Representation of inhomogeneous, non-separable covariances by sparse wavelet-transformed matrices

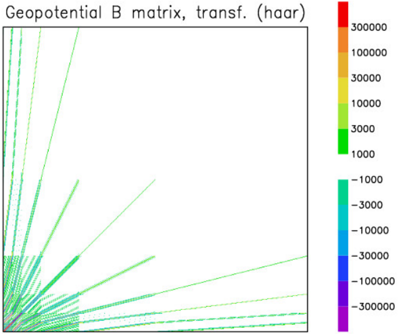
Andreas Rhodin, Harald Anlauf

German Weather Service (DWD)

Workshop on Flow-dependent aspects of data assimilation, ECMWF, Reading

June 12, 2007

Method: modelling of covariance matrices with wavelets



Wavelet transformed
zonally averaged covariance matrix
(NMC method, 480×480 coefficients)
for geopotential height
at 500 hPa, 60° N.

- Only few coefficients of wavelet transformed covariance matrices are considerably different from zero.
- Covariances are representable by extremely sparse matrices.
- No approximations beyond truncation of wavelet basis expansion.

Outline

1. Method
   - Wavelet transformation
   - Factorisation of **B**, truncated expansion
   - Multivariate covariances
2. Implementation
   - Grid
   - Truncation, factorisation, zonal averaging
   - 2D examples
   - Implementation in the 3D-Var–PSAS
3. Topics of further research
   - Statistical aspects of sample covariances and thinning
   - Flow Dependence
4. Conclusions

Discrete Wavelet transformation (DWT)

- Fast hierarchical transform with operation count $O(n)$.

coefficients (λ) of the gridded function															
$\lambda_{4,1}$	$\lambda_{4,2}$	$\lambda_{4,3}$	$\lambda_{4,4}$	$\lambda_{4,5}$	$\lambda_{4,6}$	$\lambda_{4,7}$	$\lambda_{4,8}$	$\lambda_{4,9}$	$\lambda_{4,10}$	$\lambda_{4,11}$	$\lambda_{4,12}$	$\lambda_{4,13}$	$\lambda_{4,14}$	$\lambda_{4,15}$	$\lambda_{4,16}$
			low pass ↓									high pass ↓			
$\lambda_{3,1}$	$\lambda_{3,2}$	$\lambda_{3,3}$	$\lambda_{3,4}$	$\lambda_{3,5}$	$\lambda_{3,6}$	$\lambda_{3,7}$	$\lambda_{3,8}$	$\gamma_{3,1}$	$\gamma_{3,2}$	$\gamma_{3,3}$	$\gamma_{3,4}$	$\gamma_{3,5}$	$\gamma_{3,6}$	$\gamma_{3,7}$	$\gamma_{3,8}$
		low pass ↓				high pass ↓									
$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{2,3}$	$\lambda_{2,4}$	$\gamma_{2,1}$	$\gamma_{2,2}$	$\gamma_{2,3}$	$\gamma_{2,4}$	$\gamma_{3,1}$	$\gamma_{3,2}$	$\gamma_{3,3}$	$\gamma_{3,4}$	$\gamma_{3,5}$	$\gamma_{3,6}$	$\gamma_{3,7}$	$\gamma_{3,8}$
	low pass ↓		high pass ↓												
$\lambda_{1,1}$	$\lambda_{1,2}$	$\gamma_{1,1}$	$\gamma_{1,2}$	$\gamma_{2,1}$	$\gamma_{2,2}$	$\gamma_{2,3}$	$\gamma_{2,4}$	$\gamma_{3,1}$	$\gamma_{3,2}$	$\gamma_{3,3}$	$\gamma_{3,4}$	$\gamma_{3,5}$	$\gamma_{3,6}$	$\gamma_{3,7}$	$\gamma_{3,8}$
low ↓	high ↓														
$\lambda_{0,1}$	$\gamma_{0,1}$	$\gamma_{1,1}$	$\gamma_{1,2}$	$\gamma_{2,1}$	$\gamma_{2,2}$	$\gamma_{2,3}$	$\gamma_{2,4}$	$\gamma_{3,1}$	$\gamma_{3,2}$	$\gamma_{3,3}$	$\gamma_{3,4}$	$\gamma_{3,5}$	$\gamma_{3,6}$	$\gamma_{3,7}$	$\gamma_{3,8}$
wavelet (γ) and scaling function (λ) coefficients															

- Coefficients of the transformed vector correspond to the average value (λ) and to the deviations (γ) on different scales.
- Inverse transform is a fast hierarchical transform as well.

