The Statistician (2000)
49, *Part* 1, *pp.* 1–29

Wavelet analysis and its statistical applications

Felix Abramovich,

Tel Aviv University, Israel

Trevor C. Bailey

University of Exeter, UK

and Theofanis Sapatinas

University of Kent at Canterbury, UK

[Received February 1999. Revised September 1999]

Summary. In recent years there has been a considerable development in the use of wavelet methods in statistics. As a result, we are now at the stage where it is reasonable to consider such methods to be another standard tool of the applied statistician rather than a research novelty. With that in mind, this paper gives a relatively accessible introduction to standard wavelet analysis and provides a review of some common uses of wavelet methods in statistical applications. It is primarily orientated towards the general statistical audience who may be involved in analysing data where the use of wavelets might be effective, rather than to researchers who are already familiar with the field. Given that objective, we do not emphasize mathematical generality or rigour in our exposition of wavelets and we restrict our discussion to the more frequently employed wavelet methods in statistics. We provide extensive references where the ideas and concepts discussed can be followed up in greater detail and generality if required. The paper first establishes some necessary basic mathematical background and terminology relating to wavelets. It then reviews the more well-established applications of wavelets in statistics including their use in nonparametric regression, density estimation, inverse problems, changepoint problems and in some specialized aspects of time series analysis. Possible extensions to the uses of wavelets in statistics are then considered. The paper concludes with a brief reference to readily available software packages for wavelet analysis.

Keywords: Changepoint analysis; Density estimation; Fourier analysis; Inverse problems; Nonparametric regression; Signal processing; Spectral density estimation; Time series analysis; Wavelet analysis

1. Introduction

In contrast with papers that are directed at researchers who are already working in a specialized field, our intended audience here is not those who are conversant with wavelet analysis and its applications. We wish rather to address those statisticians who, although they have undoubtedly heard of wavelet methods, may be unsure exactly what they involve, in what practical areas they can usefully be employed and where best to find out more about the subject. Our aim is therefore to give a relatively straightforward and succinct account of standard wavelet analysis and the current use of such methods in statistics.

We should immediately emphasize that wavelet analysis, in common with many other

Address for correspondence: Theofanis Sapatinas, Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury, Kent, CT2 7NF, UK.
E-mail: T.Sapatinas@ukc.ac.uk

 0039–0526/00/49001

mathematical and algorithmic techniques used in statistics, did not originate from statisticians, nor with statistical applications in mind. The wavelet transform is a synthesis of ideas emerging over many years from different fields, notably mathematics, physics and engineering. Like Fourier analysis, with which analogies are often drawn, wavelet methods are general mathematical tools. Broadly speaking, the wavelet transform can provide economical and informative mathematical representations of many different objects of interest (e.g. functions, signals or images). Such representations can be obtained relatively quickly and easily through fast algorithms which are now readily available in a variety of computer packages. As a result, wavelets are used widely, not only by mathematicians in areas such as functional and numerical analysis, but also by researchers in the natural sciences such as physics, chemistry and biology, and in applied disciplines such as computer science, engineering and econometrics. Signal processing in general, including image analysis and data compression, is the obvious example of an applied field of multidisciplinary interest where the use of wavelets has proved of significant value. Good general surveys of wavelet applications in that and other fields are given, for example, in Meyer (1993), Young (1993), Aldroubi and Unser (1996) or Mallat (1998). Within that framework of multidisciplinary interest, statisticians are among the more recent users of the technique. They bring their own particular perspective to wavelet applications in areas such as signal processing and image analysis. In addition they have explored a range of wavelet applications which are more exclusively statistical, including nonparametric regression and density estimation.

Given that background, the first point for the reader to appreciate in relation to this paper is that we make no attempt to review the subject of wavelets in its broad context but focus on wavelets from the statistician's perspective and on their use to address statistical problems. Accordingly, the majority of the references that we provide are drawn from the statistical literature and to some extent our choice of mathematical exposition is also coloured to suit a statistical audience. This applies to the degree of depth and generality adopted in our basic mathematical description of the wavelet transform, as well as the level of mathematical or algorithmic detail provided in our discussion of applications.

A second point to appreciate is that, in line with our objective of addressing a general rather than specialized statistical audience, we restrict most of our discussion to relatively well-established statistical applications of wavelets. A variety of wavelet, and wavelet-related, methods have been developed in recent years with potential applications to an increasing range of statistical problems. The full breadth of the statistical applications of wavelets is still to emerge and relative benefits over alternative methods have yet to be realistically evaluated in some cases. Given that situation, this paper restricts itself to that subset of wavelet methods and applications in statistics which we believe to be both relatively accessible and which have been established to be of real practical value.

Overall then, we acknowledge from the outset that this is not intended to be an exhaustive review of every possible statistical application of wavelets and is even less complete with respect to covering wavelets methods in a broader context. However, there is no shortage of additional material on wavelets for the interested reader to turn to, both in relation to mathematical detail and with respect to more specialized wavelet techniques and applications in statistics. Throughout this paper we provide extensive references to choose from and we should perhaps immediately mention some useful basic texts. Examples of books that represent the underlying ideas of wavelets and develop necessary formalism include Chui (1992), Daubechies (1992) and Mallat (1998); for the rigorous mathematical theory of wavelets we refer interested readers to Meyer (1992) and Wojtaszczyk (1997), while comprehensive surveys of wavelet applications in statistics are given by Ogden (1997) or Vidakovic (1999), and that subject is covered in more mathematical depth in Härdle *et al.* (1998) or Antoniadis (1999).

The remainder of the paper is organized as follows. In Section 2 we first provide some necessary mathematical background and terminology in relation to wavelets. In Section 3 we review the use of wavelets in a variety of statistical areas, including nonparametric regression, density estimation, inverse problems, changepoint problems and some specialized aspects of time series analysis. Possible extensions to these relatively well-established wavelet procedures are then discussed in Section 4. Finally, in Section 5, we conclude and briefly refer to some software packages that are available for wavelet analysis.

2. Some background on wavelets

Here we develop some mathematical background and terminology which is required to understand the wavelet applications discussed in subsequent sections. In doing so we do not emphasize mathematical generality or rigour and provide few algorithmic details. Daubechies (1992), Meyer (1992) or Wojtaszczyk (1997) may be consulted for a more rigorous and general coverage.

2.1. The wavelet series expansion

The series expansion of a function in terms of a set of orthogonal basis functions is familiar in statistics, an obvious example being the commonly used Fourier expansion, where the basis consists of sines and cosines of differing frequencies. The term *wavelets* is used to refer to a set of basis functions with a very special structure. The special structure that is shared by all wavelet bases (see below) is the key to the main fundamental properties of wavelets and to their usefulness in statistics and in other fields. A variety of different *wavelet families* now exist (see Section 2.2) which enable 'good', orthonormal wavelet bases to be generated for a wide class of function spaces. In other words, many types of functions that are encountered in practice can be sparsely (i.e. parsimoniously) and uniquely represented in terms of a wavelet series. Wavelet bases are therefore not only useful by virtue of their special structure, but they may also be applied in a wide variety of contexts.

The special structure of wavelet bases may be appreciated by considering the generation of an orthonormal wavelet basis for functions $g \in L^2(\mathbb{R})$ (the space of square integrable real functions). Following the approach of Daubechies (1992), which is that most often adopted in applications of wavelets in statistics, we start with two related and specially chosen, mutually orthonormal, functions or *parent wavelets*: the *scaling function* ϕ (sometimes referred to as the *father wavelet*) and the *mother wavelet* ψ. Other wavelets in the basis are then generated by translations of the scaling function ϕ and dilations and translations of the mother wavelet ψ by using the relationships

$$\phi_{j_0 k}(t) = 2^{j_0/2}\,\phi(2^{j_0}t - k), \qquad \psi_{jk}(t) = 2^{j/2}\,\psi(2^{j}t - k), \qquad j = j_0, j_0+1, \ldots, \qquad k \in \mathbb{Z}, \tag{1}$$

for some fixed $j_0 \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of integers.

Possible choices for parent wavelets which give rise to a suitable basis for $L^2(\mathbb{R})$, and for various alternative function spaces, are discussed in Section 2.2, but typically the scaling function ϕ resembles a kernel function and the mother wavelet ψ is a well-localized oscillation (hence the name wavelet). A unit increase in j in expression (1) (i.e. dilation) has no effect on the scaling function ($\phi_{j_0 k}$ has a fixed width) but packs the oscillations of ψ_{jk} into half the width (doubles its 'frequency' or, in strict wavelet terminology, its *scale* or *resolution*). A unit increase in k in

expression (1) (i.e. translation) shifts the location of both $\phi_{j_0 k}$ and ψ_{jk}, the former by a fixed amount (2^{-j_0}) and the latter by an amount proportional to its width (2^{-j}) .

We give an example comparing Fourier and wavelet analysis in Section 2.4, but the essential point is that the sines and cosines of the standard Fourier basis have specificity only in frequency, whereas the special structure of a wavelet basis provides specificity in location (via translation) and also specificity in 'frequency' (via dilation).

Given the above wavelet basis, a function $g \in L^2(\mathbb{R})$ is then represented in a corresponding wavelet series as

$$g(t) = \sum_{k \in \mathbb{Z}} c_{j_0 k}\, \phi_{j_0 k}(t) + \sum_{j=j_0}^{\infty} \sum_{k \in \mathbb{Z}} w_{jk}\, \psi_{jk}(t), \tag{2}$$

with $c_{j_0 k} = \langle g, \phi_{j_0 k} \rangle$ and $w_{jk} = \langle g, \psi_{jk} \rangle$, where $\langle \cdot, \cdot \rangle$ is the standard L^2 inner product of two functions:

$$\langle g_1, g_2 \rangle = \int_{\mathbb{R}} g_1(t) g_2(t)\, \mathrm{d}t.$$

The wavelet expansion (2) represents the function g as a series of successive approximations. The first approximation is achieved by the sequence of scaling terms $c_{j_0 k}\phi_{j_0 k}$ (each intuitively being a smoothed 'average' in the vicinity of $2^{-j_0}k$). The oscillating features which cannot be represented with sufficient accuracy in this way are approximated in 'frequency' and in correspondingly fine detail by sequences of wavelet terms $w_{jk}\psi_{jk}$ (each intuitively representing 'smooth wiggly structure' of 'frequency' 2^j in the vicinity of $2^{-j}k$). We may view the representation of any selected finite section of the function in the wavelet expansion by analogy with taking pictures of this section with a camera equipped with a zoom lens and two filters or masks: one designed to highlight features that are encapsulated in the shape of the scaling function and the other designed similarly in relation to the mother wavelet. We first use the scaling filter, with the lowest magnification (i.e. the greatest field of vision) of the lens which corresponds to resolution level j_0, and take a panoramic sequence of pictures covering the particular section of the function (translations of ϕ). We then change to the mother wavelet filter, leave the magnification unchanged and repeat a similar sequence of pictures (translations of ψ at resolution j_0). We then zoom in, leaving the mother wavelet filter on the lens, but doubling the magnification (i.e. restricting the field of vision), so requiring to take a sequence of twice as many pictures to cover the particular section of the function (translations of ψ at resolution $j_0 + 1$). We repeat this zooming process with the mother wavelet filter *ad infinitum*, doubling the numbers of pictures taken to cover the section of the function on each occasion. The complete set of pictures obtained then allows us to view the selected section of the function in terms of average value or any specific frequency component, knowing that for any one of these purposes we shall automatically have available a sufficient number of correctly filtered pictures at the appropriate magnification.

In many practical situations, the function to be represented as a wavelet series may be defined to be zero outside a finite interval, such as the unit interval [0, 1]. Adapting wavelets to a finite interval requires some modifications. The obvious approach of simply vanishing the underlying function outside the interval will introduce artificial discontinuities at the end points. Several papers have studied the necessary boundary corrections to retain orthonormality on an interval (see, for example, Anderson *et al.* (1993) and Cohen *et al.* (1993a, b)). However, in practice the most commonly used approaches to adapting wavelet analysis to the interval are based on periodic, symmetric or antisymmetric boundary handling (see, for example, Daubechies (1992), Jawerth and Sweldens (1994) and Ogden (1997)). The use of periodic boundary conditions effectively implies periodic wavelets and we assume periodic boundary conditions on the unit

interval as the default in later sections of this paper. In that case for $g \in L^2[0, 1]$ the wavelet series representation is then

$$g(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0 k}\, \phi_{j_0 k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} w_{jk}\, \psi_{jk}(t). \tag{3}$$

2.2. *Different wavelet bases*

The selection of any arbitrary pair of mutually orthonormal parent wavelets, followed by the reproducing process (1) discussed in the previous section, will not automatically result in a basis for $L^2(\mathbb{R})$, let alone a 'good' basis in the sense of providing a parsimonious representation of functions in terms of their corresponding wavelet series. Clearly, the parent wavelets need to be specially chosen if that is to be the case.

