Wavelet Methods for Transfer Function Modelling

Katherine Hunt

September 2002

A thesis submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the School of Mathematics of the Faculty of Science.

Abstract

This thesis considers some methods for relating one stochastic process to another. To do this we use wavelets to efficiently represent (possibly non-stationary) one and two-dimensional processes.

In the first part of the thesis we propose an extension to a wavelet packet transfer function model. The existing model is extended by including a principal components analysis in order to reduce the problems caused by high correlations between explanatory variables. The new model is applied to the modelling and prediction of wind speeds and is found to compare favourably to existing methods.

In the second part of this thesis we suggest an extension to the multiple regression Bayesian dynamic linear model. We propose including the wavelet packet coefficients of a time series as explanatory variables in the model, in place of the original time series. We apply our new method to simulated and real data sets; the inclusion of wavelet packet coefficients in the model is found to provide more reliable predictions if the data set contains certain non-stationary characteristics.

We conclude by proposing a wavelet transfer function model for two-dimensional data. We apply our new model to images obtained from a LANDSAT-type sensor and preliminary results suggest our method warrants further investigation.

Acknowledgements

The support of the University of Bristol (in particular the enthusiasm and advice from Guy Nason), EPSRC and DERA is gratefully acknowledged. I owe special thanks to my Mum for proofreading this thesis and my Dad and Rachel Garcia for their unfaltering belief in my abilities.

Thanks are due to the British Atmospheric Data Centre (BADC) for providing access to the Met Office Land Surface Observations Data.

Author's Declaration

I declare that the work in this thesis was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text and no part of the thesis has been submitted for any other degree. Any views expressed in the thesis are those of the author and in no way represent those of the University of Bristol. The thesis has not been presented to any other University for examination either in the United Kingdom or overseas.

Katherine Hunt

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis combines methodological and computational work in two areas of study: time series analysis and wavelets. Our main theme is to consider methods which relate two, possibly non-stationary, processes. Principally we use a time ordered non-decimated wavelet packet transform (NWPT) to provide a flexible representation of an explanatory time series in terms of scale, frequency and location. The resulting wavelet packet coefficients are used as variables in transfer function models to provide predictions of a response time series. The models provide valuable information about which features in the exploratory time series may contribute to the behaviour of the response time series; hence using wavelet packet coefficients in time series analysis may improve our understanding of the system.

The basic concepts behind relevant time series analyses and wavelets are presented in Chapter 2. Our review of time series analysis includes ARIMA schemes, transfer function models and Bayesian dynamic linear models. We introduce the topic of wavelets from a multiresolution approach, explain some wavelet transforms and review several roles that wavelets have played in time series analysis to date. In Section 2.9 we explain the properties of the non-decimated wavelet transforms which make non-decimated wavelet coefficients particularly suited to the role of explanatory variable. We conclude with a review of the current methods for wind speed modelling and prediction; a potential application of the methodology proposed in Chapter 4.

Although statistics has been applied to marketing data for many years, in the last twenty years the amount of data available to the researcher has been greatly multiplied due to, in part, the development of scanner technology, computer aided surveys and the internet. We analyse some short, possibly non-stationary marketing time series to determine how the price of a product drives its market share; investigating the data set through a combination of techniques including ARIMA models and the Nason and Sapatinas (2002) model. We demonstrate how using certain non-stationary models may improve our understanding of these series. Chapter 3 is the result of collaborative research work with a multinational consumer goods company.

Recent methodology developed by Nason and Sapatinas (2002) proposed modelling a response time series in terms of a NWPT of an explanatory time series, using generalised linear models. The component time series may exhibit non-stationary behaviour and the NWPT representation allows a physical interpretation to be attached to the model. In Chapter 4 we design an extension to the Nason and Sapatinas (2002) model to allow for the occasional high correlations that arise between wavelet packet coefficients; we suggest the inclusion of a principal components analysis in the model. A case study on wind speed prediction compares our new method with the Nason and Sapatinas (2002) model and a popular wind industry technique (linear regression).

A transfer function model which explicitly models non-stationary features in a data set is the multiple regression dynamic linear model (RDLM); see Harrison and West (1997). In Chapter 5, we propose modelling a response time series *not* in terms of an explanatory time series but using the non-decimated wavelet packet coefficients of the explanatory series. The time-varying regression parameters in the RDLM will have a direct interpretation in terms of the time-scale decomposition of the explanatory series. This might be beneficial for time series analysis if the explanation is more parsimonious and useful for prediction. A potential application for our new method is provided where the time series exhibit highly non-stationary behaviour.

In Chapter 6 we propose a prototype for a two-dimensional version of the wavelet packet transfer function methodology of Nason and Sapatinas (2002). Our aim is to model a response image in terms of a two-dimensional non-decimated wavelet transform (2D NWT) of an explanatory image. Using multispectral satellite imagery of Chew Valley Lake in Somerset (UK), we consider if you can predict the values of one band from a multiscale representation of another band. If you can predict one band well from only a few scales in another band you have also indicated which scales are important.

We conclude with Chapter 7 where we draw together the key results of this thesis and suggest areas for future research. Additional information pertaining to the chapters can be found in the appendices.

Chapter 2

Literature Review

2.1 Introduction

This literature review introduces the basic concepts of wavelets and certain time series analysis methods. Wavelets have been used in several areas, including signal processing, geophysics and mathematics. The interest in the utility of wavelets for time series analysis has grown in recent years and wavelets are now used in areas such as local spectral estimation (Nason *et al.*, 2000) and dynamic linear models (Zandonade & Morettin, 2001), for example.

The first half of this chapter provides an introduction to the time series methods we use. We concentrate on the aims and methods of jointly modelling two or more time series, paying particular attention to the ARIMA model. We next consider the state-space time series model known as the Bayesian dynamic linear model (DLM). The basic difference between state-space models and ARIMA models is that in a DLM the response time series y_t is *not* usually expressed by an equation including the entire past, rather y_t is determined through observation and system equations. Time series analysis is discussed extensively in, for example, Priestley (1981), Chatfield (1989), Harvey (1989) and Kendall and Ord (1990). A brief history of time series analysis is given in Tsay (2000) and a review of forecasting methods in Chatfield (1997). See, for example, Harvey (1989) or Durbin and Koopman (2001) for a statistical introduction to state-space models.

The second half of this chapter summarises some fundamental wavelet concepts. We provide motivation for wavelet analyses by considering the limitations of Fourier analyses and introduce wavelets from a multiresolution perspective. We explain various wavelet transforms, including the discrete wavelet transform (DWT) and discrete wavelet packet transform (WPT). We include discussion of *non-decimated* transforms because we wish to use such wavelets in time series analyses. In addition, we consider wavelets in two dimensions for image analysis.

In Section 2.8 we review some recent applications of wavelets in time series analysis. We discuss in detail the wavelet packet transfer function methodology of Nason and Sapatinas (2002), the model which we extend in Chapter 4. We consider the role of wavelets packets as explanatory variables (Section 2.9); in particular, providing an interpretation of our new wavelet packet combination plot. Further details on wavelets can be found in, for example, Chui (1992), Daubechies (1992), Wickerhauser (1994) and Burrus *et al.* (1998).

We conclude this chapter with discussion of the methods for modelling and predicting wind speeds. In particular, we explain the technique of Measure-Correlate-Predict (MCP), which is provided as background reading to the case study in Chapter 4.

2.2 Time series analysis

2.2.1 Introduction

A time series is a collection of observations made sequentially in time. Our first objective of time series analysis is to describe the time series. We may wish to identify trend or seasonal components in the series and check for unusual observations. Our next objective is usually to model the residual time series, i.e. the series excluding trend, seasonal and any other cyclic changes, using a statistical model. Our final objective may involve using the statistical model to provide predictions.

2.2.2 ARIMA models

We consider the stationary time series models: autoregressive (AR), moving average (MA) and autoregressive-moving average (ARMA) models.

Define

$$\begin{aligned} X_t &= Z_t + \varepsilon_1 Z_{t-1} + \ldots + \varepsilon_q Z_{t-q} && \text{as an MA(q) process and} \\ X_t &= \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t && \text{as an AR(p) process,} \end{aligned}$$

where, X_t is an observed time series, $\{Z_t\}$ is a purely random process[1] with mean zero and variance σ_z^2 and $\{\varepsilon_i\}_{i=0}^{q}$ and $\{\alpha_j\}_{j=1}^{p}$ are constants. An ARMA(p, q) scheme is defined as

$$X_t - \alpha_1 X_{t-1} - \ldots - \alpha_p X_{t-p} = Z_t - \varepsilon_1 Z_{t-1} + \ldots + \varepsilon_q Z_{t-q}.$$

If X_t is replaced by $\bigtriangledown^d X_t$ then we term the scheme an autoregressive integrated moving average (ARIMA) process, where X_t is differenced d times ($\bigtriangledown^d$) to obtain stationarity in the series. The model describing the dth differences of X_t is denoted ARIMA(p, d, q).

A popular approach to time series modelling and forecasting based on ARIMA models was developed by Box and Jenkins (1970). The first step to fitting a model is to obtain stationarity in the series. This may involve applying a linear filter to the trend or differencing the series. An appropriate model may then be fit to the stationary series using an iterative process of model identification, estimation and diagnostic checking:

Model identification The autocorrelation function (acf) and partial autocorrelation function (pacf) of the residual time series are calculated and used to identify an appropriate p and q. We can interpret the acf to identify an MA(q) process if it cuts off at lag q. The acf identifies an AR(p) or ARMA(p, q) process if it is a mixture of damped exponentials and sinusoids which die out slowly. We can identify the process as an AR(p) process if the pacf cuts off at lag p or as a MA(q) process if it attenuates (p. 52, Chatfield (1989)).

If the initial values for p and q (i.e. those suggested by the acf and pacf) provide an adequate model we try smaller values of p and q to check if a simpler model is also sufficient. If the initial values of p and q do not provide a suitable model

[1] A discrete-time process is called a purely random process if it consists of a sequence of random variables $\{Z_t\}$ which are mutually independent and identically distributed. From the definition it follows that the process has constant mean and variance and that

$$\begin{aligned} \gamma(k) &= \mathrm{Cov}(Z_t, Z_{t+k}) \\ &= 0 \text{ for } k = \pm 1, 2, \end{aligned}$$

(p. 31, Chatfield (1989))

we try fitting more complex models. By experimenting with different values of p and q we hope to identify the most appropriate, and simplest, ARIMA model.

Model estimation For our chosen values of p and q we estimate the values of the parameters $\{\varepsilon_i\}_{i=0}^{q}$ and $\{\alpha_j\}_{j=1}^{p}$. The parameters of an AR process can be calculated easily using least squares estimates (when $\{Z_t\}$ are normal these are maximum likelihood estimates). Estimating the parameters of an MA process is more complex and some form of numerical iteration must be performed. Consequently estimating the parameters of an ARIMA process may also be difficult; see Chatfield (p. 53-59, 1989) for details. We will use S-Plus to estimate the parameters, see Ansley (1979) for details of the estimation procedure.

Diagnostic checking Check the fit of the model by, for example, checking all model parameters are significant, plotting the residuals against the fitted values, looking for structure in the acf and pacf of the residuals, identifying large standard errors and computing Portmanteau statistics[2].

If the model fit is inadequate, this 3-step process is repeated, deleting or adding parameters as necessary. Several well-fitting models can be compared using Akaike's information criteria. When fitting ARIMA models difficulties may occur as the residual series can be complex and difficult to identify in practice. Refer to Jenkins (1979) for further details and examples.

2.2.3 Transfer function models

We are primarily interested in relating one time series to another, particularly where the two time series are 'causally related'. We consider one time series as an input to a linear system (X_t) and the other time series as the output of the system (Y_t). We may wish to relate Y_t and X_t in order to:

- Understand the dynamic relationships between them. One series may lead to another or feedback relationships may exist.

[2]The Portmanteau statistic is a diagnostic based on the autocorrelations, γ, taken as a whole. If the correct ARIMA model was fitted, and these data are Gaussian, then $Q = n\sum_{k=1}^{K}\hat{\gamma}_k^2$ is approximately distributed as a χ^2 random variable on $K-r$ degrees of freedom, where n is the number of observations used to compute the likelihood, K is a fixed number of lags and r is the number of parameters fit to the model (S-Plus 2000 Guide to Statistics, 1999)

- Improve the predictions of a short time series, using information from a second, longer period of historical data

Methods which relate Y_t and X_t have been considered extensively in the literature. We outline some of them below.

Modelling stationary time series

Box and Jenkins (1970) consider the case of fitting a linear relationship with noise to a finite number of observations, i.e.

$$\begin{aligned} a_0Y_t + a_1Y_{t-1} + \ldots + a_kY_{t-k} &= b_0 + b_1X_{t-1} + \ldots + b_tX_{t-l} + N_t, \\ (a_0 + a_1B + \ldots + a_kB^k)Y_t &= (b_0 + b_1B + \ldots b_tB^l)X_t + N_t, \\ \alpha(B)Y_t &= \beta(B)X_t + N_t, \\ Y_t &= \alpha^{-1}(B)\beta(B)X_t + \alpha^{-1}(B)N_t, \end{aligned}$$

where, $\alpha(B)$ is a polynomial of order k in B, $\beta(B)$ is a polynomial of order l in B and the backward shift operator B is defined by $B^lX_t = X_{t-l}$ for all l. This relates an output variable to a related input series, assuming the inputs affect the output but not vice versa. As the inputs do not fully explain the output an error component N_t is included. Box and Jenkins refer to a model in this form as a *transfer function* model. See Haugh and Box (1977), Wallis (1977), Chan and Wallis (1978) and Granger and Newbold (1986) for extensions to the Box-Jenkins procedure for modelling multiple time series.

An example of the problems of mis-specifying a time series model is given (unintentionally) by Coen, Gomme and Kendal (1969). They investigated the possibility of predicting the Financial Times ordinary share index, $x_{1,t}$, from lagged values of UK car production, $x_{2,t}$, and Financial Times commodity price index, $x_{3,t}$. They built their model by cross-correlating the detrended series and choose to use a standard regression analysis with $x_{2,(t-6)}$ and $x_{3,(t-7)}$ as the regressors. The main problem with their approach was the assumption of uncorrelated errors in the model which led to falsely significant results; see Box and Newbold (1971) for a critical evaluation. Before cross-correlating time series it is recommended that the input time series are *prewhitened*.

The cross correlation function between two autocorrelated time series will usually contain spurious values and hence can provide a misleading picture of the relationship between two time series. In other words, what the cross correlation function is trying to tell us about the relationship between an output Y_t and an input X_t can be obscured by the autocorrelation function of either series. To avoid obtaining spurious cross correlations, it is necessary to use the univariate stochastic model for the input series to convert it into random series a_t (autocorrelations are zero at all lags), that is to prewhiten the input. The same input model applied to the output series will convert it to a series β_t which, in general, will not be random.

Definition 2.1

The prewhitened cross correlation function is the cross correlation function between the prewhitened input a_{t-k} and the correspondingly filtered output β_t at lags $k = 0, 1, 2 \ldots$

The prewhitened cross correlation function can be used to identify a relationship between Y_t and X_t (p. 139, Jenkins (1979)).

Priestley (1981) considers a linear time-invariant relationship, where the current value of the output $Y(t)$ is a linear combination of the present and past values of the input $X(t)$,

$$Y(t) = \int_0^\infty g(s)X(t-s)\mathrm{d}s,$$

where $g(s)$ is some function which does not vary with time but is determined by the particular system under study. In particular, if $X(t) = A\exp(\mathrm{i}\omega t)$, a sinusoidal wave of frequency ω, then

$$\begin{aligned} Y(t) &= \int_0^\infty g(s)A\exp(\mathrm{i}\omega(t-s))\mathrm{d}s, \\ &= A.G(\omega).\exp(\mathrm{i}\omega t). \end{aligned}$$

i.e. if the input is sinusoidal then the output is also sinusoidal with exactly the same frequency but with modified amplitude A (and phase). The transfer function

$$G(\omega) = \int_0^\infty g(s)\exp(-\mathrm{i}\omega s)\mathrm{d}s$$

tells us how, for each ω, the amplitude of the input wave is transformed into the amplitude of the output wave (p. 22, Priestley (1981)).

A disadvantage of the above methods comes from the initial assumption that the involved time series are stationary (at least to order 2) or can be reduced to stationarity by some simple transformation.

Commonly, stationary signals (or rather their autocovariances) are decomposed into the Fourier basis using the Fourier transform (FT). The FT gives frequency information about the signal, but does not provide information about when the individual frequencies occur. This information is, of course, not required when the signal is stationary. In practice, for many real signals, their frequency components change over time and so the FT is often ill suited in practice.

A classical solution to this problem is to use the short time Fourier transform (STFT). In the STFT, the signal is divided into sections exhibiting stationary behaviour. A 'window' function is chosen, where the width of the window is equal to the segment of the signal where it is stationary. The STFT of the signal is a Fourier transform of the signal multiplied by the window function. For an application refer to, for example, Portnoff (1976).

The fixed time-frequency resolution of the basis functions in the STFT can be a problem. Choose a wide window and gain good frequency resolution at the expense of poor time resolution, or choose a narrow window and gain good time resolution at the expense of poor frequency resolution. Heisenberg's Uncertainty Principle states in that in modelling time-frequency phenomena one cannot be arbitrarily precise in both time and frequency simultaneously (p. 35, Vidakovic (1999)).

Modelling non-stationary time series

There has been much research on developing techniques for analysing non-stationary time series. It is important to recognise the types of non-stationarity which may occur in the data set; features in the series may be slowly varying or prone to sudden changes. We must decide whether to model the non-stationary features or to remove them; although not all features may be removed by simple operations.

Priestley (1965) developed an approach to the spectral analysis of non-stationary processes based on the theory of the *evolutionary spectra*, designed to deal with slowly-varying processes. The evolutionary spectrum is time dependent and describes the local power-frequency distribution in the neighbourhood of time t, preserving the immediate physical interpretation of the spectrum. A full account is given in Priestley (1981).

Data which exhibit short-term transient effects and/or long-term changes may be modelled by some form of intervention analysis, Bayesian method or structural model. These methods seek to avoid the problems of adjusting series prior to analysis when the series is non-stationary. We now review the Bayesian dynamic linear model in preparation for the work in Chapter 5.

2.3 Bayesian forecasting and dynamic models

In the Bayesian dynamic linear model (DLM), the response series Y_t is determined in an analytical way through *observation* and *system* equations. At each time point t, the observation and system equations are updated using a combination of prior knowledge and the observed value of Y_t and, if present, explanatory variables.

The theory of the general univariate DLM is given in Section 2.4, where we consider the general form of the observation and system equations and the concepts of online and retrospective forecasting. Section 2.5 contains the theory of the multiple regression dynamic linear regression (RDLM) with unknown variance; the model we consider in detail in Chapter 5.

For introductory texts on dynamic linear modelling refer to, for example, Pole *et al.* (1994), Chatfield (1989), Meinhold and Singpurwalla (1983) and Cooper and Harrison (1997). Detailed expositions can be found in Harrison and Stevens (1976) and Harrison and West (1997).

2.4 The general univariate DLM

2.4.1 Model form and notation

For each t, the general univariate normal dynamic linear model, represented by the quadruple $\{F_t, G_t, V_t, W_t\}$, is defined by (p. 29, Pole *et al.* (1994))

$$\begin{array}{lll} \text{Observation equation:} & Y_t = F_t'\theta_t + \nu_t, & \nu_t \sim \mathrm{N}[0, V_t], \\ \text{System equation:} & \theta_t = G_t\theta_{t-1} + \omega_t, & \omega_t \sim \mathrm{N}[0, W_t], \\ \text{Initial information:} & (\theta_0|D_0) \sim \mathrm{N}[m_0, C_0], & \end{array}$$

where, Y_t denotes the response time series at time t, F_t is a known vector of regression variables, θ_t is a vector of regression coefficients, G_t is a known matrix that defines the systematic evolution of the regression coefficients across time t and D_t is the information set at time t: $D_t = \{D_0, Y_1, \dots, Y_t\}$ or sequentially $D_t = \{D_{t-1}, Y_t\}$. Given D_{t-1}, the error sequences $\{\nu_t\}$ and $\{\omega_t\}$ are individually and mutually independent and are independent of $(\theta_0|D_0)$, where the initial information set D_0 is determined entirely by our prior specifications.

We can interpret the DLM by examining how θ_t varies over time. For instance, does it converge to a stable relationship helping us provide reliable short-term forecasts? Series containing missing data are easily handled in the DLM framework as, in the absence of new information, the posterior beliefs are simply equal to the prior beliefs.

Note that models of stationary processes (see Box and Jenkins, 1970) are forms of a special class of DLM known as the constant DLM, where $G_t = G$, $F_t = F$, $V_t = V$ and $W_t = W$ (p. 143, West and Harrison (1997)).

2.4.2 Bayesian analysis

Bayesian learning proceeds by combining information from observations, expressed through the likelihood function with the investigators existing state of knowledge before the observations are made (p. 30, Pole *et al.* (1994)).

The prior distribution of θ_t at time t (given D_{t-1}) is calculated by

$$\begin{aligned} p(\theta_t|D_{t-1}) &= \int p(\theta_t, \theta_{t-1}|D_{t-1})\mathrm{d}\theta_{t-1}, \\ &= \int p(\theta_t|\theta_{t-1}, D_{t-1})p(\theta_{t-1}|D_{t-1})\mathrm{d}\theta_{t-1}. \end{aligned}$$

The one-step forecast distribution at time t is calculated by

$$p(Y_t|D_{t-1}) = \int p(Y_t|\theta_t, D_{t-1})p(\theta_t|D_{t-1})\mathrm{d}\theta_t.$$

We may then calculate the posterior for θ_t at time t by Bayes theorem

$$\begin{aligned} p(\theta_t|D_t) &\propto p(\theta_t|D_{t-1})p(Y_t|\theta_t, D_{t-1}), \\ \text{posterior} &\propto \text{prior} \times \text{likelihood}. \end{aligned}$$

2.4.3 Updating and forecasting equations

One-step forecast and posterior distributions for the general univariate DLM are given, for each t, as follows (p. 103, Harrison and West (1997))

Posterior for θ_{t-1} at $t-1$ $(\theta_{t-1}|D_{t-1}) \sim \mathrm{N}[m_{t-1}, C_{t-1}]$,

for some mean m_{t-1} and variance matrix C_{t-1}.

Prior for θ_t at t $(\theta_t|D_{t-1}) \sim \mathrm{N}[a_t, R_t]$,

where $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t^{'} + W_t$.

One-step forecast at t $(Y_t|D_{t-1}) \sim \mathrm{N}[f_t, Q_t]$,

where $f_t = F_t^{'} a_t$ and $Q_t = F_t^{'} R_t F_t + V_t$.

Posterior for θ_t at t $(\theta_t|D_t) \sim \mathrm{N}[m_t, C_t]$,

with $m_t = a_t + A_t e_t$ and $C_t = R_t - A_t Q_t A_t^{'}$,

where $A_t = R_t F_t Q_t^{-1}$ and $e_t = Y_t - f_t$.

In Chapter 5, we provide one-step ahead forecasts of the market share for a consumer good.

2.4.4 Retrospective parametric distribution

Retrospective analysis may be used to examine what happened at times of major change. We revise estimates in view of later information in order to (hopefully) attribute effects to the contributing sources. We follow the discussion from Pole *et al.* (p. 40, 1994) to outline the retrospective estimate for yesterday given the information we have today, i.e. $p(\theta_{t-1}|D_t)$.

The filtered distribution, in terms of the joint distribution of the state for yesterday and today, is given by

$$p(\theta_{t-1}|D_t) = \int p(\theta_{t-1}|\theta_t, D_t) p(\theta_t|D_t) \mathrm{d}\theta_t.$$

Currently $p(\theta_{t-1}|\theta_t, D_t)$ is unknown but may be expressed as

$$\begin{aligned} p(\theta_{t-1}|\theta_t, D_t) &= p(\theta_{t-1}|\theta_t, y_t, D_{t-1}), \quad (\text{as } D_t = \{y_t, D_{t-1}\}) \\ &= \frac{p(y_t|\theta_{t-1}, \theta_t, D_{t-1}) p(\theta_{t-1}|\theta_t, D_{t-1})}{p(y_t|\theta_t, D_{t-1})}. \end{aligned}$$

As the first term of the numerator is equal to the denominator (y_t is conditionally independent of earlier states), the conditional filtered density for the state yesterday is thereby reduced to

$$\begin{aligned} p(\theta_{t-1}|\theta_t, D_t) &= p(\theta_{t-1}|\theta_t, D_{t-1}), \\ &= \frac{p(\theta_t|\theta_{t-1}, D_{t-1})p(\theta_{t-1}|D_{t-1})}{p(\theta_t|D_{t-1})}. \end{aligned} \tag{2.1}$$

Each distributional component in Equation 2.1 is known. We can now express $p(\theta_{t-1}|D_t)$ in terms of known components

$$\begin{aligned} p(\theta_{t-1}|D_t) &= \int p(\theta_{t-1}|\theta_t, D_t)p(\theta_t|D_t)\mathrm{d}\theta_t, \\ &= \int p(\theta_{t-1}|\theta_t, D_{t-1})p(\theta_t|D_t)\mathrm{d}\theta_t. \end{aligned} \tag{2.2}$$

As both components of Equation 2.2 are normal, standard multivariate normal theory provides the result of the intergration:

$$\theta_{t-1}|D_t \sim \mathrm{N}[a_t(-1), R_t(-1)],$$

where the defining moments are given by

$$\begin{aligned} a_t(-1) &= m_{t-1} + B_{t-1}(m_t - a_t), \\ R_t(-1) &= C_{t-1} - B_{t-1}(R_t - C_t)B_{t-1}', \end{aligned}$$

where, $B_{t-1} = C_{t-1}G_t'R_t^{-1}$.

A generalisation of this result to the marginal distribution of θ_{t-k} given D_t is provided in the next section; the results enable a full retrospective analysis.

Filtering recurrences

In the univariate DLM $\{F_t, G_t, V_t, W_t\}$, for all t, define

$$B_t = C_t G_{t+1}' R_{t+1}^{-1}.$$

Then for all k, $(1 \leq k < t)$, the filtered marginal distributions are (p. 113, Harrison and West (1997))

$$(\theta_{t-k}|D_t) \sim \mathrm{N}[a_t(-k), R_t(-k)],$$

where

$$\begin{aligned} a_t(-k) &= m_{t-k} + B_{t-k}[a_t(-k+1) - a_{t-k+1}], \\ R_t(-k) &= C_{t-k} + B_{t-k}[R_t(-k+1) - R_{t-k+1}]B'_{t-k}, \end{aligned}$$

with starting values

$$a_t(0) = m_t \qquad \text{and} \qquad R_t(0) = C_t,$$

and where as usual,

$$a_{t-k}(1) = a_{t-k+1} \qquad \text{and} \qquad R_{t-k}(1) = R_{t-k+1}.$$

2.5 The multiple regression DLM

2.5.1 Introduction

We consider the multiple regression DLM in Chapter 5. Suppose that the observed output variable Y_t is known to be linearly related to n known explanatory variables $X_{t,i}$ by

$$Y_t = a_t + \sum_{i=1}^{n} b_{t,i} X_{t,i} + \nu_t, \tag{2.3}$$

where the regression coefficients a_t and $b_{t,i}$ are allowed to evolve through time according to a random walk (so that $G_t = 1$), say, with evolution variance matrices W_t and let $\nu_t \sim N[0, V_t]$. If the $(n+1)$-vectors $\theta_t^T = [a_t, b_{t,1}, b_{t,2} \ldots, b_{t,n}]$ and $F_t^T = [1, X_{t,1}, X_{t,2}, \ldots, X_{t,n}]$, then we may characterise model (2.3) by the quadruple $\{F_t, 1, V_t, W_t\}$ (p. 270, Harrison and West (1997))

$$\begin{aligned} Y_t &= F_t'\theta_t + \nu_t, \quad \nu_t \sim \mathrm{N}[0, V_t], \\ \theta_t &= \theta_{t-1} + \omega_t, \quad \omega_t \sim \mathrm{N}[0, W_t]. \end{aligned} \tag{2.4}$$

2.5.2 Model form and notation

In practice, the variances V_t and W_t are often unknown. Information in the DLM is usually measured in terms of precision, ϕ, the inverse of the observational variance V. Assuming a constant, unknown variance, $V = 1/\phi$, W_t is scaled by the current estimate of V, i.e. S_{t-1}, at each time. For each t, the model, now based on the Normal and Student T distributions, has the form (p. 111, Harrison and West (1997)):

$$\begin{array}{lll} \text{Observation equation:} & Y_t = F_t'\theta_t + \nu_t, & \nu_t \sim \mathrm{N}[0, V = 1/\phi], \\ \text{System equation:} & \theta_t = \theta_{t-1} + \omega_t, & \omega_t \sim \mathrm{T}_{n_t}[0, W_t^* = S_{t-1}W_t], \end{array}$$

where T_n is Student's t on n degrees of freedom.

2.5.3 Updating and forecasting equations

The updating and forecasting equations are based on the following distributions (p. 111, Harrison and West (1997))

Posterior for θ_{t-1} at $t-1$ $(\theta_{t-1}|D_{t-1}) \sim \mathrm{T}_{n_{t-1}}[m_{t-1}, C_{t-1}]$,

for some mean m_{t-1} and variance matrix C_{t-1}.

Prior for θ_t at t $(\theta_t|D_{t-1}) \sim \mathrm{T}_{n_{t-1}}[a_t, R_t]$,

where, $a_t = m_{t-1}$, and $R_t = C_{t-1} + W_t$.

Prior precision $(\phi|D_{t-1}) \sim \mathrm{G}[n_{t-1}/2, d_{t-1}/2]$,

with $S_{t-1} = d_{t-1}/n_{t-1}$, $n_t = n_{t-1} + 1$, $d_t = d_{t-1} + S_{t-1}e_t^2/Q_t$.

Posterior precision $(\phi|D_t) \sim \mathrm{G}[n_t/2, d_t/2]$.

One-step forecast at t $(Y_t|D_{t-1}) \sim \mathrm{T}_{n_{t-1}}[f_t, Q_t]$,

where, $f_t = F_t'm_{t-1}$ and $Q_t = F_t'R_{t-1}F_t + S_{t-1}$,

with $S_{t-1} = d_{t-1}/n_{t-1}$.

Posterior for θ_t at t $(\theta_t|D_t) \sim T_{n_t}[m_t, C_t]$,

with $m_t = m_{t-1} + A_te_t$ and $C_t = [R_t - A_tA_t'Q_t](S_t/S_{t-1})$,

where $A_t = R_tF_t/Q_t$ and $e_t = Y_t - f_t$.

2.5.4 Retrospective parametric distribution

For all k, $(1 \leq k \leq t)$, the retrospective distributions are given by (p. 115, West and Harrison (1997))

$$(\theta_{t-k}|D_t) \sim T_{n_t}[a_t(-k), (S_t/S_{t-k})R_t(-k)],$$

$$(\phi_{t-k}|D_t) \sim G[n_t(-k)/2, d_t(-k)/2],$$

where

$$\begin{aligned} a_t(-k) &= m_{t-k} - B_{t-k}[a_{t-k+1} - a_t(-k+1)], \\ R_t(-k) &= C_{t-k} - B_{t-k}[R_{t-k+1} - R_t(-k+1)]B'_{t-k}, \end{aligned}$$

and $B_t = C_t G'_{t+1} R^{-1}_{t+1}$.

2.5.5 Specification of the evolution matrix W_t

When constructing a DLM we specify the prior information set D_0. We assign an initial mean and variance to components in the model, such as the trend, regression coefficients and the error sequence $\{\nu_t\}$. To specify an initial variance for the error sequence $\{\omega_t\}$ we use the unifying discount approach. The essential idea is to associate a discount factor, $0 < \delta_i \leq 1$, with the ith model component which expresses the local durability or rate of change of the qualified component. Discount factors are typically between 0.8 and 1 (p. 117, Harrison and West (1987)).

Given $\text{Var}[\theta_{i,t-1}|D_{t-1}] = C_{i,t-1}$, and setting $W_{i,t}$ as a fixed proportion of $C_{i,t-1}$, $W_{i,t} = C_{i,t-1}(1-\delta_i)/\delta_i$,

$$\begin{aligned} \text{Var}(\theta_{i,t}|D_{t-1}) &= R_{i,t}, \\ &= C_{i,t-1} + W_{i,t}, \\ &= C_{i,t-1} + C_{i,t-1}(1-\delta_i)/\delta_i, \\ &= C_{i,t-1}/\delta_i, \\ &= \text{Var}(\theta_{i,t-1}|D_{t-1})/\delta_i. \end{aligned}$$

i.e. the prior variance of $\theta_{i,t}$ at time t is equal to the posterior variance of $\theta_{i,t-1}$ at time $t-1$ scaled by a discount factor δ_i.

This method seeks to avoid the problem of specifying the complex evolution variance matrix W_t; see Harrison and West (p. 49-51, 1997) for further details.

2.6 Wavelets in dynamic linear models

A recent literature review highlighted the following research in wavelets (§2.7-2.9) for dynamic linear models. Zandonade and Morettin (2001) model parameters in the system matrix (F_t) of a state-space model using wavelets, but use neither wavelet packets or a non-decimated transform. The method of estimation is not Bayesian, but incorporates maximum likelihood estimation and the EM algorithm. The time series are switching models, i.e. series which exhibit sudden changes of behaviour, where F_t contains the probabilities associated with the m configurations of switching structure.

2.7 An introduction to wavelets

2.7.1 Introduction

Wavelets have gained popularity predominantly in the last twenty years, although their theoretical foundations span the last century. They are used in a wide range of applications including:

- Signal compression and denoising (Donoho and Johnstone (1994), Johnstone and Silverman (1997), Hilton (1997) and Vetterli (2001).)
- Time series analysis (Learned and Willsky (1995), Ramsey (1999), Ramsey and Lampart (1998), Walden and Contreras Cristan (1998), Nason *et al.* (2001) and Nason and Sapatinas (2002)).
- Image and texture analysis (Chang and Kuo (1993), Laine and Fan (1993), Saito and Coifman (1995) and Unser (1995)).

As we will see later, wavelet families can form an orthonormal basis for all functions in $L^2(\mathbb{R})$ (the set of functions on the interval $[-\infty, \infty]$ with finite energy) and we are primarily interested in using wavelets to represent functions. We first introduce the ideas of *efficient* function representation by recalling the limitations of a classical Fourier analysis. A sparse representation of a series, i.e. one which requires only a few non-zero coefficients, may help improve our understanding of the series and hence be useful for signal or image compression.

2.7.2 Fourier analysis

A classical method for representing a signal in terms of basis functions is using Fourier series.

Definition 2.2

A Fourier series is the infinite series

$$\frac{1}{2}a_0 + \sum a_n cos(nx) + b_n sin(nx).$$

If f is such a function the Fourier coefficients are

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)cos(nx)dx, \quad n = 0, 1, 2 \ldots$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)sin(nx)dx, \quad n = 1, 2, 3, \ldots$$

By suitable choice of the constants a_n and b_n the series can be made to converge and the sum of the series can be made equal to any periodic function x on the interval $(-\pi, \pi)$. The Fourier series can be used in the analysis of a waveform into its constituent sine and cosine waves of different frequencies and amplitudes.

Sinusoids are localized in frequency but not time, hence when used as basis functions they rely on cancellation to represent discontinuities in time. This can result in the *Gibbs effect*, i.e. an overshoot by the Fourier series at points of discontinuity. For non-stationary time series there is clearly a need for a transform which will provide a sparse and time localised representation.

For further details on Fourier analyses and applications refer to, for example, Priestley (1981), Ramirez (1985), Folland (1992), Bloomfield (2000) and Kammler (2000).

2.7.3 Multiresolution analysis

Multiresolution analysis (MRA) provides the framework for examining signals at different resolutions. As we will see later, MRA is a key concept in the formulation of the discrete wavelet transform. We define a MRA following Mallat (1989).

Definition 2.3

A multiresolution analysis of $L^2(\mathbb{R})$ is a chain of nested subspaces $\{V_j : j \in \mathbb{Z}\}$ satisfying the following conditions:

MRA 1 $V_j \subset V_{j+1} \subset L^2(\mathbb{R})$.

MRA 2 $\bigcup_{j\in\mathbb{Z}} V_j$ *is dense in* $L^2(\mathbb{R})$.

MRA 3 $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$.

MRA 4 $f(x) \in V_j \iff f(2x) \in V_{j+1}$.

MRA 5 $f(x) \in V_0 \iff f(x-k) \in V_0, \quad k \in \mathbb{Z}$.

MRA 6 *There exists a function* $\phi(x) \in V_0$ *such that the sequence* $\phi(x-k)$ *is an orthonormal basis of* V_0.

Father wavelets

Since $V_0 \subset V_1$, the function $\phi(x) \in V_0$ can be represented as a linear combination of functions from V_1, i.e.

$$\phi(x) = \sum_{k \in \mathbb{Z}} h_k \sqrt{2} \phi(2x - k),$$

for some coefficients h_k, $k \in \mathbb{Z}$. We term $\phi(x)$ the *father wavelet* or *scaling function* and h_k the *low pass filter* (p. 52, Vidakovic (1999)).

A picture of some father wavelets is given in Figure 2.1. Note that these father wavelets have compact support (this is not generally true); these are father wavelets from the Daubechies' least asymmetric and extremal phase families.

Mother wavelets

The detail space W_j is the orthogonal complement of V_j in V_{j+1}, i.e.

$$V_{j+1} = V_j \oplus W_j.$$

It is possible to define a function $\psi(x)$, termed the *mother wavelet*, such that $\{\psi(x-k)\}_{k \in \mathbb{Z}}$ is an orthonormal basis for W_0, orthogonal to all functions in V_j. Since $\psi(x) \in V_1$ (because of the containment $W_0 \in V_1$), the function $\phi(x)$ can be represented as

$$\psi(x) = \sum_{k \in \mathbb{Z}} g_k \sqrt{2} \phi(2x - k),$$

where g_k, $k \in \mathbb{Z}$ is a *high pass filter*. Due to the MRA relations, each W_j is a scaled version of W_0 in that $f(x) \in W_0 \iff f(2^j x) \in W_j$. Consequently $\{\psi_{jk}(x)\}$ is an orthonormal basis for W_j

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

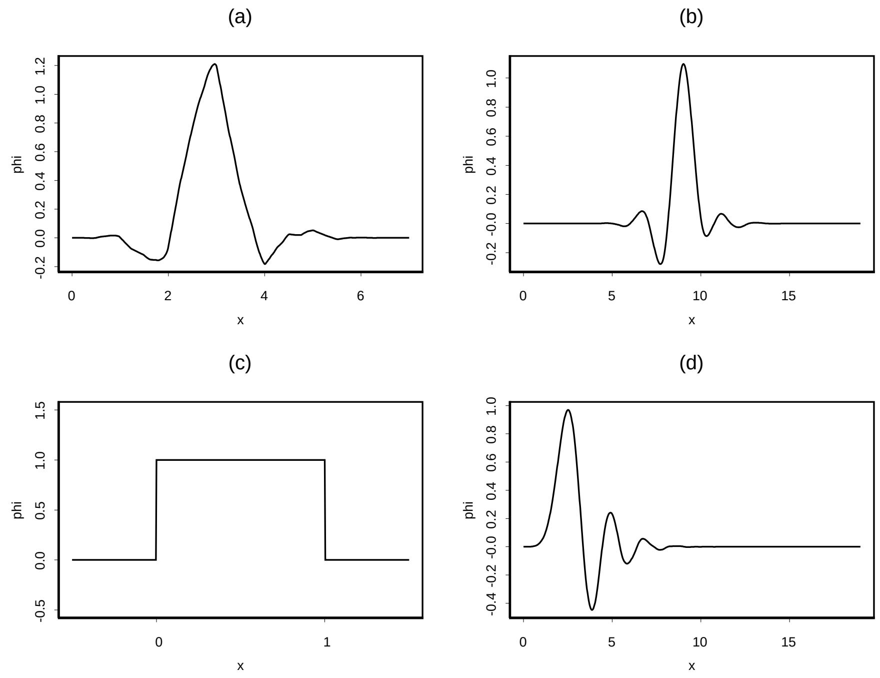

Figure 2.1: Father wavelets from a selection of Daubechies' compactly supported orthonormal wavelet bases.

The j subscript represents the dilation number and k represents the translation number. The scale factor $2^{j/2}$ maintains a constant unit norm independent of scale j, i.e. $||\psi_{jk}|| = 1$ for all $j, k \in \mathbb{Z}$. See Vidakovic (§3.3.1, 1999) on the derivation of the mother wavelet from the scaling function.

A selection of some mother wavelets from the Daubechies' least asymmetric and extremal phase families is given in Figure 2.2. The filter number corresponds to the number of vanishing moments (Definition 2.4). The number of vanishing moments corresponds to the smoothness (or differentiability) of the wavelet; the wavelet becomes more regular as vanishing moments increase.

Definition 2.4

A wavelet of order m is a wavelet which has m vanishing moments, i.e.

$$\int x^l \psi(x)dx = 0, \quad l = 0, 1, \dots, m.$$

This property ensures that fine-scale wavelet coefficients will only be large where a function or its derivatives have singularities; this property promotes the use of wavelets for compression techniques (p. 5, Nason and Silverman (1994)).

