

I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



D I S C U S S I O N
P A P E R

0902

**TREE-STRUCTURED WAVELET ESTIMATION
IN A MIXED EFFECTS MODEL FOR SPECTRA
OF REPLICATED TIME SERIES**

FREYERMUTH, J.-M., OMBAO, H. and R. von SACHS

This file can be downloaded from
<http://www.stat.ucl.ac.be/ISpub>

Tree-structured Wavelet Estimation in a Mixed Effects Model for Spectra of Replicated Time Series.

Jean-Marc Freyermuth ¹ Hernando Ombao ² Rainer von Sachs ³

29 January 2009

Abstract

This paper develops a method for estimating the spectrum of a stationary process using time series traces recorded from experimental designs. Our procedure estimates the “common” log-spectrum and the variability over the traces (or subjects) using a mixed effects model. We combine the use of spatially adaptive smoothing methods with recursive dyadic partitioning to construct a predictive model. The method is easy to implement and can handle large data sets because it uses the discrete wavelet transform which is computationally efficient. Numerical studies confirm that the proposed method performs very well despite its simplicity. The method is also applied to a multi-subject electroencephalogram data set.

KEYWORDS : log-spectrum estimation; mixed effects models; recursive dyadic partitioning; panel time series; tree-structured wavelets

¹Université catholique de Louvain, Institut de statistique, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium. Email: Jean-Marc.Freyermuth@uclouvain.be. Financial support from the contract “Projet d’Actions de Recherche Concertées” nr. 07/12/002 of the “Communauté française de Belgique”, granted by the “Académie universitaire Louvain”, is gratefully acknowledged.

²Brown University, Center for Statistical Sciences, 121 South Main Street, Providence, RI 02912, USA. Email: ombao@stat.brown.edu.

³Université catholique de Louvain, Institut de statistique, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium. Email: rvs@uclouvain.be. Financial support from the IAP research network grant P 06/03 of the Belgian government (Belgian Science Policy) is gratefully acknowledged.

1 Introduction

We develop a novel wavelet-based procedure for estimating the spectrum (or log-spectrum) of a stationary process from several time series traces. Our goal is to provide a methodological approach which uses the information included in these time series traces to both estimate the spectrum common to this collection and the deviation from this common spectrum, in an experimental situation of possibly only a small number of time series traces. In order to statistically describe this set-up, we embed it into a mixed effects model: its fixed effect corresponds to the log-spectrum whereas the variance of its zero-mean random effects models the aforementioned deviation. We recall that in this set-up, common to mixed effects modelling in general, the statistical tasks are not only to estimate trace specific and common log-spectra, including the variability of the random effects, but also to predict these effects giving reliable prediction intervals based on the former estimates.

The motivation of the general methodology we propose can be found in a neuroscience experiment where a subject is instructed to move a joystick to the right (or the left accordingly) when a cursor flashes on the right side (or left side) of a visual field. For each trial (i.e., each cursor flash), electroencephalograms (EEGs) traces are collected. The ultimate goal in the study is to investigate brain processes that are associated with motor tasks and motor intention. In particular, the scientists are interested in identifying differences in brain network for the left vs right cursor flash stimuli. We shall tackle one of the first steps toward the ultimate goal - which is to estimate the log-spectrum associated with the right stimuli using time series traces recorded from $S = 8$ subjects in the experiment. Our proposed procedure views each time series trace recorded in response to a stimulus to be a realization of a stochastic process having a log-spectrum that is unique to that stimulus. While methods for estimating the log-spectrum from a *single* time series are well-developed, there is clearly a lack of methods that are suited for an experimental setting where a *few* time series traces are available for each stimulus.

Our proposed procedure will utilize modern non-parametric curve estimation methods - which need to be tailored according to our specific situation where the ultimate goals are estimation of the common spectrum and subject-specific prediction based on possibly very few time series traces (or subjects). We do so by developing a methodology that will combine the well-established advantages of wavelet threshold estimation of curves that show important localized structure with recent approaches on Recursive Dyadic Partitioning (RDP hereafter). For the former, we exploit the improved denoising properties of "tree-structured" wavelet methods ([Baraniuk \(1999\)](#), [Autin \(2008\)](#)), whereas the use of Haar wavelets within tree-structuring will allow us to interpret our estimator as a piecewise constant fit subordinate to an RDP ([Donoho et al. \(2000\)](#)). Moreover, since wavelet coefficients are known to be closer to normality than the original non-Gaussian data, working in the coefficient domain of a wavelet-based approach allows to use classical prediction approaches based on Gaussianity.

We highlight the advantage of our approach. The "wedding" of tree-structured wavelets with RDP gives a user-friendly interpretation of the resulting estimator as "semi-linear". That is, on each of the elements of the DP it can be seen as a linear and hence kernel-like smoother. However, since the adaptively chosen DP is generally not composed of homogeneous segments, the overall estimator is non-linear and akin to a kernel method with a local smoothing parameter. We hasten to add that even though we start from piecewise constant fits along an adaptively chosen dyadic partition, our final semi-linear estimators are not restricted to remain blocky over frequency. Our procedure will apply the "average interpolation (AI)" approach suggested by [Donoho \(1993\)](#) or [Donoho et al. \(2000\)](#) to construct a smooth reconstruction along the initially found DP.

The classical approach to estimating the log-spectrum using several time series traces in an experimental design is based on multivariate analysis of variance (MANOVA). While the log-periodogram curve is defined over the *entire* interval $(0, N_f)$ (where N_f is the Nyquist frequency), MANOVA essentially reduces the log-periodogram curve to some vector of averages of the log-periodograms across disjoint bins of $(0, N_f)$. In

EEG studies, the log-spectral power is examined for the delta ($0 - 4$ Hertz), theta ($4 - 8$), alpha ($8 - 12$), beta ($12 - 30$) and gamma (> 30) bands. Our mixed-effects curve estimation approach based on wavelets, however, will offer high frequency resolution and hence will allow us to extract more refined and highly localized information within each of these broad frequency bands. Though aggregating individual estimates to construct an "over-all" curve smoother is not entirely new (see [Bunea et al. \(2006\)](#)), we believe that using a mixed-effects set-up will allow simultaneous estimation of the common log-spectrum, modelling of variation across subjects and constructing valid confidence/prediction intervals.

Our method can be seen to fall into the general area of functional mixed effects modelling - though we deal with the specific problem of log-spectral estimation using replicated time series. We briefly explain similarities and differences with existing work. Most of the ground-breaking work on functional mixed effects models is based on splines, with e.g. recent work by Guo (2002) where the subject-specific random functions are modelled non-parametrically using the same functions used to represent the fixed effect. As mentioned, the proposed procedure will utilize wavelet thresholding which has been shown to be well suited for functional data that are characterized by localized peaks and troughs ([Morris et al. \(2003\)](#), [Morris and Carroll \(2006\)](#), [Antoniadis and Sapatinas \(2007\)](#)). Moreover, as already exploited by these citations, the hierarchical structure in the coefficient domain of wavelets allows to treat multiple curves where on the one hand the locations of peaks and troughs may differ across curves but where at the same time a certain reduction of multivariate complexity is at order. Both the Bayesian approaches of Morris et al. ([Morris et al. \(2003\)](#), [Morris and Carroll \(2006\)](#)) and the inference-oriented approach of [Antoniadis and Sapatinas \(2007\)](#), using multiple likelihood-ratio tests, carry out estimation by applying the linear mixed effects model in the wavelet coefficient domain. Our approach models the log-periodogram curves directly rather than the wavelet coefficients and uses hard thresholding which has been demonstrated to keep its known optimality properties in the multiple-curve settings ([Bunea et al. \(2007\)](#)). Second, the use of wavelets within

a tree-structured RDP approach is an elegant way and actually the key feature for constructing a variance estimator of the underlying mixed-effects model. In fact, our approach can be seen as a suitably restricted non-parametric curve estimator. As in [Antoniadis and Sapatinas \(2007\)](#), the complexity of the non-parametric variance components is constrained to not exceed those of the log-spectrum. Using a hierarchic tree-structured estimation approach enables the procedure to be fully adaptive (in a non-parametric sense) to this constraint, motivated from having at our disposal only few time series traces - a situation which would not fit into the much more general set-up of the aforementioned functional approaches.

Another advantage of our approach is that it benefits from perhaps a less known property of tree-structured wavelet thresholding to control both the false positives and the false negatives in the coefficient domain much better than classical wavelet thresholding. While classical wavelet thresholding occasionally kills significant local information in the coefficient domain and accepts too much noise at locations which are supposed to correspond to smooth curve structure, we observe that our estimators suffer much less from these well-known shortcomings. Our proposed method combines information from multiple curves within a tree-structured approach - resulting in a significant improvement of our estimator of the population log-spectrum.

The rest of the paper is organized as follows. In Section 2, we describe our model set-up and deliver the necessary background on log-spectral estimation in a mixed-effects set-up, on RDP's and on wavelet tree-structured estimation. Section 3 then presents our methodology in detail. We derive point estimates of both the trace-specific and the common (or population) log-spectrum, a point estimate of the variation across traces, pointwise confidence intervals for the common log-spectrum and for the pointwise estimated predictor of each of the subject-specific log-spectra. We shall report the performance of our methodology via simulation studies in Section 4 and we apply our method to an EEG data set in Section 5. We end by a conclusion which also discusses further directions of this promising research of tree-structured spectrum estimation. An appendix section contains some additional details on our

tree algorithms and on the derivation of some theoretical justification of our proposed approach.

