Statistical Modeling by Wavelets
WILEY SERIES IN PROBABILITY AND STATISTICS
APPLIED PROBABILITY AND STATISTICS SECTION

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: Vic Barnett, Noel A. C. Cressie, Nicholas I. Fisher,
Iain M. Johnstone, J. B. Kadane, David G. Kendall, David W. Scott,
Bernard W. Silverman, Adrian F. M. Smith, Jozef L. Teugels;
Ralph A. Bradley, Emeritus, J. Stuart Hunter, Emeritus

A complete list of the titles in this series appears at the end of this volume.
Statistical Modeling by Wavelets

BRANI VIDAKOVIC
Duke University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto
# Contents

**Preface**  

**Acknowledgments**  

1 Introduction  

1.1 Wavelet Evolution  

1.2 Wavelet Revolution  

1.3 Wavelets and Statistics  

1.4 An Appetizer: California Earthquakes  

2 Prerequisites  

2.1 General  

2.2 Hilbert Spaces  

2.2.1 Projection Theorem  

2.2.2 Orthonormal Sets  

2.2.3 Reproducing Kernel Hilbert Spaces  

2.3 Fourier Transformation  

2.3.1 Basic Properties  

2.3.2 Poisson Summation Formula and Sampling Theorem  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xi</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xiii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Wavelet Evolution</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Wavelet Revolution</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Wavelets and Statistics</td>
<td>16</td>
</tr>
<tr>
<td>1.4 An Appetizer: California Earthquakes</td>
<td>19</td>
</tr>
<tr>
<td>2 Prerequisites</td>
<td>23</td>
</tr>
<tr>
<td>2.1 General</td>
<td>23</td>
</tr>
<tr>
<td>2.2 Hilbert Spaces</td>
<td>24</td>
</tr>
<tr>
<td>2.2.1 Projection Theorem</td>
<td>25</td>
</tr>
<tr>
<td>2.2.2 Orthonormal Sets</td>
<td>26</td>
</tr>
<tr>
<td>2.2.3 Reproducing Kernel Hilbert Spaces</td>
<td>29</td>
</tr>
<tr>
<td>2.3 Fourier Transformation</td>
<td>29</td>
</tr>
<tr>
<td>2.3.1 Basic Properties</td>
<td>30</td>
</tr>
<tr>
<td>2.3.2 Poisson Summation Formula and Sampling Theorem</td>
<td>32</td>
</tr>
</tbody>
</table>
2.3.3 Fourier Series 33
2.3.4 Discrete Fourier Transform 34
2.4 Heisenberg's Uncertainty Principle 35
2.5 Some Important Function Spaces 36
2.6 Fundamentals of Signal Processing 39
2.7 Exercises 40

3 Wavelets 43
3.1 Continuous Wavelet Transformation 43
3.1.1 Basic Properties 44
3.1.2 Wavelets for Continuous Transformations 48
3.2 Discretization of the Continuous Wavelet Transform 50
3.3 Multiresolution Analysis 51
3.3.1 Derivation of a Wavelet Function 57
3.4 Some Important Wavelet Bases 60
3.4.1 Haar's Wavelets 60
3.4.2 Shannon's Wavelets 63
3.4.3 Meyer's Wavelets 65
3.4.4 Franklin's Wavelets 70
3.4.5 Daubechies' Compactly Supported Wavelets 73
3.5 Some Extensions 80
3.5.1 Regularity of Wavelets 80
3.5.2 The Least Asymmetric Daubechies' Wavelets: Symmlets 85
3.5.3 Approximations and Characterizations of Functional Spaces 87
3.5.4 Daubechies-Lagarias Algorithm 89
3.5.5 Moment Conditions 91
3.5.6 Interpolating (Cardinal) Wavelets 93
3.5.7 Pollen-Type Parameterization of Wavelets 94
3.6 Exercises 96

4 Discrete Wavelet Transformations 101
4.1 Introduction 101
4.2 The Cascade Algorithm 104
4.3 The Operator Notation of DWT 109
4.3.1 Discrete Wavelet Transformations as Linear Transformations 115
4.4 Exercises 117
5 Some Generalizations
   5.1 Coiflets
      5.1.1 Construction of Coiflets
   5.2 Biorthogonal Wavelets
      5.2.1 Construction of Biorthogonal Wavelets
      5.2.2 B-Spline Wavelets
   5.3 Wavelet Packets
      5.3.1 Basic Properties of Wavelet Packets
      5.3.2 Wavelet Packet Tables
   5.4 Best Basis Selection
      5.4.1 Some Cost Measures and the Best Basis Algorithm
   5.5 \( \epsilon \)-Decimated and Stationary Wavelet Transformations
      5.5.1 \( \epsilon \)-Decimated Wavelet Transformation
      5.5.2 Stationary (Non-Decimated) Wavelet Transformation
   5.6 Periodic Wavelet Transformations
   5.7 Multivariate Wavelet Transformations
   5.8 Discussion
   5.9 Exercises