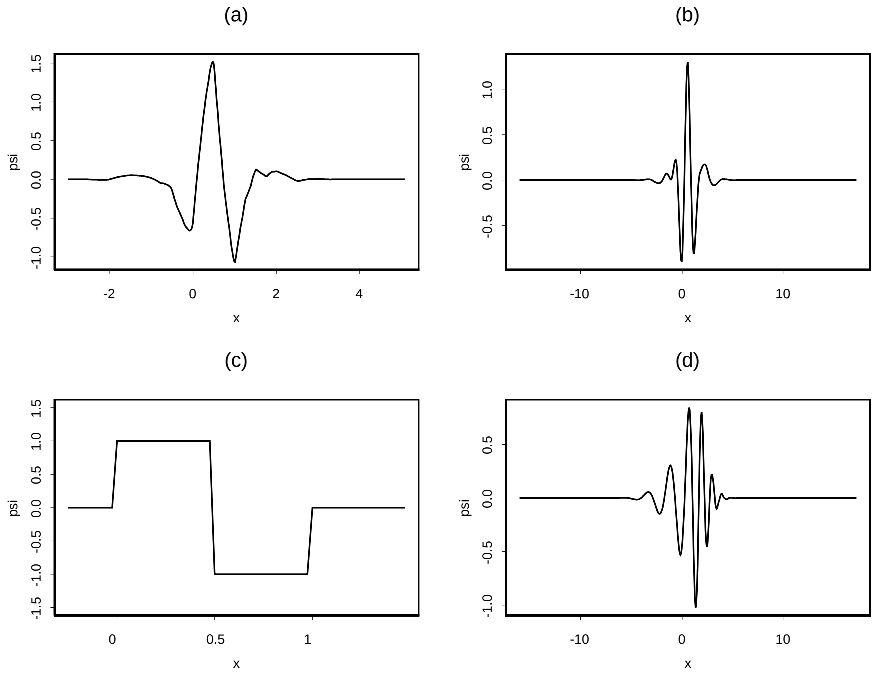

Figure 2.2: The mother wavelets corresponding to the fathers in Figure 2.1: (a) Daubechies' least asymmetric with 4 vanishing moments, (b) Daubechies' least asymmetric with 10 vanishing moments, (c) Haar wavelet (Daubechies' extremal phase with 1 vanishing moment) and (d) Daubechies' extremal phase with 10 vanishing moments.

Conjugate filters

The finite impulse response filters h and g are usually chosen to be intimately related by the relation

$$g_n = (-1)^n h_{1-n},$$

and are known as quadrature mirror filters. Using conjugate filters allows for an orthogonal exact reconstruction of the signal if they satisfy the internal orthogonal relations,

$$\sum_n h_n h_{n+2j} = 0 \qquad \text{and} \qquad \sum_n g_n g_{n+2j} = 0,$$

and the mutual orthogonality relation

$$\sum_n h_n g_{n+2j} = 0$$

(p. 282, Nason and Silverman (1995)).

2.7.4 Representing functions using wavelets

For a suitable mother wavelet, ψ, the set $\{\psi_{jk}\}_{jk}$ provides a basis which can be used to represent functions $f \in L^2(\mathbb{R})$:

$$f(x) = \sum_k c_{0k}\phi_{0k}(x) + \sum_{j=1}^{J}\sum_k d_{jk}\psi_{jk}(x),$$

where, c_{0k} is the father wavelet coefficient at the coarsest scale and d_{jk} are the mother wavelet coefficients (p. 17, Burrus *et al.* (1998)).

The orthogonal transform allows for perfect reconstruction of the time series whilst avoiding redundancy. There are certain properties of wavelets which depend on the orthogonality of the wavelet functions, such as the integral of the mother wavelet being necessarily zero. See Burrus *et al.* (p. 64, 1998) for further details.

Haar wavelets

There are a wide range of wavelet families with different smoothness, symmetry and support properties. We can pick the wavelet to suit the signal, for example, a signal which appears 'blocky' might be best represented by the Haar wavelet. In this thesis we primarily use the Haar wavelet. The father wavelet is given by

$$\phi(x) = \begin{cases} 1, & x \in (0,1), \\ 0, & \text{otherwise.} \end{cases}$$

The mother wavelet is given by

$$\psi(x) = \begin{cases} 1, & x \in [0, \frac{1}{2}), \\ -1, & x \in [\frac{1}{2}, 1), \\ 0, & \text{otherwise.} \end{cases}$$

Refer to Vidakovic (Section 3.4, 1999) for details of other wavelet families.

2.7.5 The discrete wavelet transform

There is an efficient algorithm for calculating the wavelet coefficients of a discrete sequence known as the *discrete wavelet transform* (DWT) or *pyramid algorithm*, proposed by Mallat (1989). The DWT follows the algorithm below:

Given a function f which is observed at $N = 2^J$ equally spaced time points $\{t_i\}_{i=0,\ldots.2^J-1}$, set $c_{J,i} = f(t_i)$ for $i = 0, \ldots, 2^J - 1$. The DWT of the sequence is obtained by recursively using the relations

$$c_{j-1,i} = \sum_n h_{n-2i} c_{j,n} \tag{2.5}$$

and

$$d_{j-1,i} = \sum_n g_{n-2i} c_{j,n} \tag{2.6}$$

to obtain

$$(\tilde{c}_0, \tilde{d}_0, \ldots, \tilde{d}_{j-1}),$$

where, $\tilde{d}_j = (d_{j,0}, d_{j,1}, \ldots, d_{j,2^j-1})$, $j = 0, 1, \ldots, J-1$, and h and g are low and high pass quadrature mirror filters respectively for time series of length N.

We set the father wavelets at the finest scale to be equal to samples of the function f. Most authors just assign the sample values of functions f to the finest level wavelet coefficients. However, these are not necessarily a good approximation to the true inner products $< f, \phi_{J,i} >$. A better approximation can be obtained using coiflets of some order greater than the smoothness of the space in which the function is assumed to

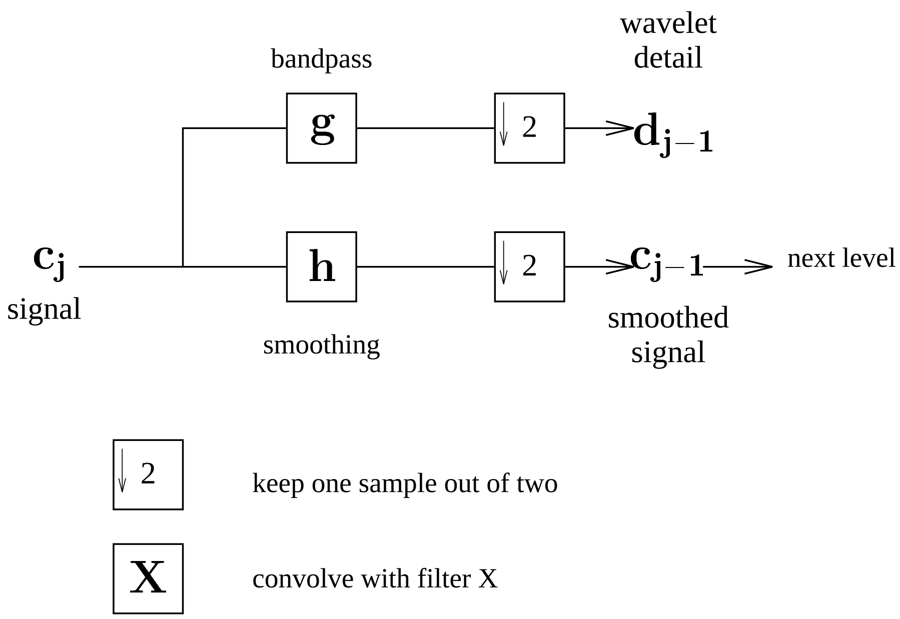

Figure 2.3: DWT decomposition of discrete signal c_j into smoothed signal c_{j-1} and wavelet coefficients d_{j-1} by use of conjugate filters h and g (after Mallat, 1989).

lie, see Antoniadis (1994). The $2i$ operation in (2.5) and (2.6) is equivalent to first filtering by h and g and then applying even dyadic decimation (which selects every even indexed observation from a sequence). Defining $\mathcal{D}_0$ to be the even decimation operator and $\mathcal{D}_1$ to be the odd decimation operator, at each stage dyadic decimation could be applied by either the $\mathcal{D}_0$ or the $\mathcal{D}_1$ operator. Permuting $\mathcal{D}_0$ and $\mathcal{D}_1$ in the sequence of consecutive decimations in a wavelet decomposition leads to the orthogonal ϵ-decimated transform (refer to Nason and Silverman (1995) for details).

The decomposition and reconstruction algorithms for the DWT are illustrated in Figures 2.3 and 2.4 respectively. The computational complexity of the DWT is $O(N)$.

Boundary conditions

A boundary problem occurs when the filter extends beyond the range of the series. This problem does not arise with the Haar wavelet but for other wavelets, for example, Daubechies' least-asymmetric wavelets, this may be a problem. Several approaches exist for resolving the boundary problem. The signal may be continued in a periodic way $(\dots, y_{n-1}, y_n | y_1, y_2, \dots)$, symmetric way $(\dots, y_{n-1}, y_n | y_{n-1}, y_{n-2}, \dots)$, padded by a constant or extrapolated as a polynomial (p. 112, Vidakovic (1999)).

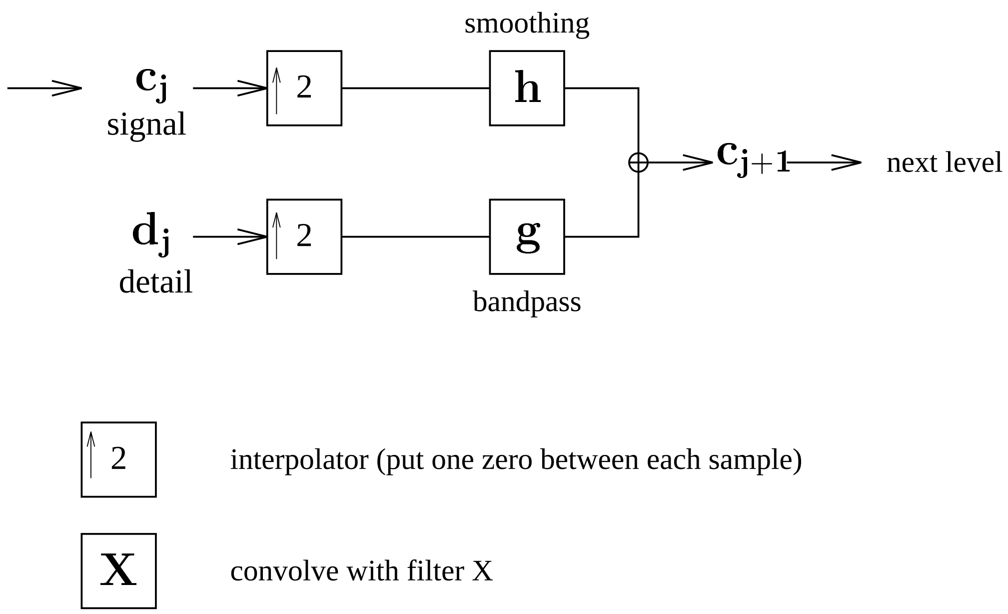

Figure 2.4: DWT reconstruction step. Producing higher resolution signal c_{j+1} from smoothed signal c_j and d_j by use of conjugate filters h and g (after Mallat, 1989).

2.7.6 The non-decimated wavelet transform

The non-decimated wavelet transform (NWT) was developed to address certain deficiencies of the DWT in areas such as time series analysis and image analysis.

A consequence of the decimation step in the DWT is that the transform is *not* translation-equivariant, i.e. a shift in the time series is not characterised by an identical shift in the DWT coefficients. For a non-decimated transform, if a certain behaviour occurs in X_t which is characterised by the NWT coefficients at time t then if the behaviour is repeated again at time $X_{t+\tau}$ the NWT coefficients at time $t+\tau$ will be the same as those at time t (Nason and Sapatinas, 2002).

In addition, the number of coefficients in each wavelet scale is the same as the number of data points in the original data. This means the NWT coefficients of a time series can be represented as a $(2N-2)$-dimensional multivariate time series, where each variable corresponds to a particular wavelet basis function (Nason and Sapatinas, 2002).

We follow Nason and Silverman (1995) to explain the NWT:

*D*enote $\mathcal{H}$ and $\mathcal{G}$ as the low and high pass filters defined by $\{h_n\}$ and $\{g_n\}$ respectively, i.e.

$$\begin{aligned} \mathcal{H}s &= \textstyle\sum_n h_{n-k}s_n, \\ \text{whilst} \quad \mathcal{G}s &= \textstyle\sum_n g_{n-k}s_n, \end{aligned}$$

where s is a doubly infinite sequence.

Let $\mathcal{Z}$ denote the operator which pads out x, a doubly infinite sequence, with zeros as follows

$$(\mathcal{Z}x)_{2j} = x_j \quad \text{and} \quad (\mathcal{Z}x)_{2j+1} = 0.$$

Define the filters $\mathcal{H}^{[r]}$ and $\mathcal{G}^{[r]}$ to have weights $\mathcal{Z}^r\mathbf{h}$ and $\mathcal{Z}^r\mathbf{g}$ respectively. Thus the filter $\mathcal{H}^{[r]}$ has weights $h^{[r]}_{2^rj} = h_j$ and $h^{[r]}_k = 0$ if k is not a multiple of 2^r. The filter $\mathcal{H}^{[r]}$ is obtained by inserting a zero between every adjacent pair of elements of the filter $\mathcal{H}^{[r-1]}$, and similarly for $\mathcal{G}^{[r]}$.

To define the stationary wavelet transform, set c^J to be the original sequence. For $j = J, J-1, \ldots, 1$ we then recursively define

$$c^{j-1} = \mathcal{H}^{[J-j]}c^j \quad \text{and} \quad d^{j-1} = \mathcal{G}^{[J-j]}c^j.$$

If the vector c^J is of length 2^J, then all the vectors c^j and d^j will be of the same length rather than getting shorter as j decreases as in the standard DWT.

The computational complexity of the NWT is $O(N\log N)$. For time series analysis the wavelets from the transform can be 'weaved' back into time order to provide the time ordered non-decimated wavelet transform; refer to Nason and Sapatinas (2002) for details.

2.7.7 The wavelet packet transform

The decomposition of a signal by the DWT (or NWT) is rather restrictive in terms of the frequency resolutions. For some time series, *wavelet packets* may be of more use because they provide a wider choice of decompositions of the frequency domain. They are particularly suited to representing signals which combine stationary and non-stationary characteristics, such as local and prolonged oscillations.

A time series is transformed into wavelet packet coefficients using the wavelet packet transform (WPT), developed by Coifman and Wickerhauser (1992). The WPT is a generalisation of the DWT (Section 2.7.5):

Given a function f which is observed at $N = 2^J$ equally spaced time points $\{t_i\}_{i=0,\ldots,2^J-1}$, set $\mathcal{W}_{J,f=0,i} = f(t_i)$ for $i = 0, \ldots, 2^J - 1$, i.e. define the set of finest scale wavelet packet coefficients to be the original time series. At scale $J-1$, the

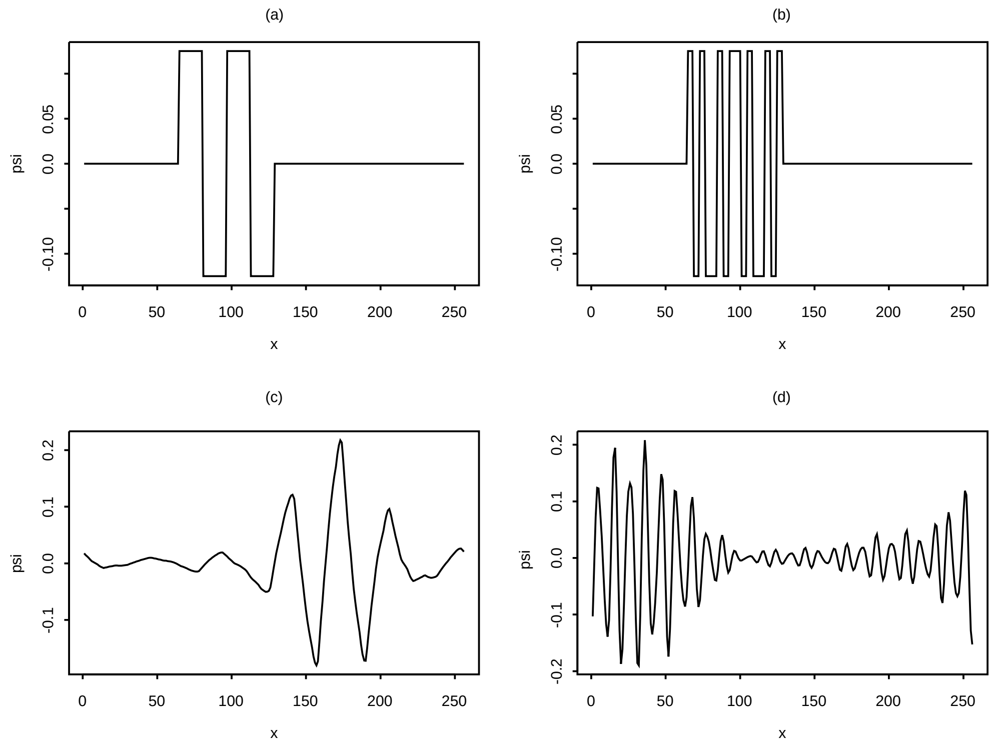

Figure 2.5: Some wavelet packets at scale $j = 2$ with frequency index f (N = 2^8): (a) based on Haar wavelet ($f = 2$), (b) based on Haar wavelet ($f = 10$), (c) based on Daubechies' least asymmetric with 4 vanishing moments ($f = 2$) and (d) based on Daubechies' least asymmetric with ten vanishing moments ($f = 10$).

wavelet packet coefficients $\mathcal{W}_{J-1,f,k}$ ($f = 0, 1$) are obtained by applying the filters h and g to $\mathcal{W}_{J,0,k}$ followed by dyadic decimation. At each further scale in the analysis the filters h and g are recursively applied (with dyadic decimation) to all wavelet packet coefficients $\mathcal{W}_{j-1,f,k}$ ($f = 0, 1, 2, \dots$) from the previous scale.

Note that each set of wavelet packet coefficients is now indexed by an additional parameter f, which corresponds to the frequency index of the wavelet packet. Wavelet packet coefficients with frequency index 0 correspond to father wavelets and wavelet packet coefficients with frequency index 1 correspond to mother wavelets.

The WPT is highly redundant, retaining $(J+1)N$ coefficients and the computational complexity of the WPT is $O(N\log N)$. Some wavelet packets at scale 2 corresponding to a sequence of length $N = 2^8$ are illustrated in Figure 2.5. Refer to Vidakovic (p. 133-140, 1999) for further details on wavelet packets.

Wavelet packets have been successfully used as the FBI standard for fingerprint image compression (Bradley *et al.*, 1993). We discuss their application to time series analysis in Section 2.8. A subset of the wavelet packets can be chosen in an adaptive way depending on the signal characteristics (but only certain choices will be orthogonal). For example, if we wish to compress the signal we may determine a cost function which allows us to reconstruct the signal to a given precision (see Vidakovic (Section 5.4, 1999) for details).

2.7.8 The non-decimated wavelet packet transform

The non-decimated wavelet packet transform (NWPT) was developed by Pesquet, Krim and Carfantan (1996). Its implementation follows the same procedure as the extension of the wavelet transform to the non-decimated wavelet transform (Section 2.7.6) but extending the WPT, where the filters $\mathcal{H}$ and $\mathcal{G}$ are now applied to both c^j and d^j recursively for both odd and even dyadic decimation. The computational complexity of the NWPT is $O(N^2)$. Figure 2.6 provides an illustration of the transform.

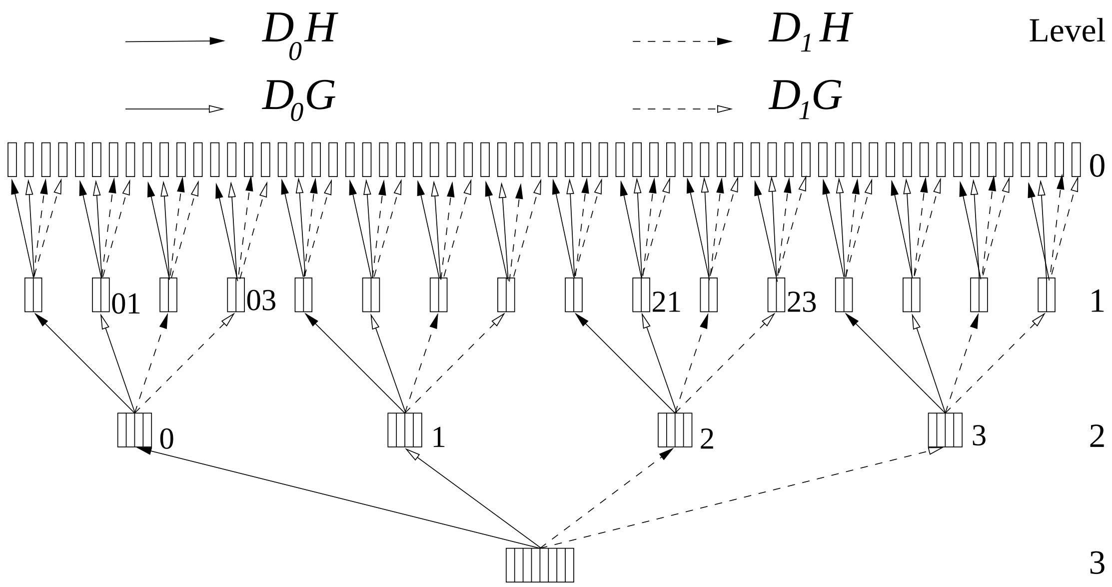

Figure 2.6: Schematic of NWPT for $N = 8$ points ($j = 3$). This figure is reproduced with permission from Nason and Sapatinas (2002).

2.7.9 Two-dimensional wavelets

Wavelets have been used in the field of image processing, prompting the introduction of two-dimensional (2D) wavelet transforms. We document the 2D wavelet transform;

refer to Section 5.7 of Vidakovic (1999) for details of a 2D multiresolution analysis.

The 2D discrete wavelet transform

Mallat (1989) proposes a 2D discrete non-decimated wavelet transform (2D NWT) based on a separable transform:

Given an image C^J of size (2^J x 2^J), the 2D DWT of the image is obtained by recursively using the relations

$$\begin{aligned} C^{j-1}_{u,e} &= \sum_m \sum_n h_{m-2u} h_{n-2e} C^j_{m,n}, \\ D^{h,j-1}_{u,e} &= \sum_m \sum_n g_{m-2u} h_{n-2e} C^j_{m,n}, \\ D^{v,j-1}_{u,e} &= \sum_m \sum_n h_{m-2u} g_{n-2e} C^j_{m,n}, \\ D^{d,j-1}_{u,e} &= \sum_m \sum_n g_{m-2u} g_{n-2e} C^j_{m,n}. \end{aligned}$$

where, the superscripts h, v and d denote horizontal, vertical and detail respectively.

The detail subscripts refer to the fact that each mother wavelet tends to extract information from a specific direction; see Vidakovic (p. 160, 1999) for an illustrative example. The algorithms for decomposition and reconstruction are illustrated in Figures 2.7 and 2.8 respectively.

The 2D non-decimated discrete wavelet transform

The 2D non-decimated discrete wavelet transform (2D NWT) follows the algorithm below:

1. Take the square matrix C^j of size 2^J x 2^J, which represents the smoothed image at level j. Apply the filters, h and g, to the rows of C^j, *retaining all columns*. Two resulting matrices are formed: hC^j and gC^j, of size 2^J x 2^J.

2. Apply filters h and g to the columns of hC^j and gC^j, *retaining all rows*. Four resulting matrices are formed: hhC^j, ghC^j, hgC^j and ggC^j, of size 2^J by 2^J.

The first matrix hhC^j, is the smoothed matrix whilst ghC^j, hgC^j and ggC^j, are the detail matrices picking out information in the vertical, horizontal and diagonal direction respectively. The 2-step process occurs recursively with hhC^j as the input matrix C^j for the J scales in the analysis.

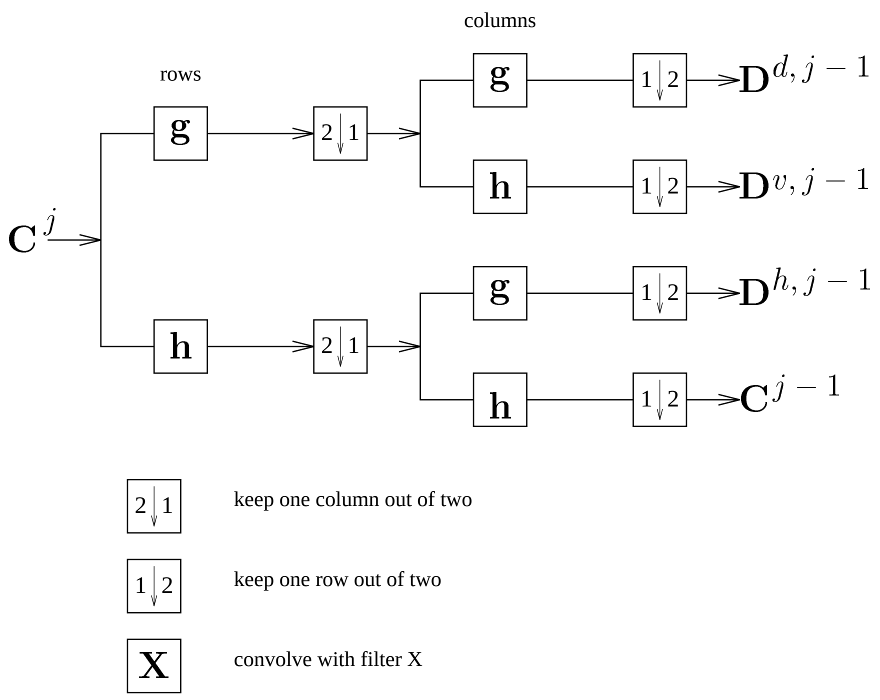

Figure 2.7: Decomposition of an image. The algorithm is based on one-dimensional convolutions along rows and columns of the image (after Mallat, 1989).

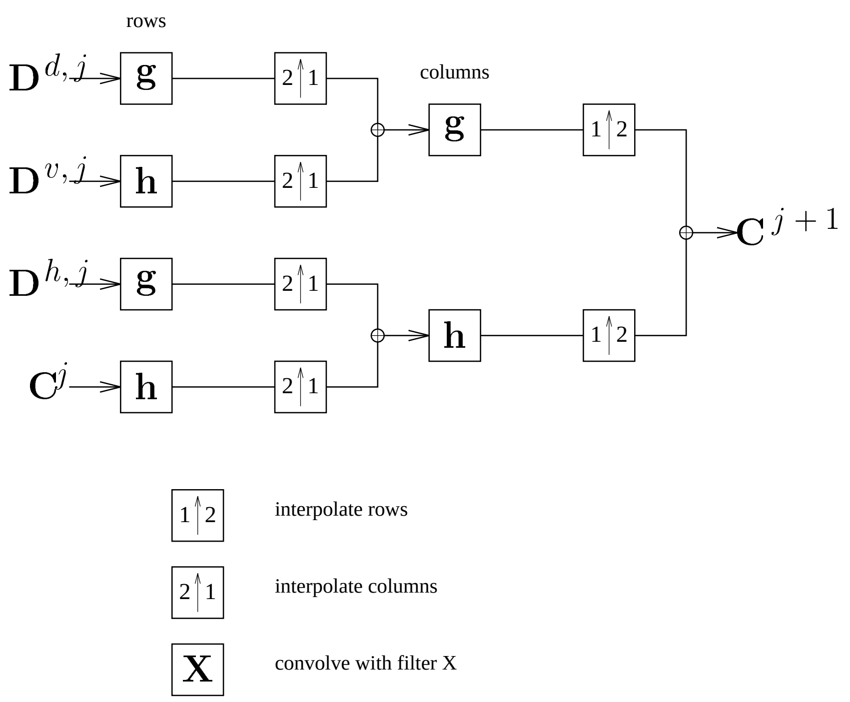

Figure 2.8: Reconstruction of an image (after Mallat, 1989).

2.7.10 Other wavelet topics

The following topics are not covered in the thesis but are included to complete this general overview of wavelets.

- Wavelet shrinkage and thresholding

  The goal of non-parametric regression is to estimate the unknown function f from the observations y_i. Suppose we have N (=2^J) noisy samples of the function f

  $$y_i = f(t_i) + \varepsilon_i, \quad \text{for } i = 1, \dots, N,$$

  where $t_i = (i-1)/N$ and the ε_i are independent, identically distributed $\mathrm{N}[0, \sigma^2]$ random variables. We wish to estimate the value of the function $\mathbf{f} = (f(t_1), \dots, f(t_n))$ by removing the effects of noise, ε.

  Wavelet shrinkage is an approach to recovering the original function, following three steps:

  1. Decomposing the data using a wavelet transform.
  2. Modifying the coefficients to remove the effects of noise.
  3. Inverting the modified wavelet coefficients to obtain an estimate of $\mathbf{f}$.

  Thresholding is a popular technique method for modifying the small wavelet coefficients which are assumed to be noise. A popular thresholding scheme is the universal threshold policy, proposed by Donoho and Johnstone (1994).

  Definition 2.5

  The universal threshold *is defined as* $\lambda = s\sqrt{2logN}$, *where* N *is the number of data points and* s *is the estimate of variation in the coefficients.*

  The policy can then be applied to the coefficients as either hard or soft thresholding. For hard, if the coefficients are smaller in absolute magnitude than a given threshold the coefficients are kept, otherwise they are removed. For soft, the coefficients are modified by the formula:

  $$d^*_{jk} = \mathrm{sgn}(d_{jk})(|d_{jk}| - \lambda)_+,$$

  where λ is the threshold (Nason & Silverman, 1994).

The universal threshold assumes the noise to be uncorrelated. Johnstone and Silverman (1997) consider the situation where the noise is correlated and stationary. One effect of the correlation is the variance of the wavelet coefficients depend on level j; they therefore propose the use of a level-dependent soft thresholding scheme in the case of correlated noise.

For further details on wavelet shrinkage and thresholding refer to, for example, Abramovich *et al.* (2000) and Vidakovic (1999).

- Locally stationary wavelet processes

Locally stationary time series exhibit patches of stationarity. If we can collect sufficient information in the region of stationarity, a sensible estimate of its statistical properties may be obtained. A review of locally stationary models is given in Nason and von Sachs (1999).

Nason *et al.* (2000) consider a locally stationary model which incorporates wavelets.

Definition 2.6

The locally stationary wavelet processes are a sequence of doubly indexed stochastic processes $\{X_{t,T}\}_{t=0,\dots,T-1}, T = 2^J \geq 1$ *having the following representation in the mean square sense:*

$$X_t = \sum_{j=-J}^{-1} \sum_k \omega_{jk}^0 \psi_{jk}(t) \xi_{jk},$$

where the ξ_{jk} *are mutually orthogonal zero mean random variables,* $\{\psi_{jk}(t)\}_{jk}$ *is a discrete non-decimated family of wavelets for* $j = -1, -2, \dots, -J(T), k = 0, \dots, T-1$ *based on a mother wavelet* $\psi(t)$ *of compact support and* ω_{jk}^0 *are the amplitudes which quantify the energy contribution to the process at scales* j *and location* k.

Nason *et al.* (2000) propose, when using the locally stationary wavelet approach, an evolutionary wavelet spectrum may be estimated which quantifies power over time and scale:

$$S_j(z) = S_j(k/T) \approx \omega_{jk}^2,$$

for $k = 0, \ldots, T-1$, where $z \in (0,1)$. This measures the local power in a LSW process at a particular (rescaled) time, z, and scale, j. Compare this to the 'usual' spectrum which measures the contribution to variance at a given frequency.

Theoretical foundations, including details of discrete wavelets, are given in Nason *et al.* (2000).

2.8 Wavelets in time series analysis

2.8.1 Introduction

We now review some of the roles wavelets may play in time series analysis. A gentle introduction to the statistical applications of wavelets (including numerous references) is provided by Abramovich *et al.* (2000). Nason and von Sachs (1999) provide a review of wavelets in time series analysis, covering topics such as wavelet denoising, long memory processes, spectral analysis and locally stationary processes. Recent books reviewing wavelets in time series analysis include Vidakovic (1999) and Percival and Walden (2000).

A favourable property of wavelets for time series analysis is that wavelet transformations tend to simplify the dependence structure in the original series. Figure 2.9 illustrates this point using some data from Chapter 3; (a) shows the first 64 observations of the market share of a consumer product, (b) shows how the autocorrelation function of the market share decays slowly and (c) shows how the DWT of the market share is uncorrelated. In this case we used the Haar wavelet, noting the dependence in the transformed domain will be influenced by the choice of decomposing wavelet.

Wavelets allow a more accurate description of features in both time and frequency because of their localised support. The wavelet coefficient d_{jk} provides information about the function near time point $2^{-j}k$ and near frequency proportional to 2^{j}.

Ramsey investigates the utility of wavelet methods in financial time series, see Ramsey *et al.* (1995), Ramsey and Zhang (1997), Ramsey (1999) and Ramsey and Lampart (1998). Motivated by the importance of time scales in economic decision making, they use wavelets and wavelet packets to decompose time series into their

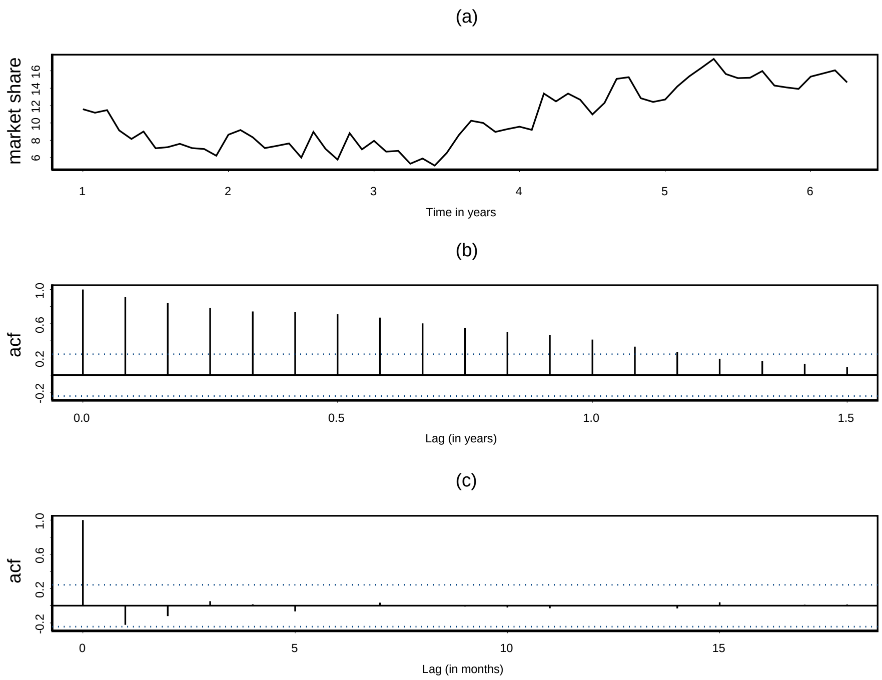

Figure 2.9: Product one: (a) market share (b) acf of the market share and (c) acf of the DWT of market share.

component time scales. For example, Ramsey and Lampart (1998) investigate the relationship between consumption and income. They study the relationship over different time scales and investigate the approximate distributional properties of the regression estimators and residuals of such models to be used in hypothesis testing and confidence intervals. Their method is exploratory; they do not directly provide a statistical model between the series.

Learned and Willsky (1995) use the WPT to extract features from a data set of naturally occurring biologically generated underwater acoustic signals in ocean noise. From the WPT of each signal they create an energy map which can be used to identify distinguishing features which determine the source of the signal. They explain the benefits of the WPT over traditional STFT methods in its use for automatic transient signal classification.

Walden and Contreras Cristan (1998) use the NWPT in the analysis of a single non-stationary series of hourly averaged Southern hemisphere solar magnetic field magnitude observations. The time series consists of 4096 hourly observations and the aim is to determine the timing of interplanetary shock waves. They compute a level 4

NWPT and, plotting the 100 largest coefficients in a time-frequency plot, demonstrate how the largest NWPT coefficients are aligned in time with the physical shock waves. They compare the NWPT method to cosine packet and STFT methods. Their aim is to determine the timing of events rather than predicting future events.

2.8.2 Wavelet packet transfer function methodology

In Chapters 3 and 4 we utilise the wavelet packet transfer function methodology of Nason and Sapatinas (2002). Nason and Sapatinas (2002) propose the use of the time ordered non-decimated wavelet packet transform (NWPT) in modelling a response time series, y_t, in terms of an explanatory time series, x_t. The explanatory time series is expressed as a heavily redundant combination of wavelet packets ψ_{jfk}, i.e.

$$\sum_j \sum_f \sum_k d_{jfk}\psi_{jfk}(x), \quad j, f, k \in \mathbb{Z}, \tag{2.7}$$

where d_{jfk} are the NWPT coefficients. Nason and Sapatinas (2002) use the Haar wavelet family because it performs well in their case study. In Nason *et al.* (2001), where an infant's sleep state is predicted from its heart rate, a smoother Daubechies' wavelet is used.

Taking a time-ordered NWPT of a time series of length N results in $2N-2$ sets of NWPT coefficients. The NWPT coefficients are stacked together to form a multivariate time series matrix. For illustration we show the multivariate time series matrix when $N = 4$, where N is the length of the time series (so $k = 1, 2, 3, 4$). We denote the NWPT coefficients as $d_{jf}(k)$ because it is clearer notation for this example. At the first scale ($j = 0$) there is one father wavelet ($f = 0$) and one mother wavelet ($f = 1$). At the second scale ($j = 1$) there is one father wavelet ($f = 0$), one mother wavelet ($f = 1$) and two wavelet packets ($f = 2$ and 3). The multivariate time series matrix is as follows:

$$\begin{pmatrix} d_{00}(1) & d_{01}(1) & d_{10}(1) & d_{11}(1) & d_{12}(1) & d_{13}(1) \\ d_{00}(2) & d_{01}(2) & d_{10}(2) & d_{11}(2) & d_{12}(2) & d_{13}(2) \\ d_{00}(3) & d_{01}(3) & d_{10}(3) & d_{11}(3) & d_{12}(3) & d_{13}(3) \\ d_{00}(4) & d_{01}(4) & d_{10}(4) & d_{11}(4) & d_{12}(4) & d_{13}(4) \end{pmatrix}$$

Each column contains NWPT coefficients corresponding to a particular wavelet packet and the k^{th} entry of the i^{th} column quantifies how similar the explanatory time series is to the corresponding wavelet packet at time point k.

As the multivariate time series matrix has $(2N-2)$ variables compared to N observations, to reduce the number of variables Nason and Sapatinas (2002) correlate the NWPT coefficients with the response time series and select 5% of the variables showing the highest correlations. Selecting just 5% of the non-decimated wavelet packet coefficients provides a data set where the number of explanatory variables is much smaller than the number of observations.

In Nason and Sapatinas (2002) the new methodology is used as a measure-correlate-predict (MCP) method to model wind speeds. Details of MCP techniques are given in Section 2.10. For the case study, a generalised linear model with gamma error and log link is used to model y_t in terms of selected NWPT coefficients now denoted $\mathcal{W}_m$, i.e.

$$y = \beta_0 + \beta_1 \mathcal{W}_1 + \beta_2 \mathcal{W}_2 + \ldots + \beta_p \mathcal{W}_p + \varepsilon, \tag{2.8}$$

where, $\{\beta_j\}_{j=0}^{p}$ are the regression coefficients, $\mathcal{W}_m$ $(m = 1, 2, \ldots, p)$ is the m^{th} set of NWPT coefficients chosen for inclusion in the model and ε is the error.

The predictions from the Nason and Sapatinas (2002) model are compared to predictions from a popular industrial method, namely a simple linear regression, i.e. $y = \beta_0 + \beta_1 x + \varepsilon$. In addition, they attach a physical interpretation to the wavelet MCP model; we discuss the interpretation of wavelet packets in the next section.

2.9 Wavelet packet coefficients as explanatory variables

2.9.1 Introduction

We consider a role that wavelet packet coefficients may play as explanatory variables. In particular, we consider the interpretation we can attach to wavelet packets and how we should treat wavelet packets which take values from the future as well as from the past.

2.9.2 Interpreting the wavelet packet combination plots

To attach an interpretation to the regression model (2.8) we use the wavelet packet combination plot, i.e. a graph of $\sum_{i=1}^{p} \beta_i \psi_i$, where β_i are the regression coefficients and ψ_i are wavelet packets represented in (2.8). In essence the wavelet packet combination plot is the impulse response function plot. The impulse response function describes how the output is related to the input of a linear system (p. 152, Chatfield (1989)). In Chapter 4 we use more complicated wavelet packet combination plots, where the plot represents (in addition) the weighting of each wavelet packet in a principal component.

Wavelet packet combination plots help provide an interpretation of the regression model. The wavelet packet provides valuable information about which components in the explanatory time series may possibly drive the response time series, i.e. which types of oscillatory behaviour in x_t may influence y_t.

As an example, a Haar wavelet packet at scale 3 and frequency index 2 is given in Figure 2.10. The wavelet packet represents the behaviour in the time series relative to time point $t = 0$ and is non-zero for a lag of sixteen time points. After a lag of 8 time points there has been one oscillation in the wavelet packet and after a lag of 16 time points there have been two oscillations in the wavelet packet. We would represent this wavelet packet in (2.8) if this behaviour is present in the explanatory time series and it helps to explain the response series.