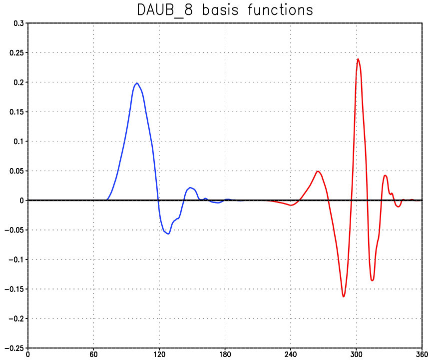
Wavelet basis functions

- The wavelet (ψ) and scaling (φ) basis functions are implicitly defined by the high and low-pass filter coefficients.
- Direct wavelet transform **W**:

  $f = \sum_k \lambda_{0,k}\, \varphi_{0,k} + \sum_{j,m} \gamma_{j,m}\, \psi_{j,m}$
- Inverse Wavelet transform $\mathbf{W}^{-1}$:

  $\lambda_{j,k} = \langle f, \hat{\varphi}_{j,k} \rangle$

  $\gamma_{j,m} = \langle f, \hat{\psi}_{j,m} \rangle$
- For orthogonal wavelet transformations the direct basis functions (ψ, φ) and duals ($\hat{\psi}, \hat{\varphi}$) are the same.



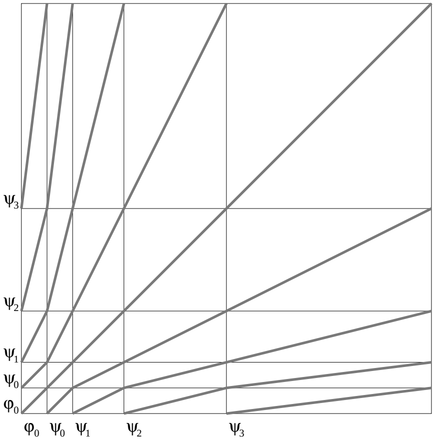
Basis functions (φ, ψ) for the orthogonal Daubechies-8 wavelet transform.

Wavelet transformed covariance matrices

- Wavelet transformations apply independently to each row and column of a matrix.
- Covariances of phenomena at the same scale are represented by the diagonal blocks.

  Covariances of phenomena at different scale are represented by off-diagonal blocks.
- Covariances of phenomena at nearby locations are represented by coefficients in the vicinity of the 'branches' (diagonals of the blocks).

  Only these coefficients are considerably different from zero.



Block structure of wavelet transformed matrices

Factorisation of $\mathbf{B}$; Truncated expansion

- Covariance matrix in wavelet representation:

$$\mathbf{B} = \mathbf{W}\hat{\mathbf{B}}\mathbf{W}^T$$

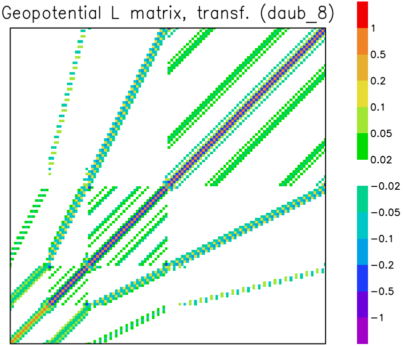
Neglection of small coefficients of $\mathbf{B}$ may lead to indefinite matrices.

- Factorise $\hat{\mathbf{B}}$ to ensure positive definite covariance matrices:

$$\hat{\mathbf{B}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T$$

- Large coefficients of the symmetric square root $\hat{\mathbf{L}}$ have similar sparse pattern as $\hat{\mathbf{B}}$.

