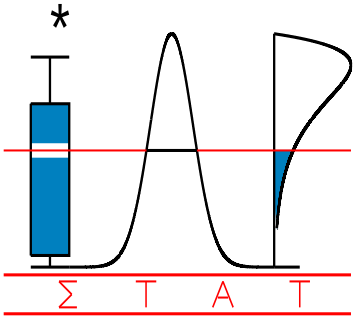
**T E C H N I C A L**

**R E P O R T**

**0621**

# ESTIMATION AND INFERENCE IN FUNCTIONAL MIXED-EFFECTS MODELS

ANTONIADIS A., and T. SAPATINAS




**I A P S T A T I S T I C S**

**N E T W O R K**

**INTERUNIVERSITY ATTRACTION POLE**

http://www.stat.ucl.ac.be/IAP

# Estimation and Inference in Functional Mixed-Effects Models

*Anestis Antoniadis*,
Laboratoire IMAG-LMC, University Joseph Fourier,
BP 53, 38041 Grenoble Cedex 9, France.
Email: Anestis.Antoniadis@imag.fr
and
*Theofanis Sapatinas*,
Department of Mathematics and Statistics, University of Cyprus,
P.O. Box 20537, CY 1678 Nicosia, Cyprus.
Email: T.Sapatinas@ucy.ac.cy

**Abstract**

Functional mixed-effects models are very useful in analyzing functional data. We consider a general functional mixed-effects model that inherits the flexibility of linear mixed-effects models in handling complex designs and correlation structures. A wavelet decomposition approach is used to model both fixed-effects and random-effects in the same functional space, meaning that the population-average curve and the subject-specific curves have the same smoothness property. A linear mixed-effects representation is then obtained that is used for estimation and inference in the general functional mixed-effects model. Adapting recent methodologies in linear mixed-effects and nonparametric regression models, we provide hypothesis testing procedures for both fixed-effects (testing whether certain fixed-effects functional components or contrasts are equal to zero) and random-effects (testing whether the random-effects functional components are equal to zero). Using classical linear mixed-effects estimation techniques, the linear mixed-effects representation is also used to obtain wavelet-based estimates for both fixed-effects and random-effects in the general functional mixed-effects model. We illustrate the usefulness of the proposed estimation and hypothesis testing procedures by means of two real-life datasets arising from physiology and proteomics.

KEYWORDS: Functional Data; Linear Mixed-Effects Models; Nonparametric Hypothesis Testing; Smoothing Spline Estimation; Wavelet Estimation.

# 1 Introduction

Nowadays, a form of data, called *functional data* (see, e.g., Ramsay & Silverman, 1997), are collected in many fields of research. Such data are encountered, for example, when units are observed over time or when, although a whole function itself is not observed, a sufficiently large number of evaluations over individual is available – a common feature of modern recording equipments. Sophisticated on-line sensing and monitoring equipments are now routinely used in research in medicine, seismology, meteorology, physiology, and many other fields. Since functional data arise as *curves* it is therefore natural to use a curve as the basic unit in functional data analysis. Functional data tend to involve a large number of repeated measurements per subject, and these measurements are usually recorded at the same, often equally spaced, time points for all subjects, and with the same high sampling rate. The aims of functional data analysis is usually of exploratory nature – to represent and display data in order to highlight interesting characteristics, perhaps as input for further analysis (see, e.g., Chapter 1 in Ramsay & Silverman, 1997). However, there are other aims, among them are the following: estimation of individual (and functional of these) curves from noisy data, characterising homogeneity and patterns of variability among curves, and assessing the relationships of shapes of curves to covariates (see Rice, 2004).

