1.2 Wavelet Analysis of Geoscientific Data

T. Gerstner, H.–P. Helfrich, A. Kunoth May 2001

Abstract

Wavelets have shown to be an indispensable tool for a scale–sensitive representation and analysis of spatial or temporal data. In this chapter we illustrate the basic mathematical background for wavelets as well as fast algorithms for concrete multivariate wavelet constructions on bounded domains. Finally, we will show an application of wavelets to analysis and compression of precipitation data.

1.2.1 Introduction

Ever since the first application of wavelet transforms in geophysics in [MAFG] for the analysis of seismic signals, there has been a vast amount of wavelet–related research, both on the theoretical as well as on the practical level. An extensive description of different activities in geophysics can be found in [KF]. The numerous applications cover not only the analysis and compression of all variants of signals and images but also the computation of solutions of partial differential equations for process simulation. These very different types of applications already indicate one of the special features of wavelets: the notion of *wavelets* does not describe a particular fixed object but rather stands for variants of a concept which is adapted according to the application at hand.

The underlying methodology may be sketched as follows. Many physical processes, signals or images, large amounts of data, or solutions of partial differential equations are in fact objects (such as functions) living on different *scales*. With these scales one may associate *meshes* of different *diameters* or mesh *sizes*. To describe the objects mathematically, one needs appropriate representations of the object. In order to obtain such representations, most applications ask for

(A) scalability;

(B) spatial locality.

Requirement (A) means that the representation is able to capture features on a coarse mesh as *efficiently* as on a fine mesh. Thus, the description of the object should in principle allow for a *multiscale representation* of the object.

Requirement (B) indicates that *local* features of an object should only be realized in a 'spatial neighborhood'. (A) and (B) together would allow for an *efficient representation* for the object. In addition, it ideally requires also only *minimal storage*.

Different mathematical descriptions of an object may satisfy one or the other of the two requirements. The concept of wavelets is that they provide in principle a methodology satisfying (A) *and* (B).

In order to specify these remarks, we mention two typical cases.

Case 1: Fourier Analysis of Data

Modern experimental methods and measuring devices deliver a huge amount of data. In order to handle the data properly, an abstract data model is required. One purpose of the model is to make the data be retrievable in an efficient manner. One idea to do that is to assume a *hierarchy* and to develop a *multiscale representation* of the data. The classical procedure is to make a *frequency analysis* of the data using the Fourier Theorem. In this way, one can decompose a (2π–periodic) function f with limited energy into an integral (or a sum) of different signals, each of which gives a precisely defined frequency, that is, f is represented by a decomposition

$$f(t) = \frac{1}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} c_j e^{ijt}, \tag{1.2.1}$$

where the coefficients c_j are determined by

$$c_j = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(t)\, e^{-ijt}\, dt. \tag{1.2.2}$$

For example, in the case of temperature data which is sufficiently dense, the Fourier model can very precisely exhibit seasonal and daily variations. All waves ω_j that generate the signal are obtained by scaling the base wave $\omega(t) = \exp(it)$, that is, one defines $\omega_j(t) := \omega(jt)$. However, in view of the above requirements, the Fourier model has very strong limitations. In the case of temperature data, the frequency space gives a clear picture of seasonal variations, but it does not show sudden climatic changes as they may be caused by the greenhouse effect. The reason is that the frequency synthesis is not localizable: each frequency spreads over all of the time axis. That is, the Fourier model satisfies requirement (A) but not (B). Apparently the reason is that the base wave ω is a function having *global support*, i.e., living on all of the (time) axis.

To overcome these restrictions, there have been various attempts to 'localize' the Fourier transform, leading to what is known as the *windowed Fourier transform.*

Case 2: Finite Elements

In the numerical simulation of physical processes, one often needs to compute an approximate solution of a partial differential equation. To this end, one can prescribe to the underlying domain where the process is to be modelled a mesh of a *uniform* mesh size. Solving the equation then corresponds to approximating the solution in terms of local piecewise polynomials, commonly called Finite Elements, which live on this mesh. Since these functions are usually *locally supported*, an expansion of the solution in terms of these functions would satisfy requirement (B). Of course, the finer the mesh is, the better one expects the solution to be. If one only works with one grid of the same mesh size, this would mean that requirement (A) is not met. Working on a very fine grid, one would then need a lot of time to solve the resulting systems of equations, and also much storage. This can be remedied by allowing for meshes of different sizes and Finite Elements living on them to represent the solution, thus, leading to a *hierachical representation* of the solution.

Now the wavelet concept provides a methodology to overcome the prescribed limitations and offers a systematic way for both a practical as well as a powerful analytical tool. In the following description of the main features, we focus first on the general principles, independent of a certain application. We will then describe later additional properties that one can use for particular applications.

The concept is perhaps most conveniently described in terms of the multiscale decomposition of a function $f : \mathbb{R} \to \mathbb{R}$ (which can, for instance, be a signal). The general idea is to split f into a sum of functions that are obtained by scaling and shifting one basis function ψ. That is, in the abbreviation

$$\psi_{j,k}(t) := 2^{j/2}\,\psi(2^j t - k), \qquad j,k \in \mathbb{Z}, \tag{1.2.3}$$

the index j stands for the *level of resolution* and k refers to the *location* of the function. In these terms, the function f is represented by a multiresolution decomposition

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_k^j \psi_{j,k}(t) \tag{1.2.4}$$

where the first sum is taken over the scales from the coarse to the finer scales and the second sum is taken over all locations. As in the case of a Fourier series all functions $\psi_{j,k}$ used in the representation are derived from one function, the *mother wavelet* ψ. In order to achieve localization in time, besides the decomposition given by the outer sum in functions with varying oscillation speed, the inner sum gives a decomposition with different localizations. Thus, one can view j as the scaling parameter or the stage which replaces the frequency or the frequency band in the Fourier synthesis and k as the localization parameter. The waves $\omega_j(t) = \exp(ijt)$ in the Fourier expansion (1.2.1) are here replaced by 'generalized' waves $\psi_{j,k}$ which have an *additional* index k representing the *location*. It is this analogy which gives the collection of all $\psi_{j,k}$ the terminology *wavelets*.

The important property of wavelets that is used in different contexts is that not only an expansion (1.2.4) exists but that in addition that a norm involving the *expansion coefficients* controls the *function norm*. In mathematical terms, this means that the collection of $\psi_{j,k}$ is a *Riesz basis*, that is, each $f \in L_2(\mathbb{R})$ (the space of square Lebesgue integrable functions, or energy space) has a *unique* decomposition (1.2.4) and

$$\|f\|^2_{L_2(\mathbb{R})} \sim \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |d^j_k|^2 \tag{1.2.5}$$

holds independent of f. Here the relation $a \sim b$ means that a can be estimated from above and below by a constant multiple of b. Recall that the Fourier expansion of f also satisfies such a *stability estimate* with the equality sign. The stability estimate implies that small disturbances of the function f result in small deviations in the coefficients and vice versa.