The simplest wavelet basis for $L^2(\mathbb{R})$ is the Haar basis (Haar, 1910) which uses a parent couple given by

$$\phi(t) = \begin{cases} 1, & 0 \leqslant t \leqslant 1, \\ 0, & \text{otherwise}, \end{cases}$$

$$\psi(t) = \begin{cases} 1, & 0 \leqslant t < \frac{1}{2}, \\ -1, & \frac{1}{2} \leqslant t \leqslant 1, \\ 0, & \text{otherwise} \end{cases}$$

(see Figs 1(a) and 1(b)).

The Haar basis was known long before wavelets developed to take their present form. However, the Haar parent couple, even though it has compact support, does not have a good time–frequency localization. Moreover, the resulting wavelet basis functions have the significant additional disadvantage of being discontinuous, which renders them unsuitable as a basis for classes of smoother functions. For almost 70 years it was thought impossible to construct alternative wavelet bases with better properties than the Haar basis, such as good time–frequency localization, various degrees of smoothness (i.e. high regularity or numbers of continuous derivatives) and large numbers of vanishing moments (which is important in enabling the sparse representation of broad classes of functions). The concepts of *multiresolutional analysis* (see, for example, Jawerth and Sweldens (1994) and Härdle *et al.* (1998)) eventually provided the mathematical framework to develop new wavelet bases.

One of the earlier developments (which was almost unnoticed at the time) was a wavelet family proposed by Stromberg (1981), whose orthonormal wavelets have infinite support, an arbitrarily large number of continuous derivatives and exponential decay. Meyer (1985), unaware of this earlier work, also developed orthonormal wavelet bases with infinite support and exponential decay. A key development was the work of Daubechies (1988, 1992) who derived two families of orthonormal wavelet bases (the so-called *extremal phase* and *least asymmetric* families) which combine compact support with various degrees of smoothness and numbers of vanishing moments. These are now the most intensively used wavelet families in practical applications in statistics. The price paid for compact support is that the corresponding wavelets have only a finite number of continuous derivatives and vanishing moments. The mother wavelets in either family are essentially indexed in terms of increasing smoothness (regularity) and numbers of vanishing moments. As the smoothness increases the support correspondingly increases. Figs 1(c) and 1(d) illustrate the parent couple from the extremal phase family with six vanishing moments.

Coiflets (Daubechies, 1993) and *spline wavelets* (Chui, 1992) are other examples of wavelet

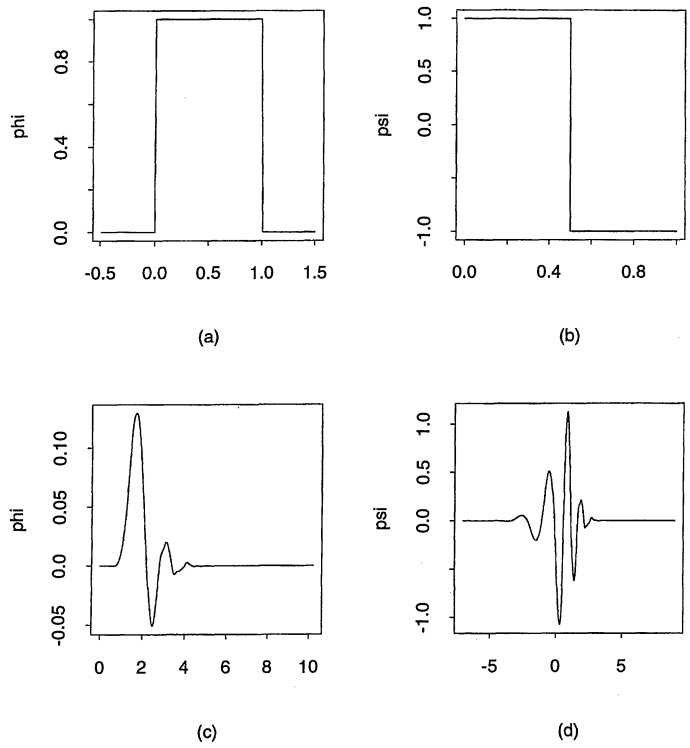

Fig. 1. (a), (b) Haar parent couple and (c), (d) extremal phase parent couple with six vanishing moments

bases that are used in practice and several additional wavelet families (orthogonal, biorthogonal and semiorthogonal) have also been developed during the last decade. Overviews of the construction of wavelet bases are, for example, given by Ogden (1997), Härdle *et al.* (1998) or Vidakovic (1999).

In general the choice of the wavelet family, and more particularly the regularity properties of the mother wavelet selected from it, will be motivated by the specific problem at hand. But the key point here is that wavelet families now exist, such as the extremal phase and least asymmetric families, which are sufficiently flexible to generate, via the choice of a suitably smooth mother wavelet, 'good' wavelet bases for a broad range of function spaces that are encountered in practical applications. These include function spaces like the Hölder and Sobolev spaces of regular functions, as well as spaces of irregular functions such as those of 'bounded variation'. These considerations are not simply esoteric but are of statistical importance, since these classes of functions typically arise in practical applications such as the processing of speech, electrocardiogram or seismic signals. Mathematically, all these function spaces can be formalized in terms of the so-called Besov spaces. For rigorous definitions and a detailed study of Besov spaces and their connection to wavelets the reader is referred to Meyer (1992), Wojtaszczyk (1997) or Härdle *et al.* (1998). Wavelet-based stochastic models which can be used to simulate a range of functions just on the boundary of membership of any particular Besov space can be found in Abramovich *et al.* (1998, 2000) and Leporini and Pesquet (1998).

In concluding this brief discussion of different wavelet bases, it is also important to appreciate that multivariate counterparts to unidimensional wavelet bases can be developed. Multivariate

wavelet bases in $\mathbb{R}^d$ ($d \geqslant 1$) are discussed in, for example, Wojtaszczyk (1997) and the construction of the two-dimensional wavelet bases that are often used in image analysis is described in, for example, Nason and Silverman (1994), Ogden (1997) or Mallat (1998).

2.3. *The discrete wavelet transform*

In statistical settings we are more usually concerned with discretely sampled, rather than continuous, functions. It is then the wavelet analogue to the discrete Fourier transform which is of primary interest and this is referred to as the *discrete wavelet transform* (DWT).

Given a vector of function values $\mathbf{g} = (g(t_1), \ldots, g(t_n))'$ at equally spaced points t_i, the DWT of $\mathbf{g}$ is

$$\mathbf{d} = W\mathbf{g},$$

where $\mathbf{d}$ is an $n \times 1$ vector comprising both *discrete scaling coefficients* $u_{j_0 k}$ and *discrete wavelet coefficients* d_{jk}, and W is an orthogonal $n \times n$ matrix associated with the orthonormal wavelet basis chosen. The $u_{j_0 k}$ and d_{jk} are related to their continuous counterparts $c_{j_0 k}$ and w_{jk} (with an approximation error of order $1/n$) via the relationships $c_{j_0 k} \approx u_{j_0 k}/\sqrt{n}$ and $w_{jk} \approx d_{jk}/\sqrt{n}$. The factor $\sqrt{n}$ arises because of the difference between the continuous and discrete orthonormality conditions. This root factor is unfortunate but both the definition of the DWT and the wavelet coefficients are now fixed by convention: hence the different notation used to distinguish between the discrete wavelet coefficients and their continuous counterpart. Because of orthogonality of W, the inverse DWT (IDWT) is simply given by

$$\mathbf{g} = W'\mathbf{d},$$

where W' denotes the transpose of W.

If $n = 2^J$ for some positive integer J, the DWT and IDWT may be performed through an efficient algorithm developed by Mallat (1989) that requires only order n operations, so both are computationally fast. In this case, for a given j_0, under periodic boundary conditions, the DWT of $\mathbf{g}$ results in an n-dimensional vector $\mathbf{d}$ comprising both discrete scaling coefficients $u_{j_0 k}$, $k = 0, \ldots, 2^{j_0} - 1$, and discrete wavelet coefficients d_{jk}, $j = j_0, \ldots, J-1$, $k = 0, \ldots, 2^j - 1$. Often for simplicity j_0 is set to 0 as the default in many software implementations and in that case there is only one discrete scaling coefficient (the coarsest, simply labelled u_0) which is an overall weighted average of the data. However, we mention here that the choice of j_0 can play an important role in statistical applications, and that setting it to 0 is not necessarily optimal. We return to this point in subsequent sections.

We do not provide technical details here of the order n DWT algorithm mentioned above. Essentially the algorithm is a fast hierarchical scheme for deriving the required inner products which at each step involves the action of low and high pass filters, followed by decimation (the selection of every even member of a sequence). The IDWT may be similarly obtained in terms of related filtering operations. For excellent accounts of the DWT and IDWT in terms of filter operators we refer to Nason and Silverman (1995), Strang and Nguyen (1996) or Burrus *et al.* (1998). It is interesting to note that the fast DWT algorithm also appears in electrical engineering as *subband filtering*, with the filters involved being referred to as *quadrature mirror filters*. Such methods were constructed during the 1980s independently of, and before, the formal DWT framework.

2.4. Wavelet analysis versus Fourier analysis

Analogies are often made between wavelet analysis and Fourier methods and we have already briefly referred to some similarities and differences in previous sections. In one sense wavelet analysis can be seen as a 'refinement' of Fourier analysis. The key point is that the basic Fourier transform highlights the spectrum of a function, signal, image etc., but this frequency decomposition is global rather than localized, whereas the wavelet transform offers a localized frequency decomposition. It provides information not only on what frequency components are present in a signal but also when, or where, they are occurring. Wavelets therefore have significant advantages over basic Fourier analysis when the object under study is non-stationary or inhomogeneous.

There are of course alternative methods to wavelets which provide a localized frequency decomposition, the most widely used being the *windowed* or *short-term* Fourier transform. Here the signal is first restricted to an interval (with smoothed edges) by multiplying it by a fixed window function; then a Fourier analysis of the product is carried out. This process is repeated with shifted versions of the window function, thus obtaining localized frequency information throughout the signal. However, since the window width is the same for all frequencies, the amount of localization remains constant for different frequencies. Wavelet analysis provides an alternative and preferable solution, since it allows the degree of localization to be automatically and appropriately adjusted—large window widths are used for analysing low frequency components, whereas small window widths are used for investigating high frequency components. The ability to achieve very good localization in time–frequency, or scale or resolution is one of the most fundamental reasons why the use of wavelets has become so popular in the last 10 years or so, especially in signal processing and image analysis.

A formal mathematical comparison of wavelet *versus* Fourier transforms is provided by Strang (1993). Here, we illustrate the local nature of wavelets as opposed to the global nature of the basic Fourier decomposition by using an idealized musical example. The illustration also demonstrates the kind of results which arise in practice from a wavelet analysis and their interpretation, so bringing together several of the concepts discussed throughout Section 2.

Consider the 'trio for violin, cello and bass' whose 'sound score' is given in Fig. 2. Here, for simplicity, we assume that each instrument is only capable of playing a single note and alternates that with pauses, and also that the sound of a note does not weaken over time and stops immediately when a note ceases to be played. The sound of a note is a sine wave of a certain frequency. The note of the violin has the highest frequency, that of the bass the lowest whereas the cello's spectrum is between the other two. Because notes are alternated with pauses the sound signal for any instrument is inhomogeneous and so is the combined sound signal produced by the trio, which is shown in Fig. 3(a).

The spectrum of the trio's sound signal obtained from a Fourier analysis is shown in Fig. 3(b). We can clearly extract three dominant frequencies corresponding to the violin, cello and bass. However, the corresponding frequency peaks are quite flat since the whole spectrum is essentially present in the Fourier representation. This is a typical situation when a Fourier analysis is applied to inhomogeneous signals. Furthermore, although we can recognize the presence of each of the three instruments, it is not possible to tell whether they are playing together all the time or separately for some of it, and, if separately, exactly when the cello was playing, etc. This is due to the global nature of the Fourier basis.

