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

Andreas Rhodin and Harald Anlauf

Deutscher Wetterdienst (German Weather Service)
D-63004 Offenbach, Postfach 10 04 65, Germany
andreas.rhodin@dwd.de harald.anlauf@dwd.de

ABSTRACT

NMC derived forecast error covariances are modelled by explicit sparse matrices in wavelet transformed representation. The number of nonzero coefficients in the sparse matrix representation is of order 10 times the number of grid-points (to be compared with the number of grid-points squared for a full representation) in 1 dimension and 30 to 50 in 2 dimensions for an accuracy of 1%. We expect that of order 100 coefficients times the number of grid-points will be required in 3-d. This makes this wavelet transform method competitive with other approaches as for instance the spectral transform.

The basic idea of the wavelet transform and its use for sparse matrix representation is described for univariate and multivariate covariances. Technical issues as the extraction of the symmetric square root and zonal averaging in higher dimensions as well as statistical aspects of sample covariance matrices are discussed. First results using a 2-dimensional implementation of the approach in the DWD 3DVAR are presented.

The method presented relies on NMC or analysis ensemble statistics using a large number of forecast differences or ensemble members in order to estimate the matrix coefficients. The prospects of the wavelet approach for flow dependent covariance modelling are discussed.

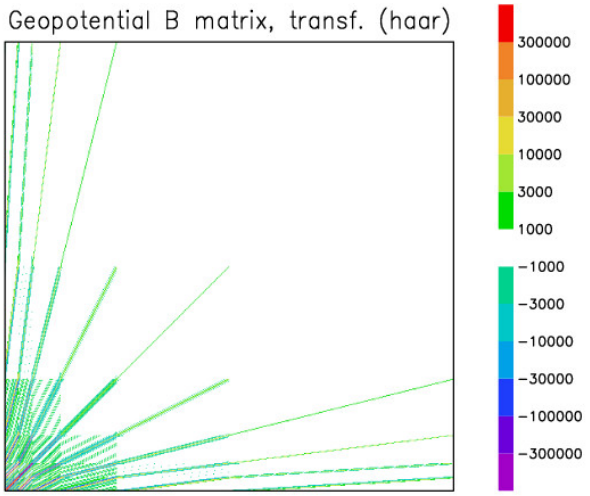
1 The Method

Figure 1a shows a wavelet transformed NMC derived covariance matrix for geopotential height in one dimension. It can be seen that only a small fraction of the coefficients are considerably different from zero. Thus, if small coefficients are neglected, covariances may be modelled by extremely sparse matrices in this representation.

The large coefficients show a distinctive pattern, reflecting the block decomposition of the wavelet transformed matrices. Substantial nonzero coefficients are located in narrow bands along the diagonal and along the off-diagonal 'branches' indicated by the gray lines in Figure 1b. The coefficients at these positions correspond to correlations between basis functions of the same or of different scales but at approximately the same location. Other correlation coefficients are almost zero and can be neglected in a truncated expansion.

The method to implement matrix operators in sparse wavelet-transformed representation has been proposed by Beylkin et al. (1991) and was applied to geostatistics by Nychka et al. (2002). Forecast error covariances in wavelet transformed diagonal representation are used by Fisher and Andersson (2001) and Deckmyn and Berre (2005). In contrast to the latter the method proposed here uses off-diagonal coefficients as well.

a)



b)

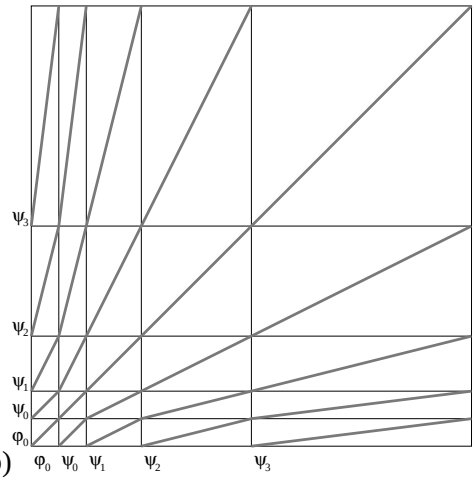


Figure 1: a) Wavelet transformed zonally averaged covariance matrix (NMC method, 480×480 coefficients) for geopotential height at 500 hPa, 60°N.
b) Block decomposition of covariance matrices in wavelet representation. Each diagonal block holds coefficients related to the covariances between scaling functions (φ_0) or wavelet functions (ψ_i) at the same scale. Each off-diagonal block holds coefficients related to covariances between scaling functions and wavelets or between wavelets of different scale.

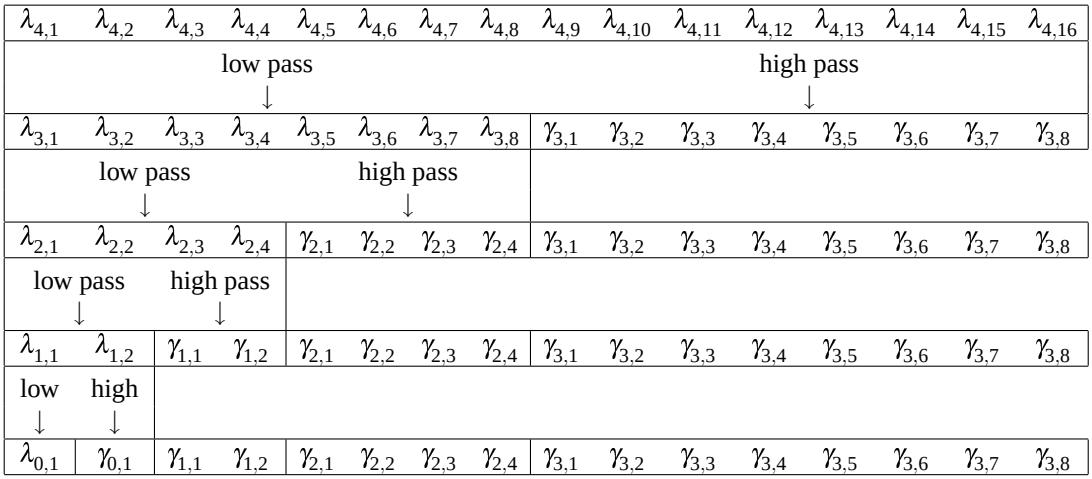
1.1 Wavelet transformation of covariance matrices

A fast hierarchical algorithm can be implemented for the Discrete Wavelet Transform (DWT). The procedure starts at the finest resolution (top level n in Figure 2a with $n=4$) there the coefficients $\lambda_{n,k}$ are identified with the coefficients of the gridded function to be transformed. In the first step of the hierarchical transform a low pass filter is used to obtain coefficients $\lambda_{n-1,l}$ of the scaling function expansion on the next coarser scale, keeping only half the number of coefficients. The complementary information is derived by application of a high pass filter yielding coefficients $\gamma_{n-1,l}$ of the wavelet function expansion. The procedure is repeated on the scaling function coefficients $\lambda_{j,l}$, leaving the wavelet coefficients $\gamma_{j,l}$ unchanged. It terminates when only one coefficient $\lambda_{0,1}$ is left or the number of coefficients cannot be divided by 2 any more. If the filter functions have compact support the number of operations required for this Discrete Wavelet Transformation scales linearly with the number of grid-points N. Thus it is faster than the Fast Fourier Transform which scales with $O(N \times \log N)$.

