Simulation studies to compare bayesian wavelet shrinkage methods in aggregated functional data

Alex Rodrigo dos Santos Sousa[1,a]

[a]Department of Statistics, State University of Campinas (UNICAMP), Brazil

Abstract

The present work describes simulation studies to compare the performances in terms of averaged mean squared error of bayesian wavelet shrinkage methods in estimating component curves from aggregated functional data. Five bayesian methods available in the literature were considered to be compared in the studies: The shrinkage rule under logistic prior, shrinkage rule under beta prior, large posterior mode (LPM) method, amplitude-scale invariant Bayes estimator (ABE) and Bayesian adaptive multiresolution smoother (BAMS). The so called Donoho-Johnstone test functions, logit and SpaHet functions were considered as component functions and the scenarios were defined according to different values of sample size and signal to noise ratio in the datasets. It was observed that the signal to noise ratio of the data had impact on the performances of the methods. An application of the methodology and the results to the tecator dataset is also done.

Keywords: wavelet shrinkage, aggregated functional data, logistic prior

1. Introduction

The statistical problem of estimating component curves from aggregated functional data has been widely studied in recent years in several areas of science. In Chemometrics, for example, there is an interest in estimating absorbance curves of substance constituents from samples of the absorbance curves of the substance itself. In this case, the absorbance curve of the substance is formed by a linear combination of the absorbance curves of its constituents according to the Beer-Lambert law (Brereton, 2003). Another example appears in the study of electricity consumption in a given region, in which the energy consumption curve over a period of time is composed of the aggregations of the individual consumption curves of households and establishments (Dias *et al.*, 2013).

The first proposed methods to estimate component curves from aggregated data approach this problem under a multivariate point of view since, in practice, such curves are observed at a finite number of locations due instrumentation limitations. In this sense, it is possible to interpret the observations as a random vector with a certain correlation structure at near locations. Principal components regression (PCR) by Cowe and McNicol (1985) and partial least squares regression (PLS) by Wold *et al.* (1983) based methods have been successfully proposed in applications. Bayesian multivariate approaches were also proposed by Brown *et al.* (1998a,b) and Brown *et al.* (2001).

Functional data analysis based methods to the aggregated functional data problem were proposed later by Dias *et al.* (2009) and Dias *et al.* (2013). These proposed methods take into account the functional structure of the observations. In this way, each component curve can be represented in

---

[1] Corresponding author: Department of Statistics, State University of Campinas (UNICAMP), Sérgio Buarque de Holanda Street, 651, 13083-859, Campinas, Brazil. E-mail: asousa@unicamp.br

terms of some convenient functional basis, such as splines basis or their variants for example, so that the problem of estimating the curve becomes a finite-dimensional problem of estimating the coefficients of the basis expansion. In this functional approach, one possibility is also to expand the functions in wavelet basis as proposed recently by Sousa (2022). The advantages of the wavelet approach are the well localization property of the wavelet coefficients and the possibility of estimating important characteristics of a function, such as peaks, oscillations, discontinuities, among others, through a few number of wavelet coefficients. Further, the sparsity property of the wavelet coefficients allows the identification of these function features by the magnitude of the nonzero coefficients at their localizations. See Vidakovic (1999) for details and properties of wavelets and its application on statistical modelling.

Sousa (2022) proposes the expansion of the component curves by wavelet basis and the use of a bayesian approach to estimate the wavelet coefficients by considering, to each wavelet coefficient, a prior distribution composed by a point mass function at zero and the symmetric around zero logistic distribution. Under this prior model and the quadratic loss function assumption, the bayesian rule to estimate the wavelet coefficients is the posterior expected value of the wavelet coefficient. Moreover, the associated bayesian rule acts by shrinking empirical wavelet coefficients and reducing the present noise effects on the coefficients. This kind of estimator is also called shrinkage rules, see Donoho and Johnstone (1994a,b and 1995) and Donoho (1993a,b and 1995a,b) for details about wavelet shrinkage procedures. Further, in Sousa (2022), a simulation study was conducted to compare the proposed method against the expansion of the component curves by B-splines basis. It was concluded that the proposed method performed better in terms of mean squared error than the B-splines expansion when the component curves have local characteristics such as discontinuities, peaks and oscillations.

Although Sousa (2022) indicates the feasibility of wavelet basis expansion application and the specific use of the bayesian shrinkage rule under logistic prior to estimate wavelet coefficients of component curves from aggregated curves, no information about performances of different bayesian shrinkage methods in aggregated functional data is available in the literature. In this sense, the present work intends to compare bayesian wavelet shrinkage methods in the problem of estimating component curves from aggregated curves. The performance comparison is evaluated in terms of averaged mean square error measure in simulation studies that cover different scenarios of sample sizes and signal to noise ratios values and different number of component functions. Five standard methods were considered in the simulatons, the rule based on the logistic prior of Sousa (2020, 2022), the rule based on the symmetric around zero beta prior proposed by Sousa *et al*. (2020), the large posterior mode (LPM) method by Cutillo *et al*. (2008), amplitude-scale invariant Bayes estimator (ABE) by Figueiredo and Nowak (2001) and Bayesian adaptive multiresolution smoother (BAMS) by Vidakovic and Ruggeri (2001). Therefore, the main motivation of this work is to provide performances in terms of averaged mean square error, of the considered bayesian shrinkage methods in different sample sizes, signal-to-noise ratios and numbers of component function for practitioners who intend to apply one of the methods in real aggregated functional data analysis.

This paper is organized as follows: The statistical model for de aggregated functional data and the estimation procedure of the component curves are defined in Section 2. The descriptions of the considered bayesian methods in the simulation studies are in Section 3. Section 4 is dedicated to the simulation studies results and analysis. Application of the methodology and results to the so called tecator dataset is done in Section 5. The paper finishes with further considerations and discussions in Section 6.

2. Statistical model and estimation procedure

Consider a univariate function $A(t) \in \mathbb{L}_2(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} | \int f^2 < \infty\}$ that can be written as

$$A(t) = \sum_{l=1}^{L} y_l \alpha_l(t) + e(t), \tag{2.1}$$

where $\alpha_l(t) \in \mathbb{L}_2(\mathbb{R})$ are unknown component functions, y_l are known real valued weights, $l = 1, \ldots, L$, and $\{e(t), t \in \mathbb{R}\}$ is a zero mean gaussian process with unknown variance σ^2, $\sigma > 0$. The estimation of the functions $\alpha_l(t)$ is considered in this work. In fact, this estimation process is usually done by multivariate methods (Brown *et al.*, 1998a,b) or functional data analysis (FDA) approach (Dias *et al.*, 2009, 2013). In this last point of view, each function α_l of (2.1) is represented in terms of some functional basis such as splines, B-splines or fourier basis, for example. Here, each component function are expanded by wavelet basis,

$$\alpha_l(t) = \sum_{j,k \in \mathbb{Z}} \gamma_{jk}^{(l)} \psi_{jk}(t), \quad l = 1, \ldots, L, \tag{2.2}$$

where $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), j, k \in \mathbb{Z}\}$ is an orthonormal wavelet basis for $\mathbb{L}_2(\mathbb{R})$ constructed by dilations j and translations k of a function ψ called wavelet or mother wavelet and $\gamma_{jk}^{(l)}$'s are unknown wavelet coefficients. Note that we consider the same wavelet family for all component functions expansion. Thus, the problem of estimating the function α_l becomes a problem of estimating the finite number of wavelet coefficients $\gamma_{jk}^{(l)}$'s of the representation (2.2). Further, the magnitudes of the wavelet coefficients allow to recover local features of the component functions, such as discontinuities points, spikes and oscillations, due the well localization in time and frequency domain of wavelets. This property does not occur in spline-based and fourier basis representations.

In practice, it is observed I samples of the aggregated curve $A(t)$ at $M = 2^J$ equally spaced locations $t_1, \ldots, t_M$, i.e, the considered dataset is $\{(t_m, A_i(t_m)), m = 1, \ldots, M$ and $i = 1, \ldots, I\}$. Thus, the discrete version of (2.1) is

$$A_i(t_m) = \sum_{l=1}^{L} y_{il}\alpha_l(t_m) + e_i(t_m), \quad i = 1, \ldots, I, \ m = 1, \ldots, M = 2^J, \tag{2.3}$$

where $e_i(t_m)$ are independent and identically normal distributed random noises with zero mean and variance σ^2, $\forall i, m$. Further, the I samples are obtained at the same locations $t_1, \ldots, t_M$ but the weights of the linear combinations are allowed to be different from one sample to another. We can rewrite (2.3) in matrix notation as

$$A = \alpha y + e, \tag{2.4}$$

where $A = (A_{mi} = A_i(t_m))_{1 \le m \le M, 1 \le i \le I}$, $\alpha = (\alpha_{ml} = \alpha_l(t_m))_{1 \le m \le M, 1 \le l \le L}$, $y = (y_{li})_{1 \le l \le L, 1 \le i \le I}$ and $e = (e_{mi} = e_i(t_m))_{1 \le m \le M, 1 \le i \le I}$.

