WAVELET SHRINKAGE WITH DOUBLE WEIBULL PRIOR

Norbert Reményi[a] and Brani Vidakovic[b]

[a] School of Industrial and Systems Engineering

Georgia Institute of Technology

765 Ferst Drive, Atlanta, Georgia 30332-0205, USA

nremenyi@gatech.edu

[b] Department of Biomedical Engineering

Georgia Institute of Technology

2101 Whitaker Building, 313 Ferst Drive, Atlanta, Georgia 30332-0535, USA

Key Words: Nonparametric regression; Wavelet shrinkage; Double Weibull distribution; Bayes estimation; Larger posterior mode.

ABSTRACT

In this paper we propose a denoising methodology in the wavelet domain based on a Bayesian hierarchical model using Double Weibull prior. We propose two estimators, one based on posterior mean (*DWWS*) and the other based on larger posterior mode (*DWWS-LPM*), and show how to calculate them efficiently. Traditionally, mixture priors have been used for modeling sparse wavelet coefficients. The interesting feature of this paper is the use of non-mixture prior. We show that the methodology provides good denoising performance, comparable even to state-of-the-art methods that use mixture priors and empirical Bayes setting of hyperparameters, which is demonstrated by extensive simulations on standardly used test functions. An application to real-word data set is also considered.

1. INTRODUCTION

In the present paper we consider a novel Bayesian model in the wavelet domain as a solution to the classical nonparametric regression problem

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where x_i, $i = 1, \ldots, n$, are equispaced sampling points, and the errors ε_i are i.i.d. normal random variables, with zero mean and variance σ^2. The interest is to estimate the function f from the observations y_i. After applying a linear and orthogonal wavelet transform, the equation in (1) becomes

$$d_{jk} = \theta_{jk} + \varepsilon_{jk},$$

where d_{jk}, θ_{jk} and ε_{jk} are the wavelet coefficients (at resolution j and position k) corresponding to y, f and ε respectively. Note that ε_i and ε_{jk} are equal in distribution due to orthogonality of wavelet transforms. Due to the whitening property of the wavelet transforms (Flandrin, 1992) many of the existing methods assume independence of the coefficients, and omit the double indices jk to work with a generic wavelet coefficient model

$$d = \theta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2). \tag{2}$$

The indices will be used when needed for clarity of the exposition.

To estimate θ in model (2) Bayesian shrinkage rules have been proposed in the literature by many authors. By a shrinkage rule the observed wavelet coefficients d are replaced with their shrunk version $\hat{\theta} = \delta(d)$. Then f is estimated as the inverse wavelet transform of $\hat{\theta}$. Empirical distributions of detail wavelet coefficients for signals encountered in practical applications are (at each resolution level) centered around and peaked at zero (Mallat, 1989). A range of models, for which unconditional distribution of wavelet coefficients mimic this observation, have been considered in the literature. The traditional Bayesian models consider prior distribution on the wavelet coefficient θ as

$$\pi(\theta) = \epsilon\delta_0 + (1 - \epsilon)\xi(\theta), \tag{3}$$

where δ_0 is a point mass at zero, ξ is symmetric about 0, unimodal distribution, and ϵ is a fixed parameter in [0,1], usually level dependent, that controls the amount of shrinkage for values of d close to 0. This type of model was considered by Abramovich et al. (1998), Vidakovic (1998), Clyde and George (1999, 2000), Vidakovic and Ruggeri (2001) and Johnstone and Silverman (2005), among others.

The above models provide good denoising performance because of their adaptivity provided by the point mass at zero. However, parameter ϵ, which controls the extent of shrinkage, needs to be specified. One of the contributions of this paper is simplification of the traditional mixture prior. We demonstrate that, in the wavelet context, a single prior can match the performance of more complex contamination priors from (3).

The paper is organized as follows. Section 2 introduces the model and discusses the advantage of using the Double Weibull prior. Section 3 explains the computation of two Bayes' estimators for our model, the posterior mean and the larger posterior mode. Section 4 contains simulations and comparisons to selected existing methods. Section 5 includes application of the method to inductance plethysmography data. Some remarks and discussion are provided in Sections 6 and 7.

2. MODEL

In our paper we consider the following Bayesian model

$$\begin{aligned} d|\theta &\sim \mathcal{N}(\theta, \sigma^2) \\ \theta &\sim \mathcal{DW}(b, c), \end{aligned} \tag{4}$$

where $\mathcal{N}(\theta, \sigma^2)$ denotes the Normal distribution and $\mathcal{DW}(b, c)$ denotes the Double Weibull distribution with probability density function

$$\pi(\theta|b, c) = \frac{c}{2b}|\theta|^{c-1} \exp\left\{-\frac{|\theta|^c}{b}\right\},$$

where b and c are the scale and shape parameters, respectively. The standard Weibull distribution is popular for analyzing lifetime data. However, its symmetric relative, the Double Weibull distribution, introduced by Balakrishnan and Kocherlakota (1985), is not extensively used in the literature, and have not been used in the wavelet shrinkage context previously. Balakrishnan and Kocherlakota (1985) considered a 3-parameter version of this distribution with location parameter a, but in our case $a = 0$ since the prior on the wavelet coefficient θ is always centered at zero, due to the definition of detail wavelet coefficients.

The Double Weibull is a flexible family, which includes the Double Exponential distribution as its special case ($c = 1$). Figure 1 shows the Double Weibull density for $b = 1$

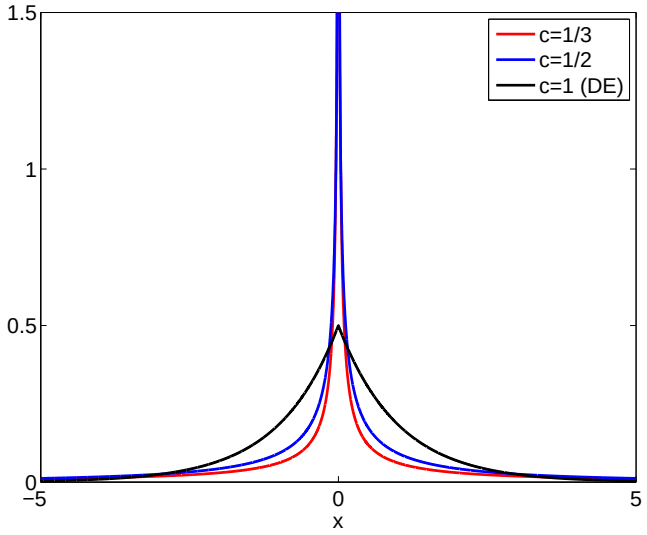

Figure 1: Double Weibull distribution for different values of c.

and $c = 1/3, 1/2, 1$. In case of $c < 1$, the Double Weibull density approaches infinity as $|\theta|$ approaches zero. This property of the prior will be crucial for the performance of the induced Bayes estimators. The singularity at zero mimics the effect of a point mass at zero in the mixture priors mentioned above. A prior with similar property was considered implicitly by Cutillo et al. (2008), and explicitly by Carvalho et al. (2010). Carvalho et al. (2010) consider the "Horseshoe" prior in form of a scale mixture of Normal densities and use it in a context of sparse estimation. The Horseshoe prior, however, does not exist in a closed form.