The analysis transform described above derives the wavelet expansion coefficients of a gridded function. In the remainder of this paper it is denoted as the inverse wavelet transform $\mathbf{W^{-1}}$. The synthesis $\mathbf{W}$, which reconstructs the function from the coefficients of the wavelet expansion, is a hierarchical transform as well. The basic algorithms of variational data assimilation use the synthesis $\mathbf{W}$ and its adjoint $\mathbf{W^T}$. The analysis $\mathbf{W^{-1}}$ is merely required for calculating the coefficients of the covariance matrices in wavelet representation.

The basis functions of the wavelet transform (scaling functions φ related to coefficients λ and wavelets ψ related to coefficients γ) are localised both in frequency and in physical space. They are implicitly defined by the filter coefficients of the DWT. Figure 2b shows the shapes of the Daubechies 8 basis functions used in this study. They are one member of a family of orthogonal wavelets (Daubechies , 1992). Orthogonality is not mandatory for this application. Bi-orthogonal wavelets are suitable as well. Basis functions merely must be sufficiently smooth to obtain a good approximation of the covariance matrices with a small number of

a)



b)

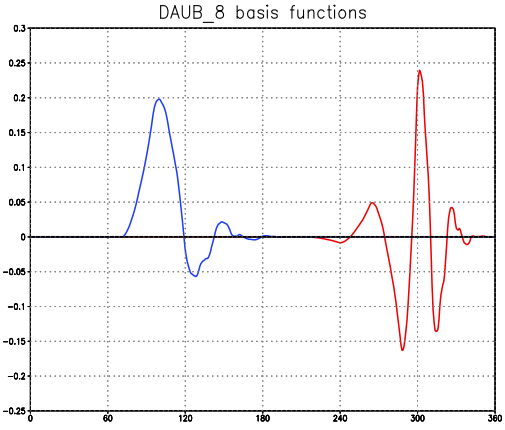


Figure 2:
a) Hierarchical Discrete Wavelet Transform: The analysis transform takes the coefficients $\lambda_{4,1}, \lambda_{4,2} \ldots$ of a gridded vector (top) and provides the coefficients $\lambda_{0,1}, \gamma_{0,1}, \gamma_{1,1} \ldots$ of the wavelet expansion (bottom).
b) Scaling functions φ (left, blue) and wavelet basis function ψ (right, red) of the Daubechies 8 basis functions used in this study.

coefficients. A more detailed discussion of different wavelet basis function properties is given in Rhodin and Anlauf (2007).

1.2 Univariate covariances

The forecast error covariance matrices used in this study were estimated by the NMC method from an ensemble of 31 forecast differences at 0 UTC in October 2004. The NMC method takes the differences between forecasts $\mathbf{u^{24}}$ and $\mathbf{u^{48}}$, valid at the same date but with different lead times (here 24 and 48 hours) as a surrogate for the differences to the truth.

$$\mathbf{B_{NMC}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{u_i^{48}} - \mathbf{u_i^{24}})(\mathbf{u_i^{48}} - \mathbf{u_i^{24}})^T \tag{1}$$

In addition to the averaging over the 31 days of the month we average over all instances of the ensemble members shifted in zonal direction by multiples of the grid spacing. This considerably improves the statistical basis but results in homogeneous covariances in the zonal direction.

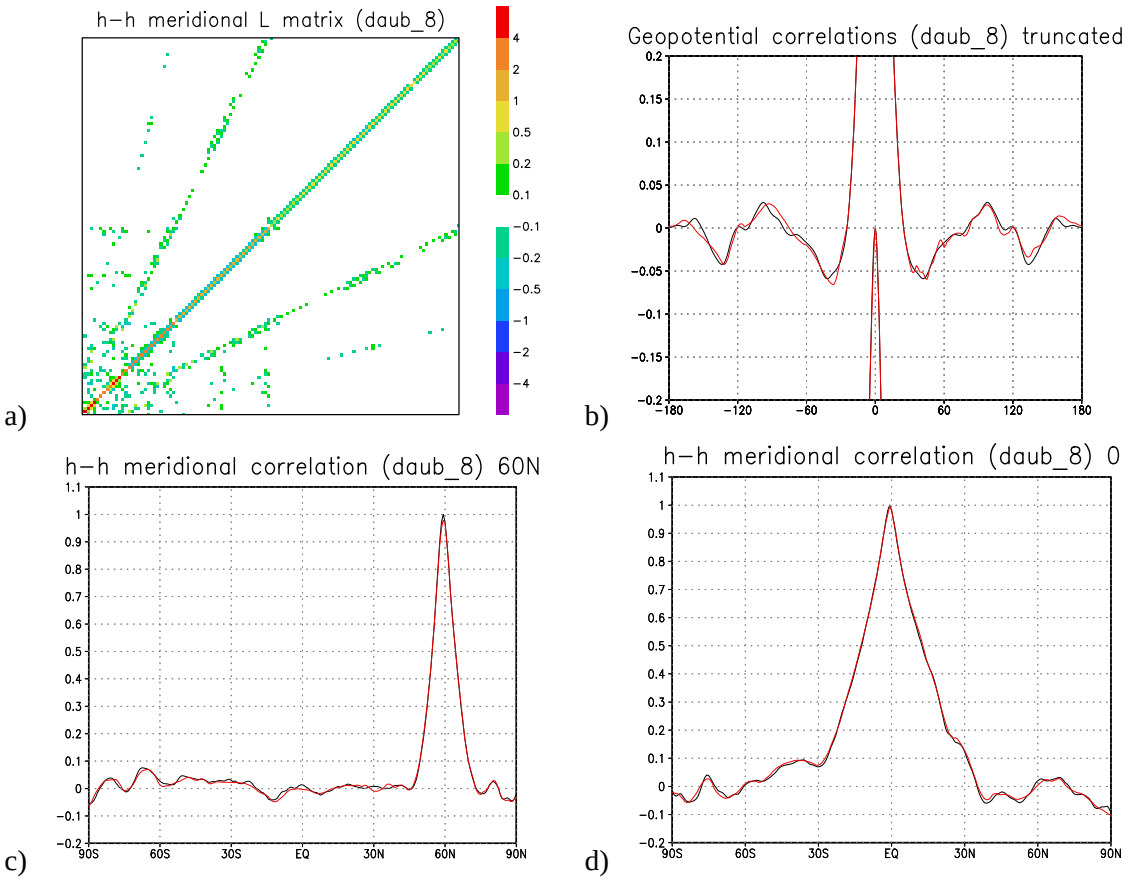


Figure 3:
a) Square root $\hat{\mathbf{L}}_{hh}$ of the zonally averaged meridional covariance matrix for geopotential height. Only the lower edge of the matrix is shown holding the 120 x 120 coefficients related to scaling functions $\varphi_{0,1\ldots15}$ and wavelets $\psi_{0,1\ldots15}$, $\Psi_{0,1\ldots15}$, $\Psi_{0,1\ldots30}$.
b) Original (black line) and approximated (red) homogeneous correlation for geopotential height in zonal direction. The curves are shown a second time with an offset of -1. to focus on the peak.
c, d) Original (black line) and approximated (red) inhomogeneous correlation for geopotential height in meridional direction with a location at 60°N and at 0°.

Forecasts from the global model GME of geopotential height and wind component *v* at the 500 hPa level interpolated to a regular longitude-latitude grid with 0.75 degree resolution (480 x 240 grid-points) are used for the investigations of Sections 1.2 and 1.3. Since 240 factorises as $2^4 \times 15$, the wavelet basis function expansion consists of 15 scaling functions φ_0 and wavelet functions ψ_0 on the coarsest scale. In this 1-dimensional study we present covariances either in zonal direction for a selected latitude (60° N), or in meridional direction on a great circle touching the poles.