Consider now the DWT of the trio's sound signal (see Fig. 3(c)). Here, as is conventional, the discrete wavelet coefficients d_{jk} (as described in Section 2.3) are displayed in the form of a 'pyramid'. The coefficients at the highest resolution level (the highest value of j) correspond to the lowest level of the pyramid, the next level up displays the coefficients at the second highest resolution and so on upwards through the pyramid. The position of a coefficient along a line is

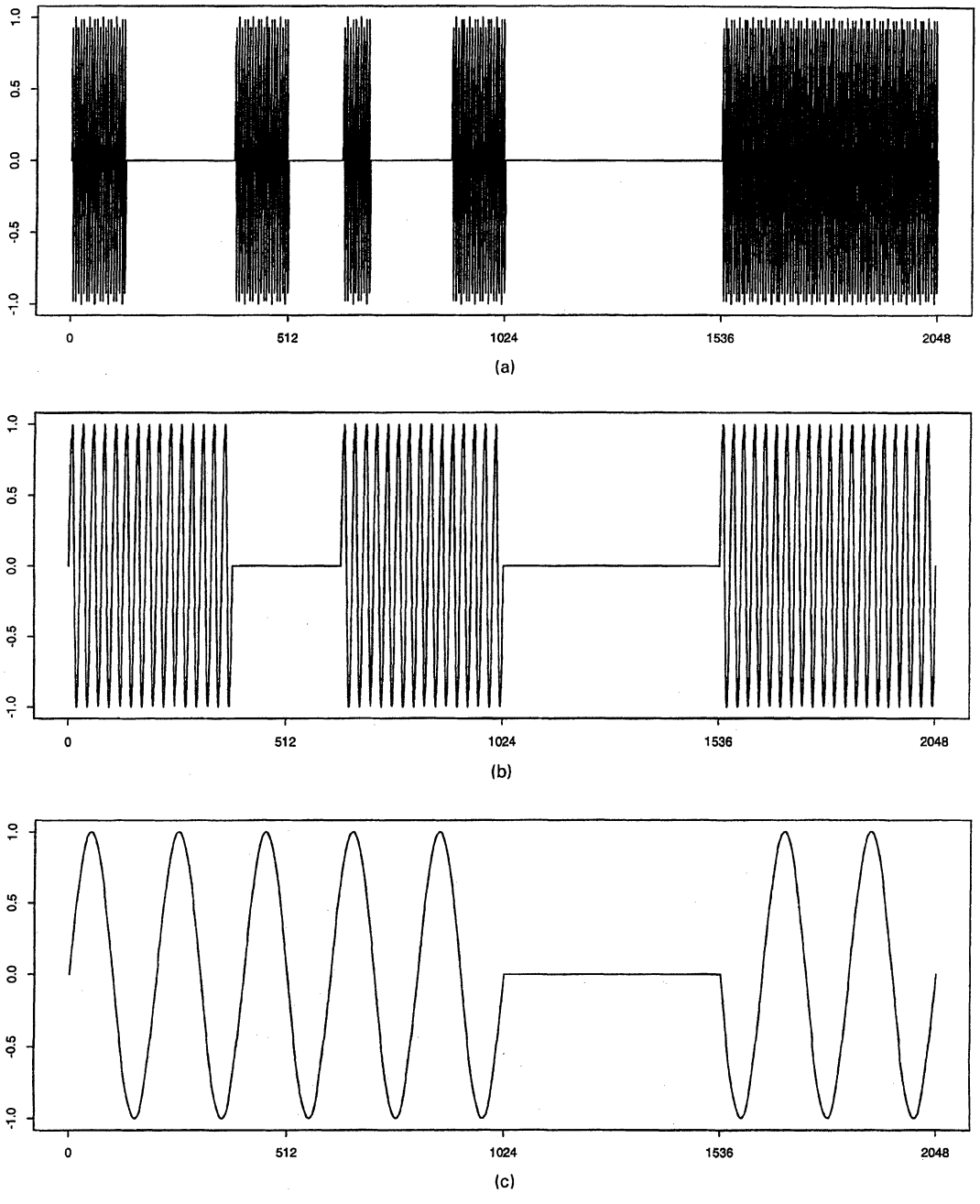

Fig. 2. Sound scores of the trio for (a) violin, (b) cello and (c) bass

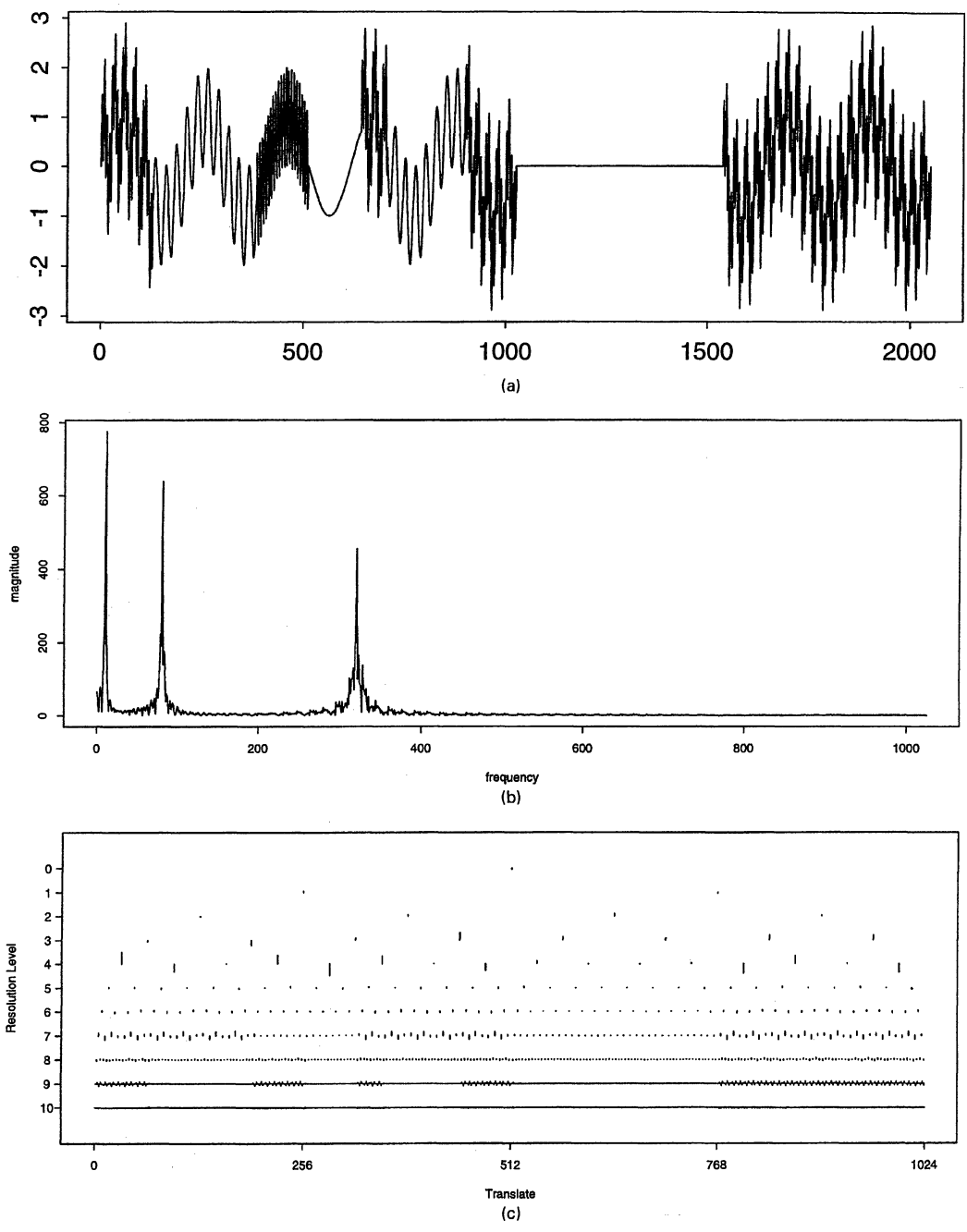

Fig. 3. (a) Combined sound of the trio, (b) spectrum of the trio's sound and (c) DWT of the trio's sound by using the extremal phase wavelet with six vanishing moments

indicative of the corresponding wavelet basis function's translate number (indexed by k). As the resolution decreases the corresponding number of coefficients is halved at each level. Coefficients are plotted as bars, up or down depending on their signs, from a reference line for each level. The sizes of the bars are relative to the magnitudes of the coefficients.

The wavelet analysis in Fig. 3(c) 'recognizes' the different instruments and separates them on different resolution levels. In this artificial example the frequency of the violin, cello and bass are respectively 320, 80 and 10, so the violin primarily appears on level 9, the cello on level 7 and the bass on level 4. In addition to separating the instruments, the magnitudes of the wavelet coefficients at different locations on the appropriate resolution level also clearly distinguish when each instrument enters and when it pauses (compare the individual sound scores in Fig. 2 and the wavelet decomposition in Fig. 3(c)). If presented solely with the wavelet analysis we might at least understand what the musical fragment might sound like, whereas that would not be so if we are presented only with the Fourier analysis of Fig. 3(b).

3. Statistical applications

Here we review some relatively well-established wavelet methods in statistical applications. These include nonparametric regression, density estimation, inverse problems, changepoint problems and some specialized aspects of time series analysis. We consider nonparametric regression in the most detail, since the basic scheme used in that case recurs with appropriate modifications in the other applications.

3.1. Nonparametric regression

Consider the standard univariate nonparametric regression setting

$$y_i = g(t_i) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{4}$$

where ϵ_i are independent $N(0, \sigma^2)$ random variables. The goal is to recover the underlying function g from the noisy data y_i without assuming any particular parametric structure for g.

One of the basic approaches to nonparametric regression is to consider the unknown function g expanded as a generalized Fourier series and then to estimate the generalized Fourier coefficients from the data. The original nonparametric problem is thus transformed to a parametric problem, although the potential number of parameters is infinite. An appropriate choice of basis for the expansion is therefore a key point in relation to the efficiency of such an approach. A 'good' basis should be parsimonious in the sense that a large set of possible response functions can be approximated well with only a few terms of the generalized Fourier expansion employed. As already discussed, wavelet series allow a parsimonious expansion for a wide variety of functions, including inhomogeneous cases. It is therefore natural to consider applying the generalized Fourier series approach by using a wavelet series.

In what follows, suppose, without loss of generality, that t_i are within the unit interval [0, 1]. For simplicity, also assume that the sample points are equally spaced, i.e. $t_i = i/n$, and that the sample size n is a power of 2: $n = 2^J$ for some positive integer J. These assumptions allow us to perform both the DWT and the IDWT using Mallat's (1989) fast algorithm. For non-equispaced or random designs, or sample sizes which are not a power of 2 or data contaminated with correlated noise, modifications will be needed to the standard wavelet-based estimation procedures discussed here (see, for example, Deylon and Juditsky (1995), Neumann and Spokoiny (1995), Wang (1996), Hall and Turlach (1997), Johnstone and Silverman (1997), Antoniadis and Pham (1998),

Cai and Brown (1998, 1999), Kovac and Silverman (2000) and von Sachs and MacGibbon (2000)).

3.1.1. *Linear estimation*

We consider first linear wavelet estimators of g. Suppose that g is expanded as a wavelet series on the interval [0, 1] as described in equation (3) with $j_0 = 0$:

$$g(t) = c_0\,\phi(t) + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} w_{jk}\,\psi_{jk}(t), \tag{5}$$

where $c_0 = \langle g, \phi\rangle$ and $w_{jk} = \langle g, \psi_{jk}\rangle$. We use equation (5) instead of equation (3) to simplify the notation in the following discussion, and because $j_0 = 0$ is the assumption in many software implementations. The modifications that are needed for the more general mathematical framework are straightforward.

Obviously we cannot estimate an infinite set of w_{jk} from the finite sample, so it is usually assumed that g belongs to a class of functions with a certain regularity. The corresponding norm of the sequence of w_{jk} is therefore finite and w_{jk} must decay to zero. Then, following the standard generalized Fourier methodology, it is assumed that

$$g(t) = c_0\,\phi(t) + \sum_{j=0}^{M}\sum_{k=0}^{2^j-1} w_{jk}\,\psi_{jk}(t) \tag{6}$$

for some $M < J$, and a corresponding *truncated* wavelet estimator $\hat{g}_M(t)$ is of the form

$$\hat{g}_M(t) = \hat{c}_0\,\phi(t) + \sum_{j=0}^{M}\sum_{k=0}^{2^j-1} \hat{w}_{jk}\,\psi_{jk}(t). \tag{7}$$

In such a setting, the original nonparametric problem essentially transforms to linear regression and the sample estimates of the scaling coefficient and the wavelet coefficients are given by

$$\begin{aligned} \hat{c}_0 &= \frac{1}{n}\sum_{i=1}^{n} \phi(t_i) y_i, \\ \hat{w}_{jk} &= \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(t_i) y_i. \end{aligned} \tag{8}$$

In practice of course these estimates may be equivalently obtained via the discrete scaling and wavelet coefficients arising from the DWT of $\mathbf{y}$ corrected by the factor $1/\sqrt{n}$ as discussed in Section 2.3.