The mother wavelet should have properties that lead to a meaningful interpretation of the decomposition. A proper choice makes it possible to split the signal recursively into a trend and a fluctuation. At each stage of the analysis, one may view the actual trend as the smooth part of the time series and the residual or fluctuations as the rough part. One obtains a decomposition ranging from very smooth to very rough components. The most important cases are the ones where the location effect can be most effective. In mathematical terms, this means that ψ has *compact support*, that is, ψ vanishes outside a fixed interval. Then each of the basis functions $\psi_{j,k}$ in the decomposition (1.2.3) has compact support which is centered around $2^j k$ with width proportional to 2^{-j}. That is, for high values of j

corresponding to a high scale, the functions $\psi_{j,k}$ represent high–frequency parts of the functions f. The easiest and for many years known example of wavelets are based on choosing ψ as the Haar wavelet, consisting of piecewise constants,

$$\psi(t) = \begin{cases} 1, & t \in [0, 1/2), \\ -1, & t \in [1/2, 1). \end{cases} \tag{1.2.6}$$

However, since the Haar wavelets lack smoothness it does not very well represent smooth functions. The triumphant march of wavelets through all sorts of applications then began in fact with the construction of smoother variants with compact support which are also L_2–orthogonal (see (1.2.10) below) in [Dau]. In these cases, ψ is no longer a piecewise polynomial and even cannot be written explicitly as a function in closed form (which is actually never needed for analysis or computation purposes.) For a compactly supported ψ, the above decomposition (1.2.4) provides not only a decomposition into different frequencies but in addition a decomposition into different local contributions of the same oscillation speed. This property of the set $\{\psi_{j,k}\}$ is commonly called *time–frequency localization.*

The above concept is by no means restricted to time series. Similar considerations can be used to analyze spatial data. Consider elevation data that come from a measurement device with a very fine resolution. The data are influenced by the roughness of the ground and disturbances caused by the vegetation. It is not possible to derive form parameters such as slope and curvature without suitable generalizations that depend on the cartographic scale used in the representation of the data. Since it is not desirable to fix the scale once for all, we have to look for a suitable scaling procedure where according to the actual scale small disturbances are smoothed out.

In this section we will show by several model problems how wavelets can be used to analyse geophysical data. We will explain some simple algorithms that can be efficiently used for different classes of problems. But first we have to collect a number of definitions and theoretic results on wavelets as an analysis tool.

1.2.2 More on the Theory of Wavelets

Like in the Fourier case, there exists both a discrete and a continuous wavelet transform, depending on the type of application.

In order to describe the continuous wavelet transform first, consider for a parameter $a \neq 0$, $a \in \mathbb{R}$, the family of functions

$$\{\psi_a : \ a \neq 0, \ a \in \mathbb{R}\}$$

which are generated by dilation of ψ as

$$\psi_a(x) := \frac{1}{\sqrt{|a|}} \, \psi\left(\frac{x}{a}\right).$$

For $|a| > 1$, the function is expanded, for $|a| < 1$ it is squeezed. Thus, a plays the role of the (inverse) frequency $1/j$ in (1.2.1) and can therefore be viewed as a *scale*. In addition, one introduces a second parameter b denoting the *location* by which ψ_a can be translated. Thus, one considers the family of functions

$$\{\psi_{a,b} : \ a \neq 0, \ a, b \in \mathbb{R}\}$$

where

$$\psi_{a,b}(x) := \frac{1}{\sqrt{|a|}} \, \psi\left(\frac{x-b}{a}\right).$$

The factor $|a|^{-1/2}$ assures that all functions have the same norm. Now the *Continuous Wavelet Transform* of a function f is an integral transformation of the form

$$(W_\psi f)(a,b) = \int_{\mathbb{R}} f(x) \, \psi_{a,b}(x) \, dx, \tag{1.2.7}$$

in terms of the scale a and the location b. It is important to note that for $f \in L_2$ and appropriately chosen ψ, the continuous wavelet transform $(W_\psi f)$ is invertible. We later describe the requirements on ψ. The *inverse continuous wavelet transform* can be determined to be

$$f(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} (W_\psi f)(a,b) \psi_{a,b}(x) \ a^{-2} \, da \, db. \tag{1.2.8}$$

The terminology *continuous* comes from the fact that a and b are real numbers. Applying the continuous wavelet transform to a function f, one obtains for each pair of parameters a and b different information on f; the information on the scale being represented by a and different values of b measuring the signal at different locations at this scale.

By the wavelet transform (1.2.7) the function f is analyzed, whereas by the inverse wavelet transform (1.2.8) the function is synthesized from its wavelet transform. Consider the case of the Haar wavelet (1.2.6). For fixed a

and b, the right hand side of (1.2.7) gives the mean deviation of the function f from its mean value, that is the fluctuation of the function, in the interval $[b, b+a]$. The reconstruction by the inverse wavelet transformation may be done by taking the limit of a sum with the fluctuations as coefficients.

Of course, different choices of the mother wavelet ψ may well influence the outcome of an analysis. It is for this flexibility that we spoke at the beginning of this section of a methodology rather than a fixed method. In order to guarantee requirement (B), locality, a good choice of ψ is a function with compact support or at least sufficiently fast decay. In addition, ψ should have zero mean or, more generally,

$$\int_{\mathbb{R}} t^r\, \psi(t)\, dt = 0, \qquad r = 0, \ldots, d-1, \tag{1.2.9}$$

where d is a natural number. In this case, one says that ψ has *moment conditions up to order* d. The moment conditions in fact require a certain oscillatory behaviour of ψ, thus, the term 'wavelet' for describing a damped wave. The amount of moment conditions plays an important role in the analysis. Indeed, the conditions (1.2.9) imply that ψ (and also all its scaled and shifted versions $\psi_{a,b}$) are *orthogonal* to all polynomials of degree $d-1$. This implies that for sufficiently smooth f which is evaluated in a Taylor series, the continuous wavelet transform only depends on the d-th remainder of the Taylor series. The quantity $(W_\psi f)(a,b)$ gives a measure of the oszillatory behavior of the function f in a vicinity of b whose size depends on a. Thus, for $a \to 0$, one has (up to a constant factor) $(W_\psi f)(a,b) \to f^{(d)}(b)$. Moreover, higher moment conditions of ψ imply a faster decay rate of $(W_\psi f)(a,b)$ as $|a| \to \infty$. This explains also why the Haar wavelet which only has moment conditions of order 1, i.e., only the mean value vanishes, is often not the best choice.