We want to obtain a representation of the matrix $\mathbf{B}$ in wavelet representation, i.e.:

$$\mathbf{W}\hat{\mathbf{B}}\mathbf{W}^T = \mathbf{B} \tag{2}$$

To derive $\hat{\mathbf{B}}$ in wavelet representation we apply the inverse transformation:

$$\hat{\mathbf{B}} = \mathbf{W}^{-1}\mathbf{B}\mathbf{W}^{-T} \tag{3}$$

Technically, the inverse transformation $\mathbf{W}^{-1}$ has to be applied to all rows and all columns of $\mathbf{B}$. The subdivision of a wavelet transformed vector in segments with coefficients $\lambda_{0,..}$, $\gamma_{0,..}$, $\gamma_{1,..}$, ... as shown in the bottom row of

Figure 2a is reflected in the respective subdivision of the rows and columns of the wavelet transformed matrix in Figure 1b.

Setting small coefficients of $\hat{\mathbf{B}}$ to zero in order to obtain a sparse representation may lead to an indefinite matrix. For that reason we extract the symmetric square root matrix $\hat{\mathbf{L}}$ with:

$$\hat{\mathbf{B}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T \tag{4}$$

The large coefficients of $\hat{\mathbf{L}}$ show a similar sparsity pattern as $\hat{\mathbf{B}}$ (Figure 3a). Now coefficients of $\hat{\mathbf{L}}$ below a certain threshold (0.5% in this example) can be neglected without the risk of an indefinite $\mathbf{B}$. The homogeneous correlation function in zonal direction represented by this truncated wavelet expansion is shown in Figure 3b. The original correlation function shows spurious fluctuations in some distance to the center of the peak. These patterns mainly consist of statistical noise due to the limited number of forecast differences used for the estimation of $\mathbf{B}$. The noise is reproduced by the truncated wavelet expansion, although the accuracy is worse than for the peak of the covariance function. It is possible to use wavelet transformed covariance matrices for filtering of noise intentionally by setting to zero certain off-diagonal matrix coefficients. Filtering by coefficient selection based on their statistical significance will be further discussed in Section 1.4. The filtering properties of diagonal matrices in wavelet transformed representation were investigated by Pannekoucke et al. (2007).

Inhomogeneous zonally averaged meridional correlation functions are shown in Figure 3c and d. Covariances with a location at the equator have considerably different shape and length scale reflecting the inhomogeneity in meridional direction. The spatial variations are reflected by respective variations of the coefficients of the $\hat{\mathbf{L}}$-matrices along the diagonals and the 'branches' in Figure 3a. The meridional covariances with a location at 60° N show enhanced oscillations due to noise near the poles. Here the zonal averaging does not increase the effective size of the ensemble.

1.3 Multivariate covariances

In atmospheric data assimilation covariances and cross-covariances are required for the prognostic variables temperature t, specific humidity q, and the wind components u and v. In general, covariances are modelled for a different set of variables which are less correlated so that cross-covariances are more easily specified. Here, these variables are chosen as geopotential height h, stream-function ψ (or its uncorrelated part ψ_u), velocity potential χ and relative humidity rh. The former prognostic variables can be derived from the latter: t by vertical differentiation of h; u and v by horizontal differentiation of ψ and χ. The most prominent cross-covariances are those between geopotential height and wind due to the geostrophic balance condition. Cross-covariances between the remaining variables are smaller and are often not accounted for.

We write the complete multivariate covariance matrix as:

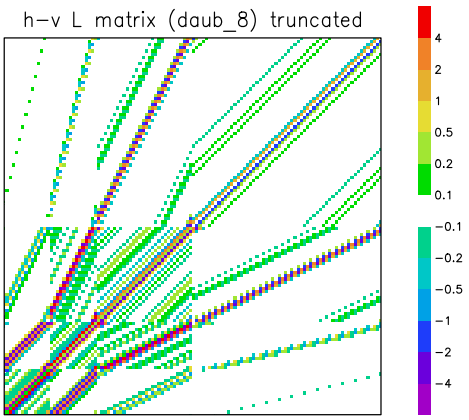
$$\mathbf{B} = \mathbf{W}\hat{\mathbf{L}}\hat{\mathbf{L}}^T\mathbf{W}^T \tag{5}$$

Here, $\mathbf{W}$ denotes the application of the wavelet transform independently to each of the variables t, rh, ψ, and χ. $\hat{\mathbf{L}}$ is chosen as the block matrix:

$$\hat{\mathbf{L}} = \begin{pmatrix} \hat{\mathbf{L}}_{hh} & & & \\ \hat{\mathbf{L}}_{\psi h} & \hat{\mathbf{L}}_{\psi_u\psi_u} & & \\ . & . & \hat{\mathbf{L}}_{\chi\chi} & \\ . & . & . & \hat{\mathbf{L}}_{rhrh} \end{pmatrix} \tag{6}$$

Further off-diagonal matrices (indicated by the dots) may be added in equation (6) if cross-covariances between the remaining variables shall be accounted for. The block matrix decomposition of $\hat{\mathbf{L}}$ does not imply

a)



b)

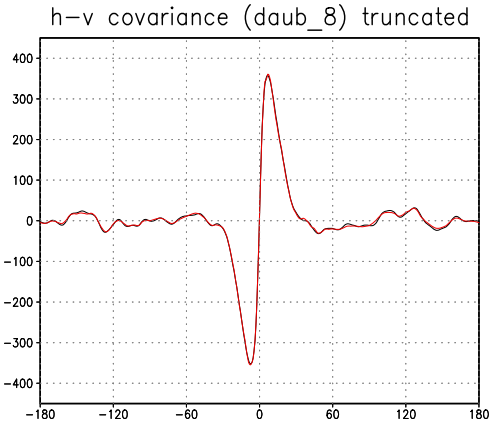


Figure 4:
a) Square root $\hat{\mathbf{L}}_{vh}$ of the zonally averaged zonal cross-covariance matrix for geopotential height and wind component v. Only the lower edge of the matrix is shown (120 x 120 coefficients).
b) Original (black line) and approximated (red) homogeneous correlation for geopotential height and wind v in zonal direction.

matrix	direction	threshold	coefficients	coef./grid-point	$\lvert\mathbf{B}_{tr}-\mathbf{B}\rvert_\infty$	$\lvert\mathbf{B}_{tr}-\mathbf{B}\rvert_2$
$\mathbf{L}_{hh}$	zonal	0.5%	2400	5.0	1.5%	0.51%
$\mathbf{L}_{vh}$	zonal	2.%	4425	9.2	6.1%	1.16%

Table 1: coefficients used and accuracy achieved for the approximated correlation functions.

any approximation beyond the errors introduced by the truncation of the wavelet expansions. The Cholesky decomposition allows to represent any positive definite matrix by a triangular square root. This is done here on a block matrix level.