2 Model set-up and background

2.1 Log-periodograms

Let $\{X_t^s, t = 1, \dots, T\}$ be a stationary time series trace for trial s , $s = 1, \dots, S$. We shall assume that the time series traces are independent replicates from a process whose log-spectrum is denoted as $h(\nu)$, $\nu \in [0, 1]$. Following the widely-used Wahba approximation of log-periodograms by log χ^2 -variates (see [Wahba \(1980\)](#)), the bias-corrected log-periodogram for trial s and frequency $\nu_\ell = \frac{\ell}{T}$, $\ell = 0, \dots, T-1$ is defined by

$$Y_\ell^s = \log \frac{1}{T} \left| \sum_{t=1}^T X_t^s \exp(-i2\pi\nu_\ell t) \right|^2 + \gamma, \quad s = 1, \dots, S, \quad (2.1)$$

where $\gamma = 0.57721$ is the Euler-Mascheroni constant. We restrict ourselves to dyadic sample sizes $T = 2^J$ to avoid any complications in the subsequent wavelet estimation. We write the mixed effects model for the (bias-corrected) log-periodogram as follows:

$$Y_\ell^s = h^s(\nu_\ell) + \epsilon_\ell^s, \quad \nu_\ell = \frac{\ell}{T}, \quad \ell = 0, \dots, T-1 = 2^J - 1, \quad s = 1, \dots, S, \quad (2.2)$$

$$= h(\nu_\ell) + z^s(\nu_\ell) + \epsilon_\ell^s, \quad (2.3)$$

where the elements of the model are as follows:

- (i.) $h(\nu_\ell)$ is the unknown stimulus-specific (or population) log-spectrum (fixed effect),
- (ii.) $h^s(\nu_\ell)$ is the trial-specific (or subject-specific) log-spectrum;
- (iii.) $z^s(\nu_\ell)$ is the deviation (random effect) of trial s log-spectrum from the over-all log-spectrum $h(\nu_\ell)$ which models variation between trials with $\mathbb{E}[z^s(\nu_\ell)] = 0$ for all s and ν , and with $V(\nu_\ell) := \text{Var}[z^s(\nu_\ell)]$ the variance at frequency ν_ℓ common to the S random subjects; and
- (iv.) $\epsilon_\ell^s \sim \log \chi_2^2$ approximately iid over $\ell = 1, \dots, T/2 - 1$ and s with $\mathbb{E}[\epsilon_\ell^s] = 0$ and with $\text{Var}[\epsilon_\ell^s] = \sigma_\epsilon^2 = \frac{\pi^2}{6}$ according to the aforementioned Wahba approximation, see also, e.g., [Gao \(1997\)](#), section 2.2.

The estimation of the log-spectrum is achieved by smoothing the log-periodogram on $[0, 1]$ (note that, due to the symmetry of the spectrum about $\nu = 0.5$, x-axes of the plots in the sequel are restricted to $[0.5, 1]$). This crucial and difficult issue requires spatially adaptive smoothing methods since the study of real data processes often reveals the presence of very localized frequency bands, i.e. the presence of peaks in the spectrum. Wavelet methods are known to be competitive in this context (see [Gao \(1997\)](#) or [Moulin \(1994\)](#)).

2.2 Basic ideas on wavelet tree-structured estimation and RDP's

Let φ and ψ respectively denote a set of compactly supported scaling and wavelet functions defined on $[0, 1]$ such that the collection $\{\varphi_{j_0,k}, k = 0, \dots, 2^{j_0} - 1; \psi_{j,k}, j \geq j_0, k = 0, \dots, 2^j - 1\}$ of their translated and dilated versions, $\varphi_{j,k}(\nu) = 2^{j/2}\varphi(2^j\nu - k)$ and $\psi_{j,k}(\nu) = 2^{j/2}\psi(2^j\nu - k)$, generates an orthonormal basis of $(L_2[0, 1], \langle \cdot, \cdot \rangle)$, the space of square integrable functions on $[0, 1]$ endowed with the inner product $\langle h, g \rangle = \int_0^1 h(\nu)g(\nu)d\nu$. We consider periodized wavelet bases on $[0, 1]$, for details, we refer the interested reader to [Mallat \(1998\)](#).

The log-periodogram of the subject s has the following multiscale representation in the wavelet basis :

$$Y^s(\nu) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k}^s \varphi_{j_0,k}(\nu) + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \hat{d}_{j,k}^s \psi_{j,k}(\nu), \nu \in [0, 1], \quad (2.4)$$

$$= \sum_{k=0}^{2^J-1} \hat{c}_{J,k}^s \varphi_{J,k}(\nu), \quad (2.5)$$

where $\hat{c}_{j,k}^s = \langle Y^s, \varphi_{j,k} \rangle$ and $\hat{d}_{j,k}^s = \langle Y^s, \psi_{j,k} \rangle$ are the empirical scaling and wavelet coefficients, respectively. We denote the “true” coefficients of the target function h^s by $c_{j,k}^s = \langle h^s, \varphi_{j,k} \rangle$ and $d_{j,k}^s = \langle h^s, \psi_{j,k} \rangle$, respectively. Analogously, all the quantities without superindex s are used to denote the empirical and true coefficients of the population log-spectrum h .

The set of translated scaling functions on scale j , $\{\varphi_{j,k}\}_k$, constitute a linear approximation space $V_j \subset L_2[0, 1]$. As usual for (periodic) spectrum estimation problems,

we consider the wavelet expansion (2.4) where $j_0 = 0$, in order to fix the primary approximation scale to be the coarsest possible, with only one scaling coefficient which represents the "mean" of the signal over the whole interval $[0, 1]$. An estimate of h^s which has been shown (e.g. by Donoho (1994)) to have very good denoising properties, is obtained by shrinking to zero the wavelet coefficients $\hat{d}_{j,k}^s$ in the equation (2.4) with magnitudes below a threshold value λ . We denote the unstructured set of wavelet coefficients of this nonlinear estimator by $I_{NL} = \{(j, k) \mid |\hat{d}_{j,k}| > \lambda\}$. The 'classical' wavelet-based estimation of log-spectra, due to Gao (1997), considers scale-dependent threshold values λ_j , based on large deviation properties of the distribution of the wavelet coefficients. These threshold values are generally too high. They ensure a nice removal of the noise but the estimation of the localized structure in the underlying curve is suboptimal (see, e.g., Jansen (2001), page 39). It would be more convenient to rather minimize the mean squared error but, in this case, nonlinear estimation is not robust enough in particular when the distribution of the noise is skewed. The estimate often shows unappealing visual artifacts (spurious bumps) due to large wavelet coefficients at fine resolution scales generated from the random noise ("false positives"). Lee (2002) showed the ability of tree-structured wavelets to improve the quality of estimation although, to the best of our knowledge, no application to the specific problem of spectrum estimation has been yet developed in the literature.

Tree-structured wavelets are based on the hierarchical interpretation of the wavelet expansion (2.4). The wavelets functions $\{\psi_{j,k}\}_{j,k}$ are arranged over a nested multi-scale structure such that the support of each $\psi_{j,k}$ contains the supports of $\psi_{j+1,2k}$ and $\psi_{j+1,2k+1}$. This induces a hierarchy among the wavelet coefficients which can be represented over a dyadic tree structure rooted to $\hat{d}_{0,0}$ (see Figure 7.1 in the appendix), with the practical implication that at the location of a singularity in the log-spectrum we observe the persistence of large wavelet coefficients over all scales. This can be used as additional information to the coefficient magnitudes in order to outperform classical nonlinear thresholding methods (see Baraniuk (1999)). The idea is to require that the set of non zero wavelet coefficients after thresholding form a *connected rooted subtree*,

this structured set of wavelet coefficients is denoted by I_T (see the Figures 7.2 and 7.3 to compare with I_{NL}). In other words, if we say that $\hat{d}_{j,k}$ is the parent of its two children $\hat{d}_{j+1,2k}$ and $\hat{d}_{j+1,2k+1}$, the hereditary constraint means that we cannot include in our estimator a large coefficient unless all its parents are large. More specifically, in this paper we use a variant of a tree-structured algorithm developed by Engel (1994) (for the description of the algorithm we defer to the appendix 7.2). The choice of the appropriate, and ideally data-driven, thresholds λ_T which lead to "optimal" sets I_T is still a matter of ongoing research.

Any kind of orthogonal wavelets can be used for tree-structured estimation but the Haar wavelets (the boxcar function $\varphi(\nu) = \mathbb{1}_{[0,1)}(\nu)$ and wavelet $\psi(\nu) = \mathbb{1}_{[0,1/2)} - \mathbb{1}_{[1/2,1)}(\nu)$) are particularly well-suited. They are naturally associated to dyadic trees since the support of the Haar functions correspond to the dyadic intervals $[k2^{-j}, (k+1)2^{-j}]$. I.e., any representation with respect to Haar Tree-Structured Wavelets (HTSW) gives rise to a Dyadic Partition (DP) of $[0, 1]$. Moreover, due to the abovementioned hierarchy of TSW, the collection of all possible HTSW representations is equivalent to a Complete Recursive Dyadic Partitioning (C-RDP) scheme on the unit interval. Thanks to this property, a HTSW estimator can be written as a unique weighted sum of boxcar functions. By slight abuse of notation, we will equivalently denote by I_T either the set of wavelets or the set of scaling coefficients in the equivalent representations (cf. equations (2.4, 2.5):

$$\hat{h}^s = \hat{c}_{0,0}^s \varphi_{0,0} + \sum_{I \in I_T} \hat{d}_I^s \psi_I = \sum_{I \in I_T} \hat{c}_I^s \varphi_I ,$$

where $I := (j, k)$ and where the collection of the supports of the scaling functions $\{\varphi_I \mid I \in I_T\}$ form a DP of $[0, 1]$.