The shrinkage estimator for the wavelet coefficient corresponding to the signal part θ, derived from (4) is fully specified by eliciting the hyperparameters σ^2, b and c. In this paper, we consider two such estimators and evaluate their performance. The first is the posterior mean, which is a traditional choice in Bayesian estimation problems and the second is the "larger posterior mode", denoted as LPM in the sequel. The shrinkage procedure based on the posterior mean will be referred as Double Weibull Wavelet Shrinker ($DWWS$), while the one based on the LPM will have the acronym $DWWS$-LPM. The existence of the LPM is an intrinsic characteristic of the considered Bayesian model (likelihood-prior). For more information on the LPM approach the reader is referred to Cutillo et al. (2008).

3. THE BAYES ESTIMATOR

In this section we provide details of how to find the posterior mean and LPM as the proposed shrinkage estimators.

3.1 Posterior Mean

It is well known that the posterior mean, as an estimator of θ, has the following form

$$\delta(d) = \frac{\int \theta g(d,\theta) d\theta}{m(d)} = \frac{\int \theta f(d|\theta)\pi(\theta) d\theta}{\int f(d|\theta)\pi(\theta) d\theta}, \tag{5}$$

where g is the joint distribution, f is the likelihood, π is the prior, and m is the marginal distribution. From the marginal distribution

$$m(d) \propto \int e^{-\frac{(d-\theta)^2}{2\sigma^2}} |\theta|^{c-1} e^{-\frac{|\theta|^c}{b}} d\theta, \tag{6}$$

it can be seen that the integral does not exist in a closed form for fixed $c < 1$. However, the integral in (6) is finite, the posterior distribution is proper, and the posterior mean exists, as well. This is true because we are convolving the Normal with the Double Weibull distribution, which is integrable and all of its moments exist (Balakrishnan and Kocherlakota, 1985).

It is possible to evaluate this integral as a convolution using the characteristic functions of the likelihood and the prior, but the characteristic function of the Double Weibull distribution does not have a simple form and involves special functions (Nadarajah, 2008). Therefore the posterior mean will be computed by numerical integration using adaptive Gauss-Kronrod quadrature, for which we utilized the function script `quadgk(fun,a,b)` in MATLAB©. It is apparent in equation (6) that the integral has a singularity at $\theta = 0$ for $c < 1$. One can significantly increase the speed and accuracy of integration by removing this singularity, which can be done with a change of variable. After a change of variable $y = \{\text{sign}(\theta)\theta\}^c$ the posterior mean becomes

$$\delta(d) = \frac{\int_0^\infty y^{1/c} e^{-y/b} e^{-\frac{\left(d-y^{1/c}\right)^2}{2\sigma^2}} dy - \int_0^\infty y^{1/c} e^{-y/b} e^{-\frac{\left(d+y^{1/c}\right)^2}{2\sigma^2}} dy}{\int_0^\infty e^{-y/b} e^{-\frac{\left(d-y^{1/c}\right)^2}{2\sigma^2}} dy + \int_0^\infty e^{-y/b} e^{-\frac{\left(d+y^{1/c}\right)^2}{2\sigma^2}} dy}. \tag{7}$$

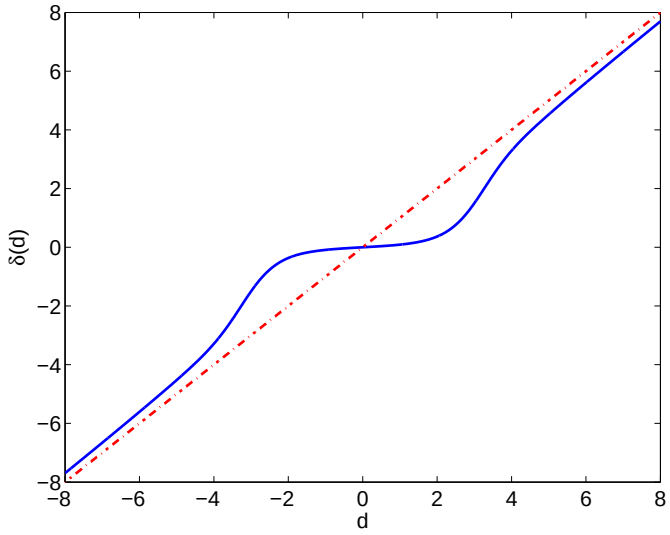

Figure 2: Posterior mean for c=1/3.

If c has a form of $1/n$, n odd, the posterior mean simplifies to

$$\delta(d) = \frac{\int_{-\infty}^{\infty} y^{1/c} e^{-|y|/b} e^{-\frac{\left(d-y^{1/c}\right)^2}{2\sigma^2}} dy}{\int_{-\infty}^{\infty} e^{-|y|/b} e^{-\frac{\left(d-y^{1/c}\right)^2}{2\sigma^2}} dy}. \tag{8}$$

Note that for any $c \in (0, 1)$ the posterior mean can be efficiently computed using 7. Figure 2 shows the posterior mean for $c = 1/3$, $b = 0.4$ and $\sigma^2 = 1$.

Figure 3 shows the marginal distribution $m(d)$, computed numerically for $c = 1/3$, $b = 1$ and $\sigma^2 = 1$. The marginal distribution is compared to a Normal distribution with mean zero and standard deviation 2.6, which arises from matching the interquartile range of the two distributions. It is a desirable property in Bayesian wavelet shrinkage to produce a marginal that matches the observed empirical distribution of wavelet coefficients. We can see from Figure 3 that the marginal distribution corresponding to model (4) exhibits heavier tails, and it is more peaked than the Normal density. This is in agreement with the observations of Mallat (1989) concerning the shape of empirical distributions of wavelet coefficients.