Like in the Fourier case, the continuous integral transform is a useful tool for theoretical purposes. For practically analyzing signals, one resorts to the *Discrete Wavelet Transform* that corresponds to the transform (1.2.7) for discrete values of a and b. In fact, it can be viewed as choosing for $j \in \mathbb{Z}, k \in \mathbb{Z}$ the parameters as $a = 2^{-j}$, $b = ak$ and computing with the notation from (1.2.3) the expansion (1.2.4),

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_k^j \psi_{j,k}(t).$$

For many applications the most interesting case is when the coefficients can be determined as

$$d_k^j = \int_{\mathbb{R}} f(t)\psi_{j,k}(t)\, dt. \qquad (1.2.10)$$

This corresponds to L_2–*orthogonal* or shortly *orthogonal wavelets* where the collection $\{\psi_{j,k}\}$ satisfies the orthogonality conditions

$$\int_{\mathbb{R}} \psi_{j',k'}(t)\, \psi_{j,k}(t)\, dt = \delta_{jj'}\delta_{kk'}, \qquad j, j', k, k' \in \mathbb{Z}. \qquad (1.2.11)$$

Note that the coefficients determined in (1.2.10) have a form corresponding to that of the Fourier coefficients (1.2.2).

If f is known, one can theoretically determine its wavelet expansion (1.2.4) by computing all the coefficients d_k^j. But even if the $\psi_{j,k}$ are known explicitly, this would be much too expensive if one explicitly computed all integrals (1.2.10). Instead one uses the *Fast Wavelet Transform* which is usually introduced through a Multiresolution Analysis of $L_2(\mathbb{R})$. This is indeed also a way to *construct* wavelets with different properties.

1.2.3 Multiresolution Analysis

The concept of Multiresolution Analysis [Ma] based on Haar wavelets may be introduced using a simple example as follows. Consider a function f which is piecewise constant on each of the intervals $[2^{-1}k,\, 2^{-1}(k+1)]$ for integer k, see Figure 1. The function f can be additively decomposed into a function g which is piecewise constant on the intervals $[k, k+1]$ for integer k and a 'remainder' h. Thus, g would capture the 'coarse' part of f by averaging over two adjacent intervals and h the difference between f and g. Then $f = g + h$ can be considered as a two–scale decomposition of f, where g consists of piecewise constants with respect to the coarse scale $j = 0$ (that is, on intervals $[k, k+1]$) and f on a finer scale $j = 1$ (being piecewise constant on the intervals $[2^{-1}k,\, 2^{-1}(k+1)]$). Both functions f and g can be represented in terms of (scaled) translates of the characteristic function $\phi(t)$ which equals 1 on $[0, 1)$ and 0 elsewhere. That is, defining

$$\phi_{j,k}(t) := 2^{j/2}\, \phi(2^j t - k), \qquad j, k \in \mathbb{Z}, \qquad (1.2.12)$$

we can write $f(t) = \sum_{k\in\mathbb{Z}} f_k\phi_{0,k}(t)$ and $g(t) = \sum_{k\in\mathbb{Z}} g_k\phi_{1,k}(t)$. In order to be able to use only one function ϕ to represent functions on different scales, it

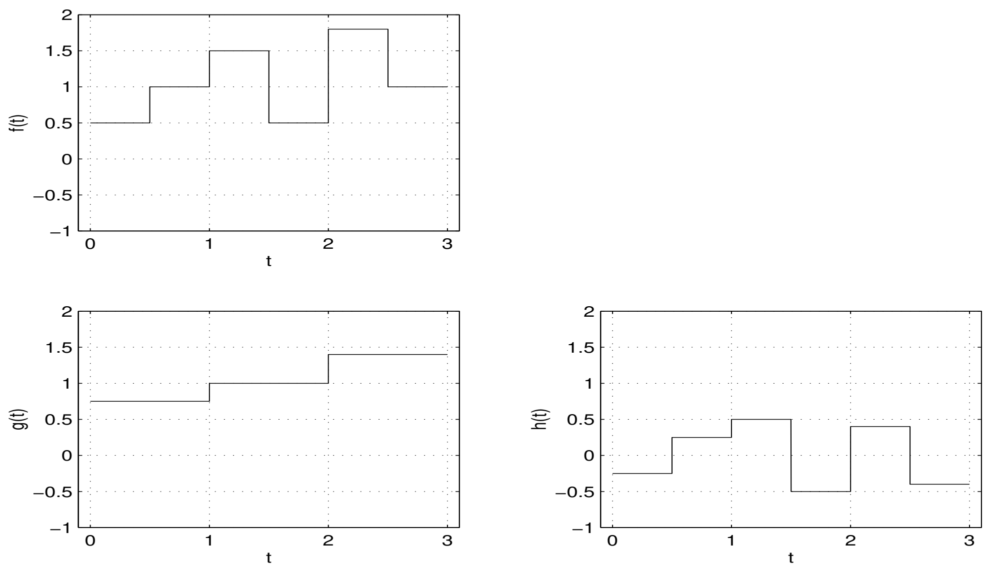


Figure 1: $f \in V_1$ (top figure) is split into $g \in V_0$ (bottom left) and $h \in W_0$ (bottom right).

is important that $\phi(t)$ can be written as a linear combination of translates of scaled versions of ϕ on the *next finer* level. This means that there are some coefficients $\mathbf{a} := \{a_m\}_{m\in\mathbb{Z}}$ such that one has a relation of the form

$$\phi_{j,k}(t) = \sum_{m\in\mathbb{Z}} a_{m-2k}\, \phi_{j+1,m}(t), \qquad j,k \in \mathbb{Z}. \tag{1.2.13}$$

This relation is often denoted as *refinement equation* for ϕ. For ϕ being the characteristic function, the coefficients are just $a_0 = a_1 = \frac{1}{\sqrt{2}}$ and $a_\ell = 0$ for all other indices ℓ. The coefficient sequence $\mathbf{a}$ is called the *mask* of ϕ.

For each j, translates of $\phi_{j,k}$ generate a space V_j,

$$V_j = \left\{ v \in L_2(\mathbb{R}) : \; v(t) = \sum_{k\in\mathbb{Z}} c_k \phi_{j,k}(t) \right\}. \tag{1.2.14}$$

For ϕ chosen as the characteristic function, the space V_j consists of all functions which are piecewise constant on each interval $[2^{-j}k, 2^{-j}(k+1)]$, $k \in \mathbb{Z}$. Because of the refinement relation (1.2.13), the spaces V_j are *nested*, i.e., $V_j \subset V_{j+1}$. As the resolution level j increases, the V_j become larger and in the limit $j \to \infty$ approximate $L_2(\mathbb{R})$. The set of all spaces $\{V_j\}_{j\in\mathbb{Z}}$ which therefore satisfy

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset L_2(\mathbb{R}) \tag{1.2.15}$$

is called a *Multiresolution Analysis* of $L_2(\mathbb{R})$. Since the function ϕ plays a particular role in this concept, it is called the *generator* of the multiresolution analysis. Because ϕ satisfies the refinement relation (1.2.13), it is sometimes also called *refinement* or *scaling function*.