We call this representation *semilinear* since on each of the elements of a DP it can be viewed as a projection of the log-periodogram on a linear approximation space. As the adaptively chosen DP will generally not be composed of homogeneous segments (of equal length), the overall representation is nonlinear and akin the result of a specific kernel estimator using a local smoothing parameter (Haerdle et al. (1998)). It provides

spatially adaptive estimators which share a *one-to-one* relation between the data and the coefficient domain; a useful property to construct models over the trial specific log-spectra. Unfortunately, this property does not hold with any other orthogonal wavelet, due to their overlapping supports (Mallat (1998)). However, we shall add a powerful post-processing step to our HSTW estimation to remedy the effect of using blocky boxcar functions without losing the attractive property of semilinearity (see Section 3.4).

3 Proposed method

In our multiple curves setting, defined by the equation (2.3), our goal is foremost to estimate the specific curves h^s , conditionally on the z^s , i.e $\mathbb{E}[h^s | z^s]$, the average-curve h and the deviation from h . A naive 'model-free' approach consists in estimating independently each spectrum $h^s, s = 1, \dots, S$, (using the HTSW) and then in computing at each frequency the mean and the empirical variance over all S curves to estimate h and V . It can be seen by the top plot of Figure 3.1 that it yields quite unsatisfactory estimation of V . We claim that a 'model-based' approach, as described further down, will improve the estimation (as in the bottom plot of the same figure), and moreover will allow us to achieve our ultimate goal of prediction, that is the derivation of prediction intervals for each of the h^s .

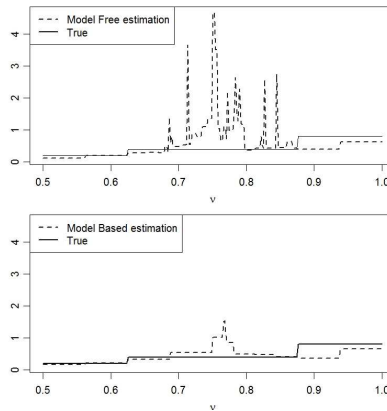


Figure 3.1: Estimation of the variance function V

3.1 Basic ingredients of our method

The HTSW estimator of each subject-specific spectrum can be represented as 0-1 dyadic tree (of zero/nonzero wavelet coefficients) associated to a partition. These trees (partitions) give an insight about the smoothness properties of the log-spectra since localized structures imply the existence of nonzero wavelet coefficients at fine scales. We use these 0-1 trees to improve the estimation of each subject-specific log-spectra h^s , $s = 1, \dots, S$ considering information from the S replicates, to construct a predictive model and particularly to apply the below-mentioned complexity constraint.

In Functional Data Analysis (FDA) the basic unit is a curve, thus, for the coherency of the interpretation it seems natural to require that the population and the subject specific curves share the same smoothness properties. Therefore, it was suggested to model both the random and the fixed effect by a unified approach (see [Guo \(2002\)](#)). It is often of interest to model FD to be curves belonging to a certain functional space in order to reduce their complexity by a development in an adapted and possibly sparse orthonormal basis (see again [Guo \(2002\)](#), or [Antoniadis and Sapatinas \(2007\)](#) who model both random and fixed effect curves to belong to the same functional space).

Driven by the same idea, we restrict our complexity by using a simpler approach based on sets of basis functions for modelling both effects. Since neither the fixed nor the random effects are directly observable, we proceed first to the estimation of the subject-specific spectra. Then we use this estimation to model the smoothness properties of h^s . We even consider the S subject-specific log-spectra estimators \hat{h}^s to be represented by the same partition which we call \hat{I}_1 , to be specified by equation (3.3). In other words, we search for an optimal set of boxcar functions $\{\varphi_I\}_{I \in \hat{I}_1}$ to model all the h^s simultaneously. Subsequently, h is modelled using the same set \hat{I}_1 to ensure that h has the same smoothness properties as the h^s which is our complexity constraint. Notice that at this step, we do not estimate separately the partitions for representing h and z^s . We just know that those are subsets of \hat{I}_1 and that, thanks to using semilinear representations in our approach, \hat{I}_1 is somehow given by the union of the unmodelled

tree representations of h and z^s .

Estimation of the variance components of the random effect is a difficult task (see also the top plot of Figure 3.1), as is nonparametric variance estimation in general. However we will benefit from a more accurate procedure which will be developed below. Similarly to the first step, this second estimation step will determine a common set of boxcar functions $\{\varphi_I\}_{I \in \hat{I}_V}$ for representing the z^s , $s = 1, \dots, S$. In addition, due to our complexity constraint, $\hat{I}_V \subseteq \hat{I}_1$, therefore, using the equivalent tree representation, \hat{I}_V is a connected rooted subtree of \hat{I}_1 . As a result, we suggest to estimate the random effect via

$$\hat{z}^s(\nu) = \sum_{I \in \hat{I}_V} \hat{c}_I^s \varphi_I(\nu), \quad s = 1, \dots, S. \quad (3.1)$$

This model is characterized by a one-to-one relation between scaling coefficients \hat{c}_I^s , $I \in \hat{I}_V$ and values of $\hat{z}^s(\nu)$ in the frequency domain: only one coefficient in the representation (3.1) is non zero for each $\nu \in [0, 1]$ due to the semilinearity. Similarly, the variance function $V(\nu)$ is modelled to be piecewise constant with values V_I estimated by the empirical variance of $\{\hat{c}_I^s, s = 1, \dots, S\}$ over each block I of \hat{I}_V .

The following section describes a simple approach to select these optimal sets of boxcar functions \hat{I}_1, \hat{I}_V .

3.2 Algorithm description

3.2.1 First step: estimation of h and $\mathbb{E}(h^s|z^s)$

From Y^s we compute the HTSW estimators \hat{h}^s of $\mathbb{E}[h^s|z^s]$ for each subject $s = 1, \dots, S$, since conditionally on the random effect, an estimator for h^s is given by

$$\hat{h}^s(\nu) = \hat{c}_{0,0}^s \varphi_{0,0}(\nu) + \sum_{I \in \hat{I}^s} \hat{d}_I^s \psi_I(\nu), \quad s = 1, \dots, S. \quad (3.2)$$

Here \hat{I}^s is the connected set of non zero wavelet coefficients to estimate the subject-specific log-spectra h^s , following the HTSW approach described in the previous section (i.e. with a specific set I_T for each specific s). The features common to h^s , $s = 1, \dots, S$, are approximately represented by the same sets of nonzero coefficients. Our investigations to provide a common partition \hat{I}_1 to represent them simultaneously suggest to

compute a trimmed union of the 0-1 trees defined by \hat{I}^s :

$$\hat{I}_1 = \text{Trimmed} \left(\cup \hat{I}^s \right), \quad (3.3)$$

where $\text{Trimmed}(\cdot)$ means that we discard the indices (j, k) from the set \hat{I}_1 if

$$\frac{1}{S} \sum_{s=1}^S \mathbb{1}_{\{\hat{d}_{j,k}^s \neq 0\}} < \beta,$$

that is if their empirical frequency is below a small value β . Once we get this, we project again each log-periodogram Y^s onto \hat{I}_1 which give coefficients $\hat{c}_I^s, I \in \hat{I}_1$. Note that if we take either the union or the intersection of rooted connected trees, the result is always a rooted connected tree, therefore, using the HTSW, the resulting estimator has a semilinear representation in terms of boxcar functions $\varphi_I(\nu)$ with coefficients \hat{c}_I^s :

$$\hat{h}^s(\nu) = \hat{c}_{0,0}^s \varphi_{0,0}(\nu) + \sum_{I \in \hat{I}_1} \hat{d}_I^s \psi_I(\nu) = \sum_{I \in \hat{I}_1} \hat{c}_I^s \varphi_I(\nu), \quad s = 1, \dots, S.$$

Finally we compute the common log-spectrum estimator by averaging the trial specific log-spectra at the frequency ν :

$$\hat{h}(\nu) = \frac{1}{S} \sum_{s=1}^S \hat{h}^s(\nu).$$

The Figures 7.4, 7.5, 7.6 motivate the construction of a trimmed union. In fact, if we form the common partition \hat{I}_1 just by the union of the sets \hat{I}^s , i.e., if $\beta = 0$, we improve the estimation of the localized structure but the impact of false positives, i.e., of erroneously active coefficients due to noise, is somehow amplified. These false positives are rare events, i.e., the concerned wavelet coefficients do not appear very frequently in the S trees. Thus, we classically choose $\beta = 5\%$ to compute the trimmed union. This yields the nicely estimated population spectrum as it can be seen in Figure 7.6.

