Bayesian False Discovery Rate Wavelet Shrinkage: Theory and Applications

ILYA LAVRIK, YOON YOUNG JUNG, FABRIZIO RUGGERI, AND BRANI VIDAKOVIC

Abstract. Statistical inference in the wavelet domain remains vibrant area of contemporary statistical research because desirable properties of wavelet representations and the need of scientific community to process, explore, and summarize massive data sets. Prime examples are biomedical, geophysical, and internet related data.

In this paper we develop wavelet shrinkage methodology based on testing multiple hypotheses in the wavelet domain. The shrinkage/thresholding approach by implicit or explicit simultaneous testing of many hypotheses had been considered by many researchers and goes back to the early 1990's. Even the early proposal, the universal thresholding, could be interpreted as a test of multiple hypotheses in the wavelet domain. We propose two new approaches to wavelet shrinkage/thresholding.

(i) In the spirit of Efron and Tibshirani's recent work on local false discovery rate, we propose the theoretical counterpart Bayesian Local False Discovery Rate, BLFDR, where the underlying model assumes unknown variances. This approach to wavelet shrinkage can also be connected with shrinkage based on Bayes factors.

(ii) The second proposal to wavelet shrinkage explored in this paper is Bayesian False Discovery Rate, BaFDR. This proposal is based on ordering of posterior probabilities of hypotheses in Bayesian testing of multiple hypotheses.

We demonstrate that both approaches result in a competitive shrinkage methods by contrasting them to some popular shrinkage techniques.

KEY WORDS: Shrinkage; Multiple Hypotheses Testing, False Discovery Rate, Bayesian Local False Discovery Rate.

1 Introduction

In this paper we introduce wavelet-based shrinkage based on two versions of false discovery rate: local FDR and Bayesian FDR based on selecting dominant posterior probabilities. The developed methodology is comparable to currently best available wavelet shrinkage methods. Even though

the two proposed methods may not achieve the minimum of MSE they possess two distinct qualities: (i) they are of thresholding type leading to most parsimonious representations desirable when dimension reduction is an issue, and (ii) the bias of obtained estimators is small.

Simultaneous testing of multiple hypotheses has always attracted the attention of statisticians (e.g., Folks, 1984) but traditionally, the number of hypotheses was modest (say, < 20). Nowadays, thousands of hypotheses need to be tested simultaneously and the traditional methods (such as Bonferroni, for example) are not sensible because of loss of specificity and power.

To illustrate the loss of specificity, consider a gene expression example. Assume that a chip contains 10000 genes and not a single gene is differentially expressed. If we test each of 10000 hypotheses separately at level $\alpha = 0.01$, we would expect that 10000 $\times 0.01 = 100$ of the tests would have p-value less than α, i.e., 100 of the tests would be falsely significant and the probability that at least one p-value will be less than α (family-wise error rate) is around 1. Thus, individual p-values are no longer valid measures of significant findings.

For controlling the FWER (family-wise error rate, Dudoit *et al.*, 2003), conservative methods such as Bonferroni correction is widely used, however this method also suffers from the lack of power when the number of hypotheses is large. For microarray data, for example, the goal is to focus on several candidate genes for further study. Thus, the low power of FWER-controlling procedure is unacceptable and it would be better to control the false discovery rate, FDR, a method that is discussed in some detail in following Sections.

To formally illustrate what happens in a testing problem when the number of hypotheses to be tested simultaneously increases, we consider paradigmatic problem of testing for the multivariate normal mean.

Suppose we wish to test

$$H_0 : \theta = 0 \ \ vs. \ \ H_1 : \theta \neq 0, \tag{1}$$

where $\mathbf{X} \sim \mathcal{MVN}_n(\theta, I_n)$ is observed. A sufficient statistics for the problem is $||\mathbf{X}||^2$. If the alternative is precise, $\theta = \theta_1$, then the α-level maximum likelihood ratio test has approximate power

$$1 - \Phi\left(\frac{z_{1-\alpha} - ||\theta_1||^2/\sqrt{2n}}{\sqrt{1 + 2||\theta_1||^2/n}}\right) \approx 1 - \Phi\left(z_{1-\alpha} - ||\theta_1||^2/\sqrt{2n}\right). \tag{2}$$

If $||\theta_1||^2$ goes to infinity, the power of the test is expected to tend to 1, however, if $||\theta_1||^2$ goes to infinity as $o(\sqrt{n})$, the power tends to α when n increases. Thus, for high-dimensional θ, the discriminatory distance $||\theta_1||^2$ is shrunk to $||\theta_1||^2/\sqrt{2n}$, and the power tends to the significance level.

This dissipation of power, while testing multiple hypotheses is discussed by many researchers. Folks (1984) gives an excellent overview of multiple hypothesis testing, including the Tippet method which can be viewed as a precursor of FDR method of Benjamini and Hochberg (1995), and local FDR discussed in Section 2. Another classical repository of methods used in multiple hypothesis testing is monograph by Miller (1981).

In his spirited paper with applications in genomics, Efron (2004) found that local FDR tends to overfit the model. He demonstrated that replacing the "theoretical null distribution" by its empirical counterpart often improves the model selection. We connect local FDR approach (Efron and

Tibshhirani, 2002; Efron, 2004) to the related model selection procedure based on Bayes factors in the context of wavelet-smoothing. Consistently with Efron's findings, the empirical H_0 density in the local FDR tends to have longer tails, leading to more parsimonious models. An efficient proposal to replace theoretical null by empirical null based on nonparametric version of Empirical Bayes estimator is proposed by Datta and Datta (2004).