So far the spaces V_j have been introduced from which f and g in Figure 1 come from. Now the question arises how the function h representing the 'detail' $f - g$ can be represented. In terms of spaces, this can be described as follows. Because of the inclusion $V_j \subset V_{j+1}$, there exists an orthogonal complement of V_j in V_{j+1} that we denote by W_j. That is, V_{j+1} can be decomposed into the direct sum

$$V_{j+1} = V_j \oplus W_j, \tag{1.2.16}$$

where V_j contains the 'low' frequencies and W_j the 'details' or differences from one resolution level j to the next finer resolution level $j+1$. Repeating this process, one ends up with an orthogonal decomposition of $L_2(\mathbb{R})$,

$$L_2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j. \tag{1.2.17}$$

Finally, every function $f \in L_2(\mathbb{R})$ can be decomposed according to

$$f(t) = \sum_{j\in\mathbb{Z}} f_j(t), \quad f_j \in W_j, \tag{1.2.18}$$

where the part f_j corresponds to the frequency j. As it turns out, the spaces W_j are spanned by translates of the wavelets $\psi_{j,k}$ defined in (1.2.3), i.e.,

$$W_j = \left\{ w \in L_2(\mathbb{R}) : \; w(t) = \sum_{k\in\mathbb{Z}} d_k \psi_{j,k}(t) \right\}. \tag{1.2.19}$$

Since by (1.2.16) $W_j \subset V_{j+1}$, the wavelets $\psi_{j,k}$ can also be represented by linear combinations of the generator on level $j+1$, i.e., for some coefficients $\mathbf{b} := \{b_m\}_{m\in\mathbb{Z}}$ one has a relation of the form

$$\psi_{j,k}(t) = \sum_{m\in\mathbb{Z}} b_{m-2k}\, \phi_{j+1,m}(t), \qquad j,k \in \mathbb{Z}. \tag{1.2.20}$$

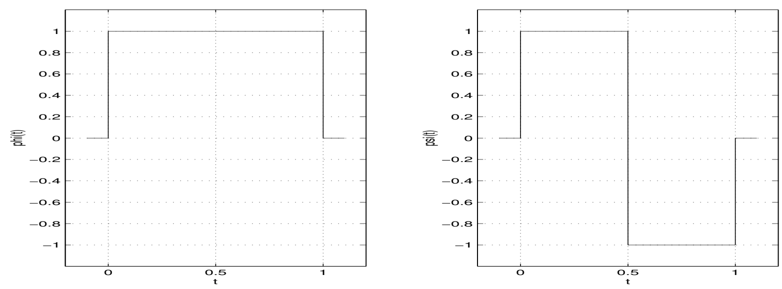


Figure 2: ϕ as characteristic function and ψ Haar wavelet.

The coefficients $\mathbf{b}$ are called the *mask* of the wavelet ψ.

If the scaling functions $\phi_{j,k}$ are orthogonal on each level j, that is,

$$\int_{\mathbb{R}} \phi_{j,k}(t)\, \phi_{j,\ell}(t)\, dt = \delta_{k,\ell}, \quad k, \ell \in \mathbb{Z}, \tag{1.2.21}$$

one can show that the coefficients b_k in (1.2.20) are indeed uniquely determined by

$$b_\ell = (-1)^{\ell-1}\, a_{1-\ell}, \tag{1.2.22}$$

where $\mathbf{a}$ is the sequence in the refinement relation (1.2.13). For ϕ being the characteristic function, one can thus see that the Haar wavelet given in (1.2.6) is the wavelet corresponding to the characteristic function, see Figure 2. Here $b_0 = \frac{1}{\sqrt{2}}$, $b_1 = -\frac{1}{\sqrt{2}}$, and all other mask coefficients vanish. Thus, given $\mathbf{a}$, one can immediately determine $\mathbf{b}$ and gets from this via (1.2.20) a representation of the wavelets $\psi_{j,k}$. Although this is no explicit representation, it is sufficient for computational purposes. In fact, the *Fast Wavelet Transform* which we describe next is only build from the mask coefficients $\mathbf{a}$ and $\mathbf{b}$ that determine $\phi_{j,k}$ and $\psi_{j,k}$.

1.2.4 The Fast Wavelet Transform

Consider a fixed parameter $J > 0$ which will stand for the highest 'desired' level of resolution and let $j = 0$ be the coarsest level. Recalling (1.2.17), the corresponding space V_J can be decomposed as

$$V_J = V_0 \oplus \bigoplus_{j=0}^{J-1} W_j. \tag{1.2.23}$$

That means, each function $v \in V_J$ can be written either in its *single–scale expansion* involving only the generator functions on level J,

$$v(t) = \sum_{k\in\mathbb{Z}} c_k\, \phi_{J,k}(t), \tag{1.2.24}$$

or in its *multi–scale expansion* involving the generators on level 0 and all wavelets on all intermediate resolution levels $0, \dots, J-1$,

$$v(t) = \sum_{k\in\mathbb{Z}} c_k^0\, \phi_{0,k}(t) + \sum_{j=0}^{J-1} \sum_{k\in\mathbb{Z}} d_k^j\, \psi_{j,k}(t). \tag{1.2.25}$$

Thus, in either representation the function v is encoded in the expansion coefficients which we abbreviate as *vectors* $\mathbf{c}$ and $\mathbf{d} = (\mathbf{c}^0, \mathbf{d}^0, \mathbf{d}^1, \dots, \mathbf{d}^{J-1})^T$ of the same length. Both representations convey different information. The coefficients c_k indicate a geometric location of v whereas the coeffcients d_k^j have the character of differences. Usually *all* coefficients c_k are needed to represent v accurately, while many of the d_k^j may in fact be very small. This fact will be used later for data compression by thresholding, see Section 1.2.7. In order to benefit from both representations, one needs a mechanism to transform between them. That is, one looks for a linear transformation

$$\mathbf{T}_J : \mathbf{d} \to \mathbf{c} \tag{1.2.26}$$

and its inverse $\mathbf{T}_J^{-1} : \mathbf{c} \to \mathbf{d}$. This transformation is indeed determined with the aid of the refinement relations (1.2.20) and (1.2.13). In order to explain the main mechanism, it is more convenient to write these relations in a matrix-vector form as follows. (This form will also be used later for describing wavelets on bounded intervals and domains.) Assemble in the (column) vector $\Phi_j(t)$ all functions $\phi_{j,k}(t)$ in their natural order,

$$\Phi_j(t) = (\dots, \phi_{j,k-1}(t), \phi_{j,k}(t), \phi_{j,k+1}(t), \dots)^T,$$

and correspondingly for $\Psi_j(t)$. Then the equations (1.2.13) and (1.2.20) can be written in abbreviated form as

$$\Phi_j(t) = \mathbf{M}_{j,0}^T\, \Phi_{j+1}(t) \tag{1.2.27}$$

and

$$\Psi_j(t) = \mathbf{M}_{j,1}^T\, \Phi_{j+1}(t). \tag{1.2.28}$$

Here the matrices $\mathbf{M}_{j,0}$ and $\mathbf{M}_{j,1}$ consist of the mask coefficients $\mathbf{a}$ and $\mathbf{b}$, that is,

$$(\mathbf{M}_{j,0}^T)_{k,m} = a_{m-2k}, \qquad (\mathbf{M}_{j,1}^T)_{k,m} = b_{m-2k}. \tag{1.2.29}$$

