Statistics and Computing **12**: 45–56, 2002

Wavelet packet transfer function modelling of nonstationary time series

GUY P. NASON* and THEOFANIS SAPATINAS†

*Department of Mathematics, University of Bristol, England
G.P.Nason@bristol.ac.uk
†Department of Mathematics and Statistics, University of Cyprus, Cyprus

Received March 2000 and accepted September 2000

This article shows how a non-decimated wavelet packet transform (NWPT) can be used to model a response time series, Y_t, in terms of an explanatory time series, X_t. The proposed computational technique transforms the explanatory time series into a NWPT representation and then uses standard statistical modelling methods to identify which wavelet packets are useful for modelling the response time series. We exhibit S-Plus functions from the freeware `WaveThresh` package that implement our methodology.

The proposed modelling methodology is applied to an important problem from the wind energy industry: how to model wind speed at a target location using wind speed and direction from a reference location. Our method improves on existing target site wind speed predictions produced by widely used industry standard techniques. However, of more importance, our NWPT representation produces models to which we can attach physical and scientific interpretations and in the wind example enable us to understand more about the transfer of wind energy from site to site.

Keywords: nonstationary transfer function, nondecimated wavelet packets, wind time series, `WaveThresh`

1. Introduction

This article models the relationship between a *response* time series, $\{Y_t\}_{t\in\mathbb{Z}}$, and an *explanatory* time series, $\{X_t\}_{t\in\mathbb{Z}}$. We hope that any model we choose might be interesting in its own right but we shall also be interested in using it to predict future values of Y_t from future values of X_t. When both time series fall into the class of ARMA type models then it is appropriate to use "transfer function" models (see, e.g. Priestley 1981, Chap. 9). However, our modelling methodology can be used when either or both time series are not stationary although it is intended for series that exhibit patches of stationarity or are locally stationary (e.g. fall into the class of oscillatory processes, see Priestley 1981, Chapter 11, locally stationary (Fourier) processes, see Dahlhaus 1997 or locally stationary wavelet processes, see Nason, von Sachs and Kroisandt 2000). Although established "transfer function" models are usually exclusively frequency domain quantities we widen the scope of the term to include the time-frequency quantities described in this article.

The models that we build first express X_t in terms of (non-decimated) wavelet packets which analyse X_t at different scales, frequencies and locations. Then standard statistical modelling techniques can be used to relate Y_t to the non-decimated wavelet packet transform (NWPT) of X_t. The selected model often reveals valuable information about which types of oscillatory behaviour in X_t influence Y_t and also supplies a method to predict future values of Y_t from *future* values of X_t. Appendix 1 provides instructions on how to compute the NWPT and use all the methods described in this article in the freeware `WaveThresh` package for S-Plus.

We do not (yet) have a theoretical formulation of our modelling procedure. Our aim is to introduce the computational method and show that it can produce interesting and verifiable results on real time series. Recently, Walden and Contreras Cristan (1998) used the NWPT in the analysis of a *single* non-stationary series of hourly averaged Southern Hemisphere solar magnetic field magnitude observations. Our work differs in that we relate a time series, Y_t, to the NWPT of *another* time series X_t. Ramsey and Lamport (1998) carried out similar analyses to ours but they only used standard *decimated* wavelets and decomposed *both* time series: X_t and Y_t which in their example were economic series of income and consumption.

They extract decimated wavelet coefficients at the same dyadic scale for each time series and then statistically model one set in terms of the other using linear regression (one for each scale). With decimated wavelets it is tricky to relate coefficients from different scales or relate coefficients from a scale to the original time series (because the number of coefficients and their location varies with scale). This article uses *non-decimated* transforms which have the same number of coefficients at each scale and coefficients within each scale are located according to the same time grid. Moreover, we use wavelet *packets* that can elicit a greater variety of behaviours than can wavelets alone.

Wavelet packets form an organized but extremely flexible class of functions of which wavelets are a subset. Section 3 gives a more detailed overview of non-decimated wavelet packets. Section 4 explains how we model Y_t in terms of the NWPT of X_t and describes the modelling advantages of using non-decimated wavelet packets. First though, in Section 2 we introduce our methodology in the context of modelling and predicting wind speeds at a proposed wind farm site from data taken at a reference wind speed site. This example contains all the essential features of our methodology: representation of the wind speeds at the reference site in terms of a NWPT; relating the wind speed at the proposed site to the NWPT and selecting the best NWPT variables for predicting the proposed site speeds and then interpreting those NWPT variables and evaluating their predictive performance.

2. Wind speed modelling and prediction

Before construction of a wind farm an 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 months. These target speeds are related to contemporaneous wind speeds measured at a nearby reference site (a Meterological Office station in the UK) and a model predicting target from 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 reference site. Modelling of this kind is described, for example, by Cook (1985) and Haslett and Raftery (1989). For reasons of cost, only one explanatory series $\{X_t\}_{t=1}^T$, for some integer $T > 0$, is usually available although our methodology could be easily extended if data from other Meteorological Office stations became available.

Figure 1 shows hourly wind speeds recorded at two Welsh Meteorological Office stations: Valley and Aberporth. Valley is located approximately 120 km north of Aberporth and they are mostly separated by Cardigan Bay. In the following example our aim is to model Valley's wind speeds ($\{Y_t\}_{t=1}^T$) in terms of those at Aberporth ($\{X_t\}_{t=1}^T$). We show how our modelling methodology can be used to predict the wind regime at Valley from

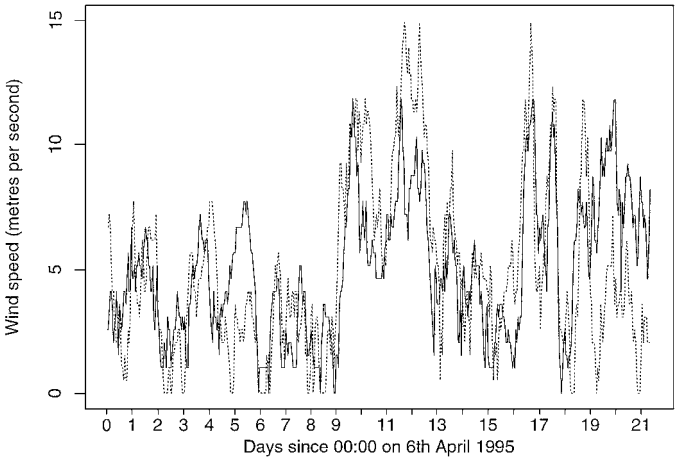

Fig. 1. *Hourly wind speeds from 00:00 on 6th April 1995 at Valley (solid line) and Aberporth (dashed line). (Data provided by M&N Wind Power)*

future Aberporth values and improves on existing methodology. More importantly, our model is highly physically interpretable (unlike existing methodology) and explains what types of wind activity at Aberporth are important for predicting Valley wind speeds.

