Routledge
Taylor & Francis Group

# Multivariate Bayes Wavelet Shrinkage and Applications

GABRIEL HUERTA

*Department of Mathematics and Statistics, University of New Mexico*

ABSTRACT    *In recent years, wavelet shrinkage has become a very appealing method for data de-noising and density function estimation. In particular, Bayesian modelling via hierarchical priors has introduced novel approaches for Wavelet analysis that had become very popular, and are very competitive with standard hard or soft thresholding rules. In this sense, this paper proposes a hierarchical prior that is elicited on the model parameters describing the wavelet coefficients after applying a Discrete Wavelet Transformation (DWT). In difference to other approaches, the prior proposes a multivariate Normal distribution with a covariance matrix that allows for correlations among Wavelet coefficients corresponding to the same level of detail. In addition, an extra scale parameter is incorporated that permits an additional shrinkage level over the coefficients. The posterior distribution for this shrinkage procedure is not available in closed form but it is easily sampled through Markov chain Monte Carlo (MCMC) methods. Applications on a set of test signals and two noisy signals are presented.*

KEY WORDS:    Bayes shrinkage, wavelets, discrete wavelet transformation, data de-noising, MCMC methods

## Introduction

Thresholding rules recently became of considerable interest when DeVore *et al.* (1992) and Donoho *et al.* (1995) applied them in the Wavelet shrinkage context. *Wavelet shrinkage* refers to a process of transforming the data or a signal with a *Discrete Wavelet Transformation* (DWT), implementing some kind of reduction to the Wavelet coefficients, and then, applying the *Inverse Wavelet Transformation* (IDWT) to reconstruct the signal. *Hard or Soft Thresholding* are a type of shrinkage in which those coefficients whose absolute value is smaller than a certain bound, are replaced by zero. Analytically simple, these rules are very efficient in data de-noising and especially data compression problems.

However, shrinkage by thresholding poorly accounts for any prior information available about the structure of the data, the signal and the noise. From a Bayesian perspective, new methods of Wavelet reduction have been developed to incorporate prior knowledge on the parameters that define a model for the signal and noise. For example, in Vidakovic (1998), a parametric model is defined for the Wavelet coefficients; a Double Exponential distribution, then the shrinkage problem is formulated as an inference problem and the goal

is that the resulting optimal actions will 'mimic desirable thresholding rules'. In the case of a Double Exponential distribution the rule corresponds to a posterior expectation that is expressed in terms of a Laplace transformation. In Müller & Vidakovic (1995) and Vannucci & Corradi (1999), the formulation is also within a parametric Bayesian model but the unrealistic assumption of independence among the Wavelet coefficients is relaxed. Applications to density estimation and regression are presented by these authors. Clyde *et al.* (1998) use a Bayesian hierarchical model that incorporates a mixture prior, which allows for each Wavelet coefficient to be zero with a positive probability (prior point mass). However, the MCMC method proposed in that paper to obtain posterior estimates of the coefficients can be computationally very intensive. In Chipman *et al.* (1997), a mixture of Normal distributions with different variances is used as a prior distribution for the Wavelet coefficients. One of the variances is set very close to zero to approximate a point mass prior. The main advantage of that approach is that it gives closed expressions for the posterior distribution of interest and so the computations can be done efficiently. In all the methods proposed in these papers, the shrinkage is achieved by replacing the original coefficients by their posterior means. Additionally, Clyde & George (2000) proposed a shrinkage approach that uses a hierarchical model with heavy-tailed error distributions. The prior specifications for some of the parameters in their model are difficult, hence an Empirical Bayes procedure is used to estimate these hyperparameters. This procedure allows us to obtain threshold shrinkage estimators based on model selection and multiple shrinkage estimators based on model averaging.

The approach proposed in this paper also falls within a Bayesian hierarchical framework. Posterior means and posterior samples are used to de-noise the data and estimate the underlying signal. However, the approach here is different from some others because it establishes prior dependency in the Wavelet domain. Additionally, the prior is robust because it is defined as a scale mixture of multivariate Normal distributions. This gives some elasticity to the approach in tuning the shrinkage via the hyperparameters of the model. Exact posterior distributions are not available in analytic form and so Markov chain Monte Carlo (MCMC) methods are necessary to implement the *Bayes shrinkage*.

The paper is organized as follows. After a short introduction to the DWT in the next section, the model and the induced shrinkage are described in the subsequent two sections. The details related to the Markov chain Monte Carlo simulation method are presented in the fifth section. Finally, three applications in data de-noising are discussed in the sixth section.

## Introduction to the Discrete Wavelet Transformation

Basics on Wavelets can be found in many different texts, monographs and papers at many different levels of exposition. The interested reader is directed to Daubechies (1992) and Meyer (1993), and the textbooks by Vidakovic (1999) and Efromovich (1999), among others. For completeness, a brief review of the Discrete Wavelet transformation (DWT) used extensively in this paper is presented.

Let $y$ be a data vector of dimension (size) $N$, where $N$ is a power of 2, and suppose that DWT is applied to the vector $y$ and transformed into a vector $d$, i.e., $d = Wy$. This transformation is linear and orthogonal and can be represented by an orthogonal matrix $W$ of dimension $N \times N$. In practice, one performs the DWT without explicitly exhibiting the matrix $W$ and by using fast filtering algorithms based on the so-called quadrature mirror filters that uniquely correspond to the Wavelet of choice. More precisely, the Wavelet

decomposition of the vector $y$ is a vector $d$ given by

$$d = (H^n y, GH^{n-1}y, \ldots, GH^2 y, GHy, Gy) \tag{1}$$

The operators $G$ and $H$ act on sequences and are defined via

$$(Ha)_k = \sum_n h_{n-2k} a_n,$$

$$(Ga)_k = \sum_n g_{n-2k} a_n \tag{2}$$

where $g$ and $h$ are high- and low-pass quadrature mirror filters corresponding to the Wavelet basis. For instance, in the case of the DAUB#2 Wavelet from the well-known Daubechies' family of wavelets, the low-pass filter is given by

$$h = (0.4829629131, 0.8365163037, 0.2241438680, -0.1294095226).$$

The elements of $d$ are called 'Wavelet coefficients'. The sub-vectors given in equation (1) represent different levels in the pyramid indexing of the Wavelet coefficients. For instance, the vector $Gy$ contains $N/2$ coefficients representing the level of finest detail, the $(n-1)$st level. These elements are represented by $(d_{n-1,0}, d_{n-1,1}, \ldots, d_{n-1,N/2-1})$.