6 Wavelet Shrinkage
   6.1 Shrinkage Method
   6.2 Linear Wavelet Regression Estimators
      6.2.1 Wavelet Kernels
      6.2.2 Local Constant Fit Estimators
   6.3 The Simplest Non-Linear Wavelet Shrinkage: Thresholding
      6.3.1 Variable Selection and Thresholding
      6.3.2 Oracular Risk for Thresholding Rules
      6.3.3 Why the Wavelet Shrinkage Works
      6.3.4 Almost Sure Convergence of Wavelet Shrinkage Estimators
   6.4 General Minimax Paradigm
      6.4.1 Translation of Minimaxity Results to the Wavelet Domain
   6.5 Thresholding Policies and Thresholding Rules
      6.5.1 Exact Risk Analysis of Thresholding Rules
      6.5.2 Large Sample Properties of \( \hat{f} \)
6.5.3 Some Other Shrinkage Rules 191
6.6 How to Select a Threshold 194
6.6.1 Mallat's Model and Induced Percentile Thresholding 194
6.6.2 Universal Threshold 195
6.6.3 A Threshold Based on Stein's Unbiased Estimator of Risk 199
6.6.4 Cross-Validation 200
6.6.5 Thresholding as a Testing Problem 202
6.6.6 Lorentz Curve Thresholding 203
6.6.7 Block Thresholding Estimators 206
6.7 Other Methods and References 207
6.8 Exercises 213

7 Density Estimation 217
7.1 Orthogonal Series Density Estimators 217
7.2 Wavelet Density Estimation 219
7.2.1 δ-Sequence Density Estimators 219
7.2.2 Bias and Variance of Linear Wavelet Density Estimators 223
7.2.3 Linear Wavelet Density Estimators in a More General Setting 224
7.3 Non-Linear Wavelet Density Estimators 225
7.3.1 Global Thresholding Estimator 227
7.4 Non-Negative Density Estimators 228
7.4.1 Estimating the Square Root of a Density 229
7.4.2 Density Estimation by Non-Negative Wavelets 234
7.5 Other Methods 236
7.5.1 Multivariate Wavelet Density Estimators 236
7.5.2 Density Estimation as a Regression Problem 239
7.5.3 Cross-Validation Estimator 240
7.5.4 Multiscale Estimator 240
7.5.5 Estimation of a Derivative of a Density 241
7.6 Exercises 243