We should emphasize that both X_t and Y_t are *not* stationary and so classical methods should not be used blindly (indeed, the wind speed relationship depends on the wind direction). It is possible that the speed relationship is piecewise stationary (linked to wind direction) so a modified form of classical "transfer function model" methodology might possibly be made to work. The obvious technical problem is how does one "join-up" the different segments of the series that exhibit similar stationarity? Piecewise stationarity, or more generally local stationarity does not cause problems for our wavelet packet methodology since the wavelet packets naturally adapt to the dominant local oscillatory behaviour.

2.1. *An established wind industry method*

Linear regression is extremely simple, effective and is widely used in practice (e.g. the measure-correlate-predict procedure from Hannah *et al.* 1996). First, the data is divided into (typically) twelve 30° *direction sectors* based on the direction of the wind at Aberporth. Then 12 separate linear regression models are computed one for each direction sector. Predictions of the wind speed are easily obtained by using the current wind direction at Aberporth to select one of the twelve regression models and then predict the windspeed at Valley by $\hat{\alpha} + \hat{\beta} X_t$ where $\hat{\alpha}$ and $\hat{\beta}$ are the fitted regression parameters for that particular sector. If the number of time series observations T is small then sometimes fewer direction sectors are chosen.

Wind speeds are usually non-normal, serially correlated and also subject to measurement error so typically a robust

regression method that takes account of the measurement error is used. However, the results from using the simple methodology described here are typically good probably because of the large number of data points used to build the regression models.

2.2. *Our wavelet packet method*

This section explains what we do with wavelet packets although we are conscious that we have not yet formally defined what wavelet packets are. To gain an overview of our methodology it is enough to know that wavelet packets are oscillatory basis functions from some large library. However, for specific details about the particular transform refer forward to Section 3.2.

Rather than build a model directly between Y_t and X_t we build a model between Y_t and a NWPT version of X_t. The NWPT generates $K = 2T - 2$ derived time series which we stack together to make a K-dimensional multivariate time series $\mathbf{X}_t$. Each variable of $\mathbf{X}_t$ quantifies how similar X_t is to that particular wavelet packet at time t. In other words each component of $\mathbf{X}_t$ tells us "how much" of each wavelet packet there is in X_t at any particular time t. The decomposition of X_t into K different wavelet packet components is extremely useful since we can subsequently model Y_t in terms of the components using standard statistical methodology. To summarize: X_t is the "explanatory" time series and $\mathbf{X}_t$ is the "collection of NWPT coefficients" of X_t.

The wavelet packet transformation analyses X_t using a diverse collection of wavelet packets at different scales, frequencies and locations. Their diversity is the reason why our methods can handle piecewise or local stationarity as the wavelet packets will activate and deactivate as particular behaviours appear and disappear in different regions. Wavelets and wavelet packets come in families. Given a particular *mother wavelet* one can derive all its wavelet packets however the choice of a mother wavelet has to be made. There are no hard rules about the choice of a mother wavelet even in areas as well developed as, say, wavelet shrinkage for curve estimation. For the examples below we shall use the Haar mother wavelet which works well and produces interesting scientific results (although in Nason, Sapatinas and Sawczenko 2001 we use a smoother Daubechies mother wavelet). How the choice of wavelet affects the final model and its interpretation is an area for future research.

To exhibit our methodology on the wind energy time series we applied the NWPT to a segment from X_t of length $T = 512$. This transform resulted in $K = 1022$ transformed time series which we stacked to make a 1022-dimensional multivariate time series $\mathbf{X}_t$ of length $T = 512$. We then modelled Y_t in terms of $\mathbf{X}_t$.

In general the statistical modelling step is completely straightforward because our *non-decimated* wavelet packet transform allows us to use many of the widely available statistical methods for modelling a response vector, Y_t, in terms of a multivariate descriptor, $\mathbf{X}_t$. Use of the standard (decimated) wavelet packet transform would not permit us to model Y_t directly in terms of the wavelet packet coefficients of X_t because the coefficients exist on different time scales to Y_t. Ramsey and Lampart (1998) demonstrate the utility of relating decimated *wavelet* coefficients of $\{Y_t = \text{consumption}\}$ and $\{X_t = \text{income}\}$ at the same scale. However, they do not directly provide a model for the series Y_t itself, nor predict Y_t from future X_t (both of which are possible to do with non-decimated transforms), nor do they use the increased flexibility of wavelet packets.

Returning to our wind energy example note that the number of variables ($K = 1022$) in the K-dimensional time series, $\mathbf{X}_t$, is always larger than the sample size $T = 512$. In this article we use a crude initial dimension reduction technique to reduce the extremely large dimensional problem into a more manageable but still large size. We then use more incisive statistical modelling techniques to reduce the dimensionality further to identify our "best" model. We stress that it is not the aim of this article to discuss and develop a comprehensive variable selection methodology and we intend to investigate alternatives.

Initially, we used various familiar statistical procedures to model Y_t in terms of $\mathbf{X}_t$. However, residual plots showed that our models were systematically in error with the error magnitude strongly related to the wind direction at Aberporth. To improve our model we inserted an extra wind direction sector factor variable: DIR. (The DIR factor has twelve levels corresponding to winds in the different 30° direction sectors. See Table 2 for a list.) We applied the following crude variable selection approach to select a subset of the $K = 1022$ variables: we selected an arbitrary 5% of variables which correlated best with Y_t and labelled the resultant $K_1 = 51$ variables which we label S1 to S51. (We admit that this procedure is somewhat rough and ready but it is an adequate initial dimension reduction step which produces interesting practical results.) Then we used generalized linear/additive modelling and CART to find a good model for the Valley wind speeds Y_t in terms of the dimension reduced $\mathbf{X}_t$. The final selected model was a generalized linear model (GLM) obtained using backwards variable selection that assumed Gamma distributed Y_t values (wind speeds are positive and skewed to the right) with a log link function. Instructions on how to perform the NWPT and select the "best" 5% of the variables in `WaveThresh` are presented in Appendix 1.

2.3. *Our wavelet packet model and interpretation*

Tables 1 and 2 show the coefficients of the final GLM model

$$\log(Y_t) \sim \mu + \text{S2} + \text{S20} + \text{S35} + \text{S39} + \text{DIR}. \tag{1}$$

Table 1. *Significant wavelet packets included in the final GLM. The table shows the value of the coefficient in the linear model along with the resolution level that the term corresponds to and its frequency index within that resolution level*

Term	Packet level	Frequency index	Term coefficient ($\times 1000$)
Intercept			8300
S2	7	0	−13
S20	2	15	38
S35	1	6	−10
S39	0	33	−47

Table 2. *GLM coefficients for the factor DIR in the final GLM along with the associated direction sectors*

Term	Direction sector (degrees)	Term coefficient ($\times 1000$)
DIR1	0–29	−31
DIR2	30–59	−18
DIR3	60–89	22
DIR4	90–119	21
DIR5	120–149	5
DIR6	150–179	−2
DIR7	180–209	12000
DIR8	210–239	−1300
DIR9	240–269	−1100
DIR10	270–299	−870
DIR11	300–329	−720
DIR12	330–359	Aliased

The final model is highly interpretable. The DIR factor can be interpreted as a multiplier reflecting the strength of association between the wind speeds at the two sites. Since the two sites are in a NNE/SSW line there is a large multiplier of 12 when the wind direction is in sector DIR7 (when the wind is at right angles to this in DIR10 the effect of all the other variables is shrunk by the multiplier −0.87). This effect is enhanced when the wind comes from a southerly or westerly ($180 \rightarrow 330^\circ$) direction rather than a northerly or easterly ($330 \rightarrow 180^\circ$) direction which is natural given the prevailing wind directions in the UK from the west and south.