The method presented above is frequently used for modelling multivariate covariances. In contrast to the usual approach we do not advocate physically motivated analytical balance operators (e.g. relate ψ to to h by a latitude dependent factor) but handle the cross-covariances consistently in the framework of sparse matrices in wavelet representation. In this formulation the cross-covariances of stream-function ψ and geopotential height h is:

$$\hat{\mathbf{B}}_{\psi h} = \hat{\mathbf{L}}_{\psi h}\hat{\mathbf{L}}_{hh}^{T} \tag{7}$$

In order to show that even complex multivariate correlations can be handled by the wavelet approach without explicitly referring to a parameterised model, we will directly derive the covariances between meridional wind component v and geopotential height $\hat{\mathbf{B}}_{vh}$ instead of modelling cross-covariances between stream-function and geopotential height:

$$\hat{\mathbf{B}}_{vh} = \hat{\mathbf{L}}_{vh}\hat{\mathbf{L}}_{hh}^{T} \tag{8}$$

$\hat{\mathbf{B}}_{vh}$ is estimated by the NMC-method similar to (1). Again $\hat{\mathbf{L}}_{vh}$ is approximated by keeping only the largest components. The results are shown in Figure 4. Neither $\hat{\mathbf{B}}_{vh}$ nor $\hat{\mathbf{L}}_{vh}$ are diagonally dominant, but the coefficients which are considerably different from zero are still located in the vicinity of the diagonal and on the 'branches' of $\hat{\mathbf{L}}$. The method is capable to represent complex cross-covariance matrices as $\mathbf{B}_{vh}$ although the number of coefficients required for an approximation of comparable accuracy is higher than for $\mathbf{B}_{hh}$. Table 1 shows the number of coefficients used and the accuracy achieved for the correlation functions.

1.4 Statistical aspects of sample covariance matrices

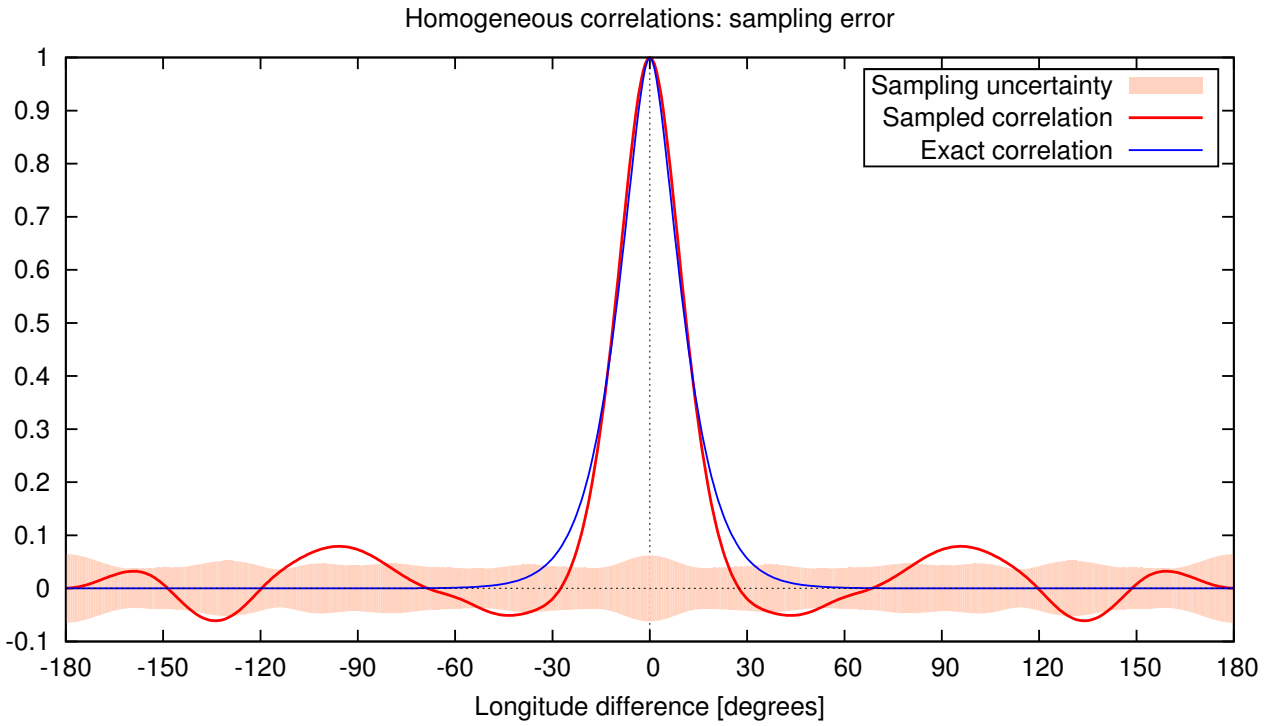


Figure 5: Zonally averaged sample correlation (red) from a random simulation with known correlation (blue). The shaded area (light red) indicates the sampling error estimated by the analysis.

The examples in the previous sections demonstrate that covariance matrices can be represented quite well by sparse matrices with a clear structure. However, the finite size of realistic samples leads to statistical fluctuations affecting coefficients both on and far from the diagonals or branches. This can be seen from the unbiased estimator of covariance,

$$\mathbf{S} = \frac{1}{N-1} \sum_{k=1}^{N} \left(\mathbf{x}_k - \overline{\mathbf{x}}\right) \left(\mathbf{x}_k - \overline{\mathbf{x}}\right)^{\mathrm{T}}, \quad \text{with} \quad \mathrm{E}\{\mathbf{S}\} = \mathbf{B}, \tag{9}$$

where $\mathbf{B}$ is the true covariance matrix and N the sample size. Assuming a Gaussian distribution of errors, we find for the variance of the sample covariance coefficients (Mallat et al. , 1998):

$$\sigma_{ij}^2 \equiv \mathrm{E}\left\{(S_{ij} - B_{ij})^2\right\} = \frac{1}{N-1}\left[B_{ii}B_{jj} + (B_{ij})^2\right] \tag{10}$$

Obviously the expected error does not only depend on B_{ij} but also on the diagonal elements B_{ii} and B_{jj}. The latter will always dominate the sampling error for those coefficients which correspond to small correlations.

One can reduce the sampling error (10) by increasing the number of independent realizations of the random process. A popular method to achieve this is zonal averaging or (weighted) averaging over neighbouring grid points, thereby trading spatial resolution for smoothness. Zonal averaging of a correlation function with correlation length L (corresponding to L_x grid points) on a periodic grid (with N_x grid points) roughly increases the effective sample size by roughly a factor $N_x/(2L_x)$. Averaging over neighbouring grid points results in somewhat smaller gains.

To illustrate the statistical effects of sampling we generated a Gaussian random field with a known truth. We chose as isotropic correlation on the sphere a 5th order function from Gaspari and Cohn (1999) with correlation length $L = 500\,\text{km}$. Figure 5 compares the zonally averaged sample correlation (ensemble size $N = 31$) at latitude 60° (red line) with the given truth (blue line). The sampling error (shaded area, light red) was determined by first calculating the zonally averaged correlation for each independent realization and then analysing the resulting 'time series' for each correlation coefficient.

Note that the estimated sampling error compares quite well with the expectations from the above reasoning. In particular, all correlation coefficients are contaminated by roughly the same amount of noise.

If we considered only those correlation coefficients statistically relevant where the 'signal' is above the 'noise level' and which are continuously connected with zero separation, the grid point representation would retain 69 coefficients (for 512 zonal grid points).[1] This number changes only slightly when stronger bounds are applied.