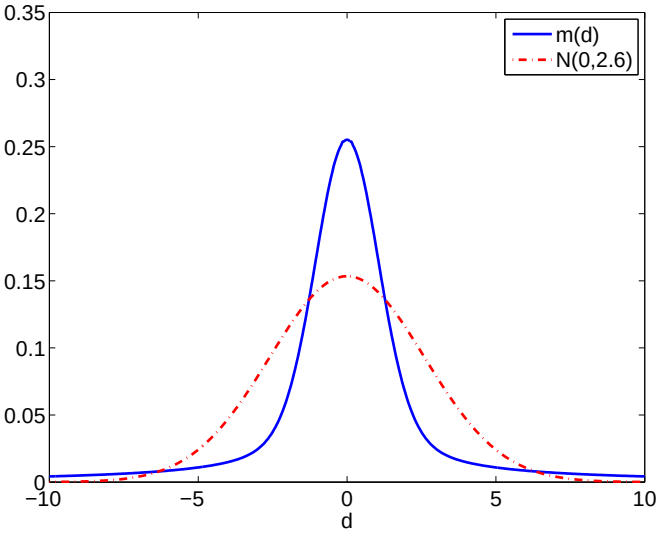

Figure 3: Marginal distribution of the wavelet coefficients.

Setting of the hyperparameters for the *DWWS* rule (5) is discussed in Section 4.1.

3.2 Larger Posterior Mode (LPM)

The LPM estimator was first introduced in the wavelet shrinkage context by Cutillo et al. (2008), and it is based on the Bayesian MAP (Maximum a Posteriori) principle. The LPM rule relates to the mode of the posterior distribution larger in absolute value. The MAP estimator of the wavelet coefficient θ is a rule maximizing the posterior $\pi(\theta|d)$, which is proportional to the joint distribution of d and θ, $g(d,\theta)$. Hence, the MAP estimator for θ also maximizes $g(d,\theta)$. For the model in (4) the joint distribution is

$$g(d,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(d-\theta)^2}{2\sigma^2}} \frac{c}{2b} |\theta|^{c-1} e^{-\frac{|\theta|^c}{b}}.$$

This leads to the posterior proportional to

$$\pi(\theta|d) \propto g(d,\theta) \propto e^{-\frac{(d-\theta)^2}{2\sigma^2}} |\theta|^{c-1} e^{-\frac{|\theta|^c}{b}}.$$

Figure 4 shows the posterior distribution for $c = 1/3$, $b = 1$, $\sigma^2 = 1$ and $d = -3, -2, -1, 1, 2, 3$. Note that the shape of posterior depends on the absolute magnitude of the observed wavelet coefficient d. If $|d|$ is small, the posterior mode is unique and equals to 0. For large values of $|d|$ there are two posterior modes and the one larger in magnitude is chosen.

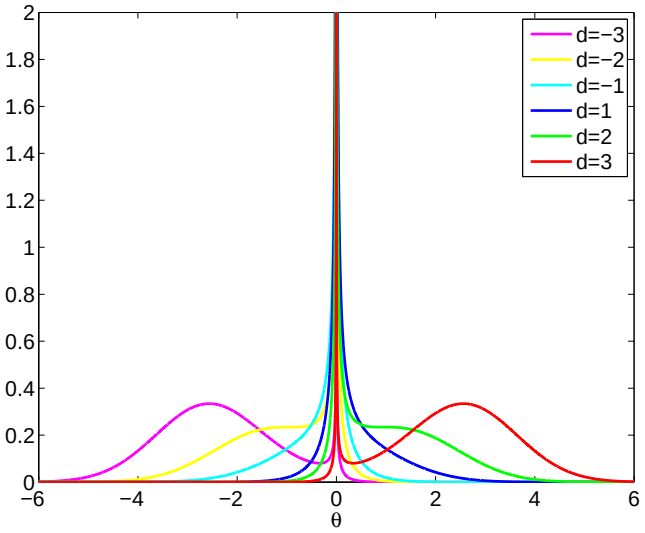

Figure 4: Posterior distribution of the wavelet coefficients.

The logarithm of the posterior is proportional to

$$l = \log \pi(\theta|d) \propto -\frac{(d-\theta)^2}{2\sigma^2} + (c-1)\log|\theta| - \frac{|\theta|^c}{b},$$

and has extrema at the solutions of the equation

$$\frac{d-\theta}{\sigma^2} + (c-1)\text{sign}(\theta)\frac{1}{|\theta|} - \frac{c}{b}\text{sign}(\theta)|\theta|^{c-1} = 0,$$

which is equivalent to the equation

$$-\frac{1}{\sigma^2}\theta^2 + \frac{d}{\sigma^2}\theta - \frac{c}{b}|\theta|^c + c - 1 = 0. \tag{9}$$

For fixed $c < 1$ and by substituting $y = |\theta|^c$, equation (9) can be modified so that the solution is equivalent to a solution of a polynomial equation of order $2/c$. We will use the following numerical algorithm to find the LPM estimator from equation (9):

(1) Find the roots of the equation $-\frac{1}{\sigma^2}y^{2/c} + \text{sign}(d)\frac{d}{\sigma^2}y^{1/c} - \frac{c}{b}y + c - 1 = 0$. Denote the roots by $y^\star$ and the real roots by $y_r^\star$.

(2) If all the roots are complex ($y_r^\star$ is empty), $\delta_{LPM}(d) = 0$.

(3) If reals roots exist, $\delta_{LPM}(d) = \text{sign}(d)[\max(y_r^\star)]^{1/c}$.