The paper is organized as follows. Section 2 introduces local false discovery rate in wavelet domain, Section 3 introduces thresholding by ordering posterior probabilities (BaFDR), Section 4 presents simulational results for four standard signals, Section 4.4 shows application of the LFDR and BaFDR to an atomic force microscopy signal.

2 Local False Discovery Rate in the Wavelet Domain

Many proposed wavelet shrinkage methods can be interpreted as multiple hypotheses testing in the wavelet domain. For example the *universal thresholding* of Donoho and Johnstone (1994), *recursive likelihood ratio tests* of Ogden and Parzen (1996), *false discovery rate* of Abramovich and Benjamini (1995, 1996) are some early references. Vidakovic (1997) proposes the use of Bayesian hypothesis testing and Bayes factors in the tasks of wavelet thresholding. Vidakovic and Ruggeri (2001) develop an adaptive Bayesian model in which the resulting Bayes rule acts as a shrinker in the wavelet domain. Their method (Bayesian Adaptive Multiscale Shrinkage, or short BAMS) is now part of *GaussianWaveDen* of Antoniadis, Bigot,and Sapatinas (2001) and allows for incorporation of prior information about the signal. We review the local false discovery rate and establish the link with Bayes factor shrinkage induced by BAMS model, all in the context of wavelet shrinkage.

Suppose the observed data $\boldsymbol{y} = (y_1, \ldots, y_n)$ represent the sum of an unknown signal $\boldsymbol{f} = (f_1, \ldots, f_n)$ and random noise $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$. Coordinate-wise, $y_i = f_i + \epsilon_i,\ i = 1, \ldots, n$. In the wavelet domain (after applying a linear and orthogonal wavelet transformation $\boldsymbol{W}$ to the observed data), expression (3) becomes $d_{jk} = \theta_{jk} + \epsilon_{jk},\ i = 1, \ldots, n$, where d_{jk}, θ_{jk}, and ϵ_{jk} are the j, k-th coordinates in the traditional scale/shift wavelet-enumeration of vectors $\boldsymbol{Wy}, \boldsymbol{Wf}$ and $\boldsymbol{W\epsilon}$, respectively. Our assumption is that the coefficients d_{jk} can be considered independently, since the wavelet transformations are decorrelating. When modeling in practice, such an assumption prove to be very reasonable. In the exposition that follows, we omit the double index jk and work with a "typical" wavelet coefficient, d. Therefore, our model is

$$d = \theta + \epsilon. \tag{3}$$

One way to select the parsimonious model is to componentwise test that the signal part of the coefficient is zero, i.e., $H_0 : \theta = 0$. If the hypothesis is rejected the coefficient is significant and retained in the model. If the H_0 is accepted, then d in the model is replaced by 0. After all n tests are conducted, the coefficients that survived the tests are back transformed to the domain of the original data.

When n simultaneous null hypotheses are tested, the corresponding test statistics (likely not all independent) will result in n p-values, $p_1, p_2, \ldots, p_n$. Under H_0 these p-values represent a sample from a uniform distribution.

It is more convenient to work with z-values,$z_1, \ldots, z_n$, where

$$z_i = \Phi^{-1}(p_i). \tag{4}$$

Under the H_0 the z_is are theoretically $N(0,1)$, and the standard normal tables could be used.

Define local false discovery rate as the ratio of $f_0(z)$, theoretical null density for zs and $f(z)$ observed empirical density for z's,

$$\textbf{lfdr}(z) = \frac{f_0(z)}{f(z)}.$$

Efron (2004) suggests to keep in the model, as *interesting*, all coefficients for which $\textbf{lfdr}(z_i) = \frac{f_0(z_i)}{f(z_i)}$ is smaller than some threshold value, say $\gamma = 0.10$. As pointed in the same paper, by dropping p_0 which is close to 1 (most of the H_0s are true, only a few coefficients are retained in the model), the probability $P(\text{Uninteresting}|z) = p_0 f_0(z)/f(z)$ is close to $\textbf{lfdr}(z) = f_0(z)/f(z)$ and represents a link with Bayes factor shrinkage proposed by Vidakovic (1997).

Next we introduce the local false discovery rate for a specific model first discussed in the wavelet context by Vidakovic and Ruggeri (2000).

Assume that $[d|\theta, \sigma^2] \sim N(\theta, \sigma^2)$ and the prior $\sigma^2 \sim \mathcal{E}(\mu), \ \ \mu > 0$, with density $f(\sigma^2|\mu) = \mu e^{-\mu\sigma^2}$. The marginal likelihood (with σ^2 integrated out) is

$$[d|\theta] \sim \mathcal{DE}\left(\theta, \frac{1}{\sqrt{2\mu}}\right), \quad \text{with density } f(d|\theta) = \frac{1}{2}\sqrt{2\mu}e^{-\sqrt{2\mu}|d-\theta|}.$$

If the prior on θ is

$$[\theta] \sim \pi_o\delta_0 + \pi_1\mathcal{DE}(0,\tau), \tag{5}$$

then the predictive distribution of d is

$$[d] \sim m(d) = \pi_0\mathcal{DE}\left(0, \frac{1}{\sqrt{2\mu}}\right) + \pi_1 m_1(d),$$

where $m_1(d)$ is

$$\frac{\tau e^{-|d|/\tau} - \frac{1}{\sqrt{2\mu}}e^{-\sqrt{2\mu}|d|}}{2\tau^2 - 1/\mu},$$