3.2.2 Estimation of the between subject variance function V

We recall that $V(\nu) = \text{Var}[z^s(\nu)]$. We first choose to smooth the random effect z^s by smoothing the residuals $\hat{R}^s = Y^s - \hat{h} = z^s + \varepsilon'^s$, where $\varepsilon'^s = (h - \hat{h}) + \varepsilon^s$. Conditionally on z^s , i.e., treating the random effect as a conditionally fixed effect, this is a problem

similar to the one encountered for the estimation of h^s . Consequently, we use the same methodology as described in the section 3.2.1, i.e., the z^s are estimated over a common partition denoted as \hat{I}_2 (using hence the information from the S subjects). To ensure that the complexity constraint $\hat{I}_V \subseteq \hat{I}_1$ is satisfied, we compute \hat{I}_V as the intersection $\hat{I}_V = \hat{I}_2 \cap \hat{I}_1$.

Then, we project each log-periodogram Y^s on \hat{I}_V , defining \hat{c}_I^s , $I \in \hat{I}_V$. Using HTSW, all the estimators of z^s have a semilinear representation (see equation (3.1)). Finally, we compute the empirical variance of the scaling coefficients at a block I over the S trials:

$$\widehat{\text{Var}}(\hat{c}_I^s) = \frac{1}{S-1} \sum_{s=1}^S (\hat{c}_I^s - \bar{c}_I^s)^2, \quad (3.4)$$

where $\bar{c}_I^s = \frac{1}{S} \sum_{s=1}^S \hat{c}_I^s$. We remark that even if we were to know the true partition I_V of z^s , (3.4) is only an asymptotically unbiased estimator of V_I (which we show in appendix 7.4). Therefore, we prefer to correct for the finite-sample bias and ensure positivity of the resulting estimator as follows:

$$\hat{V}(\nu) = \sum_{I \in \hat{I}_V} \max \left(0, \widehat{\text{Var}}(\hat{c}_I^s) - \frac{\hat{\sigma}_\varepsilon^2}{|I|} \right) \varphi_I(\nu). \quad (3.5)$$

Estimation of σ_ε^2 Accurate estimation of σ_ε^2 is important in view of the bias correction of equation (3.5) and in order to construct confidence intervals of our predictors (cf equations (3.14, 3.17)). Following a standard approach (see, e.g., Vidakovic (1999)) we estimate σ_ε by computing the Median Absolute Deviation (MAD) over the thresholded wavelets coefficients at the finest wavelet scale $J-1$ since they are supposed to carry virtually only noise:

$$\hat{\sigma}_\varepsilon = 1.4826 \text{ Median} \left(\left| \underline{\hat{d}}_{J-1} - \text{Median}(\underline{\hat{d}}_{J-1}) \right| \right), \quad (3.6)$$

where $\underline{\hat{d}}_{J-1} = \left\{ \hat{d}_{J-1,k} \mid (J-1, k) \notin \hat{I}_1; k = 0, \dots, 2^{J-1} - 1 \right\}$.

3.3 Prediction

Here, we build a predictive model in the Haar coefficient domain to construct pointwise confidence intervals of level α in the frequency domain. We denote as Q_1 and Q_V

the cardinality of the sets \hat{I}_1 and \hat{I}_V . The estimators \hat{h}^s can be represented in the coefficient domain by a vector of non zero scaling coefficients $\hat{\underline{c}}^s$ of length Q_1 . Due to the sparsity of the wavelet representation, Q_1 is generally very small compared to T . Using the semi-linear representation, we can write the mixed models equation (2.3) in the coefficient domain as follows:

$$\hat{\underline{c}}^s = \underline{c}^s + \underline{e}^s, \quad (3.7)$$

$$= \underline{c} + \Phi_1 \underline{u}^s + \underline{e}^s, \quad (3.8)$$

where \underline{c}^s and \underline{c} are the vectors of length Q_1 of true scaling coefficients obtained by the projection of h^s and h on the set of orthonormal basis functions $\{\varphi_I \mid I \in \hat{I}_1\}$. \underline{u}^s is a vector a random coefficients of length Q_V obtained by the projection of z^s on $\{\varphi_I \mid I \in \hat{I}_V\}$. The variance-covariance matrix of \underline{u}^s is $\text{Var}[\underline{u}^s] = V_u = \text{diag}(V_I, I \in \hat{I}_V)$, with V_I introduced at the end of the section 3.1. Looking at the equation (7.1) it is reasonable to assume that, conditionally on z^s , $\hat{c}_{j,k}^s \sim \mathcal{N}\left(c_{j,k}^s, \frac{\sigma_{\epsilon}^2}{2^{(J-j)}}\right)$ for j sufficiently small. Therefore, we write that e^s follows the multivariate normal distribution $\mathcal{N}_{Q_1}(\underline{0}, V_e)$, where $\text{Var}[\underline{e}^s] = V_e = \text{diag}\left(\left\{\frac{\sigma_{\epsilon}^2}{|I|}, I \in \hat{I}_1\right\}\right)$. Finally, Φ_1 is a matrix of dimension $(Q_1 \times Q_V)$, which expands \underline{u}^s from \hat{I}_V to \hat{I}_1 , i.e., it makes the correspondence between one element of the vector $\hat{\underline{u}}^s$ and at least one element of the vector \underline{c}^s (since $\hat{I}_V \subseteq \hat{I}_1$).

We first denote by \underline{r}^s the vector of non zero scaling coefficients of the subject-specific (theoretical or unobserved) residuals $R^s = Y^s - h$ in the basis \hat{I}_1 and by $\tilde{\underline{r}}^s$ and $\tilde{\underline{u}}^s$, the Best Linear Unbiased Predictors (BLUPs) of r^s and $\Phi_1 u^s$ (Carroll et al. (2003)). These BLUPs are used to model the variance of the trial-specific spectra and of the population spectrum in the sequel. We first compute the BLUPs from the joint normal distribution of $(\underline{r}^s, \Phi_1 \underline{u}^s)$:

$$\begin{pmatrix} \Phi_1 \underline{u}^s \\ \underline{r}^s \end{pmatrix} \sim N \left[\begin{pmatrix} \underline{0} \\ \underline{0} \end{pmatrix}, \begin{pmatrix} \Phi_1 V_u \Phi_1' & \Phi_1 V_u \Phi_1' \\ \Phi_1 V_u \Phi_1' & \Phi_1 V_u \Phi_1' + V_e \end{pmatrix} \right]. \quad (3.9)$$

We can compute the vector predictors $\tilde{\underline{u}}^s$ and $\tilde{\underline{r}}^s$ as follows:

$$\mathbb{E}[\Phi_1 u^s \mid r^s] = \tilde{\underline{u}}^s = \Phi_1 V_u \Phi_1' [\Phi_1 V_u \Phi_1' + V_e]^{-1} \underline{r}^s = B \underline{r}^s. \quad (3.10)$$

$$\mathbb{E}[r^s | \Phi_1 u^s] = \tilde{r}^s = \Phi_1 V_u \Phi_1' [\Phi_1 V_u \Phi_1']^{-1} \Phi_1 \underline{u}^s = \Phi_1 \underline{u}^s. \quad (3.11)$$

From the observation that the residuals in the coefficient domain $\underline{r}^s = \hat{\underline{c}}^s - \underline{c}$ could be predicted by: $\tilde{r}^s = \hat{\underline{c}}^s - \tilde{\underline{c}}$, we can write $\tilde{\underline{c}}^s$ in terms of observable and predicted quantities:

$$\tilde{\underline{c}}^s = \hat{\underline{c}}^s - \tilde{r}^s + \tilde{\underline{u}}^s. \quad (3.12)$$

Pointwise confidence intervals for subject specific spectra We compute the variance of $\tilde{\underline{c}}^s$ from the conditional joint distribution of the best linear unbiased predictors, $(\tilde{\underline{u}}^s, \tilde{r}^s) | \underline{u}^s$, see the appendix section 7.5:

$$V_{c^s} = \mathbb{V}\text{ar}[\tilde{\underline{c}}^s | \underline{u}^s] = V_e + B V_e B' + 2B V_e, \quad (3.13)$$

where $B = \Phi_1 V_u \Phi_1' [\Phi_1 V_u \Phi_1' + V_e]^{-1}$. Once we get the expression of its variance, function of V_e and V_u , we plug-in empirical estimates \hat{V}_e and \hat{V}_u . Let $\hat{V}_{c^s} = \hat{V}_e + \hat{B} \hat{V}_e \hat{B}' + 2\hat{B} \hat{V}_e$, $\hat{B} = \Phi_1 \hat{V}_u \Phi_1' [\Phi_1 \hat{V}_u \Phi_1' + \hat{V}_e]^{-1}$, and $\hat{V}_e = \text{diag}\left(\left\{\frac{\hat{\sigma}_e^2}{|I|}, I \in \hat{I}_1\right\}\right)$. Pointwise confidence intervals of level α for h^s are given by plug-in empirical estimates for the variance. Since each block of frequency is modelled independently, we just take the diagonal of the matrix \hat{V}_{c^s} .

$$CI_{h^s, \alpha}(\nu_l) = \hat{h}(\nu_l) \pm \Phi_2 \left(z_{1-\alpha/2} \sqrt{\text{diag}(\hat{V}_{c^s})} \right) (\nu_l), \quad \nu_l = \frac{l}{T}, l = 0, \dots, T-1, \quad (3.14)$$

where Φ_2 is a matrix of dimension $(T \times Q_1)$ the columns of which are scaling vectors corresponding to \hat{I}_1 , i.e., it makes the correspondence between the coefficient and the frequency domain.