2.9.3 Real-time wavelets

Wavelet packets (and wavelets) may utilise information from the future as well as information from the past. The decision to include or exclude packets which draw information from the past will depend on the aim of the analysis. If our aim is real-time prediction then we should restrict our analysis to wavelet packets which do not overlap with future data.

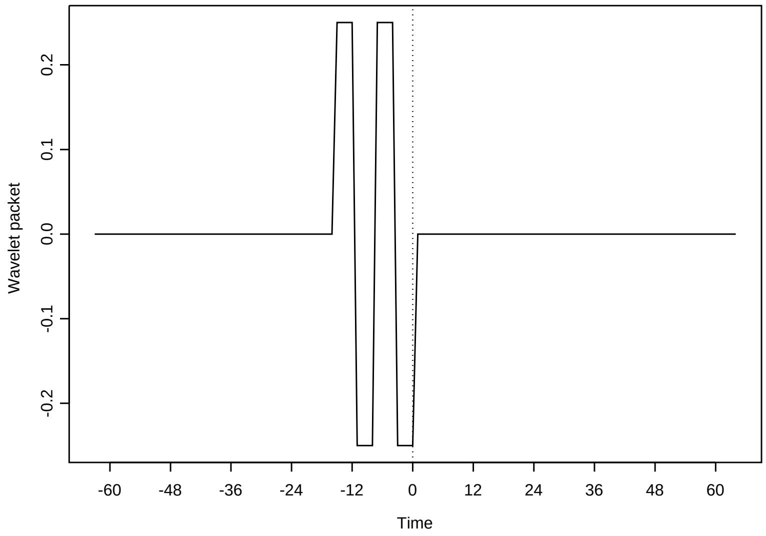

Figure 2.10: A Haar wavelet packet at scale $j = 3$ and frequency index $f = 2$ for a time series of length 128. This could be used to model time series which exhibit two oscillations where only the previous 16 observations are relevant.

2.10 Modelling and prediction of wind speeds

2.10.1 Introduction

The case study in Chapter 4 aims to model and predict wind speeds at a target location given wind speeds at a reference location. We provide a brief review of some relevant meteorological studies which have been carried out to date.

One method for analysing the univariate characteristics of wind speed series is to split the data set into components by frequency to form the spectrum. A peak in the spectrum will indicate an important contribution to variance at that frequency. The first comprehensive wind spectrum of this type was compiled by Van Der Hoven (1957). He concluded in a spectrum of hourly data we could expect to see two main features; a major peak at 0.01 cycles/hour which corresponds to the typical 4 day transit period of fully developed weather systems, called the macro-meteorological peak and a second smaller peak at 0.1 cycles/hour. Although Van Der Hoven's spectrum has been criticised because of the data set used for analysis, the general shape he identified is found in many wind speed spectrum analyses. See Cook (1986) for further information.

Many studies have considered fitting time series models to univariate data. For example, Brown *et al.* (1984) fit an AR(2) model to a time series consisting of hourly data over a one month period. Their model aims to take into account the autocorrelation, non-normal distribution and diurnal non-stationarity characteristics of the wind speed data. There are limitations in the paper; they restrict themselves to fitting an AR model (rather than the full class of ARIMA models), they provide little verification of the models predictive power (although simulation is a key part of the paper) and the model assumes stationarity over time (which may be reasonable over one month but possibly not for longer periods).

Our case study in Chapter 4 considers relating one time series to another time series and fewer studies have been carried out in this area. Most bivariate methods utilise some form of *measure-correlate-predict* (MCP) strategy. This method constructs statistical models which predict future values of a response time series from future values of an explanatory time series. Usually, a separate linear regression model is formed for 12 direction sectors so the model typically has 24 parameters. Derrick (1992) assessed potential wind farm sites by an MCP strategy using a linear regression model with general equation $y = ax + b$, where a and b are constants. Although the wind speeds y are not always linear in x, in his experience the majority of cases seemed to fit and the remainder of possibilities were covered by the power equation $y = ax^b$. Derrick suggested the strength of this method was the ease of calculating confidence limits. In addition he assesses the effects of diurnal variation and pressure systems and also the importance of time of flight (number of hours taken for winds to travel between sites). Clearly the weakness in using standard linear regression is that wind speeds are actually serially correlated.

Extensions to MCP methods have received little attention in refereed journals in recent years. One of the most recent publications is by Woods and Watson (1997). They discuss the problem of using standard MCP methods in complex terrain where the wind direction at the reference site may differ to the wind direction at the target site. This invalidates the standard assumption that the wind direction is the same at both sites. They predict wind speeds taking into account the wind direction at both sites by adjusting their predictions where there is significant 'mean veer' between sites. We will not consider using their extension because we only wish to predict the wind speed, and not the wind direction, so invalidating the assumption has little effect.

Further discussion of MCP methods is covered in the following three conference reports. Joensen *et al.* (1999) introduce a varying coefficient model estimated using local regression. For a given direction sector θ the model weights an observation in the given neighbourhood of θ, where the weight is dependent on the distance from θ to the observed wind direction. This will improve the predictive performance of the model when the wind direction falls close to the boundaries. Addison *et al.* (2000) consider using neural networks in a MCP model. Neural networks are mathematical models represented by a collection of simple computational units interlinked by a system of connections. The number of units can be very large and the connections intricate (Cheng & Titterington, 1994). Addison *et al.* (2000) discuss possible algorithms to be used in the 'hidden' layer and consider the inclusion of lagged variables. They conclude neural networks offer good levels accuracy. Manwell *et al.* (2001) provide an example of the MCP model with linear regression for a site off the coast of Southern New England. This is an applied example and does not extend the MCP model.

Haslett and Raftery (1989) estimate the long term power output at a site using long term hourly wind speeds from 12 meteorological stations in Ireland (the series are later aggregated into daily wind speeds). They use a square root transformation of the series to stabilize variance and make the marginal distributions approximately normal. They develop an ARMA model incorporating seasonality, spatial correlations, short-memory temporal dependence and temporal persistence. Problems with this approach include the model assumption that the second order moment structure of the space-time process is constant over Ireland.

2.10.2 Distributional assumptions for modelling wind speeds

The choice of distribution for the joint and marginal distributions for modelling wind speeds has been covered extensively in the literature. Although most researchers prefer to apply a transformation to the data set to obtain approximate normality there are many papers which discuss the characteristics of wind speeds that favour the fitting of a (univariate) Weibull distribution. See, for example, Stewart and Essenwanger (1978), Tuller and Brett (1984) and Lun and Lam (1999). Few papers discuss fitting the Weibull when jointly modelling two time series but for an example see Uitzinger *et al.* (1989).

The Weibull distribution is generally acknowledged to be a poor fit at low wind speeds. Takle and Brown (1977) suggest a hybrid density function $P_X^H(x)$ for describing a wind-speed distribution with non-zero probability of calms (zero wind speeds), defined as

$$P_X^H(x) = F_0\delta(x) + (1 - F_0)P^W, \qquad \text{for all real x,}$$

where, F_0 = probability of a calm, $\delta(x)$ is the Dirac delta function and P^W is the Weibull probability density function. The corresponding hybrid distribution function $F_X^H(x)$ is

$$F_X^H(x) = F_0 + (1 - F_0)F_X^W(x), \qquad x \geq 0,$$

where, $F_X^W(x)$ is the Weibull density function. Where the probability of a calm is very small it should be checked if there is actually a significant improvement in the fit of the model by using the hybrid distribution.

An example where this model has been implemented is given in Mulugetta and Drake (1996). They recommend plotting the fitted distribution curve to the observed histogram of data to identify discrepancies in certain wind intervals which may suggest when, and how, to include calms in the model.

In order to model the probability of a calm we define the following indicator:

$$h(t) = \begin{cases} 1, & \text{if wind speed} \leq 1.5\ m/s \\ 0, & \text{if wind speed} > 1.5\ m/s. \end{cases}$$

The motivation for defining a calm to be less than 1.5 m/s is the probability of having an exactly zero wind speed may be very small and it also allows for some measurement error. Due to the binary nature of the calm indicator we would use a generalised linear model with binomial family and logit link. Refer to Stern and Coe (1984) for a similar analysis on the distribution of wet and dry days in rainfall data.

2.11 Summary

We have summarised some important areas of time series analysis relevant to this thesis. In particular, we have considered regression based transfer function models for (possibly) non-stationary time series; these included ARIMA models and the Bayesian Dynamic Linear Model (DLM). The DLM is a structural model which specifies the trend, seasonal and regression components explicitly; the components are time-varying in order to capture the non-stationarity in the time series.

We considered the wavelet transforms to be utilised in later chapters: the NWT, the NWPT and the 2D NWT. Although non-decimated transforms are *not* orthogonal transforms they are translation-equivariant: a key property for our analyses. Drawing particular attention to the wavelet packet transfer function methodology of Nason and Sapatinas (2002) we have reviewed why wavelet packet coefficients may be useful as explanatory variables. We would refer the reader back to Section 2.9 when we refer to wavelet packet combination plots in later chapters.

The wavelet analysis in this thesis is performed using WaveThresh3 (Nason, 1998). In general, non-decimated wavelet transforms can be calculated on any number of points however the method of calculation (pyramid) that WaveThresh3 uses requires a time series with a length of power 2.

Chapter 3

Multiple Time Series Analysis of Marketing Data

3.1 Introduction

In this chapter we investigate how the relative price of a product drives its market share. The data set consists of the unit price and market share of six products recorded at 93 monthly intervals. By investigating the effect of price changes on the market share of competitive products it is possible to simulate alternative pricing polices.

The market researcher in the field employs various techniques to the data available. They include the efficient design of sampling schemes, multiple regression models, random effects models, cluster analysis and simulation studies. Of increasing popularity are neural networks, Bayesian methods and dynamic modelling techniques which incorporate the time-varying volatility of the market place. For examples of recent studies and numerous references refer to Rossi and Allenby (2000) who provide a detailed review of the use of statistics in marketing. However, there are very few works on modelling non-stationary relationships between time series let alone marketing series. Such non-stationary models have great potential in understanding and utilising marketing data as we hope to demonstrate.

We begin by performing some initial exploratory data analysis. We see if the data can be considered stationary and investigate the general trends in market share and log relative price (LRP) over time. We consider the relationship between LRP and market share for each product and look for relationships between products. We will then fit univariate ARIMA models to each time series and subsequently fit transfer function ARIMA models between market share and LRP for selected products.

We then consider model fitting methods which relax the stationarity assumptions required for ARIMA modelling. First, we consider the wavelet packet transfer function methodology of Nason and Sapatinas (2002). Using this method we fit models to explain the relationship between market share and LRP for products two and five individually. Second, we consider the methods of West and Harrison (1997), in particular the multiple regression dynamic linear regression model (RDLM), for product five only.

To conclude this chapter we suggest further new analyses for this data set. This chapter is the result of work produced in collaboration with a multinational consumer goods company. We have some reservations on reading too much into these data because the series are relatively short and there is evidence of non-stationarity.

3.2 The marketing data

3.2.1 Introduction

The price of a product is given relative to product one, i.e.

$$\text{Relative price of product } a = \frac{\text{Unit price of product } a}{\text{Unit price of product } 1}.$$

We have taken the log of the relative price because price movements are often multiplicative in nature, i.e. the change in price is directly proportional to the mean (p. 5, Diggle (1990)). The series consist of 93 time points, although products two and four are introduced at time points 14 and 60 respectively and product three is phased out after time point 40 and also contains missing data. The series start at year 1 and observations are recorded at monthly intervals. Note that, for these products, unit price (Figure 3.1) is comparable as they are the same houshold consumer good.

3.2.2 Initial investigation

- **Are the time series stationary?**

  Time series plots of the market share and LRP for each product indicate we

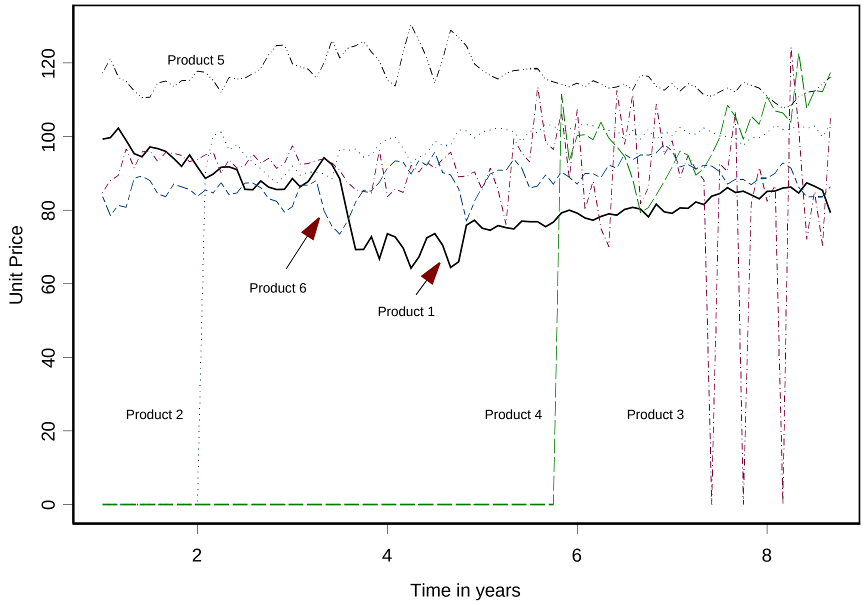

Figure 3.1: Unit Price of each product.

cannot assume that the series are stationary. If we take the first differences of the data it can be seen from Figures 3.2 and 3.3 that the first differences of the series might be stationary over certain time intervals. For example, we may split product five's LRP into three sections to obtain approximate stationarity in the variance in three time periods.

- **Market share of the six products**

  As can be seen in Figure 3.4, product one shows a small upward trend in its market share. Product two had shown a strong upward trend from its introduction at year 2 until year $5\frac{1}{2}$ when the trend increases at a slower rate. Product three began with a small market share that by year $5\frac{1}{4}$ had reduced to almost zero. Product four was introduced as the sales of product three declined; it initially shows a small increase but by year $8\frac{3}{4}$ the market share is less than 1%. Product five was initially the market leader with 60% of the market share but sales have declined steadily. Product six had shown a gradual downward trend but is now showing a slight upturn.

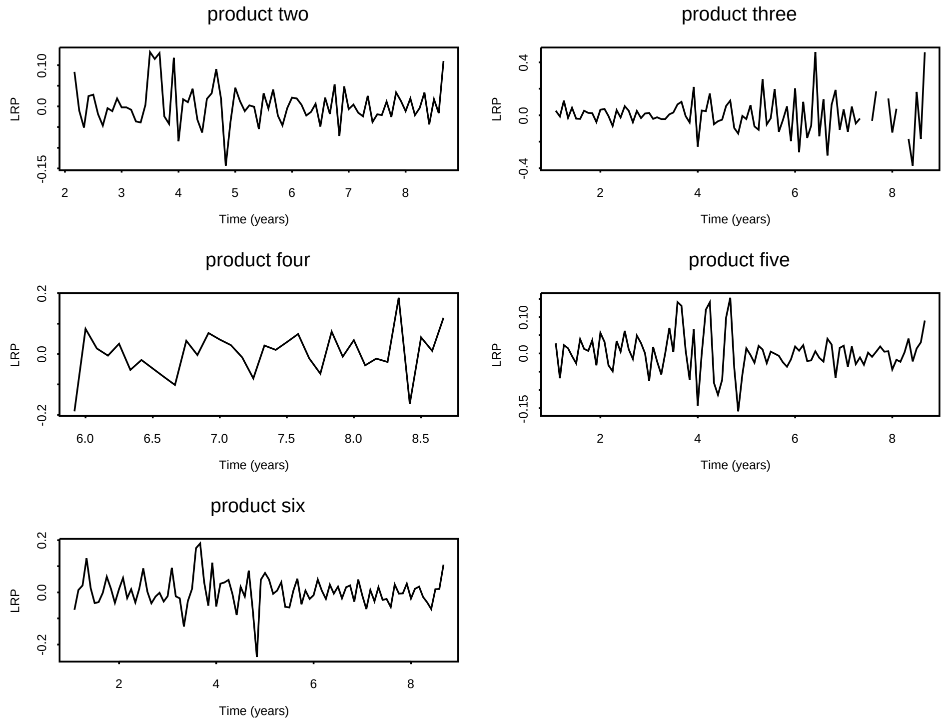

Figure 3.2: First differences of LRP for each product.

- **LRP of the six products**

  Figure 3.5 shows how LRP changes between products over time. We take our prices relative to product one because it has no missing observations and is relatively smooth. Product one started as one of the more expensive products but saw a sharp decrease in unit price around year $3\frac{3}{4}$. Although it became one of the cheapest products there has been a small upward trend since year 5. Initially, LRP for product two decreases but then increases steeply at year $3\frac{3}{4}$. After this jump (linked to a decrease in the unit price of product one) LRP has continued to decline. The LRP of product three initially increased but since year 4 has shown a variable pricing strategy. Product four was introduced around year $5\frac{3}{4}$. The LRP initially declined but has since shown an upward trend. Product five's LRP peaked around year 4. After some large fluctuations there has been a decrease in LRP, although recent observations may indicate a change to an upward trend. Product six initially increased in LRP but around year 5 began to decline.

- **Relationship between LRP and market share**

  Figure 3.6 shows product two has two interesting clusters, each demonstrating a

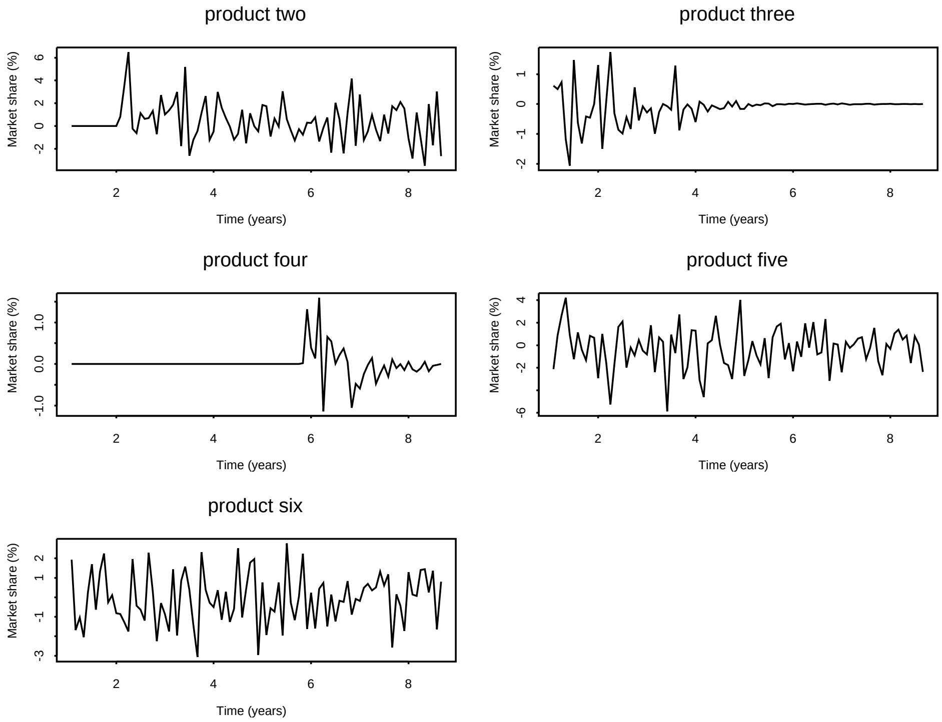

Figure 3.3: First differences of the market share for each product.

negative relationship, i.e. a high LRP is associated with a low market share and vice versa, as a data analyst might expect a priori. Product three demonstrates a negative relationship between market share and LRP, although the plot is skewed by the small market share after year $5\frac{1}{4}$. Product four has few observations but these indicate a negative relationship. Product five has two clusters with seemingly opposite relationships. At the bottom of the graph there is a cluster with a positive relationship whilst at the top of the graph there is a cluster with a negative relationship. Product six demonstrates a negative relationship.

If we produce the same plots using the raw data, i.e. unit price against market share, the relationships appear similar for all products apart from product two. Product two tends to a more positive relationship between market share and unit price.

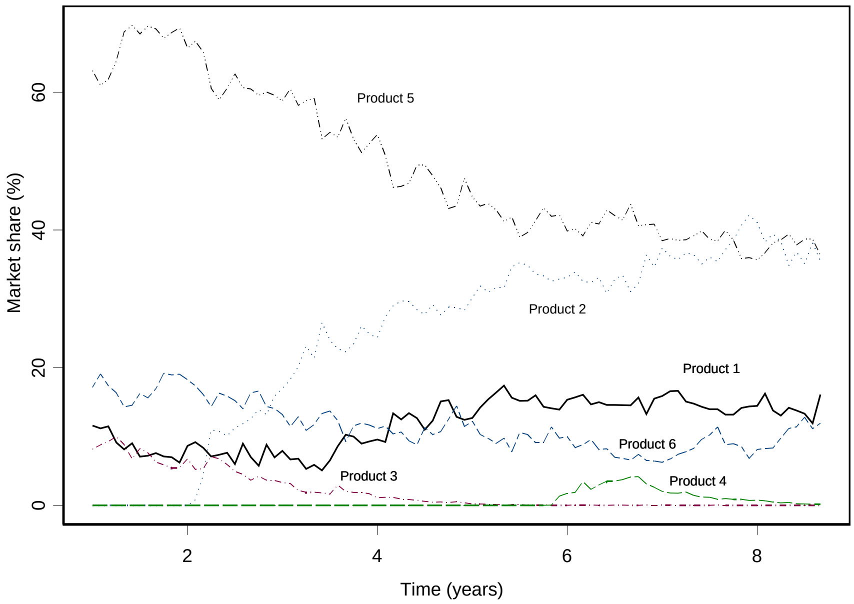

Figure 3.4: Market share of the six products.

3.2.3 Relationships between products

In Table 3.1 we look at the correlation between the first differences of unit price for the six products, excluding time points where either of the series have missing or zero values for market share. We are looking for an indication that the pricing strategies of the products may be related, excluding the general trend in price. We cannot read too much into these correlations because only the trend has been removed from the data. Only four values in the table are significant at the 5% level and we have highlighted these in bold.

Our initial look at this data set highlighted the following relationships:

1. Tentative relationship between products two and five?

   Figure 3.7 shows the relationship between product two and product five in terms of market share, LRP and unit price. It seems in the first plot that from its introduction product two has eroded the market share of product five. In the second plot we can see that the products have followed similar peaks and troughs in terms of LRP. In the third plot, where we look at unit price, we have transformed product two to show its negative relationship (plus a constant of 215

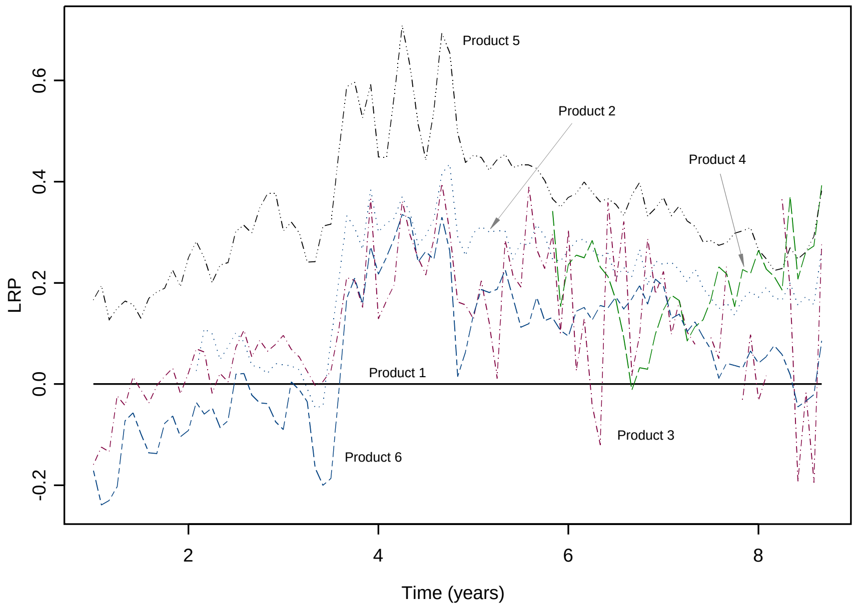

Figure 3.5: LRP of the six products (relative to product one).

to improve scaling on the graph) with product five. From this plot it appears that the unit price of the products exhibit a negative correlation. On occasions, as product five has increased its price then product two has in synchrony decreased its price; perhaps explaining why product two has increased its market share so successfully. In Figure 3.6 we picked out some unusual aspects in the relationship of LRP against market share for products two and five. In addition, we noted a significant correlation between the first differences of the unit price for these products.

	Prod. 1	Prod. 2	Prod. 3	Prod. 4	Prod. 5	Prod. 6
Prod. 1	1	-0.02	-0.13	**-0.23**	**-0.41**	**- 0.25**
Prod. 2	-	1	0.05	0.33	**-0.52**	-0.02
Prod. 3	-	-	1	0.10	0.02	0.02
Prod. 4	-	-	-	1	-0.05	**-0.35**
Prod. 5	-	-	-	-	1	**-0.42**
Prod. 6	-	-	-	-	-	1

Table 3.1: Correlation of the first differences of unit price between products.

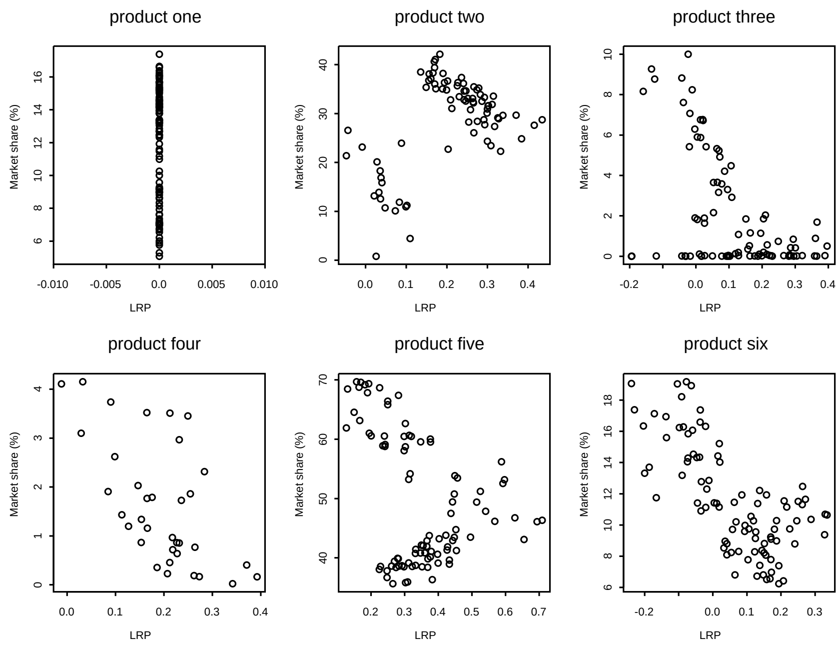

Figure 3.6: Relationship between LRP and market share.

2. Tentative relationship between products one and six?

   Whilst product one was showing a small upward trend in market share the market share of product six was declining. Around year 4, when sales of product six reduced to less than the sales of product one, product six had become the more expensive product, i.e. as product six increased its LRP relative to product one its market share declined. We note there was a significant correlation between the first differences of their unit price.

3. Tentative relationship between products three and four?

   Product four was introduced as the life cycle of product three ended.

3.3 Univariate ARIMA models

3.3.1 Introduction

We begin by fitting univariate time series models to each of the twelve time series in the data set. Univariate models provide a useful tool for checking data early on in the analysis. They increase understanding of how the time series might be generated

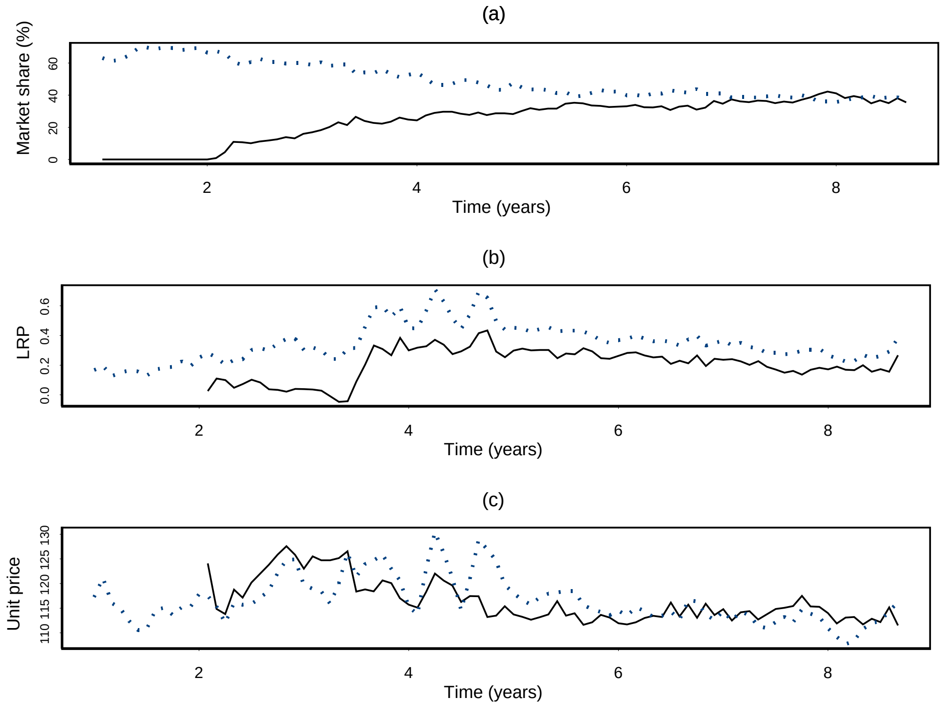

Figure 3.7: Relationship between product two (solid line) and product five (dotted line): (a) Market share, (b) LRP and (c) the inverse of the unit price for product two scaled by a constant $(-UP_{2,t} + 215)$ and the unit price for product five $(UP_{5,t})$.

and provide a comparison for more complicated models. Univariate models cannot be expected to provide reliable long term forecasts where the behaviour of possible explanatory variables (not included in the model) may change significantly in the future.

In the proceeding sections we refer to, for example, time point 13 rather than the start of year 2. To facilitate conversion between time point and year notation we provide Table 3.2, which lists the time point corresponding to the start of each year.

The notation we use in Section 3.3 is as follows:

$\bigtriangledown$	first difference operator
$X_{i,t}$	LRP of product i
$Y_{i,t}$	market share of product i

and

$$\begin{aligned} X_t &= Z_t + \varepsilon_1 Z_{t-1} + \ldots + \varepsilon_q Z_{t-q} && \text{is an MA(q) process,} \\ X_t &= \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t && \text{is an AR(p) process,} \end{aligned}$$

Time Point	Year
1	1
13	2
25	3
37	4
49	5
61	6
73	7
85	8

Table 3.2: Conversion table for time point to year. Each time point is the first month of the given year.

where Z_t is a purely random process with mean zero and variance σ_z^2 and $\{\varepsilon_i\}_{i=0}^{q}$ and $\{\alpha_j\}_{j=1}^{p}$ are the usual parameters.

3.3.2 The method

We have documented the procedure we use for fitting univariate ARIMA models in Section 2.2.2, which follows Box and Jenkins (1970). We have built the ARIMA model by adding one parameter at a time, assessing the fit at each stage using plots of the standardized residuals, the autocorrelation function (acf) and the partial autocorrelation function (pacf) of the residuals and the Portmanteau goodness of fit test statistic. We may experience problems if we wish to fit a fairly complex model due to the high number of parameters and relatively few observations in the series. Time series texts, such as Chatfield (1989), provide additional explanation of model fitting procedures.

3.3.3 Models for the LRP of each product

We fit univariate ARIMA(p, 1, q) models to the LRP of each product. We found that taking the first differences of the LRP is adequate to remove the trend in the data. Removal of other non-stationary features, such as changes in variance, were addressed by splitting the time series into sections of varying lengths.

1. Product two

   The $\bigtriangledown X_{2,t}$ can be considered stationary over 79 observations (14:93). The acf and pacf have no significant correlations so we choose to fit an ARIMA(0,1,0) model, i.e.

$$\nabla X_{2,t} = Z_t,$$

where Z_t has variance 0.0022.

2. Product three

   Figure 3.2 shows two changes in the variance of $\nabla X_{3,t}$. The first change appears around year 4 and the second change after year 7. The third section of the series corresponds to a diminishing market share, and also contains missing values, so we discard this section of time series.

   The first stretch of the series (1:35) appears to follow an ARIMA(0, 1, 0) process where Z_t has variance 0.0020.

   The second stretch of time series (36:77) has the first, fifth and sixth autocorrelations significant and the first and fourth partial autocorrelations significant. We choose to fit the simpler ARIMA(5,1,1) model constraining only the first and fifth AR terms and the first MA terms to be non-zero. The coefficients are $\alpha_1 = -0.2805$, $\alpha_5 = 0.4287$ and $\varepsilon_1 = 0.1917$. We choose this model because the acf and pacf of the residuals show no significant coefficients, although the pacf coefficients were mainly negative.

   From the models considered, the ARIMA(5, 1, 1) appeared the 'best' choice from diagnostic plots. However, we noted that the pacf coefficients were mainly negative so we should consider investigating a more complex model set. It may be possible to fit simpler models by further splitting of the time series but this would result in very short stretches of data.

3. Product four

   We consider the first differences of product four stationary from its introduction at observation 60. We fit an ARIMA(0, 1, 0) model where Z_t has variance 0.0045.

4. Product five

   We split the time series because of a change in variance. To the first stretch (1:31) we fit an ARIMA(0, 1, 0) model where Z_t has variance 0.0014. The second stretch of observations (32:47) we discard because it consists of just 15 observations. To the final stretch (48:92) we fit an ARIMA(0, 1, 0) model where Z_t has variance 0.0006.

5. Product six

   The LRP has a large negative value in the time series at time point 47. We split the time series around this point to obtain stationarity.

   The first stretch (1:46) has the third autocorrelation and third partial autocorrelation just significant. We began by fitting an ARIMA(3,1,3) (with only the third terms non-zero) but there were problems in estimating the parameters in S-Plus; namely a singularity in the asymptotic information matrix.

   Due to the non-identifiabililty problems associated with parameters in the above model, we investigated simpler models and found that an ARIMA(1,1,0) with $\alpha_1 = 0.1705$, provided a reasonable fit. Two concerns were the pacf coefficients tended to be negative and there were two consecutive outlying standardised residuals. Fitting more complicated models appeared to achieve little improvement in fit. This prompted to us to reconsider if the data could really be considered stationary over this interval. Unfortunately, splitting the data further would produce series too short to confidently fit models on.

   The remaining data (49:92) can be explained by an ARIMA(0, 1, 0) model where Z_t has variance 0.0011.

3.3.4 Models for the market share of each product

We fit univariate ARIMA(p, 1, q) models to the market share or the log market share of each product. We propose that first differences are adequate to obtain a stationary trend in the series.

1. Product two

   The $\bigtriangledown Y_{2,t}$ are stationary over observations 14:93, for which the seventh autocorrelation and partial autocorrelations are just significant. We decide to fit an ARIMA(7,1,1) model, where we constrain only the first and seventh AR term and the first MA term to be non-zero. This model appears reasonable; the acf and pacf of the residuals showed no significant coefficients, there are only two significant standardised residuals and there are no significant p-values for the Portmanteau significance test. The coefficients of the model are $\alpha_1 = -0.3836$, $\alpha_7 = -0.2083$ and $\varepsilon_1 = -0.2191$.

2. Product three

   We can assume stationarity in the first differences of the log of market share over observations 1:30. We fit an ARIMA(0,1,0) model, where Z_t has variance 0.0223. We take the log of market share because it stabilises the variability in the series.

3. Product four

   We assume the first differences of the log of market share are stationary between observations 60:93 and fit an ARIMA(0,1,0) model, where Z_t has variance 0.0525. We take the log of market share because it stabilises the variability in the series.

4. Product five

   Although some of the coefficients from the acf and pacf were on the boundary of significance, we assume the first differences of the market share are stationary over the whole series and fit an ARIMA(0, 1, 0) model, where Z_t has variance 3.47.

5. Product six

   We assume the first differences of the market share are stationary over observations 14:70 and discard the rest of the observations. The first autocorrelation and first and seventh partial autocorrelation are significant. We choose to fit an ARIMA(0,1,1), where $\varepsilon_1 = 0.4117$, for which there are two outlying standardised residuals and the pacf coefficients tended to be negative. The inclusion of additional terms (as suggested by the an acf and pacf) led to no significant improvement in model fit.

3.3.5 Summary of the univariate models

In general, the series were adequately represented by an ARIMA(0, 1, 0) model or models containing just a few AR or MA terms. In some cases it was necessary to take logs of the series in order to stabilise the variance. We note the series are quite short and show strong evidence of non-stationarity, so we treat these initial results with caution.

3.4 Transfer function ARIMA models

3.4.1 Introduction

Transfer function models contain input variables which affect all output variables unidirectionally and noise or error variables (described by a multivariate stochastic model) which enable the outputs to affect each other. In the following sections the transfer function models we will fit contain only one output time series. Transfer function models may provide a greater understanding of the behaviour of a system and result in more reliable predictions of the response series. Following the problems noted in the previous section, we decide to fit transfer function ARIMA models to products two and five only.

3.4.2 The method

We have followed the methodology described in Jenkins (1979). The time series are first prewhitened, as described in Section 2.2.3, before computing the cross-correlation function. Prewhitening the series is important because it is possible for two time series, which are unrelated, to give rise to apparently 'large' cross-correlation coefficients which are actually spurious (Chatfield, 1989). We use the information from the prewhitened cross-correlations to help us to fit the transfer function model,

$$Y_{i,t} = \omega_0 X_{i,t} + N_t,$$

where ω_0 is the regression coefficient of $X_{i,t}$ and N_t is the noise or error term. By looking at the cross-correlations for both positive and negative lags we can attempt to determine if the relationship is causal or contains feedback. As a first approximation to the noise model, N_t, we take the univariate model for Y_t. If there is no relationship between X_t and Y_t then we are simply fitting the univariate model. The parameters in the transfer function ARIMA model are simultaneously estimated using the `arima.mle` function in S-Plus.

3.4.3 Models for the relationship between LRP and market share of each product

1. Product two

   In Section 3.3.3 we saw that taking the first differences of $X_{2,t}$ for observations 14:93 converts the correlated series into an (approximately) random series. We also take the first differences of the market share in order to calculate the prewhitened cross-correlation function; see Figure 3.8 for the acf and pacf of the prewhitened series.

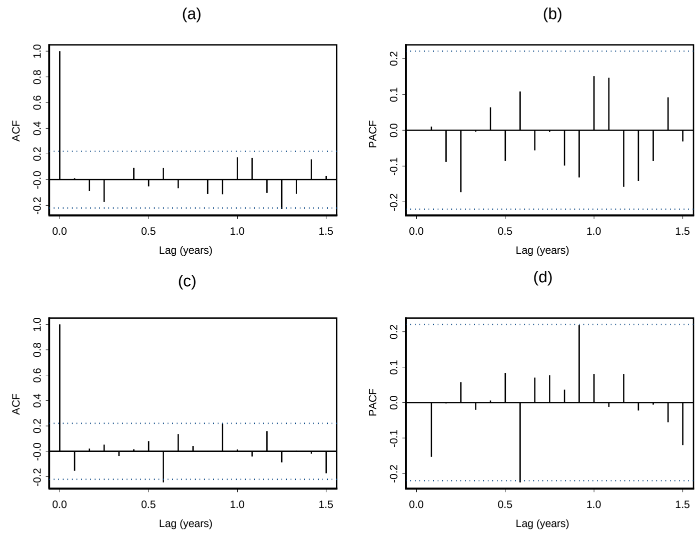

Figure 3.8: Product two (observations 14:93): (a) acf of the first differences of LRP, (b) pacf of the first differences of LRP, (c) acf of the first differences of market share and (d) pacf of the first differences of market share.

Figure 3.9 shows the prewhitened cross-correlation function has a large negative spike at lag 0 and smaller correlations for longer lags. This suggests an increase in LRP during this month decreases market share in the same period but has no influence on market share in the following periods. We propose the relationship

$$Y_{2,t} = \omega_0 X_{2,t} + N_t$$

between market share and LRP, where N_t is a noise or error term.

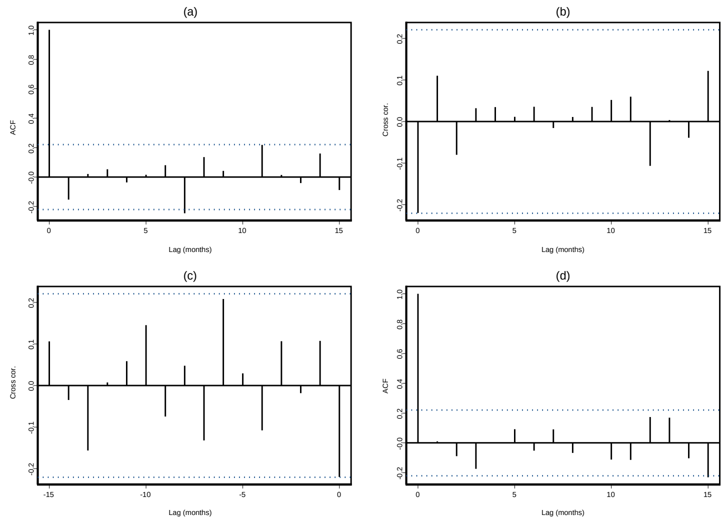

Figure 3.9: The prewhitened cross-correlation function for product two (14:93): (a) acf of LRP, (b) cross correlation between LRP and market share, (c) cross correlation between market share and LRP and (d) acf of market share.

