ELSEVIER

Computational Statistics & Data Analysis 39 (2002) 435–451

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Empirical Bayes approach to block wavelet function estimation

Felix Abramovich[a], Panagiotis Besbeas[b], Theofanis Sapatinas[c,*]

[a] *Department of Statistics and Operations Research, Tel Aviv University, Ramat Aviv 69978, Israel*
[b] *Institute of Mathematics and Statistics, University of Kent at Canterbury, Vent CT2 7 NF, UK*
[c] *Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, Nicosia CY 1678, Cyprus*

Received 1 December 2000

Abstract

Wavelet methods have demonstrated considerable success in function estimation through term-by-term thresholding of the empirical wavelet coefficients. However, it has been shown that grouping the empirical wavelet coefficients into blocks and making simultaneous threshold decisions about all the coefficients in each block has a number of advantages over term-by-term wavelet thresholding, including asymptotic optimality and better mean squared error performance in finite sample situations. An empirical Bayes approach to incorporating information on neighbouring empirical wavelet coefficients into function estimation that results in block wavelet shrinkage and block wavelet thresholding estimators is considered. Simulated examples are used to illustrate the performance of the resulting estimators, and to compare these estimators with several existing non-Bayesian block wavelet thresholding estimators. It is observed that the proposed empirical Bayes block wavelet shrinkage and block wavelet thresholding estimators outperform the non-Bayesian block wavelet thresholding estimators in finite sample situations. An application to a data set that was collected in an anaesthesiological study is also presented.

Keywords: Empirical Bayes; Block thresholding; Maximum likelihood estimation; Non-parametric regression; Wavelet transform

1. Introduction

Over the last decade, the non-parametric regression literature has been dominated by *non-linear wavelet* methods. These methods are based on the idea of thresholding,

* Corresponding author.
E-mail address: T.Sapatinas@ucy.ac.cy (T. Sapatinas).

0167-9473/02/$ - see front matter
PII: S 0167-9473(01)00085-8

which typically amounts to individual assessment of every empirical wavelet coefficient. If a coefficient is sufficiently large in magnitude, that is if its magnitude exceeds a predetermined threshold, then the corresponding term in the empirical wavelet expansion is retained (or shrunk towards zero); otherwise it is omitted. The resulting estimators are typically implemented through fast algorithms which makes them very appealing in practice.

Since the seminal papers by Donoho and Johnstone (1994) and Donoho et al. (1995), a range of alternative wavelet methods has been developed. This range includes Bayesian term-by-term wavelet shrinkage and wavelet thresholding estimators. (To introduce terminology, a *shrinkage* rule shrinks empirical wavelet coefficients to zero, whilst a *thresholding* rule shrinks and, in addition, sets to zero all empirical wavelet coefficients below a certain level.) Such estimators have been shown to be effective and have been argued to be less ad hoc than their classical counterparts. Extensive reviews and descriptions of the various classical and Bayesian term-by-term wavelet schemes can be found in, for example, the books by Ogden (1997) and Vidakovic (1999), the papers appearing in the edited volume by Müller and Vidakovic (1999), and the review papers by Antoniadis (1997), Vidakovic (1998b) and Abramovich et al. (2000).

Hall et al. (1997), Hall et al. (1998, 1999), Cai (1999) and Cai and Silverman (2001), amongst others, described another approach based on *block* thresholding. In this approach, the empirical wavelet coefficients are thresholded in blocks rather than individually. The resulting estimators have been found to possess a number of advantages over term-by-term wavelet thresholding estimators, including asymptotic optimality and better mean squared error performance in finite sample situations. The aim of this paper is to build on the advantages of block thresholding. We propose an empirical Bayes approach to incorporating information on neighbouring coefficients into wavelet function estimation that results in block wavelet shrinkage and block wavelet thresholding estimators. Simulated examples are used to illustrate the performance of the resulting estimators, and to compare these estimators with several existing non-Bayesian block wavelet thresholding estimators. It is shown that the proposed empirical Bayes block wavelet shrinkage and block wavelet thresholding estimators outperform the non-Bayesian block wavelet thresholding estimators in finite sample situations.

The paper is organized as follows. In Section 2 we introduce the wavelet paradigm to the classical non-parametric regression setting. In Section 3 we introduce the prior model imposed on the wavelet coefficients of the unknown response function, and obtain the posterior-based block wavelet shrinkage and block wavelet thresholding estimators. We also discuss an empirical Bayes approach to estimating the hyperparameters of the prior model. In Section 4, we provide simulated examples to illustrate the resulting estimators, and we compare these with several existing non-Bayesian block wavelet thresholding estimators. We also present an application to a data set that was collected in an anaesthesiological study. Some concluding remarks are made in Section 5.

2. Wavelet regression

Consider the classical non-parametric regression setting

$$y_i = g(t_i) + \sigma\varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $t_i = i/n$, $n = 2^J$ for some positive integer J, ε_i are independent and identically distributed $\mathrm{N}(0,1)$ random variables and the noise level σ may, or may not, be known. The problem is to estimate the underlying function g from the observations $\mathbf{y} = (y_1, \ldots, y_n)'$ without assuming any particular parametric structure on its form.

The wavelet approach to this problem is easily described. Given a suitable wavelet basis and *primary* resolution level $j_0 \geqslant 0$, the discrete wavelet transform (DWT) of $\mathbf{y}$ gives rise to an n-dimensional vector $\hat{\mathbf{d}}$ consisting of what are known as the *empirical scaling* coefficients $\hat{c}_{j_0k}$ ($k = 0, \ldots, 2^{j_0} - 1$) and the *empirical wavelet* coefficients $\hat{d}_{jk}$ ($j = j_0, \ldots, J-1; k = 0, 1, \ldots, 2^j - 1$). In practice, the DWT (and its inverse (IDWT)) may be performed through a computationally fast algorithm developed by Mallat (1989) that requires only $O(n)$ operations. Due to the orthogonality of the DWT, it follows from (1) that

$$\hat{c}_{j_0k} = c_{j_0k} + \sigma\varepsilon_{j_0k}, \quad k = 0, 1, \ldots, 2^{j_0} - 1, \tag{2}$$

$$\hat{d}_{jk} = d_{jk} + \sigma\varepsilon_{jk}, \quad j = j_0, \ldots, J-1, \ \ k = 0, 1, \ldots, 2^j - 1, \tag{3}$$

where the ε_{jk} are themselves independent and identically distributed $\mathrm{N}(0,1)$ random variables, and the c_{j_0k} and d_{jk} are, respectively, the true *scaling* and *wavelet* coefficients of the (unknown) vector of function values $\mathbf{g} = (g(t_1), \ldots, g(t_n))'$.