To derive $\mathbf{T}_J$, one starts from the multiscale representation (1.2.25) which is abbreviated in matrix-vector form as

$$v = (\mathbf{c}^0)^T\Phi_0 + (\mathbf{d}^0)^T\Psi_0 + (\mathbf{d}^1)^T\Psi_1 + \cdots + (\mathbf{d}^{J-1})^T\Psi_{J-1}$$

and inserts the two relations (1.2.28) and (1.2.27) until the single–scale representation (1.2.24) with respect to level J,

$$v = \mathbf{c}^T\Phi_J,$$

is reached. Then one can derive that $\mathbf{T}_J$ has the representation

$$\mathbf{T}_J = \mathbf{T}_{J,J-1}\cdots\mathbf{T}_{J,j_0}, \tag{1.2.30}$$

where each factor has the form

$$\mathbf{T}_{J,j} := \begin{pmatrix} \mathbf{M}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}. \tag{1.2.31}$$

Here for each j $\mathbf{M}_j = (\mathbf{M}_{j,0}, \mathbf{M}_{j,1})$ is a 'square' matrix consisting of the two 'halfs' from the relations (1.2.27) and (1.2.28) and $\mathbf{I}$ is the identity. Schematically $\mathbf{T}_J$ can be visualized as a pyramid scheme, see e.g. [D]. The transformation $\mathbf{T}_J$ is called the (discrete) *Wavelet Transform*. If each of the matrices $\mathbf{M}_{j,0}$, $\mathbf{M}_{j,1}$ has only a fixed finite number of entries in each row and column independent of j (which is the case when the masks $\mathbf{a}$ and $\mathbf{b}$ only involve a *finite* number of coefficients, like in the Haar basis case), it can be shown that applying the transform (1.2.30) in the above product form is an $\mathcal{O}(N)$ algorithm, where N is the number of unknowns on level J, i.e., the length of $\mathbf{c}$ or $\mathbf{d}$. Thus, applying the transform $\mathbf{T}_J$ needs even less operations than Fast Fourier Transform, justifying therefore the term *Fast Wavelet Transform.*

In the case of orthogonal wavelets, one has that $\mathbf{T}_J^{-1} = \mathbf{T}_J^T$. However, there are applications where one does not need orthogonality. In these cases, the inverse of $\mathbf{T}_J$ will also be derived as a pyramid scheme so that also the application of $\mathbf{T}_J^{-1}$ is a fast $\mathcal{O}(N)$ algorithm.

1.2.5 Different Types of Wavelets on $\mathbb{R}$ and $\mathbb{R}^n$

For many applications the orthogonal wavelets satisfying (1.2.11) with compact support constructed in [Dau] are the wavelets of choice. In these cases, the Riesz basis property (1.2.5) is satisfied with *equality* sign. One can show that wavelets with such properties cannot be made symmetric. Another reason to relax these conditions is that one sometimes asks for other properties of the wavelets. In principle, the wavelet methodology provides a mechanism to realize different species. In this subsection, a number of them which live on $\mathbb{R}$ (and can therefore be constructed by Fourier techniques) are listed.

From the application point of view, the property that is usually important is the requirement for the wavelets to have compact support, entailing that the masks $\mathbf{a}$ and $\mathbf{b}$ only have finitely many coefficients and that the correspondingly built Fast Wavelet Transform can be applied in $\mathcal{O}(N)$ operations, as described above.

Particular interest hold these types of wavelets that are constructed from generators which are *splines*, that is, piecewise polynomials. However, the different resulting *spline–wavelets* are then either globally supported or they are no longer orthogonal in the sense of (1.2.11). Relaxing this condition and only requiring orthogonality between different levels leads to *prewavelets*. These and other types have been intensively studied in [Ch]. Another class that proves at times useful are the *interpolatory wavelets* which interpolate prescribed function values. In [Be] these interpolating wavelets are used in an adaptive collocation method for the numerical solution of elliptic boundary value problems. The nice fact that these wavelets interpolate is paid for by the fact that they do not satisfy the Riesz basis property (1.2.5).

Another class that seems to be a very good compromise of different features are the *biorthogonal spline–wavelets* derived in [CDF]. In this setting, one works in fact with *two* multiresolution analyses $\{V_j\}$, $\{\tilde{V}_j\}$ of $L_2(\mathbb{R})$ (called the *primal* and *dual multiresolution analysis*, respectively) and has therefore also the flexibility of two generators ϕ for $\{V_j\}$ and $\tilde{\phi}$ for $\{\tilde{V}_j\}$. Consequently, one also obtains two mother wavelets ψ and $\tilde{\psi}$ whose dilates and translates satisfy the *biorthogonality conditions*

$$\int_{\mathbb{R}} \psi_{j',k'}(t)\, \tilde{\psi}_{j,k}(t)\, dt = \delta_{jj'}\delta_{kk'}, \qquad j, j', k, k' \in \mathbb{Z}. \tag{1.2.32}$$

The interesting point is that one can take as ϕ piecewise polynomials (specifically B–Splines or Finite Elements). The corresponding wavelets $\psi_{j,k}$ are then just linear combinations of these piecewise polynomials and are often more useful to compute with, for instance, in the numerical analysis for partial differential equations, see e.g. [DKU1]. The dual multiresolution analysis is in fact only used for analysis purposes, e.g., to determine whether the Riesz basis property (1.2.5) holds. (This in turn is important for questions like preconditioning when it comes to the efficient solution of linear systems.)

Another generalization are many variations of *multiwavelets*. There one works with instead of just one generator ϕ with a finite number of generators $\phi_1, \ldots, \phi_m$ to generate multiresolution analyses.

There are by now also a number of techniques to transform wavelets with specific properties into other ones with other properties. A special case of the general principle called *stable completion* in [CDP] is referred to as *lifting scheme* in [Sw]. This principle is used in Section 1.2.8.

So far all these investigations are for univariate functions on all of $\mathbb{R}$. For more than one dimension, it is more difficult to construct wavelets, see e.g. [LMR]. For this reason, one often just works with *tensor products* of the univariate ones to generate *multivariate wavelets*.

1.2.6 Wavelets on Bounded Domains and Manifolds

Essentially all the previous constructions of wavelets with different properties are based on the massive use of Fourier techniques. However, as soon as one needs for any application (like the numerical solution of partial differential equations) wavelets on a finite interval, a bounded domain or a manifold in $\mathbb{R}^n$, this is no longer possible. On the other hand, there are a number of very favorable properties of wavelets (among them the Riesz basis expansion into functions with compact support) that one would like to use for the discretization of partial differential or integral equations that typically live on bounded domains. A first idea to use periodization is, of course, only applicable for periodic problems.