In general, the level $j$ of the Wavelet decomposition of $y$ is a vector that contains $2^j$ elements and is represented by

$$GH^{n-j-1}y = (d_{j,0}, d_{j,1}, \ldots, d_{j,\,2^j-1}) \tag{3}$$

The main strength of the DWT in Statistics is that it unbalances the data. In addition, the transformation induces local-in-time and time-space plane divisions that form unconditional bases for a range of function spaces. These properties explain the good performance of shrinkage methods in the Wavelet domain.

## The Statistical Model

Let $y = (y_1, y_2, \ldots, y_n)$ be a vector of equally spaced observations of size $n$ whose elements satisfy the relation

$$y_i = f_i + \varepsilon_i : i = 1, 2, \ldots, n \tag{4}$$

where $f_i$ is the underlying signal generating the observed process and $\varepsilon_i$ forms a sequence of independent and identically distributed random variables with constant variance $\sigma^2$. This model can be rewritten in vector form as

$$y = f + \varepsilon \tag{5}$$

where $f = (f_1, f_2, \ldots, f_n)$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$.

From a statistical standpoint, the problem of data de-noising is addressed as the problem of estimating the unknown vector $f = (f_1, f_2, \ldots, f_n)$ given the observed data $y$.

Furthermore, after applying a DWT to the data, we have that

$$\hat{d} = d + \varepsilon' \tag{6}$$

where $\hat{d} = Wy$, $d = Wf$, and $\varepsilon' = W\varepsilon$ with $\varepsilon' = (\varepsilon'_1, \varepsilon'_2, \ldots, \varepsilon'_n)$. Since $W$ is an orthogonal matrix, if we assume that each $\varepsilon_i$ has a $N(0,\sigma^2)$ distribution, the probability distribution of $\varepsilon$ is the same as the probability distribution for $\varepsilon'$. This is a multivariate Normal with a zero mean vector, a covariance matrix $\sigma^2 I_{n \times n}$ and $I_{n \times n}$ denotes an identity matrix of dimension $n$.

For every DWT there is an Inverse Discrete Wavelet Transformation (IDWT) that permits the reconstruction of the signal from the Wavelet coefficients through the expression

$$\hat{f} = W^t \hat{d} \tag{7}$$

where $W^t$ is the transpose matrix of $W$. In consequence, inferences over the Wavelet coefficients $d$ produce inferences on the signal of interest $f$.

As mentioned in the first section, the problem of estimating $d$ can be approached from different perspectives. Here, the Bayesian paradigm is adopted for which it is necessary to specify a prior distribution on $(d,\sigma^2)$, which we denote by $\pi(d,\sigma^2)$. By Bayes theorem we obtain the joint posterior distribution for $(d,\sigma^2)$ given $\hat{d}$,

$$\pi(d, \sigma^2|\hat{d}) \propto f(\hat{d}|, d, \sigma^2)\pi(d, \sigma^2) \tag{8}$$

where $f(\hat{d}|, d, \sigma^2)$ defines the likelihood function for $(d,\sigma^2)$ which in our model is determined by equation (6).

The scale parameter $\sigma^2$ is a nuisance parameter, we marginalize the joint posterior with respect to $\sigma^2$ to obtain the marginal posterior distribution of $d$ given $\hat{d}$, i.e.,

$$\pi(d|\hat{d}) = \int_0^\infty \pi(d, \sigma^2|\hat{d})d\sigma^2 \tag{9}$$

The posterior expectation of $d$, $E(d|\hat{d})$ or the posterior median of $d$ can be used as a point estimator for $d$. However, it is very common in Bayesian Statistics that expressions for moments of posterior distributions are impossible to integrate, and so Bayesian computation methods based on Markov chain Monte Carlo (MCMC) simulation is a powerful tool to compute integrals associated with the posterior distribution.

## Likelihood and Prior Specification

For a full Bayesian analysis, it is necessary to specify the likelihood function that we define here by a multivariate Normal distribution with vector mean $d$ and covariance matrix $\sigma^2 I_{n \times n}$. That is

$$\pi(\hat{d}|d, \sigma^2) \sim N(d, \sigma^2 I_{n \times n}) \tag{10}$$

For the prior specification of $d$, Müller & Vidakovic (1995) suggest using a heavy-tailed prior distribution on $d$ with component-wise dependence. The independence assumption among the Wavelet coefficients is unrealistic, although the DWT tends to eliminate correlation in the data. Accordingly, the following hierarchical prior for $(d,\sigma^2)$ is proposed for the Bayesian shrinkage.

At the first level of the hierarchy, $d$ follows a multivariate Normal distribution with vector mean 0 and covariance matrix $\tau^2 \Sigma$, i.e.

$$\pi(d|\tau^2) \sim N(0, \tau^2 \Sigma) \tag{11}$$

where $\tau^2$ is a scale parameter and $\Sigma$ is a $n \times n$ matrix that defines the prior correlation structure among the Wavelet coefficients.

At the second level, the scale parameter $\sigma^2$ has a prior distribution that follows an Inverse Gamma with hyperparameters $\alpha_1$ and $\delta_1$ or

$$\pi(\sigma^2) \sim IG(\alpha_1, \delta_1) \tag{12}$$

Also, an Inverse Gamma prior $(\alpha_2, \delta_2)$ is specified over the scale parameter $\tau^2$, i.e.

$$\pi(\tau^2) \sim IG(\alpha_2, \delta_2) \tag{13}$$

The additional scale parameter $\tau^2$, not included by Müller & Vidakovic (1995), may induce some extra shrinkage to the Wavelet coefficients leading to a more flexible Bayes Wavelet shrinkage approach. This hierarchical prior implies a marginal multivariate $t$ prior on $d$, however there is no useful closed form expression for the marginal posterior $\pi(d|\hat{d})$.

