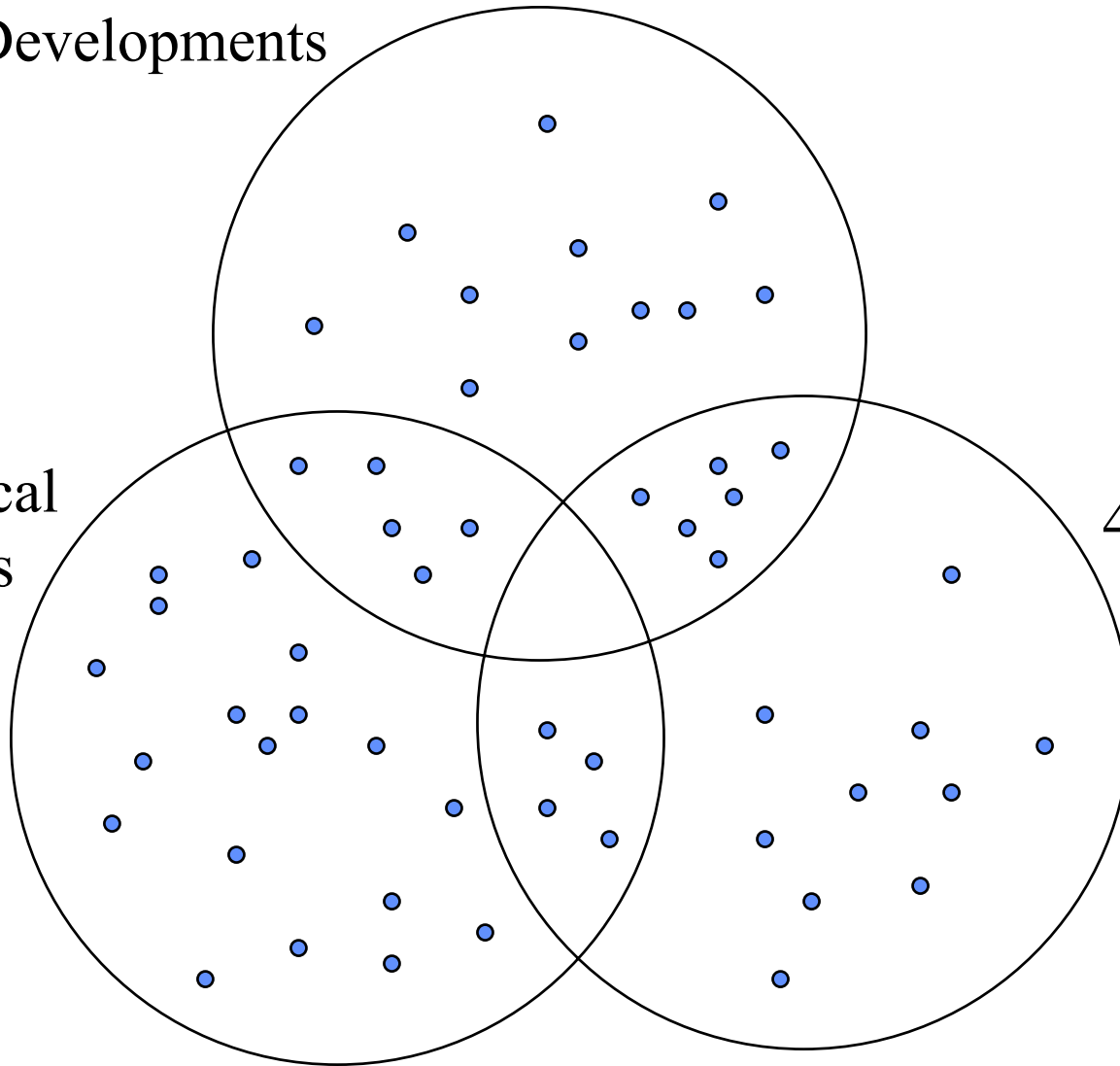
Recent Developments in Numerical Methods for 4d-Var

Mike Fisher

Recent Developments

Numerical
Methods

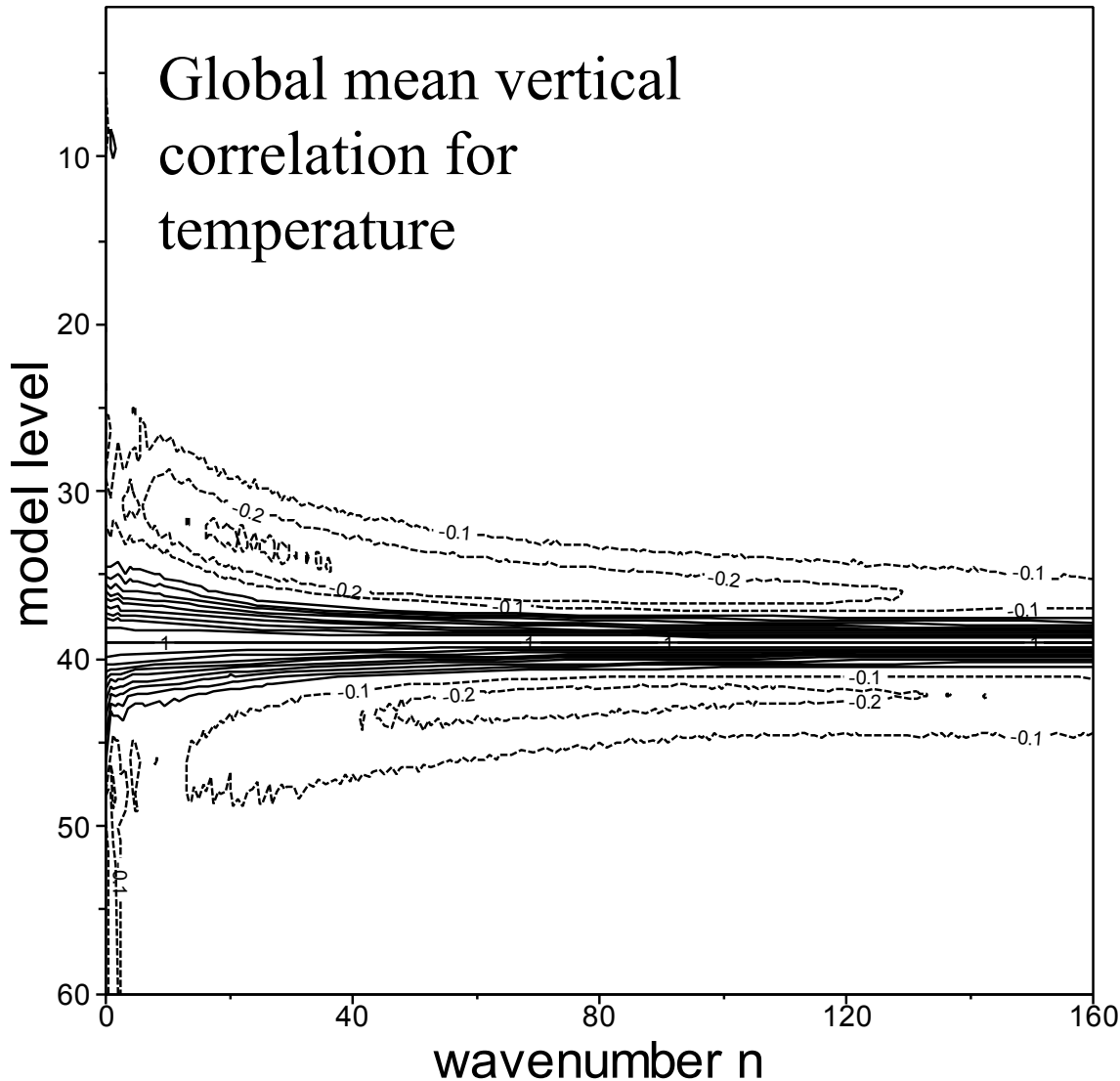
4d-Var



Outline

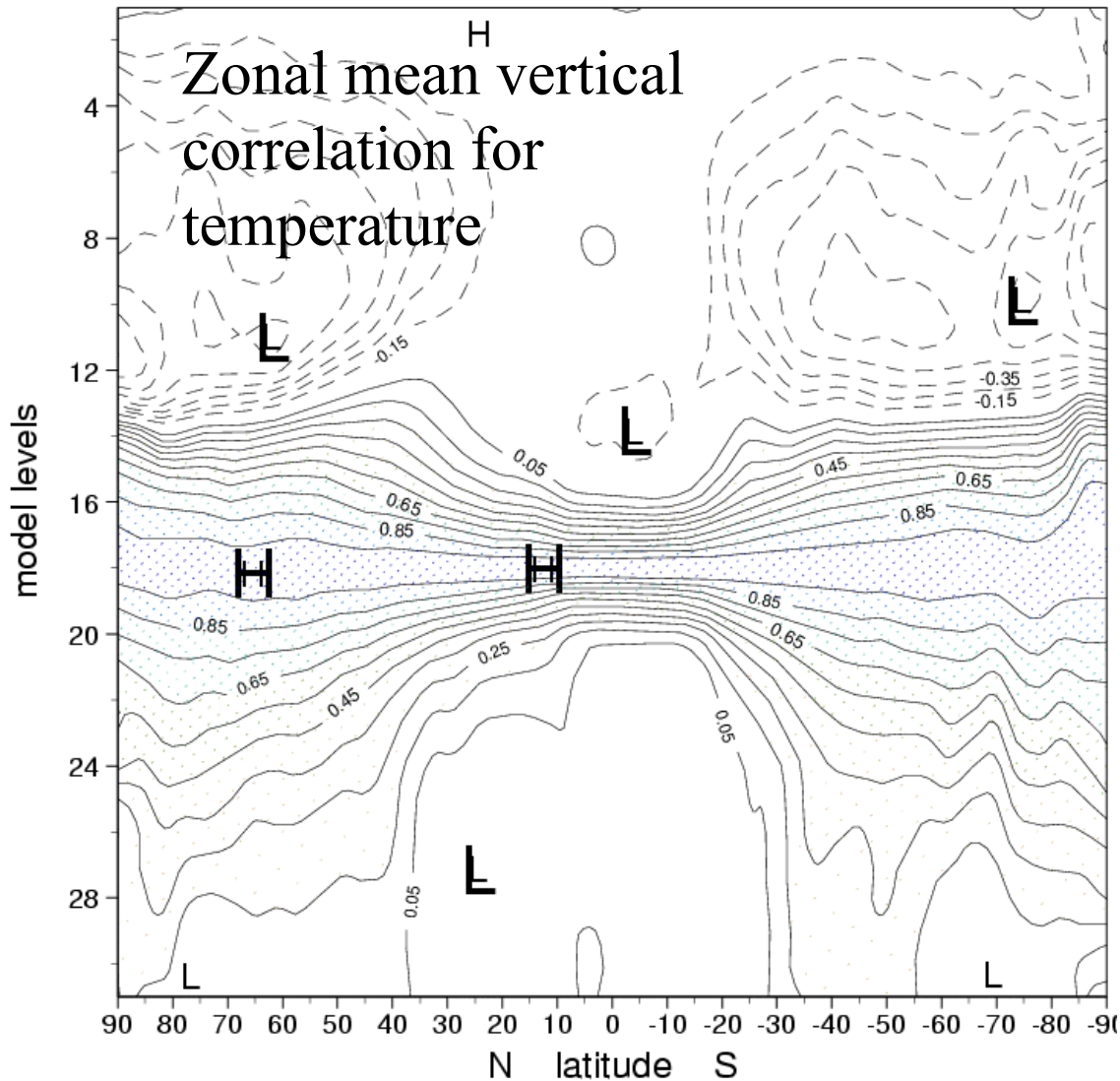
- **Non-orthogonal wavelets on the sphere:**
 - **Motivation: Covariance Modelling**
 - **Definition: Frames**
 - **Application: Wavelet J_b**

Wavelets on the Sphere – Motivation



Vertical Correlations vary with horizontal spatial scale

Wavelets on the Sphere – Motivation



Vertical Correlations vary with horizontal location.