As a first approximation to the noise model N_t, we can take the univariate model for $Y_{2,t}$, derived in Section 3.3.4:

$$\bigtriangledown N_t = \alpha_1 \bigtriangledown Y_{2,t-1} + \alpha_7 \bigtriangledown Y_{2,t-7} + \varepsilon_1 Z_{t-1} + Z_t.$$

We therefore obtain the transfer function-noise model:

$$\bigtriangledown Y_{2,t} = \omega_0 \bigtriangledown X_{2,t} + \alpha_1 \bigtriangledown Y_{2,t-1} + \alpha_7 \bigtriangledown Y_{2,t-7} + \varepsilon_1 Z_{t-1} + Z_t.$$

The ARIMA coefficients are $\alpha_1 = -0.4194$, $\alpha_7 = -0.2682$ and $\varepsilon_1 = -0.3813$ and the coefficient for the regressor variable, ω_0, is -10.7343. As the coefficient for the regressor variable is negative, it confirms our initial suggestion that there is a negative relationship between market share and LRP for product two. A t-test suggests we should consider including a constant (i.e. mean) in the model. The diagnostics of the model are reasonable.

2. Product five

   The first stretch of time series (1:31), where the $\bigtriangledown X_{5,t}$ are stationary and can be considered a random series, shows no significant prewhitened cross-correlations. We can assume there is no relationship between the market share and LRP over the first thirty-one observations for product five.

Recall we discarded the next fifteen observations.

The second stretch of time series (48:92), where the $\bigtriangledown X_{5,t}$ are stationary and represent a random series, shows no significant prewhitened cross-correlations. We can assume there is no relationship between the market share and LRP over remaining observations for product five.

In the second stretch of the time series there is a small spike (70) in the data. If we split the data around the spike the series will be too short to form a useful transfer function model. In addition, if we include the final observation (93) in the analysis there will be a significant correlation at lag zero in the prewhitened cross-correlation function. This raises the question of how much we can interpret these data using ARIMA models due to the length of data and evidence of non-stationarity.

3.4.4 Summary of the transfer function models

For product two we identified a negative relationship between market share and LRP. For product five we concluded there was no significant relationship between market share and LRP. We note the series for product five were short and may still contain some non-stationary behaviour which could have distorted our results.

3.4.5 Multivariate transfer function models

In the previous section we looked at the relationship between LRP and market share for products two and five individually. We wish to investigate this relationship further in a multivariate setting.

To fit a multivariate transfer function model to products two and five we would first fit a univariate model for $Y_{2,t}$ modelled in terms of $X_{2,t}$ and $X_{5,t}$ and a univariate model for $Y_{5,t}$ modelled in terms of $X_{2,t}$ and $X_{5,t}$. We would then cross-correlate the residuals from the two models to determine multivariate stochastic models for the residuals of the single output models. The residual multivariate model is then combined with the two single output models to provide the multiple output – multiple input models.

We decide against fitting multivariate ARIMA models because of problems previously highlighted, namely the extent of non-stationarity in the data. We cannot determine a stretch of observations long enough to build a reliable multiple input – multiple output model. We decide to investigate alternative methods which seek to avoid the problems of adjusting the series prior to analysis when the series are non-stationary.

3.5 Wavelet packet transfer function models

3.5.1 Introduction

A method which requires the data to exhibit only patches of (possibly) locally stationary behaviour is wavelet packet transfer function modelling, as proposed by Nason and Sapatinas (2002). We consider products two and five because the products are of reasonable length and have shown an interesting (possibly related) relationship to date.

3.5.2 The method

We construct a linear regression to model the market share of a product in terms of a time ordered non-decimated wavelet packet transform (NWPT) of its LRP, i.e.

$$\text{market share} = \mathcal{W}_1 + \mathcal{W}_2 + \ldots + \mathcal{W}_p + \varepsilon, \quad \varepsilon \sim \mathrm{N}[0, \sigma^2],$$

where, $\mathcal{W}_m$ $(m = 1, 2, \ldots, p.)$ are the non-decimated wavelet packet coefficients of the LRP. We refer the reader to Section 2.8.2 where this model is explained in full.

We carry out stepwise analysis for multiple regression to determine which packets may be important in modelling the market share. We decide to exclude wavelet packet coefficients which take data from the future. To obtain the final model we exclude one variable at a time, according to the size of its t-value, until all remaining packets are significant.

We found the Haar wavelet to be adequate in this analysis; the wavelet packet models we fit explain up to 96% of the variation in the response series. We took observations 14 to 93 from each time series and padded the series with zeros to obtain a time series of length 2^7 in order to compute the NWPT using WaveThresh3.

To aid interpretation of the regression models we provide details of:

Packet Scale	Months
6	2
5	4
4	8
3	16
2	32
1	64
0	128

Table 3.3: Wavelet packet scale relationship with time (for time series of length 128).

1. The coefficients, t-statistics and other diagnostics from the linear regression fit.

2. The wavelet packet coefficients included in the model. To explain their meaning we provide the wavelet packet combination plot, which may explain a component of the behaviour in the time series at any given time point (see Section 2.9). This allows us to suggest a direct interpretation for our models in terms of price elasticity of demand, i.e. the sensitivity of buyers to price changes in terms of the amount of goods they buy. For a time series of length 128 (=2^7) the scale of a wavelet packet is related to the time period shown in Table 3.3.

3. Graphs of the actual and fitted values of market share.

To begin we investigate the products individually, considering how each product's market share relates to its LRP. This effect may be causal or associative, i.e. the LRP may have a direct effect on market share or be merely associated with it; our analysis cannot tell us which.

3.5.3 Relationship between LRP and market share for product two

Following Section 3.5.2, we fit the model:

$$\text{market share} = \mathcal{W}_1 + \mathcal{W}_6 + \mathcal{W}_{11} + \mathcal{W}_{13} + \mathcal{W}_{22} + \mathcal{W}_{25} + \varepsilon, \quad \varepsilon \sim \mathrm{N}[0, \sigma^2].$$

The coefficients and t-statistics for this model are given in Table 3.4. The wavelet packets represented in the model are given in Table 3.5. The model has a multiple R^2

term of 0.94, i.e. the model explains 94% of the variation in the market share. Plots generated in S-Plus (e.g. normal probability plot, residuals versus fitted values) suggest the model is a reasonable fit. We note there are some high correlations between wavelet packet coefficients in the model (Table 3.6).

	Value	Std. Error	t-value	Pr($>\|t\|$)
Intercept	34.62	1.65	20.98	0.000
$\mathcal{W}_1$	-21.93	9.94	-2.21	0.031
$\mathcal{W}_6$	-24.63	8.57	-2.88	0.005
$\mathcal{W}_{11}$	-163.87	14.90	-11.00	0.000
$\mathcal{W}_{13}$	-93.76	27.08	-3.46	0.001
$\mathcal{W}_{22}$	92.41	10.38	8.90	0.000
$\mathcal{W}_{25}$	49.98	4.82	10.36	0.000

Table 3.4: Regression coefficients for the product two model.

The wavelet packet combination plot is given in Figure 3.10 (individual wavelet packets given in Figure 3.11). Using this graphical representation we suggest a price elasticity interpretation for our chosen model. At time $t = 0$ we see a large negative correlation between market share and LRP. This suggests a reasonable assumption that an increase in LRP may lead to an immediate decrease in market share and vice versa. We then see a smaller negative effect as, at longer lags, the initial LRP change has less effect on the number of products sold. After approximately 18 months lag there is a change to a positive relationship which is smaller in magnitude.

To further investigate this relationship we refer to Figure 3.12 which shows the changing relationship between market share and LRP over time. As noted earlier there are two clusters in the graph both exhibiting a negative relationship. We suggest

Wavelet Packet	Packet Scale	Packet Freq.
$\mathcal{W}_1$	1	0
$\mathcal{W}_6$	4	0
$\mathcal{W}_{11}$	1	3
$\mathcal{W}_{13}$	1	5
$\mathcal{W}_{22}$	1	2
$\mathcal{W}_{25}$	1	1

Table 3.5: Wavelet packets represented in the product two model.

	$\mathcal{W}_1$	$\mathcal{W}_6$	$\mathcal{W}_{11}$	$\mathcal{W}_{13}$	$\mathcal{W}_{22}$	$\mathcal{W}_{25}$
$\mathcal{W}_1$	1	0.80	-0.76	-0.56	0.18	0.19
$\mathcal{W}_6$	-	1	-0.85	-0.81	0.15	0.18
$\mathcal{W}_{11}$	-	-	1	0.56	-0.08	-0.10
$\mathcal{W}_{13}$	-	-	-	1	-0.03	-0.08
$\mathcal{W}_{22}$	-	-	-	-	1	0.77
$\mathcal{W}_{25}$	-	-	-	-	-	1

Table 3.6: Correlation of wavelet packet coefficients included in the product two model.

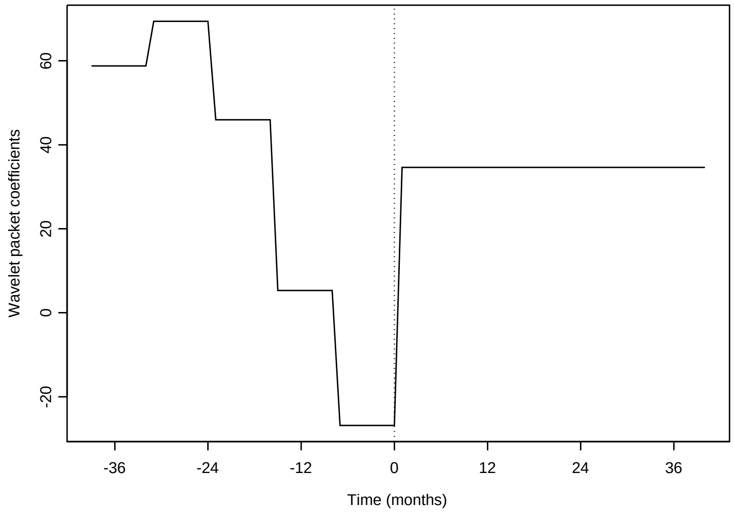

Figure 3.10: Wavelet packet combination plot of the product two model.

the model is accurately picking up the strong negative correlation between the series and the small positive effect may be caused by the shift in LRP around time points 15 to 21.

We provide a graphical representation of the fitted values given by our regression model compared to the actual market share for product two in Figure 3.13. We see the model provides a fair approximation of the actual market share.

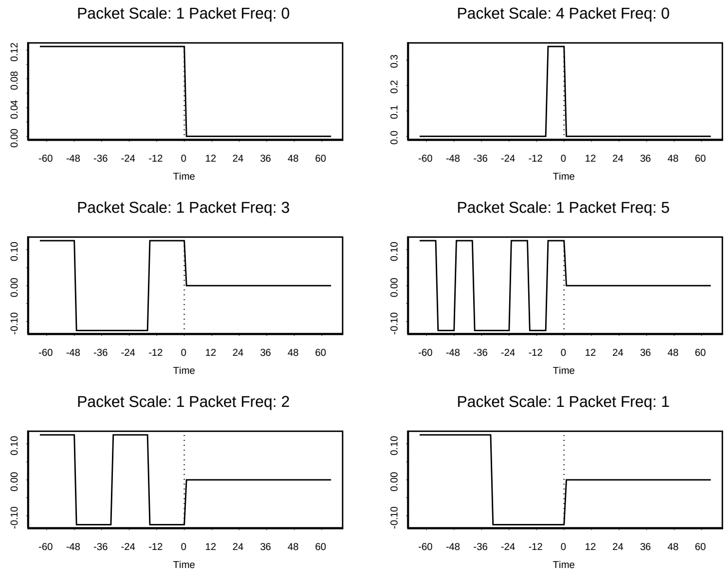

Figure 3.11: Wavelet packets in the product two model.

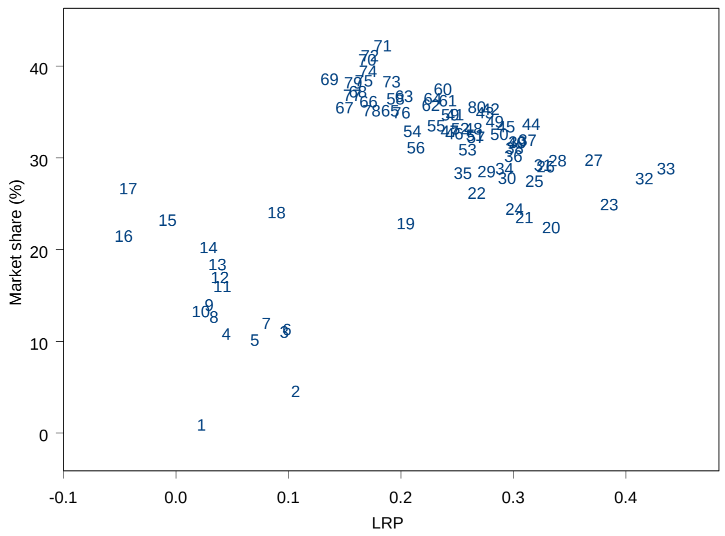

Figure 3.12: The changing relationship between LRP and market share for product two over time (this is the upper middle graph in Figure 3.6). Each number in the plot corresponds to the time ordering of the series.

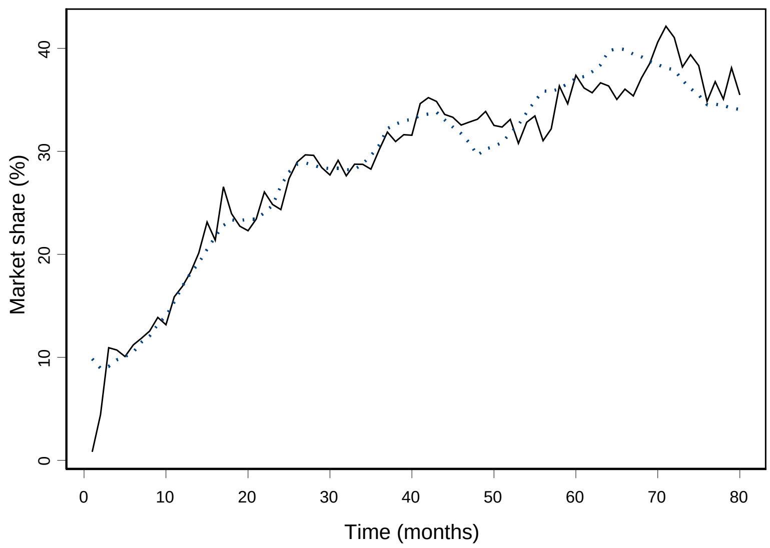

Figure 3.13: Actual and fitted values for the market share of product two model (dashed line = fitted values, solid line = actual data)

- **Effect of padding the marketing data**

  For the previous model, we padded the series in order to compute the NWPT using WaveThresh3. We now investigate whether it is preferable to taper the data rather than pad it with zeros. Tapering is a technique applied to time series to reduce the "leakage" phenomenon in spectral estimates. Leakage occurs when there is a large amplitude peak at a particular frequency f. Then the spectral estimates at frequency f can be higher than expected, and can easily obscure nearby lower amplitude peaks (Guide to Statistics, S-Plus, 2000).

  We use a form of tapering to examine the effect of the discontinuities at the end of the series caused by padding the data with zeros. Our data taper smooths the jump between the padding zeros and the actual time series. The effect on the data set is shown in Figure 3.14. Before time point 25 and after time point 104, where the discontinuity used to occur, is now a smoothly changing section. We could consider applying further tapering to the data to smooth some large jumps within the time series.

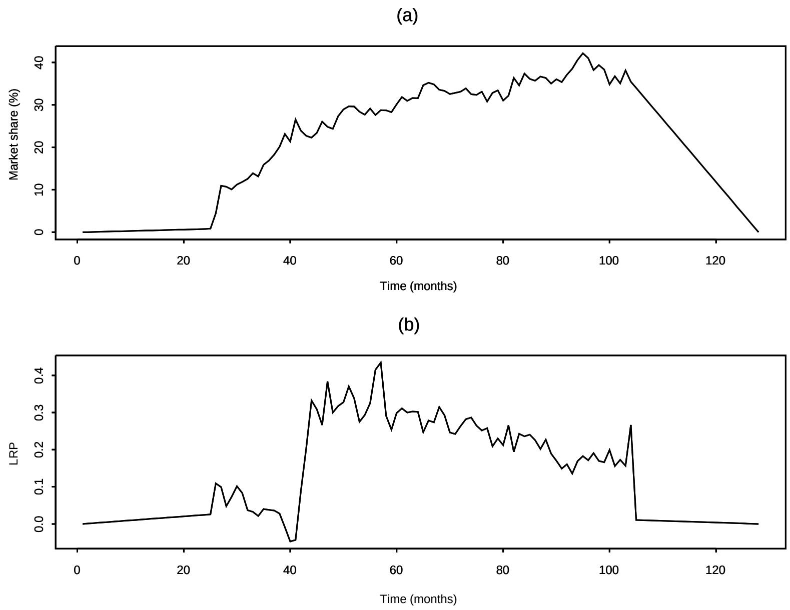

Figure 3.14: The tapered data for product two.

We perform the modelling procedures described in Section 3.5.2 to fit the model:

$$\text{market share} = \mathcal{W}_3 + \mathcal{W}_{17} + \mathcal{W}_{19} + \varepsilon, \quad \varepsilon \sim \mathrm{N}[0, \sigma^2].$$

	Value	Std. Error	t-value	Pr($>|t|$)
Intercept	19.96	1.08	18.42	0.000
$\mathcal{W}_3$	47.65	5.14	9.27	0.000
$\mathcal{W}_{17}$	47.88	7.00	6.84	0.000
$\mathcal{W}_{19}$	87.96	14.67	6.00	0.000

Table 3.7: Regression coefficients for the tapered product two model.

Wavelet Packet	Packet Scale	Packet Freq.
$\mathcal{W}_3$	2	0
$\mathcal{W}_{17}$	1	11
$\mathcal{W}_{19}$	1	2

Table 3.8: Wavelet packets represented in the tapered product two model.

The coefficients and t-statistics for this model are given in Table 3.4. The model now explains 86% of the total variation in the market share. The wavelet packets represented in this model are given in Table 3.8. The wavelet packet combination plot is given in Figure 3.15 (individual wavelet packets given in Figure 3.16). It is now much harder to reason a price elasticity interpretation for this model.

Figure 3.17 shows the fitted values compared to the actual values of tapered market share. The fitted values of market share are quite heavily smoothed and do not adequately represent the true underlying trends. We conclude that tapering should not be used in this example.

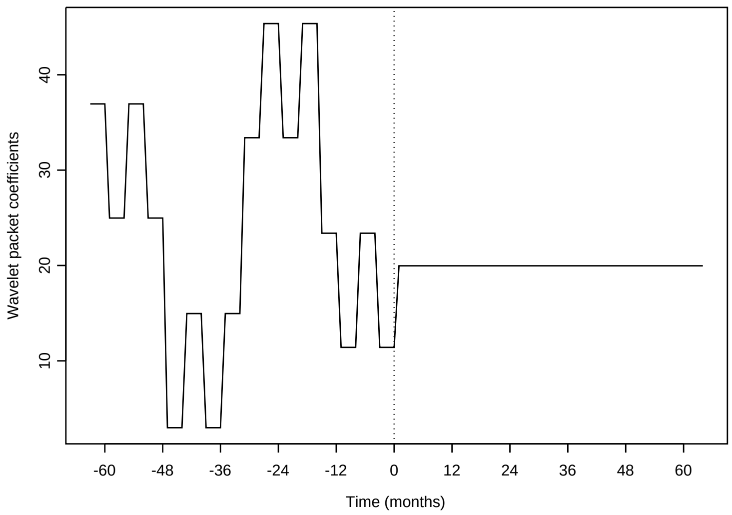

Figure 3.15: Wavelet packet combination plot of the tapered product two model. Note we have left the effect of the zeros in this plot (the actual length of the time series is 80, the graph above uses the time series padded with zeros).

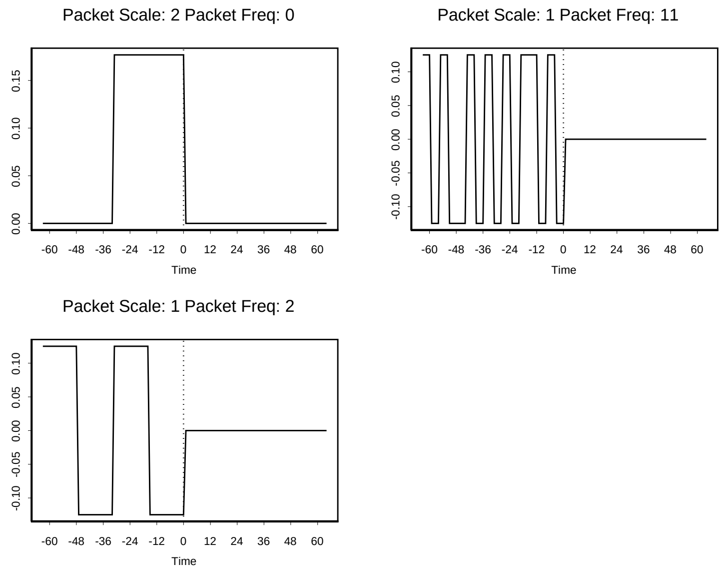

Figure 3.16: Wavelet packets in the tapered product two model.

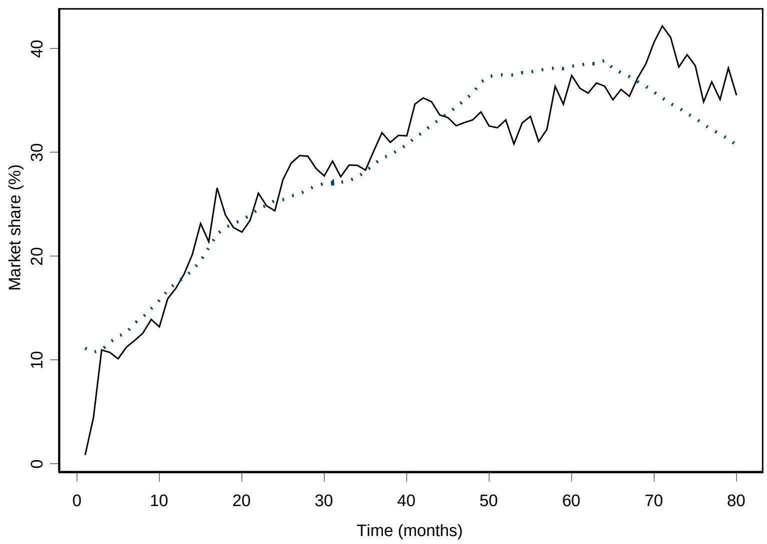

Figure 3.17: Actual and fitted values for the market share of the tapered product two model (dashed line = fitted values, solid line = actual data). The model performs well initially but deteriorates after 40 months.

3.5.4 Relationship between LRP and market share for product five

Following Section 3.5.2, we represent the market share of product five by the model:

$$\text{market share} = \mathcal{W}_9 + \mathcal{W}_{28} + \mathcal{W}_{51} + \varepsilon, \quad \varepsilon \sim \mathrm{N}[0, \sigma^2].$$

The coefficients and t-statistics for this model are given in Table 3.9. The wavelet packets represented in the model are shown in Table 3.10. The model explains 96% of the variation in the market share.

	Value	Std. Error	t-value	Pr($>|t|$)
Intercept	47.64	0.27	179.42	0.000
$\mathcal{W}_9$	23.52	4.86	4.84	0.000
$\mathcal{W}_{28}$	22.51	8.43	2.67	0.009
$\mathcal{W}_{51}$	-63.85	1.86	-34.33	0.000

Table 3.9: Regression coefficients for the product five model.

Wavelet Packet	Packet Scale	Packet Freq.
$\mathcal{W}_9$	1	3
$\mathcal{W}_{28}$	1	7
$\mathcal{W}_{51}$	1	1

Table 3.10: Wavelet packets represented in the product five model.

The correlations for the wavelet packet coefficients in this model are given in Table 3.11. We note a high correlation between $\mathcal{W}_{28}$ and $\mathcal{W}_{51}$.

	$\mathcal{W}_9$	$\mathcal{W}_{28}$	$\mathcal{W}_{51}$
$\mathcal{W}_9$	1	0.12	0.2
$\mathcal{W}_{28}$	-	1	-0.53
$\mathcal{W}_{51}$	-	-	1

Table 3.11: Correlation of wavelet packet coefficients included in the product five model.

The wavelet packet combination plot suggests an unusual split in the data (Figure 3.18, see plots of individual packets in Figure 3.19). At $t = 0$ there is a positive relationship between LRP and market share; this positive relationship (of varying magnitude) persists for a lag time of around 32 months. For longer lags we then see a negative relationship between LRP and market share. In earlier analyses a scatter plot of the LRP against market share had highlighted two clusters with these opposite relationships. Figure 3.20 displays this relationship in more detail and shows that there is initially a negative relationship which then changes (from around time point 30 onwards) into a positive relationship.

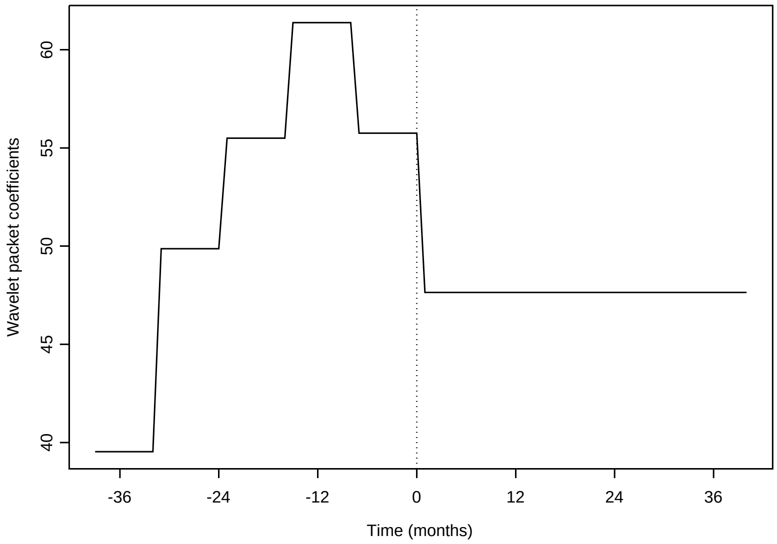

Figure 3.18: Wavelet packet combination plot of the product five model.

We construct the fitted values for the market share of product five and compare them to the actual values (see Figure 3.21). The model provides a good approximation to the trend in market share.

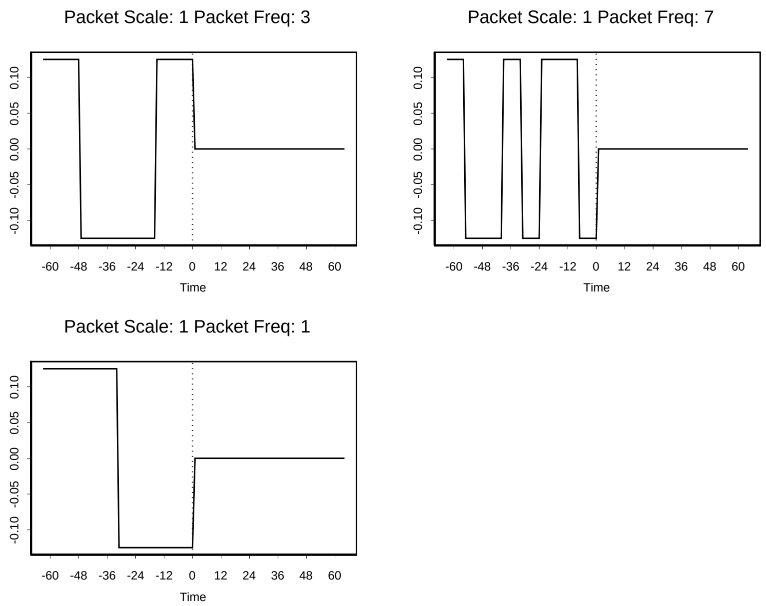

Figure 3.19: Wavelet packets in the product five model.

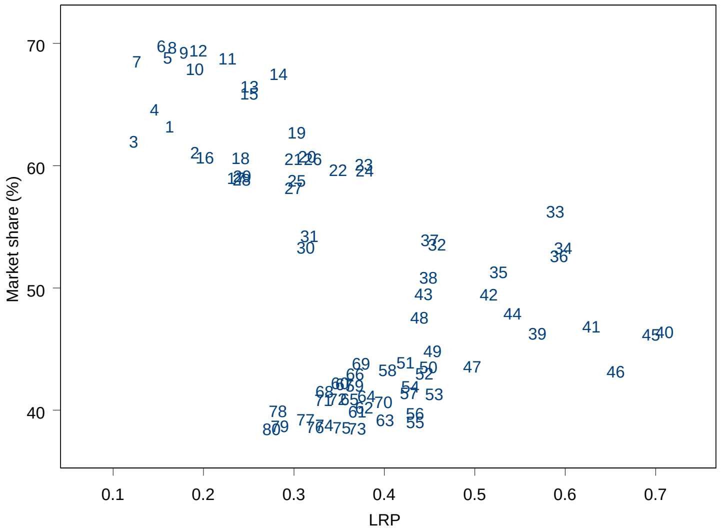

Figure 3.20: The changing relationship between LRP and market share for product five over time (this is the top middle graph in Figure 3.6). Note there is initially a negative relationship which from (approximately) time point 30 changes to a positive relationship (each number in the plot corresponds to the time ordering of the series).

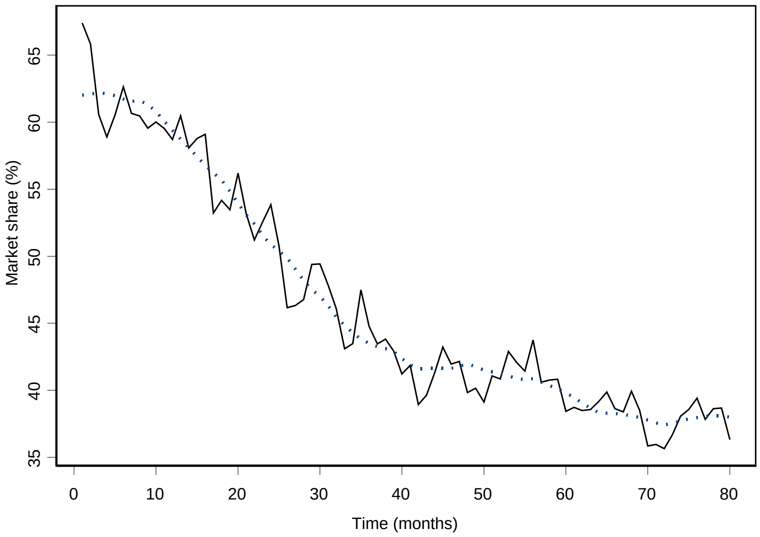

Figure 3.21: Actual and fitted values for the market share of product five model (dashed line = fitted values, solid line = actual data). We suggest the model is picking up the approximate trend in the data.

3.6 Bayesian dynamic linear modelling

In Section 2.5 we discuss the theory of the multiple regression dynamic linear model (RDLM). We will use the RDLM to model the relationship between the market share and LRP for product five. We define the RDLM as:

$$
\begin{aligned}
y_t &= \theta_t x_{t,1} + \nu_t, \qquad \nu_t \sim \mathrm{N}[0, V], \\
\theta_t &= \theta_{t-1} + \omega_t, \quad \omega_t \sim \mathrm{T}_{n_{t-1}}[0, W_t],
\end{aligned}
$$

where, the market share is y_t, the LRP is x_t, the regression coefficient for the LRP is θ_t and ν_t and ω_t are the error terms.

We define the prior information set (at $t = 0$) as follows: The trend has zero mean with a standard deviation of 1 and a discount factor of 0.9. The regression coefficient, θ_t, has zero mean, standard deviation of 0.5 and discount factor of 0.8. We assume a constant observational variance, where $V = 1$ on 1 degree of freedom. We specify W_t via the discounting approach (see Section 2.5.5). This is a vague prior information set which reflects our lack of prior knowledge of the series.

The online forecasts of market share are given in Figure 3.22. After initial uncertainty the RDLM provides accurate forecasts (the true values (y_t) are indicated by 'o' and the forecasts (f_t) by '×'). The standardised residuals are small but tend to be negative during some time periods. The retrospective forecasts of market share, given in Figure 3.23, are accurate.

Figure 3.24 is the retrospective estimate of θ_t. For the first 17 observations there is a positive relationship between LRP and market share. After observation 17 there is a negative relationship, which has increased in size until the last 9 observations. Figure 3.25 shows the observational variance V was variable for approximately the first 35 observations but has since shown a steady decline.

We conclude the RDLM has been useful in explaining the relationship between market share and LRP for product five at different time points in the data set.

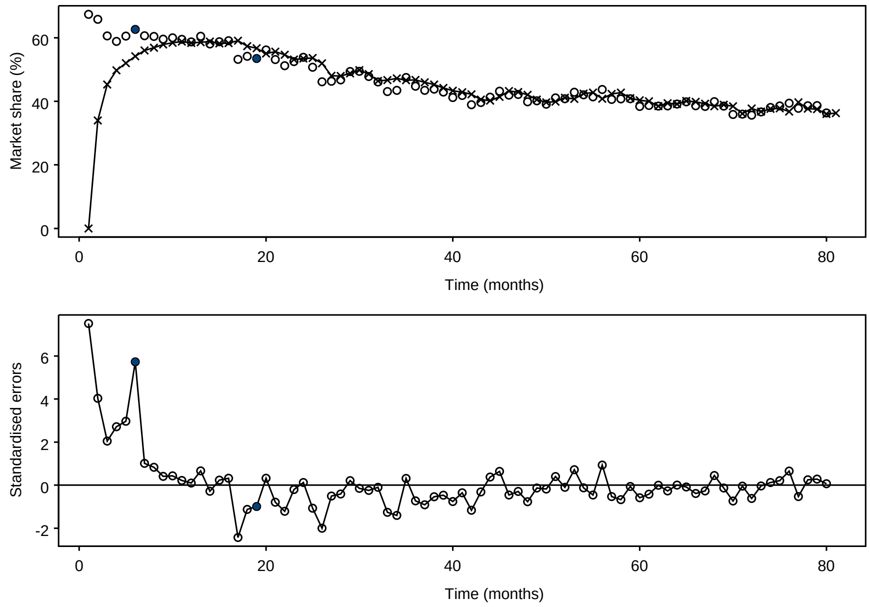

Figure 3.22: Online forecasting for product five, where 'o' denotes the true value of the response series y_t and '$\times$' denotes the online forecasts f_t (top graph). The model has provided accurate online forecasts.

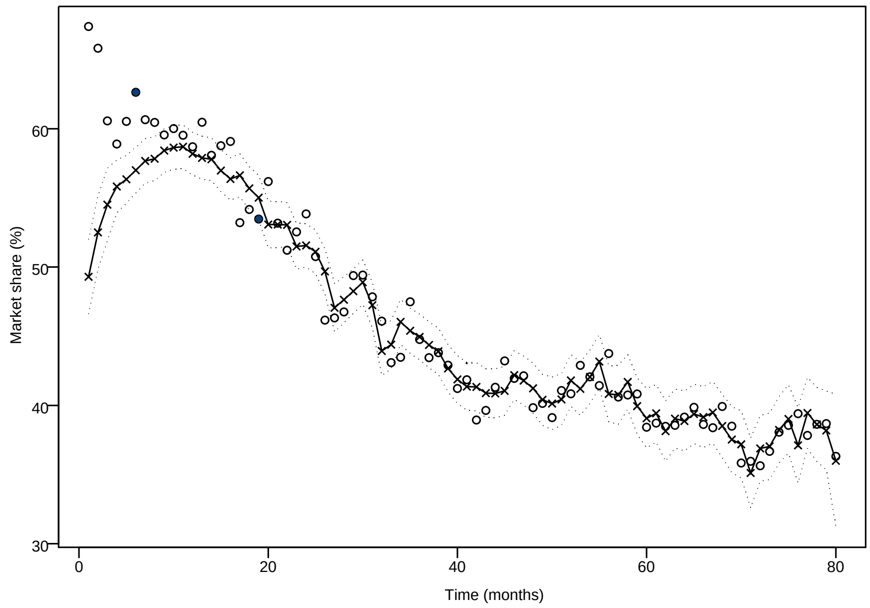

Figure 3.23: Retrospective analysis for product five, where 'o' denotes the true value of the response series y_t, '$\times$' denotes the retrospective forecasts f_t and '- - -' denotes the 2/3 confidence intervals. The RDLM has provided accurate retrospective forecasts.

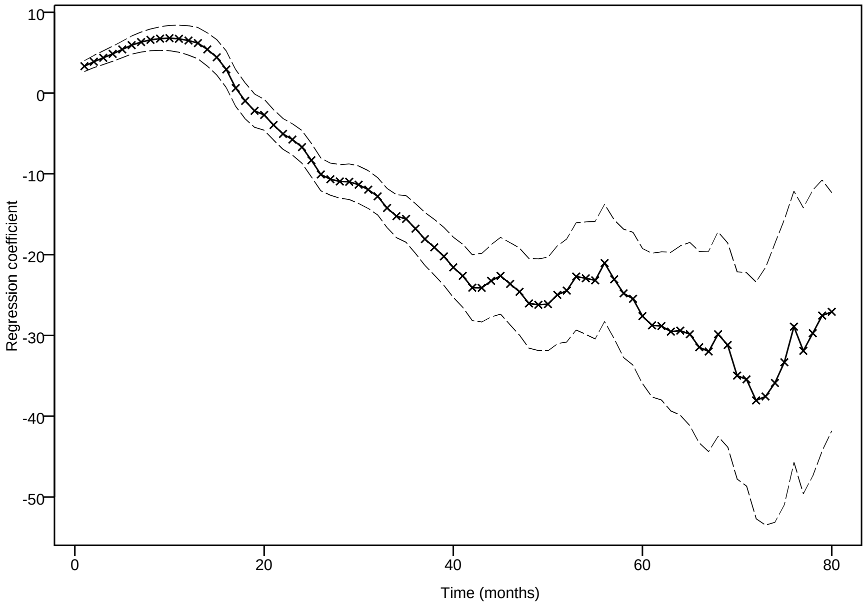

Figure 3.24: The retrospective estimate of θ_t for product five, which shows the evolution of the relationship between LRP and market share over time.

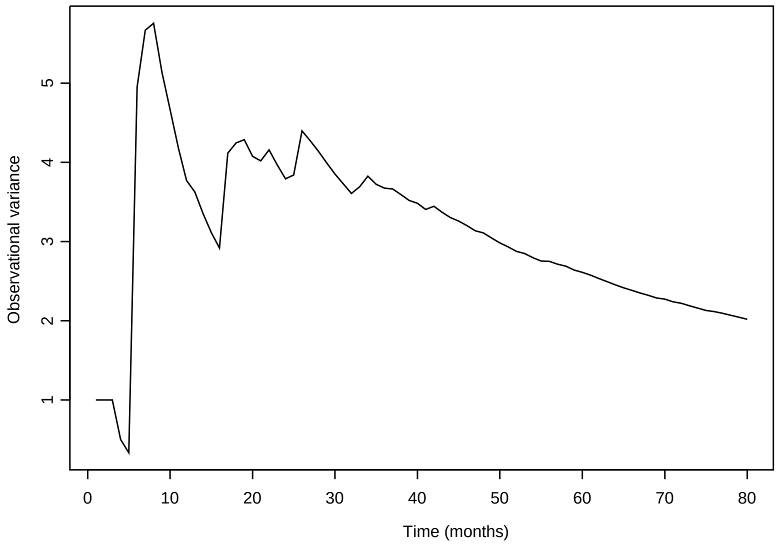

Figure 3.25: The observational variance V was initially inflated but is now decreasing steadily.

3.7 Conclusions and future work

Our aim was to analyse how the price of a product drives its market share. The data consisted of information about market share and unit price for six products. Initial analysis suggested the time series could be considered stationary (or near stationary) over certain time periods. In general, there appeared to be a negative relationship between LRP and market share, i.e. a high price corresponds with a low market share and vice versa. In addition, there appeared to be inter-relationships between products, in particular between products two and five, products one and six and products three and four.

We fitted univariate ARIMA models to the twelve time series and found they could generally be explained as white noise or simple ARIMA processes. Transfer function ARIMA models to relate market share to LRP were fit for products two and five. The model for product two suggested a negative relationship between market share and LRP. For product five we could not identify a relationship between market share and LRP. Fitting transfer function ARIMA models highlighted the problems of non-stationarity in the time series which could not be removed by simple transformations. For this reason we did not proceed to fitting multivariate transfer function models but considered alternative models to the ARIMA model.

We considered fitting models to explain the relationship between market share and LRP using the wavelet packet transfer function methodology of Nason and Sapatinas (2002); a method which allows for some non-stationarity in the data. We fitted wavelet packet models to products two and five. For product two, the wavelet packet combination plot suggests a negative relationship between market shares and LRPs close in time, which evolves into a positive relationship for time points further apart. For product five, the wavelet packet combination plot suggests a positive relationship between market shares and LRPs close in time, which evolves into a negative relationship for time points further apart. The wavelet packet models appeared to accurately model features evident in the data set; most notably, the changes in the relationship of LRP to market share for different time lags in the data set.

Finally, we applied the multiple regression dynamic linear model (RDLM) (see

Harrison and West, 1997) to forecast the market share of product five in terms of its LRP. The RDLM provided accurate online and retrospective forecasts of market share. The RDLM suggested there was initially a positive relationship between LRP and market share which evolved into a negative relationship at later time points. We conclude the RDLM may be useful for data sets of this type and should be investigated further.