The wavelet shrinkage procedure is made in the wavelet domain to estimate the coefficients γ's of (2.2). For this reason, a discrete wavelet transform (DWT) is applied on the original aggregated data, which can be represented by a $M \times M$ wavelet transformation matrix W and applied on both sides of (2.4), i.e,

$$WA = W(\alpha y + e),$$
$$WA = W\alpha y + We,$$
$$D = \Gamma y + \varepsilon, \tag{2.5}$$

where $\boldsymbol{D} = \boldsymbol{WA} = (d_{mi})_{1\leq m\leq M, 1\leq i\leq I}$ is the matrix with the empirical wavelet coefficients of the aggregated curves, $\boldsymbol{\Gamma} = \boldsymbol{W\alpha} = (\gamma_{ml})_{1\leq m\leq M, 1\leq l\leq L}$ is the matrix with the unknown wavelet coefficients of the component curves and $\boldsymbol{\varepsilon} = \boldsymbol{We} = (\varepsilon_{mi})_{1\leq m\leq M, 1\leq i\leq I}$ is the matrix with the random errors on the wavelet domain, which remain independent and zero mean normal distributed with variance σ^2 due the orthogonality property of wavelet transforms. In this sense, a particular empirical wavelet coefficient d_{mi} of $\boldsymbol{D}$ can be written as

$$d_{mi} = \sum_{l=1}^{L} y_{li}\gamma_{ml} + \varepsilon_{mi} = \theta_{mi} + \varepsilon_{mi}, \tag{2.6}$$

where $\theta_{mi} = \sum_{l=1}^{L} y_{li}\gamma_{ml}$ and ε_{mi} is zero mean normal with variance σ^2, i.e, a single empirical wavelet coefficient of the aggregated curve is also a linear combination of the unknown wavelet coefficients of the component curves plus a random error. Moreover, the weights of this linear combination at wavelet domain remain the same as the original linear combination of the curves at time domain.

The estimation of the wavelet coefficients matrix $\boldsymbol{\Gamma}$ in (2.5) is done by applying a wavelet shrinkage rule δ on each single empirical wavelet coefficient d, obtaining the matrix $\boldsymbol{\delta(D)}$ such that

$$\boldsymbol{\delta(D)} = (\delta(d_{mi}))_{1\leq m\leq M, 1\leq i\leq I}. \tag{2.7}$$

We can see the matrix $\boldsymbol{\delta(D)}$ as a denoising version of $\boldsymbol{D}$, i.e, the shrinkage rule $\delta(d)$ acts by denoising the empirical coefficient d in order to estimate θ in (2.6), $\delta(d) = \hat{\theta}$. Thus, the estimation $\hat{\boldsymbol{\Gamma}}$ of the wavelet coefficients matrix $\boldsymbol{\Gamma}$ is given by least squares method,

$$\hat{\boldsymbol{\Gamma}} = \boldsymbol{\delta(D)}\boldsymbol{y}^t\left(\boldsymbol{yy}^t\right)^{-1}, \tag{2.8}$$

and finally $\boldsymbol{\alpha}$ can be estimated at locations $t_1, \ldots, t_M$ by the inverse discrete wavelet transformation (IDWT),

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{W}^t\hat{\boldsymbol{\Gamma}}. \tag{2.9}$$

For more details about wavelet shrinkage in aggregated curves, see Sousa (2022).

3. Bayesian wavelet shrinkage methods

There are several available wavelet shrinkage methods in the literature. Most of them are thresholding, i.e, the estimator shrinks sufficiently small empirical wavelet coefficients (less than a threshold value) to exactly zero. Two extremely applied thresholding rules are the so called hard and soft rules proposed by Donoho and Johnstone (1994b). However, it is considered in this work only bayesian wavelet shrinkage methods that have been proposed in recent years. In general, bayesian methods propose a prior distribution to the wavelet coefficients and estimate them according to a loss function, as the squared loss function for example. In wavelet domain, the priors are usually symmetric around zero and peaked at zero due the sparsity property of the wavelet coefficients. The main advantage of bayesian methods is the ability to incorporate prior information about the wavelet coefficients, such as sparsity, boundedness, self similarity and others by convenient choices of prior distributions and their hyperparameters values.

In the next subsections, the bayesian shrinkage rules considered in this work are described.

3.1. Logistic shrinkage rule

The shrinkage rule under logistic prior was proposed by Sousa (2020). It assumes a mixture of a point mass function at zero and a symmetric around zero logistic distribution as prior distribution to a single linear combination of wavelet coefficients θ,

$$\pi\left(\theta; p, \tau\right) = p\delta_0\left(\theta\right) + \left(1 - p\right) g\left(\theta; \tau\right), \tag{3.1}$$

where $p \in (0, 1)$, $\delta_0(\theta)$ is the point mass function at zero and $g(\theta; \tau)$ is the logistic density function symmetric around zero, for $\tau > 0$,

$$g\left(\theta; \tau\right) = \frac{\exp\left\{-\left(\theta/\tau\right)\right\}}{\tau\left(1 + \exp\left\{-\left(\theta/\tau\right)\right\}\right)^2} \mathbb{I}_{\mathbb{R}}\left(\theta\right), \tag{3.2}$$

where $\mathbb{I}_{\mathbb{R}}(\cdot)$ is an indicator function equal to 1 for real valued argument and 0 else. Under squared loss function assumption, the associated bayesian shrinkage rule is the posterior expected value of θ, $\mathbb{E}_\pi(\theta|d)$, that under the model prior (3.1) is given by (Sousa, 2020),

$$\delta\left(d\right) = \mathbb{E}_\pi\left(\theta \mid d\right) = \frac{\left(1 - p\right) \int_{\mathbb{R}} \left(\sigma u + d\right) g\left(\sigma u + d; \tau\right) \phi\left(u\right) du}{\left(p/\sigma\right) \phi\left(d/\sigma\right) + \left(1 - p\right) \int_{\mathbb{R}} g\left(\sigma u + d; \tau\right) \phi\left(u\right) du}, \tag{3.3}$$

where $\phi(\cdot)$ is the standard normal density function. The shrinkage rule (3.3) under the model (3.2) is called logistic shrinkage rule and has interesting features under estimation point of view. First, its hyperparameters p and τ control the degree of shrinkage of the rule. Higher values of τ or p imply higher shrinkage level, i.e, the rule will reduce severely the magnitudes of the empirical coefficients. Further, as described in Sousa (2020), the logistic shrinkage rule had good performances in terms of averaged mean squared error in simulation studies against standard shrinkage or thresholding procedures.

3.2. Beta shrinkage rule

Sousa *et al*. (2020) proposed the use of a mixture of a point mass function at zero and the beta distribution with symmetric support around zero as a prior distribution to the wavelet coefficients,

$$\pi\left(\theta; p, a, m\right) = p\delta_0\left(\theta\right) + \left(1 - p\right) g\left(\theta; a, m\right), \tag{3.4}$$

and

$$g\left(\theta; a, m\right) = \frac{\left(m^2 - \theta^2\right)^{(a-1)}}{(2m)^{(2a-1)} B\left(a, a\right)} \mathbb{I}_{[-m,m]}\left(\theta\right), \tag{3.5}$$

where $B(\cdot, \cdot)$ is the standard beta function, $a > 0$ and $m > 0$ are the parameters of the distribution, and $\mathbb{I}_{[-m,m]}(\cdot)$ is an indicator function equal to 1 for its argument in the interval $[-m, m]$ and 0 else. The associated shrinkage rule under the prior model (3.4) and (3.5) is called beta shrinkage rule and it is convenient to deal with bounded wavelet coefficients on $[-m, m]$ and its expression is (Sousa *et al*., 2020)

$$\delta\left(d\right) = \mathbb{E}_\pi\left(\theta \mid d\right) = \frac{\left(1 - p\right) \int_{-m}^{m} \left(\sigma u + d\right) g\left(\sigma u + d; a, m\right) \phi\left(u\right) du}{\left(p/\sigma\right) \phi\left(d/\sigma\right) + \left(1 - p\right) \int_{-m}^{m} g\left(\sigma u + d; a, m\right) \phi\left(u\right) du}. \tag{3.6}$$
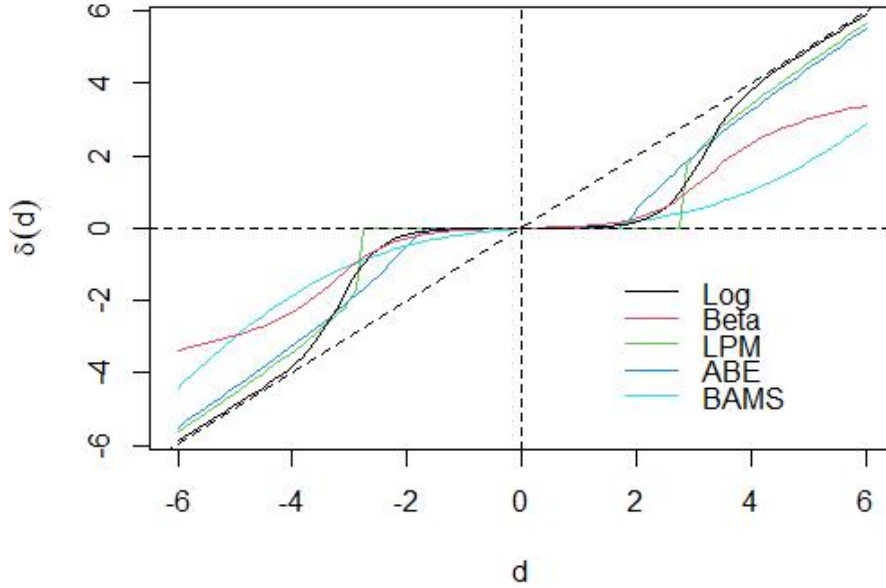
Figure 1: *Logistic (log), beta, LPM, ABE and BAMS bayesian rules. Logistic, beta shrinkage and BAMS rules are shrinkage rules while LPM and ABE are thresholding rules.*

