Statistical modelling of time series using non-decimated wavelet representations

G.P. Nason[†], T. Sapatinas
Department of Mathematics, University of Bristol, U.K.
and
A. Sawczenko
Institute of Child Health,
Royal Hospital for Sick Children, Bristol, U.K.

17th September 1997

Abstract

This article proposes the use of time-ordered non-decimated wavelet or non-decimated wavelet packet coefficients to provide a representation of a time series (explanatory). The resulting representations are then used as variables in a statistical model to provide predictions of another time series (response). The statistical model provides valuable information about which components in the explanatory time series drive the response time series.

To represent our time series we use a collection of basis functions known as wavelet packets. Our methodology modifies the standard wavelet packets representation by providing an over-determined representation of shifted wavelet packets where the shifts are not restricted to the usual wavelet grid. We introduce a fast algorithm for carrying out the time-ordered non-decimated wavelet packet transform.

The shifted wavelet packet bases can represent many classes of time series sparsely. The sparsity provides an effective dimension reduction which enables standard statistical modelling techniques to select the components of an explanatory series that best explain the variation in the response time series.

Our modelling methodology is illustrated using time series examples from two different arenas: (a) a biomedical example shows how infant sleep states can be successfully classified using the time-ordered non-decimated wavelet packet transform of heart rate and linear discriminant analysis; (b) a renewable energy example using a generalized linear model relating wind speeds at 40m to a time-ordered non-decimated wavelet packet transform of wind speeds at 10m.

[†]*Address for correspondence:* Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, England

Keywords: TIME SERIES MODELLING; WAVELETS; DIMENSION REDUCTION; BIOMEDICAL TIME SERIES; WIND TIME SERIES; LINEAR DISCRIMINANT ANALYSIS; GENERALIZED LINEAR MODELS

1 Introduction

Suppose we are interested in the relationship between a response time series, Y_t, and an explanatory time series, X_t, $t = 1, \ldots, N$. We might be interested in predicting Y_t given future values of X_t (suppose, for example, Y_t is expensive to collect but X_t is cheap). In addition we might also be interested in learning about which components of X_t influence the behaviour of Y_t. The route that we take represents X_t in terms of basis functions called wavelet packets. In fact, we use a collection of all possible time-ordered shifted wavelet packet bases (a non-decimated wavelet packet transform, NWPT) and use statistical methods for selecting the best of these packets to represent the explanatory time series for our intended purpose. The time-ordered NWPT produces a matrix $\mathbf{X}$ of N observations on K variables, where each variable contains coefficients of shifted wavelet packet basis functions. Occasionally we shall use instead a simpler version of the time-ordered NWPT called the (time-ordered) non-decimated wavelet transform (NWT). The great advantage of non-decimated wavelet representations is that they permit the use of many popular standard statistical techniques to model the response, $\mathbf{Y} = (Y_1, \ldots, Y_N)^T$, in terms of the data matrix, $\mathbf{X}$. The wavelet packet basis library contains a rich, organised, collection of basis functions that permit it to represent a wide range of time series sparsely. The sparsity of the time series representation is a sort of (deterministic) dimension reduction — the wavelets compress the time series into fewer coefficients which can be handled more efficiently by existing statistical methods.

Our motivating example is a biomedical time series arising in the field of child health. Many pathological phenomena, such as breathing problems, are most likely to occur during sleep. Figure 1 shows a segment of two time series recorded from a four month old infant who was placed to bed at night. The series shown are of heart rate, X_t, and sleep state, Y_t, (i.e. whether the infant was awake or asleep) sampled every 30 seconds. After 1.5 hours the infant eventually fell asleep only to wake around an hour later. Heart rate was automatically measured using a standard commercial ECG (electrocardiogram) monitor. Sleep state was manually determined the next day by a trained observer visually interpreting each 30 second period of the infant's EEG (electro-encephalogram — "brain-waves") and EOG (electro-oculogram — eye movements) that had been concurrently recorded, see Anders *et al.* (1971). Whilst this is an accurate and reproducible method of sleep state analysis (about 80% inter-observer agreement) the determination is time-consuming, laborious and expensive. The attachment of the recording sensors to the infants face (EOG) and scalp (EEG) may be distressing to both parents and infants, may lead to artifacts by interfering with the infants sleep, and is not practicable in the home environment. Thus such recordings must be performed in the hospital which further adds to cost and potential distress. By comparison, ECG recording is relatively unobtrusive, since the leads are attached to the infants chest and parents can be readily taught to do this. Thus a method which can

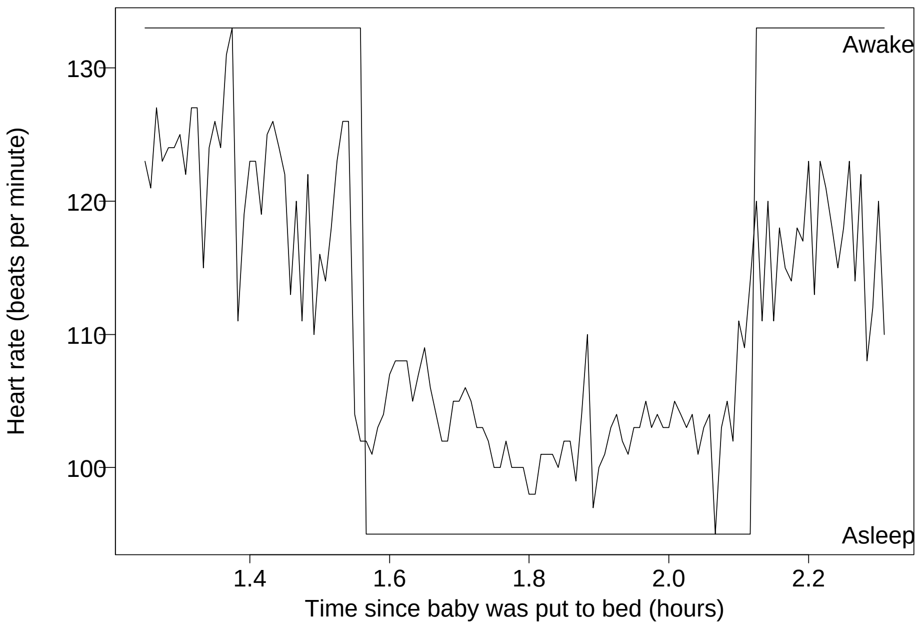

Figure 1: *Time-series of heart rate and sleep state for a four month old baby beginning at 23:09:42. Heart-rate is labelled by the left-hand axis. Sleep-state is labelled by the right-hand axis and takes only two states: asleep and awake.*

reliably predict sleep state from heart rate would be clinically valuable - our method does just this.

Figure 1 shows that the heart rate is low when the baby is asleep and high when it is awake. Therefore, the mean level of heart rate over certain time scales is likely to be important for determining sleep state. As well as prediction our model gives us valuable information about the relationship between sleep state and heart rate. The statistical model tells us which sorts of non-decimated wavelet packet functions (which sorts of components of X_t) are most important for classifying sleep state.

The paper is organised as follows. Section 2 describes the NWPT together with some important special cases, and introduces the time-ordered NWPT necessary for our modelling methodology. Section 3 shows how standard statistical modelling can be used to model a response time series, Y_t, in terms of a time-ordered NWPT of an explanatory time series, X_t, and outlines the advantages of the time-ordered NWPT over other methods. Section 4 illustrates our new methodology on two examples: the above biomedical example is discussed more fully in Section 4.1; an example to do with wind speed modelling appears in Section 4.2. Discussion of some aspects of the methodology and ideas for future work are presented in Section 5 and we conclude in Section 6.

2 Wavelets and discrete transforms

Wavelets have been found to be useful in a number of disciplines such as physics, signal processing and geophysics. In statistics wavelets have mainly been used in: *curve estimation*, see Donoho and Johnstone (1994a; 1994b; 1995), Donoho *et al.* (1995; 1996), Hall and Patil (1995), Abramovich and Benjamini (1996), Antoniadis (1996), Nason (1996), Ogden and Parzen (1996), Abramovich *et al.* (1997), Chipman *et al.* (1997), Clyde *et al.* (1997), Hall and Nason (1997), Johnstone and Silverman (1997), Neuman and von Sachs (1997) and Crouse *et al.* (1998); *time series analysis*, see Moulin (1994), McCoy *et al.* (1995), von Sachs (1996), von Sachs and Schneider (1996), von Sachs *et al.* (1996; 1997) and Gao (1997); *survival data analysis*, see Antoniadis *et al.* (1997); *inverse problems*, see Donoho (1995) and Abramovich and Silverman (1997); *classification*, see Saito (1994), Coifman and Saito (1994), Buckheit and Donoho (1995), Learned and Willsky (1995) and Saito and Coifman (1996). For a general statistical introduction to wavelets see Nason and Silverman (1994) and Bruce and Gao (1996), or see Strang (1993) for a more mathematical view. More detailed comprehensive expositions are Daubechies (1992), Meyer (1992) and Chui (1992).

Wavelets are a type of building block for constructing functions. More precisely wavelets form bases for function spaces such as $L^2(\mathbb{R})$. This article only considers orthogonal wavelets. Functions $f(t)$ will be represented by expansions such as

$$f(t) = \sum_{k \in \mathbb{Z}} c_k \phi_{0k}(t) + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(t), \tag{1}$$

where wavelets, $\psi_{jk}(t)$, are represented in terms of a *mother wavelet*, $\psi(t)$, using the formula

$$\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k) \tag{2}$$

for integers k and $j \geq 0$. The *resolution level*, j, controls the scale of the wavelet, the *translation number*, k, controls the location of the wavelet. The functions $\phi_{0k}(t)$ are obtained from a *father wavelet* $\phi(t)$ using formula (2) with $j = 0$. Typically the father wavelet looks like a statistical kernel function and the mother wavelet is a localised oscillation (hence the name wavelet). Expansion (1) can be characterised in words by: the function, $f(t)$, consists of a smooth part represented in (1) by integer translated father wavelets (kernel functions) and detail at different scales (j) represented by the wavelets $\psi_{j\cdot}(t)$. A wavelet packets representation is similar to that given in (1) except that the basis functions have an extra degree of flexibility: the number of oscillations of the basis function.