## MCMC Method

To obtain posterior inferences on the vector of Wavelet coefficients $d$, a standard Gibbs sampling procedure is adopted. Gibbs sampling is an iterative algorithm that simulates from a joint posterior distribution through iterative simulation of the full conditional distributions. For more details and various examples on Gibbs sampling and other MCMC methods see Gamerman (1997), Tanner (1996) or Casella & Robert (1999).

For the hierarchical prior and likelihood function presented in the previous section, the full conditionals for $d$, $\sigma^2$ and $\tau^2$ can be determined exactly. In fact, by a direct application of Bayes theorem it can be shown that the full conditional for $d$ given $\tau^2$, $\sigma^2$ and $\hat{d}$, is a multivariate Normal distribution with vector mean $m^*$ and covariance matrix $\Sigma^*$,

$$\pi(d|\tau^2, \sigma^2, \hat{d}) \sim N(m^*, \Sigma^*) \tag{14}$$

where $m^* = \Sigma^* \sigma^{-2} \hat{d}$ and $\Sigma^* = (\tau^{-2} \Sigma^{-1} + \sigma^{-2} I_{n \times n})^{-1}$.

Additionally, the full conditional distribution of $\sigma^2$ given $d$, $\tau^2$ and $\hat{d}$ is an Inverse Gamma distribution with parameters $\alpha_1^*$ and $\delta_1^*$,

$$\pi(\sigma^2|d, \tau^2, \hat{d}) \sim IG(\alpha_1^*, \delta_1^*) \tag{15}$$

with $\alpha_1^* = n/2 + \alpha_1$ and $\delta_1^* = (d - \hat{d})^t (d - \hat{d})/2 + \delta_1$.

Furthermore, the full conditional distribution for $\tau^2$ given $d$, $\sigma^2$ and $\hat{d}$ is also an Inverse Gamma with parameters $\alpha_2^*$ and $\delta_2^*$, i.e.

$$\pi(\tau^2|d, \sigma^2, \hat{d}) \sim IG(\alpha_2^*, \delta_2^*) \tag{16}$$

with $\alpha_2^* = n/2 + \alpha_2$ and $\delta_2^* = [d^t \Sigma^{-1} d]/2 + \delta_2$. The implementation of the Gibbs sampler with this model only requires simulation routines for the multivariate Normal and the Inverse Gamma distributions. There is a Splus/R program available from the author to implement the methodology described in this paper.

The selection of the hyperparameters for the hierarchical prior deserves attention since its values could mostly determine the resulting shrinkage and the respective reconstruction

of the signal. A convenient selection for the $\Sigma$ matrix is to fix it as block diagonal matrix with each block of the form $\lambda_k \Sigma_k$. Each of the terms $\Sigma_k$ defines a correlation structure inside the $k$th level of coefficients for the Wavelet decomposition. The values of $\lambda_k$ are intended to tune the amount of shrinkage at level $k$. The block diagonal assumption for $\Sigma$ establishes no correlation between coefficients at different levels of the Wavelet decomposition of the data. Following Müller & Vidakovic (1998), the applications presented in the next section are made with $\Sigma_k$ defined as a matrix with entries $\Sigma_{ij} = \rho^{|i-j|}$, so the larger the difference between the sub-indexes $i$ and $j$, the smaller the correlation between coefficients. $\rho$ is a scalar quantity in the interval $(0, 1)$. This structure defines the so-called *Laurent* matrix. See Vidakovic (1999) for further discussion about this matrix specification.

## Applications

### Benchmark Signals

The Bayes Wavelet shrinkage is illustrated with the test functions *HeaviSine, Blocks, Bumps* and *Doppler* used by Vannucci & Corradi (1999). These functions are often used as a benchmark in the context of Wavelet shrinkage since they represent signals with patterns that arise in several scientific fields. Figure 1 presents the four signals based on 512 observations and Figure 2 shows the same signals but corrupted by a Gaussian noise $N(0, \sigma^2)$. As in Vannucci & Corradi (1999), the value of $\sigma$ was chosen to establish a signal-to-noise ratio of 5.

Figure 3 shows the reconstructed signals using the proposed Bayes Wavelet shrinkage. The number of Gibbs sampling iterations was 2,000 and the posterior means over MCMC iterations were used to obtain point estimates of the underlying signals. The data were transformed into the Wavelet domain using the wavelet $S8$, which is one of the symmlets and default for the package Splus Wavelets. This basis provides a sensible balance between the compactness of support and smoothness of the Wavelet.
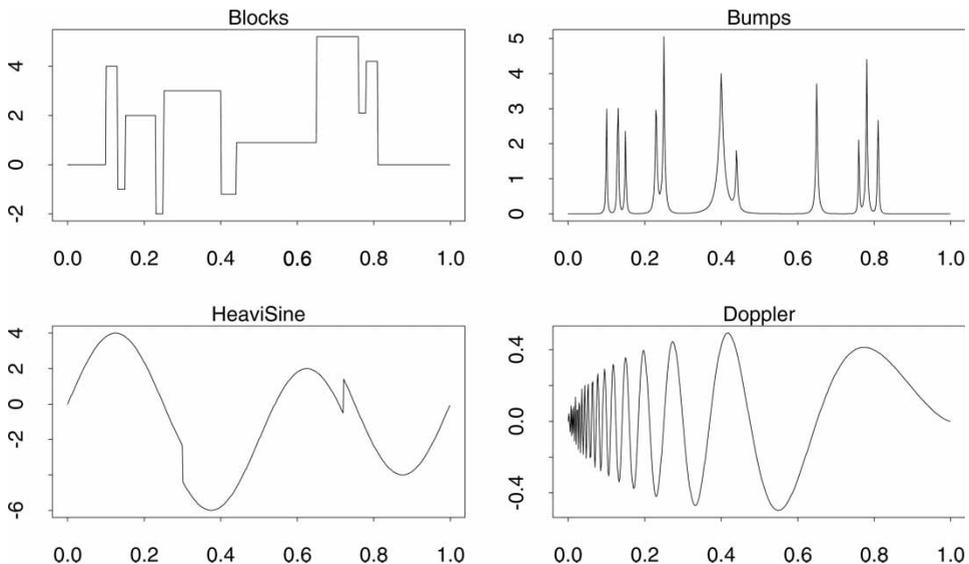