8 Bayesian Methods in Wavelets 247
8.1 Motivational Examples 247
8.2 Smooth Shrinkage 250
8.3 Bayesian Thresholding 255
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.4</td>
<td>MAP-Principle</td>
<td>257</td>
</tr>
<tr>
<td>8.5</td>
<td>Density Estimation Problem</td>
<td>259</td>
</tr>
<tr>
<td>8.6</td>
<td>Full Bayesian Model</td>
<td>262</td>
</tr>
<tr>
<td>8.7</td>
<td>Discussion and References</td>
<td>265</td>
</tr>
<tr>
<td>8.8</td>
<td>Exercises</td>
<td>267</td>
</tr>
<tr>
<td>9</td>
<td>Wavelets and Random Processes</td>
<td>271</td>
</tr>
<tr>
<td>9.1</td>
<td>Stationary Time Series</td>
<td>272</td>
</tr>
<tr>
<td>9.2</td>
<td>Wavelets and Stationary Processes</td>
<td>273</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Wavelet Transformations of Stationary</td>
<td>273</td>
</tr>
<tr>
<td></td>
<td>Processes</td>
<td></td>
</tr>
<tr>
<td>9.2.2</td>
<td>Whitening of Stationary Processes</td>
<td>274</td>
</tr>
<tr>
<td>9.2.3</td>
<td>Karhunen-Loève-Like Expansions</td>
<td>276</td>
</tr>
<tr>
<td>9.3</td>
<td>Estimation of Spectral Densities</td>
<td>279</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Gao's Algorithm</td>
<td>280</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Non-Gaussian Stationary Processes</td>
<td>284</td>
</tr>
<tr>
<td>9.4</td>
<td>Wavelet Spectrum</td>
<td>286</td>
</tr>
<tr>
<td>9.4.1</td>
<td>Wavelet Spectrum of a Stationary Time</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>Series</td>
<td></td>
</tr>
<tr>
<td>9.4.2</td>
<td>Scalogram and Periodicities</td>
<td>289</td>
</tr>
<tr>
<td>9.5</td>
<td>Long-Memory Processes</td>
<td>290</td>
</tr>
<tr>
<td>9.5.1</td>
<td>Wavelets and Fractional Brownian Motion</td>
<td>290</td>
</tr>
<tr>
<td>9.5.2</td>
<td>Estimating Spectral Exponents in Self-Similar Processes</td>
<td>292</td>
</tr>
<tr>
<td>9.5.3</td>
<td>Quantifying the Whitening Property of Wavelet Transformations for fBm Processes</td>
<td>295</td>
</tr>
<tr>
<td>9.6</td>
<td>Discussion and References</td>
<td>296</td>
</tr>
<tr>
<td>9.7</td>
<td>Exercises</td>
<td>297</td>
</tr>
<tr>
<td>10</td>
<td>Wavelet-Based Random Variables and Densities</td>
<td>299</td>
</tr>
<tr>
<td>10.1</td>
<td>Scaling Function as a Density</td>
<td>299</td>
</tr>
<tr>
<td>10.2</td>
<td>Wavelet-Based Random Variables</td>
<td>300</td>
</tr>
<tr>
<td>10.3</td>
<td>Random Densities via Wavelets</td>
<td>308</td>
</tr>
<tr>
<td>10.3.1</td>
<td>Tree Algorithm</td>
<td>308</td>
</tr>
<tr>
<td>10.4</td>
<td>Properties of Wavelet-Based Random Densities</td>
<td>310</td>
</tr>
<tr>
<td>10.5</td>
<td>Random Densities With Constraints</td>
<td>312</td>
</tr>
<tr>
<td>10.5.1</td>
<td>Smoothness Constraints</td>
<td>312</td>
</tr>
<tr>
<td>10.5.2</td>
<td>Constraints on Symmetry</td>
<td>312</td>
</tr>
<tr>
<td>10.5.3</td>
<td>Constraints on Modality</td>
<td>313</td>
</tr>
<tr>
<td>10.5.4</td>
<td>Skewed Random Densities</td>
<td>313</td>
</tr>
</tbody>
</table>
Preface

Just two months ago astronomers did not know about it. But now they are giving good odds that Hyakutake will be the most impressive comet since the invention of telescope 400 years ago. (*Herald Sun, Durham, NC, March 24, 1996.*)

One can trace the origins of wavelets back to the beginning of this century; however, wavelets, understood as a systematic way of producing local orthogonal bases, are a recent unification of existing theories in various fields and some important “discoveries.” They are mathematical objects that have interpretation and application in many scientific fields, most notably in the fields of signal processing, nonparametric function estimation, and data compression. In the early 1990s, a series of papers by Donoho and Johnstone and their coauthors demonstrated that wavelets are appropriate tools in problems of denoising, regression, and density estimation. The subsequent burgeoning wavelet research broadened to a wide range of statistical problems.

This book is aimed at graduate students in statistics and mathematics, practicing statisticians, and statistically curious engineers. It can serve as a text for an introductory wavelet course concerned with an interface of wavelet methods and statistical inference. The necessary mathematical background is proficiency in advanced calculus and algebra; consequently, this book should be useful to advanced undergraduate students as well as to graduate students in statistics, mathematics, and engineering.

This book originated from the class notes supporting the Special Topics Course on Multiscale Methods at Duke University. The content can be divided into two parts:
an introduction to wavelets (Chapters 1–5) and statistical modeling (Chapters 6–11).

An introduction and some mathematical prerequisites are presented in Chapters 1 and 2. Continuous and discrete wavelet transformations are covered in Chapters 3 and 4. Some important generalizations (coiflets, biorthogonal wavelets, wavelet packets, stationary, periodized and multivariate wavelets) are covered in Chapter 5.

Chapters 6–11 are data-oriented. Chapter 6 is the crux of the book, covering the theory and practice of wavelet shrinkage. Important theoretical aspects of wavelet density estimation are covered in Chapter 7. Chapter 8 discusses Bayesian modeling in the wavelet domain. Time series are covered in Chapter 9, while Chapter 10 contains several probabilistic and simulational properties of wavelet-based random functions and densities. Chapter 11 gives some novel and important wavelet applications in statistics.

Instead of providing appendices with data sets and programs used in the book, I opted for a more modern style. The web page:

http://www.isds.duke.edu/~brani/wiley.html

is associated with the book. This page contains all data sets, functions, and programs referred to.