The problem can now be cast as that of estimating the true wavelet coefficients d_{jk} from the empirical wavelet coefficients $\hat{d}_{jk}$. The sparseness of the wavelet expansion comes here to the foreground. Most of the information about the underlying function g in the d_{jk} is concentrated in few large wavelet coefficients while small wavelet coefficients can be attributed to the presence of noise which uniformly contaminates all wavelet coefficients. We would thus obtain a reasonable estimate of g if we could extract the wavelet coefficients of largest magnitude accurately, even if we set the rest to zero. This is the idea behind term-by-term wavelet thresholding. Here, each empirical wavelet coefficient $\hat{d}_{jk}$ is compared with a predetermined threshold and retained (or shrunk towards zero) if its magnitude exceeds the threshold; otherwise it is discarded. This approach achieves a trade-off between the variance and bias contributions to the mean squared error. However, the trade-off is not optimal and results in biased estimators (too many terms are removed from the empirical wavelet expansion) with sub-optimal L_2-risk rates (see, for example, Hall et al., 1998, 1999).

One way of increasing estimation precision is by utilising information on neighbouring empirical wavelet coefficients. In other words, at each resolution level j, the empirical wavelet coefficients $\hat{d}_{jk}$ could be thresholded in *blocks* (or groups) rather than individually. As a result, the amount of information available from the data for estimating the 'average' empirical wavelet coefficient within a block, and making a decision about retaining or discarding it, would be an order of magnitude larger than the case of a term-by-term threshold rule. This would allow threshold decisions to be made more accurately (see, for example, Hall et al., 1997; Hall et al., 1998, 1999; Cai, 1999 and Cai and Silverman, 2001).

In the next section, we propose an empirical Bayes approach to incorporating information on neighbouring coefficients into wavelet function estimation that results in *level-dependent* block wavelet shrinkage and block wavelet thresholding estimators.

3. Empirical Bayes block wavelet estimators

3.1. The prior model

Consider the sequence model given by (2) and (3). A typical characteristic of the Bayesian term-by-term wavelet shrinkage and wavelet thresholding estimation methods is the use of different but mostly *independent* priors on the wavelet coefficients (see, for example, Chipman et al., 1997; Abramovich et al., 1998; Clyde et al., 1998, Johnstone and Silverman, 1998; Vidakovic, 1998a; Abramovich and Sapatinas, 1999; Clyde and George, 1999, 2000). It is reasonable to expect that a model allowing for correlations between neighbouring wavelet coefficients would be more parsimonious. Accordingly, we suggest placing priors on *blocks* of wavelet coefficients rather than on each one individually.

More precisely, at each resolution level j ($j=j_0,\dots,J-1$), the wavelet coefficients d_{jk} are grouped into m_j *non-overlapping* blocks, b_{jK} ($K=1,\dots,m_j$), of length l_j. (The choice of l_j will be discussed in Section 4.) For each of the blocks b_{jK}, we assume the following prior model:

$$b_{jK}|\gamma_{jK} \sim \mathrm{N}(0, \gamma_{jK} V_j), \tag{4}$$

$$\gamma_{jK} \sim \text{Bernoulli } (\pi_j), \tag{5}$$

independently for $K=1,\dots,m_j$, where $0 \leqslant \pi_j \leqslant 1$. We also assume block independence across the different resolution levels j ($j=j_0,\dots,J-1$). According to this prior model, a block is either zero with probability $1-\pi_j$ or, with probability π_j, multivariate normally distributed with mean zero and variance–covariance matrix V_j. Thus, our prior model allows for the fact that a wavelet coefficient is more likely to contain signal if neighboring wavelet coefficients do also. To complete the prior formulation, we assume vague priors for the scaling coefficients c_{j_0k} ($k=0,1,\dots,2^{j_0}-1$) and in particular we take $c_{j_0k} \sim \mathrm{N}(0,\varepsilon)$, $\varepsilon \to \infty$.

Through the covariance matrix V_j, we may allow for various forms of dependency between the wavelet coefficients within each block. Since the correlation between two wavelet coefficients generally weakens as the distance between them increases, we choose V_j to be the $l_j \times l_j$ matrix with elements

$$V_j[k,l] = \tau_j^2 \rho_j^{|k-l|} \quad \text{where } |\rho_j| < 1, \;\; k,l=1,\dots,l_j. \tag{6}$$

A similar choice of covariance structure has been employed by Vidakovic and Müller (1995) and Vannucci and Corradi (1999a,b) in a different context.

The above model can be regarded as an extension of the prior mixture model of a univariate normal distribution and a point mass at zero (i.e. $l_j = 1$), used in Bayesian term-by-term wavelet function estimation by, for example, Clyde et al. (1998), Abramovich et al. (1998), Johnstone and Silverman (1998), Abramovich and Sapatinas (1999) and Clyde and George (1999, 2000).

3.2. Posterior-based block shrinkage and block thresholding estimators

At each resolution level j ($j = j_0, \ldots, J-1$), we now consider the corresponding non-overlapping blocks $\hat{b}_{jK}$ ($K = 1, \ldots, m_j$) of the empirical wavelet coefficients $\hat{d}_{jk}$. Combining the prior model (4) and (5) with the likelihood from $\hat{b}_{jK} \sim \mathrm{N}(b_{jK}, \sigma^2 I)$, and averaging over all possible γ_{jK}, yields the marginal posterior distribution

$$b_{jK} | \hat{b}_{jK} \sim \frac{1}{1+O_{jK}} N(A_j \hat{b}_{jK}, \sigma^2 A_j) + \frac{O_{jK}}{1+O_{jK}} \delta(\mathbf{0}), \tag{7}$$

where $\delta(\mathbf{0})$ is a vector of point masses at zero, $A_j = (\sigma^2 V_j^{-1} + I)^{-1}$ and O_{jK} is the posterior odds ratio that $\gamma_{jK} = 0$ versus $\gamma_{jK} = 1$, given by

$$O_{jK} = \frac{1-\pi_j}{\pi_j} \left(\frac{\det(V_j)}{\sigma^{2l_j} \det(A_j)} \right)^{1/2} \exp\left\{ -\frac{\hat{b}'_{jK} A_j \hat{b}_{jK}}{2\sigma^2} \right\}. \tag{8}$$