However, the DIR factor only multiplies the linear predictor in the final model by a fixed amount depending on the wind direction. The four wavelet packets actually model the variations in wind speed over time and they too are interpretable. Figure 2 shows pictures of the wavelet packets in the model obtained using the `drawwp.default()` function from `WaveThresh` which requires knowledge of the underlying wavelet, the resolution level and packet index of the particular wavelet (which can be obtained from the `filter, level` and `pktix` components of the `wpstRO` object described in Appendix 1). Each plot in Fig. 2 contains a vertical dashed line at $t = 0$ which serves as an origin for obtaining wavelet packet coefficients of a series. Given this, the interpretation of each of the plots in Fig. 2 is as follows:

(a) S2 is a wavelet packet (actually father wavelet) which averages X_t over the previous four hours. Inclusion of this wavelet packet indicates that the series $X_t + X_{t-1} + X_{t-2} + X_{t-3}$ is important for prediction.

(b) S20 is a wavelet packet with average oscillation frequency of just over 23 hours. We assume that this wavelet packet captures daily variation in wind speed. Note however, that the oscillation only occurs over the previous five days. So daily variation is important for prediction, but only the past five days is relevant.

(c) S35 is a wavelet packet with average oscillation frequency of 4.7 days. It is well-known that wind speeds oscillate at or near this frequency. Indeed, this frequency falls into the middle of the "macrometeorological peak" and is associated with the large-scale pressure systems passing overhead (e.g. Cook 1985, van der Hoven 1957).

(d) S39 is a wavelet packet which mostly oscillates over the whole series at a frequency of 16 hours except for the period around $t = 0$ where it averages over the immediate eight hours into the past and future. It is difficult to attach a direct meteorological interpretation to this wavelet packet, although wind takes approximately eight hours to travel between Aberporth and Valley assuming a mean wind speed of 4 to 5 ms^{-1}.

The S39 wavelet packet takes values equally from the future as well as from the past which is perfectly legitimate mathematically but would be a problem for real-time prediction. It would be possible to restrict our methodology to only include wavelet packets that do not overlap with future data.

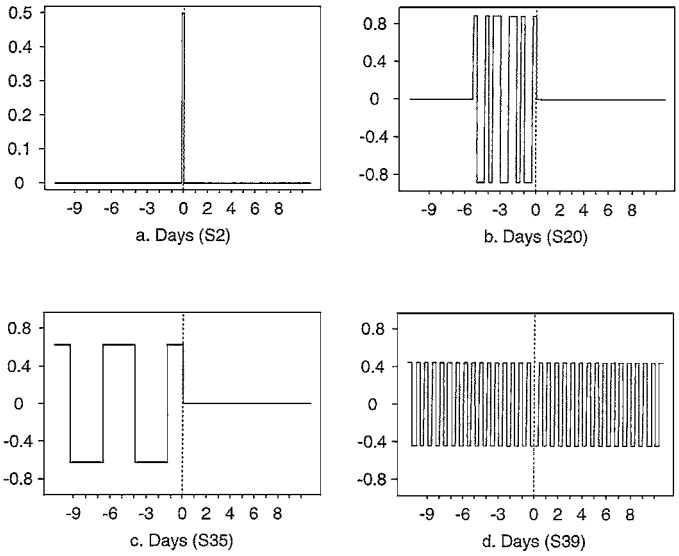

Fig. 2. *The Haar wavelet packets used in the final model. The vertical dashed line in each plot corresponds to time t. Each wavelet packet is indexed by a pair: (resolution level, frequency index within a level). They are (clockwise, from top left): a. S2: The wavelet packet (7,0) (father wavelet); b. S20: The wavelet packet (2,15); c.S35: The wavelet packet (1,6); d. S39: The wavelet packet (0,33). The wavelet packets S20, S35 and S39 are plotted on the same vertical scale which is 10× that of the scale for S2. For the physical interpretation of these wavelet packets in the model see text*

Note that the modelling above indicates that wavelet packets were useful and interpretable, indeed in this practical example none of the selected basis functions are actually wavelets (this shows the need for the more complicated wavelet *packet* transform).

2.4. *Model predictions*

For this example the differences in prediction between the established industrial method and our new methodology are not large.

In fact, both methodologies often make the same mistakes. Generally speaking residual plots show that our new methodology is slightly better (but remember that we also obtain a wealth of extra interpretable and scientific information as outlined above.) For wind energy prediction an estimate of the distribution of wind speeds at the proposed site is required. Our methodology provides a better estimate of the wind speed distribution because it provides a better model of the transfer between reference and proposed sites.

Formally our model is also better in terms of mean residual sums of squares (MRSS) for predicting wind speeds another 21 days ahead (ours is 0.088, the established industrial method is 0.094). The other interesting feature is that our model is better over the early parts of the prediction interval: over 10 days our MRSS is 0.12, the established MRSS is 0.14; over five days our MRSS is 0.19, the established MRSS is 0.23. Roughly speaking our model is 10% better than the established industrial method. Although 10% does not sound very much in absolute terms it can make a lot of difference to the wind power output statistics for a proposed wind farm (wind power output is related to the cube of the speed) and hence to the economics and viability of the proposed farm.

Instructions on how to use `WaveThresh` to generate predictions using our models appear in Appendix 1. The next section explains how to compute the NWPT and what the coefficients of the transform mean.

3. Representations using non-decimated wavelet packets

Our goal is to model the relationship between the response time series, $\{Y_t\}_{t\in\mathbb{Z}}$, in terms of the local scale and frequency properties of the explanatory time series, $\{X_t\}_{t\in\mathbb{Z}}$, which we obtain using a non-decimated wavelet packet transform. Hess-Nielsen and Wickerhauser (1996) give an excellent description of wavelet packets and explain how they reveal information about the variation of signals in time and frequency. In particular, they make a nice analogy between musical score notation which indicates the pitch, duration and position of individual notes and a wavelet packet analysis which gives information about the frequency, scale and position of energy in a time series. The analogy cannot be pushed too far as musical notation has many more subtleties.

3.1. *The time-frequency plane and wavelets*

The properties of wavelet packets can be conveniently explained by introducing the *time-frequency* plane. This two-dimensional plane represents time along the horizontal axis and frequency along the vertical axis. Waveforms (segments of time series) can be schematically represented by areas in the time-frequency plane with their width indicating duration and height indicating frequency bandwidth.