The Bayes factor in favor of testing $H_0 : \theta = 0$ versus the alternative $H_1 : \theta \neq 0$, when wavelet coefficient d is observed is

$$B_{01}(d) = \frac{f(d|0)}{m_1(d)} = \frac{\frac{1}{2}\sqrt{2\mu}e^{-\sqrt{2\mu}|d|}}{\frac{\tau e^{-|d|/\tau} - \frac{1}{\sqrt{2\mu}}e^{-\sqrt{2\mu}|d|}}{2\tau^2 - 1/\mu}} \tag{6}$$

By straightforward calculation,

$$f_0/m(d) = \frac{B_{01}(d)}{\pi_1 + \pi_0 B_{01}(d)} \tag{7}$$

and **lfdr** can be estimated by $\hat{B}_{01}(d)/(\pi_1 + \pi_0\hat{B}_{01}(d))$ where $\hat{B}_{01}$ is an empirical counterpart of Bayes factor.

Using the empirical counterpart of Bayes factor means that we choose the empirical null hypothesis instead of theoretical one. Local false discovery rate, lfdr, computed by using the empirical counterpart of Bayes factor ($\hat{B}_{01}$) corresponds to the empirical null hypothesis and lfdr computed by using the exact Bayes factor (B_{01}) corresponds to the theoretical null hypothesis.

In multiple testing problem, Efron (2004) considered the choice of an appropriate density for the null hypotheses, the point there being that large-scale situations can provide their own "empirical null", which may differ in important ways from the traditional theoretical null appropriate for any individual problem. In addition, permutation and bootstrap null density estimates should be considered as improved versions of theoretical null, rather than empirical nulls.

Remark 1. In the illustrative example of gene expression data (Jung et al., 2005), it is shown that the frequently used permutation methods can be misleading when the mean of the distribution of test statistics for most genes(non-differential genes) is not zero because the random permutations of expression levels across the control and treatment groups make the mean of the distribution definitely zero. This would yield a bias in the mean estimate and thus result in inaccurate estimation of FDR.

Jung et al. (2005) proposed the fully Bayesian mixture model-based method in meta-analysis to estimate the null distribution of test statistics and compared it with the permutation methods by computing the FDRs given the critical value. The proposed method was applied to four publicly available prostate cancer gene expression data and the results showed that the model-based approach is superior to the permutation method.

In this example, the mean of test statistics of non-differential genes was estimated 0.177, larger than zero and the null density from permutation method showed the center around zero, showing significant difference from the estimated density in the proposed method. Also for example, with critical value 1.207, the FDR estimated by the mixture model is 0.001, while the FDR estimated by the permutation method is 0.022.

Remark 2. From the form of Bayes factor in (6), it follows that $B_{01} < \alpha$ leads to a thresholding rule $|d| > \lambda$, for some $\lambda = \lambda(\alpha)$.

The Bayes factor provides a measure of data support for the H_0 and is used to calculate posterior odds of H_0 as

$$\frac{p_0}{p_1} = \frac{\pi_0}{\pi_1} B_{01},$$

where $\frac{\pi_0}{\pi_1}$ are prior odds.

A coefficient should remain in the model if the $B_{01} < \alpha$. Since adaptive Bayesian shrinkage uses level varying probability of null hypothesis, $\pi_0 = \pi_0(j)$, where j is the level in the wavelet decomposition, the local false discovery rate is equivalent to the following rule based on Bayes factors,

Keep the wavelet coefficient at level j in the model as interesting if $\quad B_{01} \leq \alpha(j).$

In automatic procedure $\alpha(j)$ is always 1, which reflects the fact that H_0 is more readily rejected. This means that the thresholding is not performed and the coefficient is retained as significant.

3 FDR Ordering of Posterior Probabilities (BaFDR)

As we hinted in the Introduction, wavelet shrinkage, in form of thresholding of wavelet coefficients, can be viewed as a multiple testing problem. For each observed wavelet coefficient $d_i = \theta_i + \sigma\epsilon_i$, consisting of signal part θ_i and the error $\sigma\epsilon_i$ the hypothesis $H_0 : \theta_i = 0$ is tested against the alternative $H_1 : \theta_i \neq 0$. If the hypothesis H_0 is rejected, the coefficient d_i is retained in the model as significant. Otherwise, it is discarded.

For example, the universal threshold can be viewed as a critical value of a test with the level

$$\alpha = P(|d_i| > \sqrt{2\log n}\,\sigma\,|H_0\,) \approx (n\sqrt{\pi\log n})^{-1}.$$

The power of this test against the alternative $H_1 : \theta_i = \theta\ (\neq 0)$ is $O\left(\frac{1}{n\sqrt{\log n}}\right)$.

Universal thresholding is equivalent to a Bonferroni-type procedure. In testing n statistical hypotheses simultaneously, the Bonferroni procedure guarantees that the overall level of the omnibus test is α by setting the levels for the individual hypotheses as $\frac{\alpha}{n}$. For large n, the individual levels $\frac{\alpha}{n}$ become unduly small, leading to loss of "strictness" and dissipation of power. This loss of strictness means that many of $H_0 : \theta = 0$ are accepted, i.e., many observed coefficients are discarded from the model leading to over-smoothing.