For $a > 1$, the density function (3.5) is unimodal around zero and as a increases, the density becomes more concentrated around zero. This is an important feature for wavelet shrinkage methods, since higher values of a imply higher levels of shrinkage in the empirical coefficients, which results in sparse estimated coefficients. Density (3.5) becomes uniform for $a = 1$, which was already considered by Angelini and Vidakovic (2004) as prior to the wavelet coefficients.

Based on the fact that much of the noise information present in the data can be obtained on the finer resolution scale, for the robust σ estimation, Donoho and Johnstone (1994a) suggest

$$\hat{\sigma} = \frac{\text{median}\left\{\left|d_{J-1,k}\right| : k = 0, \ldots, 2^{J-1}\right\}}{0.6745}. \tag{3.7}$$

Angelini and Vidakovic (2004) suggest the hyperparameters p and m be dependent on the level of resolution j according to the expressions

$$p = p(j) = 1 - \frac{1}{(j - J_0 + 1)^{\gamma}}, \tag{3.8}$$

and

$$m = m(j) = \max_{k}\left\{\left|d_{jk}\right|\right\}, \tag{3.9}$$

where $J_0 \leq j \leq J - 1$, J_0 is the primary resolution level and $\gamma > 0$. They also suggest that in the absence of additional information, $\gamma = 2$ can be adopted.

Table 1: Test functions definitions for $x \in [0, 1]$ used as component functions in the simulation studies

BUMPS
$f(x) = \sum_{l=1}^{11} h_l K((x - x_l)/w_l),$
where
$K(x) = (1 + |x|)^{-4}$
$(x_l)_{l=1}^{11} = (0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81)$
$(h_l)_{l=1}^{11} = (4, 5, 3, 4, 5, 4.2, 2.1, 4.3, 3.1, 5.1, 4.2)$
$(w_l)_{l=1}^{11} = (0.005, 0.005, 0.006, 0.01, 0.01, 0.03, 0.01, 0.01, 0.005, 0.008, 0.005)$
BLOCKS
$f(x) = \sum_{l=1}^{11} h_l K(x - x_l),$
where
$K(x) = (1 + \text{sgn}(x))/2$
$(x_l)_{l=1}^{11} = (0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81)$
$(h_l)_{l=1}^{11} = (4, -5, 3, -4, 5, -4.2, 2.1, 4.3, -3.1, 2.1, -4.2)$
DOPPLER
$f(x) = \sqrt{x(1 - x)} \sin(2.1\pi/(x + 0.05))$
HEAVISINE
$f(x) = 4 \sin(4\pi x) - \text{sgn}(x - 0.3) - \text{sgn}(0.72 - x)$
LOGIT
$f(x) = 1/(1 + \exp\{-20(x - 0.5)\})$
SpaHet
$f(x) = \sqrt{x(1 - x)} \sin(2\pi(1 + 2^{-0.6})/(x + 2^{-0.6}))$

3.3. Large posterior mode (LPM)

Cutillo *et al.* (2008) proposed a bayesian thresholding rule that is based on the maximum a posteriori (MAP) principle. Under the model (2.6) with gaussian noise, it is assumed a normal prior for θ, i.e, $\theta|t^2 \sim N(0, t^2)$, where $t^2 \sim (t^2)^{-k}$, $k > 1/2$. The LPM thresholding rule picks the larger mode in absolute value of the posterior distribution of $\theta|d$. Further, there is an interesting feature of the posterior that allows the Bayes rule to be thresholding. The posterior can be unimodal at zero or bimodal trivially at zero and at another local mode. For the unimodal case, the empirical coefficient is then shrunk to zero. For the second one, it is shrunk slightly to the local mode. The closed form of the rule is

$$\delta_{\text{LMP}}(d) = \frac{d + \text{sgn}(d)\sqrt{d^2 - 4\sigma^2(2k - 1)}}{2} \mathbb{I}_{[\lambda_{\text{LPM}}, +\infty)}(|d|), \tag{3.10}$$

where $\lambda_{\text{LPM}} = 2\sigma\sqrt{2k - 1}$.

3.4. Amplitude-Scale invariant Bayes estimator (ABE)

Figueiredo and Nowak (2001) proposed a bayesian thresholding rule that does not depend on prior hyperparameters, which must be elicited in other bayesian shrinkage/thresholding methods. It assumes an amplitude-scale invariant (noninformative) prior $\pi(\theta) \propto |\theta|^{-1}$ for θ from the model (2.6). The Bayes rule is thresholding, given by

$$\delta_{\text{ABE}}(d) = \frac{\left(d^2 - 3\sigma^2\right)_+}{d}, \tag{3.11}$$

where $(x)_+ = \max\{0, x\}$. Note that the rule depends only on the noise variance parameter σ^2.

(a) Bumps

(b) Blocks

(c) Doppler
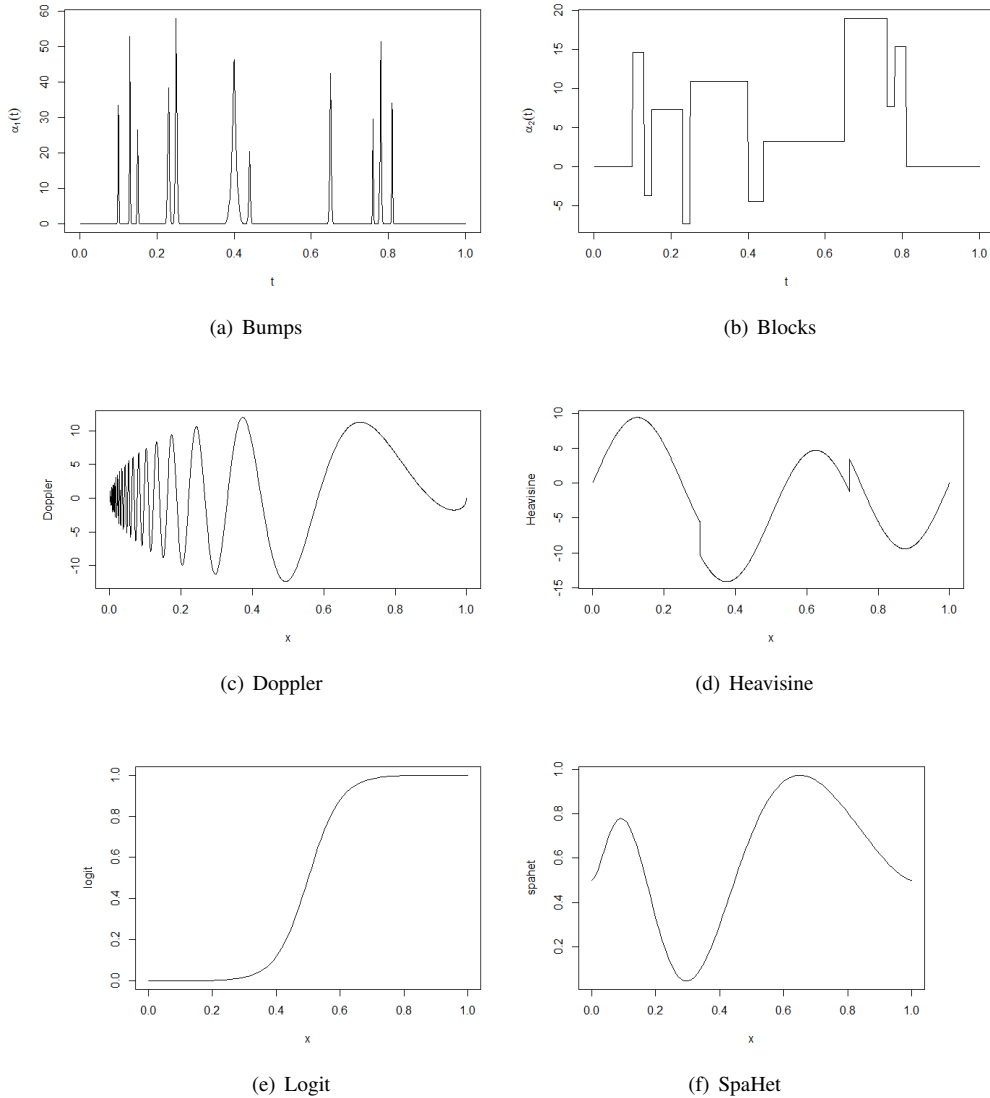
(d) Heavisine

(e) Logit

(f) SpaHet

Figure 2: *Functions used as component functions in the simulation studies.*

3.5. Bayesian adaptive multiresolution smoother (BAMS)

BAMS method was proposed by Vidakovic and Ruggeri (2001) and is obtained by considering $\sigma^2 \sim$ Exp$(1/\mu)$, $\mu > 0$, and the following prior distribution π to a single wavelet coefficient,