For the jKth block, define the vector $\hat{d}^{\star}_{jK} = A_j \hat{b}_{jK}$ and its elements $\hat{d}^{\star}_{jk}$. Applying different losses, one can generate various block shrinkage and block thresholding estimators. In particular, we have:

- *Posterior mean* (L_2*-loss*): it is straightforward to show that the posterior mean of b_{jK} is given by

$$\tilde{b}_{jK} = E(b_{jK} \mid \hat{b}_{jK}) = \frac{1}{1+O_{jK}} \hat{d}^{\star}_{jK}. \tag{9}$$

  This is a *block* (non-linear) shrinkage rule where each empirical wavelet coefficient within a block is shrunk by the same shrinkage factor depending on all coefficients within the block.
- *Marginal posterior median* (L_1*-loss*): in this case, for the posterior distribution in (7), the marginal posterior distribution of d_{jk} is given by

$$d_{jk} \mid \hat{b}_{jK} \sim \frac{1}{1+O_{jK}} N(\hat{d}^{\star}_{jk}, \sigma^2 A_j[j,j]) + \frac{O_{jK}}{1+O_{jK}} \delta(0),$$

  where $\delta(0)$ is a point mass at zero and $A_j[j,j]$ is the entry that appears on the diagonal of A_j (they are the same for all k). Hence, following the arguments of Abramovich et al. (1998) and Abramovich and Sapatinas (1999) in the Bayesian term-by-term wavelet function estimation framework, the marginal posterior median of d_{jk} has the following closed form:

$$\tilde{d}_{jk} = \text{median}\ (d_{jk} \mid \hat{b}_{jK}) = \text{sign}(\hat{d}^{\star}_{jk}) \max(0, \zeta_{jK}), \tag{10}$$

where

$$\zeta_{jK} = |\hat{d}^{\star}_{jk}| - \sigma \left(A_j[j,j]\right)^{1/2} \Phi^{-1}\left(\frac{1+\min(O_{jK},1)}{2}\right)$$

and Φ is the standard normal cumulative distribution function. This is an *individual* thresholding rule where each empirical wavelet coefficient is thresholded utilising information about neighbouring coefficients within a block.

- *Hypothesis testing* (0/1-*loss*): adjusting the Bayes Factor (BF) procedure of Vidakovic (1998a) in the Bayesian term-by-term wavelet function estimation framework to our model, test

  $$\mathrm{H}_0 : b_{jK} = 0 \quad \text{versus} \quad \mathrm{H}_1 : b_{jK} \neq 0$$

  and reject H_0 if the posterior odds ratio $O_{jK} = P(\mathrm{H}_0|\hat{b}_{jK})/P(\mathrm{H}_1|\hat{b}_{jK}) < 1$, where O_{jK} is defined in (8). This implies

  $$\tilde{b}_{jK} = \begin{cases} \hat{b}_{jK} & \text{if } O_{jK} < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

  This is a *block* thresholding rule where the *whole* block of empirical coefficients is thresholded.

Finally, due to the vague priors imposed on the scaling coefficients c_{j_0k} and using (2), any of the above losses will result in estimating c_{j_0k} by their empirical counterparts $\hat{c}_{j_0k}$. Hence, estimates of the values g at the sampling points can be obtained by inverting the DWT on the vector consisting of both the empirical scaling coefficients and the shrunken or thresholded empirical wavelet coefficients, obtained from one of (9), (10), or (11). We call the resulting estimators of g as the *BlockPostMean*, *BlockPostMed* and *BlockBF* estimators, respectively.

3.3. An empirical Bayes approach to estimating the hyperparameters

In order to apply the Bayesian block shrinkage and block thresholding estimators described in the previous section, one needs to specify the hyperparameters $\pi_j,\ \tau_j,\ \rho_j\ (j = j_0,\ldots,J-1)$ and σ. Ideally, they could be obtained from some prior information about, for example, the regularity of the unknown function g. Such an approach has been considered in Abramovich et al. (1998) and Abramovich and Sapatinas (1999) for independent priors, and extension to our model is quite straightforward. In practice, however, one can use an empirical Bayes approach to estimating the hyperparameters in order to get a completely data-based procedure (see, for example, Johnstone and Silverman, 1998; Clyde and George, 1999, 2000).

In this paper, we have used the latter approach to obtain marginal maximum likelihood estimates of the hyperparameters π_j, τ_j and ρ_j *given* σ. More specifically, at each resolution level j, since the marginal distribution of the empirical blocks $\hat{b}_{jK}$ is a mixture of two multivariate normal distributions, the marginal log-likelihood function

ℓ_j is proportional to

$$\ell_j(\pi_j, \tau_j, \rho_j, \sigma) \propto \sum_{K=1}^{m_j} \log \left\{ \pi_j (\det(B_j))^{-1/2} \exp\left(-\frac{1}{2} \hat{b}'_{jK} B_j^{-1} \hat{b}_{jK} \right) \right.$$
$$\left. + (1-\pi_j)\sigma^{-l_j} \exp\left(-\frac{1}{2\sigma^2} \hat{b}'_{jK} \hat{b}_{jK} \right) \right\}, \quad (12)$$

where $B_j = \sigma^2 I + V_j$ and V_j is defined in (6). A system of maximum likelihood equations can be formed by differentiating (12) with respect to the parameters, and setting the resulting expressions equal to zero. However, these equations cannot be solved explicitly and so numerical iterative techniques have to be adopted.

The usual iterative procedure for obtaining the maximum likelihood estimates (MLEs) in mixture distributions is the EM algorithm of Dempster et al. (1977). For independent priors, by applying the EM algorithm, one can find the MLEs for π_j and τ_j given σ (as in Johnstone and Silverman, 1998 and Clyde and George, 1999) or the MLEs for π_j, τ_j and σ (as in Clyde and George, 2000). However, in our case, the M-step of the EM algorithm is not explicit, resulting somewhat in an impractical application of the procedure.