A way to control such dissipation of massive acceptance of null hypotheses could be based on the *false discovery rate* (FDR) (Abramovich and Benjamini, 1995, 1996; Benjamini and Hochberg, 1995).

Here is a brief description. Let R be the number of wavelet coefficients retained in the model. If S of them are correctly kept, then $V = R - S$ are erroneously kept. The random variable $Q = V/R$ expresses the error in such a procedure. The false discovery rate of coefficients is the expectation of Q; that is, the expected proportion of coefficients erroneously kept. One maximizes the number of coefficients kept, subject to condition $\mathbb{E}Q \leq \alpha$, for α small.

Several Bayesian alternatives to FDR are proposed from the Bayesian stand point, a nice overview can be found in Tadesse *et al.* (2005).

Rosner and Vidakovic (2000) propose an FDR procedure in which that is based on the assessment of posterior probabilities of hypotheses. An application is given in Angelini and Vidakovic (2003). Suppose that in testing of n hypotheses H_0, we obtain a sequence of their posterior probabilities, $p_0^1, p_0^2, \ldots, p_0^n$. Let $p_{(1)}, p_{(2)}, \ldots, p_{(n)}$ be increasingly ordered posterior probabilities, and $q_{(k)} = 1 - p_{(k)},\ k = 1, \ldots, n$.

When deciding about retaining the wavelet coefficients in the model ("a discovery", "interesting coefficient", etc.) by rejecting corresponding null hypotheses $H_0 : \theta = 0$, one controls the number of hypotheses that are erroneously rejected, V. If the R hypotheses with smallest posterior probabilities are rejected, we require the expectation (with respect to the posterior measure) of $Q = V/R$ not to exceed α. Note that

$$\begin{aligned}\mathbb{E}Q &= \frac{1}{R}\sum_{i=0}^{R} iP(\text{Among } R \text{ rejected hypotheses, the number of erroneously rejected is } i)\\ &= \frac{1}{R}\sum_{i=0}^{R} iP_R(i), \qquad (8)\end{aligned}$$

where the probabilities $P_R(i)$ can be calculated efficiently as the coefficients with powers z^i in generating polynomial

$$\varphi_R(z) = \prod_{k=1}^{R}(q_{(k)} + p_{(k)}z) = \sum_{i=0}^{R} P_R(i)z^i. \tag{9}$$

Thus, this Bayesian FDR procedure (BaFDR) can be summarized as follows:

• **STEP 1.** Find the posterior probabilities p_{jk} of all hypotheses $H_0 : \theta_{jk} = 0$ and order them according to their size.
• **STEP 2.** Fix α small and set $R = 1$.
• **STEP 3.** Increase R by 1. Find $\varphi_R(z)$ using $p_{(1)}, \ldots, p_{(R)}$, and calculate $\mathbb{E}Q$.
• **STEP 4.** If $\mathbb{E}Q \geq \alpha$ then the maximum posterior probability of rejection is $p_{(R-1)}$. **STOP.**
Otherwise, if $\mathbb{E}Q < \alpha$, return to **STEP 3.**

The introduced BaFDR naturally leads to wavelet thresholding. It turns out that such shrinkage is also linked with the shrinkage based on Bayes Factors and **lfdr** discussed in the previous sections.

Note that the posterior probability p_0 of hypothesis H_0 is

$$p_0(d) = \frac{B_{01}(d)}{\frac{\pi_1}{\pi_0} + B_{01}(d)} \tag{10}$$

where d is observed wavelet coefficient and $\frac{\pi_1}{\pi_0}$ are prior odds in favor of H_1. This is an easy reformulation of the definition of Bayes Factor which links the prior and posterior odds:

$$\frac{p_0}{p_1} = B_{01} \times \frac{\pi_0}{\pi_1}.$$

If the hypotheses H_0 is rejected, by (10),

$$p_0(d) \leq \alpha \quad \text{is equivalent to} \quad B_{01}(d) \leq \frac{\alpha}{1-\alpha} \times \frac{\pi_1}{\pi_0}.$$

We provide the simulational results involving the standard test functions and the BaFDR shrinkage. Because of its global nature the resulting shrinkage is inferior to the state of art local, neighborhood-dependent shrinkage methods.

4 Simulations and Application

The same setup is used for both BaFDR, and Local Bayesian FDR in wavelet domain. Four standard test functions (`blocks, bumps, doppler` and `heavisine`) are rescaled so that an added standard normal noise produces a preassigned signal-to-noise ratio (SNR). The wavelet bases used are: Symmlet 8 for `doppler` and `heavisine`, Haar for `blocks` and Daubechies 6 for `bumps`, as standardly done. Number of levels in wavelet decomposition is 4 for signal length

of 512, 5 levels for signal length 1024 and 6 for signal length 2048. The in all three cases the smooth level contains 32 coefficients which are left intact.

One of the key challenges in shrinkage/thresholding methods based on Bayesian model is specification of hyperparameters. It is desirable to have an automatic and objective procedure amenable to a range of input signals and noises. Our method is based on Empirical Bayes moment-matching. In principle it is possible to use more formal Empirical Bayes MLII method, but for practicable models such avenue leads to a nontrivial extremal problems.

We discuss two cases in specifying the hyperparameters. In the Case 1 the parameters are specified in a global way, i.e., coefficients in all detail levels have the same model. This case is compared to two popular global methods: VisuShrink and SureShrink (Donoho and Johnstone, 1994; Donoho, 1995; and Johnstone and Donoho, 1995).