Typically a wavelet $\psi(t)$ will be localised around the origin. This implies that the derived wavelets $\psi_{jk}(t)$ are located near $2^{-j}k$. This means that for standard wavelets the location of the basis functions depends on the resolution level j (if j is large then the wavelets will be close together and each will oscillate very rapidly enabling them to represent high-frequency rapidly changing phenomena. If j is small the opposite happens). For *non-decimated* wavelets formula (2) for $\psi_{jk}(t)$ changes slightly to

$$\psi_{jk}(t) = 2^{j/2}\psi\left\{2^j(t-k)\right\}$$

for all integers k and $j \geq 0$. Thus the location of non-decimated wavelets does not depend on the resolution level: indeed non-decimated wavelets at all resolution levels, j, appear at all locations k. This is in contrast to the standard wavelets that only appear at locations $2^{-j}k$ — a dyadic grid.

The standard wavelet basis functions $\{\phi_{0k}(t), \psi_{jk}(t)\}_{j\geq 0, k\in\mathbb{Z}}$ form an orthonormal family and so the coefficients of expansion (1) can be obtained by

$$d_{jk} = \int_{\mathbb{R}} f(t)\psi_{jk}(t)\,dt \qquad (3)$$

and similarly for the c_k. The reason why wavelets are useful is that they provide *sparse* representations for a wide set of different functions, including functions that exhibit discontinuities or time-varying frequency behaviour. The sparsity of representation is due to the fact that wavelets provide *localised* information about a function: the wavelet coefficient d_{jk} provides information about a function at scale 2^j and near location $2^{-j}k$. Indeed wavelets provide a way of examining a function at different scales: a multiresolution analysis, see Mallat (1989a) or Jawerth and Sweldens (1994). The non-decimated representations are no longer orthogonal although orthogonal bases can be selected from them. However, the lack of orthogonality is not a problem for the statistical modelling we perform later.

In many cases, including time series applications, data arise as discrete observations and so discretized versions of (1) have been developed for functions on finite intervals. The best of these transforms exploit relationships between wavelets at different scales and lead to extremely fast algorithms for computing coefficients from discrete data, as we shall describe next.

2.1 Notation for discrete wavelet transforms

The notation we adopt is taken from Nason and Silverman (1995). There are four key operators: $\mathcal{H}$ — a low-pass filter; $\mathcal{G}$ — a high-pass filter; $\mathcal{D}_0$ — "even" dyadic decimation; $\mathcal{S}$ — a shifting operation. These operators are defined on doubly infinite sequences $\{x_n\}_{n\in\mathbb{Z}}$, although in practice the operators are adapted here for use on finite sequences assuming periodic boundary conditions.

The $\mathcal{H}$ filter has coefficients $\{h_n\}_{n\in\mathbb{Z}}$ and operates on an input $\{u_n\}_{n\in\mathbb{Z}}$ as follows:

$$(\mathcal{H}u)_k = \sum_n h_{n-k}u_n. \qquad (4)$$

The $\mathcal{H}$ filter is a low-pass filter. The formula for $\mathcal{G}$ is identical but uses coefficients $\{g_n\}_{n\in\mathbb{Z}}$ with a high-pass frequency response.

Most practical wavelet transforms rely on $\mathcal{H}$ and $\mathcal{G}$ being finite impulse-response filters having only finitely many non-zero coefficients. The Daubechies' (1988) compactly supported wavelets used in this article fall into this category and have

L non-zero filter coefficients, where L depends on the desired smoothness of the underlying wavelet. As a consequence the filtering operation in (4) only requires $\mathcal{O}(L)$ computations.

A discrete version of (1) using a specified wavelet $\psi(x)$ uniquely determines the filter sequences and the discrete transform itself only uses the filter sequences. (In the signal processing literature the filter sequences are known as *quadrature mirror filters*.) The filter sequences satisfy certain internal relations as a consequence of being derived from an orthonormal wavelet basis. For details on these relationships see Nason and Silverman (1995), Mallat (1989b) or Daubechies (1988; 1992).

The shift operator, $\mathcal{S}$, shifts the whole sequence along one position:

$$(\mathcal{S}u)_k = u_{k+1} \tag{5}$$

for all $k \in \mathbb{Z}$. The "odd" dyadic decimation operator, $\mathcal{D}_1$, selects every odd member of a sequence: it is equivalent to shifting followed by "even" decimation thus:

$$(\mathcal{D}_1 u)_k = (\mathcal{D}_0 \mathcal{S} u)_k = u_{2k+1} \tag{6}$$

for all $k \in \mathbb{Z}$. For simplicity the k subscript is sometimes dropped and the operator notation reduces to a vector/matrix style. On finite length sequences the decimation operators produce an output sequence which is half the length of the input.

In discrete wavelet transforms the filter operators and decimation operators are paired up to produce the operators $\mathcal{D}_0\mathcal{H}$, $\mathcal{D}_0\mathcal{G}$, $\mathcal{D}_1\mathcal{H}$ and $\mathcal{D}_1\mathcal{G}$. The vector of coefficients produced by any of these pairs is sometimes known as a "packet" of coefficients and therefore we shall refer to the four operators collectively as *packet operators*. Wavelet transforms obtain signal detail at different scales or frequency bands. They achieve this by chaining the low-pass $\mathcal{D}_0\mathcal{H}$ and the high-pass $\mathcal{D}_0\mathcal{G}$ operators together: for example, a band-pass filter can be obtained by repeated application of $\mathcal{D}_0\mathcal{H}$ followed by $\mathcal{D}_0\mathcal{G}$.

Finally all the transforms compute expansion coefficients using a fast algorithm that computes coarser level $j-1$ coefficients from finer level j coefficients. To initiate this recursive algorithm a finest resolution level J is chosen and the coefficients, c^J, are computed from the data, $\{X_t\}_{t=1}^N$, at this finest level. In general this can be done using interpolation or projection methods such as direct computation as in (3), or see Donoho (1992), Delyon and Juditsky (1995) and Kovac (1997). If there are $N = 2^J$ (for some integer J) equally spaced data points then the finest coefficients can be obtained approximately by just setting them to be equal to the data. In other words

$$c_n^J = X_n, \text{ for } n = 1, \ldots, N. \tag{7}$$

In all cases described below the *smooth at level* J, c^J, forms the input to all the wavelet transforms and coarser wavelet coefficients are rapidly computed from them.

2.2 Non-decimated wavelet packet transform

The non-decimated wavelet packet transform (NWPT) applies the four packet operators recursively to c^J the smooth at level J. This is illustrated schematically in Figure 2 for

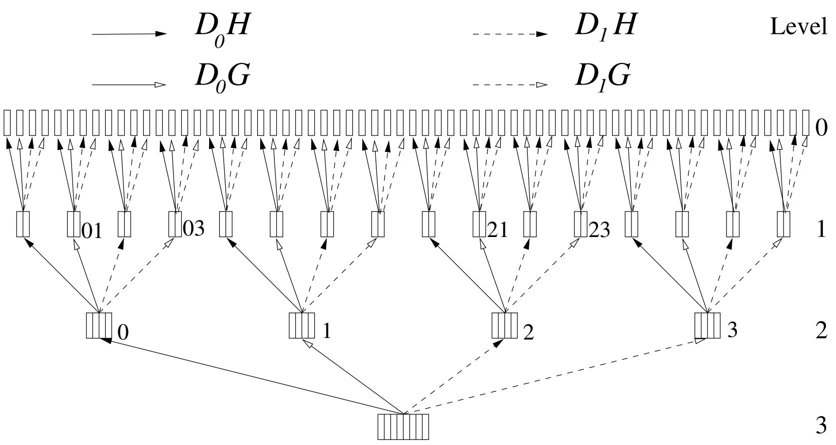

Figure 2: *Schematic of the NWPT operating on $N = 8$ points, i.e. $J = 3$ (see text for description).*

$N = 8$ data points at the root of the tree. Each packet of coefficients in the NWPT tree can be addressed by an index written in base 4. The number of digits in the index indicates the level of the packet: packets at level j have $J-j$ digits for $j = 0, \ldots, J-1$. The actual entries in the index describe how that packet was reached from the root: application of $\mathcal{D}_0\mathcal{H}$, $\mathcal{D}_0\mathcal{G}$, $\mathcal{D}_1\mathcal{H}$, or $\mathcal{D}_1\mathcal{G}$ augments a 0, 1, 2 or 3 respectively to the index. For example, the indices 0, 1, 2 and 3 at level 2 and 01, 03, 21 and 23 at level 1 are indicated in Figure 2. For example, the 23 packet is so labelled because it is obtained by the operator $\mathcal{D}_1\mathcal{H}$ followed by $\mathcal{D}_1\mathcal{G}$. There are 4^{J-j} packets each of length 2^j for levels $j = 0, \ldots, J-1$ and, therefore, the total number of coefficients and computational effort is

$$\begin{aligned}
\sum_{j=0}^{J-1} 4^{J-j} 2^j &= 2^J \sum_{j=1}^{J} 2^j \\
&= 2^{J+1}(2^J - 1) \\
&= 2N(N-1) = \mathcal{O}(N^2).
\end{aligned}$$

The NWPT was proposed by Pesquet *et al.* (1996) who also investigated its potential for curve and time-delay estimation. However, our intended statistical modelling requires coefficients in time-order and leads us to develop the time-ordered NWPT described in sections 2.3 and 2.4. The NWPT is a generalisation of the following important and well-known wavelet transforms:

NWT The non-decimated wavelet transform can be derived from the NWPT by keeping only packets whose indices are a string of even digits except for the last digit which may be odd or even (so for example, 23 is kept, but 32 is not). Figure 3 shows a schematic of the NWT. The computational cost of the NWT is $\mathcal{O}(N \log N)$.

A detailed description of the NWT can be found in Nason and Silverman (1995). See also Percival and Guttorp (1994), Bruce and Gao (1996) and Percival and Mofjeld (1997). Coifman and Donoho (1995) developed a translation-invariant curve estimation procedure using the NWT. See also Nason and Silverman (1995), Lang *et al.* (1995) and Johnstone and Silverman (1997) for more on curve estimation with the NWT. The NWT appeared earlier in the mathematical and signal processing literature, see Beylkin (1992) and Shensa (1992).