In this paper, we have considered the following approach. We have estimated σ by the *median absolute deviation* (as suggested by Donoho and Johnstone (1994) and usually applied in practice)

$$\hat{\sigma} = \frac{\text{median}(|\{\hat{d}_{J-1,k}: k = 0, 1, \ldots, 2^{J-1}-1\}|)}{0.6745}, \quad (13)$$

and have minimized $-\ell_j(\pi_j, \tau_j, \rho_j, \hat{\sigma})$ with respect to π_j, τ_j and ρ_j directly. The log-likelihood function was reparametrised with

$$\pi_j = \frac{1}{1+\exp(-\theta_{1j})}, \quad \tau_j = |\theta_{2j}| \quad \text{and} \quad \rho_j = \frac{2}{\pi} \arctan(\theta_{3j})$$

so that the parameter estimates would lie in the ranges

$$0 \leqslant \hat{\pi}_j \leqslant 1, \quad \hat{\tau}_j \geqslant 0 \quad \text{and} \quad -1 < \hat{\rho}_j < 1.$$

The algorithm that we have used for the minimization of $-\ell_j$ is the Nelder–Mead simplex search method which does not require first derivatives of $-\ell_j$ (see, for example, Everitt 1987, pp. 16–20).

The advantages of fixing the value of σ to, for example, the median absolute deviation in (13) are three-fold. First, it reduces the dimension of the parameter space. Second, it eliminates all the singularities in the likelihood function. Third, simulations indicated that the choice of initial values is not so crucial in finding the MLEs for π_j, τ_j and ρ_j. In particular, a small scale simulation study was carried out in order to examine the importance of starting values for the above estimation problem. It was found that the convergence from different starting values for π_j, τ_j and

ρ_j was always to be of the same parameter estimates, a result that held for various sets of random sample from the prior mixture model (4) and (5).

4. Applications and comparisons

The purpose of this section is to illustrate the performance of the proposed empirical Bayes block wavelet shrinkage and block wavelet thresholding estimators, and to compare these estimators with several existing non-Bayesian block wavelet thresholding estimators. Simulated samples and a data set collected in an anaesthesiological study have been used for this purpose. The computational algorithms related to wavelet analysis were performed using the WaveLab software that is freely available from `http://www-stat.stanford.edu/software/software.html`. The entire study was carried out using the Matlab programming environment.

4.1. Simulation study

The results of the simulation study will now be presented, with the remainder of this section devoted to the discussion of these results. We compare the BlockPostMean, BlockPostMed and BlockBF estimators with the non-overlapping *BlockJS* estimator of Cai (1999) and the overlapping *NeighBlock* and *NeighCoeff* estimators of Cai and Silverman (2001). These estimators have been shown to be superior in mean squared error to the *VisuShrink* (see Donoho and Johnstone, 1994), *SureShrink* (see Donoho and Johnstone, 1995) and *Translation Invariant* (see Coifman and Donoho, 1995) estimators, which have become very popular in term-by-term wavelet thresholding. The BlockJS estimator is a blockwise James–Stein rule, with properly chosen block size and threshold value. The NeighBlock and NeighCoeff estimators are, actually, variants of the BlockJS estimator, where the empirical wavelet coefficients are considered in overlapping blocks. We refer to the papers of Cai (1999) and Cai and Silverman (2001) for more details.

In this simulation study, we evaluate the various wavelet estimators using Daubechies's compactly supported wavelets *Symmlet 8* (see Daubechies, 1992, p. 198) and *Coiflet 3* (see Daubechies, 1992, p. 258), and primary resolution levels $j_0 = 3$ and 5. We have considered the *Blocks*, *Bumps*, *HeaviSine* and *Doppler* functions of Donoho and Johnstone (1994, 1995), which constitute standard tests for wavelet estimators. These functions are supposed to caricature spatially variable signals arising in a number of scientific fields, including imaging and spectroscopy. For each test function, $M = 100$ samples were generated by adding independent random noise $\epsilon \sim \mathrm{N}(0, \sigma^2)$ to $n =$ 256 (small sample size), 512 (moderate sample size) and 1024 (large sample size) equally spaced points on [0,1]. The value of σ was taken to correspond to the values 3 (high noise level), 5 (moderate noise level) and 7 (low noise level) for the root signal-to-noise ratio (RSNR)

$$\mathrm{RSNR}(g, \sigma) = \frac{\sqrt{(1/n)\sum_{i=1}^{n}(g(t_i) - \bar{g})^2}}{\sigma}, \quad \text{where } \bar{g} = \frac{1}{n}\sum_{i=1}^{n} g(t_i).$$

The goodness-of-fit for an estimator $\hat{g}$ of g was measured by its average mean squared error (AMSE) from the M simulations, defined as

$$\mathrm{AMSE}(g) = \frac{1}{nM} \sum_{m=1}^{M} \sum_{i=1}^{n} (\hat{g}_m(t_i) - g(t_i))^2;$$

its average mean absolute deviation (AMAD) from the M simulations, defined as

$$\mathrm{AMAD}(g) = \frac{1}{nM} \sum_{m=1}^{M} \sum_{i=1}^{n} |\hat{g}_m(t_i) - g(t_i)|;$$

and its average maximal absolute deviation (AMXD) from the M simulations, defined as

$$\mathrm{AMXD}(g) = \frac{1}{M} \sum_{m=1}^{M} \max_{1 \leqslant i \leqslant n} |\hat{g}_m(t_i) - g(t_i)|.$$

In order to examine the effect of the block lengths on the numerical performance of the proposed wavelet estimators, we performed a preliminary simulation study. In particular, we considered two choices of block length at each resolution level j, namely $l_j = j$ and $l_j = 2^{[\log_2(Cj)]}$, where $2^{[x]}$ is the largest dyadic integer smaller than or equal to x. For the values of C that we examined, the former choice resulted in longer blocks and it was found to be less adequate than the latter choice, whose block lengths evenly divide the empirical wavelet coefficients into non-overlapping blocks. Furthermore, we found that a value of $C \in [0.4, 0.7]$ was the most reasonable choice uniformly across various test functions, sample sizes and RSNRs. This finding is consistent with the remark of Hall et al. (1999) that the selection of block length is not critical if it is chosen within an appropriate range of values. Accordingly, and somewhat arbitrarily, we have taken $C = 0.5$ and suggest $l_j = 2^{[\log_2(j/2)]}$ to be the reference of block length at each resolution level j for the proposed empirical Bayes block wavelet shrinkage and block wavelet thresholding estimators. Thus, there will be 8 blocks of length 1 on the 3rd level, 8 blocks of length 2 on the 4th level, 16 blocks of length 2 on the 5th level, 32 blocks of length 2 on the 6th level, 64 blocks of length 2 on the 7th level, 64 blocks of length 4 on the 8th level, 128 blocks of length 4 on the 9th level and so on.