In the Case 2 the model parameters depend on detail level, thus the models are level-dependent. The level dependent shrinkage is compared to ABWS of Chipman, Kolaczyk, and McCulloch (1997) and BAMS of Vidakovic and Ruggeri (2001). Both of these methods are implemented by Antoniadis, Bigot, and Sapatinas (2001). More detailed description of this automatic hyperparameter selection is provided next.

4.1 Tuning the Model Parameters: Case 1

This is global model, i.e., hyperparameters in models for all detail coefficients are the same.

1. μ is the reciprocal of the mean for the prior on σ^2, or, equivalently, the square root of the precision for σ^2. We first estimate σ by a robust Tukey's $\texttt{pseudos} = (Q_1 - Q_3)/C$, where Q_1 and Q_3 are the first and the third quartiles of the finest level of details in the decomposition and $1.3 \leq C \leq 1.5$. We propose $\frac{1}{\texttt{pseudos}^2}$ as a default value for μ; according to the Law of Large Numbers, this ratio should be close to the "true" μ.

2. π_0 is the weight of the point mass at zero in the prior on θ and taken to be independent of level j.

3. τ is the scale of the "spread part" in the prior (5). In the case of a double exponential prior, the variance of the signal part is $2\tau^2$. Because of the independence between the error and the signal parts, we have $\sigma_d^2 = 2(1-\pi_0)^2\tau^2 + 1/\mu$, where σ_d^2 is the variance of the observations d. This yields

$$\tau = \sqrt{\max\left\{\frac{\sigma_d^2 - \frac{1}{\mu}}{2(1-\pi_0)^2}, 0\right\}}.$$

   Note when $\tau = 0$, the prior (also the posterior) put all their mass at 0, which results in $\delta(d) = 0$.

4.2 Tuning the Model Parameters: Case 2

Models are level-dependent, i.e., some hyperparameters in models for detail coefficients are the same within a level, and different for different levels.

1. μ is specified as in the Case 1.

2. π_0 is the weight of the point mass at zero in the prior on θ and should depend on level j. Depending on our prior information about smoothness, π_0 should be close to 1 at the finest level of detail and close to 0 at the coarsest levels. We propose a hyperbolic decay in j,

$$\pi_0(j) = 1 - \frac{1}{(j - \texttt{coarsest} + 1)^\gamma}, \ \texttt{coarsest} \le j \le \log_2 n,$$

   where `coarsest` is the coarsest level subjected to shrinkage.

3. Specification of τ coincides with that in Case 1 but with π_0 replaced by $\pi_0(j)$. In this case, $\sigma_d^2 = 2(1-\pi_0(j))^2\tau^2 + 1/\mu$, and

$$\tau = \sqrt{\max\left\{\frac{\sigma_d^2 - \frac{1}{\mu}}{2(1-\pi_0(j))^2}, 0\right\}}.$$

4.3 Results

Table 1 gives the mean-squared error MSE (Variance+Bias^2) for VisuShrink, SureShrink, BaFDR ($\alpha = 0.05$), and BLFDR-fixed, as procedures with a global shrinkage model and for BAMS, ABWS, and BLFDR-ld as level dependent shrinkers on standard test signals. The test signals are rescaled so that the noise variance σ^2 equals 1. Signal-to-noise ratio is 7 and sample size is 1024.

Table 1 gives MSE (Variance+Bias^2) for VisuShrink, SureShrink, BaFDR ($\alpha = 0.05$) and BLFDR as comparable global methods. In addition to superior MSE, Bayesian hard-thresholding alternatives have much smaller bias.

To illustrate performance of BLFDR and BaFDR for standard signals and SNR's we provide three tables with simulational results. Tables 2 and 3 give global and levelwise BLFDR. For the global case $p_0 = 0.95$ while in the levelwise case parameters are determined as in the Case 2 with $\gamma = 2.5$. Table 4 gives MSE value for global shrinkage induced by BaFDR with $\alpha = 0.05$ and $\pi_0 = 0.90$.

Figure 1 shows a graphical example of the application of the above concepts. For the sake of brevity we show only Doppler signal. MATLAB programs producing MSE and figures for other test signals are available at
`http://www.isye.gatech.edu/~brani/wavelets.html`.

Figure 2 shows ordered posterior probabilities (from BaFDR) for Doppler signal. Note that, as expected, for most of the coefficients the posterior probability is close to 1. On the other hand, the selection principle is robust with respect to the choice of maximal posterior probability – the number of coefficients in the model is essentially the same for all value of the posterior probability smaller than 0.9.

4.4 An Application in AFM

To illustrate features of the BLFDR and BaFDR shrinkage approaches proposed here we used measurements in atomic force microscopy (AFM).