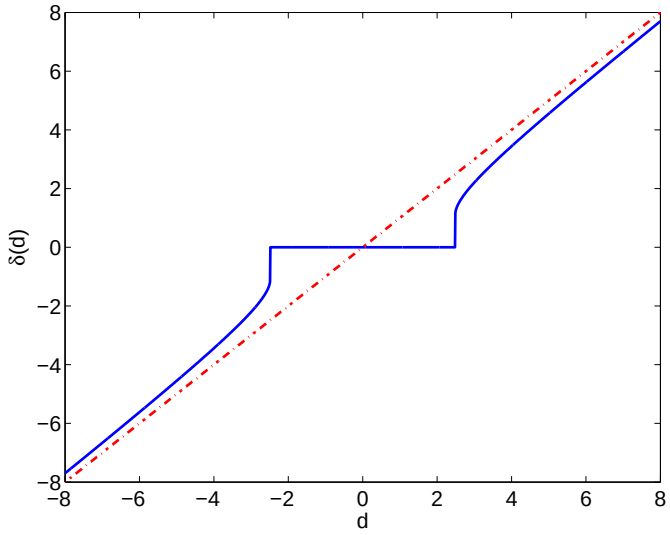

Figure 5: LPM rule for c=1/3.

Therefore, the LPM estimator for the model in (4) is

$$\delta_{LPM}(d) = \text{sign}(d)[\max(y_r^\star)]^{1/c},$$

where $\max(y_r^\star)$ is the maximum real root of the equation $-\frac{1}{\sigma^2}y^{2/c}+\text{sign}(d)\frac{d}{\sigma^2}y^{1/c}-\frac{c}{b}y+c-1 = 0$. If no real root of this equation exist, $\delta_{LPM}(d) = 0$. In general, the roots can be computed by a nonlinear equation solver for any real $c \in (0, 1)$, but for a rational $c = m/n$ the roots can be found by a polynomial root solver, which was utilized in the implementation. Figure 5 shows the LPM rule for $c = 1/3$, $b = 0.4$ and $\sigma^2 = 1$. It is apparent from the figure that the rule is thresholding.

4. SIMULATIONS

In this section we apply the proposed shrinkage estimators and compare their performance to several existing and established wavelet denoising methods. For the *DWWS* and *DWWS-LPM* estimators we first discuss the selection of hyperparameters, then we present and compare the simulation results.

4.1 Selection of Hyperparameters

In any Bayesian modeling task the selection of hyperparameters is critical for good per-

formance of the model. It is also desirable to have a default selection of the hyperparameters which makes the shrinkage procedure automatic. In the model (4) we need to specify parameters σ^2, b and c.

Parameter σ^2. Parameter σ^2 represents the variance of the random error ε. In the wavelet shrinkage literature σ^2 is frequently estimated by a robust estimator of the variance of wavelet coefficients at the finest level of detail (Donoho and Johnstone, 1994). We will adopt this practice and use the robust MAD estimator to estimate σ as $\hat{\sigma} = MAD/0.6745$. Here MAD stands for the median absolute deviation from the median of the wavelet coefficients at finest level of detail, and the constant 0.6745 calibrates the estimator to be comparable with sample standard deviation.

Parameter b. Scale parameter b accounts for the spread of the Double Weibull prior distribution. We propose a moment matching parameter specification, which was used for example by Cutillo et al. (2008) and Vidakovic and Ruggeri (2001). We propose to estimate b_j levelwise for all dyadic levels $J_0 \leq j \leq \log_2 n - 1$. Because of the linearity of wavelet transform, the i.i.d normal noise with variance σ^2 transforms stochastically unchanged to each dyadic level. In the case of the Double Weibull prior, the variance of the signal part is $\sqrt[c]{b_j^2}\Gamma(1 + 2/c)$. Since the model assumes independence of signal and error parts, we have $\sigma_{d_j}^2 = \sqrt[c]{b_j^2}\Gamma(1 + 2/c) + \sigma^2$, where $\sigma_{d_j}^2$ is the variance of the observations d_{jk} at j^{th} dyadic level. Therefore a reasonable estimator for b_j is

$$\hat{b}_j = \left\{ \frac{(\sigma_{d_j}^2 - \hat{\sigma}^2)_+}{\Gamma(1 + 2/c)} \right\}^{c/2}, \quad J_0 \leq j \leq \log_2 n - 1, \tag{10}$$

where $a_+ = \max(a, 0)$. In case $\hat{\sigma}^2 > \sigma_{d_j}^2$, we set $\hat{b}_j = 0$. Having $\hat{b}_j = 0$ is equivalent to a degenerate/point-mass-at-zero prior distribution on the wavelet coefficients. Therefore, if $\hat{b}_j = 0$, we set all the wavelet coefficients at level j to zero, similarly to Vidakovic and Ruggeri (2001).

Parameter c. Parameter c accounts for the shape of the prior distribution on the wavelet coefficients. When smaller than 1, parameter c controls the "strength of infinity" at zero. In this sense the role of c is similar to that of the point mass in the mixture prior models,

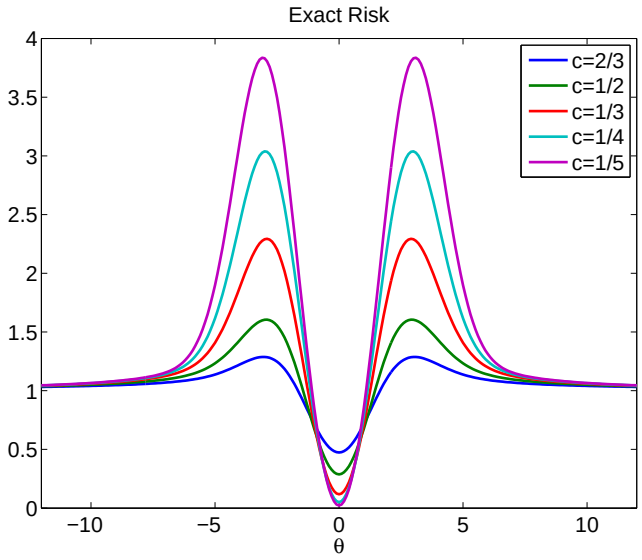

Figure 6: Exact risk plot for posterior mean, for $\sigma^2 = 1$ and $\sigma^2_{d_j} = 100$.

and should be elicited depending on the signal regularity. In addition to this, parameter c also controls the tails of the prior distribution. We used $c = 1/3$ in our simulations, which empirically was the superior universal choice. Of course, c can be adaptively set depending on the input signal under consideration, as we will do in Section 5.