WPT The wavelet packet transform can be derived from the NWPT by retaining coefficients that have been produced by even dyadic decimation (i.e. keep those packets whose indices consist only of digits that are 0 or 1). Figure 4 shows a schematic of the WPT. The computational cost of the WPT is $\mathcal{O}(N \log N)$.

An excellent reference for wavelet analysis in general and wavelet packets in particular is Wickerhauser (1994). The WPT has been used successfully for compression purposes: e.g. Coifman and Wickerhauser (1992); for curve estimation: Donoho and Johnstone (1994a), Pesquet *et al.* (1996); and for classification: Saito (1994), Learned and Willsky (1995), Coifman and Saito (1994), and Saito and Coifman (1996).

DWT The discrete wavelet transform can be derived from the NWPT by imposing both of the conditions described in the derivation of the NWT and WPT above — only even dyadic decimation is used and recursion is only applied to the $\mathcal{H}$ derived packets. A schematic of the DWT can be found in Nason and Silverman (1994). The computational cost of the DWT is $\mathcal{O}(N)$. The DWT is the intersection of coefficients from the NWT and WPT (and is therefore a subset of coefficients of the NWPT).

The DWT had been known for some time in the signal processing community as *subband coding*, see Mintzer (1982; 1985) or Smith and Barnwell (1986). The wavelet interpretation and transform was developed by Mallat (1989b; 1989a). The DWT forms the basis of most statistical methods to date (see list of papers at the beginning of this section).

The DWT is one-to-one and onto and therefore has a unique inverse. Indeed the DWT may be represented by a matrix multiplication using an orthogonal matrix although computation is always performed using the fast DWT. The DWT is essentially a discretized version of the expansion in (1). The DWT provides a representation with respect to one wavelet basis whereas the NWT, WPT and NWPT provide representations with respect to a library of bases. Recently, Cohen *et al.* (1997) describe an $\mathcal{O}(N \log N)$ "best-basis" algorithm for the NWPT. Flexible decompositions of functions using a library of bases are known in the literature as *atomic decompositions*, see Mallat and Zhang (1993) and Chen *et al.* (1995).

The non-decimated wavelet transforms (NTs) have another useful property called *translation equivariance* which means that

$$\mathrm{NT}[SX_t] = S(\mathrm{NT}[X_t]).$$

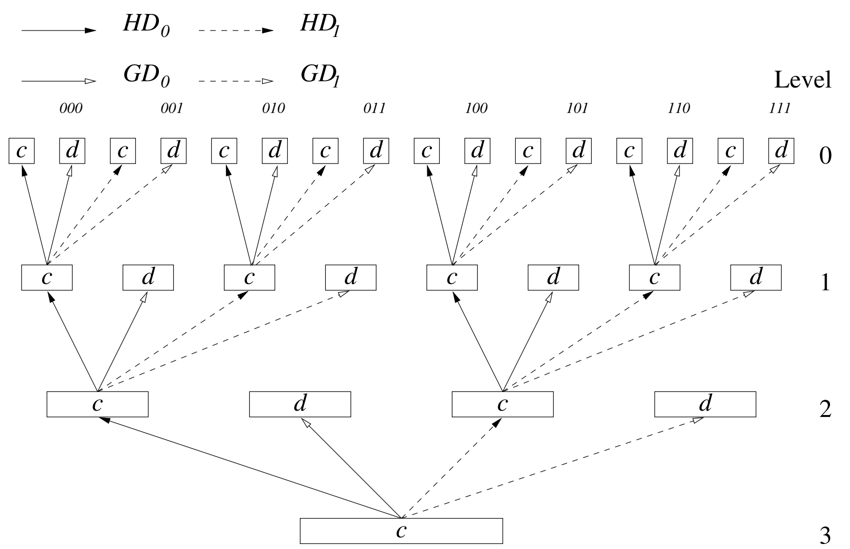

Figure 3: *Schematic of the NWT operating on $N = 8$ points, i.e. $J = 3$. The c and d boxes represent packets of father and mother wavelet coefficients respectively. Packets at level j contain 2^j coefficients labelled in each packet from 0 to $2^j - 1$ we have not shown individual coefficients — just packets. The DWT is contained within the left-most coefficients by following the solid arrows from level 3 up to level 0.*

In other words a shift in the time series is reflected by an identical shift in the time-ordered NT coefficients. The DWT and WPT are neither translation-equivariant nor translation-invariant. Finally it should be noted that it is possible to compute NTs with an arbitrary number of points N, although our transforms are restricted to N being a power of 2.

The NWT library contains all possible shifted wavelet bases and the WPT library contains a richer set of "mother" basis functions (but no shifts). The NWPT library contains all possible shifts of the richer WPT library (i.e. the NWPT is the best of both NWT and WPT worlds). Figure 5 shows four such WPT basis functions with varying time-extent, number of oscillations, location and scale.

2.3 Time-ordered coefficients from the non-decimated wavelet transforms

Figure 6 shows the NWT wavelet coefficients of a simulated reflected chirp signal. In this and all subsequent examples we use Daubechies' least-asymmetric wavelet with ten vanishing moments, see Daubechies (1992, p. 198). The chirp time series itself is shown in Figure 6 at resolution level 10 — this is the smooth at level 10 and the series is of length $N = 1024 = 2^{10}$. Chirps arise in real applications such as radar signal detection, astrophysics and biology (the Australian white-winged chough is a good biological example). Chirps are particularly interesting signals to study because

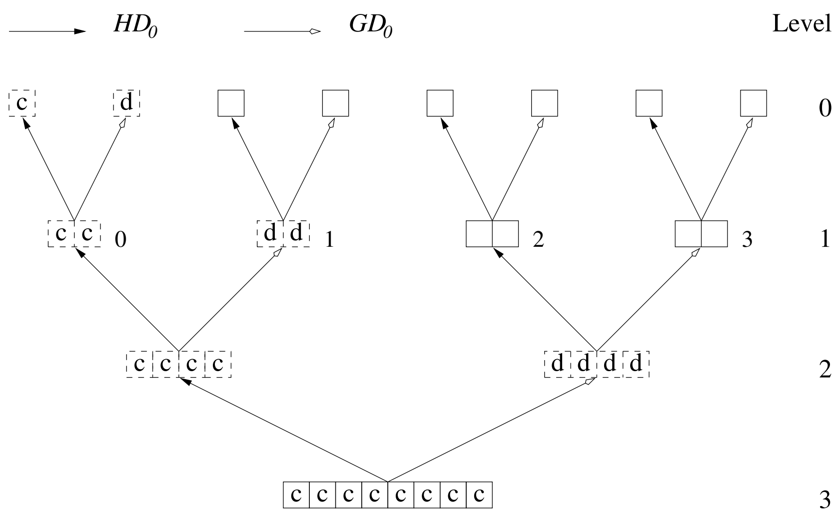

Figure 4: *Schematic of the WPT operating on* $N = 8$ *points, i.e.* $J = 3$*. At level* j *there are* 2^{J-j} *packets each containing* 2^j *points. The numbers 0,1,2,3 next to the packets at level 1 are the index numbers of packets within that level from left to right. The DWT coefficients are contained in the WPT and are shown in the dashed boxes and marked* c *and* d*. All other coefficients are with respect to other wavelet basis functions such as those illustrated in Figure 5. The total number of packets (excluding the original data) is* $2N - 2 = 14$*.*

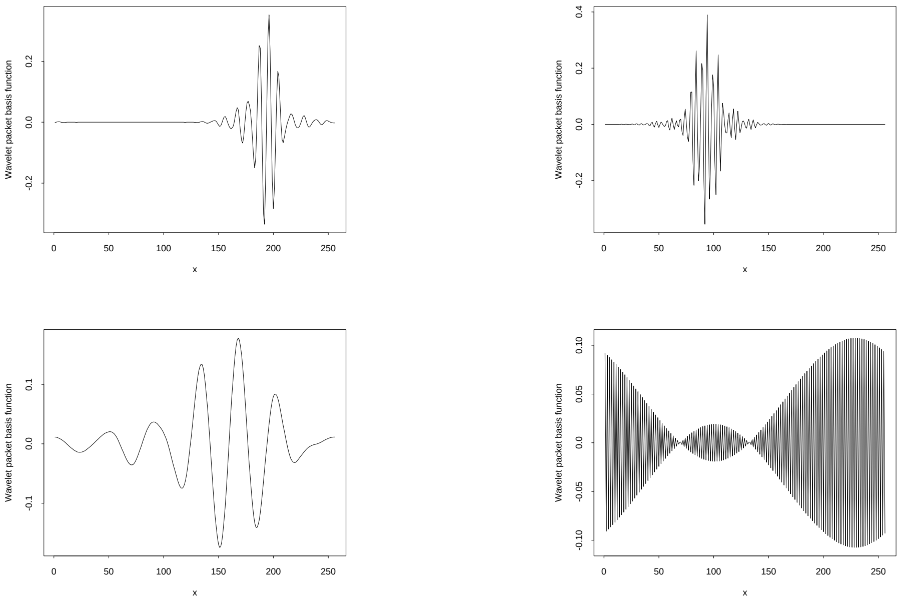

Figure 5: *Four wavelet packet basis functions for an analysis where* $N = 256$. *Packets are indexed in a tree like that in figure 4 as* (j, i) *where* j *is the resolution level and* i *is the packet index within a level. The basis functions here are (clockwise from top left) (4,2), (4,5), (2,2), (1,64) and happen to form an orthonormal set.*

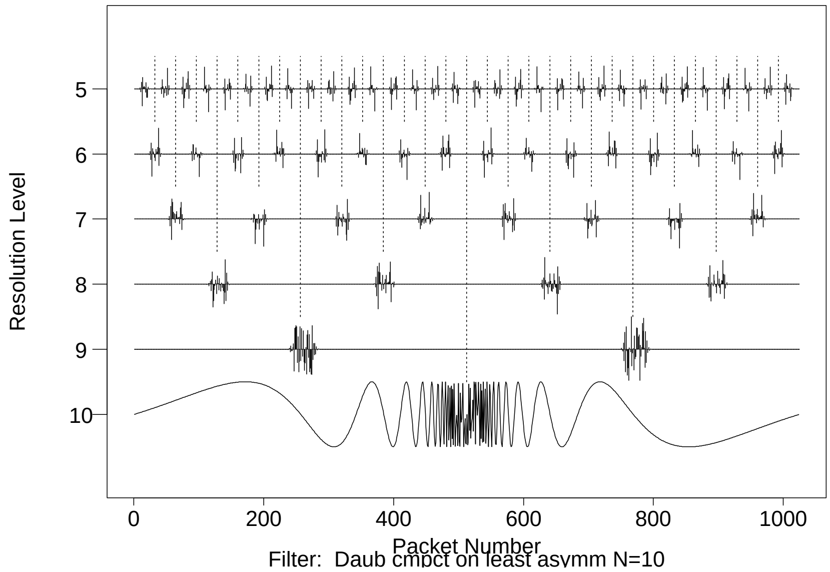

Figure 6: *Plot of wavelet coefficients from a NWT analysis of a chirp signal (see text for description).*

their frequency content changes over time and this change cannot be detected by standard Fourier-based statistical methods such as the periodogram. The resolution levels above level 10 show successively coarser resolution wavelet coefficients (the successive smooths c^j have not been shown). The layout of Figure 6 corresponds to the schematic in Figure 3, except that there are ten resolution levels, not three, and also Figure 6 has not plotted any levels with resolutions less than 5. The vertical dotted lines in Figure 6 indicate packet boundaries.