It is therefore challenging to build models for functional data that are reasonably flexible, yet feasible to fit. *Linear mixed-effects* models provide a flexible likelihood framework to model such data parametrically (see Laird & Ware, 1982). In the corresponding analysis, however, the parametric assumption in linear mixed-effects models may not always be appropriate. Extensions of linear mixed-effects models by including nonparametric fixed-effects and parametric random-effects have been considered by many researchers (see, e.g., Wang, 1998; Guo, 2002a; Durban *et al.*, 2005). The limitation of these approaches, however, is that they have used parametric random-effects, which may not be adequate to handle flexible subject-specific deviations. Various approaches to include, directly or indirectly, nonparametric methods for serial correlation in functional data analysis models have also been proposed (see, e.g., Rice & Silverman, 1991; Rice & Wu, 2001).

Although much work has been done on the *estimation* in various functional mixed-effects models, only limited work has been done regarding *inference* in these or more complex models. Both estimation and inference in a general functional mixed-effects model were recently considered by Guo (2002b). The idea behind his formulation is to model the fixed-effects as a *single* realisation of a partially diffuse integrated Wiener process, while the random-effects are modelled as *random* realisations from the same partially integrated Wiener process with proper variances. Then, an estimation procedure can be developed by taking advantage of the connection between cubic smoothing splines (at the design points) and linear mixed-effects models, a fact that originally pointed out by Speed (1991). A likelihood-ratio (LR) test was also proposed by Guo (2002b) for

testing the fixed-effects using the connection between cubic smoothing splines (at the design points) and linear mixed-effects models, and the non-standard asymptotic theory for LR tests developed by Self & Liang (1987). However, inference for the random-effects was not considered in Guo (2002b).

Although cubic smoothing splines provide a continuum of models from a trend linear in time to treating time as a factor (obtained as the smoothing parameter tends to $\infty$ and 0 respectively), the corresponding modelling methodology of Guo (2002b) seems to have its own drawbacks. Formulating cubic spline smoothing as a mixed-effects model is simply a mathematical device; the suggested logical distinction between the fixed linear trend and the random smooth variation about it is artificial, so one should not freely adopt random-effects methodology in this context (see Green, 1999). Indeed, it is well-known while the space of curves modelling the fixed-effects and the random-effects is smooth in some sense, they are drawn from a Gaussian prior process whose individual realizations (prior or posterior) do not have finite smoothness, with prior and posterior distributions entirely supported outside this space (see, e.g., Green & Silverman, 1994, pp. 53–54). More importantly, as emphasized in subsequent sections, the non-standard asymptotic theory for LR or restricted likelihood ratio (RLR) tests, which is used when the parameter under the null hypothesis is on the boundary of the parameter space (see Self & Liang, 1987), cannot be blindly applied for testing variance components in linear mixed-effects models; an approach adopted by Guo (2002b) for testing fixed-effects in a general functional mixed-effects models.

A general functional mixed-effects model, similar to the one studied by Guo (2002b), has also been recently studied by Morris & Carroll (2006). Their methodology is based on a fully Bayesian wavelet-based approach, yielding nonparametric estimates of both fixed-effects and random-effects, as well as the various between-curve and within-curve covariance matrices. Using the posterior samples for all model quantities, pointwise or joint Bayesian inference or prediction on the quantities of the model is discussed. However, formal frequentist functional hypothesis testing procedures for both fixed-effects and random-effects, which is the main focus of the present paper, is lacking from the analysis described in Morris & Carroll (2006).

In this paper, our aim is to study both *estimation* and *inference* in a general functional mixed-effects model, similar to the models studied by Guo (2002b) and Morris & Carroll (2006). Note that estimation and inference in nonparametric settings are entirely different problems since the optimal rates for smoothing parameters in nonparametric function estimation are different from the ones obtained in nonparametric hypothesis testing (see, e.g., Ingster & Suslina, 2003, Section, 1.4.4). The proposed estimation and testing procedures will be built upon an appropriate wavelet decomposition approach. Wavelet decompositions allow one to characterize different types of smoothness conditions assumed on the response function by means of its wavelet coefficients for a wider range of function classes than the ones obtained by, e.g., their Fourier or smoothing splines counterparts. In other words, these later methods may not be suitable for the spatially

heterogeneous data encountered in our motivating examples presented in Section 2.