$$\pi(\theta) = \alpha\delta_0(\theta) + (1 - \alpha)\,\mathrm{DE}(0, \tau), \qquad (3.12)$$

where $\alpha \in (0, 1)$, $\delta_0(\cdot)$ is a point mass function at zero and $\mathrm{DE}(0, \tau)$ is the double exponential density with location parameter equals to zero and scale parameter τ, $\tau > 0$, given by $g(\theta; 0, \tau) =$

$(1/2\tau) \exp\{-(|\theta|/\tau)\}$, $\theta \in \mathbb{R}$. The associated shrinkage rule under squared error loss, $\delta_{\text{BAMS}}(d) = \mathbb{E}(\theta|d)$ is

$$\delta_{\text{BAMS}}(d) = \frac{(1-\alpha)\, m(d)\, \delta(d)}{(1-\alpha)\, m(d) + \alpha \text{DE}\left(0, \left(1/\sqrt{2\mu}\right)\right)}, \tag{3.13}$$

where $m(\cdot)$ and $\delta(\cdot)$ are the predictive distribution of d and the shrinkage rule respectively under assumption that $\theta \sim \text{DE}(0, \tau)$ and given by

$$m(d) = \frac{\tau \exp\{-(|d|/\tau)\} - \left(1/\sqrt{2\mu}\right) \exp\left\{-\sqrt{2\mu}\,|d|\right\}}{2\tau^2 - (1/\mu)},$$

and

$$\delta(d) = \frac{\tau\left(\tau^2 - (1/2\mu)\right) d \exp\{-|d|/\tau\} + \left(\tau^2/\mu\right)\left(\exp\left\{-|d|/\sqrt{2\mu}\right\} - \exp\{-|d|/\tau\}\right)}{(\tau^2 - (1/2\mu))\left(\tau \exp\{-|d|/\tau\} - \left(1/\sqrt{2\mu}\right)\exp\left\{-|d|/\sqrt{2\mu}\right\}\right)}.$$

Figure 1 shows the shrinkage rules under logistic and beta priors, BAMS, LPM and ABE rules on the interval $[-6, 6]$. It is clear the shrinkage behaviour of the logistic and beta shrinkage rules and BAMS, i.e, although they shrink the empirical coefficient magnitude, the shrunk coefficient of a sufficiently small empirical coefficient is not necessarily zero. On the other hand, LPM and ABE rules are thresholding, once they shrink a sufficiently small coefficient to exactly zero. Further, the five shrinkage rules are symmetric around zero as expected.

4. Simulation studies

Simulation studies were conducted to evaluate the performances of the considered bayesian rules in estimating component curves. Six functions on $[0, 1]$ were selected as component curves: The so called Donoho-Johnstone test functions bumps, blocks, Doppler and Heavisine and the functions logit and SpaHet, that were used in simulation studies of Wand (2000), Ruppert *et al.* (2003) and Goepp *et al.* (2018). The mathematical expressions of the functions are available in Table 1 and their plots are in Figure 2. The four Donoho-Johnstone test functions have interesting local features to be recovered, such as peaks (bumps), oscillations (Doppler) and discontinuities (Blocks and Heavisine). On the other hand, logit and SpaHet are smooth functions in the whole domain $[0, 1]$. Thus, linear combinations of these functions can provide interesting datasets with several features to be estimated and that usually occur in real problems. Further, the use of wavelets representation is very suitable to estimate the component functions once wavelets are well localized in time domain and the coefficients of the representation are usually significant in positions where the component functions have important features, such discontinuities and spikes for example.

Three simulation studies were considered according to different number of component functions. In these simulations, data points were generated by aggregating the component functions according to the model (2.1). After the application of a discrete wavelet transformation on the generated data, the shrinkage rules were applied on the empirical wavelet coefficients to reduce noise and estimate the wavelet coefficients of the underlying component functions. Later, these component functions were estimated by (2.9).
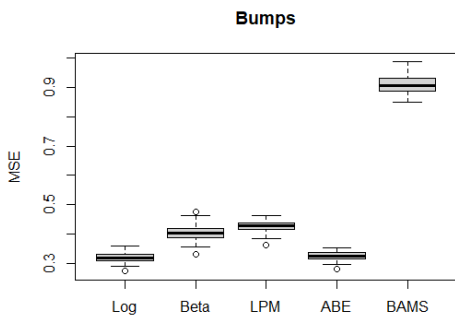
Simulation 1 was conducted by considering two component functions ($L = 2$), bumps and blocks. The Simulation 2 considered four component functions ($L = 4$), bumps, blocks, Doppler and logit.
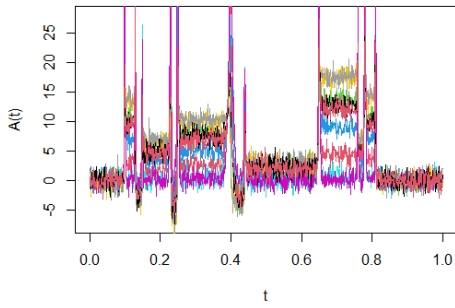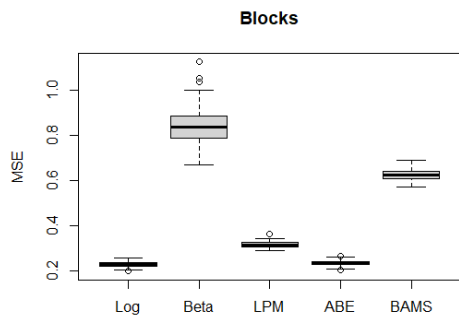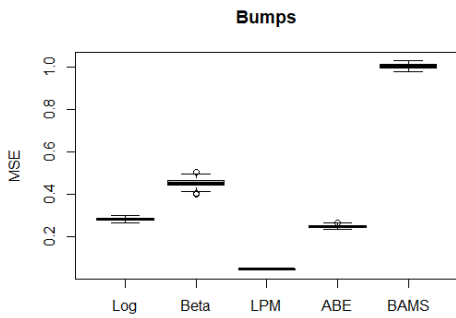
(a) Ten generated samples.
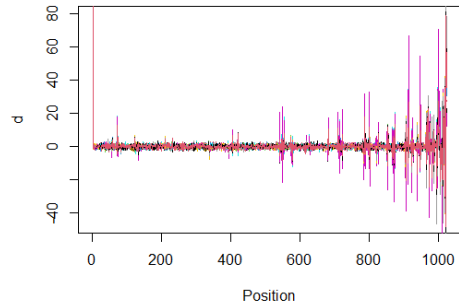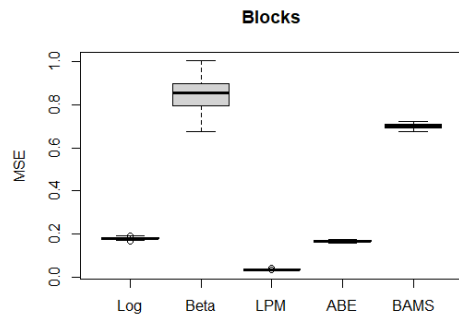

(b) Empirical wavelet coefficients.

(c) Boxplot - bumps (SNR = 3).

(d) Boxplot - blocks (SNR = 3).

(e) Boxplot - bumps (SNR = 9).

(f) Boxplot - blocks (SNR = 9).

Figure 3: *Ten samples of Simulation 1 with different weights (each color represents a sample) (a) and their respective empirical wavelet coefficients (b). Boxplots of rules MSE for bumps (c) and blocks (d) component functions for M = 1024 and* SNR = 3 *and bumps (e) and blocks (f) for* SNR = 9.