Pointwise confidence interval for the population spectrum We compute the variance of $\tilde{\underline{c}}$ from the unconditional joint distribution of the predictors $(\tilde{\underline{u}}^s, \tilde{r}^s)$, see the appendix section 7.6:

$$V_c = \mathbb{V}\text{ar}[\tilde{\underline{c}}] = \frac{1}{S} \mathbb{V}\text{ar}[\tilde{\underline{c}}^s], \quad (3.15)$$

$$= \frac{1}{S} [V_e + B V_e B' + 2B V_e + B \Phi_1 V_u \Phi_1' B']. \quad (3.16)$$

Plugging in the empirical estimates of V_e and V_u yields \hat{V}_c and the following confidence intervals:

$$CI_{h,\alpha}(\nu_l) = \hat{h}(\nu_l) \pm \Phi_2 \left(z_{1-\alpha/2} \sqrt{\text{diag}(\hat{V}_c)} \right) (\nu_l), \quad \nu_l = \frac{l}{T}, l = 0, \dots, T-1. \quad (3.17)$$

3.4 Postprocessing of the estimators

This methodology based on the HTSW yields blocky estimates and is neither visually appealing nor of small mean-squared error when estimating smooth functions. [Donoho \(1993\)](#) proposed an adapted methodology for smoothing out a piecewise constant fit subordinate to an RDP. It consists in finding a polynomial that matches the local averages given by our blocky estimator. [Donoho \(1993\)](#) proved that average-interpolation converges to a continuum limit and has some appealing estimation properties over classical Haar estimation. The *average interpolation* algorithm consists in two steps at each resolution scale j :

1. (average interpolation): at the location k , find a polynomial $\pi_{j,k}$ of even degree $D = 2L$ of each interval $I_{j,k}$ which generates the same averages in the neighborhood $(\hat{c}_{j,k'}, k' = k - L, \dots, k + L)$:

$$Ave_{j,k'} \pi_{j,k} = \hat{c}_{j,k'}, \quad -L \leq h \leq L,$$

2. (average imputation): define the mock averages at the next finer scale as averages of the AI polynomial. On the two halves of the interval we get:

$$\check{c}(j+1, 2k+h) = Ave_{j+1, 2k+h} \pi_{j,k}, \quad h = 0, 1.$$

In order to deal with spatially inhomogeneous partition, for example \hat{I}_1 , the *average interpolation* algorithm is used as follows:

From coarse scale to fine scale:

- **for** $j = j_1, \dots, J$ **do**,
 - predict the local average at the scale $j+1$ by $\check{c}_{j+1, 2k}$,
 - **if** $\hat{c}_{j+1, 2k} \notin \hat{I}_1$ **then** replace by $\check{c}_{j+1, 2k}$ **endif**.
- **enddo**

where the coarsest scale j_1 is such that $2L+1 < 2^{j_1}$.

4 Simulations

In this section we assess the finite sample performances of our methodology. We first present the way we generate our data because it is important that we actually generate the *time series* directly rather than the log-spectra. One approach consists in using the discretized Cramer's representation ([Brockwell and Davis \(1991\)](#)):

$$X_t^s = \frac{1}{\sqrt{2M}} \sum_{l=-(M-1)}^M U^s(\nu_l) \exp(i2\pi\nu_l t) \xi_l^s, \quad t = 1, \dots, T,$$

where ξ_l^s is a complex valued random variable whose real and imaginary parts are independent and such that $\xi_\ell = \xi_{-\ell}^*$, where $*$ denote the complex conjugate. U^s is the trial-specific transfer function:

$$U^s(\nu) = U(\nu) \sqrt{\exp[z^s(\nu)]},$$

where $U(\nu)$ is the transfer function of an $ARMA(p, q)$ process with AR parameters (ϕ_1, \dots, ϕ_p) and MA parameters $(\theta_1, \dots, \theta_q)$ and such that $U(\nu) = U^*(-\nu)$. A handy result is to express $U(\nu)$ as follows:

$$U(\nu) = \frac{1 + \theta_1 \exp(-i2\pi\nu) + \dots + \theta_q \exp(-i2\pi q\nu)}{1 - \phi_1 \exp(-i2\pi\nu) - \dots - \phi_p \exp(-i2\pi p\nu)}.$$

$z^s(\nu)$ is a "subject-specific random effect" which we shall express as a finite linear combination of boxcar functions:

$$z^s(\nu) = \sum_{I \in I_V} u_I^s \varphi_I(\nu), \quad (4.1)$$

where the coefficients $\{u_I^s\}_{I \in I_V} \sim \mathcal{N}(0, D)$ and D is a diagonal covariance matrix: $D = \text{diag}(\{V_I | I \in I_V\})$. The subject-specific log-spectrum is defined to be:

$$h^s(\nu) = \log |U^s(\nu)|^2, \quad (4.2)$$

$$= h(\nu) + z^s(\nu). \quad (4.3)$$

For the simulation study, we will consider the two following spectra:

- an AR(1) with parameter: $\phi_1 = 0.5$,

- an ARMA(2,2) with parameters: $\phi_1 = -0.2$, $\phi_2 = -0.9$, $\theta_1 = 0$, $\theta_2 = 1$. The log-spectrum of this ARMA(2,2) is known to have local features particularly difficult to estimate (Fryzlewicz et al. (2008)).

We define two random effect functions (see equation (4.1):

- *rand1*: few large blocks which have all the same dyadic lengths $\{\varphi_I | I \in I_V\} = \{[(j-1)/8, j/8) | j = 1, \dots, 8\}$
and $D = \text{diag}(0.2, 0.2, 0.4, 0.4, 0.4, 0.4, 0.8, 0.8)$,
- *rand2*: blocks which have not all dyadic lengths
 $\{\varphi_I | I \in I_V\} = \{[1, 1/16), [1/16, 3/16), [3/16, 7/16), [7/16, 9/16 - \iota), [9/16 - \iota, 1)\}$
and $D = \text{diag}(0.2, 0.3, 0.5, 0.6, 0.8)$, where ι is chosen such that the split occurs at a non dyadic point.

In the Figure 4.1 we can see examples of computed log-periodograms from the generated time series following this methodology. Despite its apparent simplicity such data generating process yields sufficiently complex time series (or log-spectra) to mimic some real data sets.

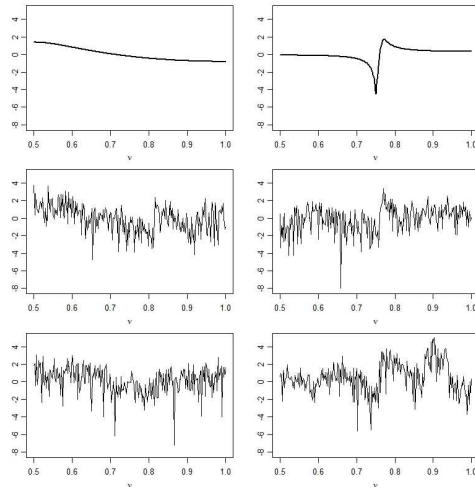


Figure 4.1: True spectra and associated log-periodograms (left: AR(1) spectrum; right: ARMA(2,2) spectrum)

For the simulation study we generate time series of length $T = 512$ observed on $S = 20$ different subjects. The level β of the trimmed union is set to 5% for both the estimation of \hat{I}_1 and of \hat{I}_2 (see section 3.2.1, 3.2.2). We compute the ISE of an estimator, for e.g., \hat{h} , on the interval $[0, 1]$, for the Monte Carlo replication m as follows :

$$ISE_{(m)}(\hat{h}) = \frac{1}{T} \sum_{l=0}^{T-1} \left(\hat{h}_{(m)}(\nu_l) - h(\nu_l) \right)^2, \quad (4.4)$$

where the number of Monte Carlo replications is $m = 1, \dots, M = 100$ and the MISE of \hat{h} is just: $MISE(\hat{h}) = \frac{1}{M} \sum_{m=1}^M ISE_{(m)}(\hat{h})$.

4.1 Estimation of h and $\mathbb{E}(h^s | z^s)$

We use the random effect function *rand1* to generate the time series. A paired Student's t-test is used to compare the MISE of our methodology to a model free approach (as described in the first paragraph of section 3). The estimation is done using HTSW without average interpolation refinement. For the ARMA(2,2), we observe on an empirical significance level below $p = 0.0001$ that $MISE(\hat{h})$ and $MISE(\hat{h}^s)$ are respectively about 43% and 55% lower with our method compared to the model free. For the AR(1), having a smooth spectrum, our method still significantly performs better for the estimation of the h^s (32% lower MISE, $p < 0.0001$), but there is no significant difference in the estimation of h . This experiment shows that our methodology provides good estimations of h and h^s , and, in particular, it succeeds in using the information from all the subjects to improve the estimation of the h^s , $s = 1, \dots, S$.

4.2 Variance function estimation

In order to explore the capabilities of our method for estimating the variance function, we use the random effect *rand1*. It allows us to easily set an estimation of the variance function over the true partition I_V , i.e., we discard the model-bias from our estimations. This yields an objective to attain in terms of ISE for our estimations performed over the partitions \hat{I}_1 , \hat{I}_2 and \hat{I}_V .