I hope the reader will find this book useful. All comments, suggestions, updates, and critiques will be appreciated.

BRANI VIDAKOVIC

Institute of Statistics and Decision Sciences
Duke University
Durham, February 1999
Acknowledgments

This book was made possible by the support of many individuals. First, I am grateful to colleagues from the Institute of Statistics and Decision Sciences (ISDS) – Duke University for supporting a wavelet-based statistics course. The attending students were patient and understood the difficulties of transferring research papers to a working copy of course notes. Their industry and enthusiasm is treasured. Duke University helped with this project through the Grant of Arts and Sciences Council 1997, and partial support was provided by the National Science Foundation Award DMS-9626159 at Duke University. Figure 1.10 is reproduced with the permission of the Salvador Dalí Museum, Inc. in St. Petersburg, Florida.

Many colleagues contributed to this project in different ways: Anestis Antoniadis, Tony Cai, Merlise Clyde, Lubo Dechevsky, Iain Johnstone, Gabriel Katul, Eric Kolaczyk, Pedro Morettin, Peter Müller, Giovanni Parmigiani, Marianna Pensky, David Rios, Fabrizio Ruggeri, Rainer von-Sacks, Naoki Saito, Yazhen Wang, and Gilbert Walter, to list a few. Collaboration with software gurus Hong-Ye Gao [TeraLogic Inc.] and Andrew Bruce [MathSoft Inc.] was fruitful. The S+Wavelets module (for S-Plus) was used for almost all of the computer examples, figures, and calculations. I am grateful to Alison Bory, Angioline Loredo, and Steve Quigley from Wiley, for their enthusiastic assistance, and to Courtney Johnson, Michael Kozdron, and Kathy Zhou, doctoral students at Duke University, for their help in proofreading the manuscript.

And most of all, I am grateful to my family for their love and strong and continuous support.
In this chapter, we give a brief overview of the history of wavelets, make a case for their use in statistics, and provide a real-life example that emphasizes specificities of wavelets in data processing problems. The wavelet method in this example is compared with its counterpart traditional approaches. The reader may encounter unfamiliar jargon or undefined objects. Some of these notions will be defined later and some are used to illustrate the general picture.

1.1 WAVELET EVOLUTION

Wavelets are developed not only from a couple of bright discoveries, but from concepts and theories that already existed in various fields. In this section, we will give a brief historic tour of some important milestones in the development of wavelets.

Functional series have a long history that can be traced back to the early nineteenth century. French mathematician (and politician!) Jean-Baptiste-Joseph Fourier [Fig. 1.1(a)] in 1807 decomposed a continuous, periodic on \([-\pi, \pi]\) function \(f(x)\) into

---

the series
\[
\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,
\]
where the coefficients \(a_n\) and \(b_n\) are defined as
\[
a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx, \quad n = 0, 1, 2, \ldots
\]
\[
b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx, \quad n = 1, 2, \ldots.
\]
It is interesting that, at the time of Fourier's discovery, the notion of a function was not yet precisely defined.

\(\text{Fig. 1.1}\) (a) Jean-Baptiste-Joseph Fourier 1768-1830 and (b) Alfred Haar 1885-1933.

The first “wavelet basis” was discovered in 1910 when Alfred Haar [Fig. 1.1(b)] showed that any continuous function \(f(x)\) on \([0, 1]\) can be approximated by
\[
f_n(x) = \langle \xi_0, f \rangle \xi_0(x) + \langle \xi_1, f \rangle \xi_1(x) + \cdots + \langle \xi_n, f \rangle \xi_n(x),\tag{1.1}
\]
and that, when \(n \to \infty\), \(f_n\) converges to \(f\) uniformly ([181]). The coefficients \(\langle \xi_i, f \rangle\) are given by \(\int \xi_i(x)f(x)\,dx\). The Haar basis is very simple:
\[ \begin{align*}
\xi_0(x) &= 1(0 \leq x \leq 1), \\
\xi_1(x) &= 1(0 \leq x \leq 1/2) - 1(1/2 \leq x \leq 1), \\
\xi_2(x) &= \sqrt{2}[1(0 \leq x \leq 1/4) - 1(1/4 \leq x \leq 1/2)], \\
\cdots \\
\xi_n(x) &= 2^{j/2}[1(k \cdot 2^{-j} \leq x \leq (k + 1/2) \cdot 2^{-j}) \\
&\quad - 1((k + 1/2) \cdot 2^{-j} \leq x \leq (k + 1) \cdot 2^{-j})], \\
\cdots 
\end{align*} \]

where \( n \) is uniquely decomposed as \( n = 2^j + k, \ j \geq 0, \ 0 \leq k \leq 2^j - 1 \), and \( 1(A) \) is the indicator of a set \( A \), i.e., \( 1(A) = 1 \), if \( x \in A \), and \( 1(A) = 0 \), if \( x \in A^c \).