The paper is organised as follows. In Section 2, we introduce two functional data sets, arising from physiology and proteomics, that have motivated our methodological thinking. In Section 3, we first provide a formulation for a general functional mixed-effects model. We then briefly recall some relevant facts about the wavelet series expansion and the discrete wavelet transform that we need further. A wavelet decomposition approach is then developed to model both fixed-effects and random-effects in the same functional space. Finally, a linear mixed-effects representation, that is subsequently used for estimation and inference in the general functional mixed-effects model, is also derived. In Section 4, adapting recent methodologies in linear mixed-effects and nonparametric regression models, we provide hypothesis testing procedures for both fixed-effects and random-effects in the general functional mixed-effects model. In Section 5, we illustrate the usefulness of the proposed testing procedures by applying them on the motivated examples described in Section 2, along with wavelet-based estimates for both fixed-effects and random-effects that are readily obtained by using classical linear mixed-effects estimation techniques. Some concluding remarks and hints for possible extensions of the proposed functional mixed-effects methodology are made in Section 6. Finally, in the Appendix, we provide outlines of the proofs of the theoretical results stated in earlier sections.

## 2 The Motivating Examples

### 2.1 Orthosis Data

Human movement data were acquired and computed at the Laboratoire Sport et Performance Motrice, Grenoble University, France (see Cahouët *et al.*, 2002). In this experiment, stepping-in-place was a relevant task to investigate how muscle redundancy could be appropriately used to cope with an external perturbation while complying with the mechanical requirements related either to balance control and/or minimum energy expenditure. For this purpose, 7 young male volunteers wore a spring-loaded orthosis of adjustable stiffness under 4 experimental conditions: a Control condition (without orthosis), an Orthosis condition (with the orthosis only), and two conditions (Spring1, Spring2) in which stepping in place was perturbed by fitting a spring-loaded orthosis onto the right knee joint. The experimental session included 10 trials of 20 seconds under each experimental condition for each subject. Data sampling started 5 seconds after the onset of stepping, and lasted for 10 seconds for each trial. So, anticipatory and joint movements induced by the initiation of the movement were not sampled. For each of the 7 subjects, 10 stepping-cycles of data were analyzed under each experimental condition. The resultant moment at the knee is derived by means of body segment kinematics recorded with a sampling frequency of 200 Hz. For

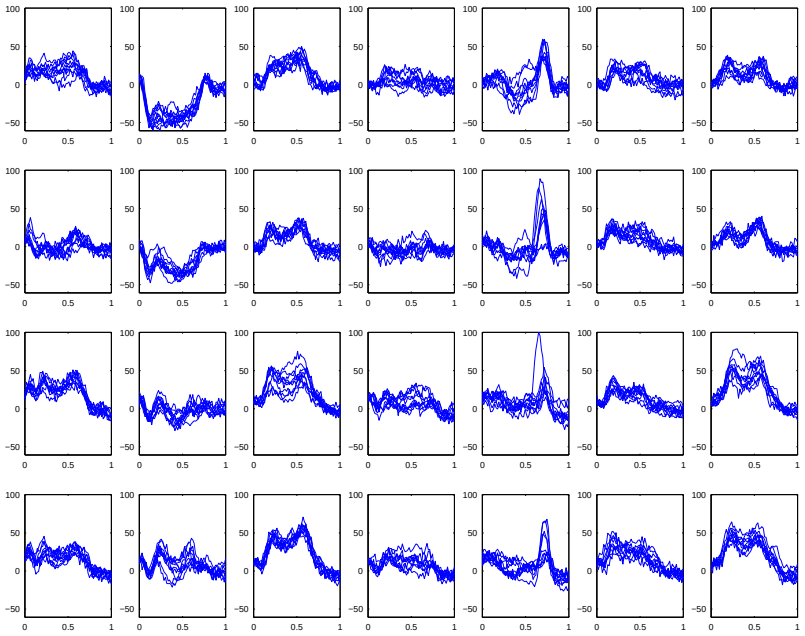


Figure 2.1: Orthosis dataset: The panels in rows correspond to 'Treatments', while the panels in columns correspond to 'Subjects' (there are 10 repeated measurements in each panel).