~200hPa

~500hPa

~850hPa

Wavelets on the Sphere – Motivation

- The variation of vertical correlation with location and with horizontal spatial scale are both important features, and both should be included in the covariance model.
- However, we are severely limited by the enormous size of the covariance matrix ($\sim 10^7 \times 10^7$).
- Essentially, the covariance matrix must be block-diagonal, with block size NLEVS \times NLEVS, and with many identical blocks.
- Currently, we specify one block per wavenumber, n .
 - variation with scale is modelled, variation with location is not.
- Alternatively, we could specify one block per gridpoint.
 - variation with location is modelled, variation with scale is not.
- Wavelets provide a way to do both (and still keep things sparse).

Wavelets on the Sphere – Frames

- It is possible to define orthogonal wavelets on the sphere by first gridding the sphere.
- Göttemann (1997, citeseer.ist.psu.edu/227230.html) defined spherical wavelets using splines on a quasi-uniform latitude-longitude grid.
- Schröder and Sweldens (1995, ACM SIGGRAPH, 161-172) defined them for a triangulation of the sphere.
- However, these approaches necessarily have special points (poles or vertices). They do not retain rotational symmetry for finite truncations of the wavelet expansion.
- If we wish to retain rotational symmetry, we must give up on orthogonality.
- I.e. we must consider frames.

Wavelets on the Sphere – Frames

● Definitions

- A family of functions, $\{\psi_j; j \in J\}$ in a Hilbert space is called a frame if there exist $A > 0$ and $B < \infty$ such that for all f in the space:

$$A \|f\|^2 \leq \sum_{j \in J} |\langle f, \psi_j \rangle|^2 \leq B \|f\|^2$$

- The condition is sufficient to ensure the existence of a dual frame, $\{\tilde{\psi}_j; j \in J\}$ with the property:

$$\frac{1}{A} \sum_{j \in J} \langle f, \psi_j \rangle \tilde{\psi}_j = f = \frac{1}{A} \sum_{j \in J} \langle f, \tilde{\psi}_j \rangle \psi_j$$

- A particularly interesting case occurs when $A=B$. This is called a tight frame. Tight frames are self-dual:

$$f = \frac{1}{A} \sum_{j \in J} \langle f, \psi_j \rangle \psi_j$$

Wavelets on the Sphere – Frames

- Tight frames share many of the properties of orthogonal bases. (An orthogonal basis is a tight frame with $\|\psi_j\|^2 = 1$ and $A=1$).
- Tight frames define a “transform”, since we may write:

$$f = \frac{1}{A} \sum_{j \in J} \langle f, \psi_j \rangle \psi_j$$

as:
$$c_j = \langle f, \psi_j \rangle \quad , \quad f = \frac{1}{A} \sum_{j \in J} c_j \psi_j$$

- c.f. Fourier series:

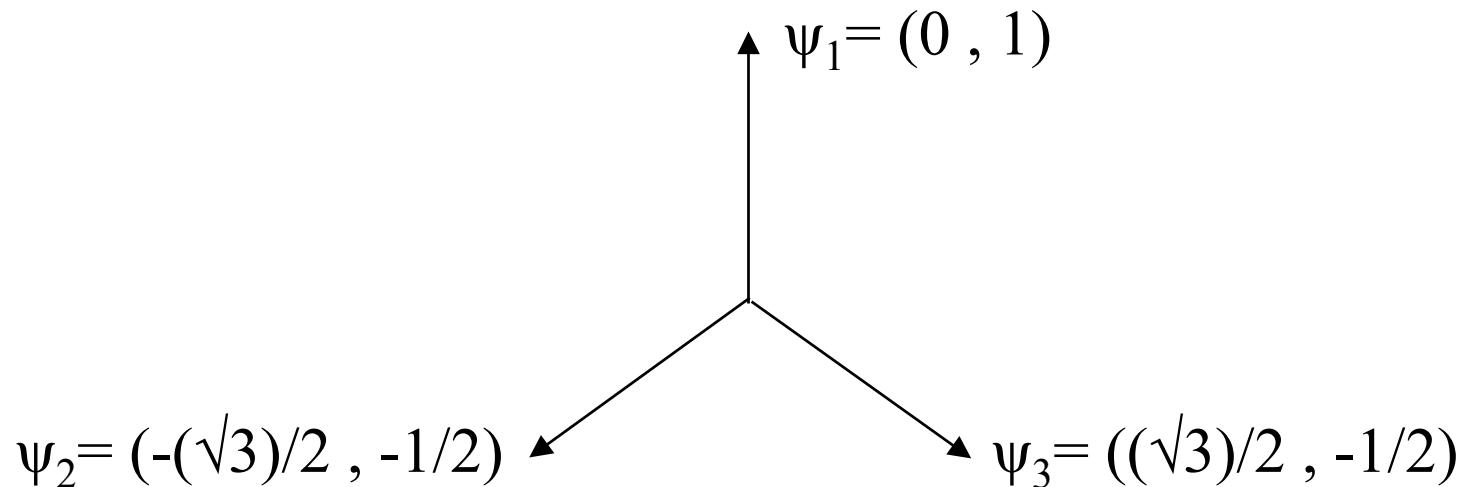
$$\hat{f}_m = \langle f, e^{2\pi i m t / (b-a)} \rangle = \frac{1}{b-a} \int_a^b f(t) e^{-2\pi i m t / (b-a)} dt$$

$$f(t) = \sum_{m=-\infty}^{\infty} \hat{f}_m e^{2\pi i m t / (b-a)}$$

Wavelets on the Sphere – Frames

- **Example: The Mercedes-Benz Frame**

- Daubechies, 1992: “Ten Lectures on Wavelets”



- **Tight frame:**
$$\sum_{j=1}^3 |\langle f, \psi_j \rangle|^2 = |f_x|^2 + \left| -\frac{\sqrt{3}}{2} f_x - \frac{1}{2} f_y \right|^2 + \left| \frac{\sqrt{3}}{2} f_x - \frac{1}{2} f_y \right|^2 = \frac{3}{2} |f|^2$$

- **Hence, for any f :**
$$\frac{2}{3} \sum_{j=1}^3 \langle f, \psi_j \rangle \psi_j = f$$

Wavelets on the Sphere – Frames

- **Example:** The discrete spherical transform.
- **Consider functions** $\hat{f}(m, n)$ **where** $0 \leq n \leq N$, $-n \leq m \leq n$.
- **Let** $\hat{\psi}_j(m, n)$ **be the** $(m, n)^{\text{th}}$ **spherical harmonic, evaluated at the** j^{th} **gridpoint of the Gaussian grid, and multiplied by sqrt(Gaussian integration weight):**

$$\hat{\psi}_j(m, n) = \sqrt{w(\phi_j)} Y_{m,n}(\lambda_j, \phi_j)$$

- **Then:** $\langle \hat{f}, \hat{\psi}_j \rangle = \sum_{m,n} \hat{f}(m, n) \sqrt{w(\phi_j)} Y_{m,n}(\lambda_j, \phi_j) = \sqrt{w_j(\phi_j)} f(\lambda_j, \phi_j)$

$$\hat{f}(m, n) = \sum_j \langle f, \hat{\psi}_j \rangle \hat{\psi}_j(m, n) = \sum_j w_j(\phi_j) f(\lambda_j, \phi_j) Y_{m,n}(\lambda_j, \phi_j)$$

- **NB:** This is a tight frame, not an orthogonal transform. There are more gridpoints than spectral coefficients.

Wavelets on the Sphere – Generalized Frames

- The concept of a frame can be generalized to the case where the number of basis functions, ψ_j is uncountable.
- The sum in the frame condition becomes an integral:

$$A\|f\|^2 \leq \int_D |\langle f, \psi_x \rangle|^2 dx \leq B\|f\|^2$$

- Most of the properties carry over from the discrete case.
- In particular, for a tight frame, we have:

$$f = \frac{1}{A} \int_D \langle f, \psi_x \rangle \psi_x dx$$

- See Kaiser, 1994 “A Friendly Guide to Wavelets”

Wavelets on the Sphere – Generalized Frames

- We will consider a specific, semi-discrete case.
- The basis functions, $\Psi_{j,\lambda,\phi}$ are labelled by 3 indices:
 - j (discrete) indicates “scale”
 - λ (continuous) is longitude
 - ϕ (continuous) is latitude
- We define:

$$\psi_{j,\lambda,\phi}(\lambda',\phi') = \Psi_j(r(\lambda',\phi',\lambda,\phi))$$

- where $r(\lambda',\phi',\lambda,\phi)$ is the great-circle distance between (λ',ϕ') and (λ,ϕ) .

Wavelets on the Sphere – Generalized Frames

- The inner product is:

$$\langle f, \Psi_{j,\lambda,\phi} \rangle = \int_{\Omega} f(\lambda', \phi') \Psi_{j,\lambda,\phi}(\lambda', \phi') \cos(\phi') d\lambda' d\phi'$$

- Let us write:

$$f_j(\lambda, \phi) = \langle f, \Psi_{j,\lambda,\phi} \rangle$$

- Then, substituting for $\Psi_{j,\lambda,\phi}(\lambda', \phi')$ we see that the inner product corresponds to a convolution on the sphere:

$$f_j(\lambda, \phi) = \int_{\Omega} f(\lambda', \phi') \Psi_j(r(\lambda', \phi', \lambda, \phi)) \cos(\phi') d\lambda' d\phi'$$

$$\text{I.e. } f_j = f \otimes \Psi_j$$

Wavelets on the Sphere – Generalized Frames

- We seek a tight frame. The condition is:

$$\sum_j \int_{\Omega} \left| \langle f, \psi_{j,\lambda,\phi} \rangle \right|^2 \cos(\phi) d\lambda d\phi = A \|f\|^2$$

- That is:
$$\sum_j \int_{\Omega} |f_j(\lambda, \phi)|^2 \cos(\phi) d\lambda d\phi = A \|f\|^2$$

- I.e.
$$\sum_j \|f_j\|^2 = A \|f\|^2$$

- Evaluating the norms in terms of spherical harmonic coefficients, we have:

$$\sum_{j,m,n} \left| \hat{f}_j(m,n) \right|^2 = A \sum_{m,n} \left| \hat{f}(m,n) \right|^2$$

Wavelets on the Sphere – Generalized Frames

$$\sum_{j,m,n} \left| \hat{f}_j(m,n) \right|^2 = A \sum_{m,n} \left| \hat{f}(m,n) \right|^2$$

- But, remember that $f_j = f \otimes \Psi_j$, where Ψ_j is a function of great-circle distance.

- Hence: $\hat{f}_j(m,n) = \hat{f}(m,n) \hat{\Psi}_j(n)$

- where $\hat{\Psi}_j(n)$ are Legendre transform coefficients.

- The condition for a tight frame is thus:

$$\sum_{j,m,n} \left| \hat{f}(m,n) \hat{\Psi}_j(n) \right|^2 = A \sum_{m,n} \left| \hat{f}(m,n) \right|^2$$

Wavelets on the Sphere – Generalized Frames

$$\sum_{j,m,n} \left| \hat{f}(m,n) \hat{\Psi}_j(n) \right|^2 = A \sum_{m,n} \left| \hat{f}(m,n) \right|^2$$

● I.e.

$$\sum_{m,n} \left(\sum_j \hat{\Psi}_j^2(n) \right) \left| \hat{f}(m,n) \right|^2 = A \sum_{m,n} \left| \hat{f}(m,n) \right|^2$$

$$\Rightarrow \sum_j \hat{\Psi}_j^2(n) = A$$

● For convenience, we will scale the basis functions appropriately, so that $A=1$:

$$\sum_j \hat{\Psi}_j^2(n) = 1 \quad \forall n$$

Wavelets on the Sphere – Generalized Frames

- If we have a tight frame, we have the transform property:

$$\begin{aligned} f &= \sum_j \int_{\Omega} \langle f, \psi_{j,\lambda,\phi} \rangle \psi_{j,\lambda,\phi} \cos(\phi') d\lambda' d\phi' \\ &= \sum_j \int_{\Omega} f_j(\lambda', \phi') \Psi_j(r(\lambda', \phi', \lambda, \phi)) \cos(\phi') d\lambda' d\phi' \end{aligned}$$

- But, the right hand side is just another convolution.
- Hence, the “transform pair” is just:

$$\begin{aligned} f_j &= \Psi_j \otimes f \\ f &= \sum_j \Psi_j \otimes f_j \end{aligned}$$

Wavelets on the Sphere – Generalized Frames

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

- The first equation defines f_j .
- We can easily verify the second equation:

$$\begin{aligned} \left(\overline{\sum_j \Psi_j \otimes f_j} \right) (m, n) &= \sum_j \hat{\Psi}_j(n) \hat{f}_j(m, n) \\ &= \sum_j \hat{\Psi}_j^2(n) f(m, n) \\ &= \hat{f}(m, n) \end{aligned}$$

Wavelets on the Sphere – Summary

- A set of functions of great-circle distance, $\{\Psi_j(r); j=0,1,2,\dots\}$ whose Legendre transform coefficients satisfy:

$$\sum_j \hat{\Psi}_j^2(n) = 1 \quad \forall n$$

define a tight generalized frame.