When constructing wavelets on bounded subsets Ω of $\mathbb{R}^n$, one cannot simply restrict the expansion (1.2.4) to Ω since one would loose the Riesz basis property (1.2.5). All the different constructions of orthogonal and biorthogonal wavelets on the interval therefore essentially use the same idea. Namely, to leave as many as possible of the functions living in the interior of the interval or the domain untouched, and then to build *fixed* linear combinations of all generators overlapping the ends of the interval in such a way that the moment conditions (1.2.9) are still satisfied. For constructing the corresponding

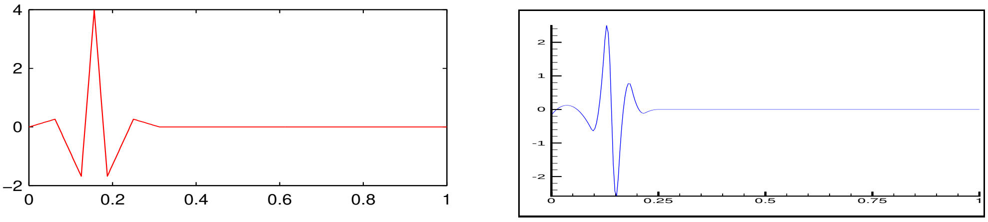


Figure 3: Examples of primal wavelets for $d = 2, 3$ adapted to the interval $[0, 1]$.

wavelets from the generators, one then uses the stable completion techniques [CDP] or the lifting scheme [Sw], see [DKU2] for a general construction of biorthogonal interval–adapted wavelets. Such generators and functions also satisfy refinement relations of the form (1.2.27) and (1.2.28) where Φ_j and Ψ_j are then vectors of finite length and the matrices $\mathbf{M}_{j,0}^T$ and $\mathbf{M}_{j,1}^T$ are up to the boundary effects of the form (1.2.29). Examples of interval–adapted primal wavelets consisting of piecewise linear and piecewise quadratic polynomials are displayed in Figure 3. For domains or manifolds in $\mathbb{R}^n$ which can be represented as the union of disjoint smooth parametric images of a cube, these constructions can then be used to generate wavelets on such domains or manifolds. Two techniques have been used there, either to *compose* the functions by glueing them together, or by using a sophisticated mathematical tool, the characterization of function spaces by Ciesielski and Figiel, see [D] for the main ideas. An example for such a piecewise linear continuous wavelet at the corner of a box (taken from a private communication by Dahmen, Harbrecht, Schneider) is displayed in Figure 4.

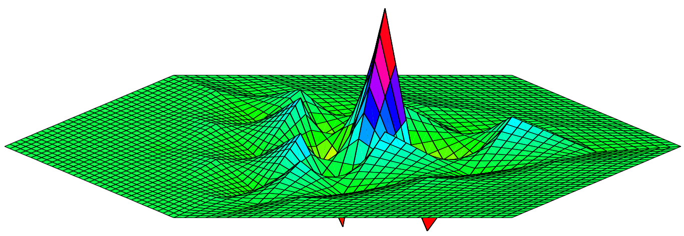


Figure 4: Piecewise linear continuous wavelet at the corner of a box.

1.2.7 Methods for Analyzing Data by Wavelets

In this section we describe some common techniques for analyzing data by wavelets, cf. [O]. Consider discrete values $\mathbf{c} = (c_1, \ldots, c_N)^T$ with $N = 2^J$ of a function that is transformed by a discrete orthogonal wavelet transform of the form

$$\mathbf{d} = \frac{1}{\sqrt{N}} \mathbf{T}_J^T \mathbf{c} \tag{1.2.33}$$

into the vector of wavelet coefficients $\mathbf{d}$ representing the signal, see (1.2.26). For simplicity, we assume that by a suitable preprocessing of the data the first sum on the right hand side of (1.2.25) is not needed in the representation, i.e., $\mathbf{c}^0 \equiv 0$. Furthermore, it has been assumed that the matrix $\mathbf{T}_J$ is orthogonal. Thus, the reconstruction of the function values is achieved by the transformation

$$\mathbf{c} = \sqrt{N}\, \mathbf{T}_J\, \mathbf{d}.$$

Now the scaling of the matrix in (1.2.33) is chosen such that

$$\mathbf{d}^T \mathbf{d} = \frac{1}{N} \sum_{i=1}^{N} c_i^2 \tag{1.2.34}$$

holds, i.e., the right hand side of (1.2.34) should be of the same order as the L_2-norm of the function.

There are several methods for analyzing the signal $\mathbf{c}$ depending on the purpose. As pointed out earlier, many of the coefficients d_k^j are very small and can be neglected without altering the original signal too much.

For the purpose of *compact storage* and *fast transmission* of data, one can suppress all coefficients that are smaller than a given threshold λ. The value of λ controls the amount of compression and the level of detail visible when transforming back the distorted signal. By this way, we get a *filtering* of the data. An application to weather data is given in the next section.

One purpose of filtering is the *recovery* of a function from *noisy*, sampled data. Assume that the data comes from discrete values $f(t_i)$ of an unknown function disturbed by white noise, i.e.,

$$c_i = f(t_i) + \epsilon_i,$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma)$ are statistically independent random variables that are normally distributed with variance σ^2. By (1.2.33) the wavelet coefficients are also normally distributed and independent. We have for $\mathbf{d} = (d_k^j)$

$$d_k^j = \theta_k^j + \frac{1}{\sqrt{N}}\, z_k^j,$$

where the unknown wavelet coefficients θ_k^j come from the function values $f(t_i)$ and $z_k^j = \mathcal{N}(0, \sigma)$ is white noise. In [DJ1] two ways of thresholding are proposed:

1. *Hard thresholding:* All coefficients with $|d_k^j| \leq \lambda$ are set to zero and the other coefficients remain unchanged.

2. *Soft thresholding:* Set

$$\hat{\theta}_k^j = \begin{cases} d_k^j - \lambda, & d_k^j \ > \lambda, \\ 0, & |d_k^j| \ < \lambda, \\ d_k^j + \lambda, & d_k^j \ < -\lambda, \end{cases}$$

   as an estimate $\hat{\theta}_k^j$ for θ_k^j.

A common choice of the threshold is

$$\lambda = \frac{\sigma}{\sqrt{N}},$$

which ensures that the dropped coefficients are of the same order of magnitude as the noise of those coefficients. In [DJ2] it is shown that a more sophisticated choice of the threshold λ leads to a smoothing procedure that in a sense is optimally smoothness–adaptive: it is near–minimax over a whole interval in the Besov scale, which characterizes smoothness properties of a different quality.

The advantage of recovery methods of functions by wavelets is that the estimation procedure can easily be adapted to handle cases where the noise level depends on time or space. This is a consequence of the good localization properties of wavelets The disadvantage of the method is that the variance σ^2 or a good estimate is needed in order to carry out a reconstruction of optimal order. But it is possible to carry over cross–validatory methods to function estimation by wavelets [HT].

Finally, we present an application of the wavelet analysis technique.