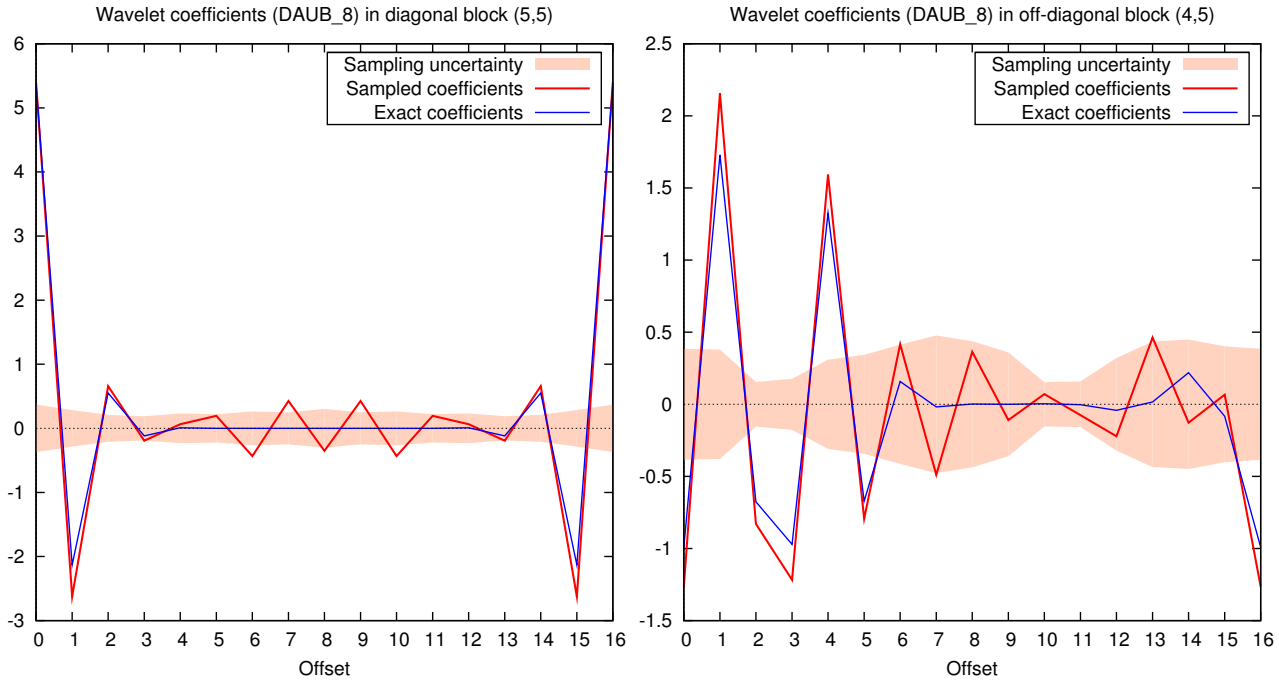


Figure 6: Wavelet coefficients as derived from the sample after zonal averaging (red) and compared to true values (blue). The typical distribution of coefficients within a diagonal block and an off-diagonal block are shown on the left and on the right, respectively. The shaded areas (light red) indicate the sampling error.

Similar patterns are found when transforming the correlation matrix to the wavelet representation. The left-hand side of figure 6 shows the characteristics of the coefficients of a matrix block which is diagonal in the wavelet scales. The largest coefficients always reside on the diagonal (offset 0 corresponds to wavelets at the same position). The corresponding features of blocks off-diagonal in wavelet scales can be seen on the right-hand side of figure 6. Again, large coefficients are found in locations close to the branches as described in the previous sections, while sampling noise is distributed roughly evenly over the coefficients within a block. This general behaviour is also found for the other wavelet scales and scale combinations. However, at smaller scales one usually finds very few coefficients to be above the noise level.

When applying a thresholding procedure to the wavelet coefficients, some care must be taken for their oscilla-

[1] The exact grid-space representation needs about 125 coefficients for 512 zonal grid points for an accuracy better than 1%.

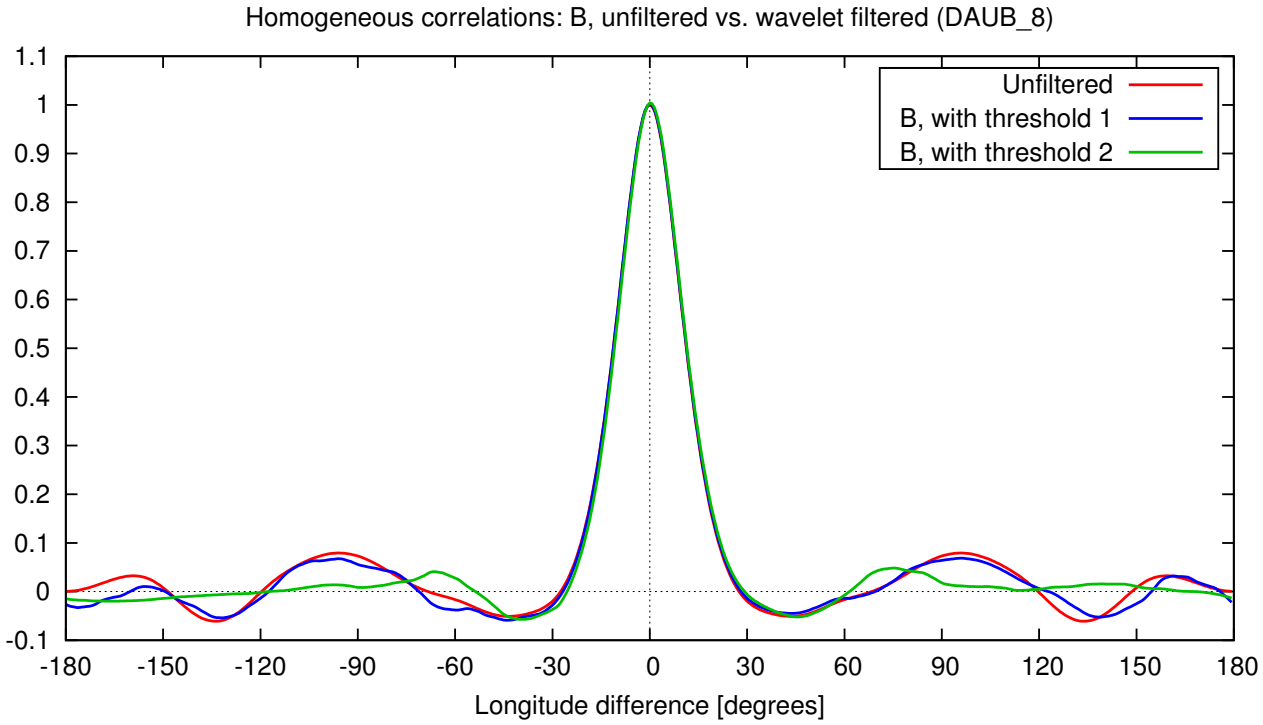


Figure 7: Filtering of homogeneous wavelet covariances by hard thresholding. The blue and green curves correspond to the two choices of threshold as described in the text.

tory behaviour. We found it convenient to define:

$$t_{ij} = \frac{S_{ij}}{\text{stddev}(S_{ij})} \cdot \sqrt{N-1}\,. \tag{11}$$

For normally distributed S_{ij} the t_{ij} follow a Student's t distribution, which can be well approximated by a normal distribution with unit variance for sufficiently large N. Furthermore, for a block of adjacent coefficients we also define a 'moving average χ^2':

$$\chi^2/\text{d.o.f} = \frac{1}{n}\sum_{\{ij\}}(t_{ij})^2\,, \tag{12}$$

which is roughly χ^2-distributed with n degrees of freedom.[2]

We used the above criteria for a hard-thresholding procedure with the following choices:

- Threshold 1: $|t| > 1$ and $\chi^2/\text{d.o.f} > 1$,
- Threshold 2: $|t| > 2$ and $\chi^2/\text{d.o.f} > 4$.