As an example consider a simple time series X_t for $t = t_i = 1, \ldots, T = 16$. Figure 3 shows two familiar tilings of the time-frequency plane. The left diagram in Fig. 3 corresponds to a representation of a time series concentrated purely in time with each vertical line representing a particular (Dirac) time basis element. The right diagram corresponds to a representation concentrated purely in frequency with each horizontal line representing a particular frequency basis element at Fourier frequencies $\exp(-i\omega_n)$ for $\omega_n = 2\pi n/16$ and $n = 0, \ldots, 15$. Each line in either of these plots could be given a separate grey-shade intensity to indicate the "contribution to variance" at each location in time (for Dirac) or frequency (for Fourier). For example, if X_{13} was a large value then the 13th line would be a dark shade of grey, if X_5 was a small value then the fifth line would be light and so on.

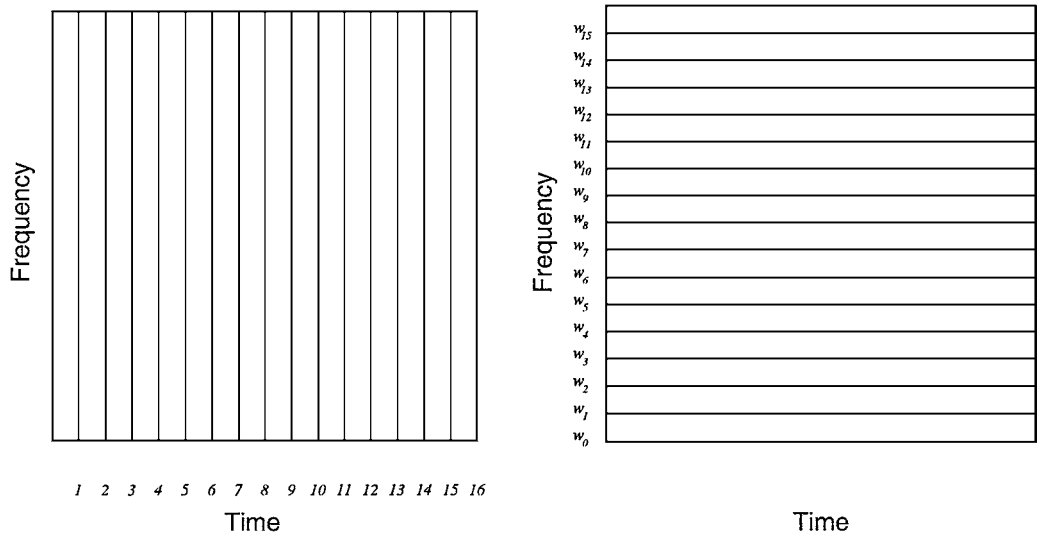

Fig. 3. *Two tilings of the time-frequency plane. Left: Dirac; Right: Fourier*

Wavelets offer an alternative, but fixed, tiling of the time-frequency plane. See Daubechies (1992) or Burrus, Gopinath and Guo (1998) for introductions to wavelets in this context or see Nason and Silverman (1994), Antoniadis (1997), Ogden (1997), Vidakovic (1999) or Abramovich, Bailey and Sapatinas (2000) for statistical introductions. Given a suitable mother wavelet, $\psi(t)$, a set of wavelets

$$\{\psi_{j,k}(t)\}_{j,k\in\mathbb{Z}} \quad \text{where} \quad \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \tag{2}$$

can form a basis for function spaces such as $L^2(\mathbb{R})$ (or indeed more complicated function spaces, see Abramovich, Sapatinas and Silverman (1998)). Since wavelets form bases we can represent functions $f(t)$ as linear combinations of wavelets by

$$f(t) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(t). \tag{3}$$

If the wavelet basis is orthonormal, as it will be throughout this article, then we can write the wavelet coefficients, $\{d_{jk}\}_{j,k\in\mathbb{Z}}$ of $f(t)$ as

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

In this article we advocate the popular Daubechies' (1992) compactly supported wavelets. The Daubechies' family is useful for several reasons: the degree of smoothness of the mother wavelet can be preselected, the associated discrete transforms are fast and efficient and smooth structure is sparsely represented. The Haar wavelet used in the wind example earlier is the least smooth Daubechies' wavelet but seemed to work best in that practical situation. Formula (3) demonstrates that $f(t)$ can be represented by basis functions, $\psi_{jk}(t)$ at different scales proportional to 2^{-j} for integers j i.e. a multiresolution analysis. The mother wavelet is usually chosen to be a short-duration oscillation and therefore localized both in time (short-duration) and in frequency (because it oscillates). The derived wavelets are scaled and translated (by $2^{-j}k$) versions of the mother wavelet: the scaling and translation operations permit analysis of time series at different times and frequencies (time-frequency analysis). Figure 4 shows how wavelets tile the time-frequency plane: as j gets larger the wavelets become finer and finer scale objects, oscillate more quickly, are packed closer together (top of figure) and the corresponding tiles get taller (they cover a wider frequency range) and thinner (their duration is less). As j gets smaller the opposite happens. The amplitude of any wavelet coefficient in a representation, $|d_{jk}|$, can be indicated on Fig. 4 by grey-scale intensity shading of the rectangle corresponding to the wavelet basis function in question.

Mallat's (1989) pyramid algorithm permits fast order $\mathcal{O}(T)$ computation of the discrete wavelet transform (DWT) of a discrete time series $X = \{X_t\}_{t=1}^T$ where $T = 2^J$ for some positive integer J. The DWT provides information about variation in a time series at different scales and locations (like the d_{jk} above). To fix notation let $\mathcal{H}$ and $\mathcal{G}$ denote the filter operators corresponding to the low- and high-pass quadrature mirror filters of some Daubechies' (1992) compactly supported wavelet with filter coefficients $\{h_k\}$ and $\{g_k\}$, both of length L. Thus the computational effort for applying $\mathcal{H}$ or $\mathcal{G}$ once is of order $\mathcal{O}(L)$. Let $\mathcal{D}_0$ denote the even dyadic decimation operator defined by $(\mathcal{D}_0 X)_k = X_{2k}$, i.e. it selects every evenly indexed observation (see Nason and Silverman 1995 for a more comprehensive discussion). Then the DWT coefficients of X at level $j = 0, \ldots, J-1$ may be obtained by $d_j = \mathcal{D}_0\,\mathcal{G}(\mathcal{D}_0\mathcal{H})^{(J-j-1)}X$. We call the concatenated operators $\mathcal{D}_0\mathcal{H}$ and $\mathcal{D}_0\,\mathcal{G}$ *packet operators*. We can again use Fig. 4 to schematically visualise the DWT of $T = 16 = 2^4$ data points. At level $j = 3$ the vector d_3 will contain 8 DWT coefficients corresponding to the eight tiles at the highest frequency in Fig. 4. At levels $j = 2$, 1 and 0 the vectors d_2, d_1 and d_0 contain 4, 2, and 1 coefficients corresponding to the tiles (wavelets) at those levels.