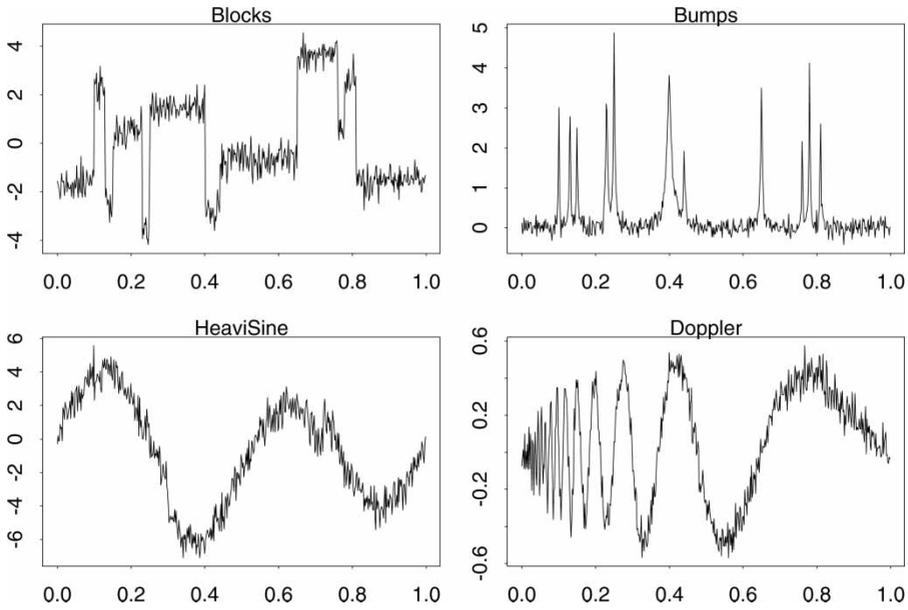**Figure 1.** The four test signals: *Blocks*, *Bumps*, *HeaviSine* and *Doppler*

**Figure 2.** The four test signals corrupted by a Gaussian white noise

For the specification of $\Sigma$, a Laurent-matrix form was adopted with $\rho = 0.5$, $\lambda_1 = 100$, $\lambda_2 = \lambda_3 = 10$, $\lambda_4 = 1$, $\lambda_5 = 0.1$, $\lambda_6 = 0.01$ and $\lambda_7 = 0.001$. This specification tries to keep the same the smooth part and first level of details after the shrinkage is performed. Coefficients at the third and fourth level will be somehow shrunk and, for the fifth, sixth and seventh levels, the coefficients will be practically made zero. All the
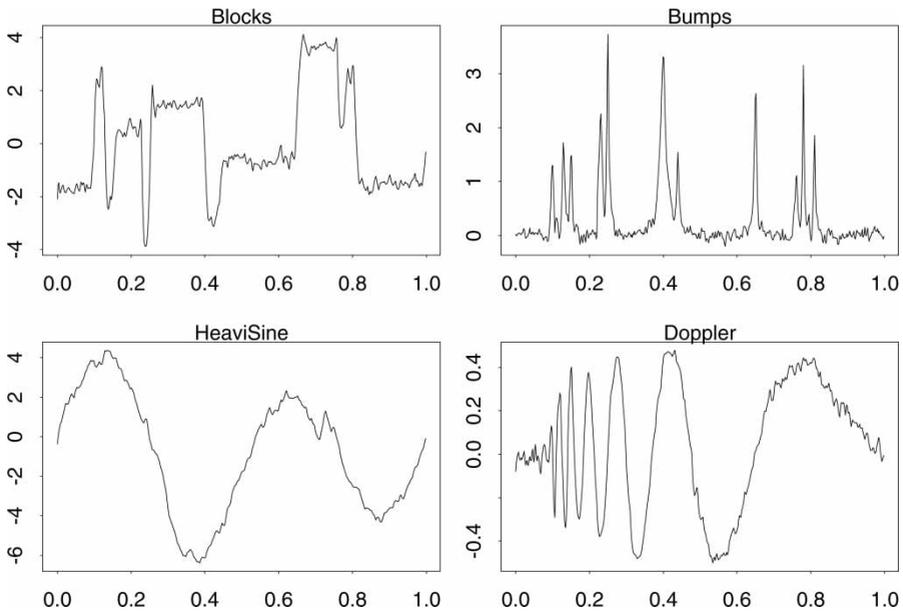


**Figure 3.** The four test signals reconstructed with the Bayesian Wavelet Shrinkage

hyperparameters of the priors for $\sigma^2$ and $\tau^2$ were set equal to one. From the results it can be noticed that the method gives a reasonable reconstruction of all the four signals, especially in the cases of *Bumps* and *HeaviSine*. However, in all the reconstructions there is a bit of noise that remains present in the estimate, particularly with *Blocks* and *Doppler*. This may be due to the Bayes rule adopted since the posterior mean only performs a shrinkage of the Wavelet coefficients, but never sets these coefficients to a zero value as in thresholding rules. The results are comparable to those reported for the same signals by Vannucci & Corradi (1999).