Figure 7 shows the resulting correlations when applying these thresholds to the wavelet representation. Note that after application of the smaller threshold there are 54 coefficients left, while after application of the second threshold only 37 coefficients remain. These numbers vary slightly with the choice of wavelet basis.

A further reduction in the number of coefficients is obtained by thresholding the remaining coefficients by their magnitude.

[2]This procedure can also be generalised to the inhomogeneous case. See Cai (2002) for details.

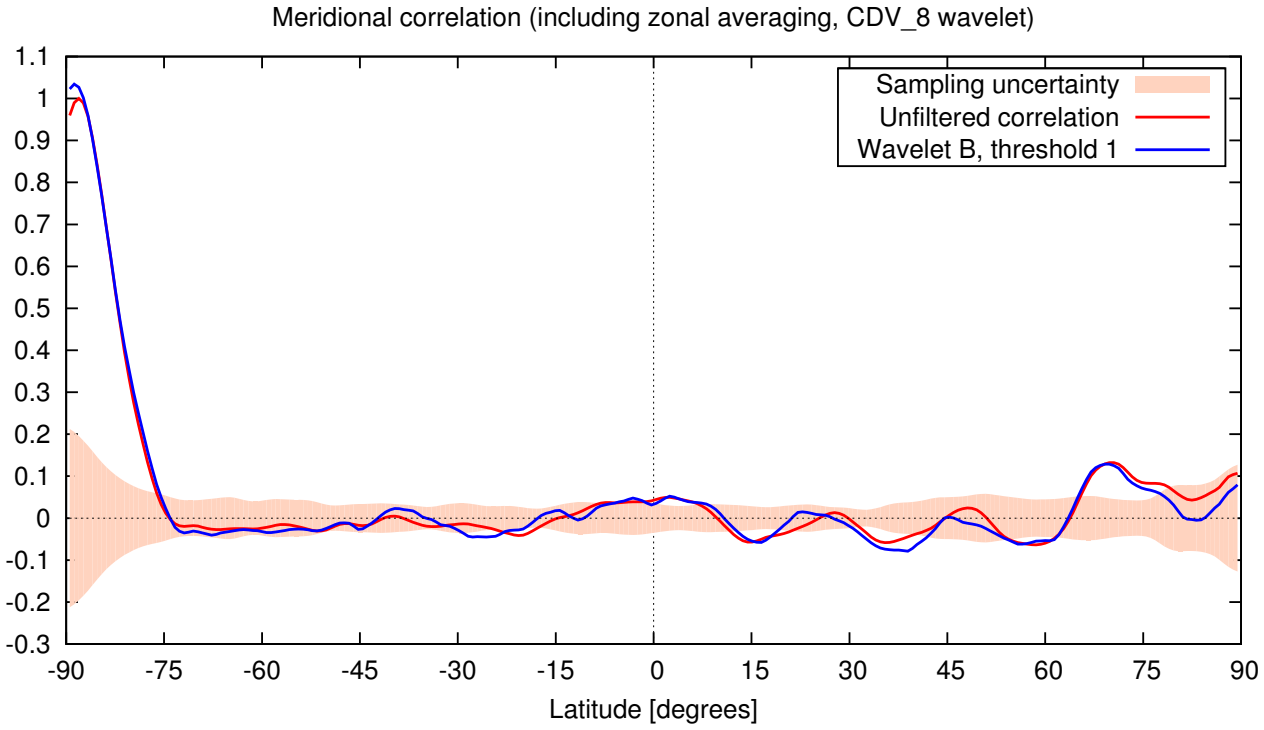


Figure 8: Filtering of meridional (inhomogeneous) wavelet covariances by hard thresholding. The red curve shows the sample correlation with 88°S, the blue curve the corresponding wavelet covariance after application of the threshold $|t| > 1$, $\chi^2/d.o.f. > 1$. The sampling error (shaded area, light red) clearly exhibits a strong increase towards the poles where zonal averaging is ineffective.

Filtering by statistical coefficient selection can also be performed for inhomogeneous correlations. Figure 8 compares the zonally averaged meridional correlations with 88°S with the hard-thresholded wavelet covariance for the same random simulation as above. Note that the sampling error strongly increases towards the poles where zonal averaging is ineffective, leading to possibly large, undesired correlations that need to be suppressed explicitly.

Although statistical selection of wavelet coefficients has the appealing property of being position- and scale-selective, there are still open questions with the approach described above and which are subject of further research. Firstly, the statistical selection is only justifiable for the elements of **B** but not for its symmetric square root **L**. Nevertheless, in those cases where the sparsity pattern of **L** resembles **B** it turns out that one can take the coefficient selection pattern from **B** and apply it to **L**, leading to very similar correlations. Secondly, removing from **L** large coefficients which are nevertheless considered noise can lead to a significant modification especially of the diagonal (variance) terms. In our application this effect is most pronounced near grid boundaries and the poles. A possible cure is renormalisation but needs further investigation.

1.5 Zonal averaging and factorisation in higher dimensions

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

$$\mathbf{W_{3D}} = \mathbf{W_x W_y W_z} \tag{13}$$

The PSAS formulation of the DWD 3D-Var does not depend on a specific grid. The covariance matrices can

be represented on a different grid than that of the forecast model. For the time being we use zonal averaging in order to increase the effective size of the NMC ensemble. To facilitate this, we chose a Gaussian grid with 512×256 grid-points in the horizontal. Vertical covariances are modelled on 64 pressure levels equidistant in $\log p$.

In higher dimensions $\hat{\mathbf{B}}$ and $\hat{\mathbf{L}}$ remain sparse. As shown in Sections 1.2 and 1.3 the number of coefficients per grid-point is of order 10 in 1 dimension. In 2-d (horizontal or vertical) 30 to 50 coefficients per grid-points are required for 1% accuracy (cf. Section 1.6). In 3-d we expect of order 100 coefficients. Thus the operation count of matrix vector products $\hat{\mathbf{L}}\mathbf{x}$ is comparable to that of the wavelet transforms $\mathbf{W}\mathbf{x}$ and competitive to other covariance modelling methods as for instance the spectral approach. Further savings in storage are possible if symmetry and homogeneity due to zonal averaging is explored.

In higher dimensions it is not feasible to estimate $\mathbf{B}$ in grid-point representation and wavelet transform the whole matrix (Equations (1), (3)) as in the 1-d case. Instead we first transform the forecast differences $\mathbf{u}$ and then directly estimate $\hat{\mathbf{B}}$ in wavelet representation:

$$\hat{\mathbf{B}}_{NMC} = \frac{1}{n} \left(\mathbf{W}^{-\mathbf{1}}\mathbf{u}\right) \left(\mathbf{W}^{-\mathbf{1}}\mathbf{u}\right)^{\mathbf{T}} \tag{14}$$

Only a limited number of coefficients is calculated and tested for relevance, starting at the diagonal and the 'branches' of $\hat{\mathbf{B}}$. Furthermore, zonal averaging and extraction of the symmetric square root of $\hat{\mathbf{B}}$ can be efficiently performed using the Fourier transform in zonal direction. In detail the procedure is:

1. Estimate $\tilde{\mathbf{B}}$ by the NMC method, using forecast differences which are Fourier transformed in the zonal and wavelet transformed in the remaining directions:

$$\tilde{\mathbf{B}} = \frac{1}{n} \left(\mathbf{W}_{\mathbf{yz}}^{-\mathbf{1}}\mathbf{F}_{\mathbf{x}}^{-\mathbf{1}}\mathbf{u}\right) \left(\mathbf{W}_{\mathbf{yz}}^{-\mathbf{1}}\mathbf{F}_{\mathbf{x}}^{-\mathbf{1}}\mathbf{u}\right)^{\mathbf{T}} \tag{15}$$

   If zonal averaging is performed, $\tilde{\mathbf{B}}$ factorises into independent block matrices for each Fourier coefficient.