	blocks	bumps
VISUSHRINK	0.6840 (0.0719 + 0.6122)	1.5707 (0.1165 + 1.4543)
SURESHRINK	0.2225 (0.1369 + 0.0856)	0.6827 (0.2660 + 0.4167)
BAFDR	0.1460 (0.1137 + 0.0322)	0.5768 (0.2880 + 0.2888)
BLFDR-FIXED	0.1244 (0.1129 + 0.0115)	0.3796 (0.2584 + 0.1212)
ABWS	0.0995 (0.0874 + 0.0121)	0.3495 (0.2228 + 0.1267)
BAMS	0.1107 (0.0965 + 0.0142)	0.3404 (0.1976 + 0.1428)
BLFDR-LD	0.1184 (0.1154 + 0.0031)	0.3828 (0.2637 + 0.1191)
	doppler	heavisine
VISUSHRINK	0.4850 (0.0523 + 0.4327)	0.1204 (0.0339 + 0.0864)
SURESHRINK	0.2285 (0.0946 + 0.1340)	0.0949 (0.0416 + 0.0534)
BAFDR	0.2489 (0.1049 + 0.1440	0.1098 (0.0463 + 0.0635)
BLFDR-FIXED	0.1817 (0.1272 + 0.0545)	0.1010 (0.0689 + 0.0320)
ABWS	0.1646 (0.1006 + 0.0640)	0.0874 (0.0442 + 0.0433)
BAMS	0.1482 (0.0899 + 0.0584)	0.0815 (0.0511 + 0.0304)
BLFDR-LD	0.1801 (0.1283 + 0.0519)	0.1070 (0.0814 + 0.0256)

Table 1: MSE (Variance+Bias^2) for VisuShrink, SureShrink, BaFDR ($\alpha = 0.05$) and BLFDR (as global methods) and ABWS, BAMS, BLFDR (as level-wise methods). The standard test signals are rescaled so that the noise variance σ^2 equals 1. SNR is 7, and sample size is 1024.

The AFM is a type of scanned proximity probe microscopy (SPM) that can measure the adhesion strength between two materials at the nanonewton scale (Binnig, Quate and Gerber, 1986). In AFM, a cantilever beam is adjusted until it bonds with the surface of a sample, and 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; in order for the data to be useful this noise must be removed.

The AFM data from the adhesion measurements between carbohydrate and the cell adhesion molecule (CAM) E-Selectin was collected by Bryan Marshal from the BME Department at Georgia Institute of Technology. The technical description is provided in Marshall, McEver, and Zhu (2001).

Figure 3 depicts the original AFM signal (Panel (a)), signal smoothed by BaFDR procedure (Panel (b)), signal smoothed with global BLFDR procedure with $\pi_0 = 0.999$ fixed for all levels (Panel (c)), and signal smoothed by BLFDR with level-dependent π_0 but γ fixed at 5.

5 Conclusion

In this paper we proposed and explored two natural approaches to threshold wavelet coefficients. The approaches are based on multiple testing of hypotheses in Bayesian fashion. They are linked with the hard thresholding paradigm and also with local false discovery rate methodology proposed and explored by Efron and Tibshirani (2002) and Efron (2004). The proposed approaches are

Function	n	SNR=3	SNR=5	SNR=7	SNR=10
Blocks	512	0.2434	0.2159	0.1982	0.1810
	1024	0.1884	0.1433	0.1244	0.1044
	2048	0.1279	0.0904	0.0698	0.0570
Bumps	512	0.5022	0.5733	0.6596	0.7407
	1024	0.3356	0.3660	0.3796	0.3946
	2048	0.2235	0.2227	0.2261	0.2384
Doppler	512	0.2439	0.2524	0.2676	0.2872
	1024	0.1684	0.1692	0.1817	0.1901
	2048	0.1180	0.1044	0.1053	0.1079
Heavisine	512	0.1510	0.1593	0.1888	0.2123
	1024	0.1120	0.0943	0.1010	0.1185
	2048	0.0897	0.0688	0.0698	0.0796

Table 2: Performance of Local False Discovery Rate in Wavelet Domain. The table shows average MSE for 1000 simulations, with parameters τ and $\pi_0 = 0.95$ fixed for all levels.

desirable when dimension reduction is important and they have small bias, as typical for hard-thresholding estimators.

The methodology leading to BLFDR is quite general and could be developed for a range of Bayesian models as well. We adhere to the concept of reproducible research. The BLFDR and BaFDR are implemented in MATLAB, and m-files with examples can be found at
`http://www.isye.gatech.edu/~brani/wavelets.html`
under BLFDR and BaFDR.

Acknowledgements: Ilya Lavrik and Brani Vidakovic acknowledge support of NSF Grant 0505490 and DoD-NSA Grant E-24-60R.

Function	n	SNR=3	SNR=5	SNR=7	SNR=10
Blocks	512	0.2859	0.2483	0.2073	0.1688
	1024	0.1997	0.1472	0.1184	0.0977
	2048	0.1101	0.0910	0.0750	0.0595
Bumps	512	0.4876	0.5540	0.6174	0.6702
	1024	0.3272	0.3698	0.3828	0.3877
	2048	0.1981	0.2210	0.2381	0.2594
Doppler	512	0.2759	0.2916	0.2982	0.3049
	1024	0.1625	0.1699	0.1801	0.1912
	2048	0.0858	0.0942	0.1081	0.1178
Heavisine	512	0.1981	0.1834	0.1900	0.1966
	1024	0.1077	0.1010	0.1070	0.1202
	2048	0.0598	0.0566	0.0606	0.0700

Table 3: Performance of Local False Discovery Rate in Wavelet Domain. The table shows average MSE for 1000 simulations, with level-dependent parameters τ and π_0, $\gamma = 2.5$.

REFERENCES

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

Abramovich, F. and Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients *Computational Statistics and Data Analysis*,**22** 351–361.

Angelini, C. and Vidakovic, B. (2003). Some Novel Methods in Wavelet Data Analysis: Wavelet Anova, F-test Shrinkage, and Γ-Minimax Wavelet Shrinkage, In: *Wavelets and their Applications*, Editors Krishna, M., Radha, R., and Thangavely, S. Allied Publishers Ltd, New Delhi, ISBN 81-7764-493-9, pp. 31–45.