Figure 6 shows the exact risks of the posterior mean estimator for different values of c. We set $\sigma^2 = 1$, $\sigma^2_{d_j} = 100$ and specified b by equation (10) depending on different c's. From the plot we can see that the choice $c = 1/3$ is a good compromise in terms of risk. For $|\theta|$ close to zero $c = 1/3$ provides smaller risk than $c = 1/2$ or $c = 2/3$, and for larger $|\theta|$ the choice $c = 1/3$ has smaller risk than $c = 1/4$ or $c = 1/5$. Note that the pattern and shape of the plot depends on the quantity $\sigma^2_{d_j} - \sigma^2$, but $c = 1/3$ was an empirically superior choice.

For $c = 1/3$, equation (9) becomes

$$-\frac{1}{\sigma^2}\theta^2 + \frac{d}{\sigma^2}\theta - \frac{1}{3b}|\theta|^{1/3} - 2/3 = 0,$$

and the algorithm to find the LPM estimator becomes equivalent to solving the equation

$$-\frac{1}{\sigma^2}y^6 + \text{sign}(d)\frac{d}{\sigma^2}y^3 - \frac{1}{3b}y - 2/3 = 0. \tag{11}$$

Note that it is possible to specify parameter c levelwise similar to specifying the weight parameter in the Bayesian mixture prior models. Therefore, if elicited levelwise, c could be set up to increase from the finest to the coarsest dyadic levels of wavelet coefficients. However, because of simplicity and the good performance provided, c was held fixed through the dyadic levels and the levelwise elicitation of parameter b provided the adaptiveness of the shrinkage rule.

4.2 Simulations and Comparisons with Selected Methods

In this section we discuss the performance of the proposed *DWWS* and *DWWS-LPM* estimators and compare them to some established wavelet-based methods for reconstructing noisy signals. Four standard test functions (`Blocks, Bumps, Doppler, Heavisine`) were considered (Donoho and Johnstone, 1994) in the simulation study. The functions were rescaled such that the added noise ($\sigma^2 = 1$) produced the preassigned signal-to-noise ratio (SNR). The test functions were simulated at $n = 512, 1024$, and 2048 points equally spaced in the unit interval. Three common SNRs were selected, SNR=3, SNR=5, and SNR=7. The standard wavelet bases were used: Symmlet 8 for `Heavisine` and `Doppler`, Daubechies 6 for `Bumps` and Haar for `Blocks`. The coarsest decomposition level was $J_0 = 3$, which matches the suggested $J_0 = \lfloor \log_2(\log(n)) + 1 \rfloor$ by Antoniadis et al. (2001). Note, that for computing the *DWWS* estimator, MATLAB's built-in Gauss-Kronrod quadrature method was used, and the *DWWS-LPM* estimator is the solution of equation (11), for which MATLAB's built-in polynomial root-solver was used.

Reconstruction of the theoretical signal was evaluated by the average mean squared error (AMSE), calculated as

$$\frac{1}{Mn} \sum_{k=1}^{M} \sum_{i=1}^{n} \left(\hat{f}_k(t_i) - f(t_i) \right)^2,$$

where M is the number of simulation runs and $f(t_i),\ i = 1, \ldots, n$ are known values of the test functions considered. We denote by $\hat{f}_k(t_i),\ i = 1, \ldots, n$ the estimator from the kth simulation run.

The proposed estimators were compared to the *EBAYES* method of Johnstone and Sil-

verman (2005) using the posterior mean, the *BAMS* method of Vidakovic and Ruggeri (2001), the *LPM* method from Model 1 of Cutillo et al. (2008), the classical *VisuShrink* and *Hybrid-SureShrink* of Donoho and Johnstone (1994, 1995), the scale invariant term-by-term *ABE* method of Figueiredo and Nowak (2001), and finally the *NeighCoeff* method of Cai and Silverman (2001). Note that methods *EBAYES*, *BAMS*, *LPM* and *ABE* are Bayesian.

Results are summarized in Tables 1 and 2, where boldface numbers indicate the smallest AMSE result for each test scenario. The number of simulations performed was $M = 1000$. From the results we can see that the proposed estimators are comparable to the established shrinkage methods. In some scenarios involving `Heavisine` signal the *DWWS* is superior. Simulations further indicate that the *DWWS* estimator outperforms the *BAMS* estimator in 64% of the cases, and the *EBAYES* method in 28% of the cases. This is remarkable considering that these Bayesian methods are based on a more complicated mixture model with a point mass, and the latter one uses an empirical Bayes procedure to estimate the hyperparameters. It is also evident from Tables 1 and 2, that the *DWWS-LPM* estimator outperforms the *LPM* estimator in 67% of the cases. Note that for the model in Cutillo et al. (2008) the posterior distribution is not proper for all values of the hyperparameter k, hence the posterior mean does not exist. For the proposed model in (4) the posterior mean always exists and the resulting *DWWS* estimator uniformly outperforms the *DWWS-LPM* estimator. However, *DWWS-LPM* is computationally more robust and faster to compute. Also note, that the authors of *LPM* select hypermarameter k separately for each simulated test function, so the results are optimal. In our simulation study we kept hyperparameter c default for each test function. It can also be seen from the results that the *DWWS-LPM* estimator outperforms the *ABE* method in 81% of the cases. The difference in AMSE was the most pronounced for signals `Doppler` and `Heavisine`. The *ABE* is also using a single prior model and the MAP approach. Finally, the proposed methods outperform the non-Bayesian methods *VisuShrink*, *Hybrid-SureShrink* and *NeighCoeff* under most test scenarios.

Graphical summary of the results is presented in Figure 7 where the boxplots of the MSE are given for $n = 1024$ and $SNR = 5$.