For our statistical modelling, to be described below, the advantage of using non-decimated wavelet transforms is that the number of coefficients at each level (NWT) or packet (NWPT) is the same as the number of data points in the original data (so in the NWT of Figure 6 there are 1024 coefficients at each resolution level). With decimated wavelet transforms (DWT, WPT) decimation halves the number of coefficients at each level (DWT) or packet (WPT). However, the coefficients in each resolution level in the NWT and NWPT are not in chronological order. For example, in Figure 6 there are two packets of coefficients at resolution level 9. The coefficients are in time-order within each packet but the left-hand coefficients correspond to even dyadic decimation and the right-hand coefficients correspond to odd. The time-ordering may be recovered by weaving the odd and even coefficients together as illustrated by Figure 7. Weaving can also be carried out at lower resolution levels to recover the time-ordering — although the weaving is not always a simple meshing of odd and even coefficients at lower resolutions. Figure 8 shows the chirp signal after weaving of coefficients at each resolution level. The horizontal axis of Figure 8 now corresponds exactly to the time-base of the original chirp signal and each resolution level is aligned to this time-

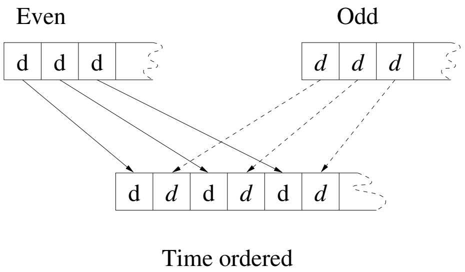

Figure 7: *How the even and odd decimated non-decimated wavelet coefficients may be woven together into time-ordering for resolution level 9 in figure 6.*

base. Moreover as the caption to Figure 8 suggests the time-ordered NWT can form the basis of a time-frequency or time-scale diagram where information about the time and frequency properties of a signal can be inferred simultaneously (in contrast to the standard periodogram which only reveals frequency content over all time). For more details about time-scale analysis with the NWT see Nason and Silverman (1995) and von Sachs *et al.* (1996; 1997).

Walden and Contreras Cristan (1997) have also developed the time-ordered NWPT. They used the time-ordered NWPT for investigating the time-scale properties of a series and related it to the *original* series. However, Walden and Contreras Cristan (1997) use a direct algorithm to produce the NWPT coefficients in time order rather than weaving as we do here. Their method uses the WPT schematic but pads out filter coefficients depending on the resolution level (similar to the technique used by Nason and Silverman (1995) in producing time-ordered NWT coefficients). Padding essentially modifies Mallat's (1989b) DWT formulae to become

$$c_k^j = \sum_n h_n c^{j+1}_{2^{J-j-1}n+k} \tag{8}$$

for $j = J-1, \ldots, 0$ (replace h_n by g_n for wavelet coefficients). Walden and Contreras Cristan (1997) also specify a phase-correction technique to bring levels into precise time-alignment. Without phase-correction each successive level of coefficients would be shifted in time with respect to the previous level. We have applied an empirically determined shift factor to time-align our levels which works well in our applications.

To summarise, the time-ordered non-decimated wavelet transforms (NWT, NWPT) convert a time series X_t into a K-dimensional multivariate time series $\mathbf{X}_t$ for $t = 1, \ldots, N$. In the time-ordered NWT each variable of $\mathbf{X}_t$ corresponds to a (non-decimated) level in the transform (e.g. each resolution level in Figure 8 corresponds to one variable in $\mathbf{X}_t$) and $K = \log_2 N$. In the time-ordered NWPT each variable corresponds to a (non-decimated) packet in the transform and $K = 2N - 2$.

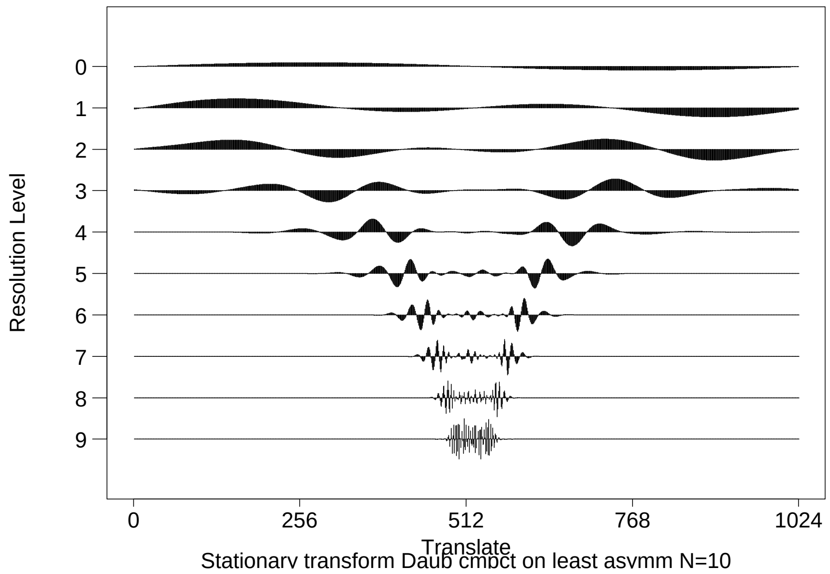

Figure 8: *Wavelet coefficients produced using the time-ordered NWT for the chirp signal. The coefficients are ordered left to right in exact alignment with the time-base in the chirp time series. The increase and decrease in frequency of the chirp over time is clearly visible with coefficients moving from low to high and then back to low resolution bands as one sweeps from left to right across the diagram.*

2.4 Weaving details for the non-decimated wavelet packet transform

A distinction must be made between an "ordinary" NWPT packet (such as the ones in Figure 2) and a time-ordered NWPT packet. Time-ordered NWPT packets follow the WPT indexing scheme and are obtained by weaving together coefficients from ordinary NWPT packets. NWPT packets are as long as the original data. For example, in the ordinary WPT at level $J-1$ there are two packets: packet 0 and 1 each of length 2^{J-1} (see Figure 4). With the ordinary NWPT at level $J-1$ there appears to be four packets (see Figure 2). However, one can also visualise the ordinary NWPT packets at level $J-1$ as two time-ordered *non-decimated* packets corresponding to the ordinary WPT by interweaving the four packets in the following way:

- weaving together the packets produced by $\mathcal{HD}_0$ and $\mathcal{HD}_1$. This produces the time-ordered *non-decimated packet* $\mathcal{H}$ and corresponds to the time-ordered non-decimated version of the ordinary WPT packet index 0.
- weaving together the packets produced by $\mathcal{GD}_0$ and $\mathcal{GD}_1$. This produces the time-ordered *non-decimated packet* $\mathcal{G}$ and corresponds to the time-ordered non-decimated version of the ordinary WPT packet index 1.

Therefore the weaving process is a two-stage procedure: choose which time-ordered non-decimated NWPT packet you require (using the WPT indexing scheme) and then identify the associated ordinary NWPT packets; weave the associated packets into time-order.

In general, to obtain the correct time ordering the ordinary NWPT packets are not taken sequentially but with reference to the root node. For example, let us refer to level 1 in Figures 4 and 2. Suppose that we wished to obtain time-ordered non-decimated wavelet packet index 1 (or in operator notation the packet produced by $\mathcal{H}$ followed by $\mathcal{G}$). This corresponds to ordinary NWPT packet indices 01, 03, 21 and 23 using the base 4 notation from Section 2.2 (each of the cases where a $\mathcal{G}$ operator follows a $\mathcal{H}$ operator regardless of decimation). To produce correct time-ordering we take coefficients successively from the ordinary NWPT packets in the order 01, 21, 03 and then 23. This ordering occurs because the shift of basis functions is finer nearer the root node. The transition from level 3 to 2 encodes a shift of one position, the transition from level 2 to 1 encodes a shift of two positions. So the "distance" of 21 to 01 is only 1, from 03 to 01 is 2 and from 23 to 01 is 3. So, relative to 01, 21 has undergone a unit shift, 03 a two unit shift and 23 both a unit and two unit (= 3 unit) shift. See the appendix for details on the ordinary NWPT packet selection and weaving algorithm.

3 Statistical modelling of time series with time-ordered non-decimated wavelet transforms

3.1 Comparison with existing modelling procedures

Both wavelets and wavelet packets have been used in time series classification problems following the usual paradigm: a training set of samples is collected; the wavelet transform is applied to all the training samples and the important coefficients for classification are identified using some statistical method (e.g. linear discriminant analysis or CART). Pioneering work by Saito (1994) and Learned and Willsky (1995) extended the method for use with the WPT by inserting a "select best-basis" step before the "best coefficients" were selected. Model verification can be performed or future samples classified by applying the wavelet (or WPT) transform and predicting the class of the sample by applying the statistical classifier to the selected (basis and) wavelet coefficients.