For brevity, we only report in detail the results for the *Bumps* function using *Symmlet 8*, $j_0 = 3$ and AMSE. Different combinations of test functions, wavelets, primary resolution levels and goodness-of-fit measures yield basically similar results. Fig. 1 contains the results of the simulation study. As observed in the figure, the BlockPostMean, BlockPostMed and BlockBF estimators produced estimates with smaller AMSE than the BlockJS, NeighBlock and NeighCoeff estimators, uniformly across the various combinations of signal morphology and RSNR. Amongst the non-Bayesian block wavelet thresholding estimators, the NeighCoeff estimator always produced the smallest AMSE whilst the BlockJS estimator always produced the largest. (This is consistent with the conclusions drawn by Cai and Silverman, 2001.) We note that each of the BlockPostMean, BlockPostMed and BlockBF estimators produced less variable estimates than the BlockJS, NeighBlock and NeighCoeff

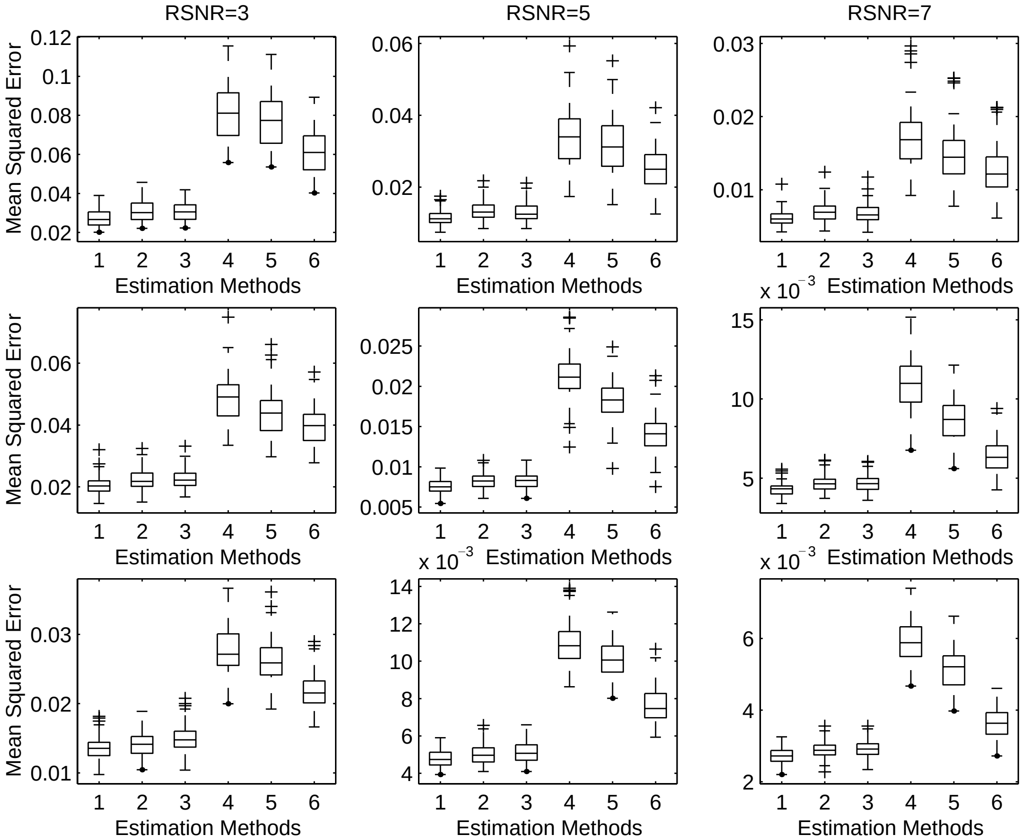

Fig. 1. Boxplots of 100 simulation results for the *Bumps* function for all nine combinations of signal points (top: 256; middle: 512; bottom: 1024) and RSNRs (left: 3; middle: 5; right: 7). In each panel, there are six boxplots indicating the mean squared error, from left to right, for the estimates produced by (1) BlockPostMean, (2) BlockPostMed, (3) BlockBF, (4) BlockJS, (5) NeighBlock and (6) NeighCoeff.

estimators in every case. Quantitatively, the BlockPostMed and BlockBF estimators are almost identical and both appear to be slightly outperformed by the BlockPost-Mean estimator in every case.

Table 1 shows the mean and standard deviation of the CPU times involved in computing 100 estimates for the *Bumps* function by each wavelet estimator and sample size. As anticipated, the BlockJS, NeighBlock and NeighCoeff estimators are superior in terms of CPU time to the BlockPostMean, BlockPostMed and BlockBF estimators. This is due to the maximisation step described in Section 3.3 that can be computationally expensive, depending on the number of iterations and the convergence criterion used. However, with the fast computing environments that exist nowadays, computing time is not a major issue.

To illustrate that the proposed empirical Bayes block wavelet shrinkage and block wavelet thresholding estimators are appealing visually as well as quantitatively, we present in Fig. 2 a noisy *HeaviSine* function sampled at $n = 1024$ equally spaced

Table 1
Mean and standard deviation (in brackets) of the CPU times involved in computing 100 estimates for the *Bumps* function by the six wavelet estimators (BlockJS, NeighBlock, NeighCoeff, BlockPostMean, BlockPostMed and BlockBF) and the three sample sizes ($n = 256$, 512 and 1024)

Method	CPU time ($n = 256$)	CPU time ($n = 512$)	CPU time ($n = 1024$)
BlockJS	0.0181 (0.0042)	0.0272 (0.0035)	0.0345 (0.0026)
NeighBlock	0.0309 (0.0051)	0.0619 (0.0032)	0.1238 (0.0022)
NeighCoeff	0.0289 (0.0048)	0.0608 (0.0033)	0.1221 (0.0024)
BlockPostMean	4.7554 (0.7123)	6.4543 (0.7256)	8.658 (0.7434)
BlockPostMed	4.8763 (0.7134)	6.5324 (0.7265)	8.7324 (0.7543)
BlockBF	4.6563 (0.7012)	6.3787 (0.7123)	8.5472 (0.7234)

points on $[0,1]$ with $\mathrm{RSNR} = 7$. Fig. 3 displays the reconstructions obtained from one simulation using the six wavelet estimators. All the wavelet estimators recover the underlying function reasonably well. The BlockPostMean, BlockPostMed and BlockBF estimators give a better reconstruction of the discontinuities in the *HeaviSine* function than the BlockJS, NeighBlock and NeighCoeff estimators. At the same time, although they appear to be slightly noisier over the regions where the underlying function is smooth, they are not noisy enough to be visually unpleasant.