The performance of the truncated wavelet estimator clearly depends on an appropriate choice of M. Intuitively it is clear that the 'optimal' choice of M should depend on the regularity of the unknown response function. Too small a value of M will result in an oversmoothed estimator, whereas setting $M = J - 1$ simply reproduces the data. In a way, the problem of choosing M has much in common with model selection, and associated methods such as Akaike's information criterion or cross-validation may be applied. The problem is also similar to the selection of a bandwidth in kernel estimation or the smoothing parameter in spline fitting.

However, it is also clear that truncated estimators will face difficulties in estimating inhomogeneous functions with local singularities. An accurate approximation of these singularities will require high level terms and, as a result, a large value of M, whereas the corresponding high

frequency oscillating terms will damage the estimate in smooth regions. In essence, truncation does not use the local nature of wavelet series and as a result suffers from problems similar to those of the use of a Fourier sine and cosine basis.

As an alternative, Antoniadis *et al.* (1994) and Antoniadis (1996) have suggested linear wavelet estimators, where instead of truncation the $\hat{w}_{jk}$ are linearly shrunk by appropriately chosen level-dependent factors. However, even these estimators will not be optimal for inhomogeneous functions. In fact, Donoho and Johnstone (1998) have shown that no linear estimator will be optimal (in a minimax sense) for estimating such functions. For further discussion on linear wavelet estimators we refer, for example, to Härdle *et al.* (1998) and Antoniadis (1999).

3.1.2. Non-linear estimation

In contrast with the *linear* wavelet estimators introduced above, Donoho and Johnstone (1994, 1995, 1998) and Donoho *et al.* (1995) proposed a *non-linear* wavelet estimator of g based on a reconstruction from a more judicious selection of the empirical wavelet coefficients. This approach is now widely used in statistics, particularly in signal processing and image analysis.

Owing to the orthogonality of the matrix W, the DWT of white noise is also an array of independent $N(0, \sigma^2)$ random variables, so from equation (4) it follows that

$$\hat{w}_{jk} = w_{jk} + \epsilon_{jk}, \qquad j = 0, \ldots, J-1, \quad k = 0, \ldots, 2^j - 1, \tag{9}$$

where ϵ_{jk} are independent $N(0, \sigma^2/n)$ random variables. The sparseness of the wavelet expansion makes it reasonable to assume that essentially only a few 'large' $\hat{w}_{jk}$ contain information about the underlying function g, whereas 'small' $\hat{w}_{jk}$ can be attributed to the noise which uniformly contaminates all wavelet coefficients. If we can decide which are the 'significant' large wavelet coefficients, then we can retain them and set all others equal to 0, so obtaining an approximate wavelet representation of the underlying function g.

The wavelet series truncation considered previously assumes *a priori* the significance of *all* wavelet coefficients in the first M coarsest levels and that *all* wavelet coefficients at higher resolution levels are negligible. As we have mentioned, such a strong prior assumption depends heavily on a suitable choice of M and essentially denies the possibility of local singularities in the underlying function. As an alternative to the assumptions implied by truncation, Donoho and Johnstone (1994, 1995) suggested the extraction of the significant wavelet coefficients by *thresholding* in which wavelet coefficients are set to 0 if their absolute value is below a certain threshold level, $\lambda \geqslant 0$, whose choice we discuss in more detail in Section 3.1.3. Under this scheme we obtain thresholded wavelet coefficients

$$\hat{w}^*_{jk} = \delta_\lambda(\hat{w}_{jk}),$$

where

$$\delta_\lambda(\hat{w}_{jk}) = \hat{w}_{jk}\, I(|\hat{w}_{jk}| > \lambda) \qquad \text{(hard thresholding)}, \tag{10}$$

$$\delta_\lambda(\hat{w}_{jk}) = \operatorname{sgn}(\hat{w}_{jk}) \max(0, |\hat{w}_{jk}| - \lambda) \qquad \text{(soft thresholding)}. \tag{11}$$

Thresholding allows the data themselves to decide which wavelet coefficients are significant. Hard thresholding is a 'keep' or 'kill' rule, whereas soft thresholding is a 'shrink' or 'kill' rule (Fig. 4). It has been shown that hard thresholding results in a larger variance in the function estimate, whereas soft thresholding has larger bias. To compromise the trade-off between variance and bias, Bruce and Gao (1997) suggested a *firm* thresholding that offers some advantages over both the hard and the soft variants. However, this has a drawback in that it requires two threshold

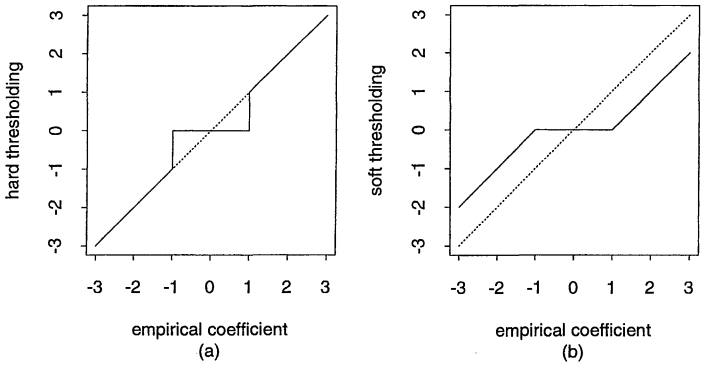

Fig. 4. (a) Hard and (b) soft thresholding rule for $\lambda = 1$

values (one for 'keep' or 'shrink' and another for 'shrink' or 'kill'), thus making the estimation procedure for the threshold values more computationally expensive.

Once obtained the thresholded wavelet coefficients $\hat{w}_{jk}^*$ are then used to obtain a selective reconstruction of the response function. The resulting estimate can be written as

$$\hat{g}_\lambda(t) = \hat{c}_0\,\phi(t) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \hat{w}_{jk}^*\,\psi_{jk}(t).$$

Often, we are interested in estimating the unknown response function at the observed data points. In this case the vector $\hat{\mathbf{g}}_\lambda$ of the corresponding estimates can be derived by simply performing the IDWT of $\{\hat{c}_0, \hat{w}_{jk}^*\}$ and the resulting three-step selective reconstruction estimation procedure can be summarized by

$$\mathbf{y} \overset{\text{DWT}}{\longrightarrow} \{\hat{c}_0, \hat{w}_{jk}\} \overset{\text{thresholding}}{\longrightarrow} \{\hat{c}_0, \hat{w}_{jk}^*\} \overset{\text{IDWT}}{\longrightarrow} \hat{\mathbf{g}}_\lambda. \tag{12}$$

As mentioned in Section 2.3, for the equally spaced design and sample size $n = 2^J$ assumed throughout this section, both the DWT and the IDWT in procedure (12) can be performed by a fast algorithm of order n, and so the whole process is computationally very efficient.

3.1.3. The choice of threshold

Clearly, an appropriate choice of the threshold value λ is fundamental to the effectiveness of the procedure described in the previous section. Too large a threshold might cut off important parts of the true function underlying the data, whereas too small a threshold retains noise in the selective reconstruction.

Donoho and Johnstone (1994) proposed the *universal* threshold

$$\lambda_{\text{un}} = \sigma\sqrt{\{2\log(n)\}}/\sqrt{n} \tag{13}$$

with σ replaced by a suitable estimate $\hat{\sigma}$ derived from the data (see later) when σ is unknown.

Despite the simplicity of such a threshold, Donoho and Johnstone (1994) showed that for both hard and soft thresholding the resulting non-linear wavelet estimator is asymptotically nearly minimax in terms of expected mean-square error of L^2-risk. Moreover, for inhomogeneous functions it outperforms any linear estimator, including the truncated and shrunk linear wavelet

estimators discussed in Section 3.1.1. The interested reader can find a rigorous mathematical treatment of these issues in Donoho and Johnstone (1994, 1995, 1998) or Härdle *et al.* (1998).

Although it has good asymptotic properties, the universal threshold depends on the data only through σ (or its estimate). Otherwise, it essentially ignores the data and cannot be 'tuned' to a specific problem at hand. In fact, for large samples it may be shown that λ_{un} will remove with high probability all the noise in the reconstruction, but part of the real underlying function might also be lost. As a result, the universal threshold tends to oversmooth in practice.

To improve the finite sample properties of the universal threshold, Donoho and Johnstone (1994) suggested that we should always retain coefficients on the first j_0 'coarse' levels even if they do not pass the threshold. Simulation results of Abramovich and Benjamini (1995) and Marron *et al.* (1998) show that the choice of j_0 may affect the mean-squared error performance for finite samples, and small or large j_0 are respectively suitable for smooth or inhomogeneous functions. In a sense j_0 plays the role of the bandwidth in kernel estimation, so theoretically its choice should be based on the smoothness of the underlying function. Hall and Patil (1996a) and Efromovich (1999) proposed to start thresholding from the resolution level $j_0 = \log_2(n)/(2r+1)$, where r is the regularity of the mother wavelet. Hall and Nason (1997) suggested that a non-integer resolution level j_0 should be considered, using a generalization of the wavelet series estimator due to Hall and Patil (1996b).

Various alternative *data-adaptive* thresholding rules have also been developed. For example, Donoho and Johnstone (1995) proposed a *SureShrink* thresholding rule based on minimizing Stein's unbiased risk estimate, whereas Weyrich and Warhola (1995), Nason (1996) and Jansen *et al.* (1997) have considered cross-validation approaches to the choice of λ. From a statistical viewpoint, thresholding is closely related to multiple-hypotheses testing, since all the coefficients are being simultaneously tested for a significant departure from 0. This view of thresholding is developed in Abramovich and Benjamini (1995, 1996) and Ogden and Parzen (1996a, b).

Further modifications of the basic thresholding scheme include *level-dependent* and *block* thresholding. In the first case, different thresholds are used on different levels, whereas in the second the coefficients are thresholded in blocks rather than individually. Both modifications imply better asymptotic properties of the resulting wavelet estimators (see, for example, Donoho and Johnstone (1995, 1998), Hall *et al.* (1998), Cai and Silverman (1998) and Cai (1999)).

Various Bayesian approaches for thresholding and non-linear shrinkage in general have also recently been proposed. These have been shown to be effective and it is argued that they are less *ad hoc* than the proposals discussed above. In the Bayesian approach a prior distribution is imposed on the wavelet coefficients. The prior model is designed to capture the sparseness of wavelet expansions that is common to most applications. Then the function is estimated by applying a suitable Bayesian rule to the resulting posterior distribution of the wavelet coefficients. Different choices of loss function lead to different Bayesian rules and hence to different non-linear shrinkages. For more details, we refer the reader to Chipman *et al.* (1997), Abramovich *et al.* (1998), Clyde and George (1998), Clyde *et al.* (1998), Crouse *et al.* (1998), Johnstone and Silverman (1998), Vidakovic (1998a) and Vannucci and Corradi (1999). Comprehensive reviews on Bayesian wavelet regression are given by Vidakovic (1998b) and Abramovich and Sapatinas (1999).

As commented earlier, in practice the noise level σ, which is needed in all these thresholding procedures, is rarely known and must therefore be estimated from the data. The most commonly employed estimate follows that used by Donoho and Johnstone (1994), where σ is robustly estimated by the median absolute deviation of the estimated wavelet coefficients at the finest level $\hat{w}_{J-1,k}$, $k = 0, \ldots, 2^{J-1} - 1$, divided by 0.6745. In the Bayesian context, a prior can be placed on σ and a hierarchical Bayesian model considered (see, for example, Clyde *et al.* (1998) and Vidakovic (1998a)).

3.1.4. An illustration of nonparametric regression

Here we use an often-quoted simulation example to illustrate the use of non-linear wavelet estimation in nonparametric regression (see Section 3.1.2) in conjunction with the standard universal threshold (see Section 3.1.3).

The 'Heavisine' function is one of four functions introduced by Donoho and Johnstone (1994, 1995) which together characterize different important features of the inhomogeneous signals arising in electrocardiogram, speech recognition, spectroscopy, seismography and other scientific fields and which are therefore useful for testing various wavelet-based estimation procedures. Fig. 5(a) shows the Heavisine function sampled at 1024 data points uniformly spaced on [0, 1], whereas Fig. 5(b) shows its wavelet decomposition using the Daubechies extremal phase wavelet with six vanishing moments. Figs 5(c) and 5(d) respectively show the Heavisine function of Fig. 5(a) corrupted with independent random noise $N(0, \sigma^2)$ (with a root signal-to-noise ratio of 7), and its corresponding wavelet decomposition based on the same wavelet. The results of using the universal threshold with σ estimated as discussed in Section 3.1.3 and the associated reconstruction of the Heavisine function are shown in Figs 5(f) and 5(e) respectively.