The advantage of using wavelet and wavelet packet functions is that they represent functions sparsely and so provide an effective non-statistical dimension reduction. The sparse representation and resultant low dimensionality of transformed data makes it easier for the statistical classifiers to do their work. Our modelling is of a somewhat different type to that just described as we build models *in situ* rather than have a large set of training samples. However, the models we propose also gain from the effective dimension reduction of the sparse wavelet representation.

3.2 Advantages of the time-ordered non-decimated wavelet packet transform for modelling

Recall that we wish to model a response time series, Y_t, in terms of an explanatory time series, X_t, where we believe certain components of X_t have a direct, but unspecified, effect on the values of Y_t.

Both the time-ordered NWT and NWPT transform X_t into a multivariate time series $\mathbf{X}_t$, where each variable contains representations of X_t with respect to particular (non-decimated) wavelet or wavelet packet basis functions. For modelling purposes think of all the K time series vectors stacked next to each other to obtain a matrix $\mathbf{X}$ with $t = 1, \ldots, N$ as rows (observations) and every column (variables) corresponding to a particular wavelet basis function. The number of variables, K, is $\log_2 N$ or $2N - 2$ if the time-ordered NWT or NWPT is used respectively. For example, the chirp signal mentioned earlier contained 1024 points. The time-ordered NWT of the chirp contains $N = 1024$ observations on $K = 10$ variables: the matrix $\mathbf{X}$ is illustrated (on its side) in Figure 8.

With time-ordered NTs (NWT, NWPT) the wavelet representation can be determined at *every* time point, t_0: it will be $\mathbf{X}_{t_0}$. Therefore time-ordered NTs will be able to detect components within X_t that occur at *any* time point. With decimated transforms it is not always possible to determine what is happening at arbitrary time points and so interesting components of X_t that arise "between" decimated coefficients might be missed.

A further advantage of the time-ordered NTs is that each resolution level (NWT) or non-decimated wavelet packet (NWPT) can be treated as a variable containing the same number of points as in the Y_t series. Therefore the problem of modelling Y_t with $\mathbf{X}_t$ is straightforward and all the power of standard statistical modelling and verification techniques can be used. Once a successful statistical model has been obtained one can interrogate the model to find out which variables are influential in the model. Since the variables contain coefficients with respect to certain wavelet (or packet) functions one can identify what sorts and shapes of component within X_t are important for modelling particular aspects of Y_t.

The final advantage of the time-ordered NTs is that they are translation-equivariant. For modelling the important consequence of translation-equivariance is that components of X_t have coefficients in $\mathbf{X}_t$ which will be the same wherever the components arise in X_t. This will not be the case for the decimated transforms where the wavelet coefficients of identical, but shifted, components could potentially be different and thus "confuse" the classifiers.

In many "standard" classification problems (such as those described in the previous section) wavelet packets work better than wavelets since packets provide a more flexible collection of basis functions. Thus packets can provide sparser representations for many more functions than can wavelets (and as mentioned before sparsity leads to increased classification efficiency).

So our transform of choice is the time-ordered NWPT which has all the advantages of non-decimated transforms and the flexibility of wavelet packets.

3.3 Statistical model selection

This section explains how a time series Y_t is modelled in terms of a series X_t using wavelets to represent important components in X_t.

The overall procedure is very simple. First, the explanatory time series X_t, is converted to a time-ordered NWT or NWPT representation, $\mathbf{X}$. Then the dependent series Y_t may modelled using $\mathbf{X}_t$ using any convenient statistical modelling method and sound statistical common sense. Once a successful model is obtained then many interesting possibilities arise. For example:

- the model often contains useful information about what components in X_t drive Y_t.
- future values of Y_t may be predicted from future values of X_t. This might be useful in the situation where predictions of Y_t are required but difficult to obtain by other means. If X_t is easy to forecast then a validated model linking Y_t with X_t might make predicting Y_t possible.

The number of variables (non-decimated levels) in the time-ordered NWT is $K = \log_2 N$. With the time-ordered NWPT the number of variables (non-decimated packets) $K = 2N - 2$ exceeds the number of observations and rules out the direct application of many useful statistical methods. In this situation we adopt a "variable-selection" approach: we usually rank the K variables according to their correlation with the response or discriminatory power and then select a "large" number, $K_1 < N$, of them to be subjected to our statistical method. Alternatively, we sometimes perform principal components on the K_1 variables and select a number, $K_2 < K_1$ of them to be the inputs of our statistical method. A better approach might be to use more sophisticated methods such as those described by Krzanowski *et al.* (1995), although even our simple methods described above are competitive in some situations. We are considerably aided in the variable selection because a number of sets of variables corresponding to the large-scale wavelet packet bases are highly correlated and so a "reasonable" number of variables can usually be selected manually without too much trouble. Usually the "reasonable number" is highly problem-specific and this is where statistical common sense is important!

Finally we sometimes transform the coefficients by taking the logarithm of the squared values of the coefficients. This particular transform is of use when local oscillatory power is thought to be important in driving the dependent time series: for instance, below in the sleep state/heart rate example, it is the power of oscillation rather than the oscillation itself that is related to changes of sleep state. Using a power-based statistic is a standard signal processing trick, see Learned and Willsky (1995). von Sachs *et al.* (1997) also advocate the use power-based coefficients in local time-scale modelling.

4 Examples

4.1 Infant sleep-state classification

This section discusses the classification of sleep state from heart rate for the four month year old baby from section 1. Computerised analyses of *adult* sleep have been documented in the medical literature using: multidimensional scaling, Burger *et al.* (1977); adaptive segmentation and fuzzy subset theory, Gath and Baron (1980); expert system approaches, Ray *et al.* (1986). In infants, attention has focussed on further subdividing sleep into two broad categories "active" and "quiet" sleep — the former being equivalent to adult rapid eye movement (REM) sleep — during which dreaming and many upper airway breathing disorders occur (DeHann *et al.* (1977); Schechtman *et al.* (1988)). Harper *et al.* (1987) developed an off-line system based on cardiac (4 variables from heart rate) and respiratory measures (3 variables from respiration). Harper *et al.* (1987) quote success rates of 85% using all 7 variables, 82% for the four cardiac and 80% for the three respiratory variables.

In all we experimented with three statistical modelling methods: linear discriminant analysis (LDA), logistic regression and CART (see Mardia *et al.* (1979), McCullagh and Nelder (1989) and Breiman *et al.* (1984) respectively). Of these three methods LDA working on the log-transformed absolute values of coefficients was most successful and is described here. We mainly used segments of length 128 but longer segments were tried with similar results.

The explanatory time series X_t from Figure 1 was transformed using the time-ordered NWPT and the $\mathbf{X}$ matrix formed. The matrix consisted of $N = 128$ observations on $K = 254$ variables (non-decimated packets). Ideally we would like to perform LDA to find the optimal combination of variables to discriminate between the asleep and awake levels of the sleep state response Y_t. However, here there are clearly many more variables than observations. We could use the methods proposed by Krzanowski *et al.* (1995) but we adopt a far simpler procedure which appears to work well in practice. For each of the 254 variables we obtain the the correlation between the variable and the sleep-state. Then we select a reasonable number (again judgement is required but 10 seems to work well) of the best variables. The identity of these variables with their correlations is indicated in Table 1. The "best" basis functions

Table 1: *Resolution levels and packet indices of the time-ordered non-decimated wavelet packets that were identified as being important for relating Y_t to X_t.*

Resolution level j	*Index within level*
4	0
5	0
3	0
6	0
1	0
2	0
1	56
2	30
2	5
2	19

Table 2: *Confusion matrix for predicted and true sleep states.*

		True	
		Asleep	Awake
Predicted	Asleep	69	13
	Awake	0	46

from Table 1 are plotted in Figure 9.

Table 1 reveals that the best six variables for one-dimensional classification all have a packet index of 0 at resolution levels 1 to 6. These variables correspond to multiple application of the smoothing filter $\mathcal{H}$ (with no application of the detail filter $\mathcal{G}$) and are father wavelets (as illustrated by the left-most functions in Figure 9). As mentioned in Section 2 these basis functions look like standard statistical kernel functions. The appearance of the father wavelets implies that simple averaging over scales 1 to 6 is important for determining sleep state. This corresponds with the observation earlier than the level of the heart rate over these scales is an important determining factor.

The top ten variables were then used in a LDA analysis which determined which linear combinations of the top ten were best for discrimination. To exercise our model we took the next 128 heart rate values, performed the NWPT analysis, extracted the same top ten variables and used the linear combinations determined by the previous LDA to predict the sleep state for the next 128 time periods (additionally the mean of the next 128 heart rate values was adjusted to be the same as the previous 128 values to prevent this affecting the analysis as it adds no discriminatory value). Figure 10 shows the new time-ordered NWPT values projected onto the first two discriminant axes, the location of the discriminant rule, and 13 misclassified observations. Table 2 shows the confusion matrix for classification of the new data achieving a 90% success rate. Figure 11 shows the new heart rate series with the true and predicted heart rate. Our method is fooled into thinking that the baby has gone to sleep just after 2.4

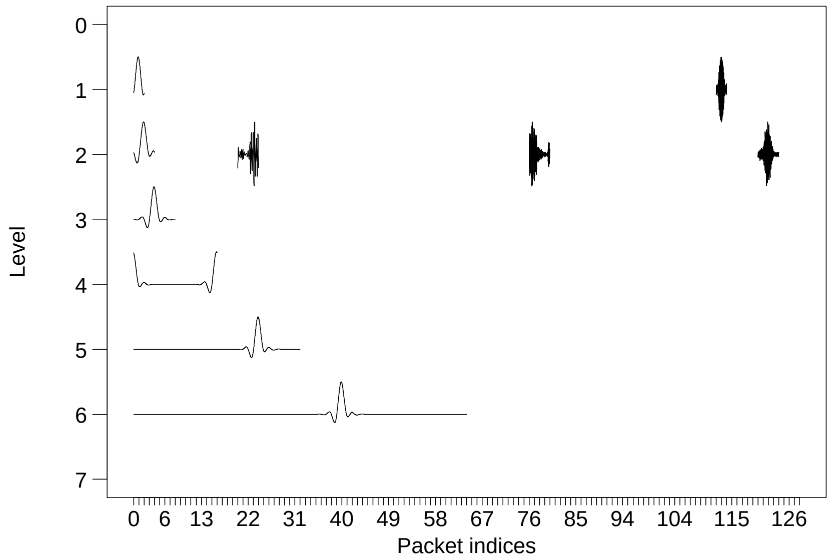

Figure 9: *Plot of the best 10 basis functions for discriminating* $\mathbf{Y}$ *from* $\mathbf{X}$. *The position of the basis function corresponds to the schematic layout in Figure 4 (although the time-ordered NWPT was used). Each basis function illustrated covers the whole time series, so in actuality each basis function should be stretched out to the full length of the series (so, for example, the basis function at level 5 should be stretched out by a factor of 2 to compare it to that at level 6). Also, since we are using periodic transforms the basis function at level 4 should look like all the other "index 0" basis functions (the hump just fell over the boundary).*