- Neglect small coefficients of $\hat{\mathbf{L}}$ for a sparse representation.

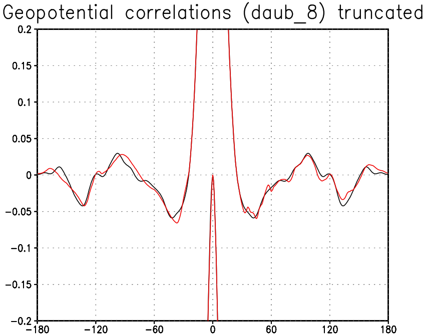


Symmetric square root $\hat{\mathbf{L}}$ of $\hat{\mathbf{B}}$ in wavelet representation.

The matrix is rescaled so that the diagonal elements are equal to 1. Only 240×240 of the 480×480 coefficients are shown.

Truncated Wavelet Expansion (zonal direction)

- Number of coefficients for a reasonable approximation of **B**: 5 to 10 times the number of grid-points.
- Variations far from the center of the covariance function consist of noise due to limited size of the NMC ensemble.
- ‘Noise’ is approximated less accurately than the signal.



Zonally averaged geopotential height correlations in 60° N.

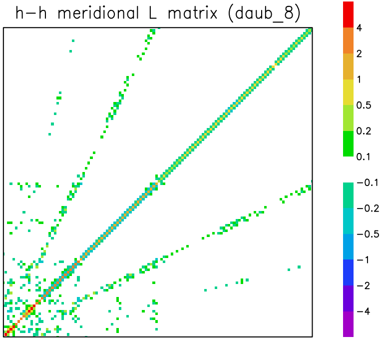
Black: NMC derived correlations (31 forecast pairs, 1 month).

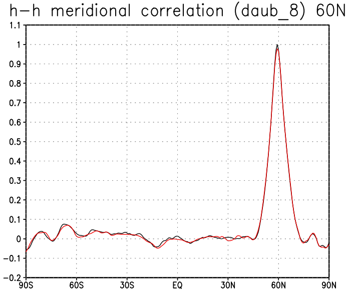
Red: Correlations approximated by the truncated wavelet expansion. The largest 2400 coefficients (0.5% threshold) of the 230400 coefficients of $\hat{\mathbf{L}}$ (1%) were used, corresponding to 5 times the number of grid-points ($N_x = 480$).

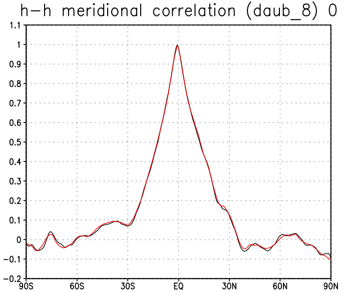
(A value of 1 has been subtracted from the central part of the plot.)

Truncated Wavelet Expansion (meridional)

- Apply method to zonally averaged meridional covariance matrices.
- Covariance functions are inhomogeneous.
- Zonal averaging has no effect at the poles.







Top: Truncated matrix $\hat{\mathbf{L}}_{hh}$ (0.5% threshold, $10.2 \times N_x$ coefficients).
Right: Correlations with geopotential height at 60° N and 0°.

Multivariate covariances

- Multivariate covariances handled by the usual block-tridiagonal approach:

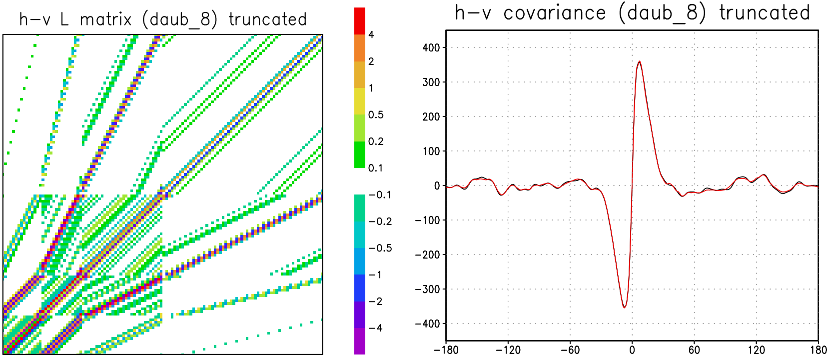
$$\hat{\mathbf{L}} = \begin{pmatrix} \hat{\mathbf{L}}_{hh} & & & \\ \hat{\mathbf{L}}_{\psi h} & \hat{\mathbf{L}}_{\psi_u \psi_u} & & \\ . & . & \hat{\mathbf{L}}_{\chi\chi} & \\ . & . & . & \hat{\mathbf{L}}_{qq} \end{pmatrix}$$