- The functions define a “transform pair”:

$$f_j = \Psi_j \otimes f$$

$$f = \sum_j \Psi_j \otimes f_j$$

Wavelets on the Sphere – Example

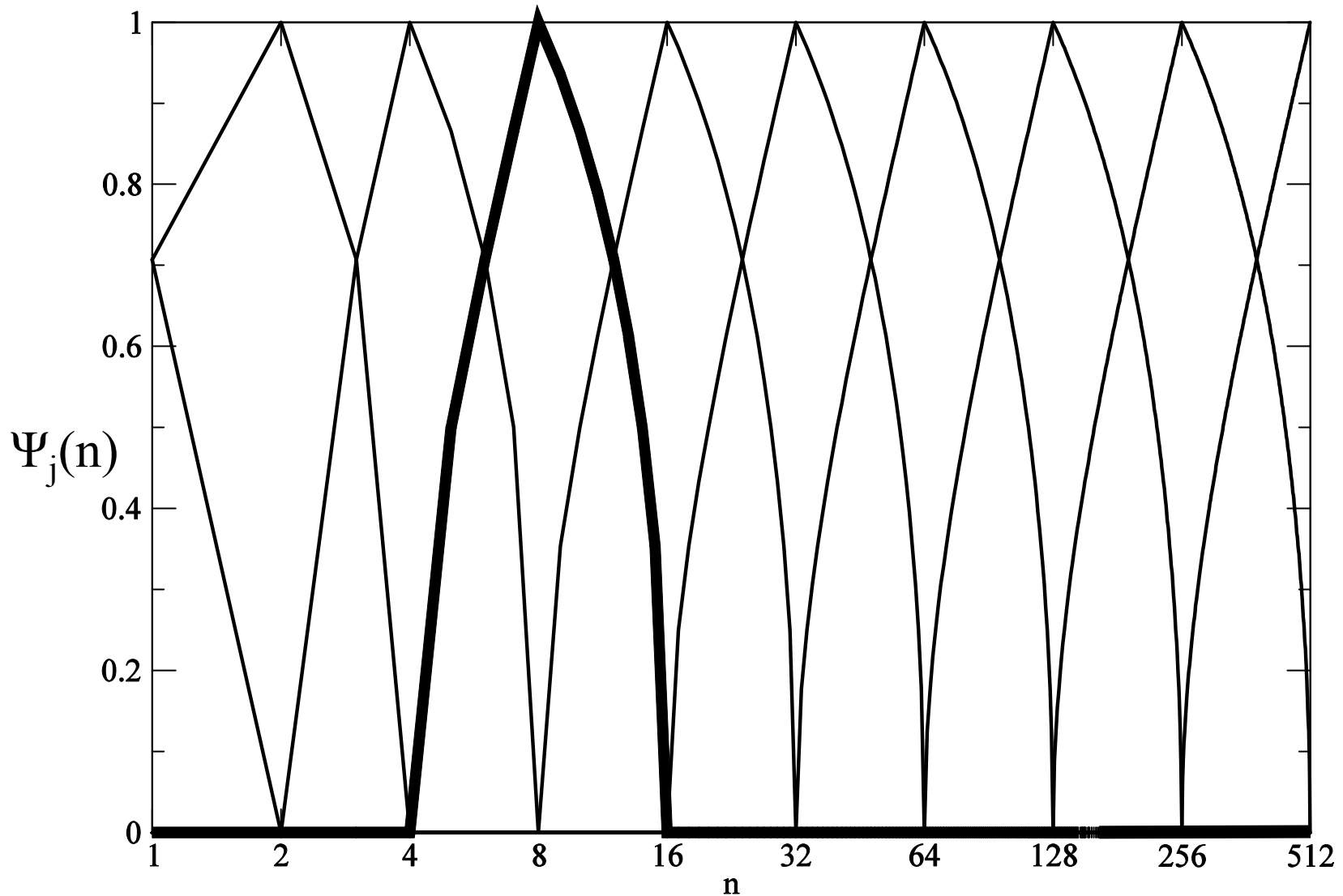
- The condition: $\sum_j \hat{\Psi}_j^2(n) = 1 \quad \forall n$

suggests we define the functions $\hat{\Psi}_j^2(n)$ to be B-splines.

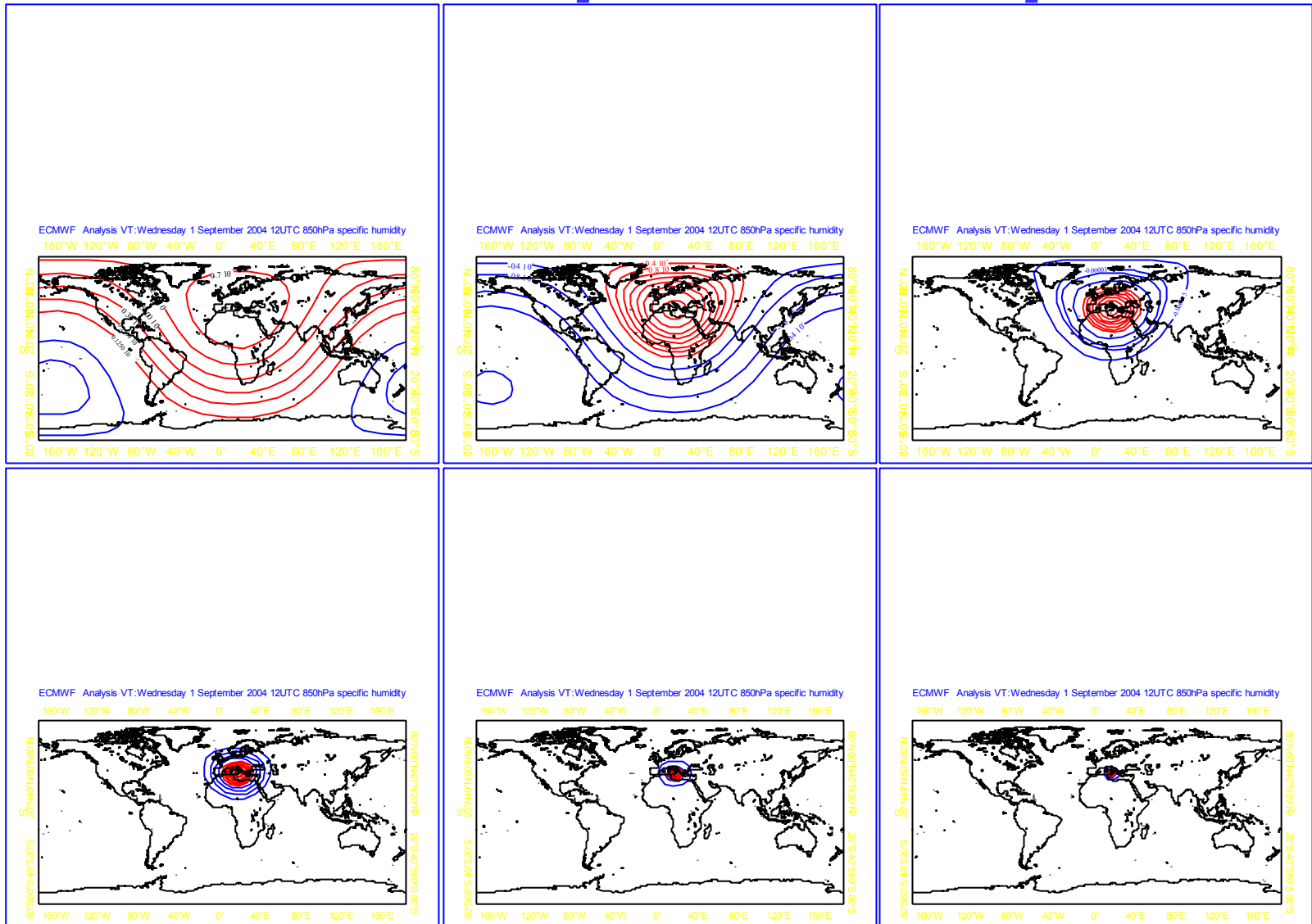
- For example, linear B-splines are triangle functions:

$$\hat{\Psi}_j^2(n) = \begin{cases} \frac{n - N_{j-1}}{N_j - N_{j-1}} & N_{j-1} < n \leq N_j \\ \frac{n - N_{j+1}}{N_j - N_{j+1}} & N_j < n < N_{j+1} \\ 0 & \text{otherwise} \end{cases}$$