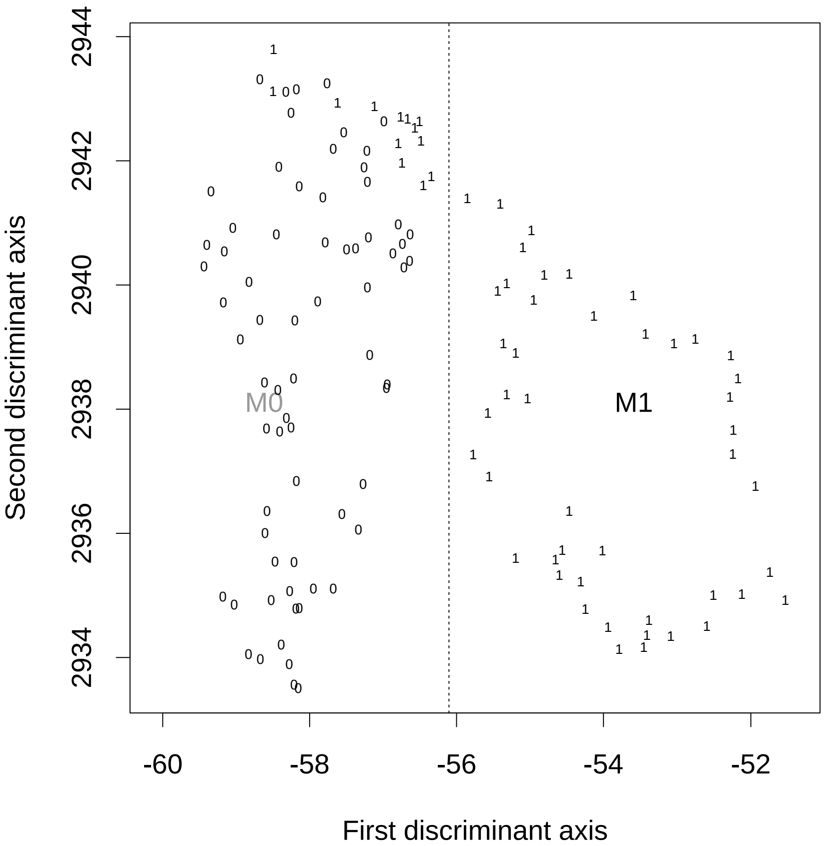

Figure 10: *Time-ordered NWPT values from the new heart rate time series projected onto first and second linear discriminant axes. The label of each point shows its true group membership. The vertical dashed line shows the discriminant rule: observations to the left are assigned to the asleep group (0), those to the right are assigned to the awake group (1). The 13 misclassified observations shown in Table 2 appear in the top-left of the plot. The M0 and M1 labels refer to the means of the asleep and awake groups used to build the discriminant model.*

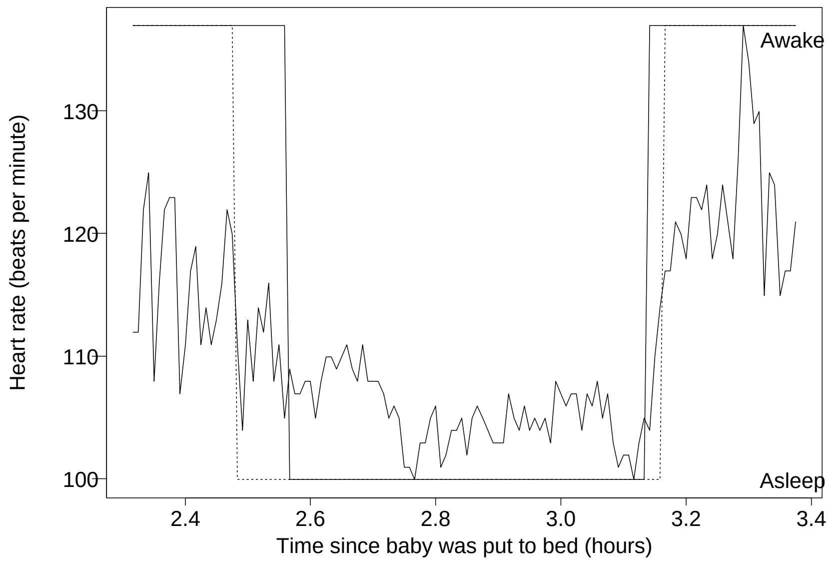

Figure 11: *New heart rate series with true sleep state (solid) and predicted sleep state (dashed)*

hours, probably by the sharp drop in heart rate. Likewise, just around 3.1 hours our method is a bit slow in noticing that the baby has awoken, but the delay in noticing is 2 minutes (however, the true record does note that during this period the human judge was uncertain about the true sleep state).

The biomedical time series used here is one of several recorded prospectively at monthly intervals from a group of infants during the first 6 months of life. As infants mature their EEG and EOG become easier to classify and conventionally determined sleep state becomes more accurate with less disagreement between observers. This was reflected by our models which became better at predicting infant sleep state with increasing age (around 75% success at 2-3 months, and 75% to 90% success at 4–5 months of age). Although we have had some success in classifying sleep state by building a model in the early part of a night and predicting what the sleep state is in later periods it may not be possible to transfer the exact model to the same infant on a different night or to a different baby of the same age. However, the same father wavelets nearly always recur in the best model suggesting that the averaging over certain time-scales is important and it is the different levels of heart rate which differ across nights or infants.

4.2 Wind speed regression

Before a wind farm is constructed a great deal of analysis is undertaken to establish whether a particular *target site* is suitable. One aspect of this analysis involves the prediction of the long-term mean wind speed at the target site. Typically, wind speeds are measured by a pilot anemometer at a height of 10m at the target site for several

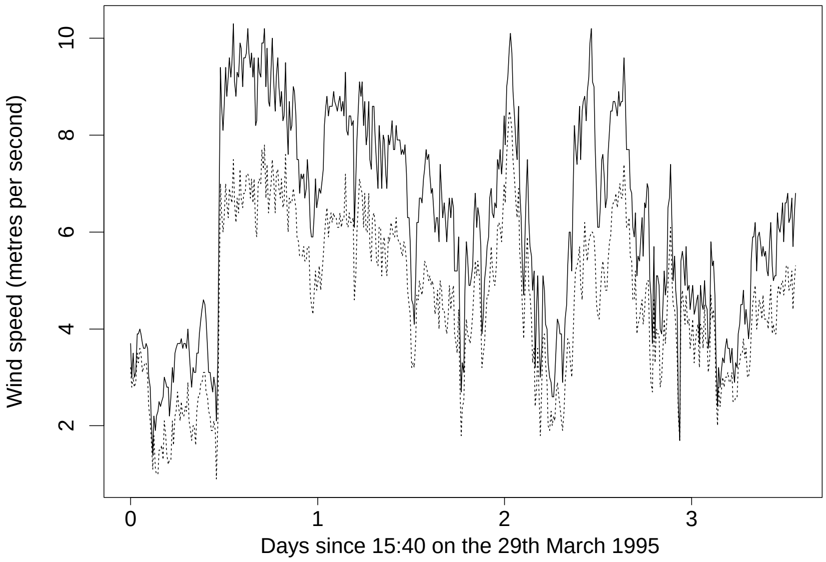

Figure 12: *Wind speeds taken every 10 minutes from 15:40 on 29th March 1995. Wind speeds at 10m (dashed line) and 40m (solid line) were collected. The whole time series corresponds to about 85 hours. (Data provided by Micon Turbines UK Ltd.)*

months. These target speeds are related to contemporaneous wind speeds measured at a nearby reference site (UK Meteorological Office station) and a model predicting target speeds from the reference speeds is constructed. The long-term mean wind speed at the target site can be estimated using the model and the long-term mean at the Met. Office site. Modelling of this kind is described by Cook (1985) and Haslett and Raftery (1989).

The actual wind turbines are often mounted at greater heights (e.g. 40m) and to assess the wind regime at these heights a second model is sometimes built based on data collected simultaneously for a shorter interval at 10m and 40m. Collecting data at 10m is cheaper than collecting it at 40m and so good models which can predict what happens at 40m from 10m are valuable.