We compute the ISE for the variance function estimation similarly as we did for the population spectrum (see equation (4.4)). Boxplot representation of the results

can be seen in the Figure 4.2. The left plot of the Figure 4.2 presents the results for the ARMA(2,2) spectrum. In this case, the partition to represent h is quite more complex than the one to represent z^s , in other words, it is mainly h which requires the finest elements of the partition \hat{I}_1 . Therefore, our two steps methodology will adapt to the lower complexity of the random effect and provides more accurate estimation of the variance components of the random effect on a coarse partition $\hat{I}_V \subset \hat{I}_1$. On the opposite, for the AR spectrum (in the right plot of the same figure), we do not observe any improvements in the estimation over the partitions \hat{I}_V compared to \hat{I}_1 because the complexity of the random effect dominates that of the fixed effect. Compared to the estimation performed over the true partition I_V , our method performs nearly as well. This shows its ability to select the best set of boxcar functions to represent the random effect.

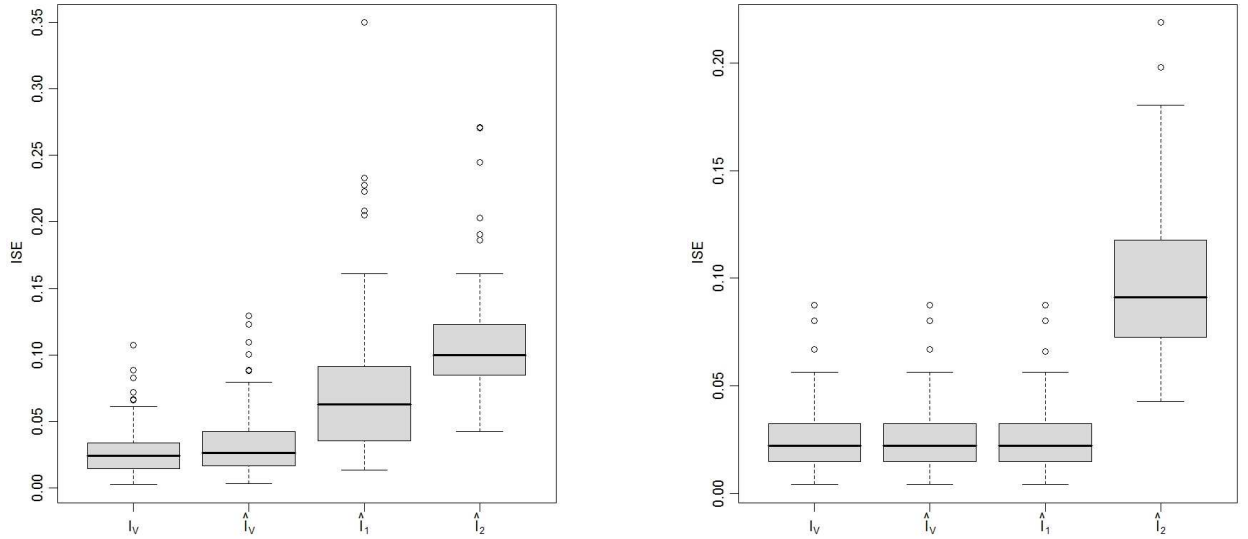


Figure 4.2: ISE when estimating the variance function over different partitions (left: ARMA spectrum; right: AR spectrum)

4.3 Confidence intervals

Here, we present some results about the coverage of our confidence intervals of nominal level 95% given by the equations (3.14, 3.17). Those confidence intervals yield blocky

upper and lower bounds in the frequency domain. Therefore, we apply the average interpolation schemes of section 3.4 with $D = 2$ to smooth them (Figure 4.3). Then we compute the coverage. We consider here the more realistic random effect given by the function *rand2*, the other parameters being exactly the same as previously set. For the AR(1) model, the coverages we obtain are very satisfying since they are about 92% and 95.2% for population and subject specific spectra, respectively. For the ARMA model, despite the localized structure, the results are still satisfying with coverages for the population and the subject specific spectra about 86% and 92%.

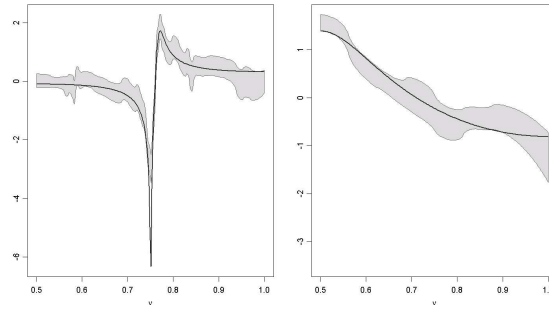


Figure 4.3: True spectrum (solide line) and 95% pointwise confidence intervals (shaded area) (left: ARMA(2,2); right: AR(1))

5 Analysis of EEG time series

This section proposes an application of our methodology to a real data analysis problem. We will show that, even if our modelling approach was made as simple as possible, it is able to deal with panels of complex time series such as encountered in the field of biomedical signals.

The data set consists of electroencephalograms recorded from $S = 8$ participants in a discrete hand movement experiment conducted at the laboratory of Professor Jerome Sanes at Brown University. This group of participants was made to be as homogeneous as possible - right handed male students in the with ages in the 18 – 30 range. The participants were instructed to move a joystick from the center-to-right (or center-to-left) when they saw a cursor flash on the right (or left) side of the screen. The ultimate

goal of the experiment was to use single trials time series to classify hand movements. The EEG was recorded over 64 channels and for our purposes we will study only the recording at the *C3* channel which roughly corresponds to the motor cortex. Each time series has length $T = 256$ and was collected over a period of 1 second (500 milliseconds prior to cursor flash until 500 milliseconds post stimulus presentation). The EEG signals were sampled at 200 Hertz. The 10-point low pass Butterworth filter was applied with a stop-frequency set to 40 Hertz.

Here, we shall (i.) estimate the common group log-spectrum corresponding to the left stimulus and (ii.) estimate the between-participant variation in the log-spectrum. The EEG and the corresponding bias-corrected raw log periodograms for Subjects 1,2,3 are shown in Figure 5.1. The estimate of the population log-spectrum in the top-left plot of the Figure 5.2 suggests some broad-band activity in the low beta range of 12-15 Hertz which is known to be involved in most cognitive tasks. The estimate of the between-participant variation is plotted in the top-right plot of the Figure 5.2 which suggests that variation remains approximately constant and stable across frequency. The Figure 5.2 establishes the nice results of our Haar-based methodology using the average interpolation refinement compared to a model free approach using smooth Daubechies (with four vanishing moments, see (Mallat (1998))) within tree-structured wavelet estimation. In addition, the second step of our methodology to reduce the complexity of \hat{I}_1 for estimating the variance function was helpful to provide an accurate estimation of V . In fact, the number of blocks in \hat{I}_V is nearly five times smaller than the number of blocks in \hat{I}_1 .

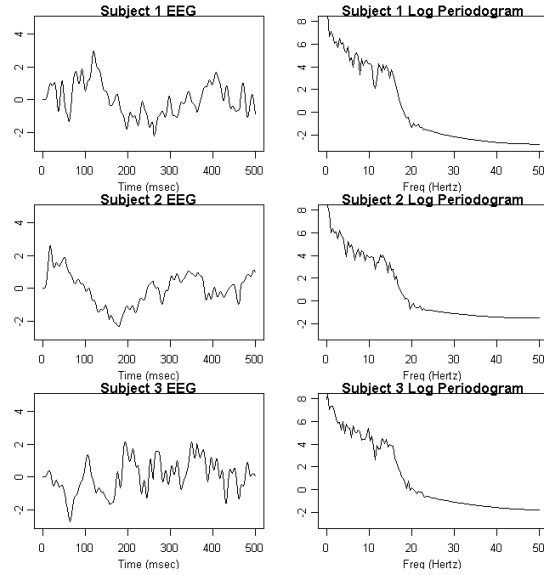


Figure 5.1: EEG Time Series and Log Periodograms for Subjects 1,2,3.

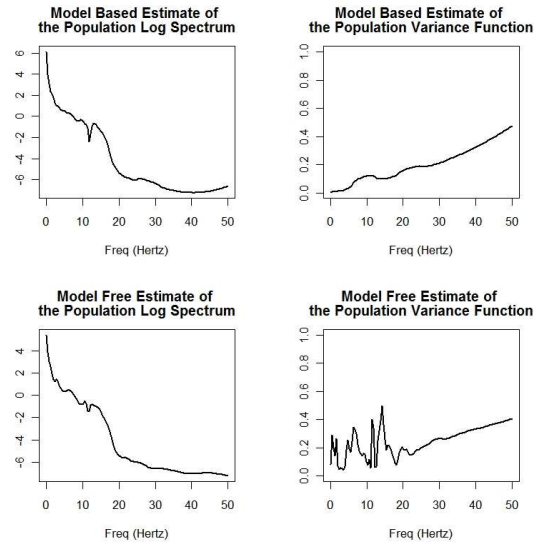


Figure 5.2: Estimation of Population log-spectrum and variance function (top: model-based; bottom: model free).

6 Conclusion and future directions

6.1 Conclusion

In this paper we have delivered a methodology to estimate non-parametrically, in the context of replicated stationary time series, their spectrum and the subject specific deviation from this ("population") spectrum. As we have embedded our approach into the appropriate framework of mixed effects modelling, we have been facing the problem of rigorously treating the random effect: using a specific curve estimation model based on the methodology of Tree Structured Wavelets, we come up with reliable estimators of the random effects variance. This is the key ingredient which allows us to also construct non-parametric (pointwise) predictors and reliable confidence intervals for the subject specific spectra as well as for the overall spectrum.