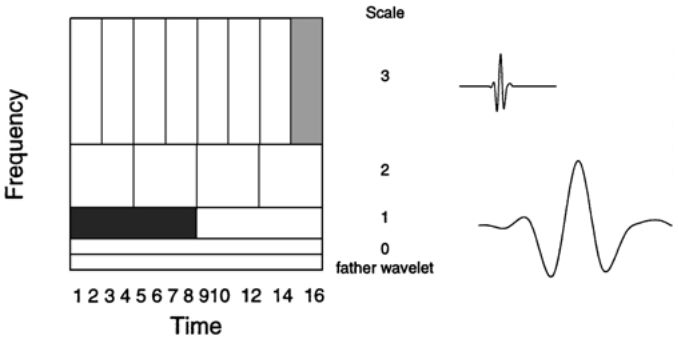

Fig. 4. *Left: wavelet tiling of the time-frequency plane. The top row of the tiling corresponds to fine scale, high frequency wavelets (top right) that "exist" over a short time scale, the bottom but one row corresponds to the largest scale, lowest frequency wavelet which "exists" over the whole time domain (for this example). The very bottom tile corresponds to the father wavelet (not shown) which again "exists" over the whole time domain at the very lowest frequency. Right: shows the wavelets (both derived from the same mother wavelet) at scales 1 and 3. At scale 3 there are eight fine scale wavelets roughly centred on locations $2k - 0.5$ for $k = 1, \ldots, 8$, at scale 1 there are two large scale wavelets centred roughly on locations 4.5 and 12.5. Note that the lighter shading in the box at scale 3 causes the wavelet to the right to have small amplitude. If the boxes had the same shade then the finer scale wavelet would actually be taller than the coarser scale one because $2^{3/2}$ is larger than $2^{1/2}$ in formula (2)*

3.2. *Wavelet packets*

Wavelets are a subset of a larger class of oscillatory functions called *wavelet packets*. A wavelet packet library is a collection of bases where the basis elements are no longer restricted to just scaling and dilation of one mother wavelet. As an illustration Fig. 5 shows four such wavelet packets from different scales with varying locations and frequencies. For applications one does not need to have a detailed knowledge of what all the wavelet packets look like or indeed need to have detailed formulae for them. The reason for this is that the statistical methodology in Section 4 picks out important basis functions and then effort can be put into interpretation of the selected functions as was done earlier in Section 2.3 using Fig. 2. A

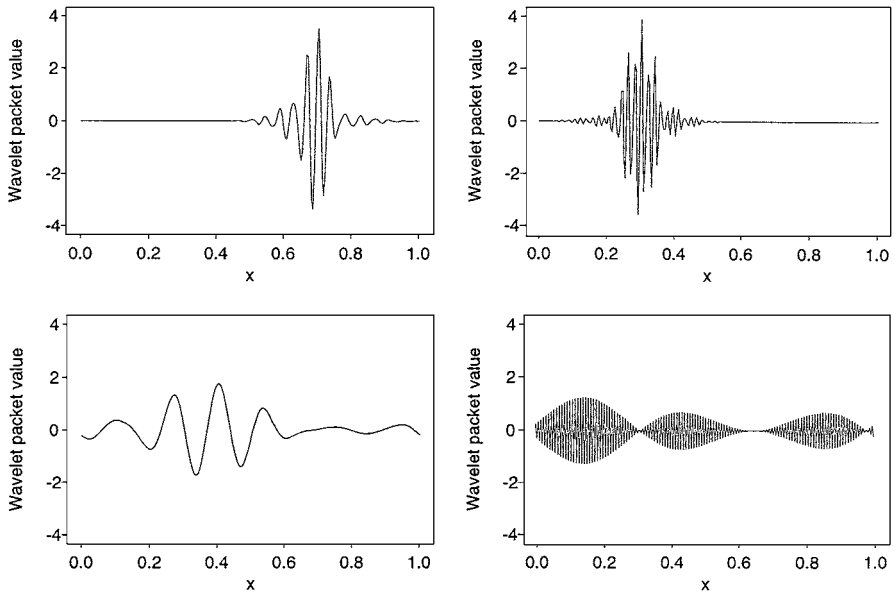

Fig. 5. *Four wavelet packets derived from Daubechies' (1992) least-asymmetric mother wavelet with 10 vanishing moments. These four wavelet packets are actually orthogonal and drawn by the* `drawwp.default()` *function in* `WaveThresh`*. The vertical scale is exaggerated by 10×*

comprehensive description of wavelet packets is beyond the scope (and length) of the present article. See Wickerhauser (1994) or Hess-Nielsen and Wickerhauser (1996) for useful technical introductions. Wavelet packets permit function representation using a basis selected from the library of wavelet packet bases where each basis element is indexed by scale, location and frequency (number of oscillations). For a discrete time series the wavelet packet transform (WPT) computes the representation of the series with respect to *all* basis functions in the library efficiently in $\mathcal{O}(T \log T)$. The WPT is computed by a full binary recursion of the $\mathcal{D}_0\mathcal{H}$ and $\mathcal{D}_0\mathcal{G}$ operators operating on the time series X (i.e. both $\mathcal{D}_0\mathcal{H}$ and $\mathcal{D}_0\mathcal{G}$ are first applied to X producing a low- and high-passed set of coefficients, then both operators are applied to each set producing four sets of coefficients, then both are applied again to each set producing eight sets and so on). Figure 7 depicts the algorithm for an initial set of eight data points. The total number of different wavelet packets (e.g. total number of rectangles in Fig. 7 not counting the original data) can be computed by realizing that there are 2 packets at scale level $J-1$, 4 packets at level $J-2, \ldots,$ 2^J packets at level 0. The total of $2+4+\cdots+2^J = 2^{J+1}-2 = 2T-2 = K$ as mentioned earlier. More details about wavelet packets and how they fit into a non-decimated scheme (next section) appear in Appendix 2.

Wavelet packets have been used in signal representation and compression. For example, Coifman and Wickerhauser (1992) use the WPT to represent a given signal in terms of all wavelet packets and then select a "best-basis" for representing that signal. Their "best-basis" is one in which the representation is sparse and an entropy criteria is used in an efficient "branch-and-bound" algorithm to select the sparsest basis. Sparse function representations are obviously useful for compression. We use wavelet packets because they can efficiently represent a wide range of time series behaviour: e.g. transients, local and prolonged oscillations.

Each wavelet packet basis provides a particular tiling of the time-frequency plane: there are a large number of possible distinct tilings/bases (let B_J denote the number of bases in a WPT tree with J levels. Then $B_1 = 1$ and $B_j = B_{j-1}^2 + 1$ which increases *very* fast). We do not claim that collections of wavelet packets optimally represent every type of time-frequency behaviour, but as Hess-Nielsen and Wickerhauser (1996) point out "it is more reasonable to correct the deficiencies in fast transforms [like wavelet packets] rather than to wait for slow mathematically perfect transforms to catch up".