Figure 12 shows wind speeds from anemometers at 10m, X_t, and at 40m, Y_t. The time-ordered NWPT was applied to a segment of X_t of length $N = 512$ and the matrix, $\mathbf{X}$, containing the log absolute values of the coefficients formed. We then used both generalized linear and additive modelling to find a good model for the 40m time series $\mathbf{Y}$ in terms of $\mathbf{X}$. The final model was a generalized linear model that assumed Gamma distributed $\mathbf{Y}$ values (wind speeds are positive and skewed to the right). There were $K = 1022$ possible basis functions (variables) that we could use. To obtain an idea of which of these functions were useful we correlated each basis function with the response and selected the top 5%. A backwards selection GLM technique was applied to these 51 and 16 variables were included in the final general linear model. The father wavelets at resolution level 8 form an important part of of the model. These wavelets

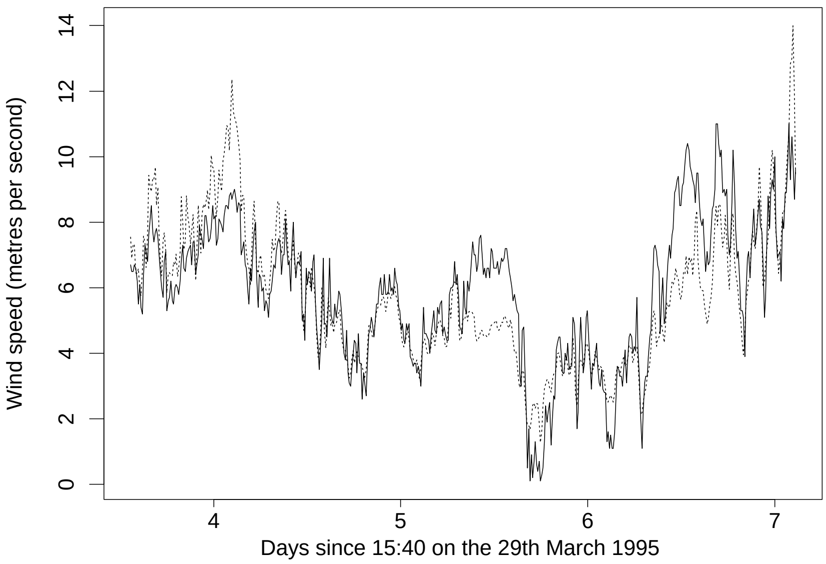

Figure 13: *Predicted wind speeds (dashed line) for 40m anemometer for every 10 minutes after 05:00 on 2nd April 1995. Actual wind speeds are shown as a solid line. (Actual data provided by Micon Turbines, UK, Ltd.)*

correspond to averaging over very short time-scales and it makes sense for them to be important in the model as both anemometers are on the same mast but just at different heights. However, in this model wavelet packets other than those corresponding to the father wavelets also play an important role.

Figure 13 shows the predicted wind speeds for the following period of 512 observations along with the actual recorded speeds for the 40m anemometer. The model does fairly well over most of the series but badly in some other areas. One possible explanation is that the model is different depending on the wind direction. Figure 14 shows a box plot of prediction errors versus wind direction and lends some support to this theory. Clearly our model is in error systematically either when the wind is in a north-westerly or south-easterly direction. There are many physical reasons why this may be the case: for example obstructions in these directions. It would be possible to introduce another explanatory time series and its time-ordered NWPT for the wind direction to try and improve the model.

5 Discussion and further work

This article only covers the basic idea of using time-ordered non-decimated wavelet transforms as providing representations for modelling a response time series in terms of an explanatory time series. There are many other interesting issues which we briefly discuss here.

It is clear with both of the real examples, and in statistical modelling in general,

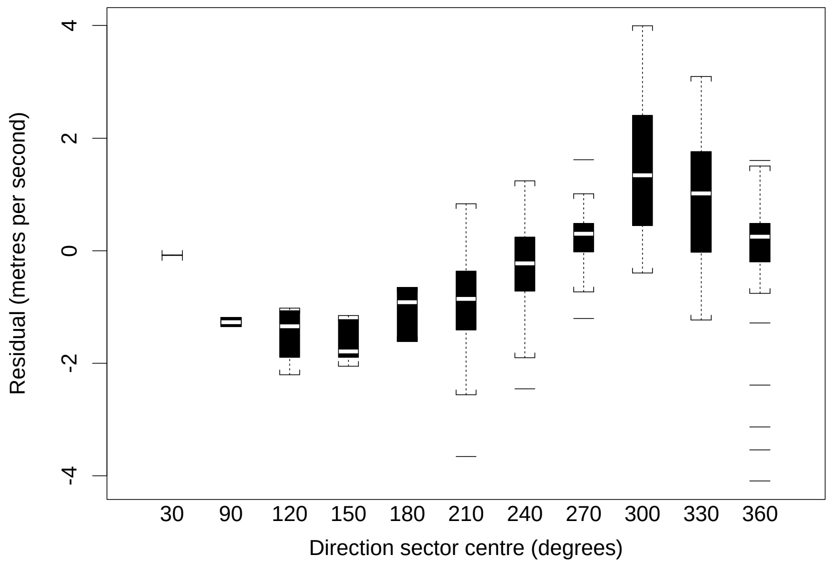

Figure 14: *Boxplot of prediction errors by direction sector at reference site for time series given in Figure 13. The circle is divided up into bins of width 30 degrees with centres indicated on the horizontal axis. (Note that no wind was recorded in the direction sector centred on 60 degrees for this time period.)*

that extra variables can only aid successful modelling. For example, with the infant example other variables such as rectal temperature and EEG are known to be important indicators of sleep state, see Sawczenko *et al.* (1997). In the wind speed modelling example wind direction is extra important variate. It is perfectly possible to include extra variables directly and/or insert their time-ordered non-decimated wavelet representations. Adding extra variables in this way could be termed a *multiple* variable representation.

Also, another extension of the methods described above would be into two or more dimensions. For example, one might have image data concerning manufacturing details and end-product of an industrial procedure and wish to relate them and provide predictions. Implementations of two-dimensional "packets" NWTs have been used, see Lang *et al.* (1995), although presumably not spatially ordered. We have implemented a two-dimensional spatially ordered NWT but have not yet used it for the modelling described in this article. Implementation of the two-dimensional spatially ordered NWPT would be a challenge.

The methods described above could be included in standard time series models or structural models, see Priestley (1981) or Harvey (1993). Indeed, explicitly lagged variables included in the models could be beneficial. For example, in the infant example there may be a lag between changes in heart rate and sleep state. However, due to the multi-scale nature of the wavelet representations time lags are implicitly included although not necessarily to the time resolution one might require. For example, lower-to mid-level coefficients contain information about some past and future events as well as the current time and the distance into the past and future increases as the resolution level decreases.

The time-ordered NWPT results in a situation where there are many more variables than observations and we have deliberately only used very simple variable selection procedures. There is much need for effective variable selection procedures driven by discrimination or regression considerations.

6 Conclusion

This article proposes the use of time-ordered non-decimated wavelet or non-decimated wavelet packet coefficients to represent an explanatory time series. The resulting representations are then used as variables in a statistical model to provide predictions of a response time series. The statistical model usually provides valuable information about which components in the explanatory time series drive the response time series. Standard statistical models can be used to their full extent as the time-ordered non-decimated wavelet transforms produce variables with as many coefficients as time series observations.

Acknowledgements

The authors would like to thank Peter Fleming and Jeanine Young of the Institute of Child Health, The Royal Hospital for Sick Children, Bristol and Piers Guy of Micon Turbines, UK, Ltd for supplying data. We would like to thank Bernard Silverman and Peter Green for reading an earlier version of this manuscript and for providing helpful comments. Nason and Sapatinas were supported in part by EPSRC grant GR/K70236. Sawczenko was supported by grant 186 from the Foundation for the Study of Infant Deaths, Shield Nationwide Ltd, The United Bristol Healthcare NHS Trust Medical Research Fund and Cot Death Research.

The calls of the Australian white winged chough and other birds can be found on the Australian National Botanic Gardens web site at `www.anbg.gov.au/anbg/birds.html`.

Appendix: Weaving algorithm

To obtain the ordinary NWPT indices associated with a time-ordered NWPT index, r at level j, say ($r = 0, \ldots, 2^{J-j} - 1$, $j = 0, \ldots, J-1$, see Figure 4 for details of the WPT indexing scheme) the following recursive procedure can be used:

1. convert the (decimal) time-ordered non-decimated wavelet packet index r into binary string s. Convert s into decimal but this time assuming s is in base 4. Call the result p (some example conversions: $\xrightarrow[b]{a}$ means convert from base a into base b.

$$1 \xrightarrow[2]{10} 1 \xrightarrow[10]{4} 1,$$

$$2 \xrightarrow[2]{10} 10 \xrightarrow[10]{4} 4,$$

$$3 \xrightarrow[2]{10} 11 \xrightarrow[10]{4} 5).$$

2. For $i = j, \ldots, J-1$ do

```
e <- 2^(2*J-2*i-1)
p <- c(p, p+e)
```

This example contains partial S code (see Becker, Chambers and Wilks (1988)). The first line sets $e = 2^{(2J-2i-1)}$ the second line uses the concatenation operator `c` that pastes together two vectors, i.e.

$$\mathtt{c}(\{x_i\}_{i=1}^n, \{y_j\}_{j=1}^m) = (x_1, \ldots, x_n, y_1, \ldots, y_m).$$

As an example suppose that again the time-ordered NWPT indices for the non-decimated wavelet packet at level 1 index 1 for the 8 point data set are required. After the binary to base 4 conversion: $p = 1$. In the loop: setting $i = 1$ we obtain $e = 8$ and $p = (1, 9)$. Then setting $i = 2$ we obtain $e = 2$ and $p = (1, 9, 3, 11)$ which are the required indices (in base 4: 01, 21, 03, 23). Time-ordered coefficients are obtained from these four ordinary packets by taking the first coefficient from each in order, then the second coefficient from each in order and so on.

References

Abramovich, F., & Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients. *Computat. Stat. Data Anal.*, **22**, 351–361.

Abramovich, F., & Silverman, B.W. (1997). The vaguelette-wavelet decomposition approach to statistical inverse problems. *Biometrika*, **84**, (to appear).

Abramovich, F, Sapatinas, T., & Silverman, B.W. (1997). Wavelet thresholding via a Bayesian approach. *J. R. Statist. Soc. B*, (under revision).

Anders, T.F., Emde, R.N., & Parmelee, A.H. (eds). (1971). *A Manual of Standardized Terminology, Techniques and Criteria for Scoring of States of Sleep and Wakefullness in Newborn Infants*. Los Angeles: UCLA Brain Information Service.

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

Antoniadis, A., Gregoire, G., & Nason, G.P. (1997). Density and hazard rate estimation for right censored data using wavelet methods. *J. R. Statist. Soc. B*, (under revision).

Becker, R.A., Chambers, J. M., & Wilks, A. R. (1988). *The New S Language*. Pacific Grove, California: Wadsworth.

Beylkin, G. (1992). On the representation of operators in bases of compactly supported wavelets. *SIAM. J. Num. Anal.*, **29**, 1716–1740.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and Regression Trees*. Belmont, Calif.: Wadsworth.

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

Buckheit, J.B., & Donoho, D.L. (1995). Improved linear discrimination using time-frequency dictionaries. *Pages 540–551 of:* Laine, A.F., Unser, M.A., & Wickerhauser, M.V. (eds), *Proc. SPIE. Wavelet Applications in Signal and Image Processing III*, vol. 2569. Bellingam, Washington: SPIE.