Because the size of wavelet coefficients can vary considerably between resolution levels, it is common to display them visually on a scale which is relative to the level rather than absolute over all levels, and this practice has been followed in Figs 5(b), 5(d) and 5(e).

3.2. Density estimation

The ideas discussed in relation to nonparametric regression can also be applied to other related statistical problems, and an obvious example is density estimation.

Let $x_1, \ldots, x_n$ be observations of an unknown density f supported on the unit interval [0, 1]. Consider the wavelet series expansion of f:

$$f(x) = c_0\,\phi(x) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} w_{jk}\,\psi_{jk}(x), \tag{14}$$

where $c_0 = \langle f, \phi \rangle$ and $w_{jk} = \langle f, \psi_{jk} \rangle$. Again, we choose $j_0 = 0$ for convenience here; in practice it may not necessarily be the best choice, but modifications for the more general case are straightforward.

For an unknown density f, sample estimates of the scaling and wavelet coefficients are obtained from the method-of-moments estimates:

$$\begin{aligned} \hat{c}_0 &= \frac{1}{n} \sum_{i=1}^{n} \phi(x_i), \\ \hat{w}_{jk} &= \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(x_i). \end{aligned} \tag{15}$$

Note that these estimates are not the same as those derived via the usual DWT in Section 3.1.1.

By analogy with the method discussed in relation to nonparametric regression in Section 3.1.1, a linear wavelet estimator for the density f is obtained by truncating the infinite series in equation (14) for some positive M:

$$\hat{f}_M(x) = \hat{c}_0\,\phi(x) + \sum_{j=0}^{M} \sum_{k=0}^{2^j-1} \hat{w}_{jk}\,\psi_{jk}(x). \tag{16}$$

Again the choice of M is a key point and may be arrived at by the use of one of several model selection procedures (for further details see, for example, Walter (1994), Tribouley (1995), Antoniadis (1999) and Vannucci and Vidakovic (1999)).

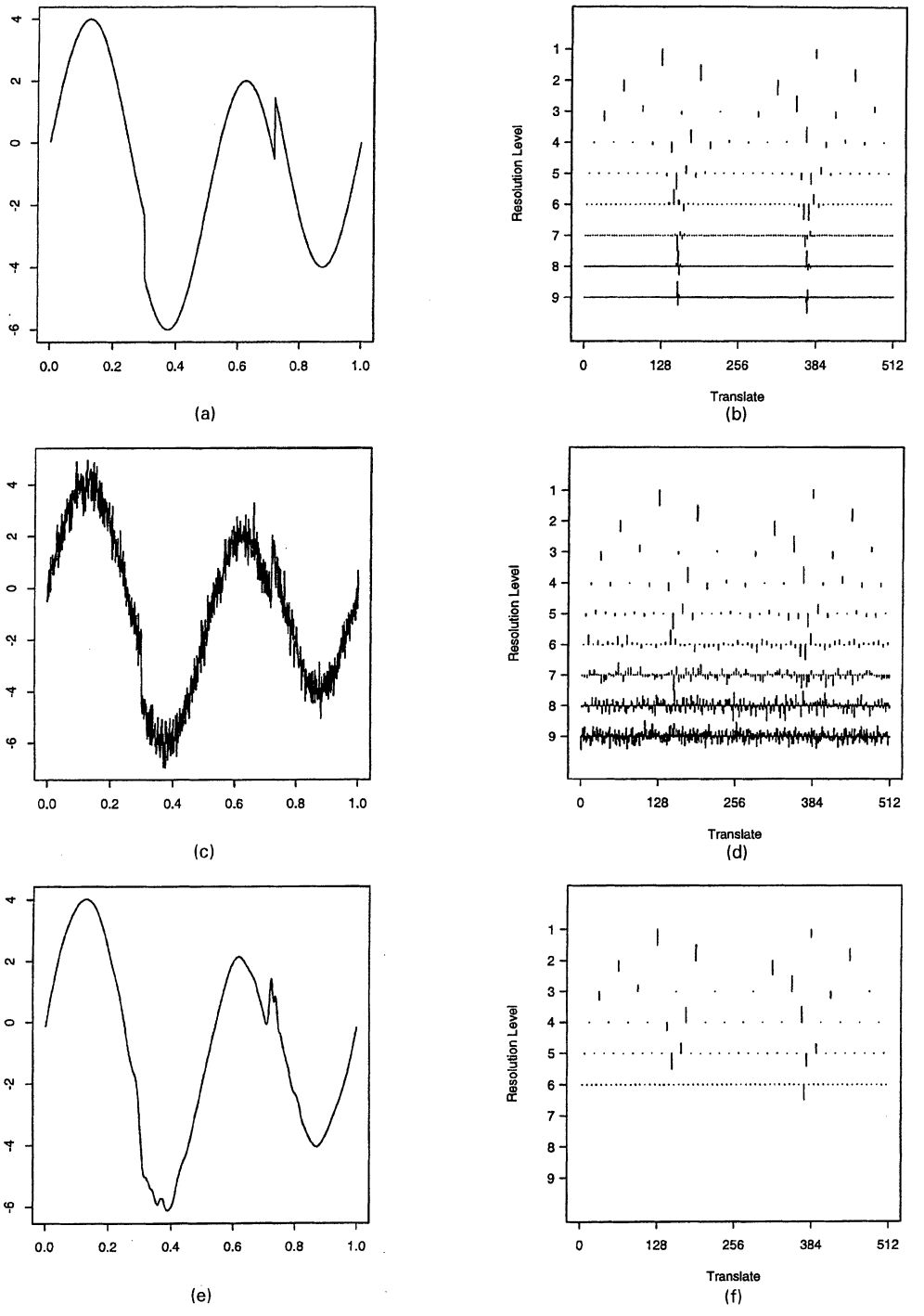

Fig. 5. (a) Heavisine function sampled at 1024 data points uniformly spaced on [0, 1]; (b) wavelet decomposition of (a) by using the extremal phase wavelet with six vanishing moments; (c) Heavisine function corrupted with independent random noise $N(0, \sigma^2)$ (root signal-to-noise ratio 7); (d) wavelet decomposition of (c) by using the extremal phase wavelet with six vanishing moments; (e) reconstruction of the Heavisine function based on (f) the universal threshold applied to (d)

As in nonparametric regression, any linear wavelet estimator (a truncated one in particular) faces difficulties when dealing with inhomogeneous densities. Donoho *et al.* (1996) and Härdle (1998) therefore proposed a non-linear thresholded estimator which is analogous to that considered for nonparametric regression in Section 3.1.2. The basic thresholding idea is the same as discussed there and the corresponding estimator takes the form

$$\hat{f}_\lambda(x) = \hat{c}_0\,\phi(x) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \hat{w}_{jk}^*\,\psi_{jk}(x). \tag{17}$$

However, it is important to note that the thresholds for density estimation need to be different from those used in nonparametric regression because the statistical model (9) for empirical wavelet coefficients does not hold in the case of density estimation. Donoho *et al.* (1995) suggested a simple threshold $\lambda = 2C\log(n)/\sqrt{n}$, with $C = \sup_{j,k}\{2^{-j/2}|\psi_{jk}(x)|\}$. As an alternative, Donoho *et al.* (1996) have considered a truncated thresholded estimator with $M = J - \log(J)$ and level-dependent thresholds $\lambda_j = A\sqrt{(j/n)}$, $j \leqslant M$, where A is an appropriately chosen constant. Both rules give rise to similar optimal properties for the corresponding wavelet estimators of the unknown density f (for rigorous definitions and statements, see Donoho *et al.* (1995, 1996) and Härdle *et al.* (1998)).

Another possible approach suggested by Donoho *et al.* (1996) is to modify the density estimation problem so that it is closer to that of nonparametric regression and then to use the thresholding techniques described in the Section 3.1.3. This is achieved by partitioning the unit interval [0, 1] into $n/4$ equally spaced intervals. Let n_i be the number of x_i falling in the ith interval and define

$$y_i = 2\sqrt{(n_i + 3/8)}.$$

Then, for large n, y_i are observations on approximately independent normal random variables with means $2\sqrt{f(i/n)}$ and unit variances. Thus, we can apply the wavelet-based thresholding procedure from the previous subsection to estimate $\sqrt{f}$ from y_i and then square the result to obtain an estimate of f.

This last method also suggests the general idea that, instead of estimating the unknown density directly, we could estimate appropriate transformations of it. For example, Pinheiro and Vidakovic (1997) have proposed an alternative method to estimate its square root. A Bayesian approach to density estimation has also been proposed by Müller and Vidakovic (1998). They parameterized the unknown density by a wavelet series on its logarithm and proposed a prior model which explicitly defines geometrically decreasing prior probabilities for non-zero wavelet coefficients at higher resolution levels.

3.3. *Inverse problems*

Some interesting scientific applications involve indirect noisy measurements. For example, the primary interest might be in a function g, but the data are only accessible from some linear transformation Kg and, in addition, are corrupted by noise. In this case, the estimation of g from indirect noisy observations $\mathbf{y} = (y_1, \ldots, y_n)'$ is often referred to in statistics as a *linear inverse problem*. Such linear inverse problems arise in a wide variety of scientific settings with different types of transformation K. Examples include applications in estimating financial derivatives, in medical imaging (the Radon transform), in magnetic resonance imaging (the Fourier transform) and in spectroscopy (convolutional transformations). Typically such problems are referred to as *ill posed* when the naïve estimate of g obtained from the inverse transform K^{-1} applied to an estimate of Kg fails to produce reasonable results because K^{-1} is an unbounded linear operator

and the presence of even small amounts of noise in the data 'blow up' when the straightforward inversion estimate is used.

3.3.1. Singular value decomposition approach

Ill-posed problems may be treated by applying regularization procedures based on a singular value decomposition (SVD); see Tikhonov and Arsenin (1977) for the general theory and O'Sullivan (1986) for a more specifically statistical discussion.

Suppose that we observe Kg at equally spaced points $t_i = i/n$, where the sample size is a power of 2. We assume that K is a linear operator and, in addition, the data are corrupted by noise, so the observed data y_i are

$$y_i = (Kg)(t_i) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{18}$$

where ϵ_i are independent $N(0, \sigma^2)$ random variables.

The basic idea of the SVD approach is the use of a pseudoinverse operator $(K^*K)^{-1}K^*$, where K^* is the adjoint operator to the operator K. The unknown g is expanded in a series of eigenfunctions e_j of the operator K^*K as say

$$g(t) = \sum_{j=1}^{\infty} \gamma_j^{-1} \langle Kg, h_j \rangle e_j(t) = \sum_{j=1}^{\infty} a_j e_j(t), \tag{19}$$

where γ_j^2 are the eigenvalues of K^*K and $h_j = Ke_j / \|Ke_j\|$. Ill-posed problems are characterized by the fact that the eigenvalues $\gamma_j^2 \to 0$. Given discrete noisy data, we find sample estimates

$$\hat{a}_j = \gamma_j^{-1} \frac{1}{n} \sum_{i=1}^{n} h_j(t_i) y_i$$

and consider the truncated SVD estimator

$$\hat{g}_M^{\mathrm{SVD}}(t) = \sum_{j=1}^{M} \hat{a}_j e_j(t)$$

for some positive M. Johnstone and Silverman (1990, 1991) showed that a properly chosen truncated SVD estimator is asymptotically the best estimator (in a minimax sense) for classes of functions that display homogeneous variation.

Note, however, that in SVD the corresponding basis is defined entirely by the operator K and essentially ignores the specific physical nature of the problem under study. For example, various stationary operators K result in the Fourier sine and cosine series which will not be appropriate if the underlying function g is inhomogeneous. For such functions, even for the optimally chosen M, difficulties will be experienced with the SVD estimator.

3.3.2. Wavelet-based decomposition approaches

In response to these limitations of the SVD, both Donoho (1995) and Abramovich and Silverman (1998) have suggested using a wavelet basis instead of the eigenfunction basis. Both of these wavelet approaches are based on similar ideas.