The approximation in (1.1) is equivalent to an approximation by step functions whose values are the averages (mean values) of the function over appropriate dyadic intervals.

Fig. 1.2 gives an exemplary function, \( f(x) = \sin \pi x + \cos 2\pi x + 0.6 \cdot 1(x > 1/2) \), and three different levels of approximation: \( f_3, f_{15}, \) and \( f_{63} \). Basis functions \( \xi_1, \xi_2, \xi_{14}, \) and \( \xi_{25} \) are shown in Fig. 1.3. Since \( \int \xi_n^2(x) \, dx = 1 \) for an arbitrary \( n \), there is a trade-off between the magnitude and the support of the basis functions in the Haar system.

Notice that for any \( n \geq 1 \) the basis function \( \xi_n \) can be expressed as a scale-shift transformation of a single function \( \xi_1 \),

\[ \xi_n(x) = 2^{j/2}\xi_1(2^j \cdot x - k), \ n = 2^j + k, \]

a property shared by critically sampled wavelets, as we will see later. The function \( \xi_0(x) \) is different in nature than the functions \( \xi_n, \ n \geq 1 \); while the functions \( \xi_n, \ n \geq 1 \) describe the details in the decomposition, the function \( \xi_0(x) \) is responsible for the "average" of the decomposed function.

The Schauder basis on \([0, 1]\) (Schauder [369]) consists of the primitives of the Haar basis functions, the triangle functions. Let \( \Delta(x) = 2x \cdot 1(0 \leq x \leq 1/2) + 2(1 - x) \cdot 1(1/2 \leq x \leq 1) \), and let \( \Delta_n(x) = \Delta(2^j x - k), \ n = 2^j + k, \ j \geq 0, \ 0 \leq k \leq 2^j - 1 \). Then \( \{1, \Delta(x), \Delta_1(x), \ldots\} \) constitutes a Schauder basis on \([0, 1]\) and, as in the case of Haar's basis, any continuous function \( f(x) \) on \([0, 1]\) can be approximated by

\[ f_N(x) = a + bx + \sum_{n=1}^{N} s_n \Delta_n(x). \quad (1.2) \]

Coefficients \( a \) and \( b \) are solutions of the system \( f(0) = a \) and \( f(1) = a + b \), while the coefficients \( s_n \) can be obtained by the simple relation
Fig. 1.2 Panels (a)-(d) show the original function $f(x) = \sin \pi x + \cos 2\pi x + 0.6 \cdot 1(x > \frac{1}{2})$, $0 \leq x \leq 1$, and three different levels of approximation in the Haar basis. Using the notation of (1.1), approximations $f_3$, $f_{15}$, and $f_{63}$ are plotted.
Fig. 1.3 Functions $\xi_1, \xi_2, \xi_3$, and $\xi_4$ from the Haar basis of $L_2([0, 1])$.

\[
s_n = f\left(\frac{k + 1/2}{2^j}\right) - \frac{1}{2} \left[ f\left(\frac{k}{2^j}\right) + f\left(\frac{k + 1}{2^j}\right)\right],
\]

where $n = 2^j + k$, $j \geq 0$, $0 \leq k \leq 2^j - 1$.

The convergence $f_N(x) \to f(x)$ is uniform and the coefficients are unique; however, the Schauder system is not orthogonal. We will see later that its orthogonalization leads to a family of wavelets, known as Franklin wavelets.

In the mid-1930s, Littlewood-Paley techniques (based on Fourier methods) [264] were broadly used in research on Fourier summability and in investigation of the behavior of analytic functions.

Prototypes of wavelets first appeared in Lusin's work in the 1930s. A standard characterization of Hardy's spaces can be given in terms of Lusin's "area" functions.

In the 1950s and 1960s, techniques by Littlewood-Paley and Lusin were developed into powerful tools for studying physical phenomena describable by solutions of differential and integral equations. Researchers realized that these techniques could be unified by the Calderón-Zygmund theory [292], now a branch of harmonic analysis.

Strömberg [390] was the first to construct an orthonormal basis of $L_2(\mathbb{R})$ of the form \(
\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \quad j, k \in \mathbb{Z}\}\), a wavelet-like basis more general than Haar's basis. Stromberg's construction uses Franklin systems which are Gram-Schmidt orthogonalized Schauder basis functions $\Delta_n(x)$. 
For more information about the historical roots of wavelets, we direct the reader to monographs by Meyer [294, 295] and Daubechies [104].