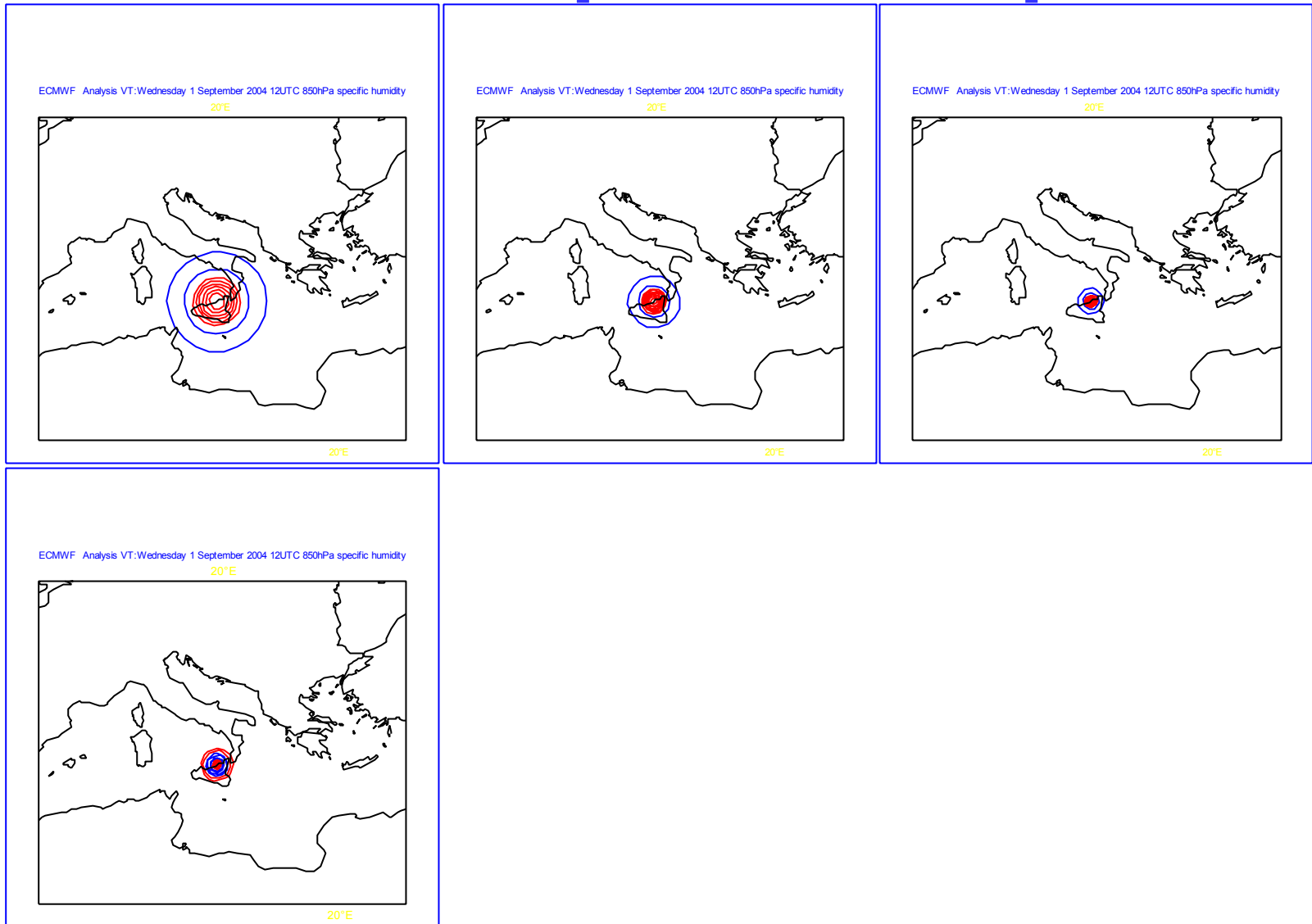
Wavelets on the Sphere – Example



Wavelets on the Sphere – Example

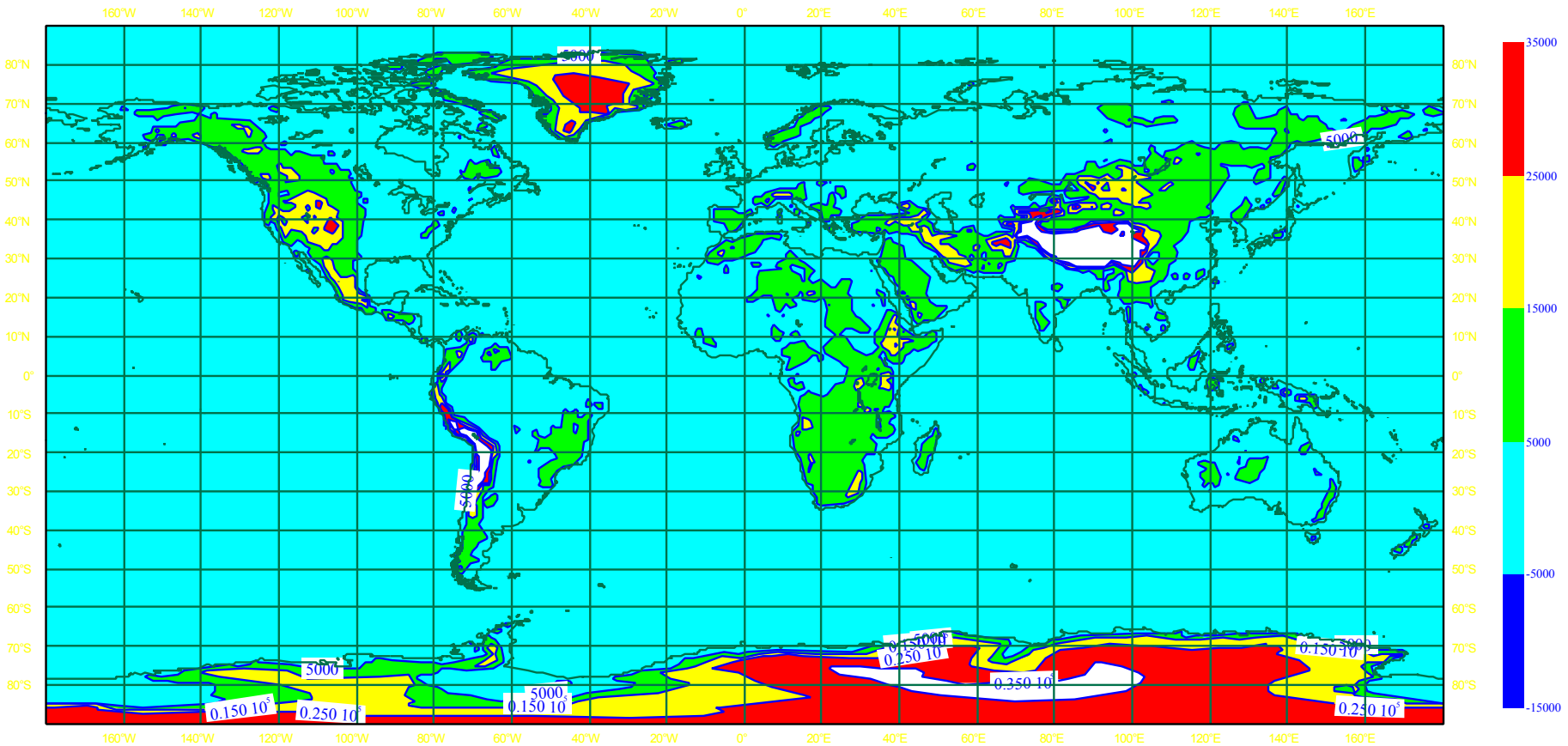


Wavelets on the Sphere – Example



Spherical Wavelets – Example

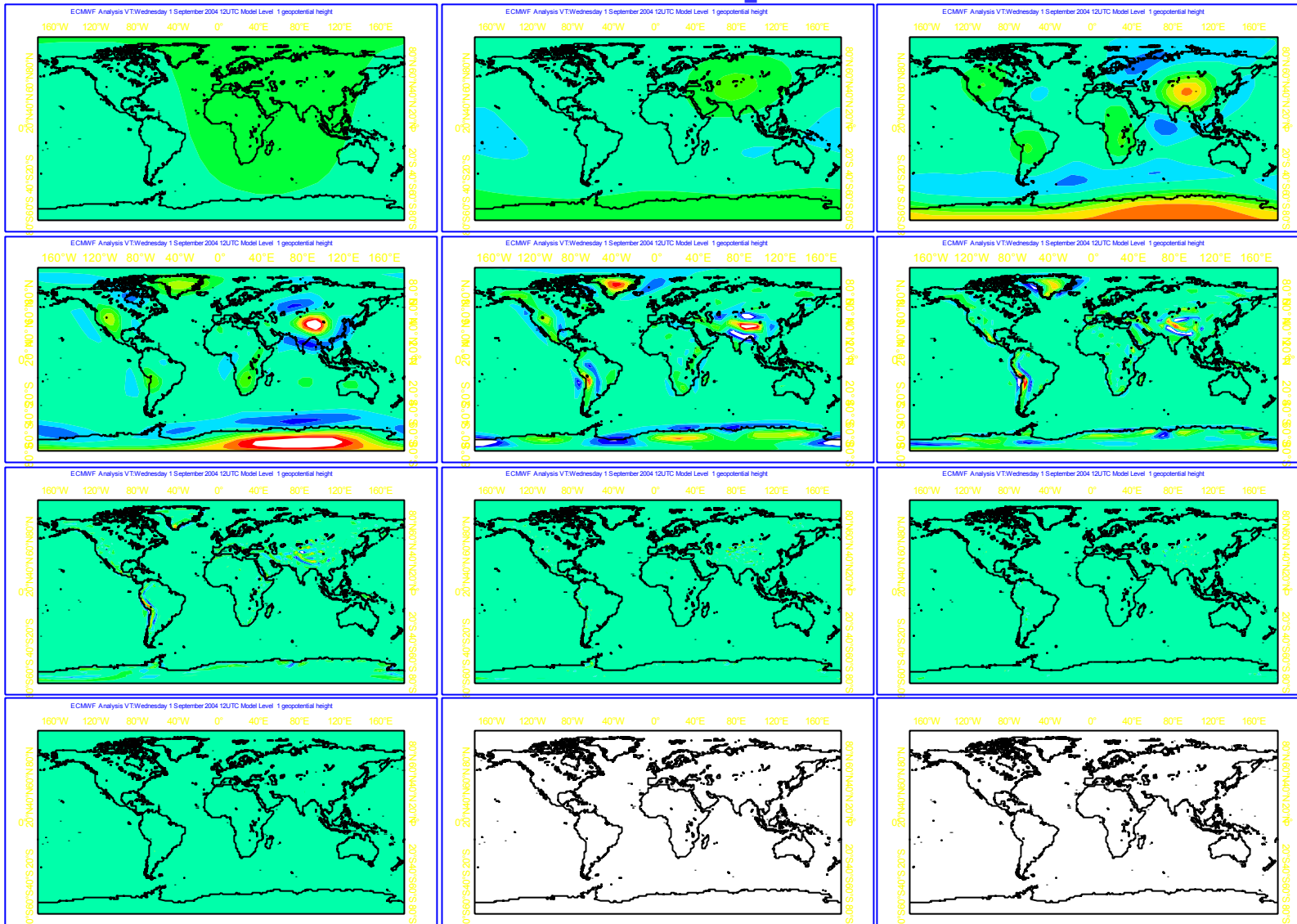
● T511 Model Orography:



Spherical Wavelets – Example

Convolve
with Ψ_j :