each stepping-in-place replication, the resultant moment was computed at 64 time points equally spaced and scaled so that a time interval corresponds to an individual gait cycle. The data set consists of 280 separate runs and involves the 7 subjects over 4 described experimental conditions, replicated 10 times for each subject. Figure 2.1 shows the available data set; typical moment plots over gait cycles. The inhomogeneous behaviour of the involved curves makes them natural candidates for the proposed functional mixed-effects methodology. One of the aim of the analysis is to understand how a subject can cope with the external perturbation, and we need to quantify the ways in which the individual mean cross-sectional functions differ over the various conditions.

## 2.2 Mass Spectrometry Proteomic Data

Mass spectrometry of proteins is a leading technology for the measurement and the detection of large biomolecules (e.g. proteins) from biological samples. This technology requires small amounts of biological material (tissue or blood samples) and it is often used for biomarker discovery, to identify proteins linked to disease status, response to therapy, or clinical prognosis. Mass spectrometry measures two properties of ion mixtures in the gas phase under a vacuum environment: the mass-to-charge ratio (m/z, in Daltons per unit charge) of ions in the mixture; and the number of ions present at different m/z values. The output is a mass spectrum, i.e., a curve where the x-axis

indicates the ratio of the weight of a specific molecule to its electrical charge and the y-axis is the signal intensity for the same molecule as a measure of the abundance of that molecule in the sample. Proteomic spectra are characterized by many spike peaks, each representing the ionized proteins or protein fragments (peptides) of a specific m/z value present in the sample. The heights of peaks and the m/z values of peaks are a fingerprint of the sample. The intrinsic structure of the measured spectra and the presence of many spike peaks is a motivating example for the proposed functional mixed-effects analysis which treats the entire spectrum as a single observation, rather than a closely spaced sequence of measurements. To illustrate the proposed methodology, we will use some proteomic data from a recent study described in Petricoin *et al.* (2002). The dataset comprises 162 cancer samples and 91 control cases. The expression profile contains measurements of 15,154 intensities on a grid of m/z ratios, but these intensities do not represent independent peptides for this instrument. The peptides yield "peaks" in the spectra, which cover a substantial number of m/z values. A typical SELDI-TOF profile might contain information on several hundred peptides. Preprocessed data are plotted in Figure 2.2. To simplify the presentation, thus avoiding many replicated curves being placed in the plot, Figure 2.2 shows only a representative subset of the preprocessed data curves for each group over a restricted m/z range. One of the aim of the analysis is to identify characteristic differences in protein expression between ovarian cancer patients and healthy patients using blood serum samples.

# 3 Functional Mixed-Effects Models

## 3.1 The General Setup

Suppose that $Y_{ij}$ $(i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, m_i)$ is the response of the $i$-th subject at point $t_{ij}$ (where $t$ is an index such as time or distance) and can be modelled as

$$Y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}(t_{ij}) + \mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t_{ij}) + \epsilon_{ij}, \tag{1}$$

where $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_p(t))^{\mathrm{T}}$ is a $p \times 1$ vector of *fixed* functions, $\boldsymbol{\alpha}^{(i)}(t) = (\alpha_1^{(i)}(t), \ldots, \alpha_q^{(i)}(t))^{\mathrm{T}}$ is a $q \times 1$ vector of stochastically independent *random* functions that are modelled as realisations of zero-mean Gaussian processes $\boldsymbol{a}(t) = (a_1(t), \ldots, a_q(t))^{\mathrm{T}}$ (a $q \times 1$ collection of such independent processes) with parametrically structured covariances modelled in the wavelet domain (see Section 3.4), $\mathbf{X}_{ij} = (X_{ij}[1], \ldots, X_{ij}[p])$ and $\mathbf{Z}_{ij} = (Z_{ij}[1], \ldots, Z_{ij}[q])$ are, respectively, $1 \times p$ and $1 \times q$ design vectors that can include dummy variables as well as covariates, and $\epsilon_{ij}$ are independent and identically distributed Gaussian random variables (independent of $\boldsymbol{a}(t)$) with zero-mean and variance $\sigma_\epsilon^2$, denoted by $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. Model (1) can be easily extended to accommodate (possibly different number of) repetitions per subject, say $l_i$ $(i = 1, 2, \ldots, n)$. Hereafter, "$^{\mathrm{T}}$" denotes the transpose of a vector or matrix.