2. Extract the square root by singular value decomposition:

$$\tilde{\mathbf{B}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\mathbf{T}} \tag{16}$$

   This step is feasible in multiple dimensions because the rank of the block matrices is limited by the number n of forecast differences.

3. Finally, transform $\tilde{\mathbf{L}}$ to wavelet representation:

$$\hat{\mathbf{L}} = \mathbf{W}_{\mathbf{x}}^{-\mathbf{1}}\mathbf{F}_{\mathbf{x}}\tilde{\mathbf{L}}\mathbf{F}_{\mathbf{x}}^{\mathbf{T}}\mathbf{W}_{\mathbf{x}}^{-\mathbf{T}} \tag{17}$$

Equation (17) is valid only for orthogonal wavelet transforms $\mathbf{W}_{\mathbf{x}}$. This is the only step where orthogonal transforms are required. Besides that, the procedure works with bi-orthogonal transforms as well. The transformation is efficient because the operation count of the Fourier and wavelet transformation is $O(N \log N)$ for homogeneous matrices $\tilde{\mathbf{L}}$, instead of $O(N \log N)^2$ or $O(N^2)$ for general matrices.

1.6 2-dimensional examples

Figure 9 shows examples of two dimensional horizontal and vertical covariance functions, reconstructed from their truncated wavelet expansions. The number of nonzero coefficients per grid-point is 40 and 30, respectively. The difference to the original NMC derived matrices is less than 1%. Inhomogeneous, anisotropic, and non-separable covariances can be represented without any restriction.

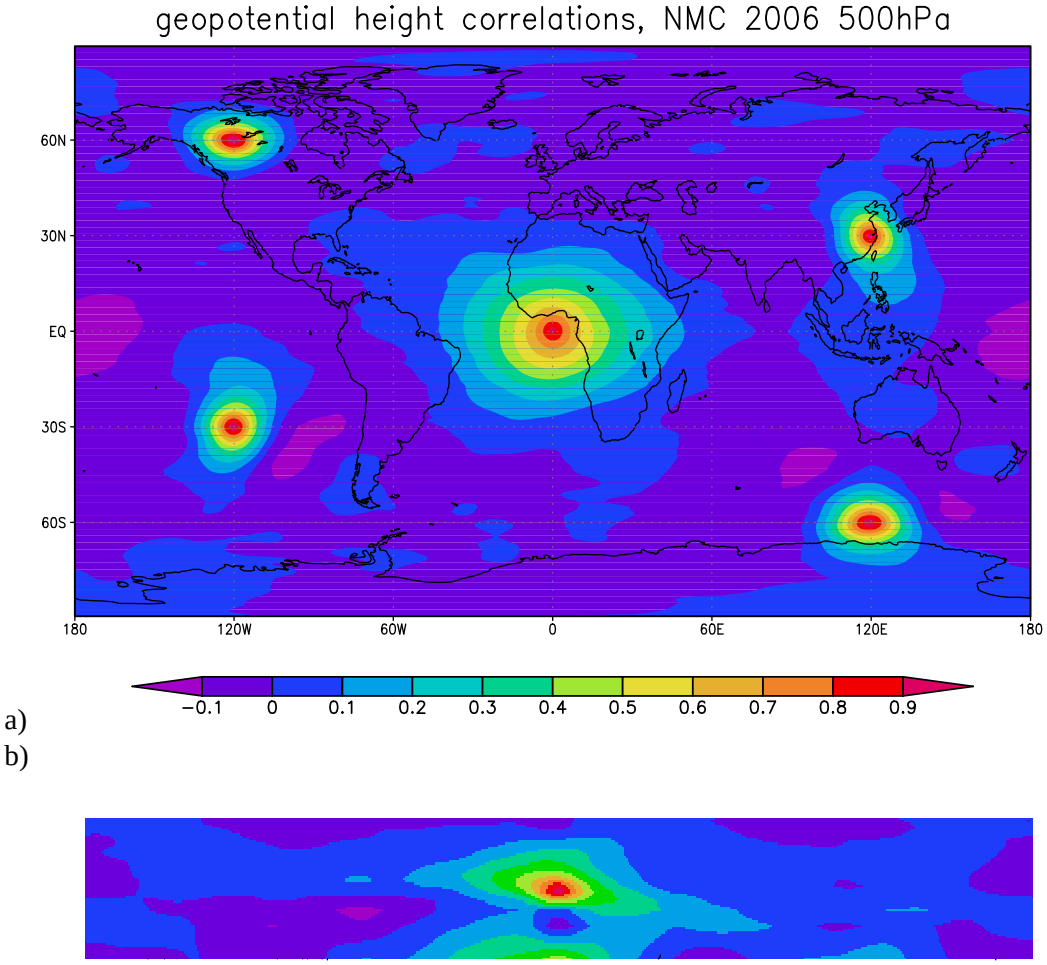


Figure 9: Two-dimensional NMC derived zonally averaged covariance matrices for geopotential height, reconstructed from their truncated wavelet expansions.
a) Horizontal correlations in 500hPa (512 x 256 grid-points).
b) Covariances with a location in 100 hPa height in a vertical slice at the equator (512 x 64 grid-points, vertical axis from 1000 to 10 hPa equidistant in log p).

2 Implementation and experimentation in the DWD 3DVAR

In the PSAS (Physical Space Assimilation System) of the 3DVAR at DWD the forecast error covariance matrix is implemented by an operator representation. For a full 3-dimensional wavelet approach the sequence of operators is:

$$\mathbf{B} = \mathbf{I}\,\mathbf{W}\,\hat{\mathbf{L}}\,\hat{\mathbf{L}}^{\mathbf{T}}\,\mathbf{W}^{\mathbf{T}}\,\mathbf{I}^{\mathbf{T}} \tag{18}$$

with	$\hat{\mathbf{L}}$	:	multivariate square root of $\hat{\mathbf{B}}$	(Eq. (6))
	$\mathbf{W}$	:	3D wavelet transform	(Eq. (13))
	$\mathbf{I}$	:	interpolation operator	

The operator $\mathbf{I}$ interpolates from the Gaussian grid to the location of the observations or to the model grid-points. It also derives temperature t from geopotential height h as well as wind components u and v from stream-function ψ and velocity potential χ by vertical or horizontal differentiation.