Future work on this type of data should include investigating multivariate techniques. Earlier we noted that product two had eroded the market share of product five as when product five had increased its price product two had decreased its price in synchrony. At present, we have only looked at the cross-correlation function for product two in order to investigate feedback relationships between market share and LRP. We wish to investigate other possible feedback relationships between products. We did note some high correlations between explanatory variables in the model and should certainly consider including some methods to reduce possible multicollinearity problems.

Chapter 4

The NWPT Regression Model with PCA

4.1 Introduction

In this chapter we extend the non-decimated wavelet packet transfer function methodology of Nason and Sapatinas (2002), which aimed to predict values of a response time series from the time ordered non-decimated wavelet packet transform (NWPT) of an explanatory time series. In particular, we deal with the occasional high correlations occurring between wavelet packet coefficients and also the choice of distribution for the regression model. We propose the addition of a principal components analysis (PCA) to the model as a possible solution to the multicollinearity problem. Thus we illustrate a new model, 'NWPT regression model with PCA', which addresses some of the deficiencies of the existing model.

We present some preliminary practical results of the new method, using hourly (10 minute averaged) wind speed time series of length 512 from three pairs of locations in the British Isles. We provide simple descriptive statistics of the sites, a time series plot, a regression model and suggest a physical interpretation for each regression model. We compare our new method with the Nason and Sapatinas (2002) model and also a simple linear regression method, i.e. the response time series is regressed on the original explanatory time series, in terms of prediction errors. The simple linear regression method is commonly used in industry at present (Addison *et al.*, 2000). The data sets were obtained from the British Atmospheric Data Centre (BADC) which is the Natural

Environment Research Council's (NERC) designated data centre for the atmospheric sciences. Its role is to assist UK atmospheric researchers to locate, access and interpret atmospheric data.

In this chapter we will label the three models under consideration as follows; the non-decimated wavelet packet transfer function methodology of Nason and Sapatinas (2002) as 'NWPTorig', a simple linear regression model as 'INDUSTRY' and the new model we propose, 'NWPT regression model with PCA' as 'NWPTpca'. We conclude by suggesting further possible extensions to our model. See Appendix B for documentation of the functions which have been written for this chapter and Hunt and Nason (2001) for an article resulting from this chapter.

4.2 Wavelet packet transfer function modelling

4.2.1 Introduction

Nason and Sapatinas (2002) propose the use of the NWPT in modelling a response time series, y_t, in terms of an explanatory time series, x_t. A generalised linear model with gamma error and log link is used to model y_t in terms of the NWPT of x_t. We term this model the 'NWPTorig' model. The model, and the role of wavelet packet coefficients as explanatory variables, is explained in detail in Chapter 2. We now consider some extensions to this model.

4.2.2 Motivating questions

Questions raised from the NWPTorig model results included:

- What allowances should be made for high correlation between wavelet packet coefficients?
- Does a gamma distribution adequately represent the joint distribution of the wind speeds in this context?
- Can the modelling procedure be implemented for longer time series? Many wind regime studies recommend the use of much longer time series.

We will now consider the first two questions: How should we resolve the multicollinearity problem (Sections 4.2.3 and 4.2.4) and what is the appropriate joint distribution for the wind speeds considered (Section 4.2.5)?

4.2.3 Correlation between wavelet packet coefficients

For the NWPTorig model, after the non-decimated wavelet packet transform is performed on the explanatory time series, the number of variables (i.e. packet coefficients) is higher than the number of observations. The number of explanatory variables is reduced by selecting a proportion of those that are highly correlated with the response variable. In the original method no consideration is made of the sometimes high correlations between packet coefficients.

Consider the degree of correlation between Haar wavelet packet coefficients when the input time series is a white noise process. We compute a NWPT of a simulated white noise process of length 128 and select 5% of the packet coefficients that correlated highly with the original white noise process. We display the correlations between the wavelet packet coefficients in Figure 4.1. The histogram shows that some large correlations exist.

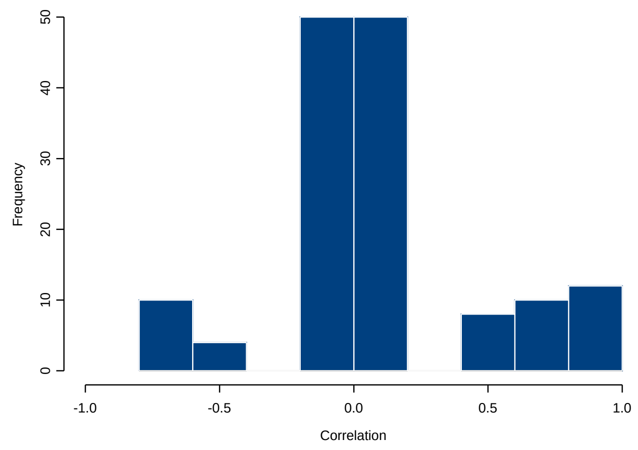

Figure 4.1: Correlations of the wavelet packet coefficients of a white noise process.

In a multiple regression model if the covariates are highly correlated the estimates of the regression coefficients can be very imprecise. As the number of explanatory variables is large, and the high degree of correlation between packet coefficients suggest the data may lie in fewer dimensions, we consider using a dimension reduction technique.

Various methods exist which we could use to make allowance for the high correlations between wavelet packet coefficients, for example projection methods or tree based methods. Tree based modelling is an exploratory technique for uncovering structure in data. When there is a single numeric response variable and a set of predictor variables the tree is called a regression tree. The model is fit using binary recursive partitioning, i.e. the data set is successively split such that it maximally distinguishes the response variable in the left and right branches. Splitting continues until it is unfeasible to continue (p. 377, Chambers and Hastie, 1992). We decided to use the projection method of PCA because it is a distribution free method and the principal components enable us to keep a physical interpretation to the models.

4.2.4 Principal components analysis

The aim of PCA is to find uncorrelated variables that are linear combinations of the original variables. The new variables are such that a few may explain most of the variation in the data, hence allowing us to reduce the number of dimensions. Components which only explain a small amount of variation can be discarded with, hopefully, little loss of information. The population principal components are defined as (p. 214, Mardia *et al.*, 1994):

Definition 4.1

If x is a random vector with mean μ and covariance matrix Σ, then the principal component transformation is the transformation

$$x \to y = \Gamma^{'}(x - \mu),$$

where Γ is orthogonal, $\Gamma^{'}\Sigma\Gamma = \Lambda$ is diagonal and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. The ith principal component of x may be defined as the ith element of the vector y, namely as

$$y_i = \gamma^{'}_{(i)}(x - \mu). \tag{4.1}$$

Here $\gamma_{(i)}$ is the ith column of Γ, and may be called the ith vector of principal component loadings. The rth element of y_i represents the score of the ith principal component on the rth observation.

Note how the eigenvalues (λ_i) are ordered such that the first principal component explains the largest variation in the data set, the second principal component explains the largest amount of variation not accounted for by the first principal component and so on. For full mathematical details refer to Mardia *et al.* (1994) or Krzanowski (1988).

As we are to use the principal components as predictors of a dependent variable we also consider their correlation with the dependent variable. Therefore, before we discard a small component we should make sure the correlation with the dependent variable is small and insignificant. Otherwise we shall be discarding information which is useful in predicting the dependent variable. So, in this case, it is not necessarily the components which explain the largest variance in the explanatory time series which are included in the model.

We also determine which of the loading coefficients are 'important' in defining the principal components we use. By removing small loading coefficients, that may not make a significant contribution to the component, we obtain a more parsimonious fit. This is a subjective decision and relates to the size of the loading coefficient in comparison to others in that component. For this reason a loading coefficient of the same size may be rejected in one component yet kept in another. We choose to include a variable in a component if its loading coefficient is greater than half the size of the largest loading coefficient in the component. See Krzanowski (1988) for further discussion on this arbitrary rule to accept or reject variables within components.

Since the principal components are uncorrelated there are no multicollinearities between them. However, if multicollinearities existed between the original variables then it appears as a principal component with a very small variance. If we include these principal components the large variances of the regression coefficients (caused by the multicollinearities) will not have been reduced. Principal components regression indicates whether a problem exists with respect to the removal of multicollinearity, i.e. whether instability in the regression coefficients can only be removed by simultaneously losing a substantial proportion of the predictability of the response variable (p. 138, Jolliffe, 1986).

We carry out PCA in S-Plus using the `princomp` function. We choose to centre the data and use the correlation matrix (rather then than the covariance matrix) in the calculations to standardise by scales as wavelet coefficients can vary greatly in size. For example, Haar father wavelet coefficients are calculated by smoothing the signal through the addition of consecutive coefficients (scaled by a constant), and so are typically large in size. Haar mother wavelets and Haar wavelet packets are smaller in size as they are calculated by a combination of addition and subtraction operators (scaled by a constant), and so typically have coefficients centred around zero at the coarser scales.

Further functions in S-Plus include the `summary` function which can be used to indicate the proportion of variance explained by the first k components (the sum of the first k eigenvalues divided by the sum of all the eigenvalues). The `plot` function can be used to plot the principal components scores and loadings. Plots of the principal components loadings and scores can show patterns, clusters or outliers which may help us to understand the structure of the data and, more importantly, highlight influential observations which may influence the number of dimensions chosen. (p. 63 and 218, Mardia *et al.*, 1994).

4.2.5 Choice of error distribution in the NWPT regression model

The data will determine which distribution we assume for the standard modelling procedures. We consider those distributions that are popular in modelling wind speed distributions, namely the normal, gamma and Weibull distributions. In the case study to follow we found that the normal distribution was the most appropriate. As we use the normal distribution we do not *have* to allow for 'calms' in our model (recall discussion of the work of Takle and Brown (1977) in Section 2.10.2) but we could model calms separately; even with the normal distribution.

4.3 The NWPT regression model with PCA

The previous section addressed some of the issues surrounding the NWPTorig model. We will extend the model by including a principal components analysis, which addresses the problem of occasional high correlation between explanatory variables in the model. We now explain the new model; NWPT regression model with PCA (NWPTpca).

As before, we perform a NWPT of the explanatory series and correlate the wavelet packet coefficients with the response time series to select a subset of wavelet packet coefficients. This gives a multivariate time series matrix, as described on p. 37. At this stage, PCA is now performed on the multivariate time series matrix. We model the response time series in terms of the principal components, i.e.

$$y = \text{pc } 1 + \cdots + \text{pc } m + \varepsilon, \quad \varepsilon \sim \mathrm{N}[0, \sigma^2], \tag{4.2}$$

where 'pc m' is the m^{th} principal component chosen for inclusion in the model. The principal components are ordered according to eigenvalues, for example pc 1 is the principal component which describes the largest amount of variation that can be accounted for in one variable.

We use stepwise subset selection for multiple regression to aid us when choosing which principal components to include in the regression model. The residual standard error and the multiple R-squared term (the fraction of the total variation in the response accounted for by the variation in the fitted values) is used to compare the fit of alternative models on the same data set. The multiple R-squared term ignores the number of coefficients in the fit so over-reliance may lead to over-fitting (p. 106, Chambers and Hastie, 1992).

4.4 Case study: Modelling and short term prediction of wind speeds

4.4.1 Introduction

We illustrate our model with the example of modelling and predicting wind speeds at a target wind farm site. The environmental impacts of conventional fossil fuel powered generators are well known and have led to increased interest in the use of sustainable sources of energy. In the UK these concerns are reflected by a Government plan to see 20% of Britain's energy needs produced from renewable sources by 2020 (Wind Directions (EWEA), May 2002.). We provide some preliminary results of our method using hourly wind speed data from three pairs of locations in the British Isles; their positions are indicated on the map in Appendix A. The data are recorded from 0.00 am on 1 January 1998 and observations are included in the analysis when there is a valid record made at both reference and target location. There are only a small number of missing observations in each example. In the Walney Island and Ronaldsway example some of the missing values are consecutive so we should consider their effect on the analysis. The time series are of length 512. We show how our model compares with the INDUSTRY and NWPTorig model in terms of prediction errors over 21 days in Section 4.6. To the regression model (Equation 4.2) a direction factor is added to take account of the variability in wind direction at the reference site. The factor records the direction bin of the wind as defined in Table 4.1. We use the Haar wavelet as for the NWPTorig model it produced interesting and effective results.

Direction Factor	Bearing (degrees)
1	0 - 29
2	30 - 59
3	60 - 89
4	90 - 119
5	120 - 149
6	150 - 179
7	180 - 209
8	210 - 239
9	240 - 269
10	270 - 299
11	300 - 329
12	330 - 359

Table 4.1: Coding of the wind direction as a direction factor.

4.4.2 St Bees Head and Valley

For our first example we use data from the St Bees Head meteorological station as the response time series and data from the Valley meteorological station as the explanatory time series. St Bees Head is a coastal location in Cumbria and Valley is a coastal location in Gwynedd. The sites are approximately 136 km (85 miles) apart on a northeast-southwest alignment and the correlation between the time series is 0.658. A plot of both time series is given in Figure 4.2. A summary of the wind speeds (metres per second) is as follows:

	Min.	1^{st} Qu.	median	mean	3^{rd} Qu.	Max.
St Bees Head	0.514	6.173	8.746	8.886	10.8	23.66
Valley	0.514	7.717	10.8	10.56	13.38	22.64

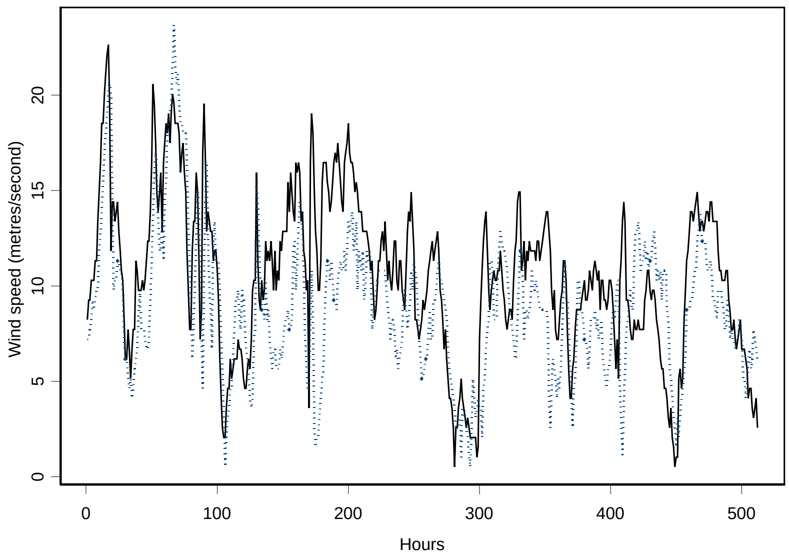

Figure 4.2: Time series plot of the St Bees Head (dashed line) and Valley (solid line) data.

We perform the modelling procedure described in Section 4.3 to fit the NWPTpca model:

$$\text{St Bees Head} = \text{pc } 1 + \text{pc } 5 + \text{direction factor}.$$

The coefficients in the model (with standard errors and t-statistics) are given in Table 4.2. The multiple R-squared statistic is 0.599, i.e. the model explains 60% of the variation in wind speeds at St Bees Head. The F-statistic is the ratio of the mean variation in the fitted values to the estimated variance; if there is no relationship between wind speeds at the two locations, this ratio has an F distribution with 13 and 498 degrees of freedom. In this case the ratio is 57.23 which is clearly significant and the true slope of the regression line is probably not zero. Diagnostic plots (residuals against fitted values, normal probability plot (NPP) in Figure 4.12(a), etc) show the model is a reasonable fit.

	Value	Std. Error	t-value	Pr(> $\|t\|$)
intercept	3.188	0.296	10.768	0.000
pc 1	0.521	0.026	20.047	0.000
pc 5	0.548	0.053	10.325	0.000
direction 1	-0.793	0.487	-1.628	0.104
direction 2	-1.126	0.279	-4.038	0.000
direction 3	0.052	0.216	0.239	0.812
direction 4	-0.240	0.349	-0.689	0.491
direction 5	-0.215	0.092	-2.333	0.020
direction 6	-0.159	0.071	-2.243	0.025
direction 7	-0.208	0.054	-3.835	0.000
direction 8	0.019	0.048	0.391	0.696
direction 9	-0.038	0.051	-0.755	0.450
direction 10	-0.054	0.050	-1.064	0.288
direction 11	-0.006	0.044	-0.139	0.889

Table 4.2: Regression coefficients for the St Bees Head and Valley model.

We have chosen to include the first and fifth principal component in the model. Both components are highly significant and the sample correlation between them is 0.06. The first principal component is a mixture of father wavelet and wavelet packet coefficients (see Table 4.3). Haar father wavelets average wind speeds over a given time, determined by the dilation index. Haar wavelet packets capture a wide variety of high and low frequency oscillations in the data, determined by both the dilation and frequency index. The fifth principal component is composed exclusively of mother wavelets (see Table 4.3). Haar mother wavelets capture a maximum of one oscillation over a given time, determined by the dilation index. The direction factor is significant when the wind is from the south/southwest (direction factor 5, 6 and 7) or northeast (direction factor 2). The southern direction factor may correspond to prevailing winds and the northeast/southwest direction factor may correspond to the alignment of the sites.

In Figures 4.3 and 4.4 we have plotted a wavelet packet combination plot of the principal components (see Section 2.9) We suggest an approximate physical interpretation for principal component 5 (Figure 4.4). In the first four hours (difference) there is a negative relationship between the two sites. As the sites are 136 km apart on a northeast-southwest alignment we would not initially expect to see the same behaviour at the two sites in the same time period. After a four hour lag there is

Packet Scale	Packet Freq.	Loading
6	0	0.24
7	0	0.24
5	0	0.22
8	0	0.24
3	3	-0.21
4	0	0.18
4	3	-0.20
3	5	-0.20
4	5	-0.19
2	5	-0.16
4	9	-0.19
2	9	-0.16
3	9	-0.18
2	6	0.15
3	17	-0.18
4	17	-0.16
1	17	-0.13
2	10	0.15
2	17	-0.13
1	9	-0.12
5	3	-0.14
2	33	-0.14
3	33	-0.15
1	65	-0.12
2	34	0.13
5	5	-0.14
2	18	0.13

Packet Scale	Packet Freq.	Loading
6	1	0.62
5	1	0.55
7	1	0.57

Table 4.3: Left: Wavelet packets represented in the first principal component of the St Bees Head and Valley model. Right: Wavelet packets represented in the fifth principal component of the St Bees Head and Valley model.

a gradual build up of the positive relationship between the two sites as wind speeds from the reference location reach the target site. This continues up until a 16 hour lag, by which time the weather system from Valley will have passed over St Bees Head. With a travelling time of 2 to 16 hours over 136 km this component corresponds to wind speeds of approximately 2.4 to 18.9 metres per second. In Figure 4.5 we can see that this reasonably incorporates most of the wind speed distribution at Valley, where dotted lines correspond to wind speeds of 2.4 and 18.9 metres per second.

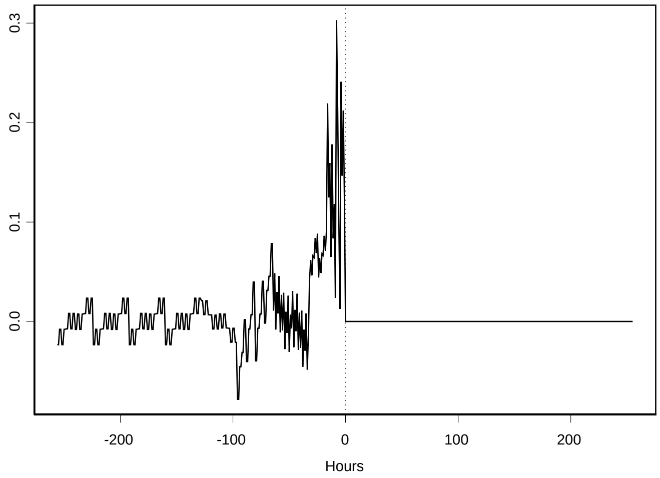

Figure 4.3: The first principal component from the St Bees Head and Valley model.

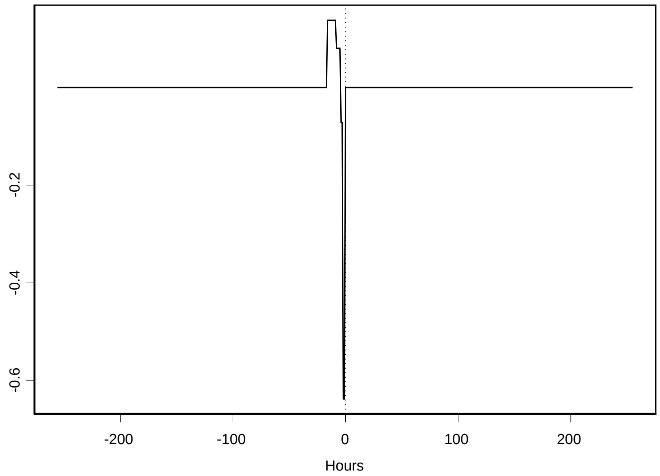

Figure 4.4: The fifth principal component from the St Bees Head and Valley model.

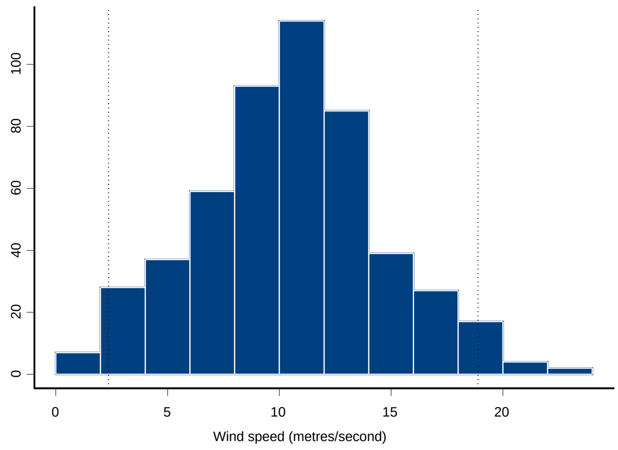

Figure 4.5: Histogram of the wind speeds at Valley. Dotted lines correspond to wind speeds of 2.4 and 18.9 metres per second. They represent the range of wind speeds highlighted as important in the fifth principal component from the regression model.

4.4.3 Milford Haven and Aberdaron

For our second example we use data from the Milford Haven meteorological station as the response time series and data from the Aberdaron meteorological station as the explanatory time series. Milford Haven is on the southwest coast of Wales and Aberdaron is on the northwest coast of Wales. Aberdaron is 136 km (85 miles) to the north of Milford Haven and the correlation between the time series is 0.762. A time series plot is given in Figure 4.6. A summary of the wind speeds (metres per second) is as follows:

	Min.	1^{st} Qu.	median	mean	3^{rd} Qu.	Max.
Milford Haven	0.514	6.688	8.746	9.169	11.32	24.69
Aberdaron	1.543	9.26	11.32	11.86	14.4	26.75

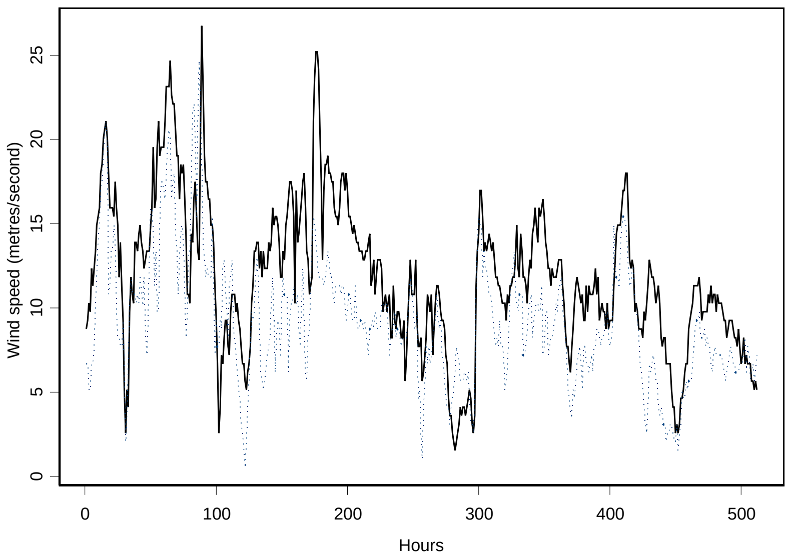

Figure 4.6: Time series plot of the Milford Haven (dashed line) and Aberdaron (solid line) data.

We fit the NWPTpca model:

$$\text{Milford Haven} = \text{pc } 10 + \text{pc } 41 + \text{direction factor}.$$

A small principal component, such as the forty-first principal component, can have a substantial correlation with the dependent variable (p. 135, Jolliffe (1986)). We chose to include this component in the model because its correlation with the response time series is large (0.76).

The coefficients in the model (with standard errors and t-statistics) are given in Table 4.4. The model explains 71% of variation in wind speeds at Milford Haven. The F-statistic is 95.91 on 13 and 498 degrees of freedom, which is significant so we conclude there is probably a relationship between the two sites. Diagnostic plots, including the NPP in Figure 4.12b, show the model is a reasonable fit.

	Value	Std. Error	t-value	Pr(> $\|t\|$)
intercept	-1.596	0.352	-4.539	0.000
pc 41	1.877	0.061	30.670	0.000
pc 10	1.390	0.320	4.344	0.000
direction 1	0.192	0.350	0.549	0.583
direction 2	1.462	0.232	6.304	0.000
direction 3	0.852	0.208	4.090	0.000
direction 4	-0.061	0.157	-0.386	0.670
direction 5	0.002	0.060	0.030	0.976
direction 6	-0.198	0.050	-3.933	0.000
direction 7	-0.397	0.043	-9.355	0.000
direction 8	-0.113	0.036	-3.114	0.002
direction 9	-0.098	0.046	-2.136	0.033
direction 10	0.051	0.040	1.281	0.201
direction 11	0.180	0.049	3.673	0.000

Table 4.4: Regression coefficients from the Milford Haven and Aberdaron model.

Both principal components are highly significant and the sample correlation between them is low (0.10). The tenth principal component is a mixture of wavelet packet coefficients (see Table 4.5) and the forty-first component is a mixture of father wavelet and wavelet packet coefficients (see Table 4.5). The direction factor is significant when the wind direction is from the northeast (direction factors 2 and 3, which we relate to site alignment) or from the southwest (direction factors 6 to 9, the direction of prevailing winds and site alignment). Direction factor 11 is also significant for which there is no immediately obvious reason.

It is difficult to suggest a clear physical interpretation for these components. The tenth principal component (Figure 4.7) is picking up more high frequency characteristics than those in the last data set. It also draws information from the future so in real-time prediction we would have to omit this packet (see Section 2.9.3). Principal component forty-one (Figure 4.8) highlights the importance of the first 100 hours in the transfer of winds between sites. This 100 hour, 4.2 day, period

Packet Scale	Packet Freq.	Loading
2	9	0.20
1	17	0.30
2	10	0.36
0	46	0.25
0	34	-0.24
2	34	0.21
0	75	0.37

Packet Scale	Packet Freq.	Loading
7	0	0.30
5	0	0.19
4	0	0.20
3	0	-0.17
3	5	0.21
3	9	-0.40
3	17	-0.31
3	3	0.25
2	33	0.21
2	17	0.28
2	18	-0.28
1	18	0.19
2	34	-0.23

Table 4.5: Left: Wavelet packets represented in the tenth principal component of the Milford Haven and Aberdaron model. Right: Wavelet packets represented in the forty-first principal component of the Milford Haven and Aberdaron model.

approximately corresponds with the typical four day transit period of fully developed weather systems, called the macrometerological peak; which was identified in a spectrum complied by Van der Hoven (1957). We also saw this feature in the first principal component for the St Bees Head and Valley model.

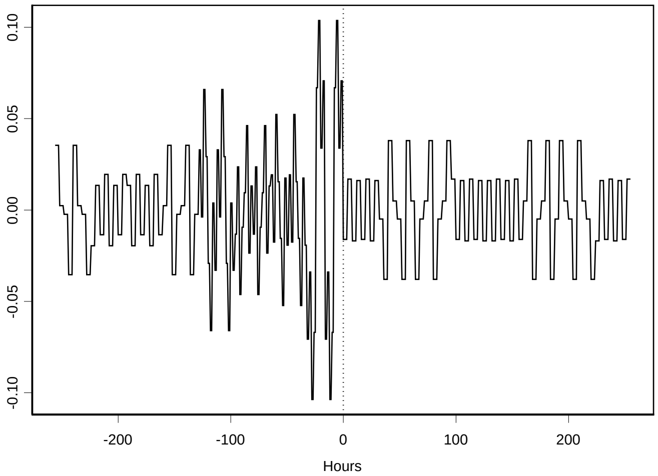

Figure 4.7: The tenth principal component from the Milford Haven and Aberdaron model.

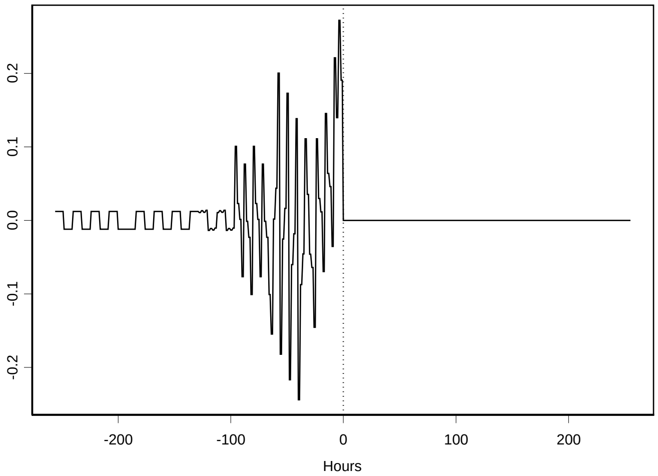

Figure 4.8: The forty-first principal component from the Milford Haven and Aberdaron model.

4.4.4 Walney Island and Ronaldsway

For our final example we take Walney Island on the northwest coast of England as the response time series and Ronaldsway on the southwest coast of the Isle of Man as the explanatory time series. The two meteorological stations are 64 km (40 miles) apart on an east-west alignment. The correlation between the two time series is 0.697 and a time series plot is given in Figure 4.9. A summary of the wind speeds (metres per second) is as follows:

	Min.	1^{st} Qu.	median	mean	3^{rd} Qu.	Max.
Walney Island	0.514	5.144	7.202	7.248	8.746	21.09
Ronaldsway	1.029	6.688	9.517	9.393	11.83	20.58

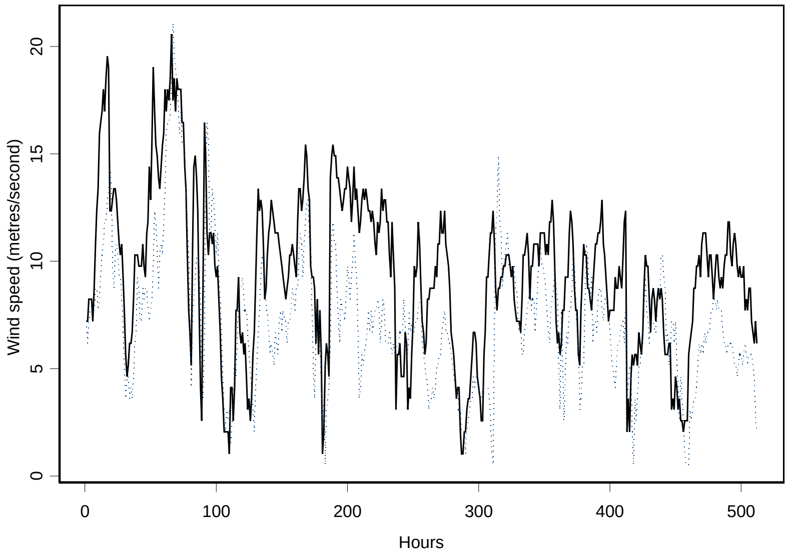

Figure 4.9: Time series plot of the Walney Island (dashed line) and Ronaldsway (solid line) data.

We fit the NWPTpca model:

$$\text{Walney Island} = \text{pc } 1 + \text{pc } 4 + \text{direction factor}.$$

The coefficients from the model (with standard errors and t-statistics) are given in Table 4.6. The model explains 73% of variation in wind speeds at Walney Island. The F-statistic is 104.4 on 13 and 498 degrees of freedom so we conclude there probably is a relationship between wind speeds at the two sites. Diagnostic plots, including the NPP in Figure 4.12c, show the model is a reasonable fit.

	Value	Std. Error	t-value	Pr(> $\|t\|$)
intercept	1.433	0.233	6.157	0.000
pc 1	0.543	0.020	27.129	0.000
pc 4	-0.477	0.038	-12.631	0.000
direction 1	-0.473	0.251	-1.884	0.060
direction 2	-0.429	0.334	-1.285	0.200
direction 3	0.230	0.194	1.188	0.235
direction 4	-0.019	0.112	-0.169	0.866
direction 5	-0.037	0.068	-0.540	0.590
direction 6	-0.314	0.043	-7.358	0.000
direction 7	-0.127	0.035	-3.627	0.000
direction 8	0.076	0.028	2.706	0.007
direction 9	0.261	0.033	7.914	0.000
direction 10	0.142	0.037	3.873	0.000
direction 11	-0.088	0.025	-3.551	0.000

Table 4.6: Regression coefficients from the Walney Island and Ronaldsway model.

Both principal components are highly significant and their sample correlation is -0.06. The first component represents a large range of wavelet and wavelet packets (see Table 4.7) whilst the fourth component represents just three mother wavelets and one coarse scale wavelet packet (see Table 4.7). Direction factors 6 to 11 are significant, corresponding to winds from a southern to a northwestern direction, encompassing both the prevailing winds and site alignment direction.

The first and fourth principal components are plotted in Figures 4.10 and 4.11 respectively. Principal component one picks up the familiar general pattern of high frequency positive behaviour between the sites decreasing to small fluctuations around zero for longer time lags. In this sense it is similar to the first principal component in the St Bees Head and Valley model, although a key difference is the component includes future observations. Principal component four picks up a large initial positive relationship between the sites indicating that the behaviour at the reference site is well matched by the behaviour at the target location. A short lag time later this decreases to a small negative relationship before there is no further relationship between the two time series. As both sites are on a similar latitude and, at only 64 km apart across sea, we suggest they may experience similar weather patterns in synchrony, hence the large positive spike.

Packet Scale	Packet Freq.	Loading
6	0	0.24
7	0	0.24
5	0	0.22
8	0	0.24
4	0	0.18
4	3	-0.18
2	9	-0.17
3	5	-0.16
4	5	-0.18
3	0	0.12
3	9	-0.16
3	3	-0.16
1	18	0.14
4	9	-0.17
2	17	-0.16
2	33	-0.17
3	17	-0.16
4	17	-0.16
5	3	-0.14
2	65	-0.15
2	5	-0.16
0	5	-0.13
3	33	-0.13
1	34	0.15
1	66	0.15
0	68	0.12
5	5	-0.13
1	130	0.13
5	9	-0.13
1	10	0.14

Packet Scale	Packet Freq.	Loading
6	1	-0.60
0	2	0.20
7	1	-0.55
5	1	-0.55

Table 4.7: Left; Wavelet packets represented in the first principal component of the Walney Island and Ronaldsway model. Right; Wavelet packets represented in the fourth principal component of the Walney Island and Ronaldsway model.

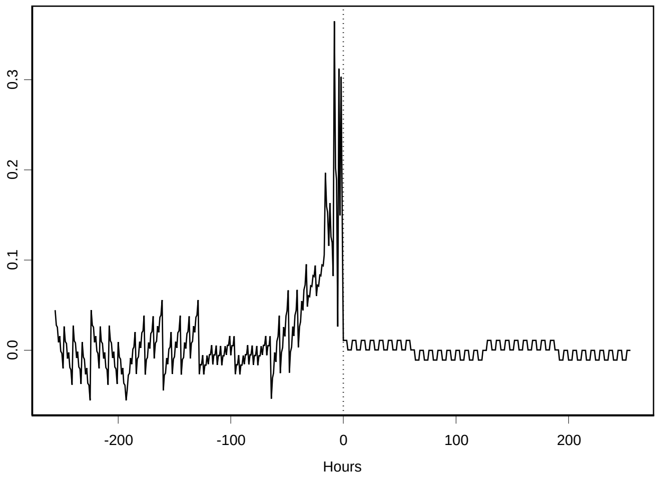

Figure 4.10: The first principal component from the Walney Island and Ronaldsway model.

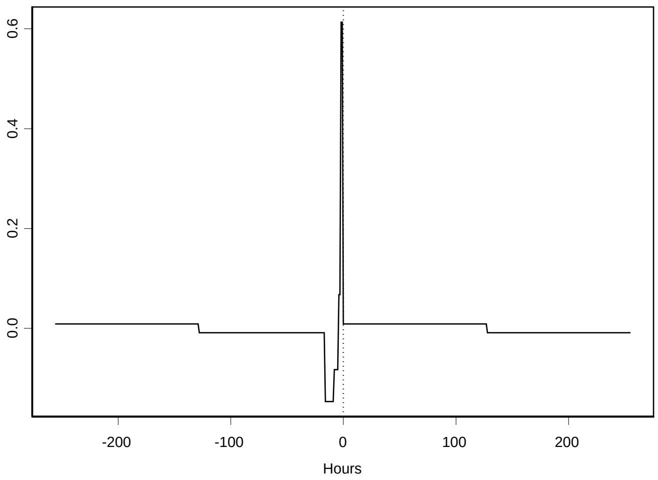

Figure 4.11: The fourth principal component from the Walney Island and Ronaldsway model.

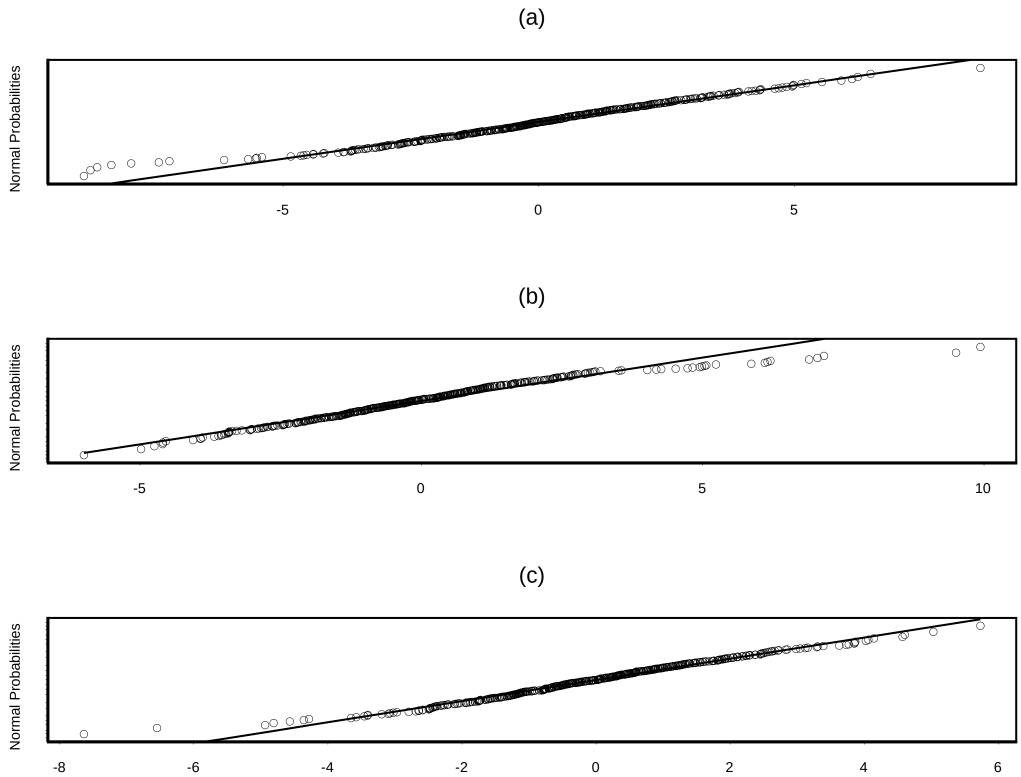

Figure 4.12: Normal probability plots: (a) St Bees Head and Valley model, (b) Milford Haven and Aberdaron model and (c) Walney Island and Ronaldsway model.

4.5 The NWPTorig model

4.5.1 Introduction

We now follow the procedure documented in Section 2.8.2 to fit the model proposed by Nason and Sapatinas (2002) for each of the previous data sets. For comparative purposes we assume a normal error distribution for the regression model. We include wavelet packet coefficients in the regression model if they lead to a decrease in the MSE of the predictions. This is contrary to our aim in fitting the NWPTpca model, where we fit the model to capture key features in the training set. We provide only brief details of the models here.