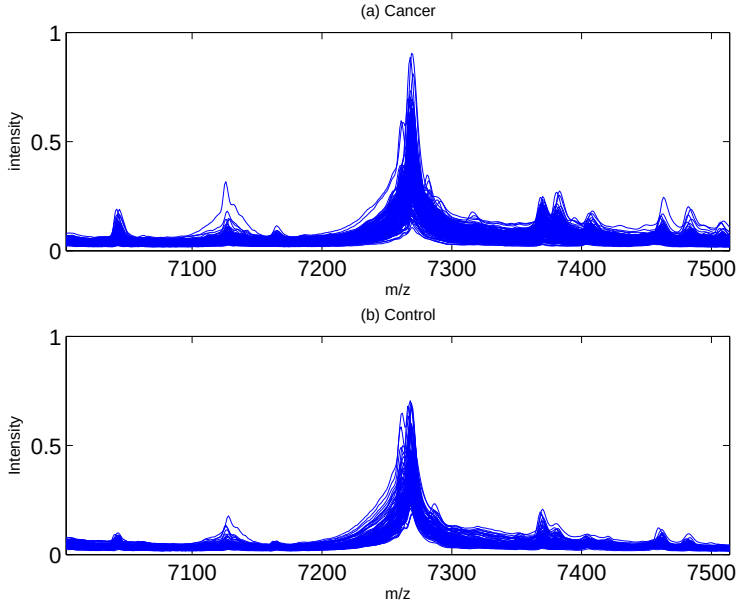



Figure 2.2: Petricoin 08-07-02 dataset: 182 randomly selected spectra (91 in each group) within an m/z range [7003,7514] with baseline subtracted.

Similar to the interpretation of linear mixed-effects models in longitudinal data settings, $\mathbf{X}_{ij}\boldsymbol{\beta}(t)$ can be interpreted as the population-average curve profile, $\mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t)$ can be interpreted as the $i$-th curve-specific deviation (also called the subject-specific deviation if each curve is from a different subject) from the population-average curve profile that accounts for correlation, and $\mathbf{X}_{ij}\boldsymbol{\beta}(t) + \mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t)$ can be interpreted as the $i$-th curve-specific function. Model (1) includes many useful models commonly used in the literature for analysing functional data, including, e.g., linear mixed-effects models, functional regression models, functional fixed-effects analysis of variance models, functional analysis of covariance models, nonparametric mixed-effects models, functional mixed-effects analysis of variance models, smoothing spline mixed-effects analysis of variance models, and nested and crossed samples of curves models (see, e.g., Guo, 2002b).

In the nonparametric analysis of functional data, both the fixed functional components of $\boldsymbol{\beta}(t)$ and the random functional components of $\boldsymbol{a}(t)$ should be modelled as nonparametric functions lying in infinite dimensional spaces (since the basic unit in functional data analysis is the curve). Since deviations from smooth effects may be present, this behaviour should also be included in the modelling formulation. A natural framework to include non-smooth effects is through wavelet decompositions, and it is developed below. We briefly recall first some relevant facts about the wavelet series expansion and the discrete wavelet transform that we need further.