Signal	N	Method	SNR=3	SNR=5	SNR=7	Signal	N	Method	SNR=3	SNR=5	SNR=7
Blocks	512	DWWS	0.2174	0.1917	0.1790	Doppler	512	DWWS	0.2002	0.2244	0.2296
		DWWS-LPM	0.2223	0.1940	0.1826			DWWS-LPM	0.2061	0.2315	0.2389
		EBAYES	0.2122	**0.1886**	**0.1670**			EBAYES	0.1962	0.2155	0.2211
		BAMS	**0.2101**	0.1943	0.1763			BAMS	0.1954	0.2131	0.2264
		LPM	0.2217	0.1949	0.1756			LPM	0.2110	0.2258	0.2353
		VISU	0.2769	0.2344	0.1945			VISU	0.2578	0.2779	0.2862
		SURE	0.3517	0.3653	0.3530			SURE	0.2743	0.3797	0.4132
		ABE	0.2221	0.2072	0.1967			ABE	0.2108	0.2240	0.2325
		NC	0.4103	0.4031	0.3679			NC	**0.1684**	**0.1784**	**0.1846**
	1024	DWWS	0.1563	0.1289	0.1241		1024	DWWS	0.1141	0.1348	0.1469
		DWWS-LPM	0.1567	0.1329	0.1281			DWWS-LPM	0.1241	0.1456	0.1561
		EBAYES	**0.1510**	**0.1207**	**0.1038**			EBAYES	0.1168	0.1363	0.1473
		BAMS	0.1583	0.1311	0.1107			BAMS	0.1180	0.1350	0.1482
		LPM	0.1596	0.1284	0.1130			LPM	0.1349	0.1584	0.1681
		VISU	0.2161	0.1510	0.1231			VISU	0.1552	0.1855	0.2085
		SURE	0.3108	0.2926	0.2274			SURE	0.1655	0.1964	0.2363
		ABE	0.1695	0.1558	0.1472			ABE	0.1554	0.1709	0.1786
		NC	0.3253	0.3088	0.2680			NC	**0.0945**	**0.1160**	**0.1241**
	2048	DWWS	0.0919	0.0816	0.0795		2048	DWWS	0.0624	0.0771	0.0884
		DWWS-LPM	0.0944	0.0852	0.0835			DWWS-LPM	0.0685	0.0846	0.0953
		EBAYES	**0.0865**	**0.0730**	**0.0603**			EBAYES	0.0642	0.0773	0.0860
		BAMS	0.0921	0.0788	0.0665			BAMS	0.0687	0.0783	0.0868
		LPM	0.0914	0.0774	0.0643			LPM	0.0755	0.0887	0.0978
		VISU	0.1172	0.0919	0.0712			VISU	0.0835	0.1003	0.1121
		SURE	0.1740	0.1815	0.1629			SURE	0.0845	0.1184	0.1514
		ABE	0.1227	0.1161	0.1108			ABE	0.1158	0.1242	0.1297
		NC	0.1938	0.1798	0.1587			NC	**0.0511**	**0.0636**	**0.0714**

Table 1: AMSE of the proposed *DWWS* and *DWWS-LPM* estimators compared to other methods for test signals `Blocks` and `Doppler`

Signal	N	Method	SNR=3	SNR=5	SNR=7	Signal	N	Method	SNR=3	SNR=5	SNR=7
Bumps	512	DWWS	0.4659	0.4733	0.4875	Heavisine	512	DWWS	**0.0793**	0.1199	0.1534
		DWWS-LPM	0.4908	0.5128	0.5270			DWWS-LPM	0.0912	0.1337	0.1696
		EBAYES	**0.4110**	**0.4417**	**0.4680**			EBAYES	0.0842	0.1205	0.1502
		BAMS	0.4834	0.5132	0.5573			BAMS	0.0957	**0.1185**	**0.1374**
		LPM	0.4606	0.4885	0.5052			LPM	0.0932	0.1445	0.1800
		VISU	0.7354	0.7630	0.8146			VISU	0.0996	0.1583	0.2028
		SURE	0.7052	0.5953	0.6497			SURE	0.0826	0.1300	0.1751
		ABE	0.4601	0.4983	0.5235			ABE	0.1315	0.1614	0.1845
		NC	0.5828	0.5273	0.4779			NC	0.0898	0.1438	0.1759
	1024	DWWS	0.2855	0.2986	0.3004		1024	DWWS	**0.0504**	**0.0683**	0.0890
		DWWS-LPM	0.3057	0.3174	0.3156			DWWS-LPM	0.0583	0.0783	0.1008
		EBAYES	**0.2713**	**0.2921**	0.2956			EBAYES	0.0536	0.0693	0.0866
		BAMS	0.2969	0.3263	0.3404			BAMS	0.0607	0.0707	**0.0815**
		LPM	0.3168	0.3318	0.3308			LPM	0.0635	0.0867	0.1121
		VISU	0.4496	0.4808	0.4884			VISU	0.0683	0.0937	0.1223
		SURE	0.3840	0.4676	0.4907			SURE	0.0534	0.0747	0.0955
		ABE	0.3004	0.3193	0.3240			ABE	0.1075	0.1233	0.1360
		NC	0.3217	0.3008	**0.2878**			NC	0.0667	0.0894	0.0989
	2048	DWWS	0.1717	0.1871	0.1905		2048	DWWS	**0.0313**	0.0457	0.0560
		DWWS-LPM	0.1836	0.1965	0.2007			DWWS-LPM	0.0376	0.0534	0.0630
		EBAYES	**0.1668**	**0.1816**	**0.1866**			EBAYES	0.0339	**0.0456**	0.0543
		BAMS	0.1823	0.1978	0.2049			BAMS	0.0402	0.0471	**0.0531**
		LPM	0.2033	0.2110	0.2120			LPM	0.0395	0.0609	0.0760
		VISU	0.2766	0.2948	0.2863			VISU	0.0416	0.0653	0.0887
		SURE	0.2438	0.2907	0.3071			SURE	0.0344	0.0506	0.0709
		ABE	0.2039	0.2132	0.2167			ABE	0.0925	0.1037	0.1103
		NC	0.1824	0.1840	0.1877			NC	0.0435	0.0543	0.0599

Table 2: AMSE of the proposed *DWWS* and *DWWS-LPM* estimators compared to other methods for test signals `Bumps` and `Heavisine`