### Microscopy Example

To illustrate the proposed shrinkage method, we use 1024 measurements in atomic force microscopy (AFM) that appears in the top frame of Figure 4. The AFM is a type of scanned proximity probe microscopy (SPM) that can measure the adhesion strength between two materials at the nano-newton scale. In AFM, a cantilever beam is adjusted until it bonds with the surface of a sample, then the force required to separate the beam and sample is measured from the beam deflection. Beam vibration can be caused by factors such as thermal energy of the surrounding air or the footsteps of someone outside the laboratory. The vibration of a beam acts as noise on the deflection signal so in order for the data to be useful this noise must be removed. Similar data were analysed via a $\Gamma-$ Minimax Wavelet Shrinkage approach in Angelini & Vidakovic (2004). The middle frame of Figure 4 shows the Bayes Shrinkage reconstruction based on the IDWT of the posterior mean of 2000 MCMC draws. The prior specification is the same as with the test functions discussed in the previous subsection. For comparison, the bottom frame of Figure 4 shows the
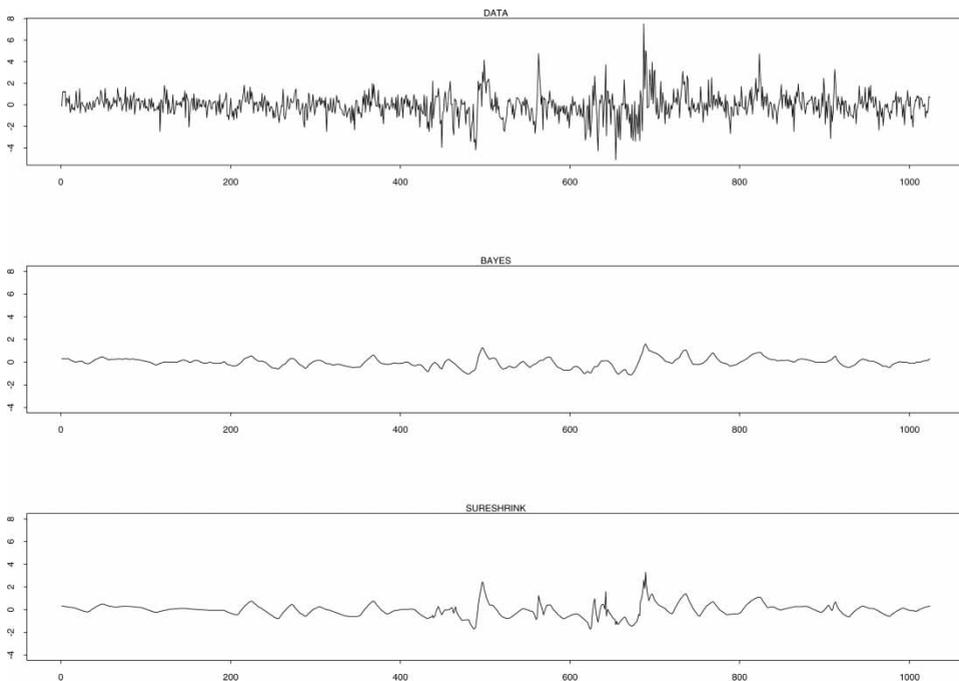


**Figure 4.** Microscopy data, Bayes Shrinkage reconstruction and Sure-shrink reconstruction

reconstruction of the signal using a *Sure-Shrink* approach. Notice that the reconstruction of the signal with the Bayes shrinkage is quite smooth, while the reconstruction with Sure-Shrink is less smooth and includes some noisy artefacts around the middle section of the data.

The multiresolution analysis (MRA) of the original data and the reconstructed signal using Bayes shrinkage appear in Figure 5. The rows indexed $d_1 - d_6$ include the values of the coefficients for the details, $d_1$ representing the finest level and $d_6$ the coarsest level. The row indexed by $s_6$ includes the coefficients of the smooth level. From this figure it can be noticed that the proposed Bayes method shrinks the values for the fine levels of details that are close to zero, while keeping practically intact coefficients corresponding to the smooth part or high levels of details ($s_6$, $d_6$).

The actual AFM measurements (circles) and 50 posterior samples of the Bayes reconstruction of the signal appear in Figure 6. For each MCMC iteration, the IDWT was applied to the sampled coefficients to obtain a realization of the posterior distribution of the signal (parameter) of interest. This plot of several realizations gives an idea of the general structure and the level of uncertainty of estimation for the underlying signal. This goes beyond summarizing results based only with a point estimate.

*Glint Data*

The Glint data presented by Bruce & Gao (1994), consists of 512 equally spaced observations. The true signal is a low frequency oscillation about zero, resulting from rotating an aeroplane model. As the model rotates, the centre of mass will shift slightly.
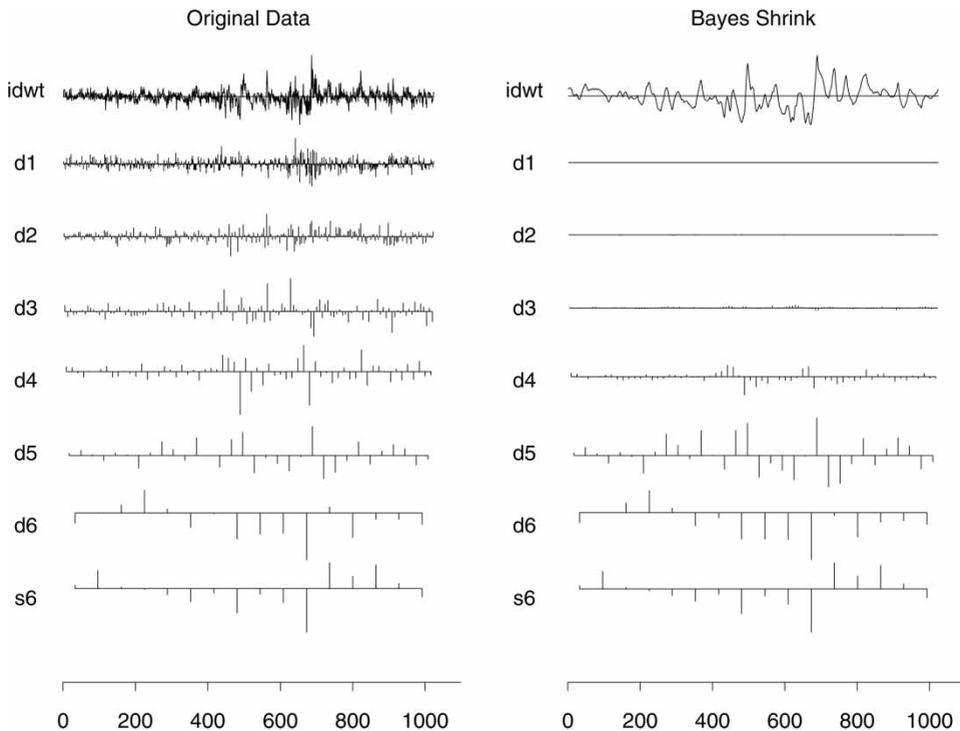