More specifically, instead of using a suboptimal (i.e. too variable) model-free approach of pointwise empirical variance estimation, we take advantage of a data-driven approach to restrict the complexity of the mixed effects to be common to both curves (fixed and, conditionally, random): using, in the first place comparatively simple box-car functions in the tree-structured wavelet representation of our denoised curves, we can reliably estimate the variance of the random effect on the segments of an "optimal" complete dyadic partition of the frequency domain. In fact, this optimal representation is determined solely from the data by a residual-based (conditional) smoothing of the random effect curves subject to the aforementioned complexity constraint. The subsequent simplicity of empirical variance estimation on this segmentation avoids needing to directly smooth squared data (known to be a harder non-parametric estimation problem). To finally overcome the blockiness of our estimates, we use the technique of average-interpolation which allows us to preserve our complexity reduced representations and all the mentioned advantages that come along with this. Hence we observe connections between our methodology of identifying blocks of frequencies chosen adaptively to the observed log-periodograms and MANOVA.

We considered the specific HAAR TSW methodology (plus post-smoothing by

average-interpolation) to benefit from the advantages and the simplicity of semilinear representations, particularly for the construction of confidence intervals. However, should one be only interested in estimating the trial-specific and common spectra, using more general TSW estimators based on orthogonal wavelets smoother than Haar (e.g. from the general Daubechies families of compactly supported wavelets), will prove useful as an improved peak-preserving denoising technique in its own right.

6.2 Future research

6.2.1 Time-varying spectra

When analyzing real time series data, the second order stationarity assumption is generally too strong. It is preferable to weaken this assumption and to only require that the process be locally stationary ([Dahlhaus \(1997\)](#)), i.e. that the variance-covariance structure of the process changes slowly over time. Consequently, the frequency analysis of those processes relies on the study of time-varying spectra. A quite appealing estimator for those curves depending now both on frequency and on time is the pre-periodogram, see [Neumann and von Sachs \(1997\)](#). The pre-periodogram is akin a very localized periodogram - it is defined as the Fourier transform of an empirical autocovariance estimator which depends only on one localized pair of lagged observations in time. Hence, this estimator has an extremely high variance, and adaptively smoothing it simultaneously over time and frequency is a very challenging task (fully non-linear wavelet thresholding, though theoretically appealing, proved to fail in finite sample situations). We expect that the methodology presented here will become very efficient in this context because it is supposed to improve on denoising in the context of highly irregular noise structure superposed onto localized signal structured ("simultaneous control of false positives and false negatives").

6.2.2 Increasing the order of approximation accuracy within estimation schemes preserving the RDP property

This work is highly motivated by the advantages of estimation subordinate to an RDP. The key to realize this objective, in this paper, has been to start from HTSW and to

overcome the blockiness of the reconstruction by an average-interpolation step which preserves the RDP property. However, this approach does not benefit from higher order approximation properties for functions which have higher regularity. To overcome this limitation, in future research we intend to borrow strength from the construction in [Donoho et al. \(2000\)](#). This approach follows the same paradigm as ours, now by application of a refinement scheme which generalizes average interpolation away from originally blocky reconstructions. This allows to get smooth estimators with faster rates of "approximation" for smoother functions (and hence an even sparser representation of an inhomogeneous function in regions without any localized signal structure).

References

- Antoniadis, A. and Sapatinas, J. (2007), "Estimation and Inference in Functional Mixed-Effects Models," *Computational Statistics and Data Analysis*, 51, 4793–4813.
- Autin, F. (2008), "On the Performances of a New Thresholding Procedure using Tree Structure," *Electronic Journal of Statistics*, 2, 412–431.
- Baraniuk, R. (1999), "Optimal Tree Approximation using Wavelets," in *Wavelet Applications in Signal Processing*, eds. Aldroubi, A. J. and Unser, M., SPIE, vol. 7, pp. 196–207.
- Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, Springer Series in Statistics.
- Bunea, F., Ombao, H., and Auguste, A. (2006), "Minimax Adaptive Spectral Estimation from an Ensemble of Signals," *IEEE Transactions on Signal Processing*, 54, 2865 – 2873.
- Bunea, F., Wegkamp, M., and Tsybakov, A. (2007), "Aggregation for Gaussian Regression," *Annals of Statistics*, 35, 1674–1697.

- Cai, T. (1999), “Adaptive Wavelet Estimation: A Block Thresholding And Oracle Inequality Approach,” *IEEE Transactions on Signal Processing*, 27, 898–924.
- Caroll, R., Ruppert, D., and Wand, M. (2003), *Semiparametric Regression*, Cambridge Series in Statistical and Probabilistic Mathematics, UK: Cambridge University Press.
- Cohen, A., Dahmen, W., Daubechies, I., and DeVore, R. (2001), “Tree Approximation and Optimal Encoding,” *Applied and Computational Harmonic Analysis*, 11, 192–226.
- Dahlhaus, R. (1997), “Fitting Time Series Models to Nonstationary processes,” *Annals of Statistics*, 25, 1–37.
- Donoho, D. (1993), “Smooth Wavelet Decompositions with Blocky Coefficient Kernels,” in *In Recent Advances in Wavelet Analysis*, Academic Press, pp. 259–308.
- (1994), “Ideal Spatial Adaptation by Wavelet Shrinkage,” *Annals of Statistics*, 81, 425–455.
- Donoho, D., Dyn, N., Levin, D., and Yu, T. P.-Y. (2000), “Smooth Multiwavelet Duals of Alpert Bases By Moment-Interpolating Refinement,” *Applied and Computational Harmonic Analysis*, 9, 166–203.
- Engel, J. (1994), “A Simple Wavelet Approach to Nonparametric Regression from Recursive Partitioning Schemes,” *Journal of Multivariate Analysis*, 49, 242–254.
- Fryzlewicz, P., Nason, G. P., and von Sachs, R. (2008), “A Wavelet-Fisz Approach to Spectrum Estimation,” *Journal of Time Series Analysis*, 29, 868–880.
- Gao, H.-Y. (1997), “Choice of thresholds for wavelet shrinkage estimate of the spectrum,” *Journal of Time Series Analysis*, 18, 231–251.
- Guo, W. (2002), “Functional mixed effects models,” *Biometrics*, 58, 121–128.

- Haerdle, W., Kerkycharian, G., D., P., and Tsybakov, A. (1998), *Wavelets, Approximation, and Statistical applications*, vol. 129 of *Lectures notes in Statistics*, Hamburg: Verlag.
- Jansen, M. (2001), *Noise Reduction by Wavelet Thresholding*, Springer.
- Lee, T. (2002), “Tree Based Wavelet Regression for Correlated Data Using the Minimum Description Length Principle,” *Australian and New Zealand Journal of Statistics*, 44, 23–39.
- Mallat, S. (1998), *A Wavelet Tour of Signal Processing*, Academic Press.
- Morris, J. and Carroll, R. (2006), “Wavelet-based functional mixed models,” *Journal Of The Royal Statistical Society Series B*, 68, 179–199.
- Morris, J., Vannucci, M., Brown, P., and Carroll, R. (2003), “Wavelet-Based Nonparametric Modeling of Hierarchical Functions in Colon Carcinogenesis,” *Journal of the American Statistical Association*, 98, 573–597.
- Moulin, P. (1994), “Wavelet thresholding techniques for power spectrum estimation,” *IEEE Transactions on Signal Processing*, 42, 3126–3136.
- Neumann, M. and von Sachs, R. (1997), “Wavelet Thresholding in Anisotropic Function Classes and Application to Adaptive Estimation of Evolutionary Spectra,” *Annals of Statistics*, 25, 38–76.
- Vidakovic, B. (1999), *Statistical Modelling by Wavelets*, New-York: Wiley Series in Probability and Statistics.
- Wahba, G. (1980), “Automatic Smoothing of the Log-Periodogram,” *Journal of the American Statistical Association*, 75, 122 – 132.

7 Appendix

7.1 Tree representation of wavelet estimators

The Figure 7.1 represents the wavelets coefficients in equation (2.4) with $j_0 = 0$ and $J = 3$. The two other figures represent the unstructured and structured set of nonzero wavelet coefficients after nonlinear and tree-structured thresholding, respectively.

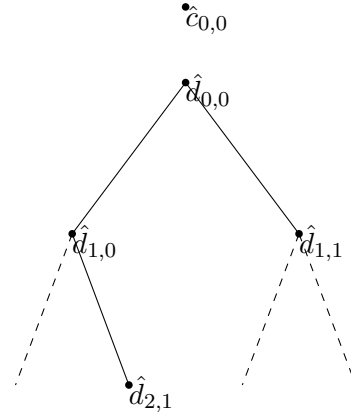
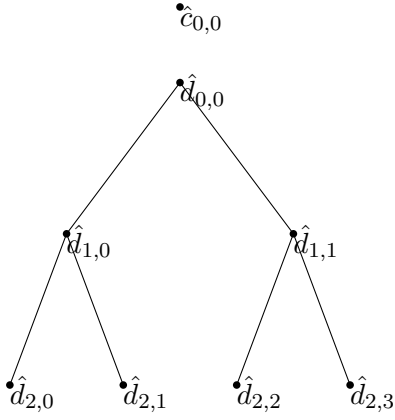


Figure 7.1: Wavelet coefficient tree (\mathcal{T})

Figure 7.2: Tree-structured estimation (I_T)

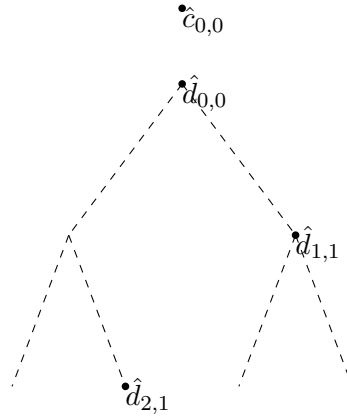


Figure 7.3: Nonlinear estimation (I_{NL})

7.2 Tree-structured algorithm

In the literature there exists plenty of thresholding methods. It is well-known that the most powerful ones reduce the number of false negatives and false positives using the

information among several wavelet coefficients rather than considering each coefficient independently. The algorithm considered in this paper (see [Engel \(1994\)](#)) takes into account the magnitude of a hierarchically structured subset of wavelet coefficients when comparing with a given threshold λ . This procedure always ensures that at each step of the algorithm the non-zero wavelet coefficients satisfies the hereditary constraint.