1.2 WAVELET REVOLUTION

The first definitions of wavelets can be attributed to Morlet et al. [300] and Morlet and Grossmann [179] and it is given in the Fourier domain: A wavelet is an $L_2(\mathbb{R})$ function for which the Fourier transformation $\Psi(\omega)$ satisfies

$$\int_{0}^{\infty} |\Psi(t\omega)|^2 \frac{dt}{t} = 1, \text{ for almost all } \omega.$$ 

The definition of Morlet and Grossmann is quite broad and over time the meaning of the term wavelet became narrower. Currently, the term wavelet is usually associated with a function $\psi \in L_2(\mathbb{R})$ such that the translations and dyadic dilations of $\psi$,

$$\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k), \quad j, k \in \mathbb{Z} \quad (1.3)$$

constitute an orthonormal basis of $L_2(\mathbb{R})$.

Calculating wavelet expansions directly is a computationally expensive task, moreover, most interesting wavelets are without a closed form. In the mid-1980s, Mallat [274, 275, 276] connected quadrature-minor filtering and pyramidal algorithms from the signal processing theory with wavelets. He demonstrated that discrete wavelet transformation can be calculated very rapidly via cascade-like algorithm. This link was of paramount importance for the practice of wavelets. Daubechies' discovery of compactly supported wavelet bases represents another important milestone in the development of wavelet theory. Daubechies' bases are versatile in smoothness and locality and represent a starting point for much of the subsequent generalizations and theoretical advances.

Wavelet theory has developed now into a methodology used in many disciplines: mathematics, geophysics, astronomy, signal processing, numerical analysis, and statistics, to list a few. Wavelets are providing a rich source of useful and sometimes intriguing tools for applications in "time-scale" types of problems. In analyses of signals, the wavelet representations allow us to view a time-domain evolution in terms of scale components. In this respect, wavelet transformations behave similarly to Fourier transformations. The Fourier transform extracts details from the signal frequency, but all information about the location of a particular frequency within the signal is lost. Time localization can be achieved by first windowing the signal, and then by taking its Fourier transform. The problem with windowing is that slices of the processed signal are of a fixed length, which is determined by the window. Slices of the same length are used to resolve both high and low frequency components. For nonstationary signals, this lack of adaptivity may lead to a local under- or over-fitting.
Table 1.1 Coefficients of the doppler function in the Haar basis plotted in levels determined by the length of support of corresponding basis functions, $\xi_n$, $n \geq 0$.

<table>
<thead>
<tr>
<th>level in Fig. 1.4</th>
<th>coefficients of support</th>
<th>$j$ and $k$ in the notation: $n = 2^j + k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$j = \frac{1}{2^9}$</td>
</tr>
<tr>
<td>d1</td>
<td>$\xi_{512} - \xi_{1023}$</td>
<td>$j = \frac{1}{2^5}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$j = \frac{1}{4}$</td>
</tr>
<tr>
<td>d8</td>
<td>$\xi_4 - \xi_7$</td>
<td>$j = \frac{1}{2}$</td>
</tr>
<tr>
<td>d9</td>
<td>$\xi_2 - \xi_3$</td>
<td>$j = 0$</td>
</tr>
<tr>
<td>s10</td>
<td>$\xi_0$</td>
<td></td>
</tr>
</tbody>
</table>

In contrast to windowed Fourier transforms, wavelets select widths of time slices according to the local frequency in the signal. This adaptivity property of wavelets is very important, and we will make it more precise later in the discussion of Heisenberg's uncertainty principle. Two panels in Fig. 1.11, on page 18, depict slicing the time-scale plane for a windowed Fourier (left) and a wavelet transformation (right).

Now we give several examples: The first example views the Haar decomposition as a wavelet decomposition and discusses connections between “levels” and resolutions of the decomposition. The subsequent four examples demonstrate important properties of wavelets: the ability to filter, “disbalance”, and “whiten” signals as well as to detect self-similarity within a signal.

Example 1.2.1 The Haar basis as a wavelet basis. To illustrate the time and scale adaptivity of wavelets, and to introduce some necessary wavelet notations and jargon, let us consider a decomposition of the function

$$y(x) = \sqrt{x(1-x)} \sin \frac{2.1\pi}{x + 0.05}, 0 \leq x \leq 1,$$

(1.4)
in Haar's basis. This function is known as the doppler test-function. Notice that frequency in the function increases as $x$ decreases.