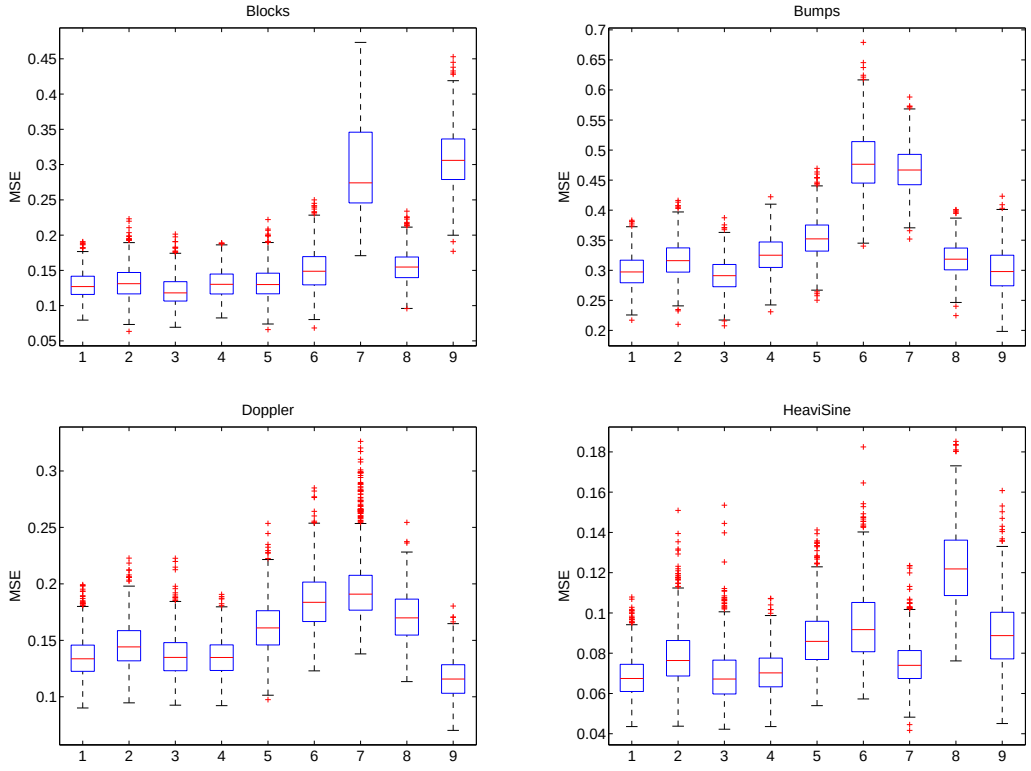

Figure 7: Boxplots of MSE for various shrinking procedures (1) *DWWS*, (2) *DWWS-LPM*, (3) *EBAYES*, (4) *BAMS*, (5) *LPM*, (6) *VisuShrink*, (7) *Hybrid*, (8) *ABE*, (9) *NeighCoeff* based on $n = 1024$ points and $SNR = 5$.

5. APPLICATION TO INDUCTANCE PLETHYSMOGRAPHY DATA

In this section we apply the proposed wavelet estimators to a real-world data set from anaesthesiology collected by inductance plethysmography. The recordings were made by the Department of Anaesthesia at the Bristol Royal Infirmary and represent measure of flow of air during breathing. This was analysed by several authors, for example Nason (1996) and Abramovich et al. (1998, 2002). For more information about the data set, we refer the reader to these two papers.

Figure 8 shows a section of plethysmograph recording lasting approximately 80 s ($n = 4096$ observations). Figure 9 shows the reconstructions of the signal with the *DWWS* and *DWWS-LPM* methods. In our reconstruction we set $c = 1/5$, which provided a smoother, visually more pleasing result, although this choice is not necessarily AMSE superior. Both methods remove the noise well, however, the *DWWS* estimator based on the posterior mean provides a slightly smoother result. Abramovich et al. (2002) report the height of the maximum peak while analysing this data set. In our case the height is 0.8410 for the *DWWS* method and 0.8421 for the *DWWS-LPM*. These are quite close to the result 0.8433, obtained by Abramovich et al. (2002), and better compared to some established methods reported in their paper.

6. REMARKS

It is worth mentioning here that a slight modification of the Double Weibull prior can lead to a Bayes rule which can be expressed as a closed form using special functions. Consider the following prior distribution on the wavelet coefficient θ:

$$\pi(\theta|b,c) = \frac{1}{2\Gamma(c)b^c}|\theta|^{c-1}\exp\left\{-\frac{|\theta|}{b}\right\},$$

which is the one dimensional special case of the more general Kotz distribution (Nadarajah, 2003) with $p = 1$, $\mu = 0$, $\Sigma = 1$, $N = (c+1)/2$, $s = 1/2$ and $r = 1/b$. Using an integral identity (p.337, Gradshteyn and Ryzhik, 1980), the marginal distribution and the posterior mean can be expressed as

$$m(d) \;=\; \frac{\sigma^c e^{-d^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}2b^c}\left\{e^{(\sigma/2b-d/2\sigma)^2}D_{-c-1}(\sigma/b-d/\sigma) - e^{(\sigma/2b+d/2\sigma)^2}D_{-c-1}(\sigma/b+d/\sigma)\right\},$$

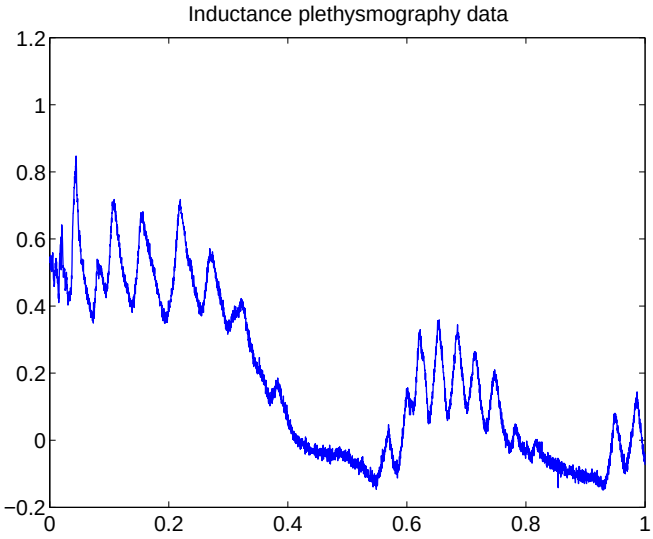

Figure 8: A section of inductance plethysmography data with $n = 4096$.