**Figure 5.** *MRA* of the Microscopy data and its Bayes Wavelet Shrinkage estimate
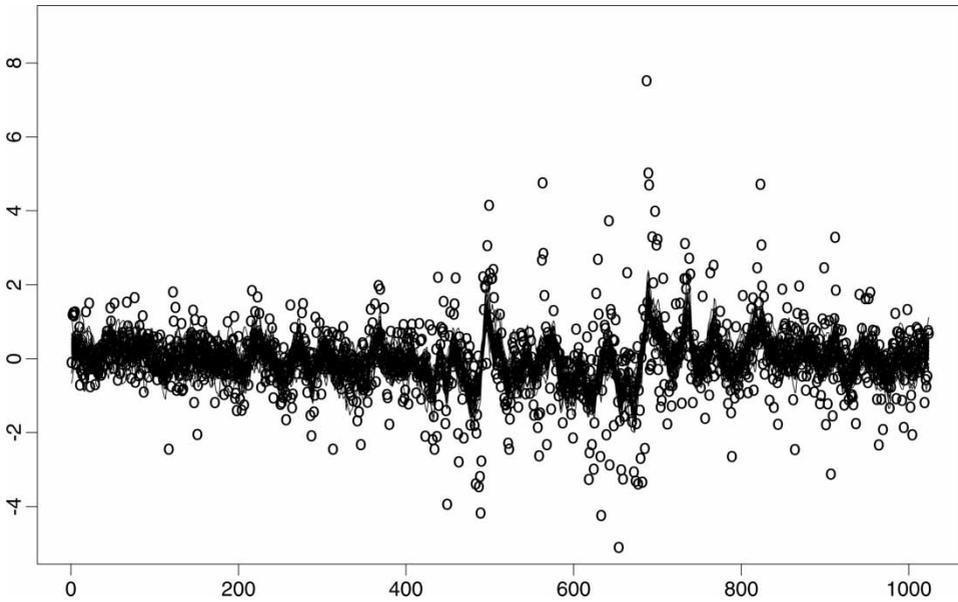
**Figure 6.** Microscopy data and 50 posterior samples of signal reconstruction

The measurements, given in angles, are subject to large errors represented by a large number of spikes, and can be off the true signal by as much as 150 degrees. Data denoising of the Glint signal is very challenging. A time series plot of the original Glint data appears as a dotted line in Figure 7.
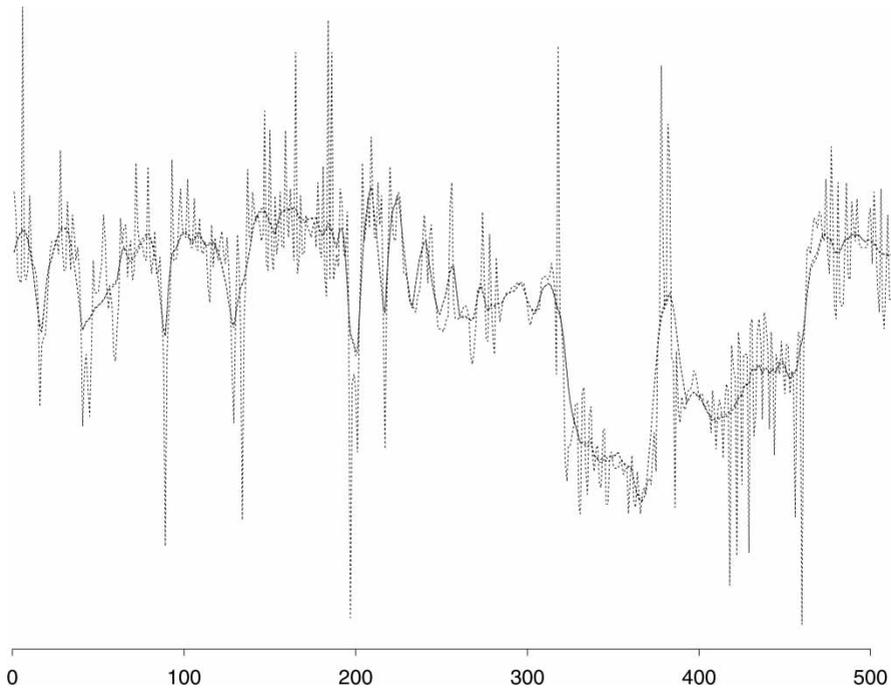


**Figure 7.** Glint data and reconstruction

The DWT was applied to Glint with the symmlet wavelet $S8$. The prior specification for all the hyperparameters is defined in the same way as in the previous applications. Based on 2,000 MCMC draws, the posterior mean was calculated to approximate $E(d|\hat{d})$. Then, the IDWT was applied to this estimated posterior mean to obtain an estimate or reconstruction of the underlying signal, which appears in Figure 7 as a solid line. Notice that the reconstruction eliminates the high peaks and produces a smooth signal around zero.

It is important to note here that the reconstruction obtained stands between universal-hard thresholding and the multiple shrinkage proposed by Clyde *et al.* (1998). Comparing Figure 7 with Figures 1–5 in that paper, we can realize that our shrinkage is not as spiky as thresholding but not as smooth as the one corresponding to multiple shrinkage. Possible this is due to the non-linearity of our Bayesian shrinkage and because our approach does not permit any coefficient to be exactly zero.