4.2. Inductance plethysmography data

We further illustrate the performance of the above wavelet estimators using a data set from anaesthesiology collected by inductance plethysmography. The recordings were made by the Department of Anaesthesia at the Bristol Royal Infirmary and measure the flow of air during breathing. The same data set has been analysed in Nason (1996) and Abramovich et al. (1998). We refer to these papers for more details.

Fig. 4 shows a section of plethysmograph recording lasting approximately 80 s ($n = 4096$ signal points). The two main sets of regular oscillations correspond to normal breathing. The disturbed behaviour in the centre of the plot, where the normal breathing pattern disappears, corresponds to the patient vomiting. Fig. 5 shows the various reconstructions of the plethysmograph recording. The recommended block length $l_j = 2^{[\log_2(j/2)]}$ for the proposed BlockPostMean, BlockPostMed and BlockBF

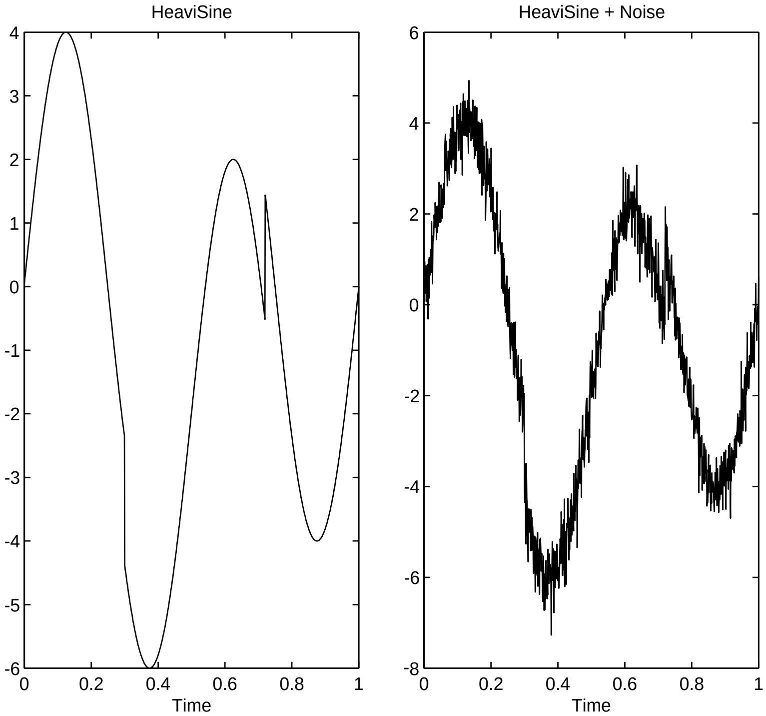

Fig. 2. (Left): the *HeaviSine* function based on 1024 signal points; (Right): a noisy version of the *HeaviSine* function with RSNR = 7.

Table 2
Highest peak value (first peak) in the data and the corresponding function estimates shown in Fig. 5 for the inductance plethysmography recording of Fig. 4

Plethysmography data	0.8472
BlockPostMean	0.8433
BlockPostMed	0.8433
BlockBF	0.8433
BlockJS	0.8268
NeighCoeff	0.8249
NeighBlock	0.8214

estimators was used. As observed in the figure, the BlockJS and NeighBlock estimators remove the noise but tend to attenuate the peaks, whilst the NeighCoeff estimator contains some high frequency effects in the region near time 0.8 and also attenuates the peaks. On the other hand, although the BlockPostMean, BlockPostMed and BlockBF estimators do not eliminate completely the large fluctuation in the region near time 0.8, they produce a satisfying smooth fit without the attenuation. This is apparent, for example, from the heights of the first peak tabulated in Table 2. It