- Diagonal blocks refer to the unbalanced variables:
  height h, stream-function ψ_u, velocity potential χ and humidity q
- Cross-correlations representable by sparse wavelet transformed matrices.
  $\hat{\mathbf{L}}_{\psi h}$ for the geostrophic balance operator.
- No approximations beyond truncation errors if lower triangle is completed.
  Block-triangular form is a Cholesky decomposition of $\hat{\mathbf{B}}$.

Multivariate covariances (wind–height)

- Complex cross-covariances representable by sparse wavelet expansion.
- Wind–height covariances (v-h) shown here:
  (instead of the simpler ψ-h covariances)

$$\mathbf{B}_{vh} = \mathbf{W}\,\hat{\mathbf{L}}_{vh}\hat{\mathbf{L}}_{hh}^{T}\,\mathbf{W}^{T}$$



Truncated matrix $\hat{\mathbf{L}}_{vh}$ (2% threshold, $9.2 \times N_x$ coefficients) and approximated wind–height (v-h) covariances.

Implementation issues: 3 dimensional Grid

The method works well in 1-D.
3-D implementation issues:

- Wavelet transforms on regular 3-D grids apply separately to each of the dimensions:

$$\mathbf{W_{3D} = W_x W_y W_z}$$

- PSAS formulation of the DWD 3D-Var does not depend on a specific grid.
- Our choice:
  - Gaussian Grid, 512×256 grid-points
  - 64 pressure levels equidistant in $\log p$

Truncation

- In higher dimensions $\hat{\mathbf{B}}$ and $\hat{\mathbf{L}}$ remain sparse.

  Number of coefficients per grid-point n/N:

Direction	n/N
1D horizontal or vertical:	10
2D horizontal:	30 - 50
2D vertical:	30
3D (expected):	100

- Operation count of $\hat{\mathbf{L}}\mathbf{x}$ comparable to $\mathbf{W}\mathbf{x}$.
- Further savings in storage possible due to symmetry and homogeneity.

Factorisation, zonal average

Zonal averaging essential in 'parameterless fit' to derive coefficients of $\hat{\mathbf{B}}$.

Procedure:

1. Estimate $\tilde{\mathbf{B}}$ in Fourier representation from ensemble vectors $\mathbf{u}$ (NMC method), including zonal averaging:

$$\tilde{\mathbf{B}} = \frac{1}{n} \left(\mathbf{W}_{\mathbf{yz}}^{-1}\mathbf{F}_{\mathbf{x}}^{-1}\mathbf{u}\right) \left(\mathbf{W}_{\mathbf{yz}}^{-1}\mathbf{F}_{\mathbf{x}}^{-1}\mathbf{u}\right)^{\mathbf{T}}$$

2. Extract square root:

$$\tilde{\mathbf{B}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\mathbf{T}}$$

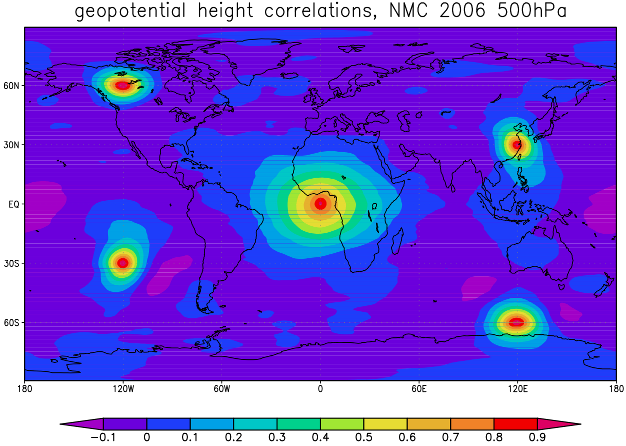
3. Transform $\tilde{\mathbf{L}}$ to wavelet representation:

$$\hat{\mathbf{L}} = \mathbf{W}_{\mathbf{x}}^{-1}\mathbf{F}_{\mathbf{x}}\,\tilde{\mathbf{L}}\,\mathbf{F}_{\mathbf{x}}^{\mathbf{T}}\mathbf{W}_{\mathbf{x}}^{-\mathbf{T}}$$

   - Valid only for $\mathbf{W}_{\mathbf{x}}^{\mathbf{T}} = \mathbf{W}_{\mathbf{x}}^{-1}$ (orthogonal wavelet transform).
   - Operation count of Fourier and wavelet transformations is $O(N \log N)$ for homogeneous matrices.
   - Only a limited number of coefficients calculated and tested for relevance, starting at the diagonal and the 'branches' of $\hat{\mathbf{B}}$.

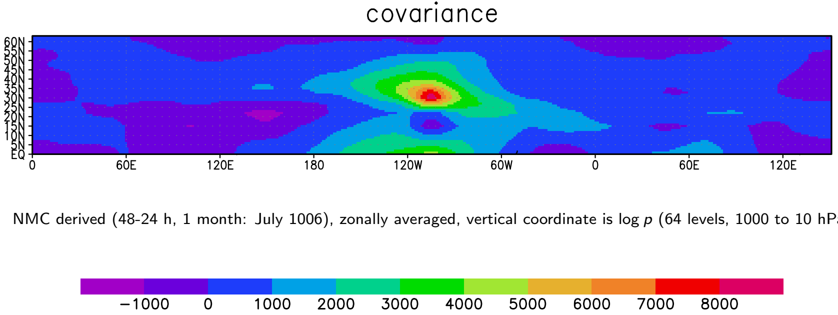
Example: 2d horizontal covariances

- Geopotential height horizontal covariance matrix, NMC derived (48-24 h, 1 year: 2006), zonally averaged.
- Number of coefficients: $40 \times N_x \times N_y$ (Gaussian grid, 512×256).



Example: 2d vertical covariances

- Zonally averaged vertical (zonal-height) covariance matrix at 0°. Covariances with geopotential height in 100 hPa.
- Number of coefficients: $30 \times N_x \times N_z$ (grid: 512×64),
- No separability assumed.
- Tilted structures can be represented.



NMC derived (48-24 h, 1 month: July 1006), zonally averaged, vertical coordinate is log p (64 levels, 1000 to 10 hPa).

Implementation in the 3D-Var–PSAS of DWD

- Full 3D implementation of the Wavelet approach:

$$\mathbf{B} = \mathbf{I}\,\mathbf{W}\,\hat{\mathbf{L}}\,\hat{\mathbf{L}}^{\mathbf{T}}\,\mathbf{W}^{\mathbf{T}}\,\mathbf{I}^{\mathbf{T}}$$

with $\hat{\mathbf{L}}$: multivariate square root of $\hat{\mathbf{B}}$
$\mathbf{W}$: 3D wavelet transform
$\mathbf{I}$: interpolation operator, differentiation: $h \rightarrow t; \quad \psi, \chi \rightarrow u, v$

- Intermediate step:
The former explicit (separable) covariance model is replaced by the equivalent wavelet formulation:

$$\mathbf{B} = \mathbf{I}\,\mathbf{K}\,\mathbf{W_v}\,\hat{\mathbf{L}}_\mathbf{v}\,\mathbf{W_h}\,\hat{\mathbf{L}}_\mathbf{h}\,\hat{\mathbf{L}}_\mathbf{h}^{\mathbf{T}}\mathbf{W}_\mathbf{h}^{\mathbf{T}}\hat{\mathbf{L}}_\mathbf{v}^{\mathbf{T}}\mathbf{W}_\mathbf{v}^{\mathbf{T}}\mathbf{K}^{\mathbf{T}}\mathbf{I}^{\mathbf{T}}$$

with $\mathbf{K}$: geostrophic balance operator relating h and ψ.
$\hat{\mathbf{L}}_\mathbf{v}\,\hat{\mathbf{L}}_\mathbf{h}$: square root of vertical and horizontal univariate covariances.