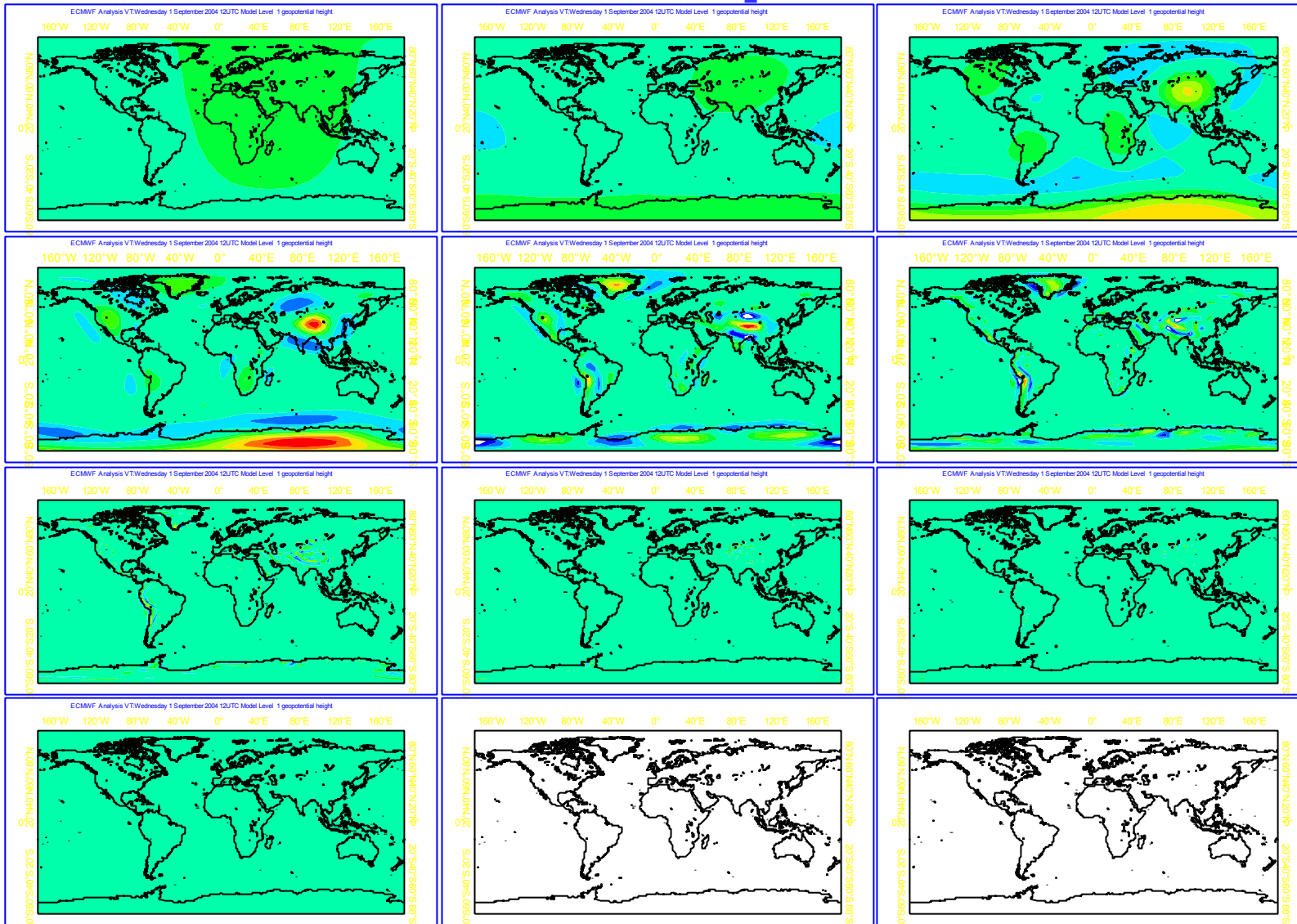
$$f_j = \Psi_j \otimes f$$



Spherical Wavelets – Example

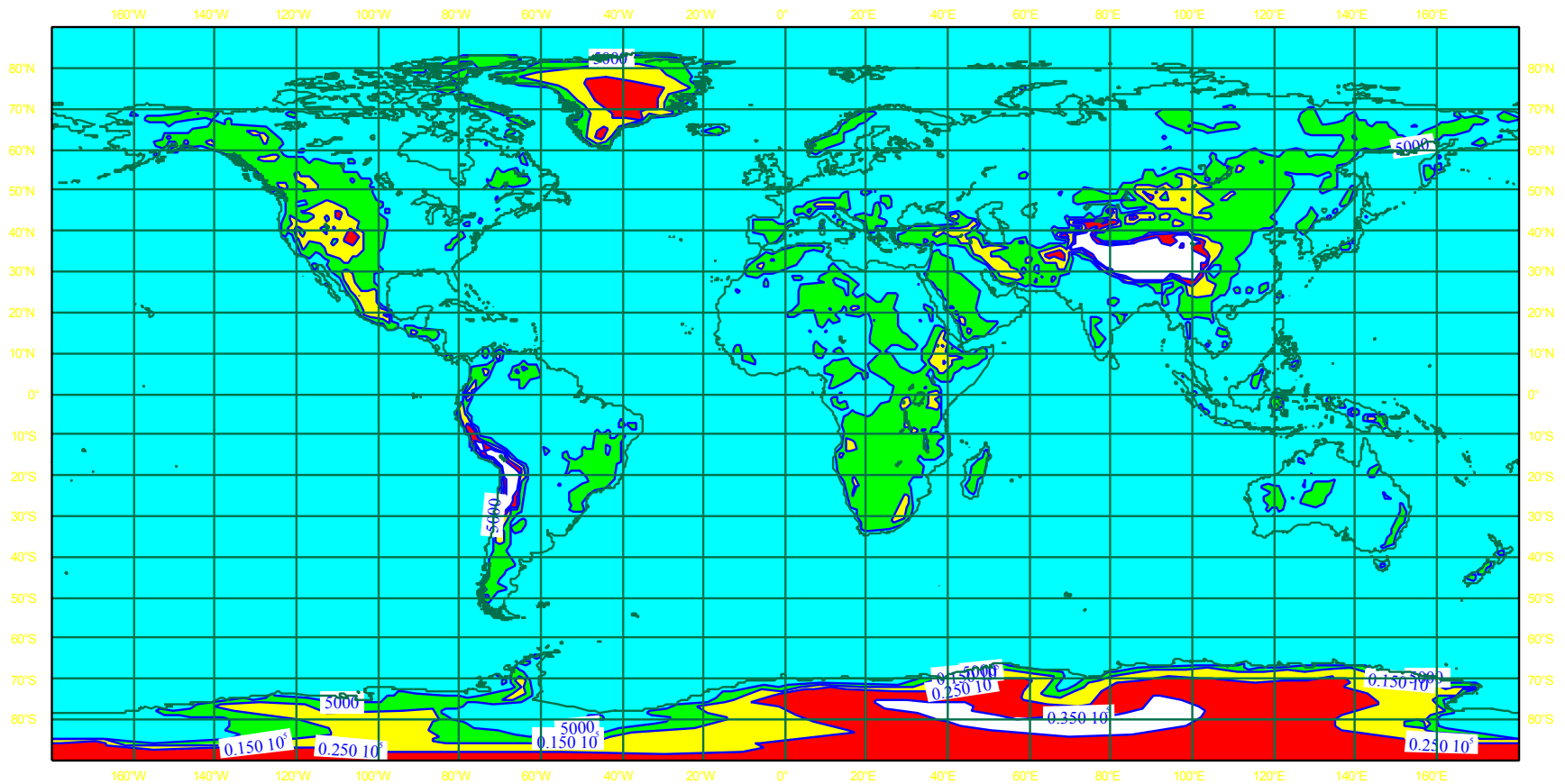
Convolve
again...

$$\Psi_j \otimes f_j$$



Spherical Wavelets – Example

- ... and add, to retrieve the original field: $f = \sum_j \Psi_j \otimes f_j$



Spherical Wavelets – Other Approaches

- **The wavelets we have derived are similar to those of Freeden and Windheuser (1996, Adv. Comp. Math. 51-94.**
 - They are a special case of the very broad range of spherical wavelet decompositions described by Freeden et al. (1998, “Constructive Approximation on the Sphere, OUP).
- **A different approach, using group theory, is taken by Antoine and Vandergheynst (1999, Appl. and Comput. Harm. Anal. 262-291).**
 - They define spherical wavelets as coherent states of the product group of rotations on the sphere, and dilations on the polar-stereographic tangent plane.
- **Mhaskar et al. (2000, Adv. Comput. Math.) describe polynomial wavelet frames on the sphere.**
 - and cite 11 papers, each defining a different approach to the construction of wavelets on the sphere.

Wavelet \mathbf{J}_b

- So, how does all this help us formulate a covariance model?
- First, let's review the main idea behind the current \mathbf{J}_b .
- 3d/4d-Var determine the analysis by minimizing a cost function:

$$J = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{d} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b))^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b))$$

- Usually, we do not minimize directly in terms of \mathbf{x} , but formulate the problem as:

$$J = \boldsymbol{\chi}^T \boldsymbol{\chi} + (\mathbf{d} - \mathbf{H}\mathbf{L}\boldsymbol{\chi})^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H}\mathbf{L}\boldsymbol{\chi})$$

$$\mathbf{x} = \mathbf{x}_b + \mathbf{L}\boldsymbol{\chi}$$

- NB: \mathbf{L} defines the background covariance matrix: $\mathbf{B} = \mathbf{L}\mathbf{L}^T$.

Wavelet J_b

- L defines the background covariance matrix: $B=LL^T$.
- To keep things simple, consider a 2d, univariate model.
- The current ECMWF covariance model boils down to:

$$L = \Sigma D$$

- Σ is diagonal in grid space, and corresponds to multiplying each gridpoint by a standard deviation.
 - It accounts for the spatial variation of background error.
- D is diagonal in spectral space, and corresponds to multiplying each wavenumber, n, by a standard deviation, which is a function of n, only.
 - It accounts for the variation of background error with scale.

Wavelet J_b

$$\mathbf{L} = \mathbf{\Sigma D}$$

- The current covariance model separates the spatial and spectral variation of background error.
- In particular, D corresponds to a convolution.
- D defines the horizontal correlation of background error for the covariance model.
- Because D is a convolution, the horizontal correlations are the same everywhere.
- This is a major shortcoming of the covariance model.

Wavelet \mathbf{J}_b

- To define a covariance model using wavelets, note first that there is no requirement for \mathbf{L} to be square.
- A rectangular matrix \mathbf{L} still defines a valid covariance model, $\mathbf{B}=\mathbf{L}\mathbf{L}^T$.
- We define the control variable as:

$$\boldsymbol{\chi} = \begin{pmatrix} \chi_1 \\ \chi_2 \\ \vdots \end{pmatrix}$$

where the χ_j correspond to different scales.

- The change of variable matrix is defined by:

$$\mathbf{x} - \mathbf{x}_b = \mathbf{L}\boldsymbol{\chi} = \sum_j \hat{\Psi}_j \otimes \boldsymbol{\Sigma}_j \chi_j$$

Wavelet \mathbf{J}_b

$$\mathbf{x} - \mathbf{x}_b = \mathbf{L}\boldsymbol{\chi} = \sum_j \hat{\Psi}_j \otimes \boldsymbol{\Sigma}_j \boldsymbol{\chi}_j$$

- The matrices $\boldsymbol{\Sigma}_j$ are diagonal in grid space, and account for the spatial variation of background error for each scale.
- Let us write the convolution with $\hat{\Psi}_j$ explicitly as a matrix operator $\mathbf{S}^{-1}\hat{\Psi}_j\mathbf{S}$, where \mathbf{S} is the spherical transform, and $\hat{\Psi}_j$ is diagonal.
- Using the symmetry of $\hat{\Psi}_j$, and $\boldsymbol{\Sigma}_j$, and the orthogonality of \mathbf{S} , we can write the covariance matrix implied by \mathbf{L} as:

$$\mathbf{B} = \mathbf{L}\mathbf{L}^T = \sum_j \mathbf{S}^{-1}\hat{\Psi}_j\mathbf{S}\boldsymbol{\Sigma}_j^2\mathbf{S}^{-1}\hat{\Psi}_j\mathbf{S}$$

- We will illustrate the covariance structures generated by wavelet \mathbf{J}_b by applying this matrix to delta functions.

Wavelet \mathbf{J}_b

$$\mathbf{B} = \mathbf{L}\mathbf{L}^T = \sum_j \mathbf{S}^{-1} \Psi_j \mathbf{S} \Sigma_j^2 \mathbf{S}^{-1} \Psi_j \mathbf{S}$$

- Consider the case where there is no spatial variation in the standard deviations: $\Sigma_j = \sigma_j \mathbf{I}$.
- Then, $\Psi_j \mathbf{S} \Sigma_j^2 \mathbf{S}^{-1} \Psi_j$ is diagonal, with elements $\sigma_j^2 \hat{\Psi}_j^2(n)$.
- But, $\sum_j \hat{\Psi}_j^2(n) = 1$, so $\sum_j \sigma_j^2 \hat{\Psi}_j^2(n)$ is a weighted average of σ_j 's
- If we choose $\hat{\Psi}_j^2(n)$ to be B-splines, then the variation of variance with n is an interpolation between the prescribed σ_j 's.

Wavelet J_b

- Suppose we want approximately Gaussian structure functions, with length scale that is a smoothly-varying function of latitude and longitude.
- Weaver+Courtier (2000) give the following expression for the modal variances corresponding to convolution with a quasi-Gaussian function with lengthscale L :

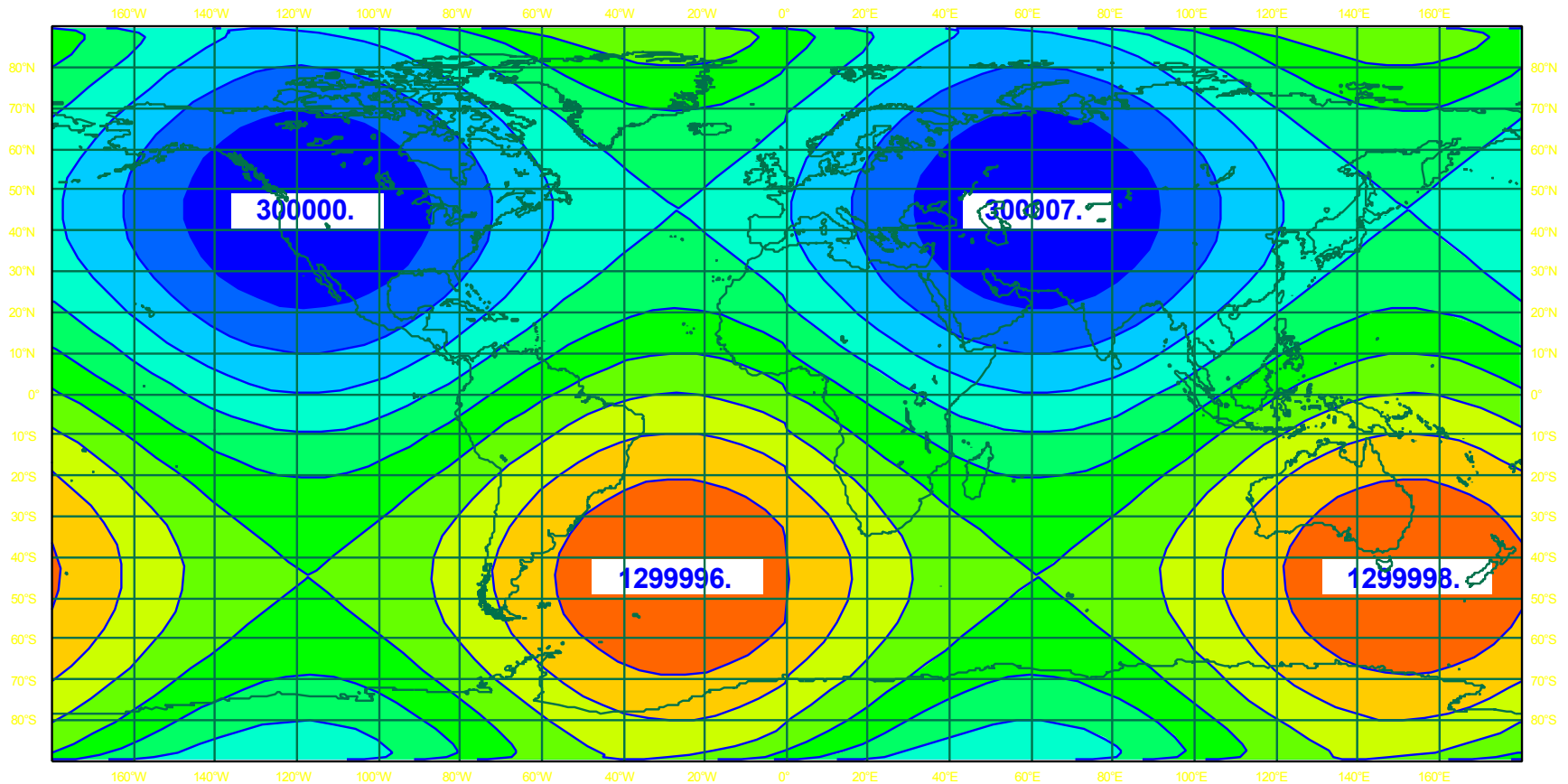
$$b^2(n; L) = \frac{\exp\left[-\left(L^2 n(n+1)\right) / 2a^2\right]}{\sum_n (2n+1) \exp\left[-\left(L^2 n(n+1)\right) / 2a^2\right]}$$