1.2.8 Application: Wavelet Analysis of Precipitation Data

In order to improve the prediction of the weather forecast for Germany, inside the global model GME a non–hydrostatic local model LM is considered. The current resolution of the local model LM is 7 km and provides a weather prediction for the next two days. Since the system is chaotic and very unstable, depending on the current weather situation the size of the prediction errors may vary strongly. Therefore it is important to include into a prediction also an estimate for its probability. For this in turn it is important to adequatetely *filter* the data, see [TH] for details. This is done here in such a way that first the given data (which can be viewed as data living on the smallest resolution level) is transformed into multiscale data, living on different levels, by means of the Fast Wavelet Transform defined in (1.2.26), (1.2.30). Recalling that the new data has the character of difference information from one scale to the next, discarding the coefficients that live below a certain threshold provides then the desired filtering. For the wavelet transform, the wavelets constructed in terms of the lifting scheme [Sw] are used [G1].

Figures 5 and 6 show two situations where this filtering has been applied. The original data is visualized in the lower right graphic. The other three graphics each show the filtered data, filtered at different levels. `Level=6` means that wavelet coefficients belonging to resolution level 7 or higher are discarded. Accordingly, as `Level` becomes higher, more and more detail of the data is seen.

One in fact obtains different results when thresholding of the wavelet coefficients in different norms is compared. For another set of data [Gr] we have compared wavelet compression with respect to the L_1–norm (Figure 7) and with respect to the L_∞–norm (Figure 8).

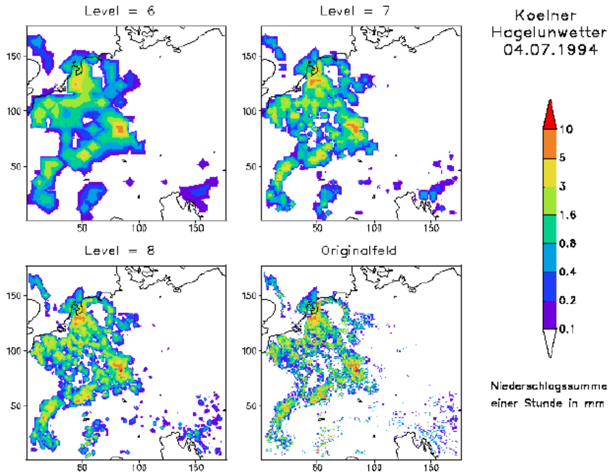


Figure 5: Wavelet analysis of precipitation data “Kölner Hagelunwetter”.

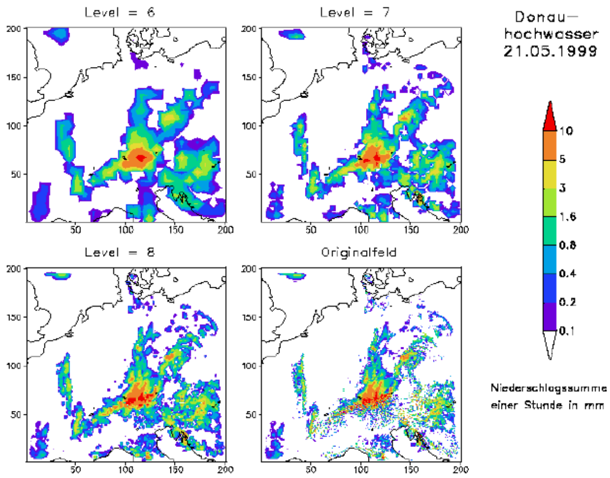
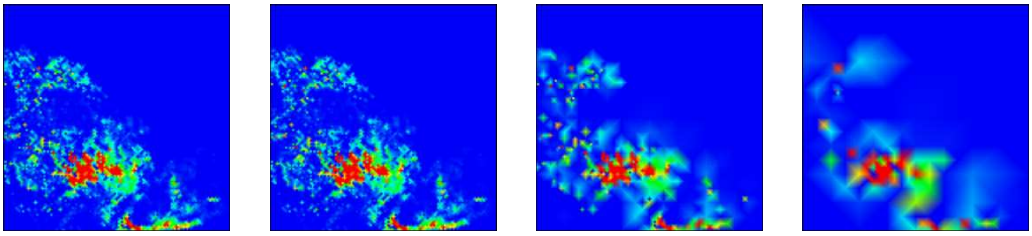
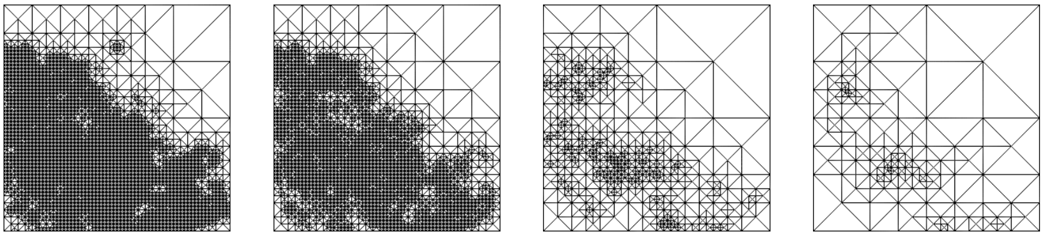


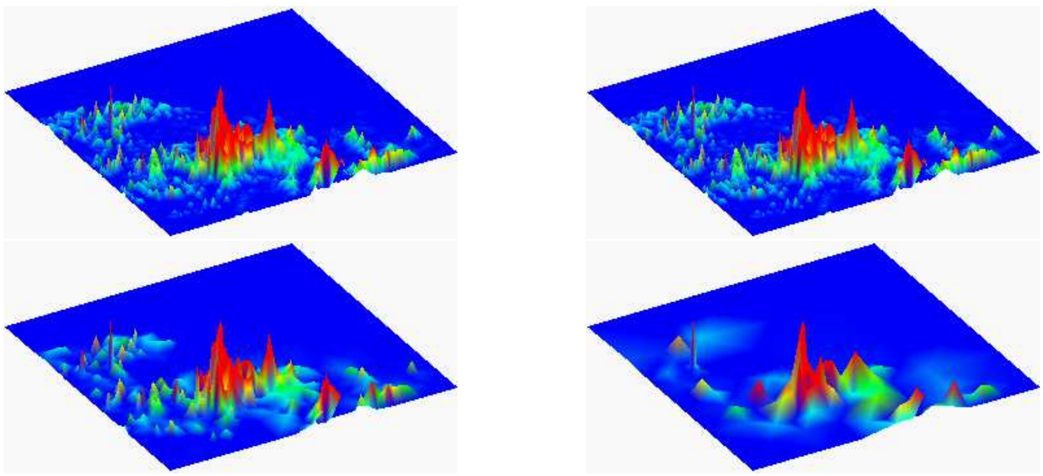
Figure 6: Wavelet analysis of precipitation data “Donauhochwasser”.