Experimentation in the 3D-Var–PSAS

Status:

- Wavelet representation of the former analytic formulation
  - analysis increment differences of 1% (truncation errors).
- Simulation of noise as expected from NMC ensembles of 3 months
  - degradation of the analyses.

Next steps:

- Further reduce noise by:
  - 1 year statistics.
  - Specific filtering at the poles.
  - Thresholding based on statistical reasoning.
- Use horizontal covariances from NMC statistics.

Statistical aspects of sample covariances and thinning

- Ensemble $x = x_{\text{model}} - x_{\text{true}}$, true covariance: $\mathbf{B} = \mathrm{E}\left\{x\,x^\dagger\right\}$
- Unbiased estimate: sample covariance

$$\mathbf{S} = \frac{1}{N-1}\sum_{m=1}^{N}\left(x^{(m)} - \bar{x}\right)\left(x^{(m)} - \bar{x}\right)^\dagger \;, \quad \mathrm{E}\left\{\mathbf{S}\right\} = \mathbf{B}$$

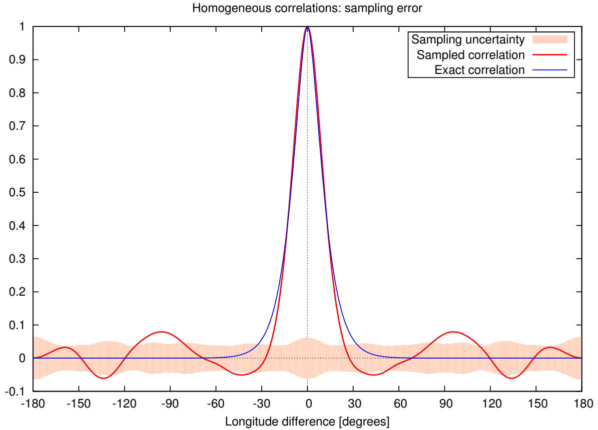
- Variance of sample covariance coefficients (Gaussian errors)

$$\sigma_{ij}^2 \;\equiv\; \mathrm{E}\left\{(S_{ij} - B_{ij})^2\right\} = \frac{1}{N-1}\left[B_{ii}B_{jj} + (B_{ij})^2\right]$$

  Error of off-diagonal coefficients dominated by diagonal elements!
- Zonal averaging increases "effective sample size" by roughly $N_x/(2L_x)$, where L_x is the correlation length scale in grid points

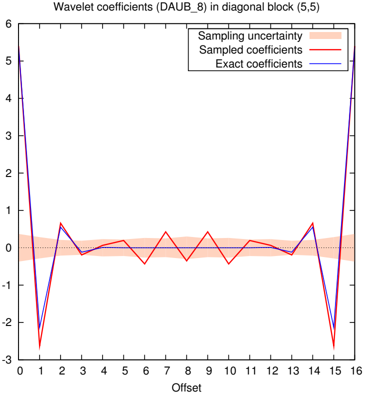
Estimation of a homogeneous correlation in grid space

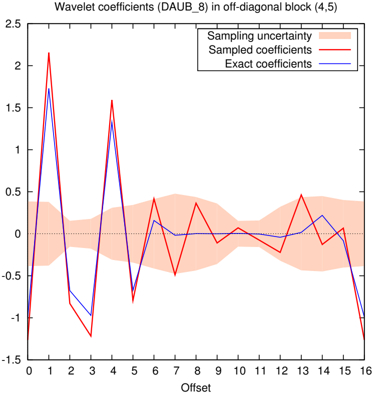
- Monte-Carlo simulation with known truth
  - Isotropic correlation on sphere ($L = 500\,\text{km}$), compactly supported, Gaspari&Cohn
  - Test sample: ensemble with $N = 31$, latitude: 60°
  - Grid space representation to 1% needs 125 coefficients ($N_x = 512$)!