- We simply set: $\sigma_j = (N_{\max} + 1)b(N_j; L)$

but allow L (and hence σ_j) to vary with latitude and longitude.

Wavelet J_b

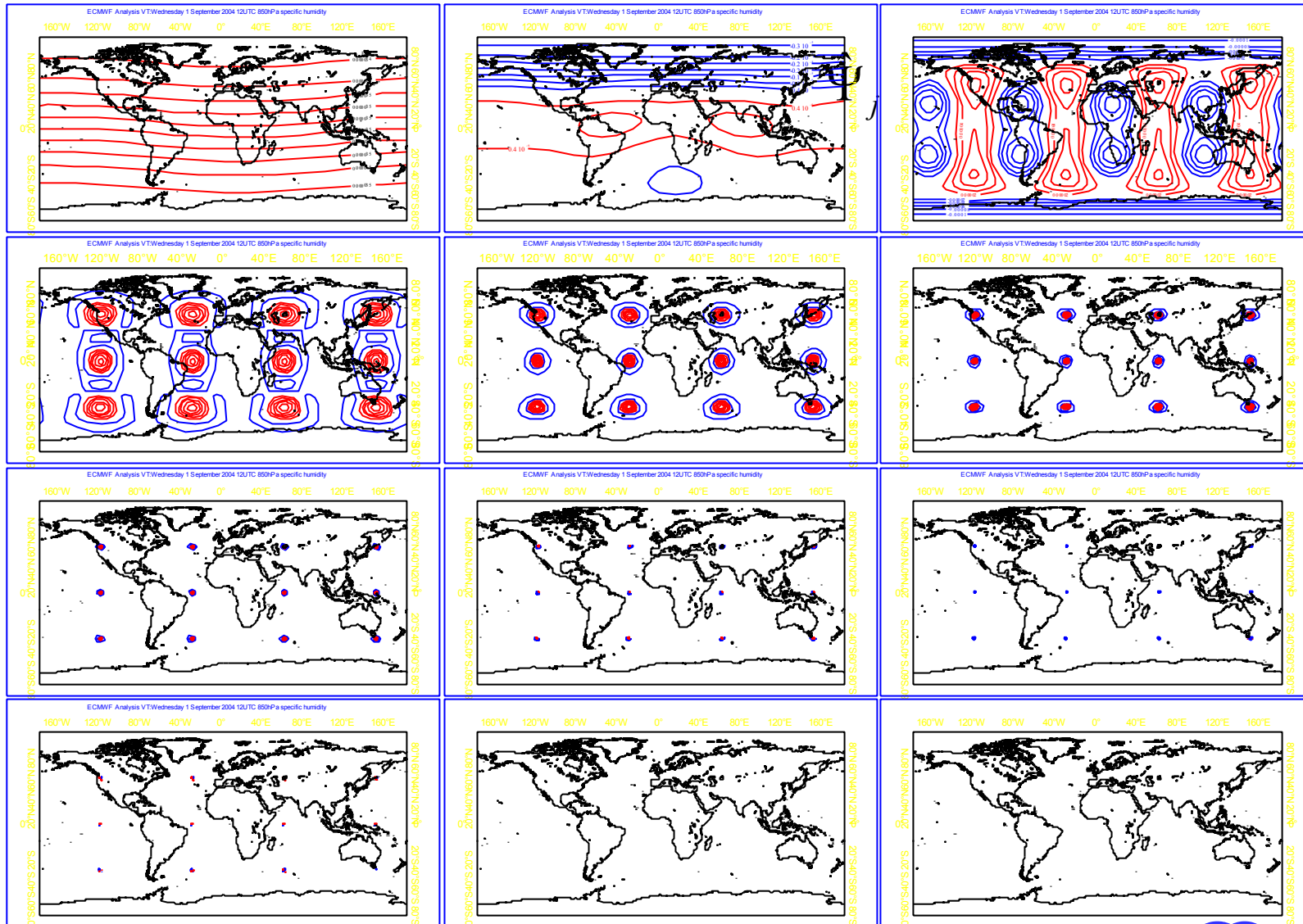
- Desired Length scale (300km – 1300km):



Wavelet J_b

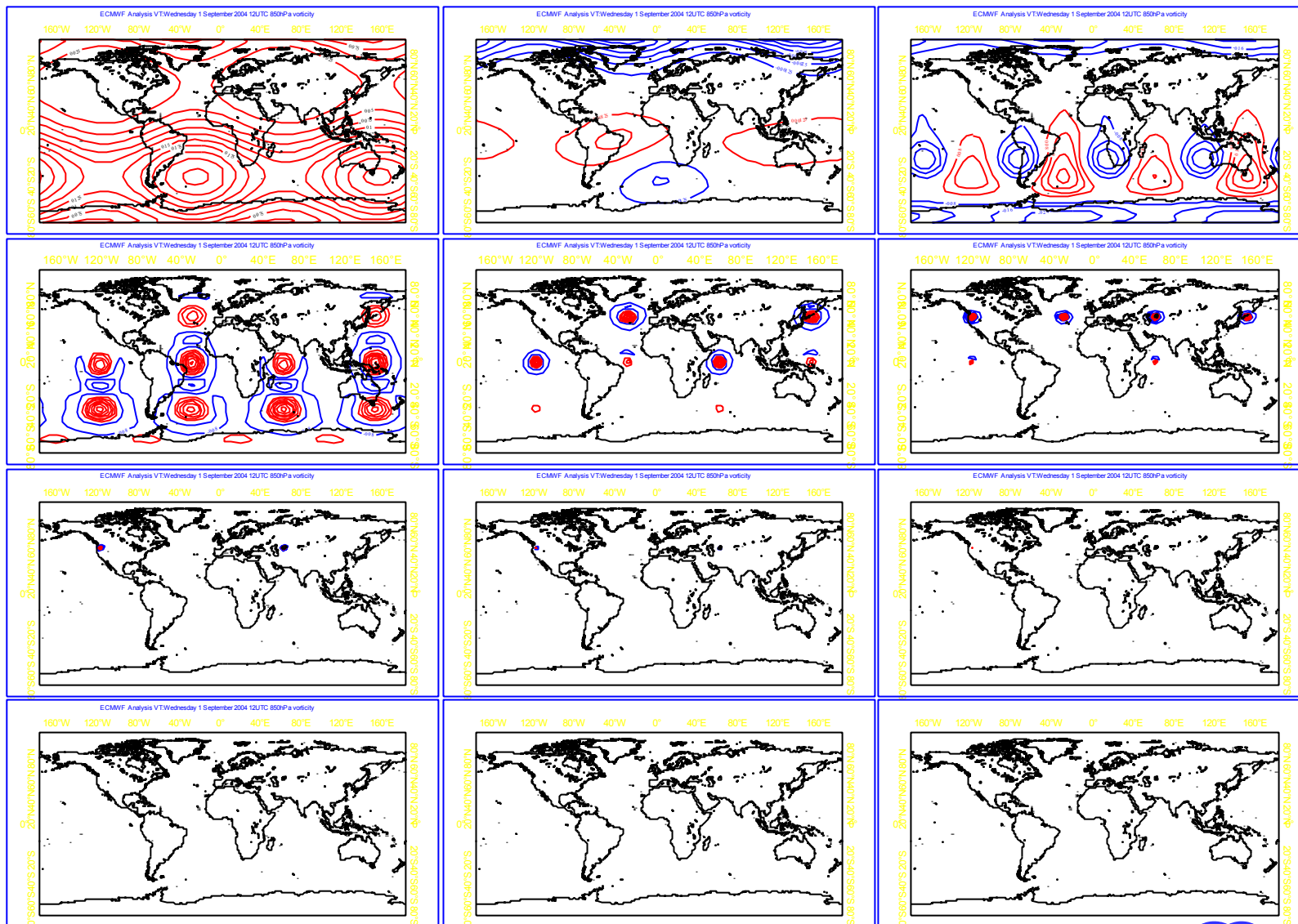
$$\Psi_j \otimes \mathbf{x}$$

Where \mathbf{x} is a set of delta functions



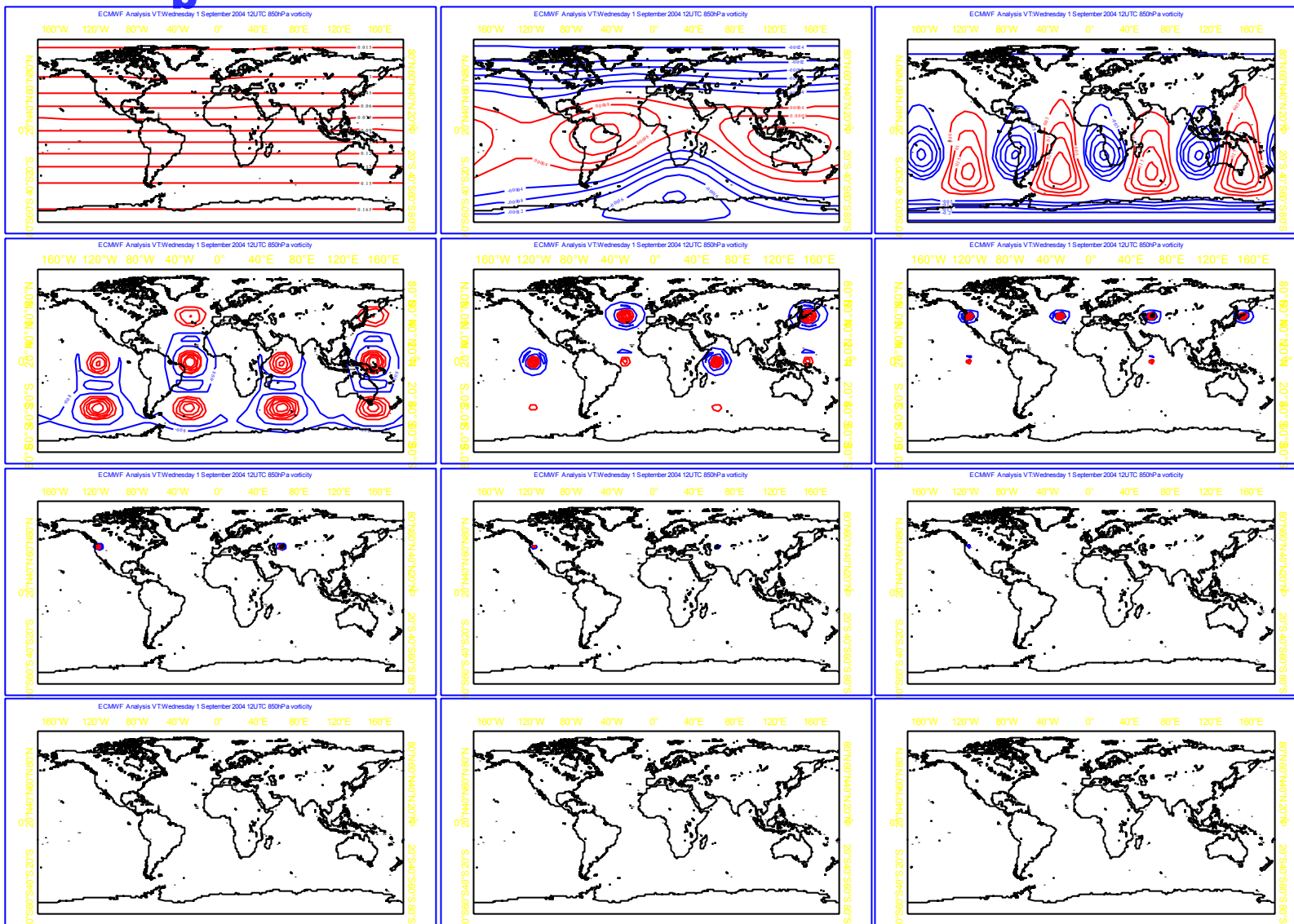
Wavelet J_b

Scale with σ_j^2
at each
gridpoint, and
for each scale.