In the Abramovich and Silverman (1998) approach Kg is expanded in a wavelet series:

$$(Kg)(t) = c_0 \phi(t) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j - 1} w_{jk} \psi_{jk}(t), \tag{20}$$

where the scaling function ϕ and wavelets ψ_{jk} are constructed so that they belong to the range of

K for all j and k, and where $c_0 = \langle Kg, \phi \rangle$ and $w_{jk} = \langle Kg, \psi_{jk} \rangle$. As in nonparametric regression, empirical estimates $\hat{c}_0$ and $\hat{w}_{jk}$ can be obtained via the DWT of the data vector $\mathbf{y}$ and, as in equation (9), we have $\hat{w}_{jk} = w_{jk} + \epsilon_{jk}$, where ϵ_{jk} are independent $N(0, \sigma^2/n)$ random variables.

Following the general scheme discussed in Section 3.1.2, $\hat{w}_{jk}$ are thresholded via some suitable rule $\hat{w}^*_{jk} = \delta_\lambda(\hat{w}_{jk})$ providing an estimate for Kg as

$$\widehat{Kg}(t) = \hat{c}_0 \, \phi(t) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \hat{w}^*_{jk} \, \psi_{jk}(t). \tag{21}$$

To obtain an estimate of g itself, equation (21) is now mapped back by K^{-1}, obtaining

$$\hat{g}(t) = \hat{c}_0 \, \Phi(t) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \hat{w}^*_{jk} \, \Psi_{jk}(t), \tag{22}$$

where $\Phi = K^{-1}\phi$ and $\Psi_{jk} = K^{-1}\psi_{jk}$. As can be seen, the estimator (22) is essentially a 'plug-in' estimator.

In most cases of practical interest the operator K is homogeneous, i.e. for all t_0 it satisfies

$$K[g\{a(t - t_0)\}] = a^{-\alpha}(Kg)\{a(t - t_0)\}$$

for some constants a and α, where α is referred to as the index of the operator. Examples of homogeneous operators include integration, fractional integration, differentiation and (in the two-dimensional case with an appropriate co-ordinate system) the Radon transform. Ill-posed problems are characterized by a positive index for K. For homogeneous operators K we have

$$\Psi_{jk}(t) = K^{-1}\{\psi_{jk}(t)\} = 2^{\alpha j} 2^{j/2} \, \Psi(2^j t - k),$$

so Ψ_{jk} are also translations and dilations of a single mother function $\Psi = K^{-1}\psi$. Unlike wavelets, Ψ_{jk} are not mutually orthogonal and such 'wavelet-like' sets of functions have been referred to as *vaguelettes*. Because of this, the approach of Abramovich and Silverman (1998) is sometimes referred as *vaguelette–wavelet decomposition* (VWD).

A practically important example of an inverse problem is the estimation of the derivative of a function g. This fits into the above framework if we think of Kg as being Kg' in equation (18) where K is the integration operator ($\alpha = 1$). The resulting VWD estimator of g' will then be of the form

$$\hat{g}'(t) = \hat{c}_0 \, \phi'(t) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \hat{w}^*_{jk} \, \psi'_{jk}(t),$$

i.e.

$$\hat{g}'(t) = \hat{c}_0 \, \phi'(t) + \sum_{j=0}^{J-1} 2^{(3/2)j} \sum_{k=0}^{2^j-1} \hat{w}^*_{jk} \, \psi'(2^j t - k), \tag{23}$$

where $\hat{c}_0$ and $\hat{w}_{jk}$ are derived via the DWT of a data vector $\mathbf{y}$ of observed values of g.

It is intuitively clear that for ill-posed problems the estimator (21) of Kg to be 'plugged into' equation (22) should be smoother than that chosen if the purpose was simply to estimate Kg itself. Thus, a larger threshold should be used in deriving the thresholded wavelet coefficients $\hat{w}^*_{jk}$ in inverse problems than that used in nonparametric regression. Indeed, for a homogeneous operator of index α, Abramovich and Silverman (1998) have shown that a universal threshold which is asymptotically optimal (in the minimax sense) is given by

$$\lambda^\alpha_{\text{un}} = \sigma \sqrt{\{2(1 + 2\alpha) \log(n)\}} / \sqrt{n}, \tag{24}$$

which is somewhat larger than that used in standard nonparametric regression, given in equation (13) (for rigorous formulations and details, see Abramovich and Silverman (1998)). In particular, for estimation of a function's derivative in equation (23) the corresponding universal threshold would be $\sigma\sqrt{\{6\log(n)\}}/\sqrt{n}$.

In contrast with the approach of Abramovich and Silverman (1998), that of Donoho (1995) is often referred to as *wavelet–vaguelette decomposition* (WVD), since it is the unknown function g itself, rather than Kg, which is expanded in a wavelet series. One then constructs a vaguelette series of $\Psi_{jk} = K\psi_{jk}$ for Kg, finds corresponding empirical vaguelette coefficients and then applies suitable thresholding rules to them to derive a vaguelette estimator of Kg. Mapping this back by K^{-1} yields the resulting estimator of g. A good example of a practical application is given by Kolaczyk (1996) who considered in detail the WVD for the Radon transform that arises in positron emission tomography. It is important to note that, unlike wavelets, vaguelettes are not mutually orthogonal or normalized: hence vaguelette coefficients in the WVD are not in general independent; nor do they have equal variances and the thresholding procedure therefore needs to be correspondingly modified. For a homogeneous operator of index α, Donoho (1995) has shown that the variances of the vaguelette coefficients on the jth level increase like $2^{2\alpha j}$ as a direct consequence of the ill-posedness of the inverse problem and therefore that level-dependent thresholds proportional to $2^{\alpha j}$ should be used. It follows from the results in Abramovich and Silverman (1998) that level-dependent thresholds $\lambda_j = 2^{\alpha j}\lambda_{\text{un}}^{\alpha}$, where $\lambda_{\text{un}}^{\alpha}$ is defined as in equation (24), result in an optimal WVD estimator. Johnstone (1999) has also considered the WVD with the SureShrink thresholding rule mentioned in Section 3.1.3.

3.4. Changepoint problems

The good time–frequency localization of wavelets provides a natural motivation for their use in changepoint problems. Here the main goal is the estimation of the number, locations and sizes of a function's abrupt changes, such as sharp spikes or jumps. Changepoint models are used in a wide set of practical problems in quality control, medicine, economics and physical sciences. The detection of edges and the location of sharp contrast in digital pictures in signal processing and image analysis also fall within the general changepoint problem framework (see, for example, Mallat and Hwang (1992) and Ogden (1997)).

The general idea used to detect a function's abrupt changes through a wavelet approach is based on the connection between the function's local regularity properties at a certain point and the rate of decay of the wavelet coefficients near this point across increasing resolution levels (see, for example, Daubechies (1992), section 2.9, and Mallat and Hwang (1992)). Local singularities are identified by 'unusual' behaviour in the wavelet coefficients at high resolution levels at the corresponding location. Such ideas are discussed in more detail in, for example, Wang (1995) and Raimondo (1998). Bailey *et al.* (1998) have provided an example of related work in detecting transient underwater sound signals.

Antoniadis and Gijbels (2000) have also proposed an 'indirect' wavelet-based method for detecting and locating changepoints before curve fitting. Conventional function estimation techniques are then used on each identified segment to recover the overall curve. They have shown that, provided that discontinuities can be detected and located with sufficient accuracy, detection followed by wavelet smoothing enjoys optimal rates of convergence.

Bayesian approaches using wavelets have also been suggested for estimating the locations and magnitudes of a function's changepoints and are discussed by Richwine (1996) and Ogden and Lynch (1999). These place a prior distribution on a changepoint and then examine the posterior distribution of the changepoint given the estimated wavelet coefficients.

3.5. Time series analysis

Some of the wavelet techniques discussed in previous sections are clearly of potential use in various aspects of the analysis of time series and excellent overviews of wavelet methods in time series analysis generally can be found in, for example, Morettin (1996, 1997), Priestley (1996) or Percival and Walden (1999). In this section we focus on the use of wavelet methods to address some more specialized aspects of time series analysis, in particular the exploration of spectral characteristics.

3.5.1. Spectral density estimation for stationary processes

Consider a real-valued, stationary, Gaussian random process $Y_1, Y_2, \ldots$ with zero mean and covariance function $R(j) = \mathrm{cov}(Y_k, Y_{k+j})$. An important tool in the analysis of such a process is its spectral density (or power spectrum function) $f(\omega)$, which is the Fourier transform of the covariance function $R(j)$:

$$f(\omega) = R(0) + 2\sum_{j=0}^{\infty} R(j)\cos(2\pi j\omega). \tag{25}$$

Given n observations $y_1, \ldots, y_n$, the sample estimator of the spectral density is the sample spectrum or *periodogram* $I(\omega)$. This is essentially the square of the discrete Fourier transform of the data and it is typically computed at the 'fundamental' frequencies $\omega_j = j/n$, $j = 0, \ldots, n/2$, in which case

$$I(\omega_j) = \frac{1}{n}\left|\sum_{k=1}^{n} y_k \exp\{-2i\pi(k-1)\omega_j\}\right|^2, \qquad \omega_j = j/n, \quad j = 0, \ldots, n/2.$$

$I(\omega)$ is an asymptotically unbiased estimate of $f(\omega)$, but it is not a consistent estimate and it is usually suggested that the periodogram should be smoothed to estimate the spectral density. Wahba (1980) proposed as an alternative that $\log\{I(\omega)\}$ should be smoothed to estimate $\log\{f(\omega)\}$, since the log-transformation stabilizes the variance, and this modification is now commonly adopted in practice.

The conventional smoothing is based on kernel estimation, but wavelet-based techniques can also be used for estimating $\log\{f(\omega)\}$ from $\log\{I(\omega)\}$. Moulin (1993) and Gao (1997) proposed wavelet shrinkage procedures for estimating $\log\{f(\omega)\}$, and then exponentiating to obtain an estimate of $f(\omega)$. Using ideas similar to those discussed in previous sections, one expands $\log\{I(\omega)\}$ as a wavelet series, thresholds the resulting empirical wavelet coefficients $\hat{w}_{jk}$ and then performs the inverse transform to estimate $\log\{f(\omega)\}$. Wahba (1980) showed that at coarse levels $\hat{w}_{jk}$ are approximately $N(w_{jk}, \pi^2/6n)$, where w_{jk} are the 'true' wavelet coefficients of $\log\{f(\omega)\}$, so thresholds for model (9) in nonparametric regression can be applied. For high resolution levels, the above approximation does not hold and so some caution is needed when choosing appropriate thresholds for those levels. Gao (1997) suggested level-dependent thresholds:

$$\lambda_j = \max\left[\frac{\pi\sqrt{\{2\log(n)\}}}{\sqrt{(6n)}}, \frac{\log(2n)}{2^{(J-j-3)/4}}\right], \qquad j = 0, 1, \ldots, J-2, \tag{26}$$

where $n = 2^J$. The first element of the left-hand side of equation (26) is just the universal threshold (13) which is essentially used for coarse levels where Gaussian approximation applies; the second term is used for fine resolutions. The highest resolution level specified in equation (26) is $J - 2$, since the sample spectral density is computed at only $n/2$ points.

Moulin (1993) has suggested alternative level-dependent thresholds based on a saddlepoint approximation to the distribution of the sample wavelet coefficients of $\log\{I(\omega)\}$. In addition, Walden *et al.* (1998) have suggested a wavelet thresholding procedure for spectral density estimation which is based on an alternative approach to the use of the log-periodogram.

3.5.2. Wavelet analysis of non-stationary processes

The conventional theory of spectral analysis discussed in Section 3.5.1 applies only to stationary processes. The spectral characteristics of non-stationary processes change over time, so these processes do not have a spectral density as defined through equation (25). Instead, it is natural to consider the evolution of their spectral properties in time. Owing to their localization in both the time and the frequency, or scale or resolution domains, wavelets are a natural choice for this purpose. Recently, there has been growing interest in wavelet analysis of non-stationary processes, particularly in so-called 'locally stationary' processes (see, for example, Neumann (1996), von Sachs and Schneider (1996), Neumann and von Sachs (1997) and Nason *et al.* (2000)).

By analogy with the periodogram which is the square of the discrete Fourier transform of the data, we can define a *wavelet periodogram* as the square of the DWT of the data (in fact, it is preferable to define the wavelet periodogram via the *non-decimated wavelet transform* (NWT) of the data mentioned subsequently in Section 4). The wavelet periodogram may then be used as a tool for analysing how the spectral characteristics of a process change in time. As in the case of stationary time series, the wavelet periodogram needs to be denoised to obtain a consistent estimate, and that is achieved through appropriate thresholding procedures. For a discussion and more details we refer to Nason and von Sachs (1999). More general coverage of wavelet methods in the analysis of non-stationary time series may be found in Priestley (1996), Morettin (1997) or Percival and Walden (1999).