Estimation of wavelet coefficients

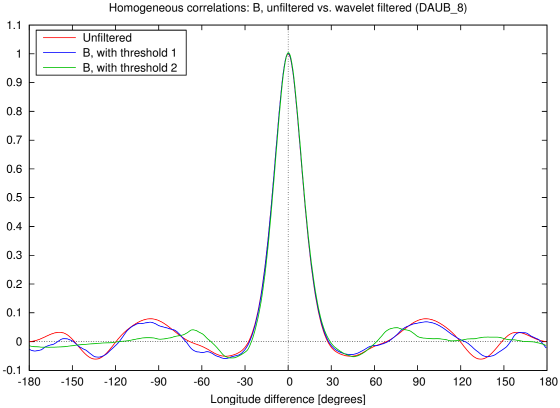
- Wavelet transformation collects information in few coefficients
  - Noise is almost evenly distributed at each scale combination
- Statistical filtering: selection of a (connected) coefficient block
  - Student's t test (single coefficient) or χ^2 test (multiple coefficients)





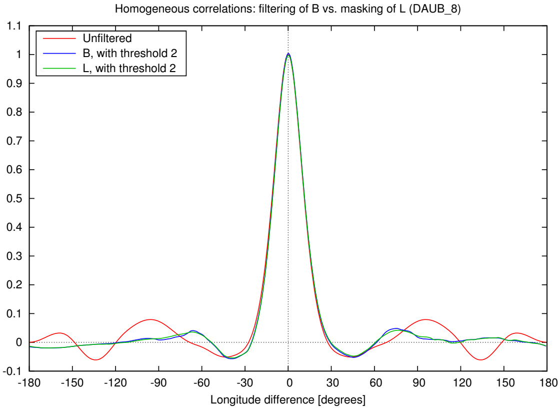
Statistical thinning using wavelets: homogenous case

- Choose *threshold* according to the desired confidence level for significance testing
  - Example 1: $|t| > 1$ and $\chi^2/\text{d.o.f} > 1$
  - Example 2: $|t| > 2$ and $\chi^2/\text{d.o.f} > 4$



Statistical thinning using wavelets: homogenous case

- Thinning of **B** destroys positive-definiteness!
- Observation: sparsity pattern of *symmetric square root* **L** very close to that of **B**
  - Take mask from **B**, but apply to **L**
  - Reconstruct **B**



Flow Dependence:

- Up to now static covariance matrices, derived from (NMC) forecast differences over long periods (1 month to 1 year).
- Method (parameterless fit) cannot be applied to EnKF ensembles ($n < 100$) because noise is not eliminated.
- Aim: blend static covariance matrix derived from large ensemble with information from a smaller EnKF ensemble.

  2 possible approaches:

  - Add free parameters (diagonal matrix **D**) to the static covariance model.
    Adjust (fit) the free parameters so that the statistical properties of the EnKF ensemble are met.

    $$\mathbf{B} = \mathbf{I}\,\mathbf{W}\,\hat{\mathbf{L}}\,\mathbf{D}\,\hat{\mathbf{L}}^{\mathbf{T}}\,\mathbf{W}^{\mathbf{T}}\,\mathbf{I}^{\mathbf{T}}$$

  - Estimate coefficients of $\hat{\mathbf{B}}$ or $\hat{\mathbf{L}}$ from both the large ensemble and the small EnKF ensemble and estimate an optimal value from the respective values and spreads.