Table 2: AMSE (standard deviation) of simulation study 1 for aggregated data generated with the component functions bumps and blocks ($L = 2$ component functions)

		Simulation study 1 - $L = 2$			
		$M = 512$		$M = 1024$	
Function	**Method**	**SNR = 3**	**SNR = 9**	**SNR = 3**	**SNR = 9**
	LOG	**0.2932 (0.0187)**	0.3310 (0.0103)	**0.3214 (0.0158)**	0.2827 (0.0077)
	BETA	0.3761 (0.0271)	0.4671 (0.0186)	0.4044 (0.0259)	0.4556 (0.0211)
Bumps	LPM	0.3083 (0.0170)	**0.0342 (0.0019)**	0.4266 (0.0172)	**0.0471 (0.0020)**
	ABE	0.3091 (0.0195)	0.3016 (0.0088)	0.3267 (0.0152)	0.2493 (0.0059)
	BAMS	11.2350 (0.0411)	12.7690 (0.0160)	0.9092 (0.0300)	10.061 (0.0112)
	LOG	**0.2711 (0.0169)**	0.2643 (0.0087)	**0.2275 (0.0114)**	0.1796 (0.0046)
	BETA	0.8775 (0.0699)	0.9446 (0.0664)	0.8391 (0.0797)	0.8505 (0.0712)
Blocks	LPM	0.3045 (0.0188)	**0.0340 (0.0022)**	0.3151 (0.0130)	**0.0349 (0.0016)**
	ABE	0.2853 (0.0175)	0.2525 (0.0083)	0.2331 (0.0111)	0.1669 (0.0039)
	BAMS	0.9784 (0.0362)	11.2562 (0.0155)	0.6231 (0.0252)	0.7002 (0.0095)

Finally, Simulation 3 considered the six component functions ($L = 6$). For each simulation study, $I = 50$ aggregated curves were generated from model (2.3) for $M = 512 = 2^9$ and $1024 = 2^{10}$ equally spaced points on $[0, 1]$. The zero mean gaussian noises were generated with variances according to signal to noise ratio (SNR) values SNR = 3 and 9. Smaller SNR value implies higher amount of noise in the data. For each scenario of M and SNR values, $N = 100$ replicates were done and the component functions points were estimated according to (2.9) for each considered bayesian rule. The mean squared error (MSE) of the j-the replication of a component function $\alpha_l(t)$,

$$\text{MSE}_l^{(j)} = \frac{1}{M} \sum_{i=1}^{M} \left[\hat{\alpha}_l^{(j)}(t_i) - \alpha_l(t_i) \right]^2$$

was calculated for each replicate and the averaged mean squared error (AMSE),

$$\text{AMSE}_l = \frac{1}{N} \sum_{j=1}^{N} \text{MSE}_l^{(j)}$$

was considered as performance measure, for each component function $\alpha_l(t)$, $l = 1, \ldots, L$. According to the AMSE measure criteria, lower AMSE implies in a better performance. In all generated datasets, the DWT under Daubechies basis with ten null moments (Vidakovic, 1999) was applied to obtain the empirical coefficients vector. The choice of the wavelet basis defines the specific wavelet function $\psi(t)$ in the representation (2.2) and the matrix W of the DWT in (2.5).

Moreover, for the three simulation studies, the chosen weights in model (2.1) have sum equals to one, i.e, $\sum_l y_l = 1$ and the chosen hyperparameters values were $\tau = 5$ of the logistic shrinkage rule in (3.3), $a = 5$ of the beta shrinkage rule in (3.6), $k = 1$ of LPM rule in (3.10), $\tau = 2$ and $\mu = 0.5$ for BAMS rule in (3.14). The estimate of σ was according to (3.7).

4.1. Simulation study 1

In this simulation study, the 50 samples were generated by aggregation of bumps and blocks functions ($L = 2$ component functions) with different weights for each function in the linear combination of model (2.3). Figures 3(a) and (b) show ten generated samples and their empirical wavelet coefficients respectively for $M = 1024$ and SNR = 3. It is possible to observe from Figure 3(a) the local features of both component functions in the samples, i.e., the spikes of bumps, discontinuities and piecewise
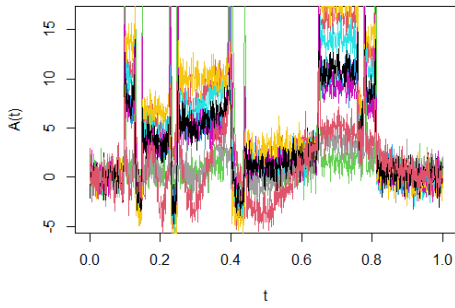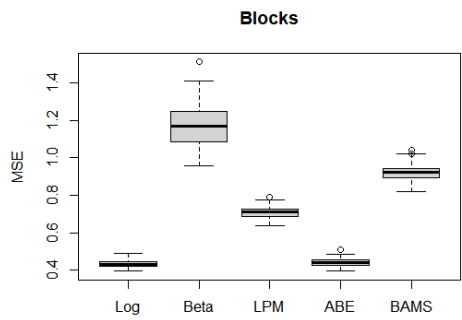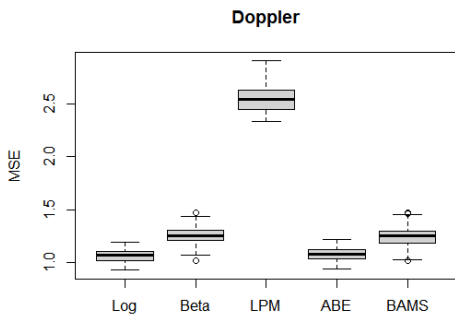
(a) Ten generated samples.
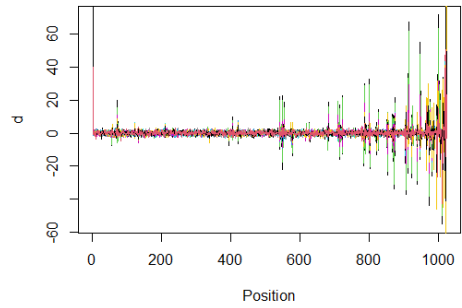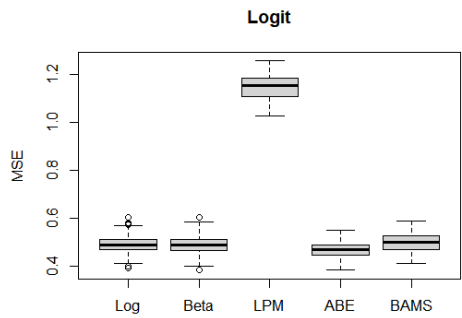
(b) Empirical wavelet coefficients.
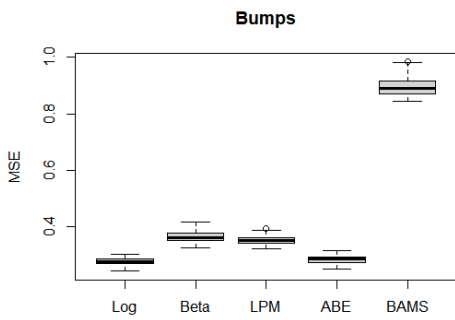
(c) Boxplot - bumps.

(d) Boxplot - blocks.

(e) Boxplot - doppler.

(f) Boxplot - logit.

Figure 4: *Ten samples of Simulation 2 with different weights (each color represents a sample) (a) and their respective empirical wavelet coefficients (b). Boxplots of rules MSE for bumps (c), blocks (d), Doppler (e) and logit (f) component functions for $M = 1024$ and $SNR = 3$.*

Table 3: AMSE (standard deviation) of simulation study 2 for aggregated data generated with the component functions bumps, blocks, Doppler and logit ($L = 4$ component functions)

| | | Simulation study 2 - $L = 4$ | | | |
| | | $M = 512$ | | $M = 1024$ | |
Function	Method	SNR = 3	SNR = 9	SNR = 3	SNR = 9
	LOG	**0.3653 (0.0232)**	0.4129 (0.0134)	**0.2776** (0.0133)	0.2736 (0.0070)
	BETA	0.4686 (0.0351)	0.5709 (0.0199)	0.3648 (0.0200)	0.4265 (0.0163)
Bumps	LPM	0.4021 (0.0275)	**0.0449 (0.0028)**	0.3536 (0.0144)	**0.0392 (0.0017)**
	ABE	0.3785 (0.0234)	0.3643 (0.0115)	0.2848 (0.0138)	0.2430 (0.0056)
	BAMS	13.432 (0.0557)	14.910 (0.0206)	0.8957 (0.0279)	0.9937 (0.0112)
	LOG	**0.4702 (0.0370)**	0.3657 (0.0141)	**0.4336 0.0199**	0.2702 (0.0092)
	BETA	11.348 (0.1024)	11.184 (0.0917)	11.769 (0.1066)	10.870 (0.1029)
Blocks	LPM	0.5930 (0.0377)	**0.0660 (0.0038)**	0.7112 (0.0309)	**0.0786 (0.0031)**
	ABE	0.4821 (0.0377)	0.3474 (0.0123)	0.4411 (0.0200)	0.2562 (0.0080)
	BAMS	13.636 (0.0813)	14.666 (0.0253)	0.9203 (0.0429)	0.9474 (0.0158)
	LOG	13.489 (0.1123)	0.7043 (0.0411)	**10.592 (0.0623)**	0.4292 (0.0218)
	BETA	19.660 (0.1966)	14.773 (0.1358)	12.548 (0.0873)	0.7968 (0.0609)
Doppler	LPM	26.068 (0.1635)	**0.2925 (0.0168)**	25.441 (0.1104)	**0.2829 (0.0127)**
	ABE	**13.015 (0.0995)**	0.5585 (0.0319)	10.732 (0.0616)	0.3584 (0.0173)
	BAMS	15.768 (0.1279)	13.576 (0.0378)	12.383 (0.0923)	10.250 (0.0288)
	LOG	0.9529 (0.0905)	0.4387 (0.0309)	0.4866 (0.0399)	0.1840 (0.0110)
	BETA	0.9929 (0.1061)	0.5040 (0.0408)	0.4858 (0.0427)	0.2250 (0.0172)
Logit	LPM	16.164 (0.1065)	**0.1785 (0.0104)**	11.475 (0.0495)	**0.1269 (0.0063)**
	ABE	**0.8913 (0.0816)**	0.3554 (0.0217)	**0.4660 (0.0335)**	0.1453 (0.0081)
	BAMS	15.619 (0.1488)	13.097 (0.0463)	0.4981 (0.0415)	0.3333 (0.0149)