4. Extensions

Section 3 has reviewed some of the more commonly used wavelet methods in statistics. Here we very briefly discuss various extensions and developments to those techniques.

Perhaps the first point to emphasize is that in Section 3 we have focused only on unidimensional wavelet methods. As commented earlier, in Section 2, multivariate wavelet bases may be constructed (see, for example, Wojtaszczyk (1997)) and multidimensional counterparts to some of the methods discussed in Section 3 may therefore be developed. A primary use of such techniques has been in image analysis (see, for example, Nason and Silverman (1994), Ogden (1997) and Mallat (1998)).

Previous sections of this paper have also focused on the application of the DWT in various statistical problems. One particular problem with the DWT is that, unlike the discrete Fourier transform, it is not translation invariant. This can lead to Gibbs-type phenomena and other artefacts in the reconstruction of a function. Very recently, alternatives to the DWT have been developed and used in statistical applications with the objective of obtaining a better representation of the object under study (function, image, spectral density etc.). The NWT (also called the *undecimated*, or *translation invariant* or *stationary* wavelet transform) is one example.

The NWT of the data vector $\mathbf{y} = (y_1, \ldots, y_n)'$ at equally spaced points $t_i = i/n$ is defined as the set of all DWTs formed from the n possible shifts of the data by amounts i/n, $i = 1, \ldots, n$. Thus, unlike the DWT, where there are 2^j coefficients on the jth resolution level, in the NWT there are n equally spaced wavelet coefficients

$$\tilde{w}_{jk} = n^{-1} \sum_{i=1}^{n} 2^{j/2} \psi\{2^j(i/n - k/n)\} y_i, \qquad k = 0, \ldots, n-1,$$

on each resolution level j. This results in $\log_2(n)$ coefficients at each location. As an immediate consequence, the NWT enjoys translation invariance. Owing to its structure, the NWT implies a finer sampling rate at all levels (especially for low and medium resolution levels) and, as a result, provides a better exploratory tool for analysing changes in the scale or frequency behaviour of the underlying signal in time. For comparison, Figs 6(a) and 6(b) respectively show the DWT and NWT of the trio for violin, cello and bass considered in Section 2.4.

However, one should appreciate that the set of wavelet coefficients arising from the NWT is overcomplete ($n \log_2(n)$) and therefore does not imply a unique inverse transform. There are two possible approaches to the construction of an inverse transform to the NWT: one is based on choosing the 'best' shifted DWT basis among all present in the NWT, whereas the second averages all reconstructions obtained by all shifted DWTs. Fast algorithms of the order $n \log_2(n)$ for performing NWT and 'inverse' NWT exist (see, for example, Percival and Guttorp (1994), Coifman and Donoho (1995) and Nason and Silverman (1995)). The advantages of the NWT over the DWT in nonparametric regression and time series analysis are demonstrated, for example, in Coifman and Donoho (1995), Nason and Silverman (1995), Lang *et al.* (1996), Johnstone and

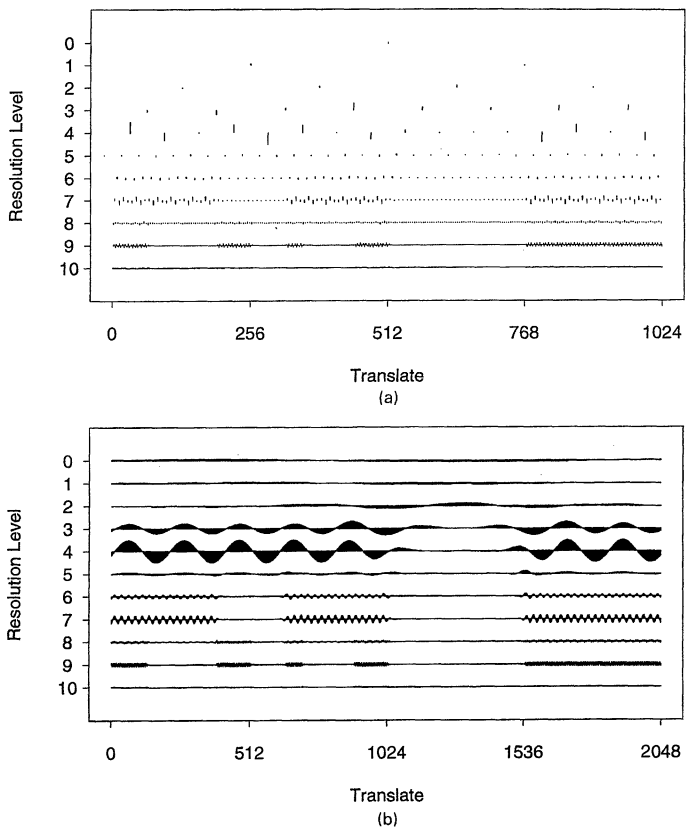

Fig. 6. (a) DWT and (b) NWT of the trio's sound in Fig. 3(a) by using the extremal phase wavelet with six vanishing moments

Silverman (1997, 1998), Percival and Mofjeld (1997), Berkner and Wells (1998) and Nason *et al.* (2000).

Further and less-well-developed extensions to standard wavelet analyses in statistical applications also exist which we do not follow up here. The reader is referred, for example, to Donoho (1993), McCoy and Walden (1996), Serroukh *et al.* (1998), Antoniadis *et al.* (1999), Percival and Walden (1999) and Silverman (1999) for discussions of some of these possibilities. We should particularly mention that, although wavelets are the best-known alternative to a traditional Fourier representation, they are not only one. Multiple wavelets, wavelet packets, cosine packets, chirplets and warplets are examples from the ever growing list of available waveforms of different types that may be used to decompose a function in interesting and informative ways. As Mallat (1998) pointed out, there is a danger that creating new basis families may become just 'a popular new sport' if it is not motivated by applications. However, some of these techniques have already found use in statistical problems (see, for example, Downie and Silverman (1998), Walden and Contreras Cristan (1998) and Nason and Sapatinas (1999)) and that trend will undoubtedly continue.

5. Conclusions

The objective of this paper was to increase familiarity with basic wavelet methods in statistical analysis, and to provide a source of references to those involved in the kinds of data analysis where wavelet techniques may be usefully employed. We hope that we have demonstrated that wavelets can be, and are being, used to good effect to address substantial problems in a variety of areas, and also that such methods have reached the stage where they may be considered another standard tool of the applied statistician, rather than a research novelty.

We also emphasize that wavelet methods are now easily accessible. We have deliberately not referred much in this paper to specific wavelet software, for the simple reason that, like most other software, once described it tends to be superseded by more versatile and powerful alternatives. Nevertheless we should make it clear that there are many readily available wavelet-based software packages which are being continuously developed. Among those commonly used by statisticians the most versatile include `WaveThresh` (see Nason and Silverman (1994)), `S+Wavelets` (see Bruce and Gao (1996)) and `Wavelab` (see Buckheit and Donoho (1995)). The first two are designed to run in conjunction with S-PLUS, the latter within MatLab.

Finally, in summary to what has been a fairly lengthy and wide ranging paper, we note that the numerous references included clearly demonstrate a wide and growing interest in the use of wavelets in statistics. There have already been three major international conferences on wavelet methods in statistics (at Grenoble, in November 1994, Canberra, in June 1995, and Durham, North Carolina, in October 1997) and two edited volumes (Antoniadis and Oppenheim, 1995; Müller and Vidakovic, 1999). Wavelets have also been the subject of several important sessions in other major international conferences or specialized discussion meetings (e.g. 'Wavelets: the key to intermittent information', organized by the Royal Society, London, in February 1999). Such high profile exposure demonstrates the scientific importance that is attached to the subject and a significant part of that interest is concerned with statistical applications. Although there is scope for the development and refinement of wavelet-based methodologies in statistics, there is no doubt that the 'smooth wiggly function' is here to stay.

Acknowledgements

We would like to thank Dr Guy Nason for helpful discussions on an earlier version of this paper. The work was partially supported by Engineering and Physical Sciences Research Council grant

GR/K70236 relating to Dr Theofanis Sapatinas's previous appointment as a Research Associate in the School of Mathematics at the University of Bristol, UK.

References

Abramovich, F. and Benjamini, Y. (1995) Thresholding of wavelet coefficients as multiple hypotheses testing procedure. *Lect. Notes Statist.*, **103**, 5–14.

———(1996) Adaptive thresholding of wavelet coefficients. *Comput. Statist. Data Anal.*, **22**, 351–361.

Abramovich, F. and Sapatinas, T. (1999) Bayesian approach to wavelet decomposition and shrinkage. *Lect. Notes Statist.*, **141**, 33–50.

Abramovich, F., Sapatinas, T. and Silverman, B. W. (1998) Wavelet thresholding via a Bayesian approach. *J. R. Statist. Soc.* B, **60**, 725–749.

———(2000) Random functions defined by expansions in an overcomplete wavelet dictionary. *Probab. Theory Reltd Flds*, to be published.

Abramovich, F. and Silverman, B. W. (1998) The vaguelette-wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85**, 115–129.

Aldroubi, A. and Unser, M. (eds) (1996) *Wavelets in Medicine and Biology.* Boca Raton: CRC.

Anderson, L., Hall, N., Jawerth, B. and Peters, G. (1993) Wavelets on closed subsets of the real line. In *Recent Advances in Wavelet Analysis* (eds L. L. Schumaker and G. Webb), pp. 1–61. New York: Academic Press.

Antoniadis, A. (1996) Smoothing noisy data with tapered coiflets series. *Scand. J. Statist.*, **23**, 313–330.

———(1999) Wavelets in statistics: a review (with discussion). *J. Ital. Statist. Soc.*, **6**, in the press.

Antoniadis, A. and Gijbels, I. (2000) Detecting abrupt changes by wavelet methods. *J. Nonparam. Statist.*, **10**, in the press.

Antoniadis, A., Grégoire, G. and McKeague, I. W. (1994) Wavelet methods for curve estimation. *J. Am. Statist. Ass.*, **89**, 1340–1353.

Antoniadis, A., Grégoire, G. and Nason, G. (1999) Density and hazard rate estimation for right-censored data by using wavelet methods. *J. R. Statist. Soc.* B, **61**, 63–84.

Antoniadis, A. and Oppenheim, G. (eds) (1995) Wavelets and statistics. *Lect. Notes Statist.*, **103**.

Antoniadis, A. and Pham, D. T. (1998) Wavelet regression for random or irregular design. *Comput. Statist. Data Anal.*, **28**, 353–369.

Bailey, T. C., Sapatinas, T., Powell, K. J. and Krzanowski, W. J. (1998) Signal detection in underwater sound using wavelets. *J. Am. Statist. Ass.*, **93**, 73–83.

Berkner, K. and Wells, Jr, R. O. (1998) Smoothness estimates for soft-threshold denoising via translation invariant wavelet transforms. *Technical Report 98-01*. Computational Mathematics Laboratory, Rice University, Houston.

Bruce, A. G. and Gao, H.-Y. (1996) *Applied Wavelet Analysis with S-Plus*. New York: Springer.

———(1997) WaveShrink with firm shrinkage. *Statist. Sin.*, **4**, 855–874.

Buckheit, J. B. and Donoho, D. L. (1995) WaveLab and reproducible research. *Lect. Notes Statist.*, **103**, 55–81.

Burrus, C. S., Gonipath, R. A. and Guo, H. (1998) *Introduction to Wavelets and Wavelet Transforms: a Primer.* Englewood Cliffs: Prentice Hall.

Cai, T. T. (1999) Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.*, **27**, 898–924.

Cai, T. T. and Brown, L. D. (1998) Wavelet shrinkage for nonequispaced samples. *Ann. Statist.*, **26**, 1783–1799.

———(1999) Wavelet estimation for samples with random uniform design. *Statist. Probab. Lett.*, **42**, 313–321.

Cai, T. T. and Silverman, B. W. (1998) Incorporating information on neighboring coefficients into wavelet estimation. *Technical Report 98-13*. Department of Statistics, Purdue University, West Lafayette.

Chipman, H. A., Kolaczyk, E. D. and McCulloch, R. E. (1997) Adaptive Bayesian wavelet shrinkage. *J. Am. Statist. Ass.*, **92**, 1413–1421.

Chui, C. K. (1992) *An Introduction to Wavelets*. New York: Academic Press.

Clyde, M. and George, E. I. (1998) Robust empirical Bayes estimation in wavelets. *Discussion Paper 98-21*. Institute of Statistics and Decision Sciences, Duke University, Durham.