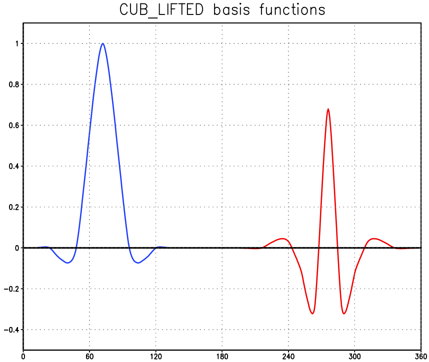
Conclusions

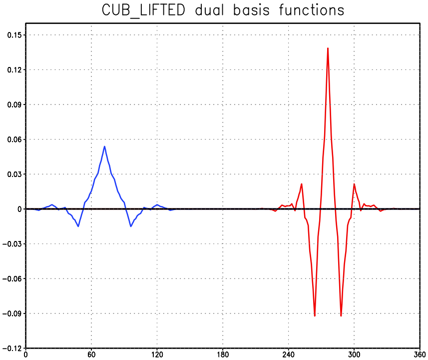
- NMC derived covariances are represented with 1% accuracy by truncated wavelet basis expansions.
- The method has been implemented and tested in 1 and 2 dimensions.
- The cost is comparable to that of other methods (spectral transform).
- The accuracy of the analyses is limited by the noise introduced by the NMC statistics.
  - More efficient filtering procedures are tested.
- Up to now static covariances were derived.
  - Flow dependent extensions are investigated.

Spare slides

...

Bi-Orthogonal Wavelet basis functions



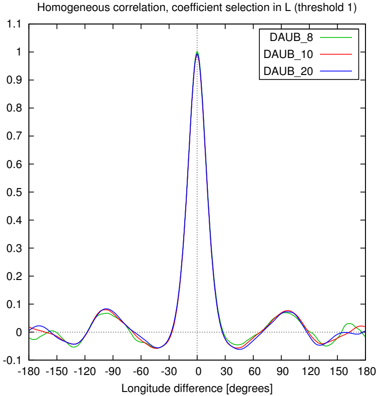


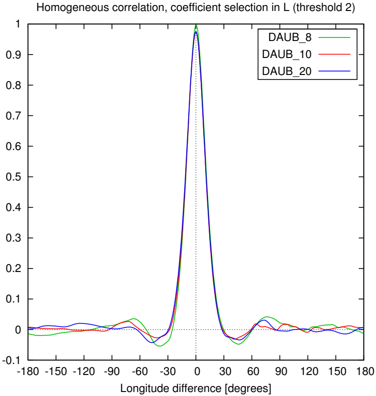
Scaling φ and wavelet basis functions ψ for a bi-orthogonal transform (left) and the respective dual basis functions $\hat{\varphi}$, $\hat{\psi}$ (right).

- 2-dimensional transforms on irregular grids can be constructed with bi-orthogonal wavelets by the Lifting Scheme.

Statistical thinning: dependence on wavelet basis

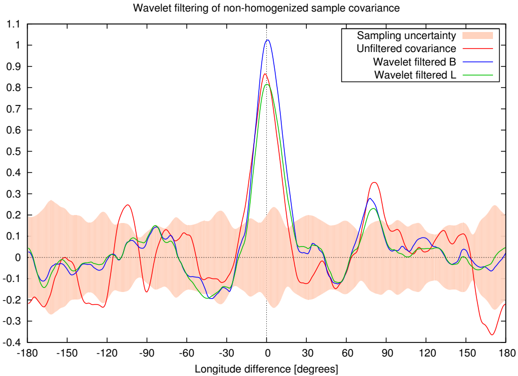
- Thinning of **L** using sparsity pattern derived from analysis of **B**
- Filtering properties depend on wavelet smoothness, support length
  - Smoother wavelets lead to smoother results, but have longer 'tails'!





Statistical thinning using wavelets: inhomogeneous case

- More sophisticated estimation techniques necessary (e.g. T. Cai)
- A simple test: just take over mask from homogeneous case
  - Smoothing quite effective
  - Thinning of **L** degrades normalization of **B**!
  - A posteriori adjustment necessary (e.g. renormalization in 1d)
  - Not clear yet how to do in 2d and 3d, but unavoidable



Statistical thinning: meridional covariances

- Meridional covariances are noisier, even with zonal averaging
    - Unfavorably large correlations between poles difficult to suppress on statistical grounds only
    - Ad-hoc or model-based elimination of affected coefficients impairs normalization