3.3. *Non-decimated wavelet packets*

Our goal is to spot any relationship between variation in the response time series and the time-scale behaviour of the explanatory time series expressed though its wavelet-packet coefficients. For the modelling that we have in mind, it is not possible to use the WPT algorithm directly because of the dyadic decimation—the consequence of which is to reduce the number of wavelet packet coefficients by a factor of two for each coarser scale level computed (in the continuous representation this just means that "large scale" basis functions are kept further apart than "small scale" basis functions because functions are translated by shifts of $2^{-j}k$ which is larger for coarse scale wavelet packets). The NWPT uses exactly the same basis functions as the WPT but arranges for the wavelet packets on every scale to be present at *all* time locations. In other words for every time point in the explanatory series there is a corresponding wavelet packet coefficient for each of the wavelet packets at all scales. The

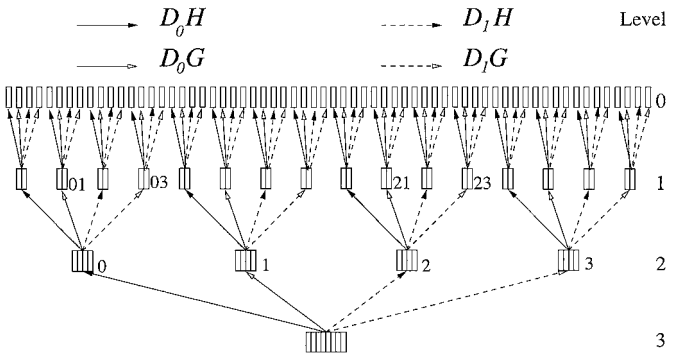

Fig. 6. *Schematic of NWPT for $T = 8$ points ($J = 3$)*

resulting NWPT representation is heavily overdetermined and non-orthogonal but the structure is advantageous for the modelling described in Section 4. Next we briefly summarise the key components of the NWPT.

First, let $\mathcal{D}_1$ denote the "odd" dyadic decimation operator which selects every odd indexed observation from a sequence. The NWPT simply applies the four packet operators $\mathcal{D}_0\mathcal{H}$, $\mathcal{D}_0\mathcal{G}$, $\mathcal{D}_1\mathcal{H}$ and $\mathcal{D}_1\mathcal{G}$ recursively to the time series to form a tree where each node has 4 children corresponding to the each of the packet operators. The NWPT was proposed by Pesquet, Krim and Carfantan (1996) and developed for curve estimation (wavelet shrinkage) by Cohen, Raz and Malah (1997). The NWPT is illustrated schematically in Fig. 6 for $T = 8$ data points input 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 indicate 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 Fig. 6; the 23 packet is so labelled because it is obtained by the operator $\mathcal{D}_1\mathcal{H}$ followed by $\mathcal{D}_1\mathcal{G}$. The computational cost of the NWPT can be found by noting that there are 4^{J-j} packets each of length 2^j for levels $j = 0, \ldots, J - 1$. Thus the total number of coefficients is

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

Therefore since each coefficient computed costs $\mathcal{O}(L)$ the total effort for the NWPT is $\mathcal{O}(T^2L)$. It is possible to compute non-decimated transforms with an arbitrary number of points T, but our software restricts us to data sets of length a power of 2. In practice, this is not a draconian restriction and may be overcome by padding with zeroes, for example (due to the time localization of the wavelet transform the extra zeroes do not affect the majority of the coefficients except at very coarse scales where they get included in their calculation. This is unlike standard spectral estimation where zero padding causes a different spectral interpolation (see Priestley 1981, Section 7.6).

3.3.1. Translation-equivariance of the NWPT

The NWPT has another important property called *translation equivariance* which means that if S is the cyclic shift operator then

$$\text{NWPT}[SX_t] = S(\text{NWPT}[X_t]).$$

In other words a shift in the time series is reflected by an identical shift in the transform coefficients and the relevance of this property for our modelling is explained in Section 4. Note that both the DWT and WPT are neither translation-equivariant nor translation-invariant and cannot be used for our particular purpose.

3.4. *Time-ordered non-decimated wavelet packets*

Although the NWPT produces as many wavelet packet coefficients as there are data points at each scale, the coefficients produced by the recursive algorithm are not delivered in time order. For example, packets 0 and 2 in Fig. 6 both correspond to the first stage of low-pass filtering but the coefficients in each of the NWPT packets would have to be interleaved to produce a sequence of time-ordered coefficients (i.e. both sets are father wavelet coefficients obtained by filtering with $\mathcal{H}$ but then one set has every even element selected by $\mathcal{D}_0$ and the other has every odd element selected by $\mathcal{D}_1$). Interweaving these two sets produces a set of eight coefficients *in time order* where each coefficient is associated with a corresponding father wavelet. Other interweavings are necessary to obtain time-ordered NWPT coefficients for other wavelet packets and details on how to achieve them are presented in Appendix 2.

Finally we note that the NWPT coefficients need to be phase-corrected to bring them into perfect time-alignment before use. Walden and Contreras Cristan (1998) specify a phase-correction technique but we use phase-shifts determined empirically here (by observing how delta functions are shifted) which work very well in practice. Note that the computational scheme in Walden and Contreras Cristan (1998) produces time-ordered sequences by default.

4. Modelling using NWPT

For every wavelet packet basis function the NWPT of X_t computes the coefficient of that basis function at *every* time location $t = 1, \ldots, T$—a large coefficient at t indicates that the X_t is behaving coherently with the particular wavelet packet function at that point. With the standard (decimated) WPT it is not possible to obtain coefficients at *all* time points.

The translation-equivariance of the NWPT is also critical in the following sense. If a certain behaviour occurs in X_t which is reflected in coefficients in the NWPT at t then if that behaviour

again happens at, say, $X_{t+\tau}$ it turns up again in exactly the same way in the NWPT at $t+\tau$. For a standard (decimated) DWT this again does not happen as the same behaviour in X at t and $t+\tau$ might we be reflected in completely different coefficients at these two locations. Earlier we described modelling Y_t in terms of the transformed version of X_t and clearly the lack of translation-equivariance would have caused problems (as certain behaviour in X at different times would not be represented consistently).

The final major advantage of using the NWPT as opposed to the (decimated) WPT is that for each wavelet packet basis function we obtain T coefficients—the same as the number of observations in the response time series $\{Y_t\}_{t=1}^T$. This means that we can represent the NWPT coefficients of $\{X_t\}_{t=1}^T$ as a K-dimensional multivariate time series $\{\mathbf{X}_t\}_{t=1}^T$ where each variable corresponds to the coefficients of a particular wavelet basis function. It is this property which enables us to make good use of the huge variety of statistical techniques (e.g. CART, multiple regression, logistic regression, GLMs, GAMs, Bayesian variable selection techniques) for modelling a response vector $(Y_1, \ldots, Y_T)$ in terms of multivariate explanatory variables $(\mathbf{X}_1, \ldots, \mathbf{X}_T)$ where each $\mathbf{X}_i$ is K-dimensional.