constant parts of blocks. In fact, the local features of one component function are clearer as higher its associated weight is in the linear combination (2.3). Further, most of the empirical wavelet coefficients of the aggregated data are zero or very close to zero, as it is shown in Figure 3(b) and the significant coefficients are associated to locations where bumps and blocks have local features to be recovered, i.e, the spikes from bumps and the discontinuities points from blocks.

Table 2 shows the obtained AMSEs (with standard deviations of MSE) of the bayesian rules for each scenario of M and SNR. The smallest AMSE of each scenario is in bold. In fact, the logistic shrinkage rule was the best for all the scenarios with SNR = 3 for both component functions, while LPM rule was the best one for the scenarios with SNR = 9. In the opposite sense, BAMS rule and the beta shrinkage rule did not work well in general, even considering that the component functions are bounded and suitable for the beta shrinkage rule application.

The number of sample points M did not have meaningful impact in the estimation process but the rules performed better for SNR = 9 as expected. Figures 3(c) and (d) present the boxplots of MSEs for bumps and blocks component functions respectively for $M = 1024$, SNR = 3 and Figures 3(e) and (f) for SNR = 9 scenarios. Note that BAMS and beta shrinkage rule did not have good performances in comparison with the other rules in almost all the replications. Finally, it is important to note that ABE rule had good performances for scenarios with SNR = 3 and their results were very close to the logistic shrinkage rule ones for both component functions and scenarios. Similar results were obtained for $M = 512$.

4.2. Simulation study 2

In this study, 50 samples were generated according to model (2.3) by considering bumps, blocks, Doppler and logit as underlying component functions ($L = 4$). Figures 4(a) and (b) show ten generated samples and their respective empirical wavelet coefficients. Note that it is more difficult to visualize

local features of the component functions in the aggregated data, but a good estimation process should recover the spikes of bumps, the discontinuities and piecewise constant parts of blocks, the oscillations of Doppler and the smoothness of logit.

Table 3 presents the AMSEs and their standard deviations of Simulation 2. For the scenarios under SNR = 9, LPM rule had the best performance for the four component functions and both M values, as in Simulation study 1. Further, ABE had the second best performance in all the scenarios under SNR = 9 and the logistic shrinkage rule was the third best one. For SNR = 3, the logistic shrinkage rule and ABE rule had the best performances, the first one was the best for bumps and blocks functions and the second one was the best for Doppler and logit functions. However, their overall works were very close with each other, which it is important, since one rule is applied to estimate all the component functions in practice. Therefore, although LPM and logistic shrinkage rule had the best performances under SNR = 9 and 3 respectively, it should be noted that ABE worked well in both SNR values, which suggests flexibility of this rule independently of the noise amount in the data.

On the other hand, BAMS rule did not work well in all the scenarios and the beta shrinkage rule had bad results mainly for blocks and Doppler component functions. Figures 3(c)–(f) show the boxplots of MSE for bumps, blocks, Doppler and logit functions respectively for $M = 1024$ and SNR = 3. Note again that BAMS did not have good performance for the four underlying functions in almost all the replications. Moreover, note that LPM rule did not have good behaviour for SNR = 3 scenarios, although it was the best one for scenarios with lower noise, i.e, SNR = 9. Finally, the boxplots and the standard deviations of MSE show little variation of the MSE's in the replications for the five rules.

4.3. Simulation study 3

Finally, simulation study 3 considered all the six functions of Table 1 as component functions ($L = 6$) for aggregated data generation. As in simulation study 2, the local features of the functions are not so clear by visualization of the generated samples, as it is shown in Figure 5(a). Further, the rules should detect, in addition to the local features of the component functions of simulation 2, the discontinuity point of Heavisine function and the smoothness of SpaHet function. Figure 5(b) also present the empirical wavelet coefficients of ten generated samples. High empirical coefficients values mean important locations of component functions features to be recovered by the shrinkage rules.

Table 4 shows the AMSEs and the standard deviations of MSEs of the considered methods for all the scenarios. The logistic shrinkage rule and ABE had good overall performances for SNR = 3 scenarios as in simulation studies 1 and 2. Under this context, the logistic shrinkage rule was the best for bumps and blocks functions, while ABE was the best one for Doppler function. The novelty is that BAMS method was the best for Heavisine, logit and SpaHet functions, although it did not work well in the other functions. Since the chosen method estimates simultaneously the six functions, logistic shrinkage rule and ABE would be preferable under this context again.

For scenarios under SNR = 9, LPM had an overall good performance, being the best one for bumps, blocks and Doppler functions, while ABE was the best for estimating the remaining ones. Note therefore that ABE had good performances for both SNR values as in the previous simulation studies, which indicates its application independently of the noise amount in the data. Figures 5(c)–(h) present the boxplots of MSEs for the six component functions in the scenario of $M = 1024$ and SNR = 3. It should be noted that the beta shrinkage rule did not again have good performances and that LPM rule was also not well succeeded in estimating Doppler, Heavisine and the smooth functions logit and SpaHet.
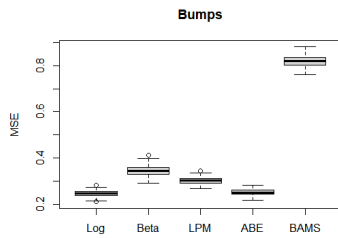
In general, the three simulation studies suggest that the signal to noise ratio (SNR) of the generated datasets has impact on the bayesian wavelet shrinkage rule to be chosen in applications. For SNR =
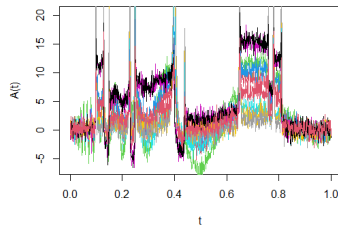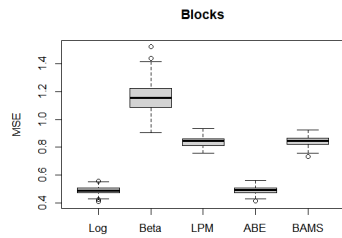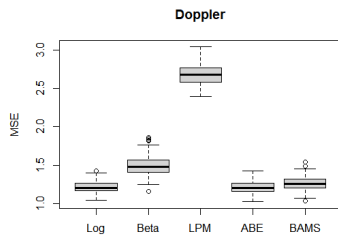
(a) Ten generated samples.
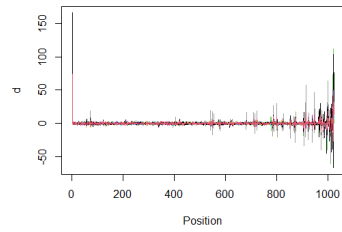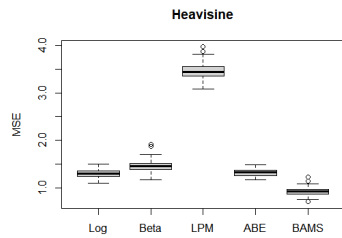
(b) Empirical wavelet coefficients.
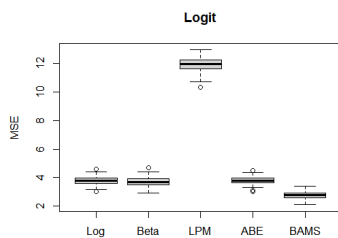

(c) Boxplot - bumps.
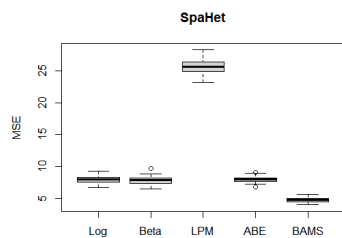
(d) Boxplot - blocks.