In Table 1.1, $n$ is represented as $2^j + k$ where $j$ is a level and $k$ is a shift within the level. Notice that all functions $\xi$ within a level have supports of the same length. The support of a function is defined as closure of the set at which the function differs from zero.

When $j \to \infty$, the number of coefficients in the level increases and the length of support of the corresponding basis functions decreases. For example, the level indexed by $j = 5$ has $2^5 = 32$ coefficients and the supports are of length $2^{-5}$. The shifts within a level are indexed by $k$, where $k$ ranges from 0 to $2^j - 1$. For
Fig. 1.4 The doppler function and its Haar basis decomposition.
an exact description of an arbitrary function, the number of levels is infinite. The coefficient corresponding to $\xi_0$ is called the "smooth" coefficient ($s_0$ in Fig. 1.4) and the coefficients corresponding to $\xi_n$, $n > 1$ are called "detail" coefficients. Level $j = 9$ ($d_1$ in Fig. 1.4) contains coefficients corresponding to "fine" details.

When dealing with functions that are given by their sampled values, it is customary to set the sampled values to be "smooth" coefficients at the level $j = J$. The subsequent "detail" levels denoted by $d_1, d_2, \ldots$, correspond to $j = J - 1, J - 2, \ldots$.

We provide four more examples that emphasize the most interesting features of wavelet transformations. Occasionally, we will use terms like "fine and coarse levels", "wavelet domain", and "energy", which have not been previously defined and will be defined in the subsequent chapters. However, the intended messages of the examples should be clear even without precise definitions of such terms.

**Example 1.2.2 Wavelets generate local bases.** Classical orthonormal bases (Fourier, Hermite, Legendre, etc.) have been used with great success in applied mathematics for decades. However, there is a serious limitation shared by many classical bases, which is non-locality. A basis is non-local when many basis functions are substantially contributing at any value of a decomposition. The convergence of non-local classical decompositions often relies on a multitude of cancellations.

Local bases are desirable since they are more adaptive and parsimonious. In 1946, Gabor [161] suggested localizing Fourier bases by modulating and translating an appropriate "window" function $g$. More precisely, Gabor suggested bases in the form

$$\{g_{m,n}(x) = e^{2\pi i m x} g(x - n)\},$$

where $m$ and $n$ are integers and $g$ is a square-integrable function. An example of a function $g$ that produces an orthonormal basis of $L^2(\mathbb{R})$ is $\sin(\pi x)/(\pi x)$.

The Balian-Low theorem stipulates limitations of Gabor bases. If the Gabor basis is orthogonal and $\hat{g}(\omega)$ is the Fourier transformation of the window $g(x)$ then, by the Balian-Low theorem, either $\int x^2 |g(x)|^2 dx = \infty$ or $\int \omega^2 |\hat{g}(\omega)|^2 d\omega = \infty$. In other words, orthogonal Gabor bases are non-local either in time or in scale (frequency). Modulations and translations of the Gaussian window $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (which is well localized in both time and frequency, and for which the above integrals are finite) will not produce an orthonormal basis.

Locality of wavelet bases comes from their construction. Most of the wavelets that are used in statistics now are either compactly supported or decay exponentially. An exception are Meyer-type wavelets (with a polynomial decay) used in deconvolution problems.

**Example 1.2.3 Wavelets filter data.** To illustrate the action of wavelets as a filtering device, we generate two periodic functions with different frequencies, $y_1 = \sin x + \cos 2x$, and $y_2 = \frac{1}{2} \arcsin(\sin 20x)$, where $x \in [-2\pi, 2\pi]$. These are shown in panel (a) in Fig. 1.5. Our goal is to filter out the component $y_2$ from the given sum $y_1 + y_2$. 
Since the periods of $y_1$ and $y_2$ are different, the functions are described by wavelets with different supports (and whose coefficients belong to different levels). Fig. 1.5(c), depicts the level-wise energies (sums of squares of wavelet coefficients). The support of wavelets associated with level 1 is 32 times larger than the support of wavelets associated with level 5. This means that almost all the energy in levels 0, 1, and 2 comes from signal $y_1$, and the energy in level 5 comes from $y_2$, thus allowing an easy separation. The filtered components are depicted in Fig. 1.5(d).

**Example 1.2.4 Wavelets “disbalance” energy in data.** The term “disbalance” is coined and it relates to an uneven distribution of energy in a signal. Disbalancing is desirable since a signal can be well described by only a few energetic components.