In Section 2 we noted that the number of variables ($K = 2T-2$) was greater than the number of observations (T). Many standard statistical techniques require $K < T$. Again, the aim of the current article is to show the utility of the NWPT and not dwell on the problem of "more variables than observations". However, we mention two techniques we have used. The first, very simple, approach is just to select a suitable number of those wavelet packet variables in $\mathbf{X}_t$ that correlate best with Y_t and then use standard techniques as described in more detail above. Although naive this approach is fast and has worked well in practice. As a further development, since the variables in $\mathbf{X}_t$ are correlated, it might be worth subjecting $\mathbf{X}_t$ to a principal components analysis and then use the new PCA projected variables to model Y_t. Secondly, Nason, Sapatinas and Sawczenko (2001) used the antedependence models of Krzanowski, Jonathan, McCarthy and Thomas (1995) to perform discrimination in the case of singular covariance matrices (when the number of variables is larger than the number of observations).

It is possible to use a non-decimated wavelet transform instead of a NWPT. Indeed, Nason, Sapatinas and Sawczenko (2001) demonstrated how non-decimated wavelets could be used to model infant sleep-state from ECG (electro-cardiogram) signals. In this case only wavelet functions are identified and the full generality of wavelet packets is not needed. However, we did not know that *a priori* only wavelets would be needed so generally speaking we use the full flexibility of the NWPT.

5. Conclusions and further work

This article has introduced a computational method for building a transfer function model using non-decimated wavelet packets for non-stationary time series. The transfer function model may well be interesting in itself or be useful for predicting future values of a response time series from future values of an explanatory time series.

Our methods could be easily extended to the case where there is more than one explanatory time series by using multivariate statistical procedures. Our methods could, in principle, also be extended to build non-decimated wavelet packet models between multidimensional objects although in practice algorithms for computing non-decimated wavelet packet transforms are probably only practical for low numbers of dimensions (1, 2 or 3). However, a two-dimensional non-decimated wavelet packet model may well be useful for relating images in many applied areas such as industrial inspection and control.

We intend to extend our work to both the multivariate and multidimensional settings. Although the main aim of this article is to introduce the (time-ordered) non-decimated wavelet packet transform as a tool for building transfer function models we recognise that the naive variable pre-selection in this article could be improved upon. Further, although the main goal of this article was to describe a computational technique we intend to supply a mathematical framework for the modelling described above: our intention is to represent both response and explanatory time series as locally stationary Fourier or wavelet processes and see how slowly varying models can be constructed according to the principles described above.

Appendix 1: S-Plus functions

We now describe the S-Plus functions used to implement the modelling methodology described in this article. These functions may be computed using Version 3 of the free `WaveThresh` package for S-Plus and R (see Nason and Silverman 1994 for details on Version 2 or `www.stats.bris.ac.uk/~wavethresh` for Version 3).

1. The NWPT can be carried out using the `wpst()` function:

   ```
   wpst(data, filter.number, family,
     FinishLevel)
   ```

   which computes the NWPT on the data set (time series) `data` using any of the Daubechies' compactly supported wavelets with `filter.number` vanishing moments from the `family` series up to a coarse resolution level specified by `FinishLevel`.
2. The NWPT and selection of the "best" 5% wavelet packet variables (described in Section 2.2) can be carried out using the `makewpstRO()` function:

   ```
   makewpstRO(timeseries, response,
     filter.number, family, trans, percentage)
   ```

   which applies the NWPT to `timeseries` (a vector containing the explanatory time series X_t) then finds the best K_1-sized subset of wavelet packet variables by correlating with the response time series contained in the vector

`response`. The wavelet with `filter.number` vanishing moments from the `family` series is used for the NWPT. The argument `trans` permits a mathematical transform to be applied to the NWPT coefficients before correlation with the response time series (much in the same way that a `log()` or `sqrt()` transform is used to stabilize variance in regression). The `percentage` argument specifies the proportion of the wavelet packet variables that are returned in this initial crude dimension reduction step. The `makewpstRO()` function returns a `wpstRO` class object which is an S-Plus `list` with the following components:

- `df`: a data frame containing the "best" K_1 wavelet packet variables;
- `ixvec`: an indexing vector which for each of the selected wavelet packets in `df` references the position of that packet in the matrix version of the NWPT of `timeseries`;
- `level` and `pktix`: the resolution level and packet index of each selected packet in the NWPT of `timeseries`;
- `nlevels`: the total number of resolution levels in the NWPT of `timeseries`;
- `cv`: a vector containing K_1 correlation coefficients between `response` and each of the selected K_1 variables;
- `filter` and `trans`: details of the wavelet filter and transform specified in the call to `makewpstRO()`.

Of these components the `$df` component can be supplied directly to commands such as `glm()` or `gam()` to e.g. model `response` on the K_1 selected NWPT coefficients of `timeseries` as described in Section 2.3.

3. Prediction of future values of Y_t from future values of X_t as described in Section 2.4 can be carried out using the `wpstREGR()` function:

```
wpstREGR(newTS, wpstRO)
```

which takes a new segment of X_t (possibly detrended to remove its mean) called `newTS` and uses the information stored in the existing `wpstRO` object to construct a new data frame containing the same K_1 variables as in the `wpstRO` but computed with the new time series data in `newTS`. The statistical model that was constructed (e.g. like (1)) can then be applied to the new data frame to obtain predicted values for Y_t.

Appendix 2: Weaving wavelet packets

There is no closed-form formula for the continuous wavelet packets derived from Daubechies' compactly supported wavelets. However their Fourier transform for $j = 0, \ldots, J-1$ can be written (see, for example, Daubechies 1992, p. 333) as

$$\hat{\psi}_{j;\epsilon_1,\ldots,\epsilon_{J-j}}(\omega)\left[\prod_{p=1}^{J-j} m_{\epsilon_p}(2^{-p}\omega)\right]\hat{\psi}(2^{-J-j}\omega), \qquad (4)$$

where $m_0(\omega) = \frac{1}{\sqrt{2}}\sum_k h_k e^{-i\omega k}$ and $m_1(\omega) = \frac{1}{\sqrt{2}}\sum_k g_k e^{-i\omega k}$, and $\hat{\psi}$ is the Fourier transform of the particular Daubechies' compactly supported wavelet determined by the quadrature mirror filters. The sequence $\epsilon_p = 0$ or 1 forces selection of the m_0 or m_1 at each level (in the WPT this is equivalent to following a $\mathcal{D}_0\mathcal{H}$ or $\mathcal{D}_0\mathcal{G}$ convolution branch in the binary tree respectively). The wavelet packets shown in Fig. 5 correspond to ϵ_p sequences of (clockwise from top left) 0010, 0101, 000010 and 1000001. Formula (4) also shows that the number of distinct wavelet packets at level $j = 0, \ldots, J-1$ is given by 2^{J-j}.