Burger, D., Cantani, P., & West, J. (1977). Multidimensional analysis of sleep electrophysiological signals. *Biol. Cybern.*, **26**, 131–139.

Chen, S.S., Donoho, D.L., & Saunders, M.A. (1995). *Atomic decompositions by basis pursuit*. Tech. rept. 479. Department of Statistics, Stanford University, Stanford.

Chipman, H.A., Kolacyzk, E.D., & McCulloch, R.E. (1997). Adaptive Bayesian Wavelet Shrinkage. *J. Am. Statist. Ass.*, **92**, (to appear).

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

Clyde, M., Parmigiani, G., & Vidakovic, B. (1997). Multiple shrinkage and subset selection in wavelets. *Biometrika*, **84**, (to appear).

Cohen, I., Raz, S., & Malah, D. (1997). Orthonormal shift-invariant wavelet packet decomposition and representation. *Sig. Proc.*, **57**, 251–270.

Coifman, R.R., & Donoho, D.L. (1995). Translation-invariant de-noising. *Pages 125–150 of:* Antoniadis, A., & Oppenheim, G. (eds), *Wavelets and Statistics, Lecture Notes in Statistics 103*. New-York: Springer-Verlag.

Coifman, R.R., & Saito, N. (1994). Constructions of local orthonormal bases for classification and regression. *Compt. Rend. Acad. Sci. Paris Ser. A*, **319**, 191–196.

Coifman, R.R., & Wickerhauser, M.V. (1992). Entropy-based algorithms for best-basis selection. *IEEE Trans. Inf. Theor.*, **38**, 713–718.

Cook, N.J. (1985). *The Designer's Guide to Wind Loading of Building Structures*. London: Butterworths.

Crouse, M.S., Nowak, R.D., & Baraniuk, R.G. (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Sig. Proc.*, **46**, (to appear).

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

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

DeHann, R., J., Patrick, Chess, G., & Jaco, N. (1977). Definition of sleep state in the newborn infant by heart rate analysis. *Am. J. Obstet. Gynecol.*, **127**, 753–758.

Delyon, B., & Juditsky, A. (1995). Estimating wavelet coefficients. *Pages 151–168 of:* Antoniadis, A., & Oppenheim, G. (eds), *Wavelets and Statistics, Lecture Notes in Statistics 103*. New-York: Springer-Verlag.

Donoho, D.L. (1992). *Interpolating wavelet transforms*. Tech. rept. 408. Department of Statistics, Stanford University, Stanford.

Donoho, D.L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harm. Anal.*, **2**, 101–126.

Donoho, D. L., & Johnstone, I.M. (1994a). Ideal denoising in an orthonormal basis chosen from a library of bases. *Compt. Rend. Acad. Sci. Paris Ser. A*, **319**, 1317–1322.

Donoho, D.L., & Johnstone, I.M. (1994b). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D.L., & Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Ass.*, **90**, 1200–1224.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., & Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. R. Statist. Soc. B*, **57**, 301–337.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., & Picard, D. (1996). Density-estimation by wavelet thresholding. *Ann. Statist.*, **24**, 508–539.

Gao, H.-Y. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum. *J. Time Series Anal.*, **18**, (to appear).

Gath, I., & Baron, E. (1980). Computerized method for scoring of polygraphic sleep recordings. *Comput. Prog. Biomed.*, **11**, 217–223.

Hall, P., & Nason, G.P. (1997). On choosing a non-integer resolution level when using wavelet methods. *Statist. Probab. Lett.*, **34**, 5–11.

Hall, P., & Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.*, **23**, 905–928.

Harper, R.M., Schechtman, V.L., & Kluge, K.A. (1987). Machine classification of infant sleep state using cardiorespiratory measures. *Electroenceph. Clin. Neurophysiol.*, **67**, 379–387.

Harvey, A. (1993). *Time Series Models*. 2nd edn. New York: Harvester Wheatsheaf.

Haslett, J., & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resource. *J. R. Statist. Soc. C*, **38**, 1–50.

Jawerth, B., & Sweldens, W. (1994). An overview of wavelet based multiresolution analyses. *SIAM Rev.*, **36**, 377–412.

Johnstone, I.M., & Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise. *J. R. Statist. Soc. B*, **59**, 319–351.

Kovac, A. (1997). *Wavelet thresholding for unequally spaced data*. Ph.D. thesis, Department of Mathematics, University of Bristol, Bristol.

Krzanowski, W.J., Jonathan, P., McCarthy, W.V., & Thomas, M.R. (1995). Discriminant analysis with singular covariance matrices: methods and application to spectroscopic data. *J. R. Statist. Soc. C*, **44**, 101–115.

Lang, M., Guo, H., Odegard, J.E., Burrus, C.S., & Wells, R.O. (1995). Nonlinear processing of a shift invariant DWT for noise reduction. *Pages 640–651 of:* Szu, H.H. (ed), *Proc. SPIE. Wavelet Applications II*, vol. 2491. Bellingam, Washington: SPIE.

Learned, R.E., & Willsky, A.S. (1995). A wavelet packet approach to transient signal classification. *Appl. Comput. Harm. Anal.*, **2**, 265–278.

Mallat, S. G. (1989a). Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Trans. Am. Math. Soc.*, **315**, 69–87.

Mallat, S. G. (1989b). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattn Anal. Mach. Intell.*, **11**, 674–693.

Mallat, S.G., & Zhang, Z. (1993). Matching pursuit in a time-frequency dictionary. *IEEE Trans. Sig. Proc.*, **41**, 3397–3415.

Mardia, K.V., Kent, J.T., & Bibby, J.M. (1979). *Multivariate Analysis*. London: Academic Press.

McCoy, E.J., Percival, D.B., & Walden, A.T. (1995). *Spectrum estimation via wavelet thresholding of multitaper estimators*. Tech. rept. TR-95-14. Statistics Section, Imperial College, London.

McCullagh, P., & Nelder, J.A. (1989). *Generalized Linear Models*. 2nd edn. London: Chapman and Hall.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.

Mintzer, F. (1982). On half-band, third-band and Nth band FIR filters and their design. *IEEE Trans. Acoust. Speech Sig. Proc.*, **30**, 734–738.

Mintzer, F. (1985). Filters for distortion-free two-band multirate filter banks. *IEEE Trans. Acoust. Speech Sig. Proc.*, **33**, 626–630.

Moulin, P. (1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Sig. Proc.*, **42**, 3126–3136.

Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *J. R. Statist. Soc. B*, **58**, 463–479.

Nason, G.P., & Silverman, B.W. (1994). The discrete wavelet transform in S. *J. Comput. Graph. Statist.*, **3**, 163–191.

Nason, G.P., & Silverman, B.W. (1995). The stationary wavelet transform and some statistical applications. *Pages 281–300 of:* Antoniadis, A., & Oppenheim, G. (eds), *Wavelets and Statistics, Lecture Notes in Statistics 103*. New-York: Springer-Verlag.

Neumann, M.H., & von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.*, **25**, 38–76.

Ogden, T., & Parzen, E. (1996). Change-point approach to data analytic wavelet thresholding. *Statist. Comput.*, **6**, 93–99.

Percival, D.B., & Guttorp, P. (1994). Long-memory processes, the Allan variance and wavelets. *Pages 325–344 of:* Foufoula-Georgiou, E., & Kumar, P. (eds), *Wavelets in Geophysics*. San Diego: Academic Press.

Percival, D.B., & Mofjeld, H. (1997). Multiresolution and variance analysis using wavelets with application to subtidal coastal seas. *J. Am. Statist. Ass.*, **92**, (to appear).

Pesquet, J.C., Krim, H., & Carfantan, H. (1996). Time-invariant orthonormal wavelet representations. *IEEE Trans. Sig. Proc.*, **44**, 1964–1970.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Ray, S., Lee, W., Morgan, C., & Airth-Kindree, W. (1986). Computer sleep state scoring — an expert system approach. *Int. J. Biom. Comput.*, **19**, 43–61.

Saito, N. (1994). *Local feature extraction and its applications using a library of bases*. Ph.D. thesis, Yale University, New Haven.

Saito, N., & Coifman, R.R. (1996). On local feature-extraction for signal classification. *Z. Angew. Math. Mech.*, **76**, 453–456.

Sawczenko, A., Young, J., & Fleming, P. (1997). Variation of rectal and body temperature with sleep state. *Early Hum. Dev.*, **47**, (To appear).

Schechtman, V.L., Kluge, K.L., & Harper, R.M. (1988). Time-domain system for assessing variation in heart-rate. *Med. Bio. Eng. Comput.*, **26**, 367–373.

Shensa, M.J. (1992). Discrete wavelet transforms: wedding the a trous and Mallat algorithms. *IEEE Trans. Sig. Proc.*, **40**, 2464–2482.

Smith, M.J.T., & Barnwell, T.P. (1986). Exact reconstruction techniques for tree-structured subband coders. *IEEE Trans. Acoust. Speech Sig. Proc.*, **34**, 434–441.

Strang, G. (1993). Wavelet transforms versus Fourier transforms. *Bull. (New Series) Am. Math. Soc.*, **28**, 288–305.

von Sachs, R. (1996). Adaptively wavelet-smoothed Wigner estimates of evolutionary spectra. *Z. Angew. Math. Mech.*, **76**, 71–74.

von Sachs, R., & Schneider, K. (1996). Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *Appl. Comput. Harm. Anal.*, **3**, 268–282.

von Sachs, R., Nason, G.P., & Kroisandt, G. (1996). Spectral representation and estimation for locally-stationary wavelet processes. *Proceedings of the workshop "Spline functions and wavelets": Montreal*. (to appear).

von Sachs, R., Nason, G.P., & Kroisandt, G. (1997). *Adaptive estimation of the evolutionary wavelet spectrum*. (in preparation).

Walden, A.T., & Contreras Cristan, A. (1997). *The phase-corrected undecimated discrete wavelet packet transform and the recurrence of high latitutde interplanetary shock waves.* Tech. rept. TR-97-03. Statistics Section, Imperial College, London.

Wickerhauser, M.V. (1994). *Adapted Wavelet Analysis from Theory to Software.* Wellesley, Massachusetts: A.K. Peters.