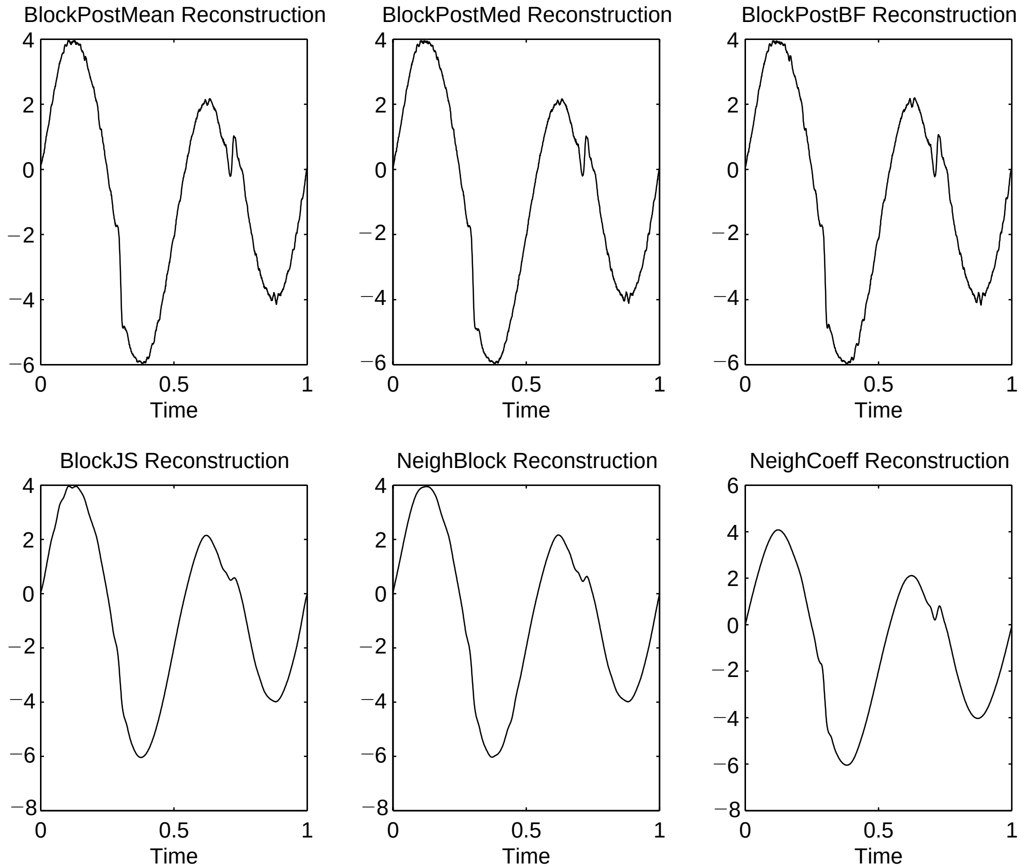

Fig. 3. Reconstructions of the *HeaviSine* function from one simulation based on 1024 signal points and RSNR = 7 for the estimates produced by, from left to right, (top) BlockPostMean, BlockPostMed and BlockBF, (bottom) BlockJS, NeighBlock and NeighCoeff.

is interesting to note that the heights of the first peak value yielded by the Block-PostMean, BlockPostMed and BlockBF estimators are identical. This common value is also equal to two decimal places to the value given by the *BayesThresh* method of Abramovich et al. (1998), in which adjustment of one of the hyperparameters of their prior distribution was necessary (see Eq. (9) and Table 1 in this paper for more details). The main point here is that the BlockPostMean, BlockPostMed and BlockBF estimators automatically produce estimates similar to that obtained by BayesThresh after ad hoc tuning of the hyperparameters.

5. Concluding remarks

An empirical Bayes approach to incorporating information on neighbouring empirical wavelet coefficients into function estimation that results in block wavelet

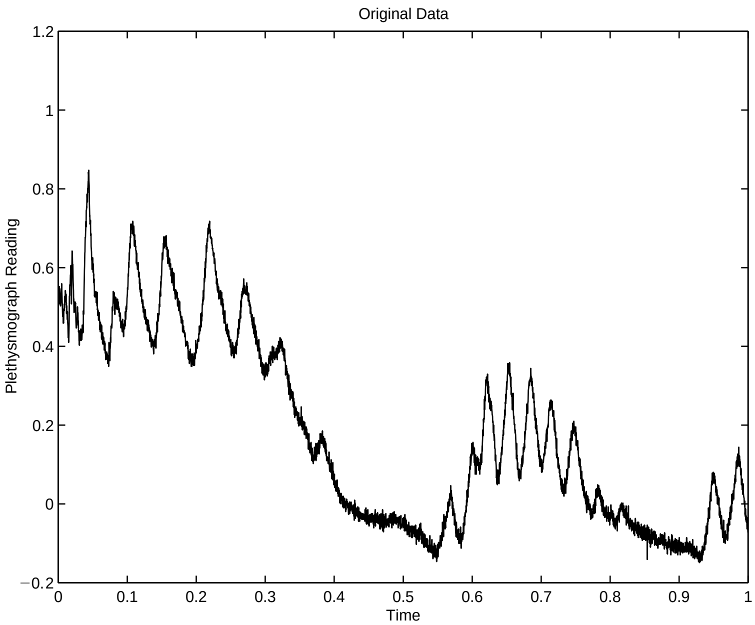

Fig. 4. Section of an inductance plethysmography recording lasting approximately 80 s.

shrinkage and block wavelet thresholding estimators has been discussed. A simulation study has been conducted to illustrate the resulting wavelet estimators, and these estimators have been compared with several existing non-Bayesian block wavelet thresholding estimators. An application to a data set collected in an anaesthesiological study has also been presented.

It has been demonstrated that, with an adequate choice of block length at each resolution, the empirical Bayes block wavelet shrinkage and block wavelet thresholding estimators are superior to several existing non-Bayesian block wavelet thresholding estimators in finite sample situations. However, due to the maximization inherent in their methodology described in Section 3.3, they are computationally more expensive. The proposed empirical Bayes block wavelet shrinkage and block wavelet thresholding estimators could been presented as a possibly useful addition to the growing range of wavelet-based function estimation tools.

The position of the blocks has a significant effect on the performances of the proposed estimators. Although not presented here, by averaging over different block centers, the resulting empirical Bayes block wavelet shrinkage and block wavelet thresholding estimators have been found to have superior numerical performances, at the cost of higher computational complexity. This technique was also used for the classical block wavelet thresholding estimators discussed in Hall et al. (1997), Cai (1999) and Cai and Silverman (2001) to show improvements in mean squared error over the standard position of the blocks. However, we believe that, by first choosing the *positions* and the *lengths* of the blocks adaptively and then applying our empirical Bayes wavelet estimators to the resulting blocks will produce better