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

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J.R. Stat. Soc. Ser. B Stat. Methodol. 57 289–300.

Binnig, G., Quate, C.F., and Gerber, Ch. (1986). Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933.

Chipman, H., Kolaczyk, E., and McCulloch, R. (1997). Adaptive Bayesian Wavelet Shrinkage, *Journal of the American Statistical Association*,**92**, 1413–1421.

Datta, S. and Datta, S. (2005). Empirical Bayes screening of many p-values with applications to microarray studies. *Bioinformatics*, **21**, 1987–1994.

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

Donoho, D. L. (1995). De-noising by soft thresholding. *IEEE Transactions on Information Theory*,**41**, 613–627.

Dudoit, S., Shaffer, J., and Boldrick, J. (2003). Multiple hypothesis testing in microarray experiments, *Statistical Science*,bf 18, 71–103.

Function	n	SNR=3	SNR=5	SNR=7	SNR=10
Blocks	512	0.4114	0.3012	0.2527	0.2130
	1024	0.2687	0.2098	0.1460	0.1099
	2048	0.1322	0.1025	0.0743	0.0573
Bumps	512	1.0058	0.9367	0.9541	1.0605
	1024	0.4902	0.5479	0.5768	0.5408
	2048	0.2785	0.3130	0.3117	0.3028
Doppler	512	0.3875	0.3657	0.3940	0.4091
	1024	0.1842	0.2085	0.2489	0.2938
	2048	0.0814	0.1031	0.1260	0.1464
Heavisine	512	0.1078	0.1440	0.1966	0.3213
	1024	0.0617	0.0873	0.1098	0.1556
	2048	0.0391	0.0609	0.0720	0.0989

Table 4: Performance of the BaFDR. The average MSE for 1000 simulations with $\alpha = 0.05$ and $\pi_0 = 0.90$ `coarsest`=5 for all.

Efron, B. (2004). Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis *Journal of the American Statistical Association*, **99**, 96–104.

Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiology*, **23**, 70–86.

Efron, B., Tibshirani, R., Storey, J. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. J. Amer. Statist. Assoc. 96 1151–1160.

Folks, J. L. (1984). Combination of Independent Tests, In *Handbook of Statistics, Vol 4.*, P. R. Krishniah and P. K. Sen Editors, Elsevier Scince Publishers, 113–121.

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

Marshall, B., McEver, R. and Zhu, C. (2001). Kinetic rates and their force dependence on the P-Selectin/PSGL-1 interaction measured by atomic force microscopy. Proceedings of ASME 2001, Bioengineering Conference, BED - Vol. 50.

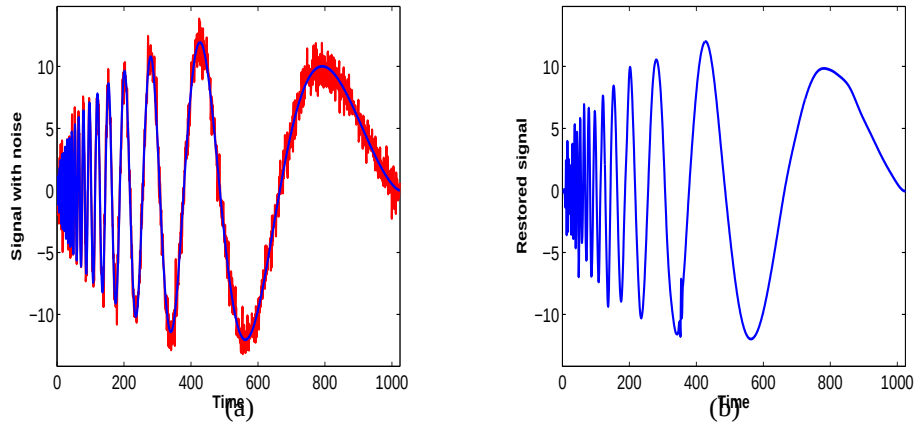
Miller, R. (1981). *Simultaneous Statistical Inference*, Second Edition. Springer-Verlag, NY.

Ogden, T. and Parzen, E. (1996). Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Computational Statistics and Data Analysis*, **22**, 53–70.

Rosner, G. and Vidakovic, B. (2000). Wavelet Functional ANOVA, Bayesian False Discovery Rate, and Longitudinal Measurements of Oxygen Pressure in Rats, Technical Report 1/2000, ISyE, Georgia Institute of Technology.

Tadesse, M., Ibrahim, J., Vannucci, M., and Gentleman, R. (2005). Wavelet Thresholding with Bayesian False Discovery Rate Control, *Biometrics*,**61**, 25–35.

Vidakovic, B. and Ruggeri, F. (2001). BAMS Method: Theory and Simulations. *Sankhyā, Series B*, **63**,*2* (Special Issue on Wavelets), 234–249.