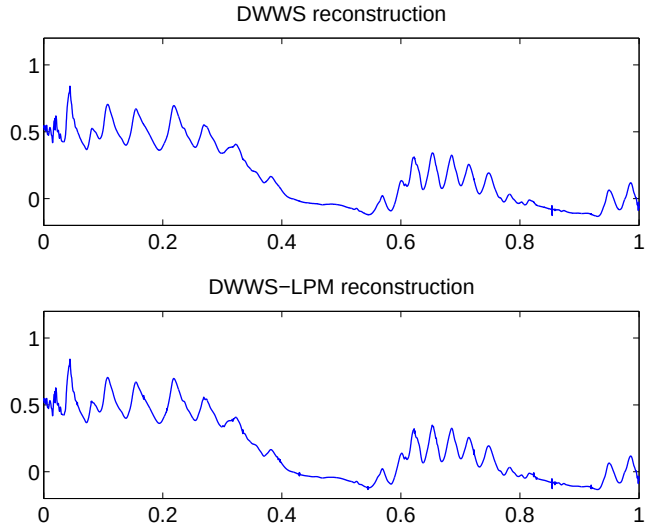

Figure 9: Reconstruction of inductance plethysmography data obtained by the *DWWS* and *DWWS-LPM* methods.

$$\delta(d) \;=\; c\sigma\frac{e^{-d/2b}D_{-c-1}(\sigma/b-d/\sigma)-e^{d/2b}D_{-c-1}(\sigma/b+d/\sigma)}{e^{-d/2b}D_{-c}(\sigma/b-d/\sigma)-e^{d/2b}D_{-c}(\sigma/b+d/\sigma)},$$

where $D_v(x)$ is the parabolic cylinder function (Abramowitz and Stegun, 1964).

Because the marginal distribution is available in a closed form, the empirical Bayes procedure is a possibility for eliciting the hyperparameters of the prior. However, in practice, this estimator is computationally more expensive than *DWWS*, *DWWS-LPM*, and the performance in terms of AMSE is somewhat inferior.

7. CONCLUSIONS

In this paper we have proposed a methodology for Bayesian wavelet denoising. A hierarchical model was specified in which the Double Weibull distribution was utilized as the prior on the locations of wavelet coefficients. In contrast to mixture priors used by some state-of-the-art methods, the wavelet coefficients were modeled by a density with single expression. The flexibility of the Double Weibull distribution was able to mimic the characteristics of mixture priors consisting of a point mass at zero and a heavy tailed spread part. Two Bayesian estimators were proposed, one as the posterior mean (*DWWS*) and the other as the larger posterior mode (*DWWS-LPM*). We also showed how to compute them efficiently. Simulations on standard test functions and comparisons with numerous existing methods demonstrated that the methodology provides good and comparable denoising performance, even compared to state-of-the-art methods that use mixture priors and empirical Bayes setting of hyperparameters. Once again, we emphasize that the aim was the simplicity of the model, and demonstration that a carefully selected single prior could match the performance of more complex mixture priors. An application to real-word data set (inductance plethysmography) was also considered. The methodology performed well in both denoising and preserving the important features of the real data.

Future improvements of the method are possible by specifying hyperparameter c based on dyadic levels and signal regularity. Another avenue for future improvement can be the approximation of integral in (5) to evaluate the posterior mean. However, if approximations are asymptotic, this would work satisfactorily only in the case of shrinkage of multiple related

signals (Chang and Vidakovic, 2002).

In the spirit of reproducible research we made MATLAB scripts used in simulation for *DWWS* and *DWWS-LPM* available at http://gtwavelet.bme.gatech.edu/.

Acknowledgement. The authors thank the editor and two anonymous referees for constructive comments that improved the paper.

BIBLIOGRAPHY

Abramovich, F., and Benjamini, Y. (1995). Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In Wavelets and Statistics (Antoniadis A. and Oppenheim G. Eds.) Lect. Notes Statist., **103**, 5–14, Springer-Verlag, New York.

Abramovich, F., Besbeas, P., and Sapatinas T. (2002). Empirical Bayes Approach to Block Wavelet Function Estimation. *Computational Statistics and Data Analysis*, **39**, 435–451.

Abramovich, F., Sapatinas T., and Silverman B.W. (1998). Wavelet thresholding via a Bayesian Approach. *Journal of the Royal Statistical Society, Series B* **60**, 725–749.

Abramowitz, M., and Stegun, I. (1964). *Handbook of Mathematical Functions*. Dover, New York.

Antoniadis, A., Bigot, J., and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, **6**, 1–83.

Balakrishnan, N., and Kocherlakota S. (1985). On the Double Weibull distribution: Order statistics and estimation. *Sankhyā: The Indian Journal of Statistics, Series B* **47**, 161–178.

Cai, T.T., and Silverman, B.W. (2001). Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhyā: The Indian Journal of Statistics, Series B* **63**, 127–148.

Carvalho, C.M., Polson, N.G., and Scott, J.G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, **97**, 465–480.

Chang, W., and Vidakovic, B. (2002). Wavelet estimation of a base-line signal from repeated noisy measurements by vertical block shrinkage. *Computational Statistics and Data Analysis*, **40**, 317–328.

Clyde, M., and George, E.I. (1999). Empirical Bayes estimation in wavelet nonparametric regression. In Bayesian Inference in Wavelet Based Models, Muller, P. and Vidakovic, B. (Eds.), Lect. Notes Statist., 141, pp. 309–322, New York: Springer-Verlag.

Clyde, M., and George, E.I. (2000). Flexible Empirical Bayes Estimation for Wavelets. *Journal of the Royal Statistical Society, Series B* **62**, 681–698.

Cutillo, L., Yung, J.J., Ruggeri, F., and Vidakovic, B. (2008). Larger Posterior Mode Wavelet Thresholding and Applications. *Journal of Statistical Planning and Inference*, **138**, 3758–3773.

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

Donoho, D.L., and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, **90**, 1200–1224.

Figueiredo, M.A.T., and Nowak, R.D. (2001). Wavelet-based image estimation: an empirical Bayes approach using Jeffrey's noninformative prior. *IEEE Transactions on Image Processing*, **10**, 1322–1331.

Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, **38**, 910–917.

Gradshteyn, I.S., and Ryzhik, I.M. (1980). *Table of integrals, series, and products.* Academic Press, New York.

Johnstone, I.M., and Silverman, B.W. (2005). Empirical Bayes selection of wavelet thresholds. *The Annals of Statistics*, **33**, 1700–1752.

Mallat, S.G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

Nadarajah, S. (2003). The Kotz-type distribution with applications. *Statistics*, **37**, 341–358.

Nadarajah, S. (2008). Discussion Letter to the Editor. *Coastal Engineering*, **55**, 189–190.

Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society, Series B* **58**, 463–479.

Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *Journal of the American Statistical Association*, **93**, 173–179.

Vidakovic, B., and Ruggeri, F. (2001). BAMS Method: Theory and Simulations. *Sankhyā: The Indian Journal of Statistics, Series B* **63**, 234–249.