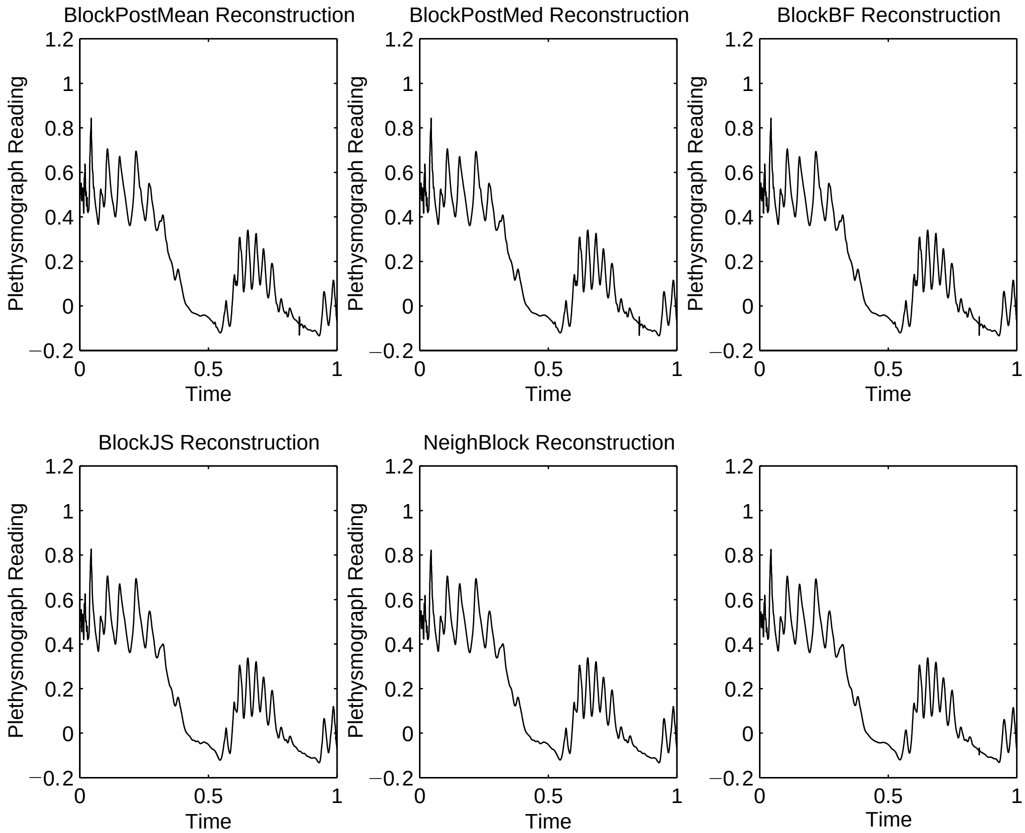

Fig. 5. Smooth estimates of the inductance plethysmography recording shown in Fig. 4 obtained by, from left to right, (top) BlockPostMean, BlockPostMed and BlockBF (bottom) BlockJS, NeighBlock and NeighCoeff.

estimates of the unknown response function. This is an interesting topic for further research we hope to address elsewhere.

Acknowledgements

We would like to thank Anestis Antoniadis and Jeremie Bigot (Laboratoiré IMAG-LMC, University Joseph Fourier, France) for helpful discussions. The inductance plethysmography data was kindly supplied by Andrew Black (Sir Humphry Davy Department of Anaesthesia, Bristol Royal Infirmary, United Kingdom). We would also like to thank the two anonymous referees and Erricos Kontoghiorghes (Co-Editor) whose valuable comments and recommendations helped to improve this paper.

References

Abramovich, F., Sapatinas, T., 1999. Bayesian approach to wavelet decomposition and shrinkage. In: Müller, P., Vidakovic, B. (Eds.), Bayesian Inference in Wavelet Based Models, Lecture Notes in Statistics, Vol. 141. Springer, New York, pp. 33–50.

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

Abramovich, F., Sapatinas, T., Silverman, B.W., 1998. Wavelet thresholding via a Bayesian approach. J. Roy. Statist. Soc. B 60, 725–749.

Antoniadis, A., 1997. Wavelets in statistics: a review (with discussion). J. Ital. Statist. Soc. 6, 97–144.

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

Cai, T.T., Silverman, B.W., 2001. Incorporating information on neighboring coefficients into wavelet estimation. Sankhyā Ser. B 63, 127–148.

Chipman, H.A., Kolaczyk, E.D., McCulloch, R.E., 1997. Adaptive Bayesian wavelet shrinkage. J. Am. Statist. Assoc. 92, 1413–1421.

Clyde, M., George, E.I., 1999. Empirical Bayes estimation in wavelet nonparametric regression. In: Müller, P., Vidakovic, B. (Eds.), Bayesian Inference in Wavelet Based Models, Lecture Notes in Statistics, Vol. 141. Springer, New York, pp. 309–322.

Clyde, M., George, E.I., 2000. Flexible empirical Bayes estimation for wavelets. J. Roy. Statist. Soc. B 62, 681–698.

Clyde, M., Parmigiani, G., Vidakovic, B., 1998. Multiple shrinkage and subset selection in wavelets. Biometrika 85, 391–402.

Coifman, R.R., Donoho, D.L., 1995. Translation-invariant de-noising. In: Antoniadis, A., Oppenheim, G. (Eds.), Wavelets and Statistics, Lecture Notes in Statistics, Vol. 103. Springer, New York, pp. 125–150.

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

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). J. Roy. Statist. Soc. B 39, 1–38.

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

Donoho, D.L., Johnstone, I.M., 1995. Adapting to unknown smoothness via wavelet shrinkage. J. Amer. Statist. Assoc. 90, 1200–1224.

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

Everitt, B.S., 1987. Introduction to Optimization Methods and their Application in Statistics. Chapman & Hall, London.

Johnstone, I.M., Silverman, B.W., 1998. Empirical Bayes approaches to mixture problems and wavelet regression. Technical Report, School of Mathematics, University of Bristol, UK.

Hall, P., Kerkyacharian, G., Picard, D., 1998. Block threshold rules for curve estimation using kernel and wavelet methods. Ann. Statist. 26, 922–942.

Hall, P., Kerkyacharian, G., Picard, D., 1999. On the minimax optimality of block thresholded wavelet estimators. Statist. Sinica 9, 33–50.

Hall, P., Penev, S., Kerkyacharian, G., Picard, D., 1997. Numerical performance of block thresholded wavelet estimators. Statist. Comput. 7, 115–124.

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

Müller, P., Vidakovic, B. (Eds.), 1999. Bayesian Inference in Wavelet Based Models. Lecture Notes in Statistics, Vol. 141. Springer, New York.

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

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

Vannucci, M., Corradi, F., 1999a. Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. J. Roy. Statist. Soc. B 61, 971–986.

Vannucci, M., Corradi, F., 1999b. Modelling dependence in the wavelet domain. In: Müller, P., Vidakovic, B. (Eds.), Bayesian Inference in Wavelet Based Models, Lecture Notes in Statistics, Vol. 141. Springer, New York, pp. 173–186.

Vidakovic, B., 1998a. Non-linear wavelet shrinkage with Bayes rules and Bayes factors. J. Amer. Statist. Assoc. 93, 173–179.

Vidakovic, B., 1998b. Wavelet based nonparametric Bayes methods. In: Dey, D.D., Müller, P., Sinha, D. (Eds.), Practical Nonparametric and Semiparametric Bayesian Statistics, Lecture Notes in Statistics, Vol. 133. Springer, New York, pp. 133–155.

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

Vidakovic, B., Müller, P., 1995. Wavelet shrinkage with affine Bayes rules with applications. Technical Report, Institute of Statistics and Decision Sciences, Duke University, USA.