Clyde, M., Parmigiani, G. and Vidakovic, B. (1998) Multiple shrinkage and subset selection in wavelets. *Biometrika*, **85**, 391–401.

Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993a) Multiresolution analysis, wavelets and fast algorithms on an interval. *C. R. Acad. Sci. Paris*, Ser. I, **316**, 417–421.

Cohen, A., Daubechies, I. and Vial, P. (1993b) Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harm. Anal.*, **1**, 54–81.

Coifman, R. R. and Donoho, D. L. (1995) Translation-invariant de-noising. *Lect. Notes Statist.*, **103**, 125–150.

Crouse, M. S., Nowak, R. D. and Baraniuk, R. G. (1998) Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Process.*, **46**, 886–902.

Daubechies, I. (1988) Orthonormal bases of compactly supported wavelets. *Communs Pure Appl. Math.*, **41**, 909–996.

———(1992) *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics.

———(1993) Orthonormal bases of compactly supported wavelets: II, Variations on a theme. *SIAM J. Math. Anal.*, **24**, 499–519.

Delyon, B. and Juditsky, A. (1995) Estimating wavelet coefficients. *Lect. Notes Statist.*, **103**, 151–168.
Donoho, D. L. (1993) Non-linear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data. *Proc. Symp. Appl. Math.*, **47**, 173–205.
———(1995) Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harm. Anal.*, **2**, 101–126.
Donoho, D. L. and Johnstone, I. M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.
———(1995) Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Ass.*, **90**, 1200–1224.
———(1998) Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**, 879–921.
Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia (with discussion)? *J. R. Statist. Soc.* B, **57**, 301–369.
———(1996) Density-estimation by wavelet thresholding. *Ann. Statist.*, **24**, 508–539.
Downie, T. R. and Silverman, B. W. (1998) The discrete multiple wavelet transform and thresholding methods. *IEEE Trans. Signal Process.*, **46**, 2558–2561.
Efromovich, S. (1999) Quasi-linear wavelet estimation. *J. Am. Statist. Ass.*, **94**, 189–204.
Gao, H.-Y. (1997) Choice of thresholds for wavelet shrinkage estimate of the spectrum. *J. Time Ser. Anal.*, **18**, 231–251.
Haar, A. (1910) Zur Theorie der orthoganalen Funktionensysteme. *Math. Ann.*, **69**, 331–371.
Hall, P., Kerkyacharian, G. and Picard, D. (1998) Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.*, **26**, 922–942.
Hall, P. and Nason, G. P. (1997) On choosing a non-integer resolution level when using wavelet methods. *Statist. Probab. Lett.*, **34**, 5–11.
Hall, P. and Patil, P. (1996a) Effect of threshold rules on performance of wavelet-based curve estimators. *Statist. Sin.*, **6**, 331–345.
———(1996b) On the choice of smoothing parameter threshold and truncation in nonparametric regression by non-linear wavelet methods. *J. R. Statist. Soc.* B, **58**, 361–377.
Hall, P. and Turlach, B. A. (1997) Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Ann. Statist.*, **25**, 1912–1925.
Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998) Wavelets, approximation, and statistical applications. *Lect. Notes Statist.*, **129**.
Jansen, M., Malfait, M. and Bultheel, A. (1997) Generalised cross validation for wavelet thresholding. *Signal Process.*, **56**, 33–44.
Jawerth, B. and Sweldens, W. (1994) An overview of wavelet based multiresolution analyses. *SIAM Rev.*, **36**, 377–412.
Johnstone, I. M. (1999) Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statist. Sin.*, **9**, 51–84.
Johnstone, I. M. and Silverman, B. W. (1990) Speed estimation in positron emission tomography and related inverse problems. *Ann. Statist.*, **18**, 251–280.
———(1991) Discretization effects in statistical inverse problems. *J. Complex.*, **7**, 1–34.
———(1997) Wavelet threshold estimators for data with correlated noise. *J. R. Statist. Soc.* B, **59**, 319–351.
———(1998) Empirical Bayes approaches to mixture problems and wavelet regression. *Technical Report*. School of Mathematics, University of Bristol, Bristol.
Kolaczyk, E. D. (1996) A wavelet shrinkage approach to tomographic image reconstruction. *J. Am. Statist. Ass.*, **91**, 1079–1090.
Kovac, A. and Silverman, B. W. (2000) Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *J. Am. Statist. Ass.*, **95**, in the press.
Lang, M., Guo, H., Odegard, J. E., Burrus, C. S. and Wells, Jr, R. O. (1996) Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Process. Lett.*, **3**, 10–12.
Leporini, D. and Pesquet, J.-C. (1998) Multiscale regularization in Besov spaces. In *Proc. 32nd Asilomar Conf. Signals, Systems, and Computers*, vol. 1, pp. 723–726. Pacific Grove.
Mallat, S. G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattn Anal. Mach. Intell.*, **11**, 674–693.
———(1998) *A Wavelet Tour of Signal Processing*. San Diego: Academic Press.
Mallat, S. G. and Hwang, W. L. (1992) Singularity detection and processing with wavelets. *IEEE Trans. Inf. Theory*, **38**, 617–643.
Marron, J. S., Adak, S., Johnstone, I. M., Neumann, M. H. and Patil, P. (1998) Exact risk analysis of wavelet regression. *J. Comput. Graph. Statist.*, **7**, 278–309.
McCoy, E. J. and Walden, A. T. (1996) Wavelet analysis and synthesis of stationary long-memory processes. *J. Comput. Graph. Statist.*, **5**, 26–56.
Meyer, Y. (1985) Principe d'incertitude, bases hilbertiennes et algébres d'opérateurs. *Sém. Bourbaki*, 1985/86, no. 662.
———(1992) *Wavelets and Operators*. Cambridge: Cambridge University Press.
———(1993) *Wavelets: Algorithms & Applications*. Philadelphia: Society for Industrial and Applied Mathematics.
Morettin, P. A. (1996) From Fourier to wavelet analysis of time series. In *Proc. Computational Statistics* (ed. A. Pratt), pp. 111–122, New York: Physica.
———(1997) Wavelets in statistics. *Resenhas*, **3**, 211–272.
Moulin, P. (1993) Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Signal Process.*, **42**, 3126–

3136.
Müller, P. and Vidakovic, B. (1998) Bayesian inference with wavelets: density estimation. *J. Comput. Graph. Statist.*, **7**, 456–468.
———(eds) (1999) Bayesian inference in wavelet based models. *Lect. Notes Statist.*, **141**.
Nason, G. P. (1996) Wavelet shrinkage using cross-validation. *J. R. Statist. Soc.* B, **58**, 463–479.
Nason, G. P. and von Sachs, R. (1999) Wavelets in time series analysis. *Phil. Trans. R. Soc. Lond.* A, **357**, 2511–2526.
Nason, G. P., von Sachs, R. and Kroisandt, G. (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc.* B, **62**, 271–292.
Nason, G. P. and Sapatinas, T. (1999) Wavelet packet modelling of nonstationary wind energy time series. *Technical Report 99/31*. Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury.
Nason, G. P. and Silverman, B. W. (1994) The discrete wavelet transform in S. *J. Comput. Graph. Statist.*, **3**, 163–191.
———(1995) The stationary wavelet transform and some statistical applications. *Lect. Notes Statist.*, **103**, 281–300.
Neumann, M. H. (1996) Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.*, **17**, 601–633.
Neumann, M. H. and von Sachs, R. (1997) Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.*, **25**, 38–76.
Neumann, M. H. and Spokoiny, V. (1995) On the efficiency of wavelet estimators under arbitrary error distributions. *Math. Meth. Statist.*, **4**, 137–166.
Ogden, R. T. (1997) *Essential Wavelets for Statistical Applications and Data Analysis*. Boston: Birkhäuser.
Ogden, R. T. and Lynch, J. D. (1999) Bayesian analysis of change-point models. *Lect. Notes Statist.*, **141**, 67–82.
Ogden, R. T. and Parzen, E. (1996a) Change-point approach to data analytic wavelet thresholding. *Statist. Comput.*, **6**, 93–99.
———(1996b) Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Comput. Statist. Data Anal.*, **22**, 53–70.
O'Sullivan, F. (1986) A statistical perspective on ill-posed inverse problems (with discussion). *Statist. Sci.*, **1**, 502–527.
Percival, D. B. and Guttorp, P. (1994) Long-memory processes, the Allan variance and wavelets. In *Wavelets in Geophysics* (eds E. Foufoula-Georgiou and P. Kumar), pp. 325–344. San Diego: Academic Press.
Percival, D. B. and Mofjeld, H. (1997) Multiresolution and variance analysis using wavelets with application to subtidal coastal seas. *J. Am. Statist. Ass.*, **92**, 868–880.
Percival, D. B. and Walden, A. T. (1999) *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.
Pinheiro, A. and Vidakovic, B. (1997) Estimating the square root of a density via compactly supported wavelets. *Comput. Statist. Data Anal.*, **25**, 399–415.
Priestley, M. B. (1996) Wavelets and time-dependent spectral analysis. *J. Time Ser. Anal.*, **17**, 85–104.
Raimondo, M. (1998) Minimax estimation of sharp change points. *Ann. Statist.*, **26**, 1379–1397.
Richwine, J. (1996) Bayesian estimation of change-points using Haar wavelets. *MSc Thesis*. Department of Statistics, University of South Carolina, Columbia.
von Sachs, R. and MacGibbon, B. (2000) Nonparametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Statist.*, **27**, in the press.
von Sachs, R. and Schneider, K. (1996) Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *Appl. Comput. Harm. Anal.*, **3**, 268–282.
Serroukh, A., Walden, A. T. and Percival, D. B. (1998) Statistical properties of the wavelet variance estimator for non-Gaussian/non-linear time series. *Technical Report 98-03*. Department of Mathematics, Imperial College of Science, Technology and Medicine, London.
Silverman, B. W. (1999) Wavelets in statistics: beyond the standard assumptions. *Phil. Trans. R. Soc. Lond.* A, **357**, 2459–2474.
Strang, G. (1993) Wavelet transforms versus Fourier transforms. *Bull. Am. Math. Soc.*, new ser., **28**, 288–305.
Strang, G. and Nguyen, T. (1996) *Wavelets and Filter Banks*. Wellesley: Wellesley–Cambridge Press.
Strömberg, J. O. (1981) A modified Franklin system and higher order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces. In *Proc. Conf. in Honor of Antoni Zygmund* (eds W. Beckner, A. P. Calderón, R. Fefferman and P. W. Jones), pp. 475–493. New York: Wadsworth.
Tikhonov, A. N. and Arsenin, V. Y. (1977) *Solutions of Ill-posed Problems*. New York: Wiley.
Tribouley, K. (1995) Practical estimation of multivariate densities using wavelet methods. *Statist. Neerland.*, **49**, 41–62.
Vannucci, M. and Corradi, F. (1999) Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. *J. R. Statist. Soc.* B, **61**, 971–986.
Vannucci, M. and Vidakovic, B. (1999) Preventing the Dirac disaster: wavelet based density estimation. *J. Ital. Statist. Soc.*, **6**, in the press.
Vidakovic, B. (1998a) Non-linear wavelet shrinkage with Bayes rules and Bayes factors. *J. Am. Statist. Ass.*, **93**, 173–179.
———(1998b) Wavelet based nonparametric Bayes methods. *Lect. Notes Statist.*, **133**, 133–155.
———(1999) *Statistical Modeling by Wavelets*. New York: Wiley.
Wahba, G. (1980) Automatic smoothing of the log periodogram. *J. Am. Statist. Ass.*, **75**, 122–132.
Walden, A. T. and Contreras Cristan, A. (1998) The phase-corrected undecimated discrete wavelet packet transform and its applications to interpreting the timing of events. *Proc. R. Soc. Lond.* A, **454**, 2243–2266.

Walden, A. T., Percival, D. B. and McCoy, E. J. (1998) Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Trans. Signal Process.*, **46**, 3153–3165.
Walter, G. G. (1994) *Wavelets and Other Orthogonal Systems with Applications*. Boca Raton: CRC.
Wang, Y. (1995) Jump and sharp cusp detection by wavelets. *Biometrika*, **82**, 385–397.
———(1996) Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.*, **24**, 466–484.
Weyrich, N. and Warhola, G. T. (1995) De-noising using wavelets and cross-validation. *NATO Adv. Study Inst.* C, **454**, 523–532.
Wojtaszczyk, P. (1997) *A Mathematical Introduction to Wavelets*. Cambridge: Cambridge University Press.
Young, R. K. (1993) *Wavelet Theory and Its Applications*. Boston: Kluwer.