(e) Boxplot - doppler.

(f) Boxplot - heavisine.

(g) Boxplot - logit.

(h) Boxplot - SpaHet.

Figure 5: *Ten samples of simulation 3 with different weights (each color represents a sample) (a) and their respective empirical wavelet coefficients (b). Boxplots of rules MSE for bumps (c), blocks (d), Doppler (e), Heavisine (f), logit (g) and SpaHet (h) component functions for M = 1024 and SNR = 3.*

Table 4: AMSE (standard deviation) of simulation study 3 for aggregated data generated with the component functions bumps, blocks, Doppler, Heavisine, logit and SpaHet ($L = 6$ component functions)

Simulation study 3 - $L = 6$					
		$M = 512$		$M = 1024$	
Function	**Method**	**SNR = 3**	**SNR = 9**	**SNR = 3**	**SNR = 9**
Bumps	LOG	**0.3899 (0.0262)**	0.4188 (0.0118)	**0.2447 (0.0134)**	0.2455 (0.0059)
	BETA	0.4998 (0.0366)	0.5840 (0.0206)	0.3457 (0.0243)	0.4056 (0.0181)
	LPM	0.4411 (0.0286)	**0.0489 (0.0028)**	0.3020 (0.0141)	**0.0334 (0.001)**
	ABE	0.4032 (0.0268)	0.3690 (0.0101)	0.2512 (0.0135)	0.2191 (0.0045)
	BAMS	1.374 (0.0609)	1.512 (0.0204)	0.8199 (0.0255)	0.9190 (0.0088)
Blocks	LOG	**0.5138 (0.0402)**	0.4015 (0.0151)	**0.4880 (0.0272)**	0.2285 (0.0104)
	BETA	1.176 (0.1184)	1.176 (0.0960)	1.164 (0.1200)	1.007 (0.1235)
	LPM	0.6998 (0.0451)	**0.0771 (0.0042)**	0.8401 (0.0366)	**0.0927 (0.0037)**
	ABE	0.5243 (0.0407)	0.3787 (0.0134)	0.4906 (0.0262)	0.2173 (0.0089)
	BAMS	1.382 (0.0793)	1.493 (0.0272)	0.8439 (0.0411)	0.8255 (0.0163)
Doppler	LOG	0.8734 (0.0703)	0.5543 (0.0310)	1.214 (0.0785)	0.4121 (0.0207)
	BETA	1.274 (0.1104)	1.071 (0.0727)	1.495 (0.1384)	0.9149 (0.0890)
	LPM	1.384 (0.0819)	**0.1543 (0.0100)**	2.671 (0.1278)	**0.2944 (0.0119)**
	ABE	**0.8518 (0.0637)**	0.4450 (0.0240)	**1.213 (0.0759)**	0.3647 (0.0183)
	BAMS	1.534 (0.1068)	1.556 (0.0394)	1.264 (0.0893)	1.028 (0.0318)
Heavisine	LOG	5.090 (0.4850)	1.164 (0.1098)	1.304 (0.0816)	0.3282 (0.0215)
	BETA	5.566 (0.5527)	1.889 (0.2260)	1.462 (0.1226)	0.6707 (0.0736)
	LPM	11.156 (0.7521)	1.226 (0.0756)	3.469 (0.1575)	0.3852 (0.0180)
	ABE	4.965 (0.4490)	**0.9368 (0.0810)**	1.320 (0.0732)	**0.2709 (0.0188)**
	BAMS	**4.296 (0.4739)**	2.184 (0.1267)	**0.9275 (0.0877)**	0.5113 (0.0256)
Logit	LOG	10.561 (0.8675)	2.606 (0.2831)	3.794 (0.2898)	0.9159 (0.075)
	BETA	10.228 (0.8989)	2.611 (0.3215)	3.709 (0.3132)	1.186 (0.1754)
	LPM	23.801 (1.177)	2.658 (0.1600)	11.931 (0.4723)	1.318 (0.0552)
	ABE	9.756 (0.7668)	**1.659 (0.1897)**	3.806 (0.2565)	**0.6812 (0.0562)**
	BAMS	**6.454 (0.7391)**	2.084 (0.2118)	**2.775 (0.2471)**	1.326 (0.0789)
SpaHet	LOG	3.580 (0.3248)	0.8431 (0.1008)	7.916 (0.4996)	1.851 (0.1427)
	BETA	3.635 (0.3418)	1.021 (0.1392)	7.817 (0.5727)	2.677 (0.2301)
	LPM	7.797 (0.5035)	0.8733 (0.0526)	25.654 (1.059)	2.838 (0.1359)
	ABE	3.398 (0.3027)	**0.6180 (0.0668)**	7.934 (0.4271)	**1.100 (0.0975)**
	BAMS	**3.319 (0.4242)**	1.783 (0.0999)	**4.749 (0.4070)**	1.933 (0.1313)

3 (high amount of noise in the data) scenarios, the logistic shrinkage rule and ABE were the best in terms of AMSE and should be considered in high noise datasets applications. For SNR = 9 (low amount of noise in the data) scenarios, the LPM rule was the best one in terms of AMSE, although ABE also worked well under this context. Despite the six component functions are bounded, the beta shrinkage rule did not have good performances for both SNR scenarios, as BAMS rule did not too. The sample size and the number of component functions in the aggregated data did not have impact on the estimators performances.

5. Application - Tecator dataset

Aggregated curves frequently arise in chemometrics, specifically in spectroscopy. According to the Beer - Lambert law (Brereton, 2003), the absorbance curve $A(t)$ at wavelengths t of a given substance can be represented as a linear combination of the absorbance curves $\alpha_l(t), l = 1, \ldots, L$ of its L constituents, where the weights y_l are the known concentrations of the constituents. Thus, the relationship between the absorbance of the substance and the absorbances of its constituents can be modelled by the aggregated functional model (2.1).
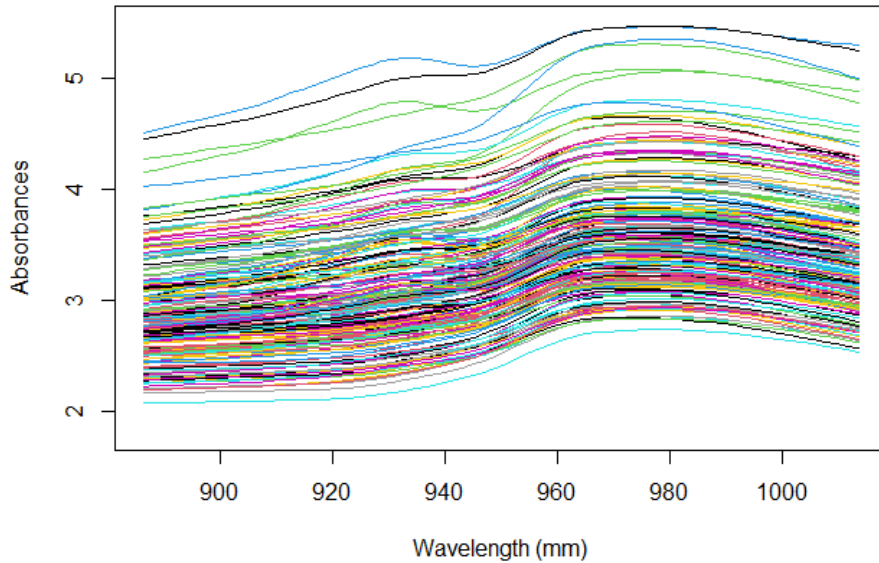
Figure 6: *Tecator dataset. Each color represents an absorbance curve of a sample meat. The dataset consists of 215 samples.*

In this sense, it is considered the well known tecator dataset as an application of the methodology and results of the present work. This dataset is available in the R-package fda.usc (Bande and de la Fuente, 2012) and consists of absorbance curves of $I = 215$ meat samples. It is considered here the absorbances in $M = 64$ equally spaced wavelength points from 850 mm to 1050 mm. Figure 6 show the absorbance curves of the samples.

Each meat sample is composed by fat, water and protein constituents ($L = 3$) in different concentrations. The goal is to estimate the absorbance curves of those constituents. It is known by the measurement process that low noise is present in the data, i.e, it is considered high SNR in the dataset. This information is relevant to choose a bayesian shrinkage rule with a good performance for high SNR, which leads to LPM rule choice according to the simulation studies results of Section 4.

It was applied a DWT (Daubechies basis with ten null moments) on the dataset and the matrix D of the empirical wavelet coefficients was obtained. The LPM rule was applied on D to remove noise of the empirical coefficients and the estimated wavelet coefficients of the constituents absorbance curves matrix $\hat{\Gamma}$ was obtained by (2.8). Finally, the points of the constituents absorbance curves $\hat{\alpha}$ were estimated by (2.9).