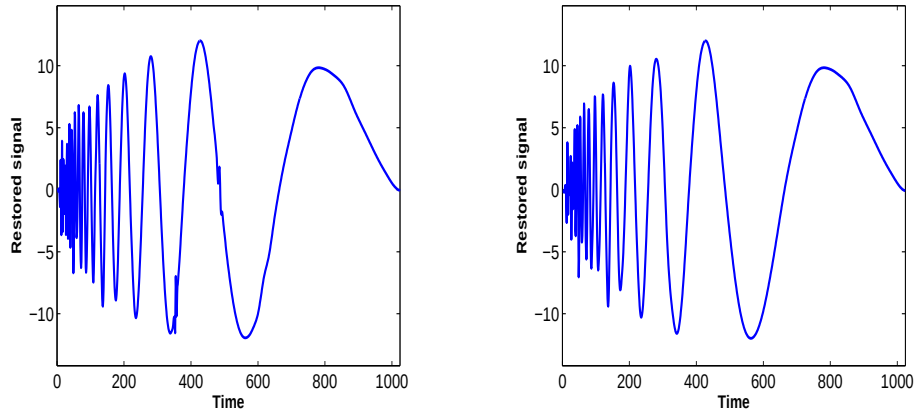


Figure 1: (a) Doppler signal with noise (SNR=7); (b) BLFDR with $p_0 = 0.95$; (c) BLFDR with levelwise p_0; and (d) BaFDR with $\alpha = 0.05$.

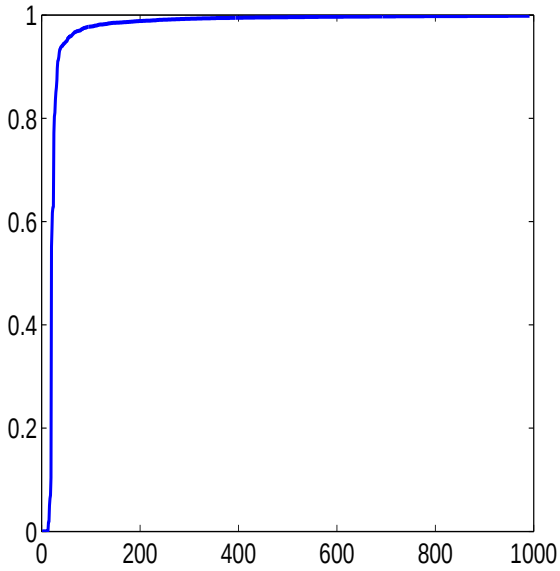


Figure 2: Ordered posterior probabilities (from BaFDR) for Doppler signal.

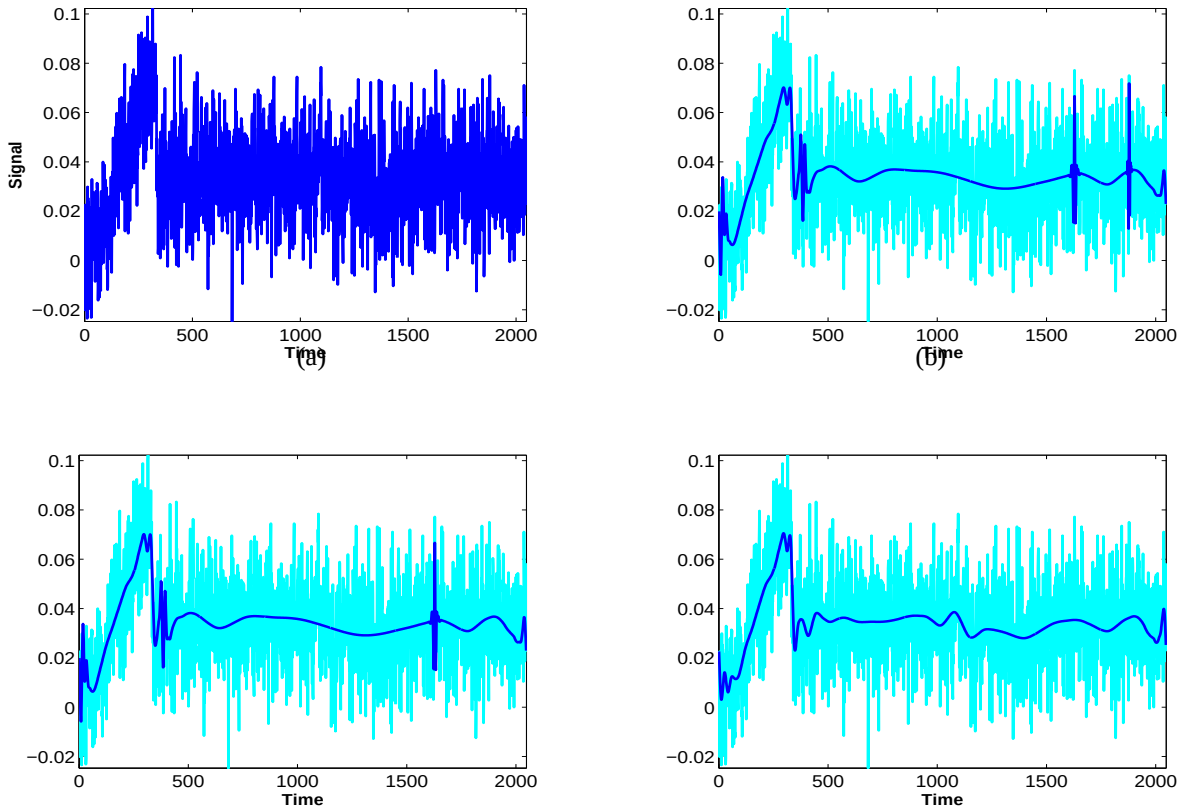


Figure 3: (a) Original AFM signal; (b) Smoothing with BaFDR; (c) Smoothing with BLFDR with $\pi_0 = 0.999$ fixed for all levels; and (d) Smoothing with BLFDR with level-dependent π_0 but fixed $\gamma = 5$.