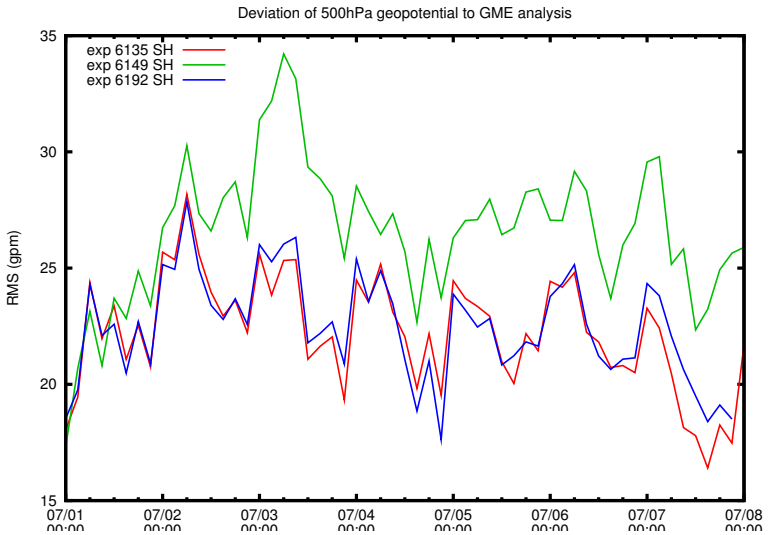


Figure 10: RMS differences of geopotential height analysis using conventional observations only, compared to the operational analysis in the southern extra-tropics: experiment with analytical covariances (red, Exp.6135), NMC derived covariances equivalent to 3 months statistics (green, Exp.6149), and NMC derived covariances equivalent to 12 months statistics and meridional localisation (blue, Exp.6192).

As an intermediate step we have replaced the former explicit separable covariance model by the equivalent wavelet formulation:

$$\mathbf{B} = \mathbf{I}\mathbf{K}\mathbf{W}_\mathbf{v}\hat{\mathbf{L}}_\mathbf{v}\mathbf{W}_\mathbf{h}\hat{\mathbf{L}}_\mathbf{h}\hat{\mathbf{L}}_\mathbf{h}^\mathbf{T}\mathbf{W}_\mathbf{h}^\mathbf{T}\hat{\mathbf{L}}_\mathbf{v}^\mathbf{T}\mathbf{W}_\mathbf{v}^\mathbf{T}\mathbf{K}^\mathbf{T}\mathbf{I}^\mathbf{T} \tag{19}$$

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

Wavelet representations of the former analytic formulation (isotropic covariances based on a Bessel function expansion or on compactly supported covariance functions from Gaspari and Cohn (1999)) have been derived. Analysis differences with respect to the old formulation due to the truncated wavelet representation are negligible (of order 1% of typical analysis increments).

As the covariance matrix coefficients are directly derived from the NMC statistics without applying a parameterised model, a large number of forecast differences is required to minimise statistical noise. In order to study the effect of limited ensemble size, random forecast differences were generated with the statistical properties of the analytic forecast error covariances. Then, wavelet representations were derived from these artificial forecast differences by the NMC method for different ensemble sizes.

Covariance matrices derived from NMC ensembles of limited size lead to spurious analysis increments especially in data sparse regions. Assimilation experiments have been conducted using conventional observations only. Figure 10 shows RMS differences of geopotential height analyses compared to the operational analyses. Covariances derived from 3 months of forecast differences degrade the analyses significantly (Exp.6149). Analysis differences are especially large in the vicinity of the South pole as expected from the discussions in Section 1.4, Figure 8. A significant reduction of noise is achieved (Exp.6192) if covariance matrices are derived from one year of forecast differences and a weak meridional localisation is applied, to suppress correlations between the North and the South pole.

As a next step of the development, covariance matrices derived from true NMC forecast differences of a one year period will be used. We also investigate application of the methods advocated in Section 1.4 to be able to

use smaller ensemble sizes and to proceed towards the complete 3-dimensional implementation.

3 Prospects for flow dependent covariance modelling

The method presented here cannot be applied without modifications to analysis ensembles of typical size $n < 100$ because statistical noise is not eliminated efficiently. There are two possible approaches to apply the method within an Ensemble Kalman Filter framework. The aim is to blend the information derived from a large ensemble of NMC forecast differences or EnKF members taken from a long period with that from a small ensemble representative for the time of the analysis:

1. Add free parameters (for instance a diagonal matrix $\mathbf{D}$) to the static covariance model (18) and adjust (fit) the free parameters (diagonal coefficients d_{ii}) so that the statistical properties of the EnKF ensemble are met.

$$\mathbf{B} = \mathbf{I}\,\mathbf{W}\,\hat{\mathbf{L}}\,\mathbf{D}\,\hat{\mathbf{L}}^{\mathbf{T}}\,\mathbf{W}^{\mathbf{T}}\,\mathbf{I}^{\mathbf{T}} \tag{20}$$

   This approach has been applied to the ECMWF analysis ensemble. Preliminary results indicate that adjusting just diagonal coefficients d_{ii} is not sufficient to represent the statistical properties of the ensemble (variances and length scales) reasonably well.

2. Estimate coefficients of $\hat{\mathbf{B}}$ or $\hat{\mathbf{L}}$ from both the large ensemble and the small analysis ensemble and estimate an optimal value from the respective values and spreads. This approach is attractive because

   (a) The number of coefficients which may be relevant is limited and the locations of these coefficients are known. They are mainly the same as those selected for the static covariance matrix as the sparsity pattern of wavelet transformed covariance matrices is universal and not specific to the situation.

   (b) This approach basically is a localisation in wavelet representation which corresponds to both a localisation in spectral and in grid-point representation (cf. the contributions of Mark Buehner and Olivier Pannekoucke & Loïk Berre within this volume).

References

Beylkin, G., R. Coifman, and V. Rokhlin (1991). Fast wavelet transforms and numerical algorithms I, *Communications on Pure and Applied Mathematics XLIV*, 141–183,

Cai, T.T. (2002). On Block Thresholding in Wavelet Regression: Adaptivity, Block Size, and Threshold Level, *Statistica Sinica 12*, 1241–1273.

Deckmyn, A. and L. Berre (2005). A wavelet approach to representing background error covariances in a limited-area model, *Mon. Wea. Rev. 133*, 1279–1294.

Daubechies, I. (1992). Orthonormal Bases of Wavelets and Multiresolution Analysis, *Ten Lectures on Wavelets, Chapter 6*, SIAM.

Fisher, M. and E. Andersson (2001). Developments in 4D-Var and Kalman filtering, *ECMWF Tech. Memo. 347*

Gaspari, G. and S.E. Cohn (1999). Construction of correlation functions in two and three dimensions, *Quart. J. Roy. Meteor. Soc. 125*, 723–757.

Majewski, D., D. Liermann, P. Prohl, B. Ritter, M. Buchhold, T. Hanisch, G. Paul, W. Wergen, and J. Baumgardner (2002). The operational global icosahedral-hexagonal gridpoint model GME: Description and high-resolution tests, *Mon. Wea. Rev. 130*, 319–338.

Nychka, D., C. Wikle, and J. A. Royle (2002). Multiresolution models for nonstationary spatial covariance-functions, *Stat. Modell. 2*, 315–332.

Mallat, S., G. Papanicolaou, and Z.F. Zhang (1998). Adaptive covariance estimation of locally stationary processes, *Annals of Statistics 26*, 1–47.

Pannekoucke, O., L. Berre and G. Desroziers (2007). Filtering properties of wavelets for local background-error correlations, *Quart. J. Roy. Meteor. Soc. 133*, 363-397

Rhodin, A. and H. Anlauf (2007). Sparse Representation of Forecast Error Covariance Matrices by a Truncated Wavelet Expansion: 1-D Feasibility Study, *in preparation*.