Color shaded data

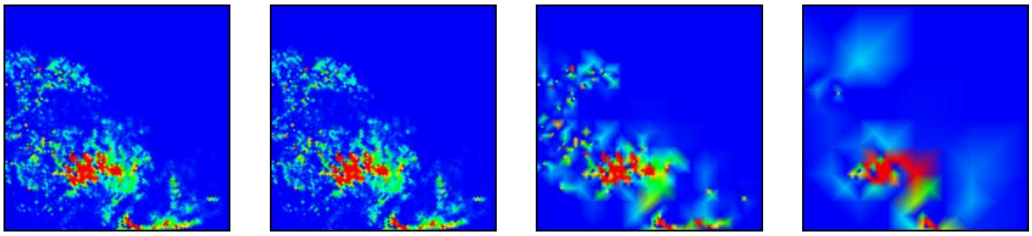


Corresponding grids

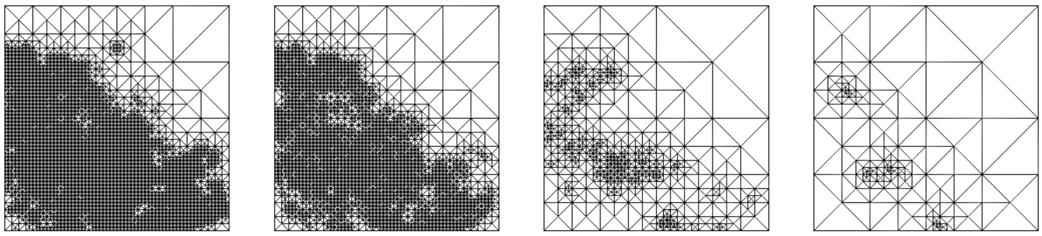


Color shaded 3D view

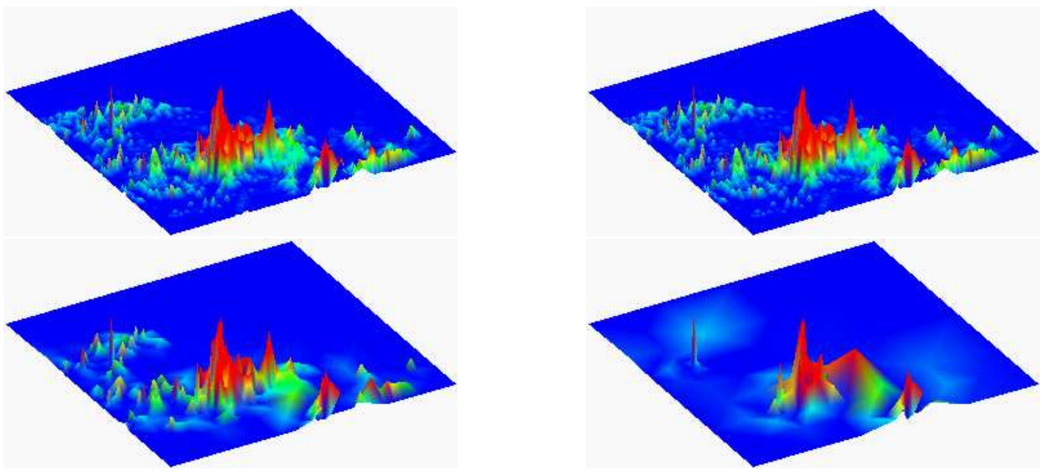
Figure 7: Wavelet analysis of precipitation data compressed in L_1–norm.



Color shaded data



Corresponding grids



Color shaded 3D view

Figure 8: Wavelet analysis of precipitation data compressed in L_∞–norm.

References

[Be] S. Bertoluzza, *An adaptive collocation method based on interpolating wavelets*, in: Multiscale Wavelet Methods for PDEs, W. Dahmen, A. J. Kurdila, P. Oswald (eds.), Academic Press, 1997, 109–136.

[CDP] J.M. Carnicer, W. Dahmen, J.M. Peña, *Local decomposition of refinable spaces*, Appl. Comp. Harm. Anal. 3, 1996, 127–153.

[Ch] C.K. Chui, *An Introduction to Wavelets*, Academic Press, 1992.

[CDF] A. Cohen, I. Daubechies, J.-C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math. 45, 1992, 485–560.

[D] W. Dahmen, *Wavelet and Multiscale Methods for Operator Equations*, Acta Numerica 1997, 55–228.

[DKU1] W. Dahmen, A. Kunoth, K. Urban, *A wavelet Galerkin method for the Stokes problem*, Computing 56, 1996, 259–302.

[DKU2] W. Dahmen, A. Kunoth, K. Urban, *Biorthogonal spline-wavelets on the interval – Stability and moment conditions*, Appl. Comput. Harm. Anal. 6, 1999, 132–196.

[Dau] I. Daubechies, *Orthonormal bases of wavelets with compact support*, Comm. Pure and Appl. Math. 41, 1987, 909-996.

[DJ1] D.L. Donoho, I.M. Johnstone, *Ideal spatial adaption via wavelet shrinkage*, Biometrika, 81, 1994, 425–455.

[DJ2] D.L. Donoho, I.M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Am. Statist. Ass. 90, 1995, 1200–1214.

[G1] T. Gerstner, *Adaptive hierarchical methods for landscape representation and analysis*, in Process Modelling and Landform Evolution, S. Hergarten, H.J. Neugebauer (eds.), Lecture Notes in Earth Sciences 78, Springer, 1999, 75–92.

[G2] T. Gerstner, *Multiresolution visualization and compression of global topographic data*, in revision for GeoInformatica, 2000 (shortened version in Proc. Spatial Data Handling 2000, to appear, also as SFB 256 report 29, Universität Bonn, 1999).

[Gr] P. Gross, Private communication, Meteorologisches Institut, Universität Bonn.

[KF] P. Kumar, E. Foufoula–Georgiou, *Wavelet analysis for geophysical applications*, Reviews of Geophysics 35, 4, 1997, 385–412.

[HT] C.M. Hurvich and C.–L. Tsai, *A crossvalidatory* AIC *for hard wavelet thresholding in spatially adaptive function estimation*, Biometrika, 85,701 – 710, 1998.

[LMR] A. Louis, P. Maass, A. Rieder, *Wavelets. Theory and Applications*, Wiley, 1997.

[Ma] S. Mallat, *Multiresolution approximation and wavelet orthonormal bases of* L^2, Trans. Amer. Math. Soc. 315, 1989, 69–88.

[O] R. T. Ogden, *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser, Boston, 1997.

[Me] Y. Meyer, *Wavelets: Algorithms & Applications*, SIAM, Philadelphia, 1993.

[MAFG] J. Morlet, G. Arens, E. Fourgeau, D. Giard, *Wave propagation and sampling theory, Part 1. Complex signal and scattering in multilayered media*, Geophysics 47, 2, 1982, 203–221; *Part 2, Sampling theory and complex waves*, Geophysics 47, 2, 1982, 222–236.

[Sw] W. Sweldens, *The lifting scheme: A custom-design construction of biorthogonal wavelets*, Appl. Comput. Harmon. Anal. 3, 1996, 186–200.

[TH] S. Theis, A. Hense, *Filterung von LM–Vorhersagefeldern mittels einer Wavelettransformation* (in German), Manuscript, Meteorologisches Institut, Universität Bonn, April 2000.