To illustrate the disbalancing action that is typical of wavelets, we first introduce some necessary notation. Given a vector $g = (a_1, a_2, \ldots, a_n)$ let $\|g\|^2 = \sum_i a_i^2$ be the total energy of $g$ and let $a_i^2$ be the $i$th energy component. Let $a_1^2, a_2^2, \ldots, a_n^2$ be increasingly ordered energy components. The standard measure of disbalance used in economics is the Lorentz curve. The Lorentz curve was introduced at the beginning of the century. It was used by economics researchers to assess inequality of distribution of wealth in a country, region, or among people within a particular population group.

One definition of the Lorentz curve, in terms of energy components, is

$$L(p) = \frac{1}{\|g\|^2} \cdot \sum_{i=1}^{np} a_{(i)}^2, \quad p \in [0, 1],$$

where $\lfloor x \rfloor$ is the largest integer smaller than $x$. In Fig. 1.6(a), an observed time series (turbulence data set) is given. Below is its wavelet transformation represented in a vector form beginning with coarse coefficients. Orthogonality of the transformation preserves the total energy, $\|g\|^2$. However, the energy in the wavelet domain is more disbalanced, as indicated by the Lorentz curves in Fig. 1.6(b). Notice that 90% of energy is contained in about 6-7% of the components in the wavelet-transformed data set compared to nearly 50% of the components in the original (time) domain.

**Example 1.2.5 Wavelets whiten data.** In this example, we show another interesting property of wavelets. Orthogonal wavelet transformations map white noise to white noise, which is a consequence of orthogonality. However, signals that are correlated in the time domain become almost uncorrelated in the wavelet domain. Informally, the wavelet transformation acts as an approximation to the Karhunen-Loève transformation. To exemplify this statement, a time series of 256 components was generated from a random process with stationary increments, ARIMA(1,1,1) process. Such processes exhibit long-range dependence and their autocovariance functions (Fig. 1.7(a)) show slow decay. The autocovariance function of the wavelet-transformed time series exhibits very different behavior. Only the covariances at the first few lags are significant at a 5% significance level.

Related discussion can be found in Johnstone and Silverman [222], Mallat [277].
Fig. 1.5 Filtering property of wavelets. Two functions $y_1 = \sin x + \cos 2x$ and $y_2 = \frac{1}{2} \arcsin(\sin 20x)$, and their sum $y_1 + y_2$ are plotted in panels (a) and (b). Panel (c) shows the separation of “energy” to different levels in wavelet decomposition, while panel (d) shows filtered functions.
Fig. 1.6  (a) Atmospheric turbulence measurements of $u$ velocity component (upper panel) and their wavelet transformation (lower panel). (b) Lorentz curves of the original and transformed measurements. The curve corresponding to transformed measurements has higher curvature.
Fig. 1.7  Illustration of the whitening effect of wavelet transformations. Autocovariance function for a time series [ARIMA(1,1,1)] in the time domain [panel (a)] and the wavelet domain [panel (b)].
Fig. 1.8 Self-similarity of a turbulence time series.

Walter [439], and Wornell [461].

Example 1.2.6 Wavelets detect self-similar phenomena. Being self-similar themselves, wavelets are especially apt to describe phenomena exhibiting self-similarity in different scales (Fig. 1.8). Early research on wavelets was generated to address related problems in geophysics, especially in turbulence. An overview can be found in Kumar and Foufoula-Georgiou [250]. A curious phenomenon is that atmospheric turbulence measurements of different physical quantities, such as air velocities, ozone and humidity concentrations, temperature, and so on, follow identical power laws (as predicted by Kolmogorov's [242] theory). Such laws describe the energy transport in the inertial range of turbulent flows. A nice reference is a book by Frisch [160].

One of the theoretical laws is the “$-\frac{5}{3}$” law. It states that the log-power spectrum in the inertial range decreases linearly, with the slope of $-\frac{5}{3}$. Fig. 1.9 shows the wavelet-spectrum of air velocity measurements and it's near-perfect compliance with the $-\frac{5}{3}$ law.

There are problems in which wavelets should be used with caution. For instance, in the wavelet domain, the dependence structure in the transformed time series is influenced by the choice of the decomposing wavelet. In some cases, the extent of such non-robustness hinders practical generalizations. When non-robustness is of particular concern, researchers usually fix a good wavelet for a class of problems, as is the case with the prevalent use of the Haar and Walsh bases in processing the turbulence data.
Fig. 1.9 Wavelet power-spectrum and Kolmogorov's $-\frac{5}{3}$ law. Dots represent the logarithms of the cumulative level-energies. The variable $\log(\text{wavenumber})$ is linearly related to the level $j$. 