A distinction must be made between an "ordinary" NWPT packet (such as the ones in Fig. 6) 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. Time-ordered 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 Fig. 7). With the ordinary NWPT at level $J-1$ there appears to be four packets (see Fig. 6). However, one can also visualise the ordinary NWPT packets at level $J-1$ as two time-ordered *non-decimated* packets corresponding to the WPT by interweaving the four packets in the following way:

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

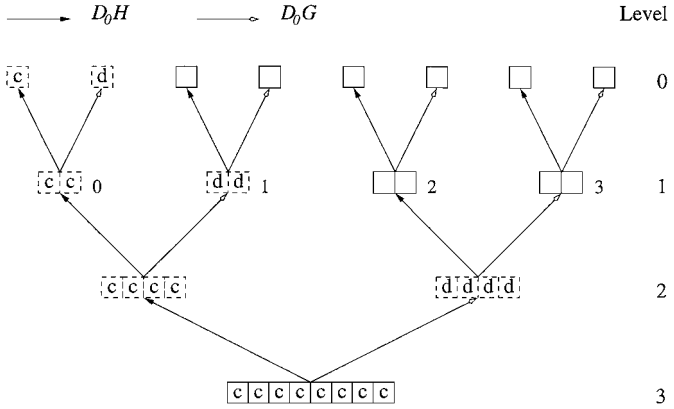

Fig. 7. *Schematic of the WPT operating on $T = 8$ points, i.e. $J = 3$. At level j there are 2^{J-j} packets each containing 2^j points. The numbers 0, 1, 2, 3 next to the packets at level 1 are the indices of packets within that level from left to right. The DWT coefficients are contained in the WPT and are shown in the dashed boxes and marked c and d. All other coefficients are with respect to other wavelet packets such as those illustrated in Fig. 5. The total number of wavelet packets (excluding the original data) is $2T - 2 = 14$. An arrow corresponds to convolution with the appropriate labelled operator*

Therefore the weaving process is a two-stage procedure: choose which time-ordered 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 Figs. 6 and 7. Suppose that we wished to obtain time-ordered NWPT packet of 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 3.3 (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 wavelet packets 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.

To obtain the ordinary NWPT indices associated with a time-ordered NWPT packet of index, r at level j, say ($r = 0, \ldots, 2^{J-j} - 1$, $j = 0, \ldots, J-1$, see Fig. 7 for details of the WPT indexing scheme) the following recursive procedure can be used:

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

$$1 \xrightarrow[2]{10} 1 \xrightarrow[10]{4} 1; 2 \xrightarrow[2]{10} 10 \xrightarrow[2]{10} 4; 3 \xrightarrow[10]{4} 11 \xrightarrow[10]{4} 5\Big).$$

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 S 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 of 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.

Finally we mention that the functions `getpacket.wpst ()` and `accessD.wpst ()` could be used to extract ordinary non-decimated wavelet packets and time-ordered non-decimated wavelet packets respectively (see the help on the `WaveThresh` package from the Web site mentioned in Appendix 1).

Acknowledgments

This research was supported by EPSRC Grant GR/K70236. The authors would like to thank Piers Guy of N&M Wind Power for providing the data and Stuart Barber, Steve Brooks, Peter Green, Bernard Silverman for providing helpful comments on earlier versions of this manuscript. We would like thank the referee for providing numerous comments which improved the article.

References

Abramovich F., Bailey T.C., and Sapatinas T. 2000. Wavelet analysis and its statistical applications. The Statistician 49: 1–29.

Abramovich F., Sapatinas T., and Silverman B.W. 1998. Wavelet thresholding via a Bayesian approach. Journal of the Royal Statistical Society, Ser. B. 60: 725–749.

Antoniadis A. 1997. Wavelets in statistics: A review (with discussion). Journal of the Italian Statistical Society 6: 97–144.

Becker R.A., Chambers J.M., and Wilks A.R. 1988. The New S Language. Wadsworth, Pacific Grove, CA.

Burrus C.S., Gopinath R.A., and Guo H. 1998. Introduction to Wavelets and Wavelet Transforms: A Primer. Prentice Hall, Upper Saddle River, NJ.

Cohen I., Raz S., and Malah D. 1997. Orthonormal and shift-invariant wavelet packet decomposition. Signal Processing 57: 251–270.

Coifman R.R. and Wickerhauser M.V. 1992. Entropy-based algorithms for best-basis selection. IEEE Transactions on Information Theory 38: 713–718.

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

Dahlhaus R. 1997. Fitting time series models to nonstationary processes. Annals of Statistics 25: 1–37.

Daubechies I. 1992. Ten Lectures on Wavelets. SIAM, Philadelphia.

Hannah P., Palutikof J.P., Rainbird P.B., and Shein K. 1996. Prediction of extreme wind speeds at wind energy sites. Harwell, Oxfordshire, OX11 0RA, UK, ETSU Report W/11/00427/REP.

Haslett J. and Raftery A.E. 1989. Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource (with discussion). Applied Statistics 38: 1–50.

Hess-Nielsen N. and Wickerhauser M.V. 1996. Wavelets and time-frequency analysis. Proceedings of the IEEE 84: 523–540.

van der Hoven I. 1957. Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. Journal of Meteorology 14: 160–164.

Krzanowski W.J., Jonathan P., McCarthy W.V., and Thomas M.R. 1995. Discriminant analysis with singular covariance matrices: Methods and application to spectroscopic data. Applied Statistics 44: 101–115.

Mallat S.G. 1989. A theory for multiresolution signal decomposition. IEEE Transactions on Pattern Analysis and Machine Intelligence 11: 674–693.

Nason G.P., Sapatinas T., and Sawczenko A. 1999. Wavelet packet modelling of infant sleep state using heart rate data. Sankhyā, Series B 63: 199–217.

Nason G.P. and Silverman B.W. 1994. The discrete wavelet transform in S. Journal of Computational and Graphical Statistics 3: 163–191.

Nason G.P. and Silverman B.W. 1995. The stationary wavelet transform and some statistical applications. In: A. Antoniadis and G. Oppenheim (Eds.). Lecture Notes in Statistics, vol. 103. Springer-Verlag, New York, pp. 281–300.

Nason G.P., von Sachs R., and Kroisandt G. 2000. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. Journal of the Royal Statistical Society, Ser. B. 62: 271–292.

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

Pesquet J.C., Krim H., and Carfantan H. 1996. Time-invariant orthonormal wavelet representations. IEEE Transactions on Signal Processing 44: 1964–1970.

Priestley M.B. 1981. Spectral Analysis and Time Series. Academic Press, London.

Ramsey J.B. and Lampart C. 1998. The decomposition of economic relationships by time scale using wavelets: Expenditure and income. Studies in Nonlinear Dynamics and Econometrics 3: 23–42.

Vidakovic B. 1999. Statistical Modeling by Wavelets. John Wiley, New York.

Walden A.T. and Contreras Cristan A. 1998. The phase-corrected undecimated discrete wavelet packet transform and its application to the timing of events. Proceedings of the Royal Society, Series A. 454: 2243–2266.

Wickerhauser M.V. 1994. Adapted Wavelet Analysis from Theory to Software. A.K. Peters, Wellesley, MA.