In fact, in Figure 8 we have the time series of the original coefficients (a) with three shrinkage curves (b)–(d). The shrinkage curve is defined by a scatter plot of the original
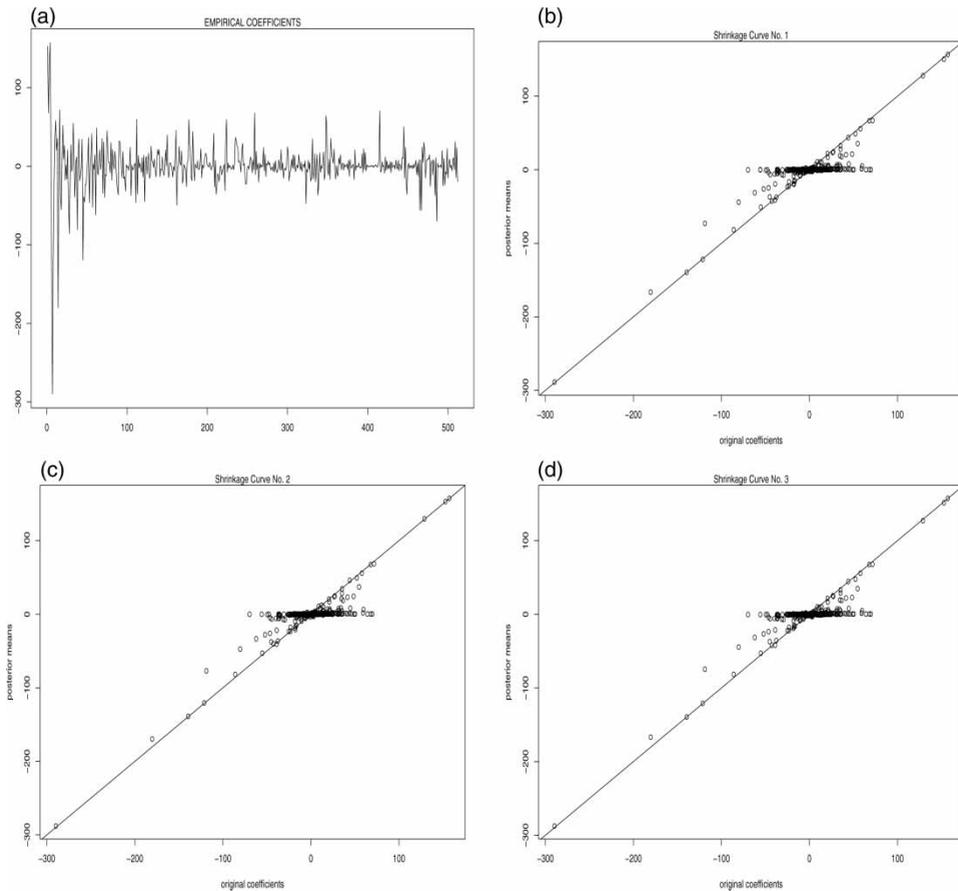


**Figure 8.** (a) Top left: Shrinkage curve no. 1, $\lambda_1 = 100$; $\lambda_2 = 10$; $\lambda_3 = 10$; $\lambda_4 = 1$; $\lambda_5 = 0.1$ and $\lambda_6 = 0.01$. (b) Top right: Shrinkage curve no. 2, $\lambda_1 = 10$; $\lambda_2 = 10$; $\lambda_3 = 1$; $\lambda_4 = 1$; $\lambda_5 = 0.1$ and $\lambda_6 = 0.1$. (c) Bottom left: Shrinkage curve no. 3, $\lambda_1 = 10$; $\lambda_2 = 1$; $\lambda_3 = 1$; $\lambda_4 = 0.1$; $\lambda_5 = 0.01$ and $\lambda_6 = 0.01$. (d) Bottom right: Shrinkage by levels, $\lambda_1 = 1000$; $\lambda_2 = 1000$; $\lambda_3 = 10$; $\lambda_4 = 1$; $\lambda_5 = 0.1$ and $\lambda_6 = 0.01$
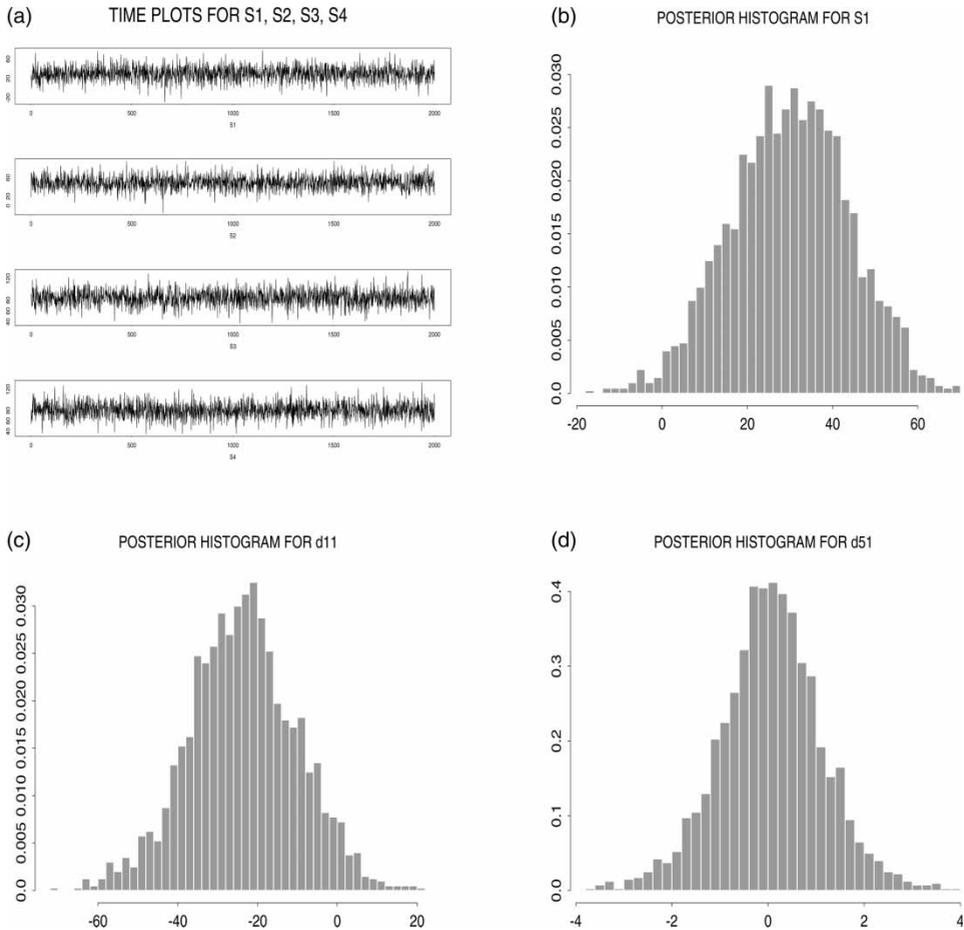
**Figure 9.** (a) Top left: 2000 iterations of the MCMC for the coefficients of the smooth part. (b) Top right: Posterior histogram for the coefficient S1 corresponding to the smooth part. (c) Bottom left: Posterior histogram for the coefficient d11 corresponding to the first level of details. (d) Bottom right: Posterior histogram for the coefficient d51 corresponding to the fifth level of details.