Although the theoretical performances of this algorithm has been not yet studied, there exists results about the optimality of several methods which forebode its performances. [Cohen et al. \(2001\)](#) proved that tree-structured wavelet schemes have nearly the same approximation power as nonlinear schemes. From a maxiset point of view, [Autin \(2008\)](#) showed that the estimation using tree-structured wavelets outperformed nonlinear methods (by the reduction of false negatives). Concerning the optimality of thresholding methods which use information among neighbored coefficients, we refer the interested reader to [Cai \(1999\)](#) or [Autin \(2008\)](#).

We denote \mathcal{T} the complete wavelet coefficient tree with $J - 1$ scales, by $C(j, k)$ the set of all children of the coefficient located at (j, k) and itself, i.e.,

$$C(j, k) = \mathcal{T} \cap \{(j, k), (j + 1, 2k), (j + 1, 2k + 1), \dots, \\ (J - 1, 2^{J-1-j}k), (j + J - 1, 2^{J-j-1}k + 1), \dots, (J - 1, 2^{J-j}k - 1)\},$$

and by $|C(j, k)|$ the cardinality of the $C(j, k)$. The tree-structured thresholding algorithm can be written as follows:

```

• for  $j = (J - 1), \dots, 0$  do,
  - for  $k = 0, \dots, 2^j - 1$  do,
    * if  $\sum_{(\mu, \kappa) \in C(j, k)} \hat{d}_{\mu, \kappa}^2 \leq |C(j, k)| \lambda$ ,
    * then prune the tree  $\mathcal{T}$ , i.e, set  $\mathcal{T} = \mathcal{T} \setminus C(j, k)$  endif,
  - enddo,
• enddo.
```

7.3 Estimation over a trimmed union

The following figures show the estimation of the h^s , $s = 1, \dots, S$ over different sets of wavelet coefficients and of the population log-spectrum, computed as an average over each frequency of the previously estimated h^s .

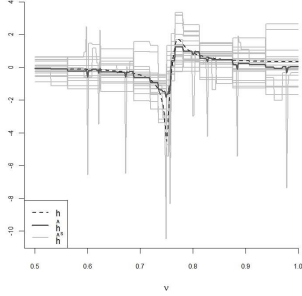


Figure 7.4: Estimation over \hat{I}^s .

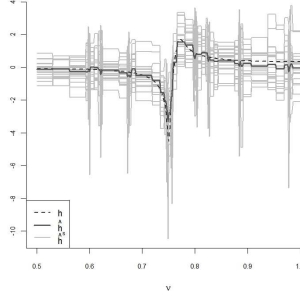


Figure 7.5: Estimation over $\cup \hat{I}^s$.

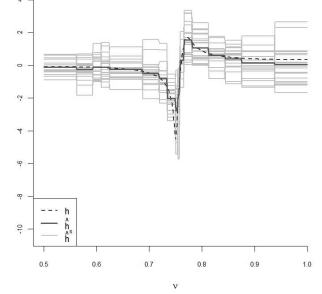


Figure 7.6: Estimation over $\text{Trimmed}(\cup \hat{I}^s)$.

7.4 Finite sample bias correction of the variance estimator

For the following calculations, we consider that the random effect is a finite linear combination of boxcar functions ($\{\varphi_I, I \in I_V\}$). We limit our investigations to study the behavior of the variance estimator over one block I of I_V , where we ignore the potential model-bias as the partition I_V is not observable in practice. First recall that an estimate of the scaling coefficients is given by:

$$\begin{aligned} \hat{c}_I^s &= \langle Y^s, \varphi_I \rangle, \\ &= \frac{1}{|I|} \sum_{l \in I} Y_l^s \varphi_I(\nu_l). \end{aligned} \quad (7.1)$$

Since, in our model, the subject specific deviation is modelled as piecewise constant over blocks, see equation (4.1), $\text{Cov}(z_l^s, z_{l'}^s) = \text{Var}[z_l^s] = V_I$, for all $l, l' \in I$. The variance of \hat{c}_I^s is:

$$\begin{aligned} \text{Var}(\hat{c}_I^s) &= \text{Var}\left(\frac{1}{|I|} \sum_{l \in I} Y_l^s \varphi_I(\nu_l)\right), \\ &= \frac{1}{|I|} \sigma_\varepsilon^2 + \frac{1}{|I|^2} \left(\sum_{l \in I} \text{Var}(z_l^s) + 2 \sum_{l, l'; l < l'} \text{Cov}(z_l^s, z_{l'}^s) \right), \\ &= \frac{1}{|I|} \sigma_\varepsilon^2 + V_I. \end{aligned} \quad (7.2)$$

The empirical variance estimator $\widehat{\text{Var}}(\hat{c}_I^s)$ is an unbiased estimator of $\text{Var}(\hat{c}_I^s)$, therefore:

$$\mathbb{E} \left[\widehat{\text{Var}}(\hat{c}_I^s) \right] = \frac{1}{|I|} \sigma_\varepsilon^2 + V_I.$$

Finally, to unbiasedly estimate V_I , we propose to use $\text{Var}[\hat{c}_I^s] - \frac{\sigma_\varepsilon^2}{|I|}$, where σ_ε^2 has to be estimated according to the section 3.2.2.

7.5 Details for the confidence intervals for $\mathbb{E}(h^s|z^s)$

From the joint distribution $(\Phi_1 \underline{u}^s, \underline{r}^s)$ and the expression of the predictors, see equations (3.10, 3.11) we compute the joint distribution of \tilde{u}^s and \tilde{r}^s conditionally on \underline{u}^s :

$$\begin{pmatrix} \tilde{u}^s \\ \tilde{r}^s \end{pmatrix} \mid \underline{u}^s \sim N \left[\begin{pmatrix} B\Phi_1 \underline{u}^s \\ \Phi_1 \underline{u}^s \end{pmatrix}, \begin{pmatrix} BV_e B' & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} \right]. \quad (7.3)$$

Using the equation (3.12) we can compute $V_{c^s} := \text{Var}[\tilde{c}^s | \underline{u}^s]$ as follows:

$$\begin{aligned} V_{c^s} &= \text{Var}[\tilde{c}^s - \tilde{r}^s + \tilde{u}^s | \underline{u}^s], \\ &= \underbrace{\text{Var}[\tilde{c}^s | \underline{u}^s]}_{=V_e} + \underbrace{\text{Var}[\tilde{r}^s | \underline{u}^s]}_{=0} + \underbrace{\text{Var}[\tilde{u}^s | \underline{u}^s]}_{BV_e B'} \\ &\quad - 2 \underbrace{\text{Cov}(\tilde{c}^s, \tilde{r}^s | \underline{u}^s)}_{=0} + 2 \underbrace{\text{Cov}(\tilde{c}^s, \tilde{u}^s | \underline{u}^s)}_{=BV_e} - 2 \underbrace{\text{Cov}(\tilde{r}^s, \tilde{u}^s | \underline{u}^s)}_{=0}, \\ &= V_e + BV_e B' + 2BV_e. \end{aligned} \quad (7.4)$$

7.6 Details for the confidence intervals for h

We consider the unconditional joint normal distribution in the coefficient domain:

$$\begin{pmatrix} \tilde{u}^s \\ \tilde{r}^s \end{pmatrix} \sim N \left[\begin{pmatrix} \underline{0} \\ \underline{0} \end{pmatrix}, \begin{pmatrix} B\Phi_1 V_u \Phi_1' B' + BV_e B' & B\Phi_1 V_u \Phi_1' \\ B\Phi_1 V_u \Phi_1' & \Phi_1 V_u \Phi_1' \end{pmatrix} \right]. \quad (7.5)$$

$$\begin{aligned} \text{Var}[\tilde{c}] &= \underbrace{\text{Var}[\hat{c}^s]}_{\Phi_1 V_u \Phi_1' + V_e} + \underbrace{\text{Var}[\tilde{r}^s]}_{\Phi_1 V_u \Phi_1'} + \underbrace{\text{Var}[\tilde{u}^s]}_{B\Phi_1 V_u \Phi_1' B' + BV_e B'} \\ &\quad - 2 \underbrace{\text{Cov}[\hat{c}^s, \tilde{r}^s]}_{\Phi_1 V_u \Phi_1'} + 2 \underbrace{\text{Cov}[\hat{c}^s, \tilde{u}^s]}_{B\Phi_1 V_u \Phi_1' + BV_e} - 2 \underbrace{\text{Cov}[\tilde{r}^s, \tilde{u}^s]}_{B\Phi_1 V_u \Phi_1'}, \\ &= V_e + B\Phi_1 V_u \Phi_1' B' + BV_e B' + 2BV_e. \end{aligned} \quad (7.6)$$