Figures 7(a), (c) and (e) show the estimated absorbance curves of fat, water and protein constituents respectively and Figures 7(b), (d) and (f) present the estimated wavelet coefficients of the respective absorbance curves. Note that the absorbance curves are smooth, so they can be represented by a few number of nonzero wavelet coefficients in the wavelet domain. In fact, the significant coefficients are related to specific features of the curves, such as minimum and maximum local regions.
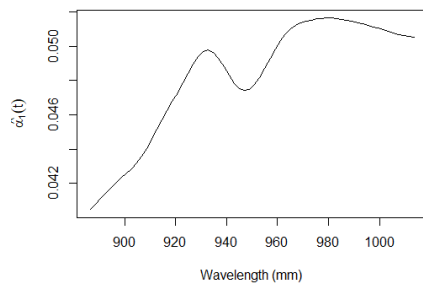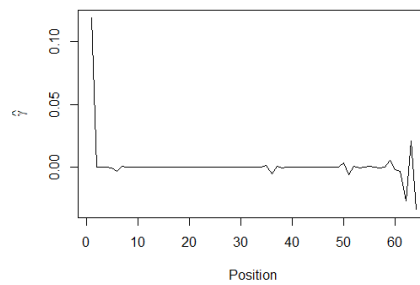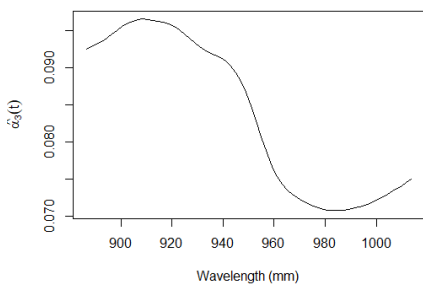
(a) Fat curve.
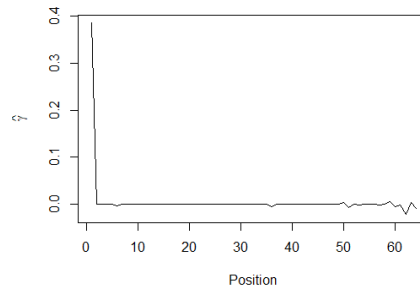
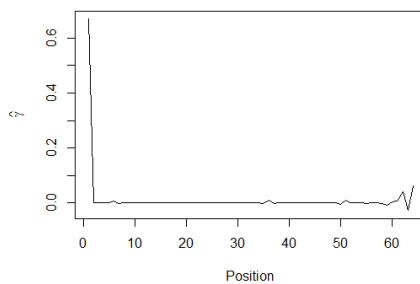(b) Estimated wavelet coefficients - fat curve.
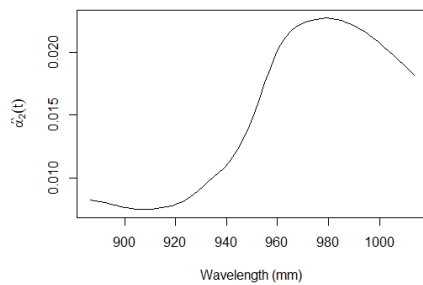
(c) Water curve.

(d) Estimated wavelet coefficients - water curve.

(e) Protein curve.

(f) Estimated wavelet coefficients - protein curve.

Figure 7: *Estimated constituents absorbance curves (a), (c) and (e) and their respectives wavelet coefficients (b), (d) and (f) of Tecator dataset. The estimation process was done by the application of LPM shrinkage rule, once it is suitable for high SNR datasets according to results of Section 4.*

6. Final considerations

The present paper had the main goal to compare the performance of five bayesian wavelet shrinkage methods, the shrinkage rules under logistic and beta priors, large posterior mode (LPM), amplitude-

scale invariant Bayes estimator (ABE) and Bayesian adaptive multiresolution smoother (BAMS), in estimating component curves of aggregated functional data. Simulation studies involving the so called Donoho-Johnstone test functions, the logit and SpaHet functions as component functions were conducted, for different sample sizes (M = 512 and 1024), signal to noise ratios (SNR = 3 and 9) and number of component curves (L = 2, 4 and 6). In these simulations, data points were generated by aggregating the component functions according to the model (2.1). After the application of a discrete wavelet transformation on the generated data, the shrinkage rules were applied on the empirical wavelet coefficients to reduce noise and estimate the wavelet coefficients of the underlying component functions. Later, these component functions were estimated by (2.9). The mean squared error was calculated in each replication and the averaged mean squared error was obtained for each shrinkage method as the performance measure.

For scenarios with SNR = 9, the LPM method had the best performances in terms of averaged mean squared error measure. This result suggests the use of LPM for aggregated functional data with low noise presence in the dataset. On the other hand, for scenarios with SNR = 3 (high noise presence), the shrinkage rule under logistic prior and ABE method had good performances. BAMS and the shrinkage rule under beta prior did not performed well in general. As expected, the overall performances were improved by sample size M = 1024 in relation to to M = 512 but there was not detected meaningful difference in the methods performances for different number of component curves.

Further comparisons with standard and no bayesian shrinkage or thresholding rules, the performances in terms of other measures such as averaged median absolute error and the impact of the chosen wavelet basis in the estimation process in aggregated curves are topics of interest for future works.

References

Angelini C and Vidakovic B (2004). Gama-Minimax wavelet shrinkage: A robust incorporation of information about energy of a signal in denoising applications, *Statistica Sinica*, **14**, 103–125.

Bande MF and de la Fuente MO (2012). Statistical computing in functional data analysis: The R-package fda.usc., *Journal of Statistical Software*, **51**, 1–28.

Brereton RG (2003). *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, John Wiley and Sons, Chichester.

Brown PJ, Vannucci M, and Fearn T (1998a). Bayesian wavelength selection in multicomponent analysis, *Journal of Chemometrics*, **12**, 173–182.

Brown PJ, Vannucci M, and Fearn T (1998b). Multivariate Bayesian variable selection and prediction, *Journal of the Royal Statistical Society*, Series B, **60**, 627–641.

Brown PJ, Vannucci M, and Fearn T (2001). Bayesian wavelet regression on curves with application to a spectroscopic calibration problem, *Journal of the American Statistical Association*, **96**, 398–408.

Cowe IA and McNicol JW (1985). The use of principal components in the analysis of near-infrared spectra, *Applied Spectroscopy*, **39**, 257–266.

Cutillo L, Jung YY, Ruggeri F, and Vidakovic B (2008). Larger posterior mode wavelet thresholding and applications, *Journal of Statistical Planning and Inference*, **138**, 3758–3773.

Dias R, Garcia NL, and Martarelli A (2009). Non-Parametric estimation for aggregated functional data for electric load monitoring, *Environmetrics*, **20**, 111–130.

Dias R, Garcia NL, and Schmidt A (2013). A hierarchical model for aggregated functional data,

Technometrics, **55**, 321.

Donoho DL (1993a). Nonlinear wavelet methods of recovery for signals, densities, and spectra from indirect and noisy data, *Proceedings of Symposia in Applied Mathematics*, volume 47, American Mathematical Society, Providence: Rhode Island.

Donoho DL (1993b). Unconditional bases are optimal bases for data compression and statistical estimation, *Applied and Computational Harmonic Analysis*, **1**, 100–115.

Donoho DL (1995a). De-Noising by soft-thresholding, *IEEE Transactions on Information Theory*, **41**, 613–627.

Donoho DL (1995b). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Applied and Computational Harmonic Analysis*, **2**, 101–126.

Donoho DL and Johnstone IM (1994a). Ideal denoising in an orthonormal basis chosen from a library of bases, *Comptes Rendus-Academie des Sciences Paris Serie*, **319**, 1317–1322.

Donoho DL and Johnstone IM, (1994b). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**, 425–455.

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

Figueiredo MAT and Nowak RD (2001). Wavelet-Based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior, *IEEE Transactions on Image Processing*, **10**, 1322–1331.

Goepp V, Bouaziz O, and Nuel G (2018). Spline Regression with Automatic Knot Selection, Available from: *arXiv: 1808.01770v1*

Ruppert D, Wand M, and Carroll RJ (2003). *Semiparametric Regression*, Cambridge University Press, Cambridge.

Sousa ARS (2020). Bayesian wavelet shrinkage with logistic prior, *Communications in Statistics: Simulation and Computation*, **51**, 4700–4714, Available from: http://doi: 10.1080/03610918.2020.1747076

Sousa ARS, Garcia NL, and Vidakovic B (2020). Bayesian wavelet shrinkage with beta prior, *Computational Statistics*, **36**, 1341–1363.

Sousa ARS (2022). A wavelet-based method in aggregated functional data analysis, *arXiv preprint [stat.ME]*, Available from: *arXiv:2205.15969v1*

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

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

Wand MP (2000). A comparison of regression spline smoothing procedures, *Computational Statistics*, **15**, 443–462.

Wold S, Martens H, and Wold H (1983). The multivariate calibration problem in chemistry solved by PLS. In A Ruhe and B Kagstrom (Eds), *Matrix Pencils*, (pp. 286–293), Springer, Heidelberg.