4.5.2 St Bees Head and Valley

We fit the NWPTorig model:

$$\text{St Bees Head} = \mathcal{W}_1 + \mathcal{W}_{28} + \mathcal{W}_{33} + \mathcal{W}_{30} + \text{direction factor},$$

where, $\mathcal{W}_1$ represents a father wavelet at scale 6, $\mathcal{W}_{28}$ represents a father wavelet at scale 3, $\mathcal{W}_{33}$ represents a wavelet packet at scale 0 with frequency index 30 and $\mathcal{W}_{30}$ represents a wavelet packet at scale 4 with frequency index 15. The wavelet packets are shown in Figure 4.13. The model explains 48% of the variation in the response time series. The correlations between wavelet packet coefficients are:

	$\mathcal{W}_1$	$\mathcal{W}_{28}$	$\mathcal{W}_{33}$	$\mathcal{W}_{30}$
$\mathcal{W}_1$	1	0.61	0.22	-0.28
$\mathcal{W}_{28}$	-	1	0.06	0.13
$\mathcal{W}_{33}$	-	-	1	-0.75
$\mathcal{W}_{30}$	-	-	-	1

4.5.3 Milford Haven and Aberdaron

We fit the NWPTorig model:

$$\text{Milford Haven} = \mathcal{W}_6 + \mathcal{W}_{42} + \mathcal{W}_{24} + \text{direction factor},$$

where, $\mathcal{W}_6$ represents a father wavelet at scale 3, $\mathcal{W}_{42}$ represents a wavelet packet at scale 1 with frequency index 23 and $\mathcal{W}_{24}$ represents a wavelet packet at scale 0

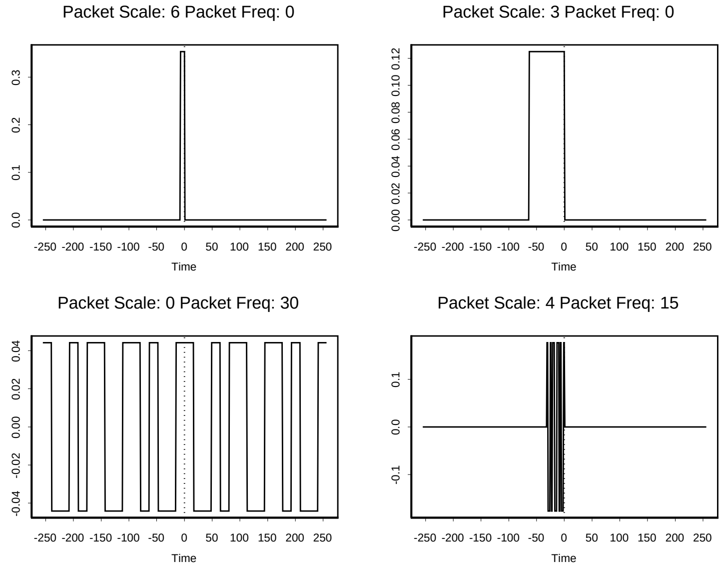

Figure 4.13: Wavelet packets in the St Bees Head and Valley model.

with frequency index 257. The wavelet packets are shown in Figure 4.14. The model explains 42% of the variation in the response time series. The correlations between wavelet packet coefficients are:

	$\mathcal{W}_6$	$\mathcal{W}_{42}$	$\mathcal{W}_{24}$
$\mathcal{W}_6$	1	0.27	-0.75
$\mathcal{W}_{42}$	-	1	0.00
$\mathcal{W}_{24}$	-	-	1

4.5.4 Walney Island and Ronaldsway

We fit the NWPTorig model:

$$\text{Walney Island} = \mathcal{W}_{14} + \mathcal{W}_{51} + \text{direction factor},$$

where, $\mathcal{W}_{14}$ represents a father wavelet at scale 2 and $\mathcal{W}_{51}$ represents a wavelet packet at scale 2 with frequency index 6. The wavelet packets are shown in Figure 4.15. The model explains 39% of the variation in the response time series. The correlation between the explanatory variables is -0.08.

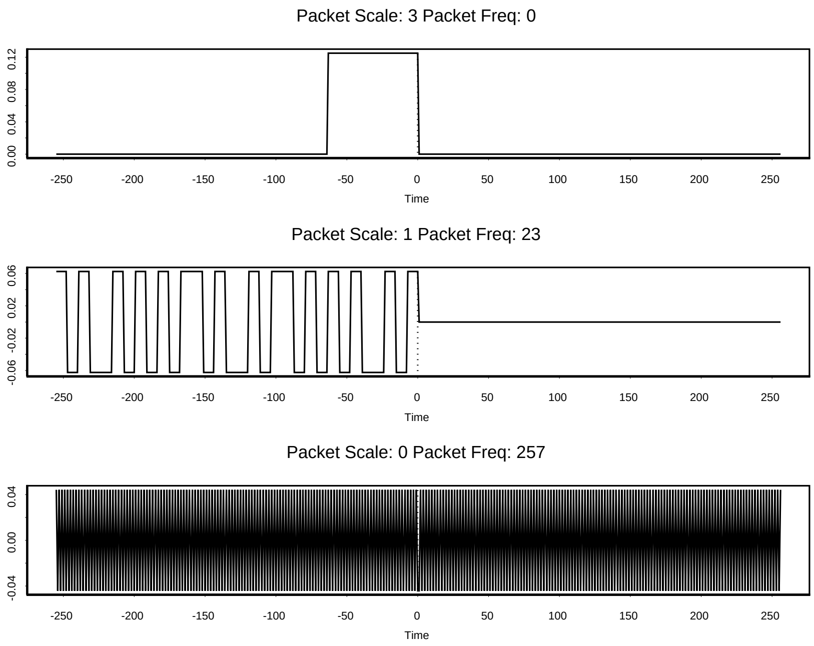

Figure 4.14: Wavelet packets in the Milford Haven and Aberdaron model.

4.5.5 The NWPTorig model: Summary

Each NWPTorig model represents at least one father wavelet and one wavelet packet. Mother wavelets were absent from the models. We note high correlations between explanatory variables for both the St Bees Head and Milford Haven models, which also contain wavelet packet coefficients that draw information from the future.

4.6 Evaluation of the model: Predicting future wind speeds

We now compare our method with the INDUSTRY and NWPTorig model by seeing how the models perform when predicting future values of the response series y_t from future values of the explanatory time series x_t. We compare predictions over three time periods; five, ten and twenty-one days, to provide a measure of how the models perform over the short-term.

We measure the accuracy of our method by looking at the mean squared error (MSE), $\sum e_t^2/n$, of the predictions of future values of y_t, denoted $\hat{y}_t$, given future

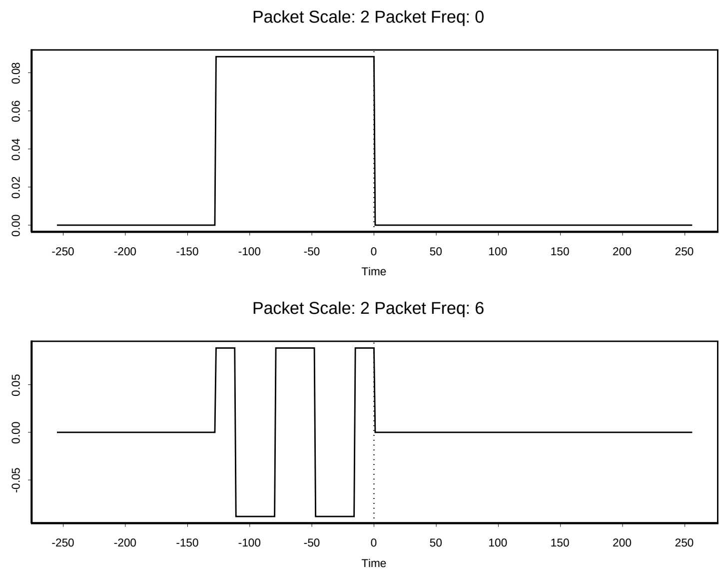

Figure 4.15: Wavelet packets in the Walney Island and Ronaldsway model.

values of x_t, where $e_t = y_t - \hat{y}_t$, $t = 1, \ldots, n$. The models were fitted using a training data set of 512 observations and a consecutive data run of 512 observations is used for the prediction. For further analysis we provide graphs of the predictions from the INDUSTRY and NWPTpca models.

There is some dependence in the NWPT predictions due to the mean correction of the series used in the prediction process (see Nason *et al.*, 2001). We adjust the mean of the prediction time series to be the same as the mean of the training data set. If we had used a shorter time series for prediction, say 128, the mean correction of the prediction time series would be slightly different.

It can be seen from Table 4.8 that the NWPTpca model, in comparison with the INDUSTRY model, performs significantly better over the early prediction period. The predictions for the longer prediction period (where we estimate 504 observations) may be unreliable due to only a small stretch of data (512 observations) being used to formulate the models. Wind power output is related to the cube of the speed so even a small improvement in the prediction of the wind speeds may make a significant difference to the estimated power output of a site (p. 15, Haslett and Raftery, 1989).

The NWPTorig model compares favourably against both the NWPTpca model and the INDUSTRY model, particularly over the shorter prediction period. For two out of

Location	Time Period	NWPTpca	INDUSTRY	NWPTorig
St Bees Head	5 days	6.20	8.08	4.60
and	10 days	5.40	6.89	4.74
Valley	21 days	5.46	6.48	5.37
Milford Haven	5 days	4.80	16.69	1.39
and	10 days	4.36	12.74	2.22
Aberdaron	21 days	6.95	9.09	4.85
Walney Island	5 days	3.29	7.00	1.70
and	10 days	3.59	6.57	3.73
Ronaldsway	21 days	4.39	5.51	5.01

Table 4.8: Mean Squared Error (MSE) of the predictions.

three models the NWPTorig model provides a lower MSE for each prediction period. For the Walney Island example, the NWPTpca model provides better predictions over 10 and 21 days.

We believe the difference in MSE between the NWPTpca and NWPTorig predictions is due to the model fitting strategy. For the NWPTpca model our aim was to explain key features in the data set. On average the NWPTpca models explained 68% of variation in the response time series. This is in contrast to our strategy for the NWPTorig model, where we choose the regression equation according to its prediction abilities. The NWPTorig models explained on average 43% of the variation in the response series. To test this belief, we construct a NWPTpca model for the St Bees Head and Valley data set, where the strategy is to optimise the MSE of the predictions. We chose to include three principal components in the regression equation, none of which had been present in the previous model. The MSE of the predictions for time periods of 5, 10 and 21 days are 4.05, 4.59 and 5.32 respectively. These are lower than the MSE of predictions from the NWPTorig model.

The number of parameters in the models are given in Table 4.9. We have included all the direction factors in the total number of parameters, although not all directions are significant.

Figure 4.16 provides a graphical representation of the predictions from the INDUSTRY model and the NWPTpca model for St Bees Head and Valley. We can see that the models performed fairly similarly but that occasionally the INDUSTRY model tended to underestimate wind speeds. As shown in Figure 4.17 the INDUSTRY model

Response Location	NWPTpca	Trans. NWPTpca	INDUSTRY	NWPTorig
St Bees Head	14	42	24	16
Milford Haven	14	32	24	15
Walney Island	14	46	24	14

Table 4.9: Number of parameters in the regression models. Trans. NWPTpca refers to the number of parameters if we expressed the NWPTpca model in terms of the original variables, i.e. wavelet packet coefficients.

for Milford Haven and Aberdaron consistently overestimates the wind speeds until the last 100 observations. The NWPTpca model initially followed the time series but also tended to overestimate the wind speeds for later observations. Both of these models probably overestimated because of a decrease in wind speeds during the prediction period that was different to the behaviour evident in the training data set. Figure 4.18 is the graphical representation of the predictions for the Walney Island and Ronaldsway model. Both models perform in a broadly similar manner, but the INDUSTRY model tends to overestimate more in the first part of the time series.

We believe these preliminary results suggest our NWPTpca model is worth investigating as an alternative method for the short term prediction of wind speeds.

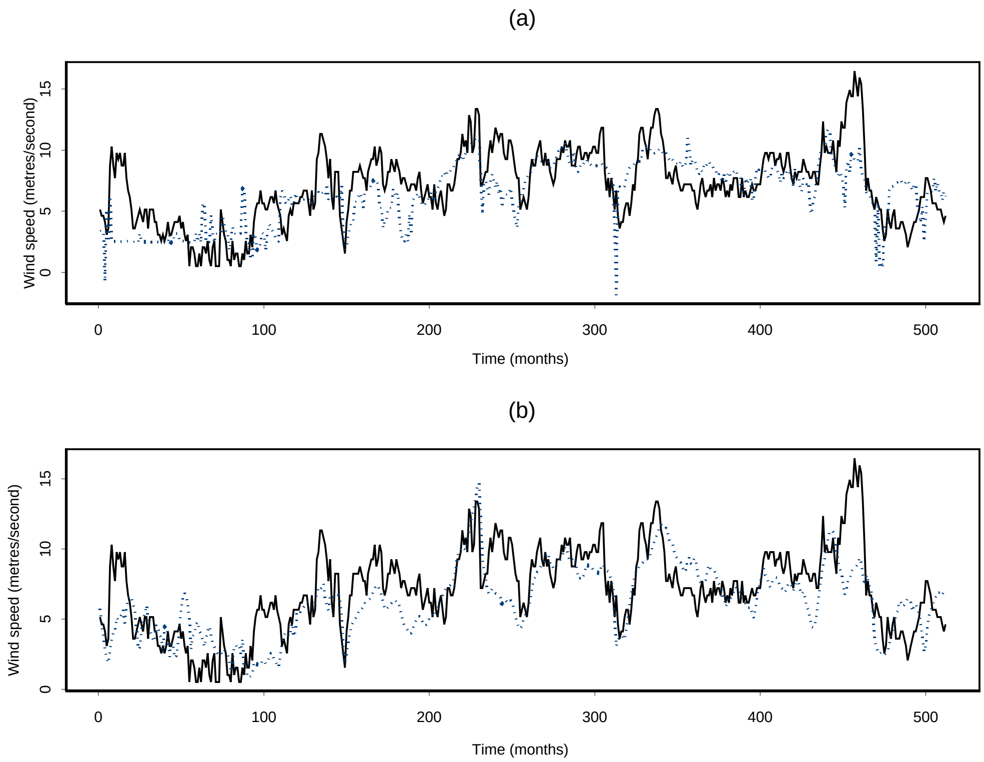

Figure 4.16: Predictions for the St Bees Head and Valley models: (a) compares how the INDUSTRY model performs and (b) compares how well the NWPTpca model performs. In each graph the solid line represents the true wind speeds of the response time series.

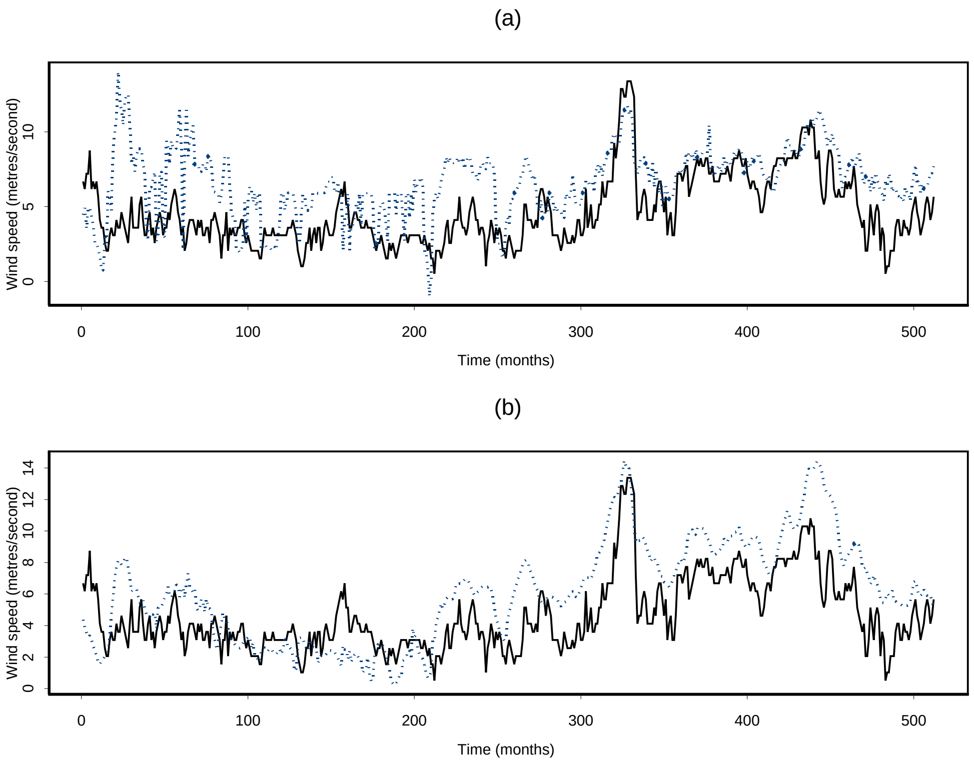

Figure 4.17: Predictions for the Milford Haven and Aberdaron models: (a) compares how the INDUSTRY model performs and (b) compares how well the NWPTpca model performs. In each graph the solid line represents the true wind speeds of the response time series.

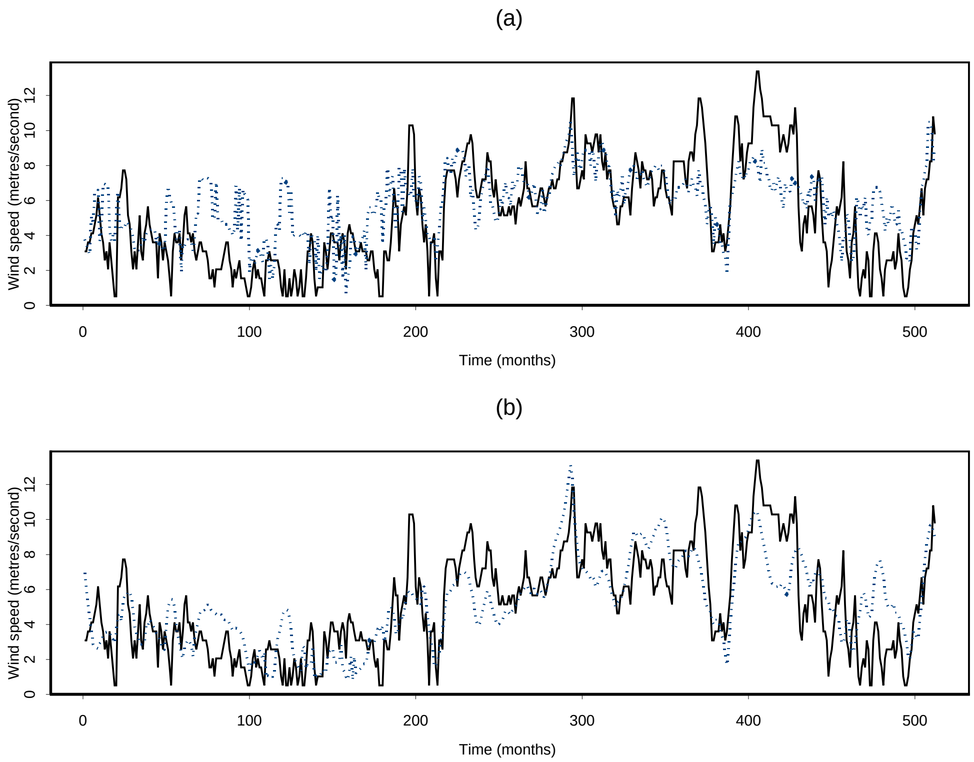

Figure 4.18: Predictions for the Walney Island and Ronaldsway models: (a) compares how the INDUSTRY model performs and (b) compares how well the NWPTpca model performs. In each graph the solid line represents the true wind speeds of the response time series.

4.7 Conclusions and future work

In this chapter we have proposed extensions to the wavelet packet transfer function methodology of Nason and Sapatinas (2002) (the NWPTorig model). The NWPTorig model took a NWPT of an explanatory time series and used the generated wavelet packet coefficients as explanatory variables in a regression model, in order to predict values of a response time series. There were several deficiencies in the NWPTorig model, including the occasional high correlation between wavelet packet coefficients. The new model, 'NWPT regression model with PCA' (the NWPTpca model), has sought to address some of the deficiencies. The introduction of a principal components analysis in the model has been useful in terms of reducing the correlation between regression parameters. The principal components have provided a detailed physical interpretation for the models, suggesting complex interactions in weather systems between sites. In comparison with the INDUSTRY model, the NWPTpca model achieves an improvement in the MSE of predictions. The NWPTpca model compared less favourably with the NWPTorig model in terms of the MSE of predictions, but we suggest this may be partly explained by our different strategies for fitting the models (for which we provide brief justification).

We believe the NWPTpca model provides a good alternative method for the prediction of a response time series given an explanatory time series. We appreciate that we have to continue this work for much longer time periods than 21 days if this model is to be considered further for predicting wind speeds. However the computational complexity of fitting the model with, for example, one years' observations is high. To consider longer time series ($n > 1500$) we could initially compute coefficients of individual packets by direct filtering rather than the 'fast' algorithm proposed by Pesquet, Krim and Carfantan (1996) (which simultaneously computes all). For a given packet we could compute a goodness of fit parameter, for example the correlation between the packet and the response time series. This would help us to identify which wavelet packet coefficients may be most useful for predicting the response time series and, in future, we may only need to calculate a specific part of the wavelet packet tree (illustrated on p. 30), hence reducing computational complexity.

We would like to develop the initial variable selection techniques, where a selection of the wavelet packet coefficients are chosen through their correlation with the response times series. In addition, we could also take a NWPT of the response time series and investigate relating the transforms of both series. A further extension is the inclusion of more than one explanatory time series in the model. The extension to incorporating two explanatory time series has since been successfully implemented by Barber (personal communication, 2001) to analyse the patterns of behaviour in different muscles involved in repetitive tasks.

Chapter 5

Bayesian Dynamic Linear Modelling using Wavelets

5.1 Introduction

In this chapter we investigate a role that wavelets may play in Bayesian dynamic linear modelling. We propose using a multiple regression dynamic linear model (RDLM) (see Harrison and West, 1997) to model a response time series y_t not in terms of the explanatory time series $x_{1t}, x_{2t}, \ldots$ but utilising the non-decimated wavelet packet coefficients of $x_{1t}, x_{2t}, \ldots$ The inclusion of wavelet packet coefficients in the RDLM is used here as a wavelet 'preprocessing' step. In this initial investigation we restrict attention to one explanatory time series.

Intervention in the RDLM may be necessary when there is a change in the underlying nature of the model. For example, the component time series may be highly non-stationary or the relationship between series may change rapidly. We suggest that the inclusion of a wavelet preprocessing step may be a suitable alternative to intervention when the change in the underlying nature of the model cannot be anticipated. As wavelets are well-known for efficiently representing discontinuities in signals, we suggest they may provide a framework for analysing certain types of time series behaviour. Our aim here is not to extend the RDLM theory but rather show the possible benefits of combining wavelet and RDLM methods.

We provide two data sets to compare how the RDLM performs using wavelet time-scale variables over standard variables. Our first example is a simulated study to assess the performance of the RDLM when we model a step response series in terms of an explanatory time series. The discontinuities in the response function result from the rapid evolution of the regression coefficients. Our second example is taken from Harrison and Reed (1999) and aims to predict the number of cases of BSE in cows from an explanatory variable. In this example the component time series are non-stationary.

For each data set we provide graphs of online predictions f_t, retrospective forecasts f_t^R and regression coefficients θ_t. We provide a brief residual analysis, which includes the mean absolute deviation and mean square error of residuals and a normal probability plot. We conclude by discussing the limitations of our proposed model and suggest areas for future work.

This chapter assumes that the reader is familiar with the work of Harrison and West (1997), which we review in Chapter 2.

5.2 The wavelet packet preprocessing step

5.2.1 Introduction

For data sets such as those with non-stationarity in the component time series, the main option for improving model fit is to introduce an intervention strategy. This may involve including an additional parameter in the model to allow for a mean shift, increasing the observational variance V to reflect the uncertainty or simply ignoring observations which are not representative of underlying trends. An example of a time series requiring intervention is given in West and Harrison (1997): the time series records the monthly total sales of a tobacco product in the UK and is piecewise linear due to abrupt changes caused by factors such as company announcements that prices are to rise. It is possible to anticipate some changes, for example those caused by legislation in the Government Budget. In general, we can classify information into three main types; 'weak form' which consist of historical information, 'semi-strong form' which consist of publicly available information and 'strong form' where only certain groups have monopolistic access to the relevant information (Fama, 1969). We believe the method we propose may be useful when information is limited; we do not attempt to forecast the jump but rather provide a way for the RDLM to adapt quickly to changes.

5.2.2 The method

As an alternative to intervention we wish to improve the fit of the RDLM by taking a time ordered non-decimated wavelet packet transform (NWPT) of the explanatory time series and including the resulting wavelet packet coefficients as explanatory variables in the model in place of the original explanatory time series.

To determine which wavelet packet coefficients to include in the RDLM we fit a normal linear regression model to the response series, using the wavelet packet coefficients as explanatory variables. We do not include all the wavelet packet coefficients (recall for a time series of length N there are 2N-2 packets), only those highlighted by a stepwise analysis. We exclude explanatory variables on the basis of their t-value, until all remaining variables are significant. This is a rather naive method of variable reduction, but has proved adequate for this initial investigation. The RDLM is then fit using the Harrison and Reed (1999) software using the new NWPT variables.

We will refer to the RDLM which uses wavelet packet coefficients as explanatory variables as 'RDLMwp'. It is hoped the RDLMwp will be more efficient at representing certain types of behaviour; such as non-stationary behaviour in the component time series or rapidly evolving regression coefficients. If required, it is possible to include both wavelet and ordinary variables in the RDLM.

5.3 Comparison of the RDLM and RDLMwp

5.3.1 Introduction

We now provide some examples to illustrate the utility of the RDLMwp. Our first example predicts a step response function from one explanatory time series, where we obtain the average fit of the model from 100 simulations. Our second example is taken from Harrison and Reed (1999), where the aim is to forecast the number of confirmed cases of BSE amongst 5 year old cows based upon the actual number of cases in 4 year old cows. In the first example we wish to capture the discontinuities in the regression coefficient over time. In the second example we wish to capture the non-stationary behaviour of the component time series.

We have used the Haar wavelet in the NWPT because it performed well for our examples. In Example one, there appeared little difference in the fit of the RDLMwp for each of the Daubechies' extremal phase wavelets. In Example two, the residual diagnostics for the models deteriorated slightly as we increased the number of vanishing moments.

For each model we specify a prior information set; this includes specifying a prior distribution for the regression coefficients and a variance model for the observational and evolution error terms. Note we specify the evolution variance W_t via the discounting approach (see Section 2.5.5).

Numerical comparisons using the one-step ahead forecast errors, $e_t\,(t = 1, \ldots, n)$, such as mean absolute deviation $(\sum |e_t|/n)$ and mean square error $(\sum e_t^2/n)$, are used to compare models in terms of simple forecasting accuracy. A section of negative residuals may indicate a dynamic model that is too adaptive, i.e. the discount factors are too small. The result is the posterior distributions adapt markedly to each observation, and thus short-term predictive accuracy is lost (p. 350, Harrison and West (1997)).

5.3.2 Example 1: Simulation results

We propose the RDLMwp may provide a better representation of a series with non-smooth regression coefficients. To investigate we follow the process below to obtain the average fit of the standard RDLM.

1. Construct a simulation for time series of length $n = 128$ as follows:

   A response time series is constructed as

   $$Y_t = \Theta_t X_t + E_t,$$

   where Θ_t is chosen to be

   $$\begin{aligned} \Theta_t = \quad & \mathrm{N}[2, (0.25)^2], \quad 1 < t \leq 32 \\ & \mathrm{N}[3, (0.25)^2], \quad 32 < t \leq 64 \\ & \mathrm{N}[4, (0.25)^2], \quad 64 < t \leq 96 \\ & \mathrm{N}[5, (0.25)^2], \quad 96 < t \leq 128. \end{aligned}$$

The explanatory time series, X_t, is generated from N[5, $(0.2)^2$] iid.

The error, E_t, is generated from N[0, $(0.2)^2$] iid.

When estimating the RDLM we treat X_t and Y_t as fixed realisations of the time series x_t and y_t.

2. Compute the RDLM estimates of θ_t using the Harrison and Reed (1999) software. The estimation model is of the form:

$$\begin{aligned} y_t &= \theta_t x_t + \nu_t, \quad \nu_t \sim \mathrm{N}[0, (0.2)^2], \\ \theta_t &= \theta_{t-1} + \omega_t, \quad \omega_t \sim \mathrm{T}_{n_{t-1}}[0, W_t]. \end{aligned}$$

We specify the prior information set (at $t = 0$) as follows: We constrain the trend to be zero, i.e. there is no intercept in this model. We do not standardise the regression variable and the regression coefficient θ_t has zero mean, standard deviation of 1.5 and an associated discount factor of 0.8. By setting the prior mean of θ_t to zero we assume the explanatory variable has no initial effect on the response variable. The observational variance V is fixed at 0.04. A residual analysis was used to examine the validity of different starting parameters.

3. Repeat steps one to two 100 times to obtain the average RDLM estimates. We denote y_t as the true value of the response, $\bar{f}_t = \sum_{n=1}^{100} f_t/100$ as the average online forecast and $\bar{f}_t^R = \sum_{n=1}^{100} f_t^R/100$ as the average retrospective forecast. Similarly, the standardised residuals are calculated as $\tilde{e}_t = \bar{e}_t/\mathrm{sd}(\sum_1^{t-1} \bar{e}_t)$.

We provide the average online forecasts, i.e. the averaged true values $\bar{y}_t$ and the average one-step ahead forecasts $\bar{f}_t$ in Figure 5.1. The true values of the response series are indicated by the solid line and forecasts by the dotted line. We can see the RDLM does not capture well the behaviour of the step function, responding slowly to changes in the regression coefficient. The standardised residuals remain positive because the RDLM adapts slowly to change points in the data set. The online analysis has failed to adequately represent the response time series. We could consider a lower discount factor for the regression coefficient but low discount factors, such as $\delta = 0.8$, reflect a rapidly changing quantitative relationship and can signal an inadequate model (p. 196, Harrison and West (1997)).

We provide the average retrospective estimates in Figure 5.2, where the true values y_t and the average retrospective forecasts $\bar{f}_t^R$ are represented by solid and dotted lines respectively. The RDLM approximates the response time series with a smoothed step-like function which fails to capture the true behaviour in the time series. Figure 5.3 is a normal probability plot of the standardised residuals $\tilde{e}_t$, which shows that the normality assumption is invalid.

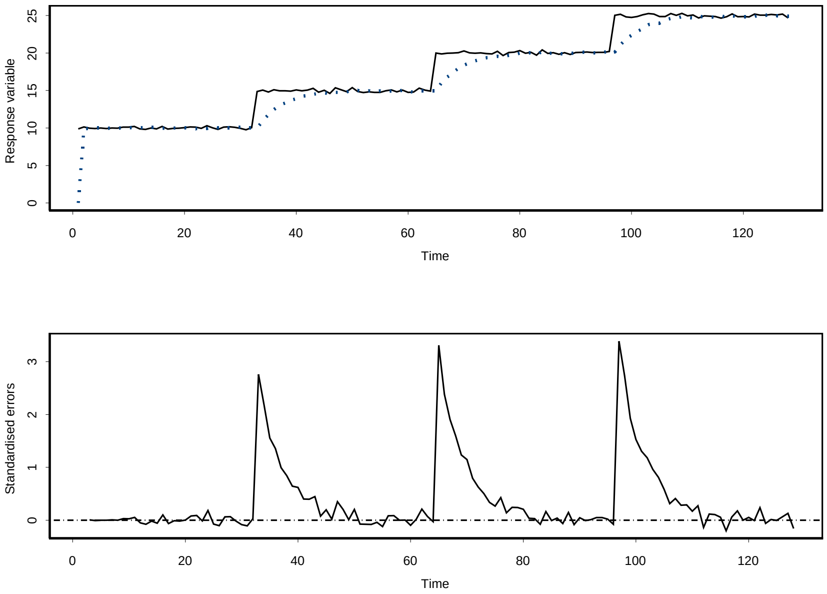

Figure 5.1: Online forecasting for Example 1. In the top graph, the averaged true values of the response series, $\bar{y}_t$, are represented by the solid line and the average online forecasts, $\bar{f}_t$, by the dotted line. The lower graph illustrates the standardised residuals $\tilde{e}_t$. The RDLM has not provided accurate online forecasts and consequently the standardised residuals tend to be positive.

We conclude the RDLM performs poorly for this data set. As an alternative to an intervention policy implemented at each suspected change point in the regression coefficient, we investigate next if including wavelet packet coefficients as explanatory variables in the RDLM will improve the accuracy of predictions.

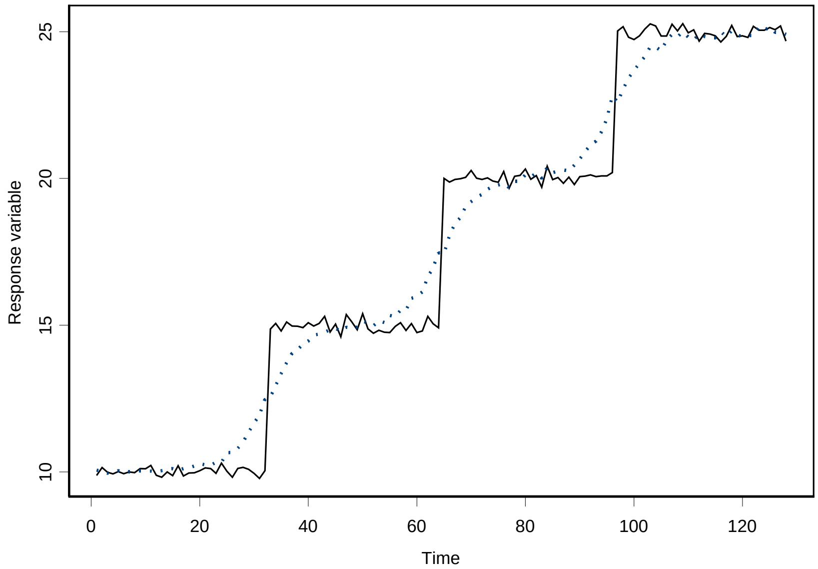

Figure 5.2: Retrospective analysis for Example 1. The averaged true values $\bar{y}_t$ are represented by the solid line and the average retrospective forecasts $\bar{f}_t^R$ by the dotted line. The RDLM has not provided accurate retrospective forecasts.

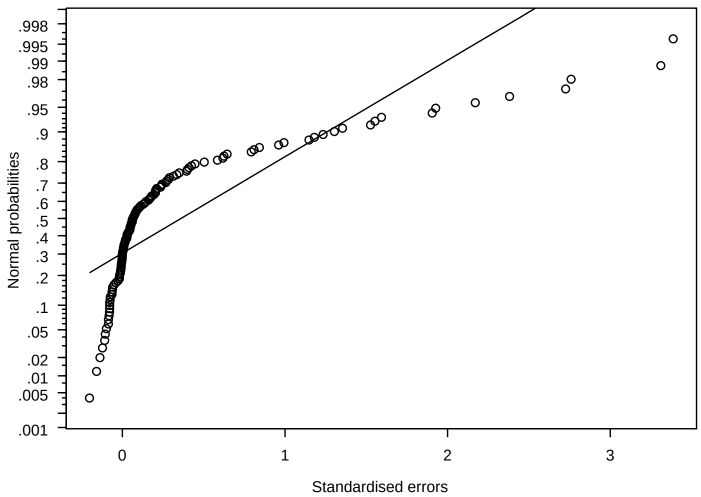

Figure 5.3: The normal probability plot of the standardised residuals $\tilde{e}_t$ for Example 1. The standardised residuals are clearly non-normal.

Example 1 with wavelet packet preprocessing

We investigate our initial supposition that there may be a significant improvement in model fit by incorporating wavelet packet coefficients as explanatory variables. We perform the procedure below to obtain the average fit of the RDLMwp.

1. Simulate the response and explanatory series as given at the start of Section 5.3.2.

2. Perform the wavelet preprocessing step.
   As human interaction is strictly limited we choose to include all wavelet packet coefficients highlighted by stepwise analysis, i.e. we do not attempt to find the 'best' model in each case.

3. Compute the RDLMwp estimates using the Harrison and Reed (1999) software. The prior information set is defined as before, except each regression coefficient has zero mean, a standard deviation of 0.25 and associated discount factor of 0.9.

4. Repeat steps one to three 100 times to obtain the average RDLMwp estimates.

The model is of the form:

$$\begin{aligned} y_t &= \textstyle\sum_{i=1}^{n} \theta_{t,i} \mathcal{W}_{t,i} + \nu_t, & \nu_t &\sim \mathrm{N}[0, (0.2)^2], \\ \theta_{t,i} &= \theta_{t-1,i} + w_t, & w_t &\sim \mathrm{T}_{n_{t-1}}[0, W], \end{aligned}$$

where $\mathcal{W}_{t,i}$ is the i^{th} wavelet packet coefficient and n is the number of wavelet packet coefficients in the model (which may differ for each simulated data set).

We provide the online forecasting results in Figure 5.4, where the solid line indicates the true values y_t and the dashed line indicates the forecasted values $\bar{f}_t$. We can see the online forecasts are very close to the true values. There is only a very short time period after the change point in the response series before the forecasted and actual values converge. The standardised residuals are smaller than those of the previous model (see Figure 5.1). The RDLMwp has adapted much more quickly to change points in the data set by using wavelet packet coefficients as explanatory variables. Results of the retrospective analysis are given in Figure 5.5. We can see the retrospective forecasts $\bar{f}_t^R$ are accurate, in fact almost identical to the true values y_t.

The normal probability plot of the standardised residuals is given in Figure 5.6. The changes in the regression coefficient have resulted in some non-normality of the

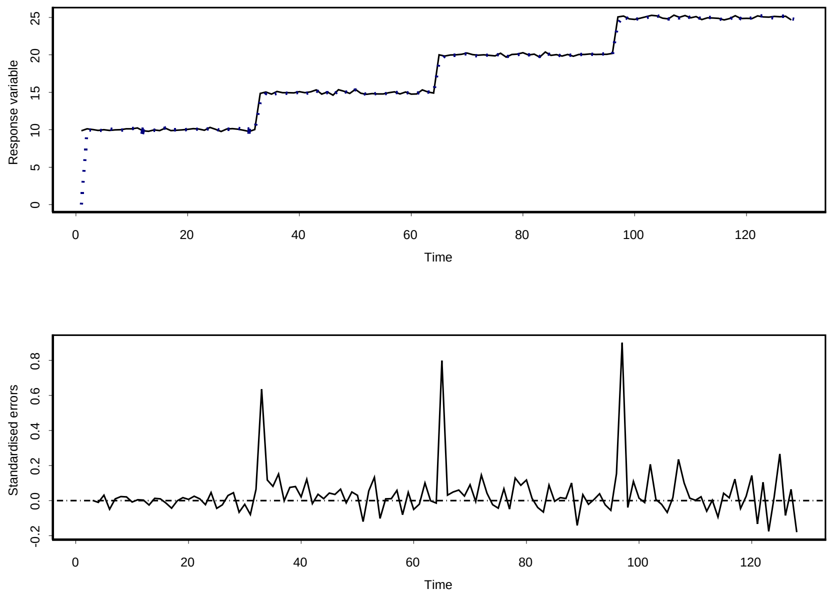

Figure 5.4: Online forecasting for Example 1 (RDLMwp). In the top graph, the averaged true values of the response series, $\bar{y}_t$, are represented by the solid line and the average online forecasts, $\bar{f}_t$, by the dotted line. The lower graph illustrates the standardised residuals $\tilde{e}_t$. The RDLMwp has provided accurate online forecasts and the standardised residuals are large only at change points.

standardised residuals. For further comparison between models we provide a table of residual diagnostics:

	RDLM	RDLMwp
mean absolute deviation	93	22
mean square error	308	101

We conclude the RDLMwp has produced a model that closely reflects the true behaviour of the data set. As human interaction in the simulations is strictly limited we simply included all the explanatory variables highlighted by stepwise analysis in the model. This would hopefully capture the behaviour of the response time series but would not necessarily provide the best fit in terms of diagnostic plots. The diagnostics of the simulation results may be improved by a better decision rule for the inclusion of wavelet packet coefficients in the model. Although the true change in the regression coefficient may not be easily quantified in the RDLMwp, the wavelet packet coefficients

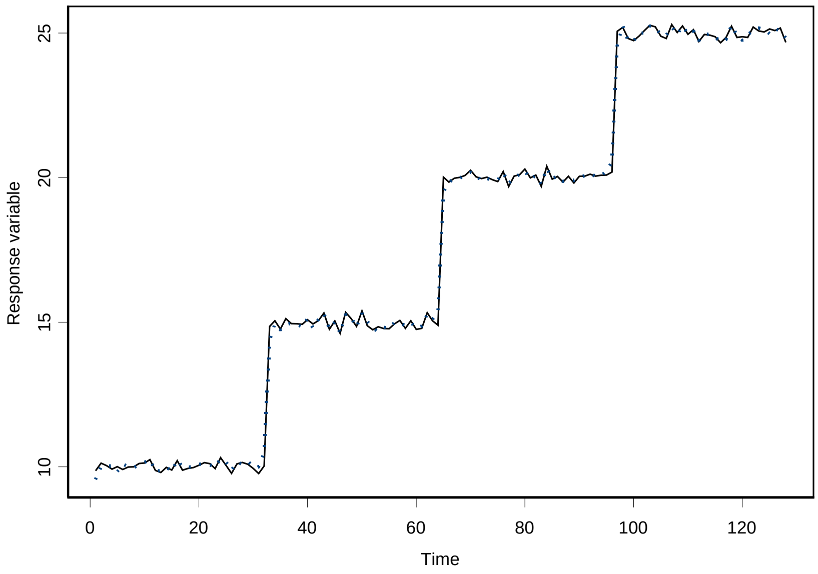

Figure 5.5: Retrospective forecasting for Example 1 (RDLMwp). The averaged true values $\bar{y}_t$ are represented by the solid line and the average retrospective forecasts $\bar{f}_t^R$ by the dotted line. The RDLMwp has provided accurate retrospective forecasts.

have provided a good model that may be reliable over a reasonable period (reflected by the high discount factor for the regression coefficients).