coefficients versus their posterior means. Notice that the non-linearity in the transformation and the non-monotonicity is enforcing a stronger shrinkage at higher level of details. Another point to observe from this figure is that the shrinkage curves correspond to different combinations of the $\lambda_k$ parameters. This explores the sensitivity of the results with respect to the hyperparameters specification. Essentially, no changes are observed across the three shrinkage curves, and so the corresponding reconstructions look exactly as the one shown in Figure 7.

Figure 9(a) presents a time series plots for the 2,000 iterations of the simulated values for some of the Wavelet coefficients. The MCMC moves rapidly through the parameter space, and this implies that there is no need to iterate more the Gibbs sampler to achieve posterior convergence. In Figures 9(b)–(d) histograms of the posterior simulations are presented for a coefficient corresponding to the smooth part, a coefficient of the second level of details and a coefficient of the last level of details. All the features

of the posterior distribution will appear in such pictures (symmetry, fat tails, etc). In this example, it can be observed that the smooth part and first level of details will remain the same while, for the last level, the posterior distribution has most of its mass centred at zero.

## Concluding Remarks

The goal of this paper is to present a hierarchical Bayesian model that allows for data de-noising using Wavelet shrinkage through Gibbs sampling and that has the possibility of tuning the shrinkage with the hyperparameters of the model. The shrinkage method is illustrated with three different data sets. For the Glint signal, no differences were observed for the resulting shrinkage for different values of the hyperparameters $\lambda_k$. For other applications it is possible that the model may be more sensitive to these parameters On the other hand, it is feasible to learn about the hyperparameters $\lambda_k$ and $\rho$ by adding another level or hierarchy into the model. For example, Vannucci & Corradi (1999) propose a Beta prior on $\rho$ and an Inverse Gamma prior on $\lambda$. In this scenario, the proposed MCMC method of the fifth section becomes a hybrid algorithm that requires a Metropolis–Hastings step to simulate the full conditional distribution of $(\rho, \lambda)$. This extension is proposed for future research; however, preliminary results under such model specifications prove to be of little improvement with respect to pre-specified hyperparameter values. The method presented here is computationally more intensive than other Bayes shrinkage methods due to the simulations of the multi-variate Normal distribution but it has the advantage of being robust against large errors in the data.

## Acknowledgement

## References

Angelini, C. & Vidakovic, B. (2004) Gamma-minimax wavelet shrinkage: a robust incorporation of information about energy of a signal in denoising applications. *Statistica Sinica*, 14, pp. 103–125.

Bruce, A. & Gao, H.-Y. (1994) *S + Wavelets, Users Manual* (Seattle, WA: StatSci).

Casella, G. & Robert, C. (1999) *Monte Carlo Statistical Methods* (New York: Springer-Verlag).

Chipman, H., Kolaczyk, E. & McCulloch, R. (1997) Adaptive Bayesian wavelet shrinkage, *Journal of the American Statistical Association*, 92, pp. 1413–1421.

Clyde, M. & George, E. I. (2000). Flexible empirical Bayes estimation for wavelets, *Journal of the Royal Statistical Society, Series B*, 62, pp. 681–698.

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

Daubechies, I. (1992) *Ten Lectures on Wavelets* (Society for Industrial and Applied Mathematics).

DeVore, R. A., Jawerth, B. & Lucier, B. J. (1992) Image compression through wavelet transform coding, *IEEE Transactions on Information Theory*, 38(2), pp. 719–746.

Donoho, D., Johnstone, I., Kerkyacharian, G. & Picard, D. (1995) Wavelet shrinkage: asymptopia?, *Journal of the Royal Statistical Society, Series B*, (2), 301–369.

Efromovich, S. (1999). *Nonparametric Curve Estimation* (New York: Springer Verlag).

Gamerman, D. (1997) *Markov Chain Monte Carlo: Stochastic Simulation for Posterior Distributions* (Chapman and Hall).

Meyer, Y. (1993) *Wavelets. Algorithms and Applications* (SIAM).

Müller, P. & Vidakovic, B. (1998) Bayesian inference with wavelets: density estimation. *Journal of Comput. Graph. Statist*, 7(4), pp. 456–468.

Tanner, M. A. (1996) *Tools for Statistical Inference: Methods for Likelihood Approximations and Posterior Distributions*, 3rd edn (New York: Springer Verlag).

Vannucci, M. & Corradi, F. (1999) Covariance structure on wavelet coefficients: theory and models in a Bayesian perspective, *Journal of the Royal Statistical Society, Series B*, 61(4), pp. 971–986.

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

Vidakovic, B. (1999) *Statistical Modeling by Wavelets* (New York: Wiley).