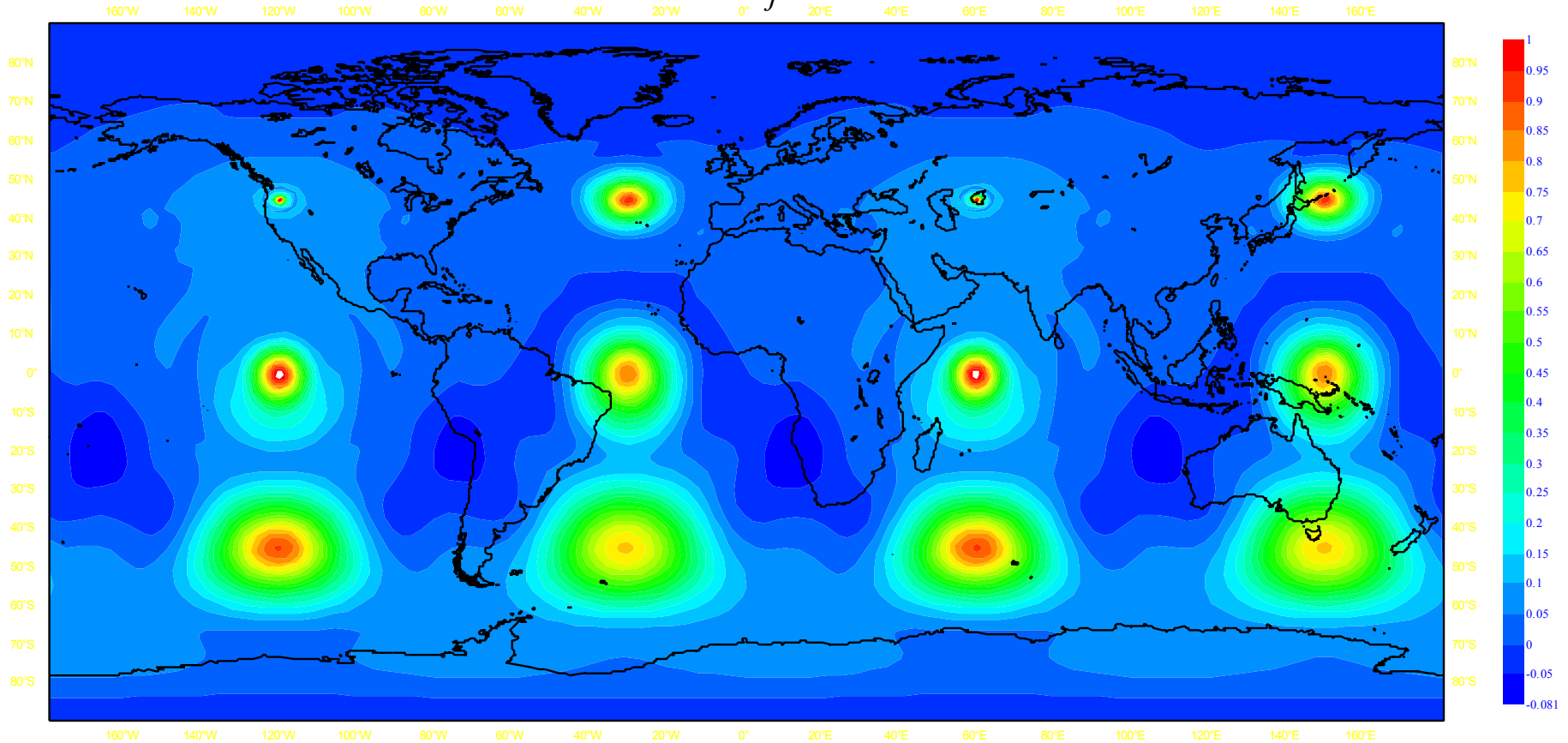
Wavelet J_b

Convolve
again with
 $\hat{\Psi}_j(n)$



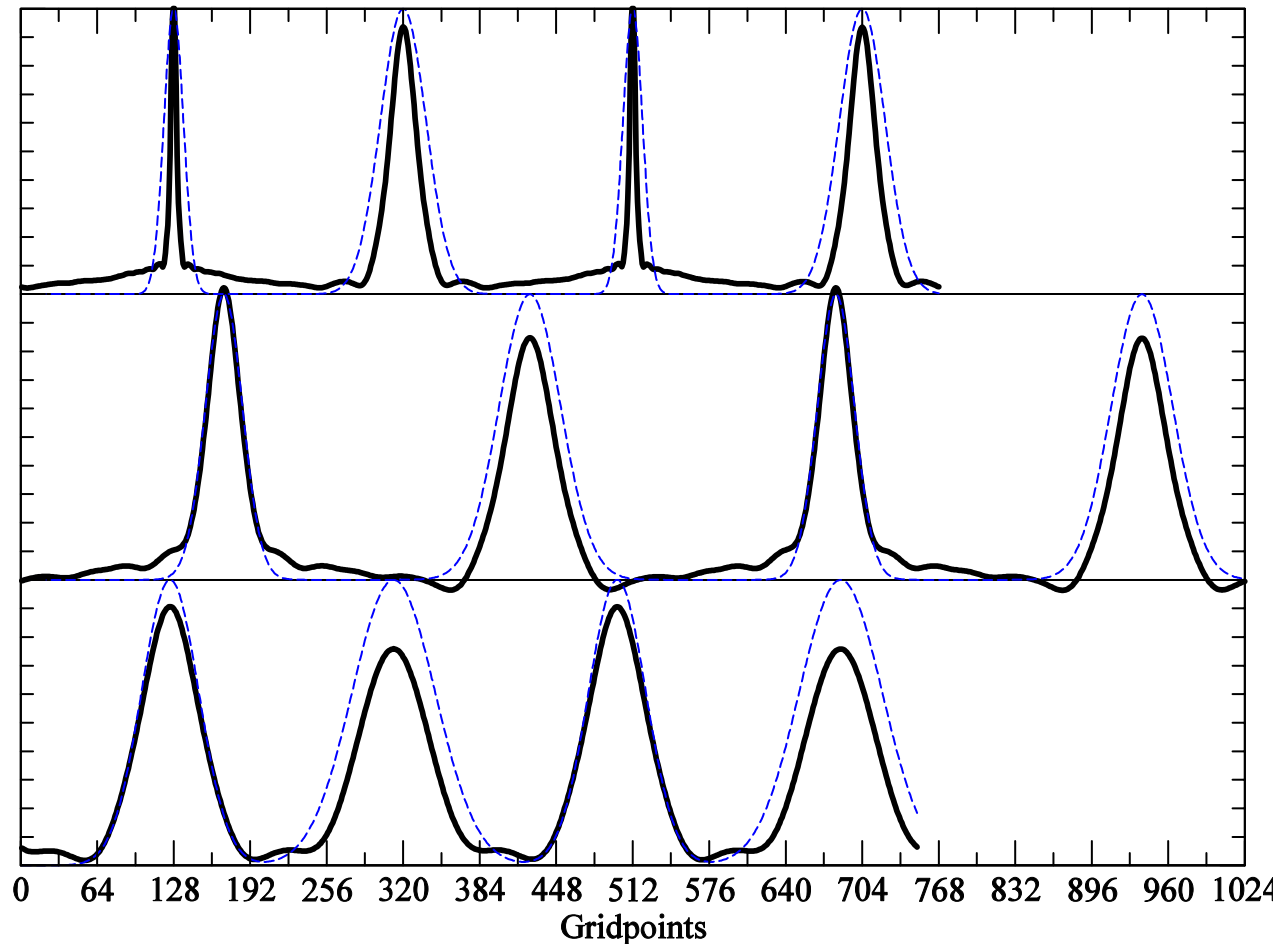
Wavelet J_b

● Add together to give: $B_X = \sum_j S^{-1} \Psi_j S \Sigma_j^2 S^{-1} \Psi_j S X$



Wavelet J_b

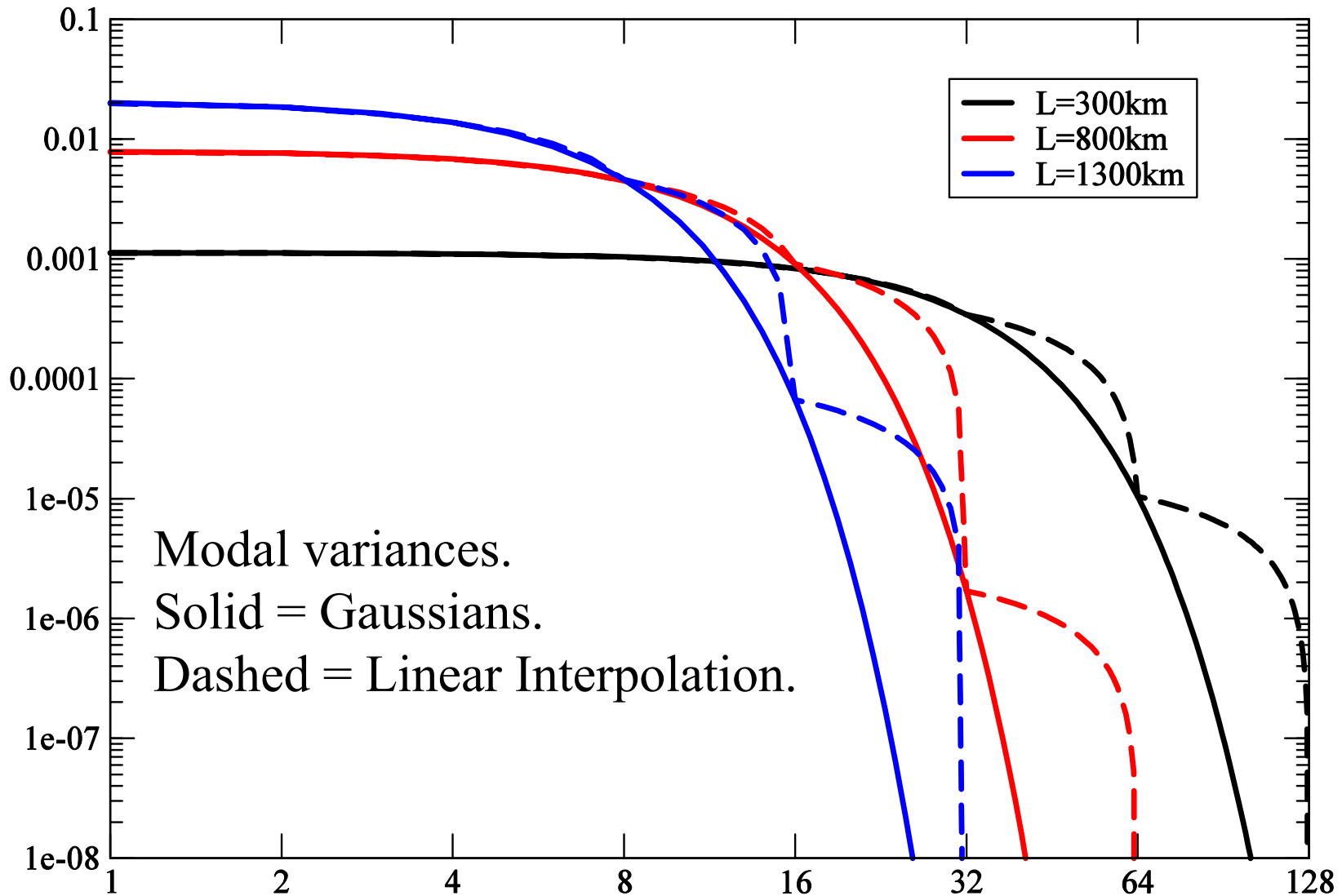
- E-W cross-sections. and the Gaussians we wanted:



Wavelet J_b

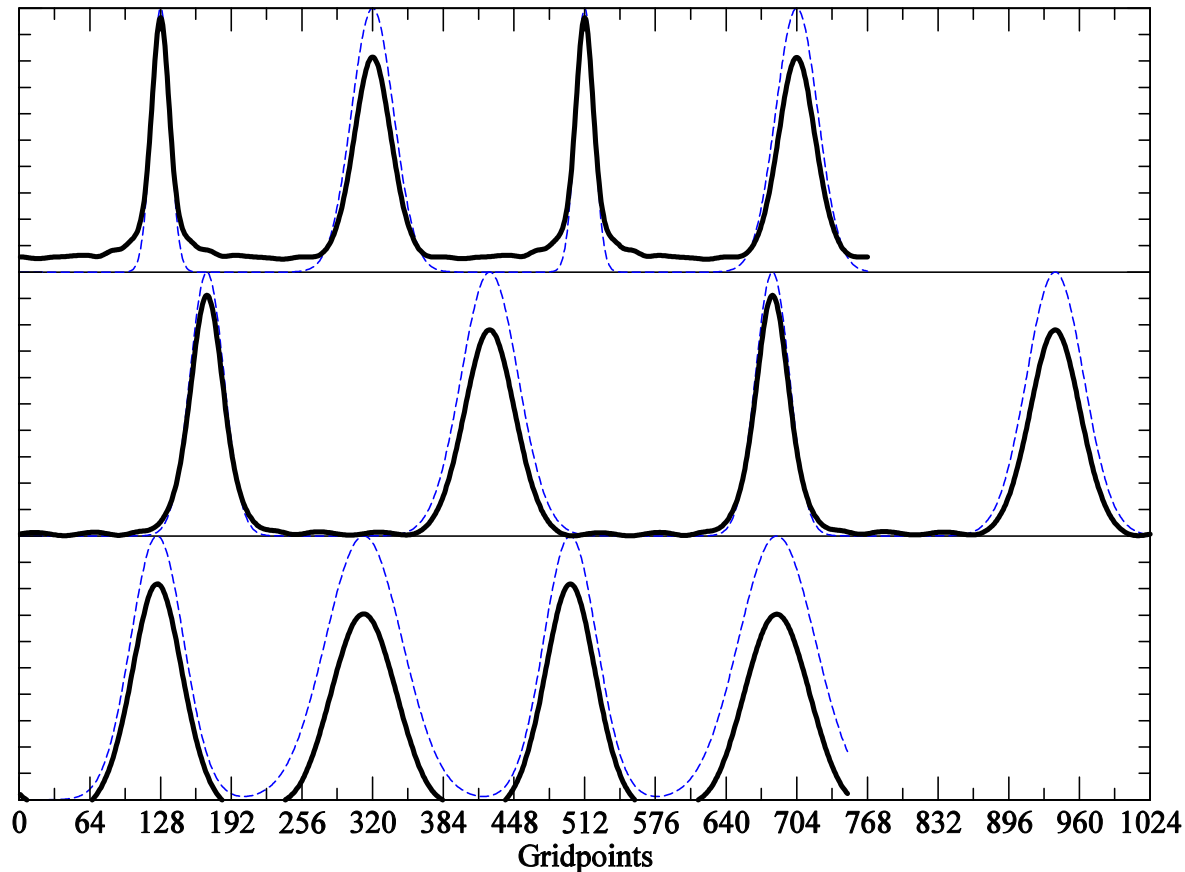
- **The agreement between the Gaussians we wanted, and the functions we got is not perfect!**
- **But note:**
 - **No attempt was made to tune the cut-off wavenumbers to improve the accuracy with which the Gaussian structure functions were modelled.**
 - **The implied modal variances are effectively linearly interpolated between those for wavenumbers 0,2,4,8,16,...**
 - **For large lengthscale (1300km), there is little variance beyond about $n=10$, so there are very few nodes in the interpolation.**
 - **More spectral resolution at large scales may improve the approximation.**
 - **Higher order B-splines would also help.**

Wavelet J_b



Wavelet J_b

- E-W cross-sections for a different choice of spectral bands (0,1,3,6,10,16,25,39,60,91,138,208,313,471,511):



Wavelet J_b

- We demonstrated Wavelet J_b by producing structure functions of a given analytic form.
- However, a significant advantage of Wavelet J_b over other approaches to covariance modelling on the sphere (digital filters, diffusion operators, etc.) is that we can calculate the coefficients of the covariance model directly from data.
- The covariance model is: $\mathbf{x} - \mathbf{x}_b = \sum_j \hat{\Psi}_j \otimes \Sigma_j \chi_j$
- The transform property gives us: $\Sigma_j \chi_j = \hat{\Psi}_j \otimes (\mathbf{x} - \mathbf{x}_b)$
- But χ_j has covariance matrix = I, so (in 2d), Σ_j is simply the matrix of gridpoint standard deviations of $\hat{\Psi}_j \otimes (\mathbf{x} - \mathbf{x}_b)$
- This is easily generated, given a sample of bg errors.

Wavelet J_b

- Extension of Wavelet J_b to 3 spatial dimensions is straightforward.

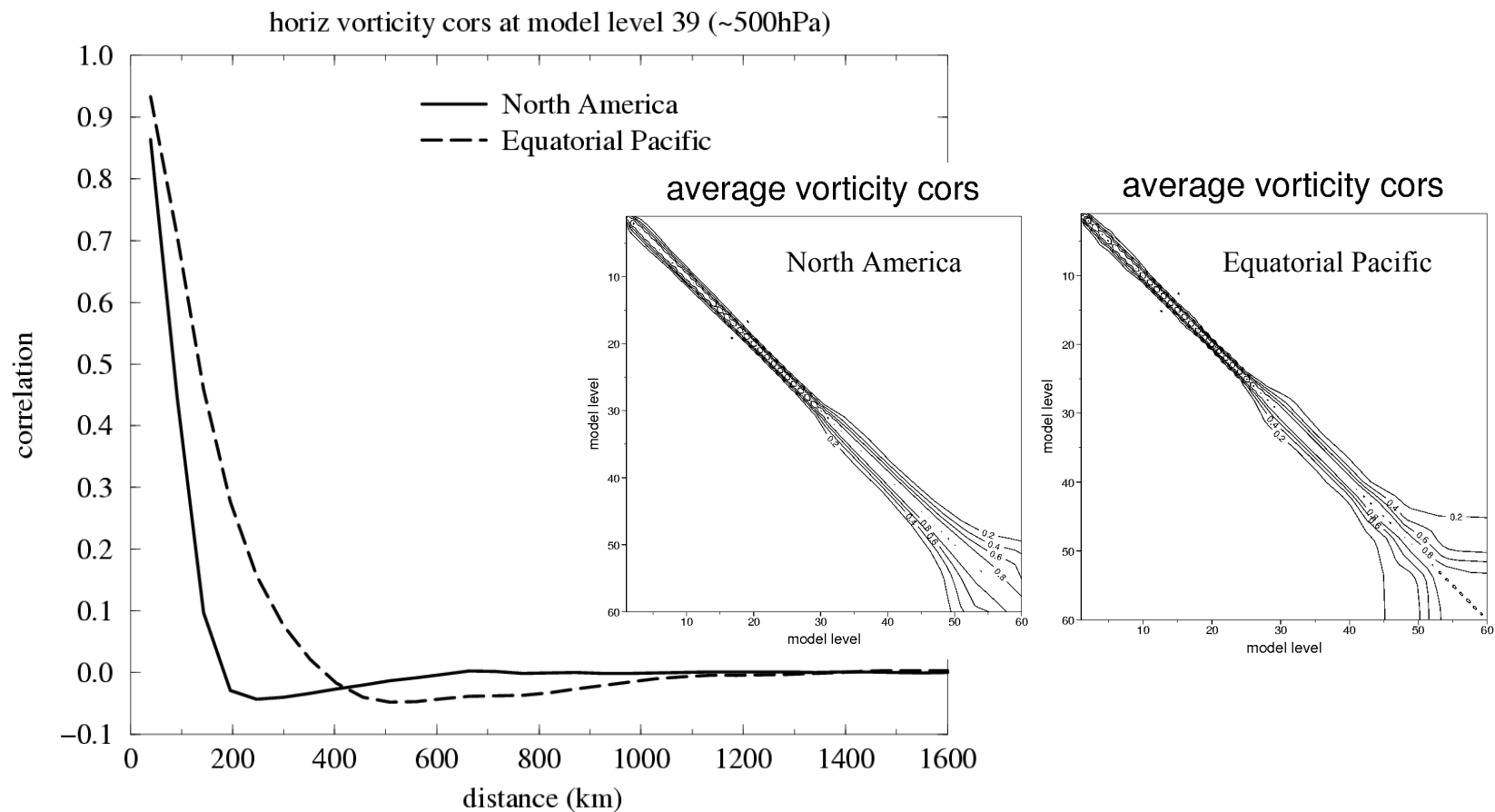
- In 2 dimensions, we have: $\mathbf{x} - \mathbf{x}_b = \sum_j \hat{\Psi}_j \otimes \Sigma_j \chi_j$

where Σ_j is diagonal. The diagonal elements are standard deviations.

- In 3 dimensions, Σ_j becomes block-diagonal, with blocks of dimension NLEVS×NLEVS. The diagonal blocks are symmetric square-roots of vertical covariance matrices.
- This is not fully general, but is sufficient to capture the variation of vertical correlation with horizontal scale, and with location.

Wavelet J_b

● Example: Horizontal and vertical Vorticity Correlations.



Wavelet J_b – Memory and Cost

- At first sight, the memory requirements for Wavelet J_b appear high.
- However, if $\hat{\Psi}_j(n) = 0$ for $n > N_{j+p}$ then the χ_j are strictly band-limited, and may be represented on Gaussian grids of appropriate resolution.

- If we arrange for:
$$N_j \leq \frac{N_{j+1}}{\sqrt{2}}$$

then the memory requirement for the control vector is at most $p+2$ full model grids, where p is the order of the B-splines.

- Only $p+1$ full-resolution spherical transforms are required.

Wavelet J_b – Memory and Cost

- Storage for the vertical covariance matrices is potentially enormous. But, can be reduced to manageable levels by reducing their spatial resolution (e.g. one matrix for every 10 gridpoints).
- The main CPU requirement is in handling the increased-length control vector.
- The bottom line is that Wavelet J_b adds about 5% to the cost of a 4d-Var analysis.
- I think it's worth it!

Summary

- **Tight frames provide a useful mathematical construct.**
- **They share many of the desirable properties of orthogonal bases, but allow considerable flexibility.**
- **Using tight frames, a flexible family of wavelets may be defined on the sphere.**
- **Unlike grid-based wavelets, there is no pole problem.**
- **There is a lot of scope for tuning of the spectral and spatial resolution. (This remains to be explored.)**
- **Wavelet J_b is expected to be implemented operationally in the ECMWF 4d-Var system by the end of this year.**