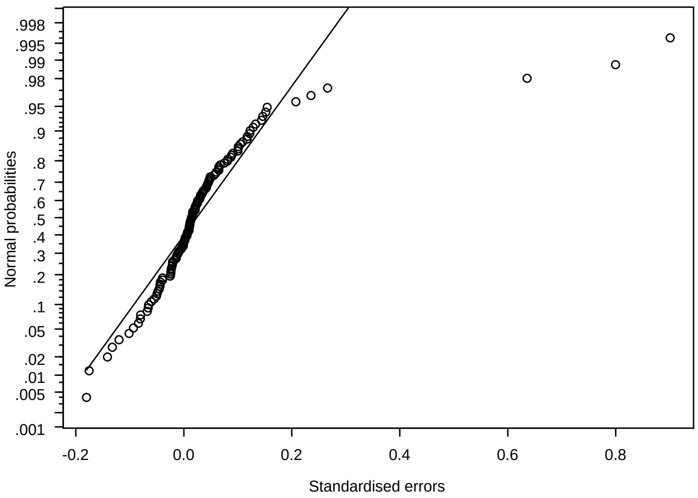

Figure 5.6: The normal probability plot of the standardised residuals $\tilde{e}_t$ for Example 1 (RDLMwp). The standardised residuals show some deviation from the normal distribution.

5.3.3 Example 2: BSE data (Harrison and Reed, 1999)

Our second example, taken from Harrison and Reed (1999), comprises of a data set where the component time series appear non-stationary. We wish to assess how the inclusion of wavelet packet coefficients as explanatory variables will affect the RDLM. The aim is to forecast the number of confirmed cases of BSE amongst 5 year old cows which are recorded in the fourth quarter ahead (y_t), based upon the number of confirmed cases of 4 year olds recorded in the last quarter (x_t). The time series are of length 31 and are plotted in Figure 5.7.

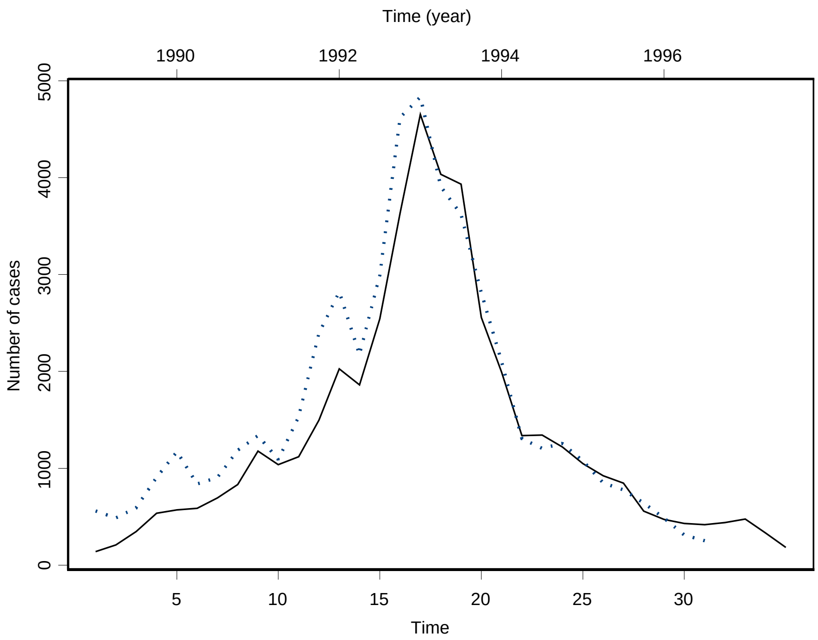

Figure 5.7: The number of confirmed cases of 4 year olds with BSE lagged four quarters (solid line) and the number of confirmed cases of 5 year olds with BSE (dashed line). On the x-axis time t corresponds to observation number t in the time series.

We construct the RDLM following Harrison and Reed (p. 7, 1999). The model is of the form:

$$\begin{aligned} y_t &= \theta_t x_t + \nu_t, \quad (\nu_t|V) \sim \mathrm{N}[0, V], \\ \theta_t &= \theta_{t-1} + \omega_t, \quad (\omega_t|V) \sim \mathrm{T}_{n_{t-1}}[0, W_t]. \end{aligned}$$

For the prior information set the regression variable is not standardised and has a regression coefficient with mean and standard deviation of 1 and an associated discount factor of 0.8. The observational error terms, ν_t, are assumed constant with $V = 100$

on 1 degree of freedom.

We plot the online and retrospective forecasting results in Figures 5.8 and 5.9. We can see for online analysis the RDLM performs well along the smoothly changing segments of the time series but has problems in approximating the spike (time point 17). Solid dots on the online and retrospective forecasting graphs indicate the software's automatic monitoring system suggesting intervention. This may occur when, for example, there has been a large increase in the observational variance V or an observation occurs in the tail of the assumed forecast distribution. The software suggests intervention in three places (time point 7, 17 and 19) but no action was taken in Harrison and Reed (1999). For retrospective analysis the RDLM captures the spike effectively but performs poorly in the period before the spike.

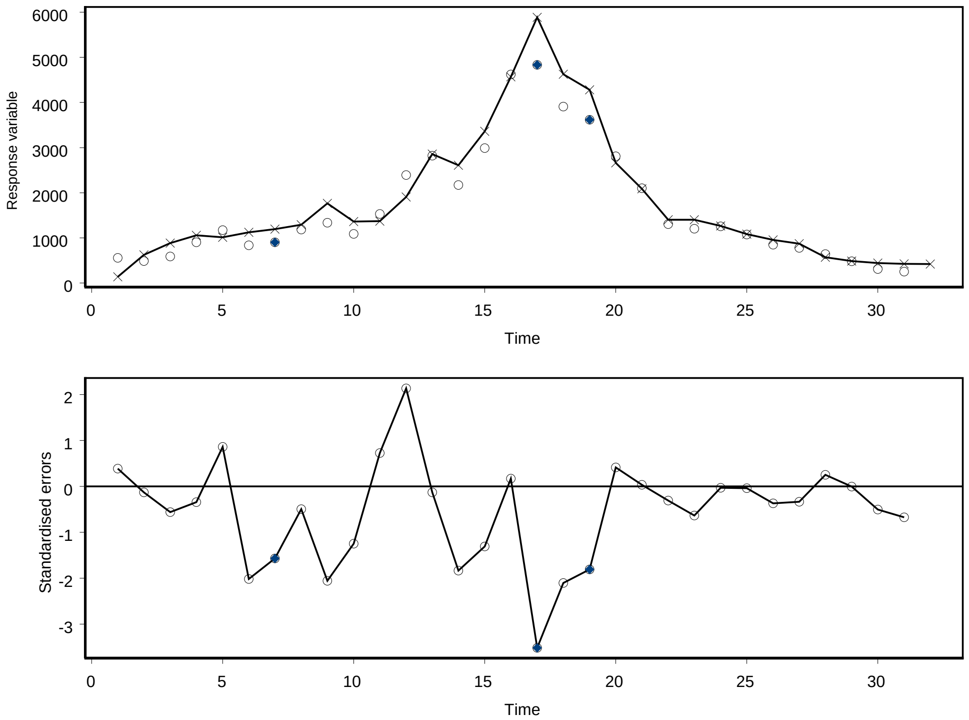

Figure 5.8: Online forecasting for Example 2, where 'o' denotes the true values of the response series y_t and '$\times$' denotes the online forecasts (f_t) in the top graph. Note the RDLM has problems approximating the spike at time point 17. The solid dots indicate the software's automatic monitoring system suggests intervention at this time point.

The retrospective estimate of the regression coefficient θ_t is given in Figure 5.10. The regression coefficient exhibits smooth evolution; apparently converging to around 1 which seems reasonable from Figure 5.7.

Figure 5.11, a plot of the observational variance V, shows how V becomes highly

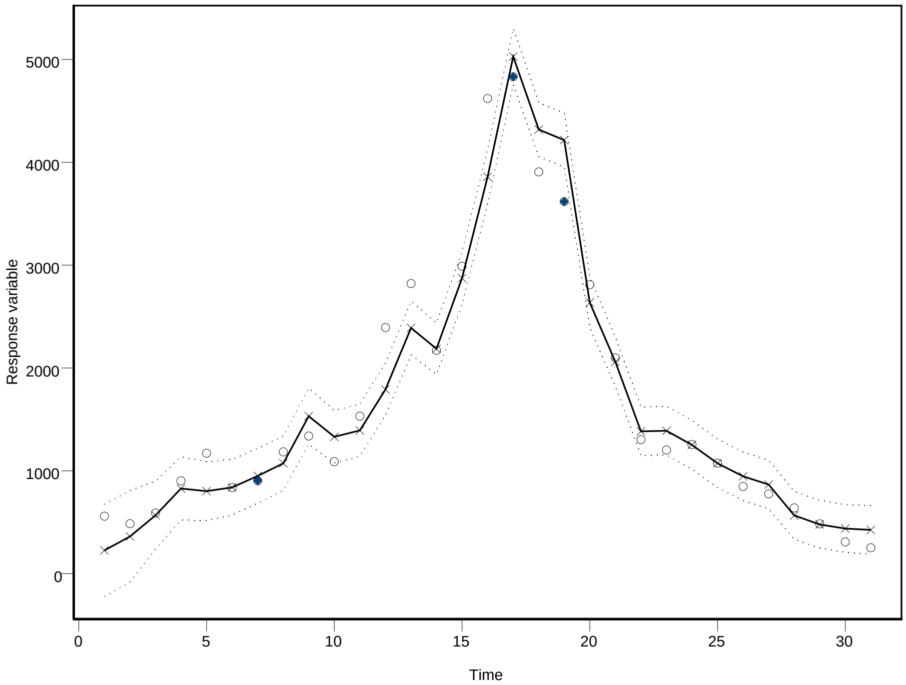

Figure 5.9: Retrospective analysis for Example 2, where 'o' denotes the true value of the response series y_t, '$\times$' denotes the retrospective forecasts f_t^R and '- - -' denotes the 2/3 confidence intervals. The RDLM has provided reasonably accurate retrospective forecasts.

inflated around the spike. Refer to Cooper and Harrison (1997) for a detailed discussion of this model and associated developments.

We conclude the RDLM provides an adequate fit of the data set but we will investigate if a better model may be obtained by including wavelet packet coefficients as explanatory variables in the RDLM.

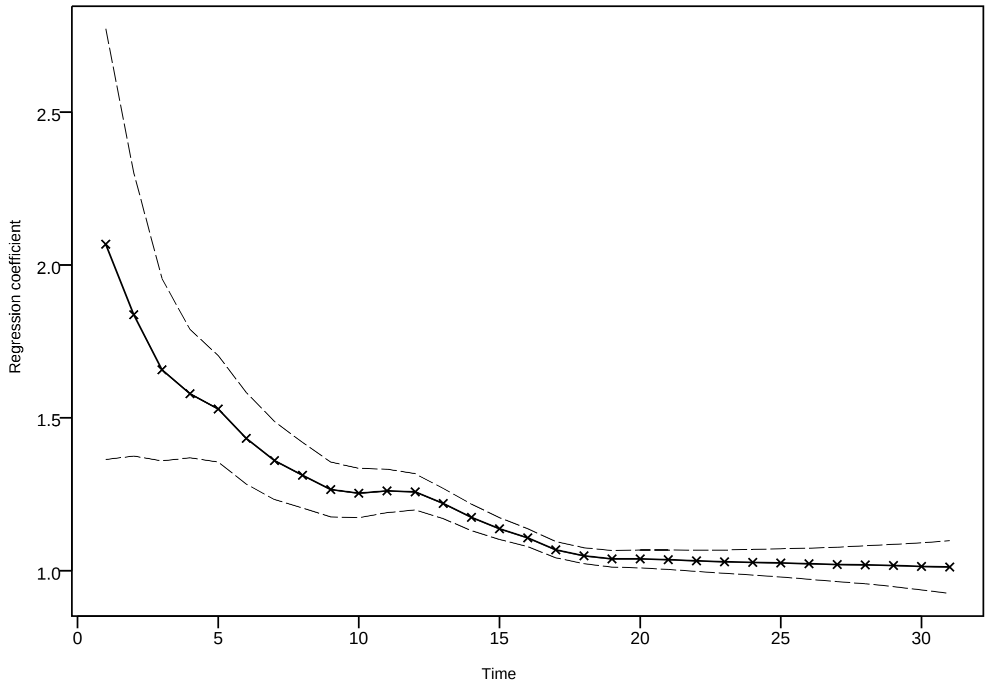

Figure 5.10: The retrospective estimate of θ_t for Example 2. The regression coefficient appears to have converged to 1.

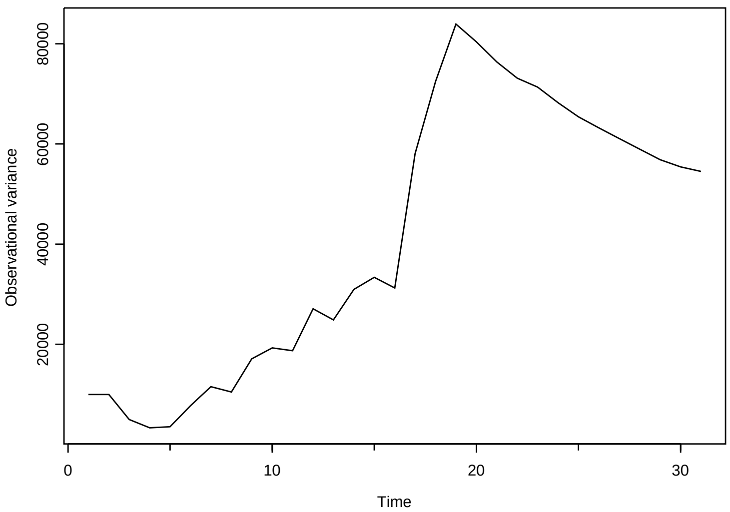

Figure 5.11: The observational variance V; the biggest increase occurs around time point 17 (coinciding with a spike in the data set).

Example 2 with wavelet packet preprocessing

We investigate if using wavelet packet coefficients as explanatory variables will improve the fit of the model for the BSE data. We follow the same fitting procedure as described in Section 5.2, although we pad the series with zeros on either side to length 64 in order to perform the NWPT in WaveThresh3. We choose to include four variables in the model; $\mathcal{W}_{t,1}$, $\mathcal{W}_{t,41}$, $\mathcal{W}_{t,42}$ and $\mathcal{W}_{t,31}$ and their correlations are given below:

$$\begin{pmatrix} 1.00 & 0.08 & -0.32 & -0.23 \\ - & 1.00 & 0.00 & 0.39 \\ - & - & 1.00 & -0.01 \\ - & - & - & 1.00 \end{pmatrix}$$

A description of the wavelet packets represented in the model is given in Table 5.1 and a plot of the wavelet packets in Figure 5.12. The wavelet packet coefficients $\mathcal{W}_{t,41}$ and $\mathcal{W}_{t,42}$ exhibit high frequencies and draw observations from the future. $\mathcal{W}_{t,1}$ is a father wavelet which averages over the last 2 observations.

Explanatory Variable	Packet Scale	Packet Frequency
$\mathcal{W}_{t,1}$	5	0
$\mathcal{W}_{t,31}$	3	6
$\mathcal{W}_{t,41}$	0	48
$\mathcal{W}_{t,42}$	0	27

Table 5.1: Wavelet packets represented in Example 2.

The model is of the form:

$$\begin{aligned} y_t &= \theta_{t,1}\mathcal{W}_{t,1} + \theta_{t,41}\mathcal{W}_{t,41} + \theta_{t,42}\mathcal{W}_{t,42} + \theta_{t,31}\mathcal{W}_{t,31} + \nu_t, && (\nu_t|V) \sim \mathrm{N}[0, V], \\ \theta_{t,i} &= \theta_{t-1,i} + w_t, && (w_t|V) \sim \mathrm{T}_{n_{t-1}}[0, W_t]. \end{aligned}$$

When constructing the RDLMwp we use a similar prior information set as before because we wish to provide a model that is comparable with the Harrison and Reed (1999) model. The regression variables are not standardised and have regression coefficients with a mean and standard deviation of 1 and an associated discount factor of 0.8. The observational variance V is assumed constant with $V = 100$ on 1 degree of freedom.

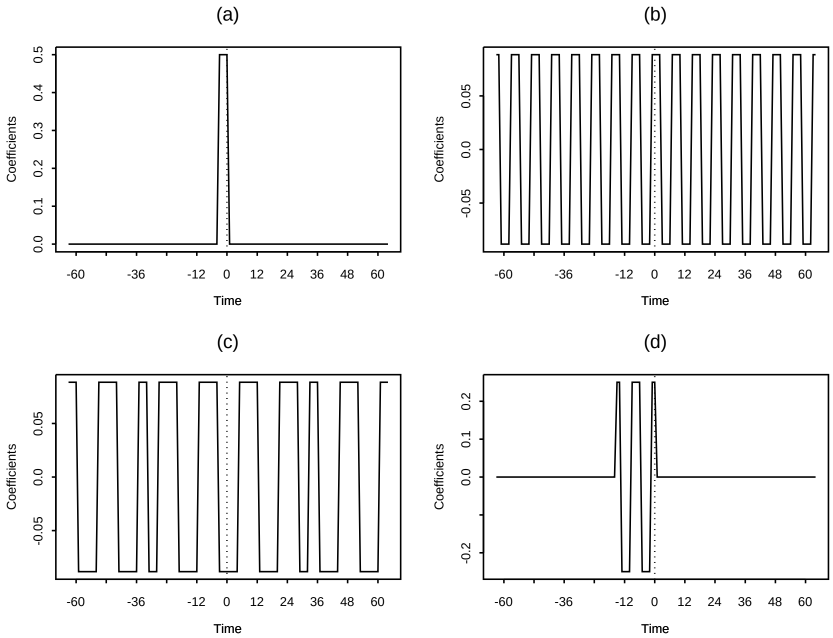

Figure 5.12: Wavelet packets represented in Example 2: (a) $\mathcal{W}_{t,1}$, (b) $\mathcal{W}_{t,41}$, (c) $\mathcal{W}_{t,42}$ and (d) $\mathcal{W}_{t,31}$.

We plot the online and retrospective forecasting results in Figures 5.13 and 5.14. The online analysis performs similarly to the last model along the smoothly changing segments of the time series, but captures the behaviour of spike (time point 17) far more effectively. For retrospective analysis the RDLMwp again performs better; the forecasts f_t^R are closer to the true values y_t and the confidence limits are much narrower. The softwares automatic monitoring system suggests intervention for only one observation, compared with three observations highlighted in the RDLM.

We have plotted how the regression parameters evolve over time for the four explanatory variables in the model; $\mathcal{W}_{t,1}$ in Figure 5.15, $\mathcal{W}_{t,31}$ in Figure 5.16, $\mathcal{W}_{t,41}$ in Figure 5.17 and $\mathcal{W}_{t,42}$ in Figure 5.18. The effect of $\mathcal{W}_{t,1}$ decreases in magnitude initially and then remains quite stable, converging to around 1.1. The evolution of $\mathcal{W}_{t,1}$, which represents a father wavelet, appears similar to the evolution of θ_t in the RDLM; the variables may contain similar information. The coefficients of $\mathcal{W}_{t,31}$ and $\mathcal{W}_{t,42}$ decreased in magnitude and may not have converged. The regression coefficient for $\mathcal{W}_{t,41}$ is stable initially but then evolves as a large spike; it may be capturing the spike (time point 17) in the explanatory series.

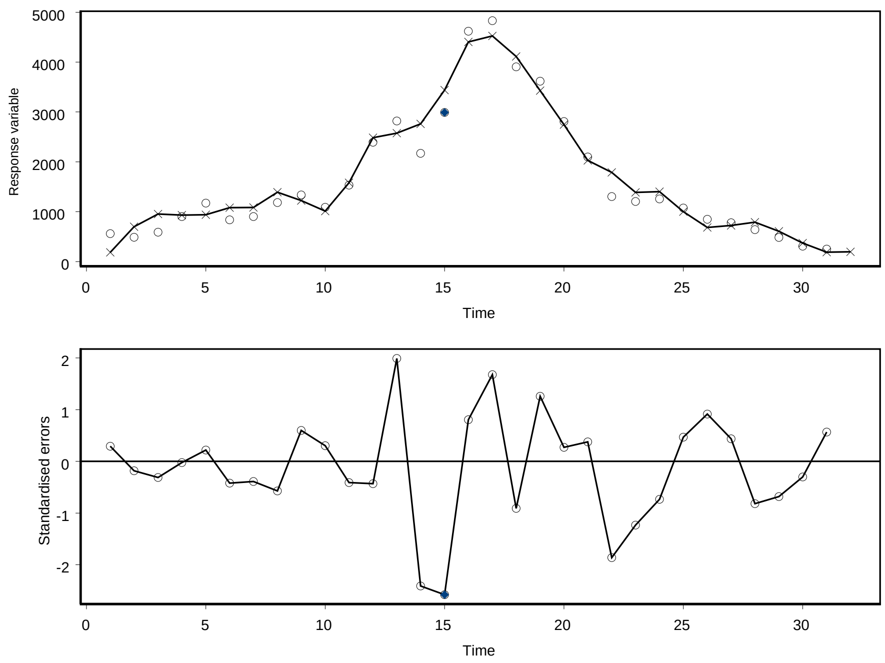

Figure 5.13: Online forecasting for Example 2 (RDLMwp), where 'o' denotes the true values of the response series y_t and '$\times$' denotes the online forecasts (f_t) in the top graph. The RDLMwp has provided accurate online forecasts.

Figure 5.19 shows how the observational variance V is less inflated than before. For further comparison between models we provide a table of residual diagnostics:

	RDLM	RDLMwp
mean absolute deviation	244	194
mean square error	115,160	56,353

We note that if we had included $\mathcal{W}_{t,1}$ as the sole regressor in the RDLMwp the mean absolute deviation would be 224 and the mean square error would be $86,176$. So, using just one explanatory variable in the RDLMwp would have offered gains over the RDLM.

We conclude, for this data set, the inclusion of wavelet packet coefficients has made an improvement to the model; the behaviour of the spike in the data set was captured effectively and residual diagnostics showed a reduction in forecasting error. To aid comparison between models we did not attempt to pick the optimum prior information set so further improvements may be possible for this data set.

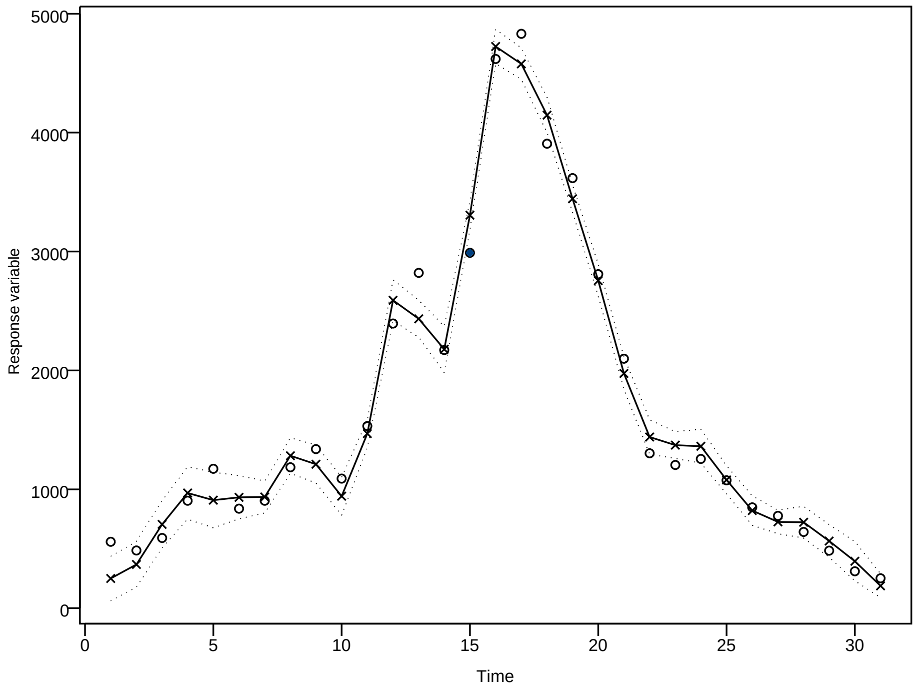

Figure 5.14: Retrospective analysis for Example 2 (RDLMwp), where 'o' denotes the true value of the response series y_t, '×' denotes the retrospective forecast (f_t^R) and '- - -' denotes the 2/3 confidence intervals. The RDLMwp has provided accurate retrospective forecasts.

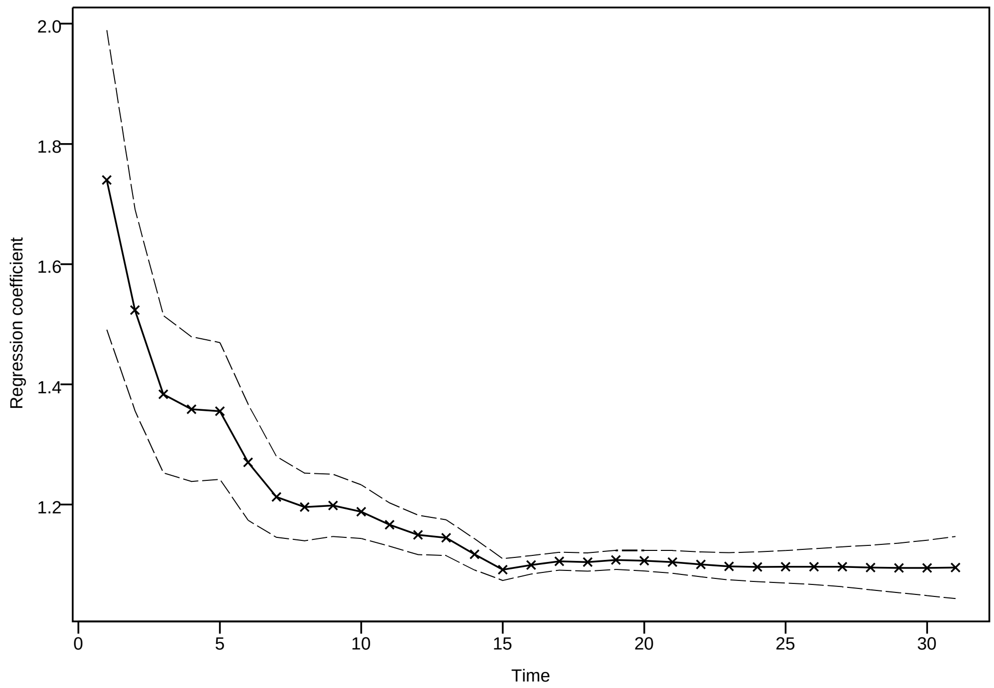

Figure 5.15: The retrospective estimate of $\theta_{t,1}$ (Example 2). The regression coefficient appears to have converged.

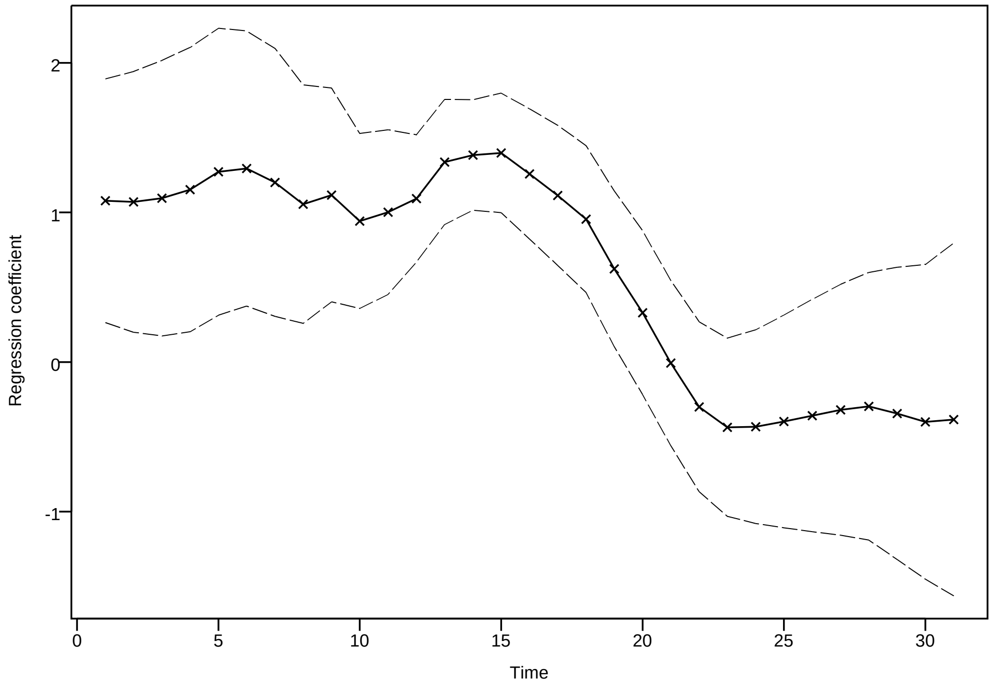

Figure 5.16: The retrospective estimate of $\theta_{t,31}$ (Example 2).

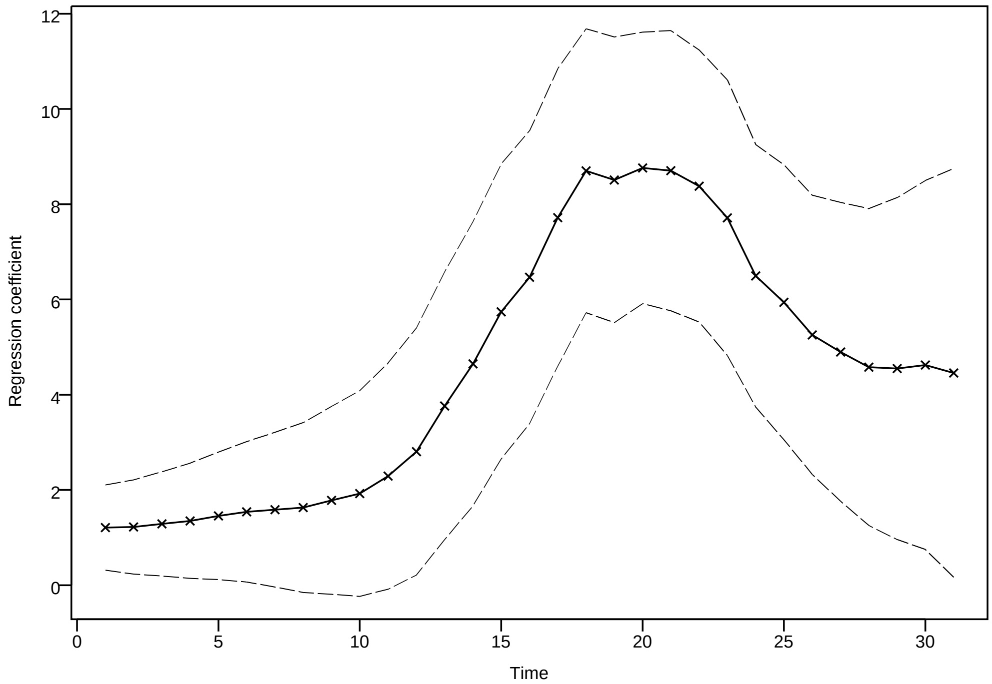

Figure 5.17: The retrospective estimate of $\theta_{t,41}$ (Example 2).

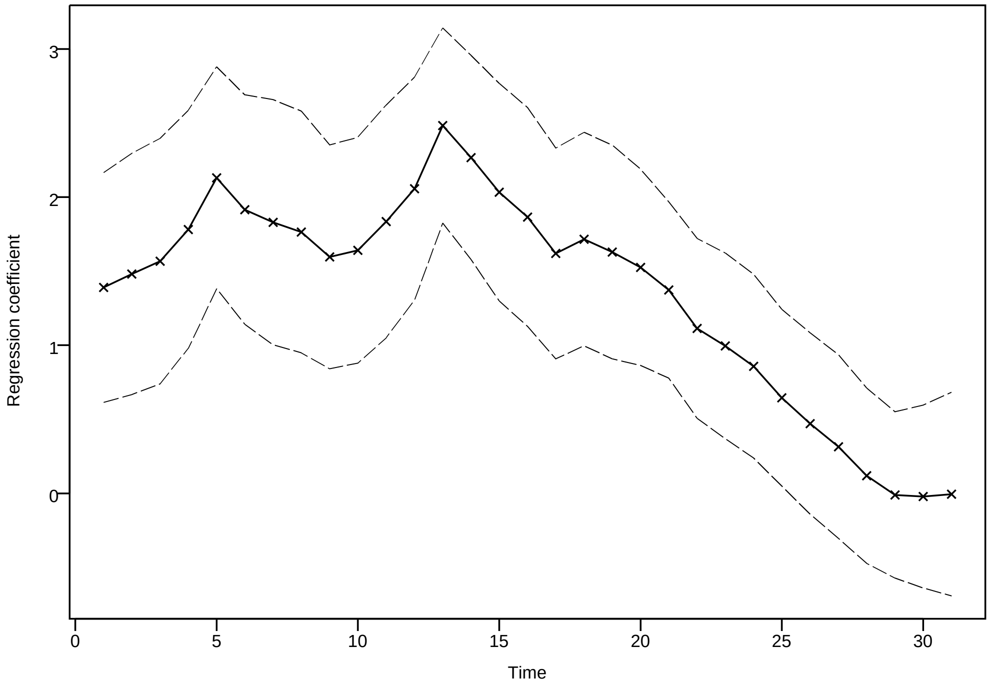

Figure 5.18: The retrospective estimate of $\theta_{t,42}$ (Example 2).

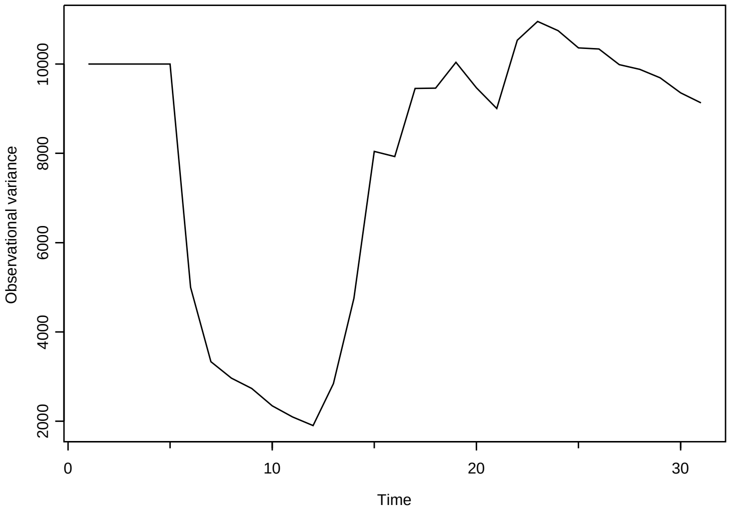

Figure 5.19: The observational variance V for Example 2 (RDLMwp).

5.4 Conclusions and future work

In this chapter we have investigated multiple regression Bayesian dynamic linear modelling (see Harrison and West, 1997). We have proposed adding a wavelet packet preprocessing step to the RDLM and termed the model 'RDLMwp'. The wavelet packets represented in the RDLMwp may indicate which features of the explanatory time series, x_t, may be useful in predicting the response times series, y_t. As wavelet packets are particularly effective at representing time series with discontinuities we considered models with a rapidly changing relationship between the response and explanatory time series. We have considered two types of relationship; non-smooth regression coefficients and non-stationarity in the component time series.

We provided two examples to compare the effectiveness of the RDLM with and without a preprocessing step. Simulation studies of a step response function with one explanatory time series (Example 1) concluded that including wavelet packet coefficients in the RDLM has been beneficial in modelling a rapid change in the regression coefficients. The second example, when predicting the number of 5 year old cows with BSE from the number of confirmed cases in 4 year olds, demonstrated that the RDLMwp may provide more accurate forecasts when the component time series are non-stationary.

There are other aspects of the RDLMwp that we wish to investigate in the future. We may wish to restrict ourselves to 'real-time' wavelet packets; obviously of importance when online forecasting is required (see Section 2.9.3). We could investigate a wider range of time series behaviour to assess which data sets may be better represented by the RDLMwp. In addition, we made a decision on which wavelet packet coefficients may be useful using the whole of the time series. In practice, we may choose which wavelet packet coefficients to include based on training data, taking into account the occasional high correlation between wavelet packet coefficients.

Chapter 6

Image Regression using Wavelets

6.1 Introduction

Our aim is to construct a statistical model to relate two images. We wish to investigate the following: can you predict the intensity values of one image from another image? By taking a wavelet transform of the explanatory image, we consider if you can predict one image from only a few scales in another image, hence indicating which scales are important. We may wish to relate two images in order to predict the response image, restore missing data or to 'calibrate' an image. For example, we may wish to calibrate images from different dates to remove time dependent effects.

We discuss some current regression models before introducing the model we propose: the two-dimensional non-decimated wavelet transform (2D NWT) regression model. The 2D NWT model is a (simplified) 2D version of the wavelet packet transfer function methodology of Nason and Sapatinas (2002). We do not use the full wavelet packet transform because we feel the wavelet transform is adequate for an initial investigation. See Nason, Sapatinas and Sawczenko (2001) for an example where the full flexibility of the wavelet packet transform was not utilised.

Multispectral imagery of Chew Valley Lake in Somerset (UK) is used to assess the performance of the new model. We compare our new model to a simple linear regression model in terms of the mean squared error (MSE) of predictions. In the first example, we model a response image using a selection of frequencies from an explanatory image. In the second example, we obtain the average fit of the response image in the presence of noise from 100 simulations. To conclude, we suggest extensions to our model and consider alternative methods for image analysis.

6.2 Image processing

Image processing is implemented in a diverse range of fields, for example, engineering, chemistry, astronomy, medicine and geography. As a result there are a good selection of introductory texts; such as the technical handbook of Joyce-Loebl (1985) which covers topics such as image capture, object detection and image analysis. As much of the research to date involves the development of statistical techniques, image analysis has seen increasing prevalence in many areas of statistical research.

Image capture involves converting an image into an array of numbers that represents the intensity of light (or alternative form of energy). We may wish to analyse the captured scene in order to classify images, describe the spatial correlation, identify the shape and structure of features or restore missing data. Techniques such as regression modelling, linear discriminant analysis, cluster analysis, Bayesian approaches and neural networks have all been implemented. References for these methods include: Geladi and Esbensen (1991) for regression models, Griffith (1988) for spatial theory, Barone *et al.* (1992) for a selection of papers on statistical methods in image analysis and Glasbey and Glidewell (2001) for further discussion and references on multivariate statistical methods in image analysis.

6.3 Two-dimensional regression models

6.3.1 Introduction

We are particularly interested in the use of regression models in image analysis. We treat the response image y, a matrix of size $(n \text{ x } n)$, as a vector, i.e. the matrix

$y = y(\text{row } i, \text{column } j)$ is treated as a (n^2 x 1) column vector:

$$y = \begin{pmatrix} y(1,1) \\ y(1,2) \\ \vdots \\ y(1,n) \\ y(2,1) \\ \vdots \\ y(n,n) \end{pmatrix}$$

The explanatory image, x, is treated similarly.

6.3.2 Current models

We first review some of the existing regression models proposed for jointly modelling two images.

1. Simple linear regression

   This model relates a response image, y, to one or more explanatory images x_m ($m = 1, 2, \ldots, p$), i.e.

   $$y = x_1 + \ldots + x_p + \varepsilon, \quad \varepsilon \sim \mathrm{N}[0, \sigma^2]. \tag{6.1}$$

2. Principal components regression

   An extension to simple linear regression is principal components regression. A principal components analysis is carried out on the set of explanatory images and the resulting orthogonal principal components, pc m ($m = 1, 2, \ldots, p$), are regressed on the dependent image, i.e.

   $$y = \text{pc } 1 + \ldots + \text{pc } p + \varepsilon, \quad \varepsilon \sim \mathrm{N}[0, \sigma^2].$$

   As we will see later, multispectral images can be highly correlated and this method could be particularly useful to reduce any multicollinearity problems.

Refer to Geladi and Esbensen (1991) for discussion, examples and references for the simple linear regression and principal components regression models. Further details can be found in Chapter 4 where we implement a principal components regression.

6.4 The 2D NWT regression model

We now introduce the 2D NWT regression model. We begin by taking a 2D NWT of the (n x n) explanatory image. At each scale j ($j = 0, 1, \ldots, (J-1)$, where $n = 2^J$) there is one father wavelet coefficient and three mother wavelet coefficients at each location which capture smooth, vertical, diagonal and horizontal detail.

We construct a linear regression of the form:

$$y = \mathcal{W}_1^{2D} + \ldots + \mathcal{W}_p^{2D} + \varepsilon, \quad \varepsilon \sim \mathrm{N}[0, \sigma^2], \tag{6.2}$$

where, $\mathcal{W}_m^{2D}$ ($m = 1, 2, \ldots, p$) are the 2D non-decimated wavelet coefficients of the explanatory image. Stepwise analysis is used to highlight the wavelet coefficients that may be useful in 'explaining' the response image. We fit the final model by excluding wavelet coefficients with small t-values in turn, until all remaining variables are significant. We label our model as the 2DWT model.

Further details

The 2DWT model does not contain spatial information beyond that afforded by the 2D NWT. We do not take into account that coefficients which are far apart in a vector may be highly correlated, i.e. there is no spatial factor. We could introduce more spatial information into the 2DWT model by, for example, instead of the regression $y(2,2)$ on $x(2,2)$, we could regress $y(2,2)$ on points in the neighbourhood of $x(2,2)$, i.e.

$$y(2,2) = \sum_{i=-1}^{1} \sum_{j=-1}^{1} x(2+i, 2+j).$$

Discussion on spatial models is given in numerous texts, see for example, Haslett (1991), Griffith (1988) and Cressie (1991).

We chose to use the Haar wavelet in this analysis. During simulations (given in Section 6.7) we experimented with smoother wavelets but found the Haar wavelet to provide the best results. If we had used an image showing quite different characteristics this may not have been the case. We note that Unser (1995) who uses wavelets for texture classification writes that "increasing the number of vanishing moments ... of the basis functions does not have any real advantage for texture analysis". To detect edges in images we are effectively looking for a region of rapid change. The level of

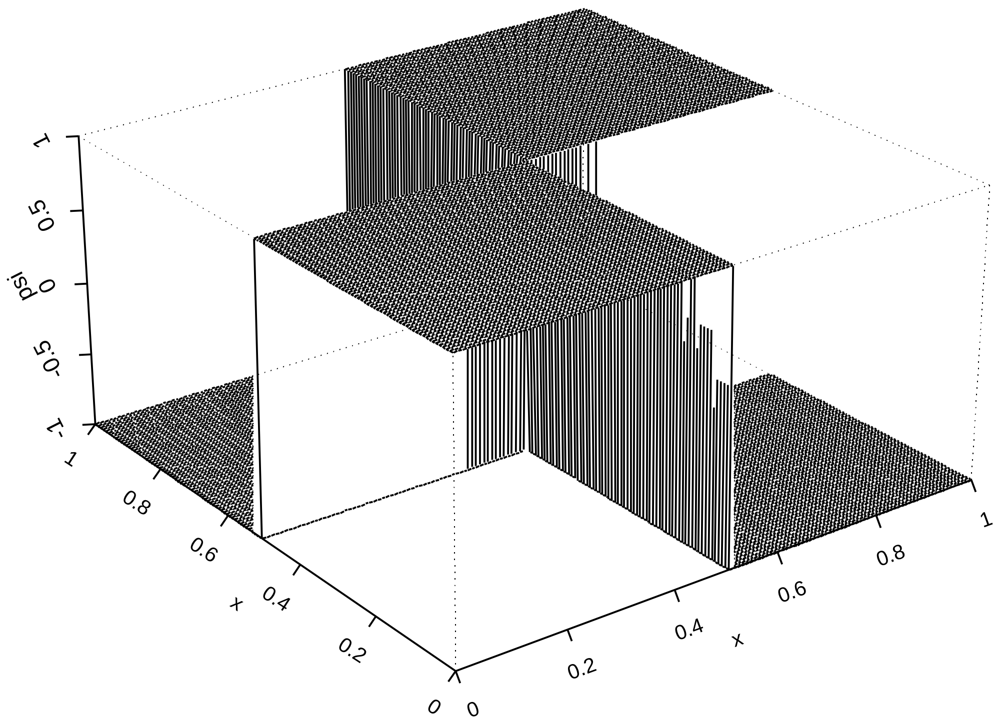

Figure 6.1: The 2D Haar diagonal mother wavelet.

change can be detected by simply subtracting the intensity value of one pixel from the intensity value of the pixel to its right (p. 89, Joyce-Loebl, 1985). This edge finding function, which seeks to find the maximum change, is similar to the one-dimensional Haar high pass filter. A picture of the diagonal 2D Haar wavelet is given in Figure 6.1.

6.5 The image data

The data set for this chapter consists of a multispectral image of size (512 x 512) of Chew Valley Lake in Somerset (UK). The scene is measured at 12 parts of the spectrum (Table 6.1) and the quality of the images is generally good. Figure 6.2 shows two (128 x 128) sections of the image sampled at channel 11. The left hand image shows some buildings and neighbouring fields and the right hand image shows a section of the water line (the lake is on the right hand side of the image). It is clear from a visual inspection that the images are non-stationary.

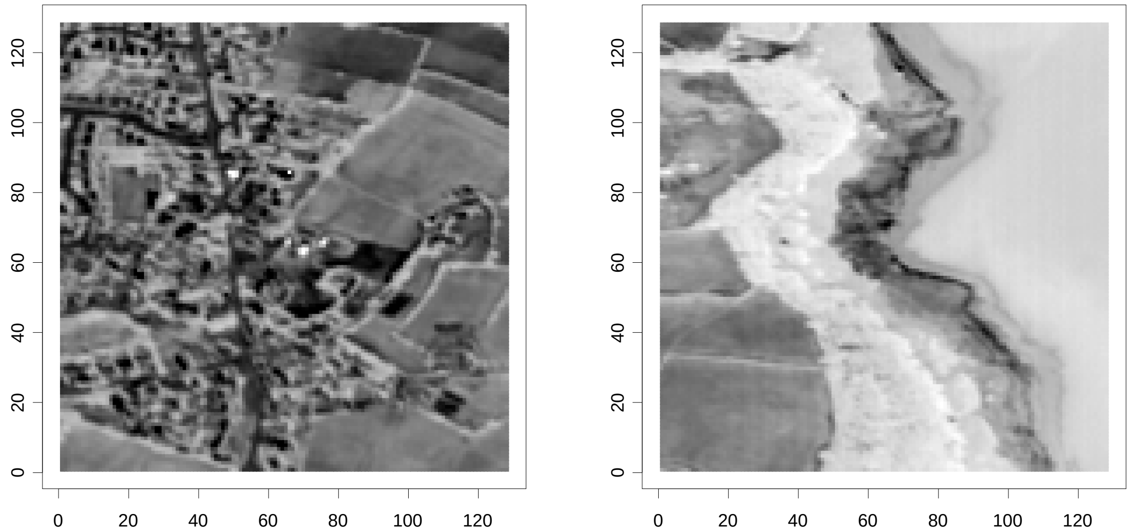

Figure 6.2: Sections from the Chew Valley Lake image.

Consider the correlation matrix given below for a (512 x 512) image between channels 2 to 7. In general, the correlations decrease as the images are further apart in spectral frequency. Clearly nearby bands contain similar types of information and, therefore, it is reasonable to relate images in nearby bands.

$$\begin{pmatrix} 1.00 & 0.97 & 0.96 & 0.92 & 0.50 & 0.52 \\ & 1 & 0.99 & 0.96 & 0.63 & 0.61 \\ & & 1 & 0.98 & 0.55 & 0.58 \\ & & & 1 & 0.69 & 0.68 \\ & & & & 1 & 0.86 \\ & & & & & 1 \end{pmatrix}$$

6.6 2D NWT regression model: Initial analysis

6.6.1 Introduction

Our aim is to investigate if the 2DWT model can accurately model a response image given a multiscale representation of an explanatory image. As a comparison we will compare the 2DWT model with the linear regression model (Equation 6.1) using one explanatory variable ($m = 1$); which we term the LINREG model. We experimented with the 2DWT model on different sections of the Chew Valley Lake image; the

Channel	Wave Length (μm)	Notes
1	0.42-0.45	noisy
2	0.45-0.52	
3	0.52-0.60	
4	0.605-0.625	
5	0.63-0.69	
6	0.695-0.75	
7	0.76-0.90	
8	0.91-1.05	noisy
9	1.55-1.75	
10	2.08-2.35	
11	8.50-13.00	
12	8.50-13.00	same as 11

Table 6.1: Frequencies of the Chew Valley Lake images.

following example, using just one small part of the image, illustrates our initial results.

6.6.2 Example One

Pictured in Figure 6.3 is a (128 x 128) segment of the Chew Valley Lake image; this particular image is at channel 3. We take this same sub-image at different channels and see if we can predict the intensity values in one channel from the intensity values in another channel.

We perform the procedure described in Section 6.4, allocating the image in one channel as the response image and the same image in another channel as the explanatory image for a selection of channels. The results, i.e. the channels of the response and explanatory image, the residual standard error (RSE) and R^2 coefficients, are given in Table 6.2.

6.6.3 Initial conclusions

We do not provide the regression equations here but summarise the results.

For images that are very highly correlated (> 0.95), including five sets of wavelet coefficients as explanatory variables in the regression equation proved adequate. In each regression equation, four of the explanatory variables were the finest scale wavelet coefficients. The remaining variable represented either a coarse scale father wavelet or

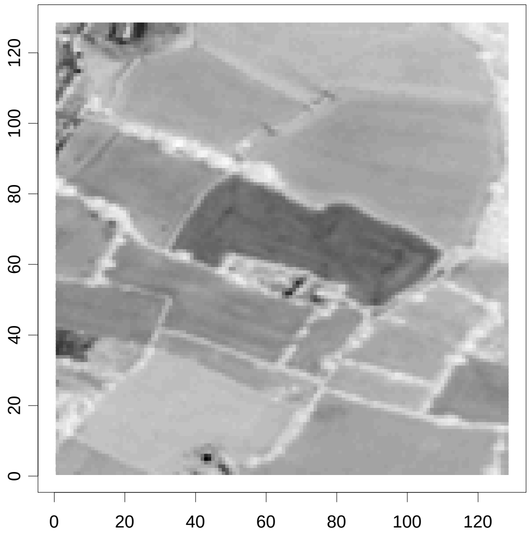

Figure 6.3: The image from Example One.

a coarse scale mother wavelet capturing horizontal information.

For less correlated images (< 0.85) a large range of wavelet coefficients from a number of scales were included in the regression equation. In each regression equation, the finest scale wavelet coefficients were always present.

For all the regression equations we fitted, the detail coefficients appeared equally important from their regression coefficients. We can see from Table 6.2 that, if the images are highly correlated, then both models perform broadly similar in terms of RSE and the multiple R^2 coefficient. For images that are not as highly correlated ($< 90\%$), we can see differences of up to 6% in the multiple R^2 coefficient; this means the 2DWT model explains up to 6% more of the variation in the response image.

6.7 2D NWT regression model: Simulation results

6.7.1 Introduction

Initial results have shown that the methods are equally effective when the images are highly correlated but that the 2DWT model may be more effective when the images are less correlated. An occasion where the images may not be highly correlated is where the response image may be noisy. We investigate this situation in the next section.

Response image	Explanatory image	Correlation	2DWT: RSE	R^2	LINREG: RSE	R^2
channel 3	channel 4	0.98	1.121	96.29	1.212	95.67
channel 3	channel 2	0.98	1.229	95.54	1.260	95.32
channel 4	channel 3	0.98	1.579	96.53	1.764	95.67
channel 4	channel 2	0.97	1.791	95.53	1.977	94.55
channel 5	channel 2	0.96	4.451	93.31	4.811	92.49
channel 9	channel 5	0.83	13.400	73.63	14.620	68.56
channel 9	channel 3	0.82	14.230	70.26	15.030	66.80
channel 9	channel 4	0.80	14.550	68.91	15.63	64.08
channel 9	channel 2	0.76	15.620	64.14	16.880	58.09

Table 6.2: Results from Example One (ordered by the degree of correlation between images).

6.7.2 Example Two

We now consider how the models perform when predicting the 'true' response image if the response image is noisy. We construct a simulation where we add noise to the response image. The images from Example One at channels 3 and 4 are used as, from a visual inspection, they appear relatively noise free.

We model the response image for five levels of zero mean noise with standard deviation, σ, equal to 1, 5, 10, 20 and 30. We note, with an average pixel intensity value of 62 for the underlying true image, the larger standard deviations produce very noisy response images. The process below is performed 100 times to provide the average fit of the 2DWT model and the LINREG model.

1. Assign channel 3 to be the response image, y, and channel 4 to be the explanatory image, x.

2. Add noise, $\varepsilon \sim \mathrm{N}[0, \sigma^2]$, to the response image, i.e. $\tilde{y} = y + \varepsilon$.

3. Build a LINREG model between the noisy channel 3 and channel 4, i.e. $\tilde{y} = x + \varepsilon$.

4. Build a 2DWT model between the noisy channel 3 and channel 4, i.e. $\tilde{y} = \mathcal{W}_1^{2D} + \cdots + \mathcal{W}_p^{2D} + \varepsilon$, where, $\mathcal{W}_m^{2D}$ $(m = 1, 2, \ldots, p)$ are the 2D non-decimated wavelet coefficients of channel 4.

   Stepwise analysis was used to highlight the important $\mathcal{W}_m^{2D}$. We choose to include

Std. dev. of noise (σ)	2DWT model	LINREG model
1	1.08	1.47
5	1.12	1.47
10	1.22	1.48
20	1.58	1.51
30	2.18	1.60

Table 6.3: Mean Squared Error (MSE) of the predictions (simulation results). Note the average intensity value of the true response series is 62.

every $\mathcal{W}_m^{2D}$ highlighted by stepwise, even if it may not be significant in the fitted model.

5. Assess the two models by comparing mean squared error (MSE), i.e. the fitted values, $\hat{y}$, of channel 3 with the 'true' values, y, of channel 3.
   MSE $= \sum(y - \hat{y})^2/n^2$, where n^2 is the number of pixels in the image.

6.7.3 Simulation conclusions

The MSE of predictions, for each level of noise, are given in Table 6.3. The results show that for noise levels less than 20 the 2DWT model provides the better fit. For high levels of noise, the LINREG model provides a better representation of the data. We conclude, the 2DWT model provides a marginally better representation than a LINREG model for low levels of noise.

6.8 Alternatives to the 2D NWT

6.8.1 Introduction

We were interested to discover why the 2DWT model did not offer greater gains against the simpler LINREG model compared to the performance improvements in one-dimension (1D) given in Chapter 4. We provide a brief summary of one further method that may be useful for image analysis, which may explain why the 2DWT model may be deficient.

6.8.2 Non-separable wavelet bases

The 2D wavelet transform we use is a separable transform; the 2D wavelets are simply tensor products of 1D wavelets. Non-separable bases have become increasingly popular in the field of image analysis. They are more complicated constructions because the multidimensional wavelets are not tensor products of 1D wavelets. Schemes based on a true 2D transform are expected to perform better than a scheme constructed from two 1D transforms (Do and Vetterli (2001) and Bonnet *et al.* (2000)).

A 2-D MRA is a generalization to $L^2(\mathbb{R}^2)$ of the 1D proposed by Mallat (1989). The basic dilation is a matrix D with integer entries such that all its eigenvalues have absolute value strictly greater than 1. An example of the so called *quincunx* matrix is given by:

$$\mathbf{D} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The quincunx matrix is of interest in image processing because it treats the directions more homogeneously then the separable schemes (p. 322, Daubechies, 1992).

Refer to Borup and Nielsen (2002) for details of the multiresolution associated with dilation matrix $\mathbf{D}$. For further details refer to Meyer (1992) or Daubechies (1992). Examples of the utility of the non-separable wavelet and wavelet packet transforms are given in Borup and Nielsen (2002) and Bonnet *et al.* (2000).

For further details refer to Candès and Donoho (1999). Do and Vetterli (2001) have also developed an orthonormal basis in the Radon domain; which is invertible, non-redundant and based on fast algorithms.

6.9 Conclusions and future work

In this chapter we have investigated a new method for modelling a response image in terms of an explanatory image; the 2D NWT regression model (the 2DWT model). By taking a 2D NWT of the explanatory image we hoped to obtain an efficient representation of the image, uncovering structure which may be useful in predicting the response image.

We assessed our 2DWT model using imagery of Chew Valley Lake in Somerset (UK). The images are sections of (128 x 128) pixels, which represent the same scene but measured at different parts of the spectrum. We compare our 2DWT model with a simple linear regression model (the LINREG model). Initial results from the 2DWT model were promising, suggesting that our model may be effective at relating two images when the correlation between them is relatively low. Simulation results, when predicting the response image in the presence of noise, showed the 2DWT model outperformed the LINREG model for low levels of noise. For particularly noisy data the LINREG model provided a better fit to the data.

To conclude, we suggested an alternative method for image analysis; non-separable wavelet transforms. This provided some insight into why our method did not offer greater gains over the LINREG model; namely that 2D wavelets are not fully efficient at representing 1D singularities. Wavelets are efficient at representing point discontinuities; they are good at picking up a discontinuity across an edge, but not the smoothness along an edge (p. 2, Do and Vetterli, 2001). We have not explored the topic in depth but the method may warrant further investigation.

Extensions to the 2DWT model include using a wavelet packet transform, which may represent a wider range of image characteristics. Chang and Kuo (1993), Saito and Coifman (1995) and Laine and Fan (1993) have successfully used the 2D wavelet packet transform for classification of textures, noting that much of the important information may be contained in the middle frequencies. We could also extend the 2DWT model to include more than one explanatory image and, if the explanatory variables are highly correlated, we should use, for example, a principal components regression.

Chapter 7

Conclusions and Future Directions

The main theme of this thesis has been to consider relating one stochastic process to another, where the data set may be one or two-dimensional. In particular, we have considered transfer function models to relate two (possibly) non-stationary time series.

Many methods exist for modelling one process in terms of other explanatory processes. We considered both classical and novel approaches to explain (using data obtained from a collaborator) how the price of a consumer product may drive its market share. In Chapter 3, we found the utility of methods which assume the time series are stationary, such as ARIMA modelling, were limited due to the length of the data which could be considered stationary. Methods which weaken the stationarity assumptions, such as the Nason and Sapatinas (2002) model and the multiple regression Bayesian dynamic linear model (RDLM), were employed with some success; although our results were again limited due to the length of the time series. The RDLM suggested the relationship between price and market share changed over time; hence showing that methods which model the non-stationary features in the data set may be more informative.

On production of new observations from the marketing time series we would perform further multivariate analysis. In particular, we would wish to jointly model products two and five and investigate the possibility of feedback relationships between price and market share; for example, an increase in a product's market share may enable a higher price to be charged for that product.

Occasionally, high correlations occur between the wavelet packet coefficients of a process. Problems may occur when the highly-correlated wavelet packet coefficients are *included as explanatory variables in a regression model. By performing a principal components analysis on a matrix of wavelet packet coefficients we reduced the problems associated with high correlation between explanatory variables.* In Chapter 4, we use the resultant principal components in a regression model and the new model (NWPTpca) is shown to compare favourably with the Nason and Sapatinas (2002) model. Wavelet packet combination plots provided a detailed physical interpretation of some behavioural components in the explanatory time series from our model.

Future work on the NWPTpca model may include improving the initial variable selection technique, where a selection of wavelet packet coefficients are chosen through their correlation with the response variable. It may also be informative to take a NWPT of the response time series and investigate its relationship with a NWPT of the explanatory time series.

The RDLM relates a response time series to one or more explanatory time series, where the regression coefficients may evolve over time. Generally, the RDLM performs well if the relationship between the series is slowly-varying. The RDLM is less efficient if the component time series contain discontinuities or if the relationship between them changes rapidly. We proposed the inclusion of wavelet packet coefficients as explanatory variables in the RDLM to improve its efficiency when modelling data sets exhibiting some forms of discontinuity. In Chapter 5, we evaluate our model (RDLMwp) using two examples; a simulated data set with non-smooth regression coefficients and a data set from Harrison and Reed (1999) with non-stationarity in the component time series. For each data set, the RDLMwp provided large improvements over the standard RDLM in terms of online and retrospective forecasts of the response series.

In future we would like to restrict ourselves to a RDLMwp which uses only real-time wavelet packet coefficients as explanatory variables; online analysis is a key feature of the Bayesian dynamic linear model. It is important to obtain additional results using real and simulated data sets to investigate how widely applicable our model is, i.e. what kind of time series behaviour can the RDLMwp efficiently handle.

Wavelet packet coefficients have been successfully used as explanatory variables in a time series regression model. We designed a two-dimensional extension to the wavelet packet transfer function model (Nason and Sapatinas, 2002) and investigated whether successful results may be obtained when modelling images. In Chapter 6, we took a

2D non-decimated wavelet transform of an explanatory time series and modelled a response image in terms of the resultant 2D non-decimated wavelet coefficients. A simulation study to predict the true response image in the presence of noise indicated we could expect moderate success with this technique in comparison with a simple linear regression model.

The model proposed in Chapter 6 is in the early stages of development. We would wish to investigate, for example, whether the full wavelet packet transform may be more useful.

Appendix A

Map of the British Isles

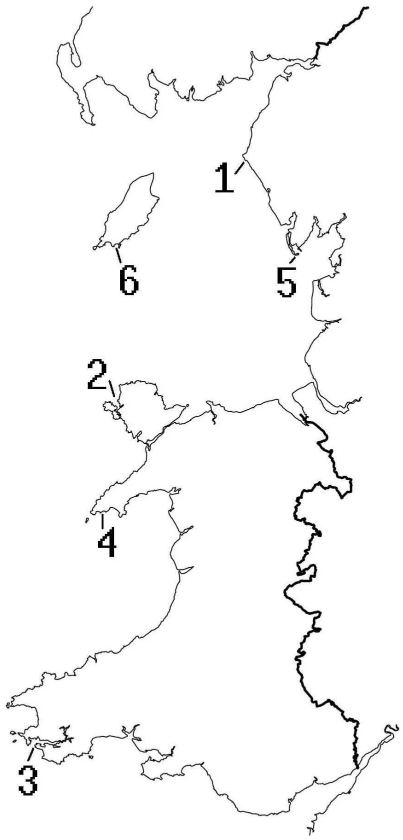

Figure A.1: Location of the six meteorological stations referenced in Chapter 4: 1. St Bees Head 2. Valley 3. Milford Haven 4. Aberdaron 5. Walney Island 6. Ronaldsway. Reproduced from Ordnance Survey map data by permission of the Ordnance Survey ©Crown copyright 2001.

Appendix B

S-Plus Code for Chapter 4

This appendix provides brief documention of the code used in Section 4.4. MCP analysis was implemented using the mcp software (Nason, 1996). The `makewpstRO` function from WaveThresh3 (Nason, 1998) performs the NWPT and initial variable selection (documented at `www.stats.bris.ac.uk/∼makh/makewpstRO.html`).

The function `WindAnalysis` incorporates principal components analysis in the wavelet packet transfer function model of Nason and Sapatanis (2001), as described in Sections 4.2.4 and 4.3. Two arguments are required: the name of the response time series (response) and the name of the explanatory time series (explanatory).

```
WindAnalysis(explanatory, response)
{
# COMPUTE THE NON-DECIMATED WAVELET PACKET TRANSFORM
# REGRESSION OBJECT #
#
data.nwptRO <- makewpstRO(explanatory, response,
 filter.number = 1, family = "DaubExPhase",
 trans = identity, per = 10) #
#
# CARRY OUT PRINCIPAL COMPONENTS REGRESSION #
# DELETE ANY SMALL EIGENVALUES #
#
data.pca <- princomp(data.nwptRO$df[,
```

```
 2:(round(0.1 * length(response)) + 1)], cor = T)
zapped.pca <- data.pca$loadings
len.pca <- sqrt(length(data.pca$loadings))
max.pca <- 0
for(i in 1:len.pca)
max.pca[i] <- max(zapped.pca[, i])
for(j in 1:len.pca)
{
 for(k in 1:len.pca)
  if(Mod(zapped.pca[k, j]) < Mod((max.pca[j]/2)))
zapped.pca[k, j] <- 0
}
data.pca.rotated <- data.frame(as.matrix(data.nwptRO$df[,
2:(round(0.1 * length(response)) + 1)]) %*% as.matrix(zapped.pca))
#
# RETURN OUTPUT #
#
output <- list(response = response, wpstRO = data.nwptRO,
 RotatedData = data.pca.rotated, PCAAnalysis = zapped.pca)
return(output)
}
```

Appendix C

Software Documentation for Chapter 5

Software by Harrison and Reed (1999) is available from the StatLib website[1] for fitting a multiple regression dynamic linear model (RDLM). The `Bts.model.build` function, which constructs the RDLM, requires the following arguments:

1. The forecast variable
   This is the response series, Y_t, in Equation 2.4.

2. The trend
   For a multiple regression DLM we assume the trend is level; we attribute any sustained growth to the regressor variables. We provide the prior mean, standard deviation and associated discount factor of the trend component.

3. The regression variables
   These are the explanatory variables, $X_{t,i}$, contained in the regression vector $F_t^{'}$ in Equation 2.4. They are (usually) standardised to be orthogonal with the intercept term to facilitate comparison of coefficients. We provide the prior mean, standard deviation and associated discount factor of the regression coefficients.

4. The variance model
   We would normally specify the observational variance V to be constant; a

[1] `http://www.mirror.ac.uk/sites/lib.stat.cmu.edu/`

situation often found in practice. We provide a prior estimate of V and specify the degrees of freedom parameter.

The RDLM is modified using the `Bts.model.edit` function. `Bts.model.forecast` calculates one-step ahead forecasts and generates a plot of the standardised residuals, which can be examined for inadequacies in the model. Retrospective analysis is performed by `Bts.model.retro`. `Bts.model.graph` can be used to plot the output from `Bts.model.forecast` and `Bts.model.retro`. For an object of class 'BtsModelForecastOutput' the function plots a graph of the observed response variable Y_t (denoted 'o') against the one-step ahead forecast means f_t (denoted '$\times$') with 2/3 (66%) confidence intervals (denoted '`---`'). For an object of class 'BtsModelRetroOutput' the function plots a graph of Y_t against the retrospective forecast f_t^R with 2/3 confidence intervals. `Bts.model.graph` also graphs the regression parameters and their effect on the response variable (i.e. the variable multiplied by its estimated regression coefficient).

This software was adapted for use in the simulations of Section 5.3.

Appendix D

S-Plus Code for Chapter 6

This appendix provides S-Plus code for the simulation method outlined in Section 6.7. Three arguments are required: the name of the response image (response), the name of the explanatory image (explanatory) and the variance level of the noise added to the response image (sigma). One further argument is optional: the number of iterations (sims).

```
image(response, explanatory, sigma, sims=100)
{
wav.err<-linreg.err<-0
#
# SIMULATION #
#
for(i in 1:sims)
{
StepwiseResults<-response.noise<-LinReg<-explanatory.imwd<-0
response.noise.resp<-formula<-wav<-response.wav.fit<-0
response.linreg.fit<-explanatory.exp<-explanatory.data<-vars<-0
EXPformula<-vector(length=0)

response.noise <- response+rnorm(length(response),0, sigma)

LinReg<-lm(as.vector(response.noise[173:300,173:300])
```

```
~as.vector(explanatory[173:300,173:300]))

explanatory.imwd<-imwd(explanatory[173:300,173:300], type="station",
filter.number = 1, family = "DaubExPhase")
response.noise.resp<-as.vector(response.noise[173:300,173:300])
explanatory.exp<-data.frame(explanatory.imwd$w6L4, explanatory.imwd$w6L1,
explanatory.imwd$w6L2, explanatory.imwd$w6L3, explanatory.imwd$w5L4,
explanatory.imwd$w5L1, explanatory.imwd$w5L2, explanatory.imwd$w5L3,
explanatory.imwd$w4L4, explanatory.imwd$w4L1, explanatory.imwd$w4L2,
explanatory.imwd$w4L3, explanatory.imwd$w3L4, explanatory.imwd$w3L1,
explanatory.imwd$w3L2, explanatory.imwd$w3L3, explanatory.imwd$w2L4,
explanatory.imwd$w2L1, explanatory.imwd$w2L2, explanatory.imwd$w2L3,
explanatory.imwd$w1L4, explanatory.imwd$w1L1, explanatory.imwd$w1L2,
explanatory.imwd$w1L3, explanatory.imwd$w0L4, explanatory.imwd$w0L1,
explanatory.imwd$w0L2, explanatory.imwd$w0L3, explanatory.imwd$w0Lconstant)
explanatory.data<-data.frame(explanatory.exp, response.noise.resp)

StepwiseResults<-stepwise(explanatory.exp, response.noise.resp,size.max=10)
ExpVariables<-vector(length=0)
for(k in 1:length(explanatory.exp))
{
for(i in 1:max(StepwiseResults$size))
if(StepwiseResults$which[i, k]==T)
 ExpVariables<-rbind(ExpVariables, k)
}

WaveletsChosen<-unique(ExpVariables)
for(j in 1:length(WaveletsChosen))
vars[j]<-as.array(paste("X",WaveletsChosen[j], sep=""))
for(j in 1:(length(WaveletsChosen)-1))
EXPformula<- paste(vars[j], "+",EXPformula)
EXPformula<- paste(EXPformula, vars[length(WaveletsChosen)])
```

```
formula<- paste("response.noise.resp ~",EXPformula)

wav<-lm(formula, data=explanatory.data)

response.wav.fit<-as.vector(predict.lm(wav))
response.linreg.fit<-as.vector(predict.lm(LinReg))

wav.err<-wav.err + sum((response[173:300,173:300]-response.wav.fit)^2)
linreg.err<- linreg.err +
sum((response[173:300,173:300]-response.linreg.fit)^2)
}
#
# RESULTS #
#
cat("wavelet MSE\n", round(wav.err/(128*128)/sims, 4) ,"\n" )
cat("linear regression MSE\n",round(linreg.err/(128*128)/sims, 4))
}
```

Bibliography

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

Addison, J.F.D., Hunter, A., Bass, J., & Rebbeck, M. 2000. A neural network version of the measure correlate predict algorithm for estimating wind energy yield. *Published at COMADEM Conference.*

Ansley, C. F. 1979. An algorithm for the exact likelihood of a mixed autoregressive–moving average process. *Biometrika*, **66 (1)**, 59–65.

Antoniadis, A. 1994. Smoothing noisy data with coiflets. *Statistica Sinica*, **4**, 651–678.

Barone, P., Frigessi, A, & Piccioni, M. 1992. *Stochastic Models, Statistical Methods and Algorithms in Image Analysis.* London: Springer.

Bloomfield, P. 2000. *Fourier Analysis of Time Series.* New York: Wiley.

Bonnet, S., Peyrin, F., Turjman, F., & Prost, R. 2000. Tomographic reconstruction using nonseparable wavelets. *IEEE Transactions on Image Processing*, **9(8)**, 1445–1450.

Borup, L., & Nielsen, M. 2002. Nonseparable wavelet packets. *Pages 51–61 of:* Chui, C. K., Schumaker, L. L., & Stckler, J. (eds), *Approximation Theory X: Wavelets, Splines and Applications.* Vanderbilt University Press.

Box, G. E. P., & Jenkins, G. M. 1970. *Time Series Analysis Forecasting and Control.* San Francisco: Holden-Day.

Box, G. E. P., & Newbold, P. 1971. Some comments on a paper of Coen, Gomme and Kendall. *Journal of the Royal Statistical Society (Series A)*, **134 (2)**, 229–240.

Bradley, J.N., Brislawn, C.M., & Hopper, T. 1993. The FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression. *Visual Info. Process. II, SPIE, Orlando, FL.*

Brown, B. G., Katz, R.W., & Murphy, A. H. 1984. Time series models to simulate and forecast wind speeds and wind power. *Journal of Climate and Applied Meteorology*, **23**, 1184–1195.

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

Candès, E. J., & Donoho, D.L. 1999. Ridgelets: a key to higher-dimensional intermittency? *Phil. Trans. R. Soc. Lond. A.*, **357**, 2495–2509.

Chambers, J. M., & Hastie, T. J. 1992. *Statistical Models in S.* California: Wadsworth and Brooks.

Chan, W. Y. T., & Wallis, K. F. 1978. Multiple time series modelling: Another look at the mink-muskrat interaction. *Applied Statistics*, **27 (2)**, 168–175.

Chang, T., & Kuo, C. C. J. 1993. Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on Image Processing*, **2 (4)**, 429–444.

Chatfield, C. 1989. *The Analysis of Time Series: An Introduction.* London: Chapman and Hall.

Chatfield, C. 1997. Forecasting in the 1990s. *Statistician*, **46 (4)**, 461–473.

Cheng, B., & Titterington, D. M. 1994. Neural networks: A review from a statistical perceptive. *Statistical Science*, **9 (1)**, 2–30.

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

Coen, P. G., Gomme, E.D., & Kendall, M. G. 1969. Lagged relationships in economic relationships. *Journal of the Royal Statistical Society (Series A)*, **132**, 133–152.

Coifman, R. R., & Wickerhauser, M. V. 1992. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, **38 (2)**, 713–718.

Cook, N. J. 1986. *The Designer's Guide to Wind Loading of Building Structures.* London: Butterworths.

Cooper, J. D., & Harrison, P. J. 1997. A Bayesian approach to modelling the observed bovine spongiform encephalopathy epidemic. *Journal of Forecasting*, **16**, 355–374.

Cressie, N. 1991. *Statistics for Spatial Data.* New York: Wiley.

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

Derrick, A. 1992. Development of the measure-correlate-predict strategy for site assessment. *Proc 14th British Wind Energy Association Conference*, 259–265.

Diggle, P. J. 1990. *Time Series: A Biostatistical Introduction.* Oxford: Clarendon Press.

Do, M. N., & Vetterli, M. 2001. The finite ridgelet transform for image representation. *IEEE Transactions on Image Processing*, accepted 2001.

Donoho, D.L., & Johnstone, I. M. 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81 (3)**, 425–455.

Durbin, J., & Koopman, S.J. 2001. *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Fama, E. F. 1969. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25 (2)**, 383–417.

Folland, G. B. 1992. *Fourier Analysis and its Applications.* Pacific Grove: Brooks/Cole.

Geladi, P., & Esbensen, K. 1991. Regression on multivariate images: Principal component regression for modelling, prediction and visual diagnostic tools. *Journal of Chemometrics*, **5**, 97–111.

Glasbey, C. A., & Glidewell, S. M. 2001. The application of multivariate statistical methods to NMR imaging. *Computers and Electronics in Agriculture*, **32**, 85–100.

Granger, C. W. J., & Newbold, P. 1986. *Forecasting Economic Time Series.* Orlando: Academic Press Inc.

Griffith, D. A. 1988. *Advanced Spatial Statistics.* Dordrecht: Kluwer Academic Publishers.

Harrison, P. J., & Reed, R. J. 1999. *Dynamic Linear Modelling with SPLUS.* Department of Statistics, University of Warwick, UK.

Harrison, P. J., & Stevens, C. F. 1976. Bayesian forecasting. *Journal of the Royal Statistical Society (Series B)*, **38 (3)**, 205–247.

Harrison, P. J., & West, M. 1987. Practical Bayesian forecasting. *Statistician*, **36 (2/3)**, 115–125.

Harrison, P. J., & West, M. 1997. *Bayesian Forecasting and Dynamic Models.* New York: Springer.

Harvey, A.C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Haslett, J. 1991. Spatial data analysis — challenges. *Statistician*, **41 (3)**, 271–284.

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

Haugh, L. D., & Box, G. E. P. 1977. Identification of dynamic regression (distributed lag) models connecting two time series. *Journal of the American Statistical Association*, **72 (357)**, 121–130.

Hilton, M. L. 1997. Wavelet and wavelet packet compression of electrocardiograms. *IEEE Transactions on Biomedical Engineering*, **44**, 394–402.

Hunt, K., & Nason, G.P. 2001. Wind speed modelling and short-term prediction using wavelets. *Wind Engineering*, **25 (1)**, 55–61.

Jenkins, G. M. 1979. *Practical Experiences with Modelling and Forecasting Time Series.* Jersey: Gwilym Jenkins and Partners [Overseas] Ltd.

Joensen, A., Landberg, L., & Madsen, H. 1999. A new measure-correlate-predict approach for resource assessment. *Proceedings EWEC99, Nice, France.*

Johnstone, I. M., & Silverman, B.W. 1997. Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society (Series B)*, **59**, 319–351.

Jolliffe, I. T. 1986. *Principal Component Analysis.* New York: Springer Verlag.

Joyce-Loebl. 1985. *Image Analysis Principles and Practice.* Gateshead: Joyce-Loebl.

Kammler, D. W. 2000. *A First Course in Fourier Analysis.* Upper Saddle River, N.J.: Prentice Hall.

Kendall, M., & Ord, J. K. 1990. *Time Series.* Sevenoaks: Edward Arnold.

Krzanowski, W. J. 1988. *Principles of Multivariate Analysis.* Oxford: Clarendon Press.

Laine, A, & Fan, J. 1993. Texture classification by wavelet packet signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15 (11)**, 1186–1191.

Learned, R. E., & Willsky, A. S. 1995. A wavelet packet approach to transient signal classification. *Applied and Computational Harmonic Analysis*, **2**, 265–278.

Lun, I. Y. F., & Lam, J. C. 1999. A study of Weibull parameters using long-term wind observations. *Renewable Energy*, **20**, 145–153.

Mallat, S. G. 1989. A theory for multiresolution signal decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11 (7)**, 674–693.

Manwell, J.F., Rogers, A., McGowan, J.G., & Bailey, B. 2001. Assessment of the Southern New England offshore energy resource. *AWEA Conference 2001.*

Mardia, K. V., Kent, J. T., & Bibby, J. M. 1994. *Multivariate Analysis.* London: Academic Press Inc.

Meinhold, R. J., & Singpurwalla, N. D. 1983. Understanding the Kalman filter. *The American Statistician*, **37 (2)**, 123–127.

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

Mulugetta, Y., & Drake, F. 1996. Assessment of solar and wind energy resources in Ethiopia. II: Wind energy. *Solar Energy*, **57 (4)**, 323–334.

Nason, G. P. 1996. *A Measure-Correlate-Predict Package for S-Plus for Windforce Turbines Limited.* Department of Mathematics, University of Bristol, UK.

Nason, G. P. 1998. *WaveThresh3 Software.* Department of Mathematics, University of Bristol, UK.

Nason, G. P., & Sapatinas, T. 2002. Wavelet packet transfer function modelling of non-stationary time series. *Statistics and Computing*, **12(1)**, 45–56.

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

Nason, G. P., & von Sachs, R. 1999. Wavelets in time series analysis. *Phil. Trans. R. Soc. Lond. A*, **357**, 2511–2526.

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

Nason, G. P., Sapatinas, T., & Sawczenko, A. 2001. Wavelet packet modelling of infant sleep state using heart rate data. *Sankhya (Series B)*, **63**, 199–217.

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

Percival, D. B., & Walden, A.T. 2000. *Wavelet Methods for Time Series Analysis.* Cambridge: Cambridge University Press.

Pesquet, J. C., Krim, H., & Carfantan, H. 1996. Time-invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, **44 (8)**, 1964–1970.

Pole, A., West, M., & Harrison, P. J. 1994. *Applied Bayesian Forecasting and Time Series Analysis.* New York: Chapman and Hall.

Portnoff, M. R. 1976. Implementation of the digital phase vocoder using the fast Fourier transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **ASSP-24 (3)**, 243–248.

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

Priestley, M.B. 1965. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society (Series B)*, **27**, 204–237.

Ramirez, R. W. 1985. *The FFT: Fundamentals and Concepts*. Englewood Cliffs, N.J.: Prentice-Hall.

Ramsey, J. B. 1999. Regression over time scale decompositions: A sampling analysis of distributional properties. *Economic Systems Research*, **11 (2)**, 163–183.

Ramsey, J. B., & Lampart, C. 1998. The decomposition of economic relationships by time scale using wavelets: Expenditure and income. *Studies in Nonlinear Dynamics and Econometrics*, **3 (1)**, 23–42.

Ramsey, J. B., & Zhang, Z. 1997. The analysis of foreign exchange data using waveform dictionaries. *Journal of Empirical Finance*, **4**, 341–372.

Ramsey, J. B., Usikov, D., & Zaslavsky, G. M. 1995. An analysis of U.S. stock price behaviour using wavelets. *Fractals*, **3(2)**, 377–389.

Rossi, P.E., & Allenby, G. M. 2000. Statistics and marketing. *Journal of the American Statistical Association*, **95 (450)**, 635–638.

Saito, N., & Coifman, R. R. 1995. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, **5**, 337–358.

Starck, J., Candès, E. J., & Donoho, D. L. 2002. The curvelet transform for image denoising. *Annals of Statistics*, **30**, 784–842.

Stern, R. D., & Coe, R. 1984. A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society (Series A)*, **147 (1)**, 1–34.

Stewart, D. A., & Essenwanger, O. M. 1978. Frequency distributions of wind speed near the surface. *Journal of Applied Meteorology*, **17**, 1633–1642.

Takle, E. S., & Brown, J. M. 1977. Note on the use of Weibull statistics to characterize wind-speed data. *Journal of Applied Meteorology*, **17**, 556–559.

Tsay, R. S. 2000. Time series and forecasting: Brief history and future research. *Journal of the American Statistical Association*, **95 (450)**, 638–643.

Tuller, S. E., & Brett, A. C. 1984. The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis. *Journal of Climate and Applied Meteorology*, **23**, 124–134.

Uitzinger, M.J., Verkuijlen, E., & van Mierlo, B. 1989. On the impractibility of Weibull statistics in the comparison of wind time series. *Wind Engineering*, **13 (4)**, 207–212.

Unser, M. 1995. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*, **4 (11)**, 1549–1560.

Van der Hoven, I. 1957. Power spectrum of horizontal wind speeds spectrum in the frequency range from 0.0007 to 900 cycles per hour. *Journal of Meteorology*, **14**, 160.

Vetterli, M. 2001. Wavelets, approximation and compression. *IEEE Signal Processing Magazine*, **September**, 59–73.

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

Walden, A. T., & Contreras Cristan, A. 1998. The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events. *Proc. R. Soc. Lond. A*, **454**, 2243–2266.

Wallis, K. F. 1977. Multiple time series analysis and the final form of econometric models. *Econometrica*, **45 (6)**, 1481–1497.

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

Woods, J. C., & Watson, S. J. 1997. A new matrix method of predicting long-term wind roses with MCP. *Journal of Wind Engineering and Industrial Aerodynamics*, **66**, 85–94.

Zandonade, E., & Morettin, P. A. 2001. Wavelets in state space models. (submitted 2001).