### 3.2 The wavelet series expansion and the discrete wavelet transform

Throughout the paper we assume that we are working within an orthonormal basis generated by dilations and translations of a compactly supported scaling function, $\phi(t)$, and a compactly supported mother wavelet, $\psi(t)$, associated with an $r$-regular ($r \geq 0$) multiresolution analysis of $\big(L^2[0,1], \langle\cdot,\cdot\rangle\big)$, the space of squared-integrable functions on $[0,1]$ endowed with the inner product $\langle f, g\rangle = \int_{[0,1]} f(t)g(t)\,dt$. For simplicity in exposition, we work with periodic wavelet bases on $[0,1]$ (see, e.g., Mallat, 1999, Section 7.5.1), letting

$$\phi^{\mathrm{p}}_{jk}(t) = \sum_{l\in\mathbb{Z}} \phi_{jk}(t-l) \quad \text{and} \quad \psi^{\mathrm{p}}_{jk}(t) = \sum_{l\in\mathbb{Z}} \psi_{jk}(t-l), \quad \text{for} \quad t \in [0,1],$$

where $\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k)$ and $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$. For any given primary resolution level $j_0 \geq 0$, the collection

$$\{\phi^{\mathrm{p}}_{j_0k},\ k = 0, 1, \ldots, 2^{j_0} - 1;\ \psi^{\mathrm{p}}_{jk},\ j \geq j_0;\ k = 0, 1, \ldots, 2^j - 1\}$$

is then an orthonormal basis of $L^2[0,1]$. The superscript "p" will be suppressed from the notation for convenience. Despite the poor behavior of periodic wavelets near the boundaries, where they create high amplitude wavelet coefficients, they are commonly used because the numerical implementation is particular simple. Therefore, for any $f \in L^2[0,1]$, we denote by $u_{j_0k} = \langle f, \phi_{j_0k}\rangle$ ($k = 0, 1, \ldots, 2^{j_0} - 1$) the scaling coefficients and by $w_{jk} = \langle f, \psi_{jk}\rangle$ ($j \geq j_0;\ k = 0, 1, \ldots, 2^j - 1$) the wavelet coefficients of $f$ for the orthonormal periodic wavelet basis defined above; the function $f$ is then expressed in the form

$$f(t) = \sum_{k=0}^{2^{j_0}-1} u_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} w_{jk}\psi_{jk}(t), \quad t \in [0,1].$$

In statistical settings, we are more usually concerned with discretely sampled, rather than continuous, functions. It is then the wavelet analogy to the discrete Fourier transform which is of primary interest and this is referred to as the discrete wavelet transform (DWT). Given a vector of function values $\mathbf{f} = (f(t_1), ..., f(t_n))'$ at equally spaced points $t_i$, the discrete wavelet transform of $\mathbf{f}$ is given by $\mathbf{d} = W_{n\times n}\mathbf{f}$, where $\mathbf{d}$ is an $n \times 1$ vector comprising both discrete scaling coefficients, $c_{j_0k}$, and discrete wavelet coefficients, $d_{jk}$, and $W_{n\times n}$ is an orthogonal $n \times n$ matrix associated with the orthonormal periodic wavelet basis chosen. The $c_{j_0k}$ and $d_{jk}$ are related to their continuous counterparts $u_{j_0k}$ and $w_{jk}$ (with an approximation error of order $n^{-1}$) via the relationships $c_{j_0k} \approx \sqrt{n}\,u_{j_0k}$ and $d_{jk} \approx \sqrt{n}\,w_{jk}$. Note that, because of orthogonality of $W_{n\times n}$, the inverse DWT (IDWT) is simply given by $\mathbf{f} = W^{\mathrm{T}}_{n\times n}\mathbf{d}$, where $W^{\mathrm{T}}_{n\times n}$ denotes the transpose of $W_{n\times n}$. If $n = 2^J$ for some positive integer $J$, the DWT and IDWT may be performed through a computationally fast algorithm (see, e.g., Mallat, 1999, Section 7.3.1) that requires only order $n$ operations.

### 3.3 A wavelet-based Model Specification for the Fixed and Random Effects

An approach to modelling the fixed and random effects, that allows a wide range of irregular effects (for both fixed-effects and random-effects), is through the sequence space representation of Besov spaces. The (inhomogeneous) Besov spaces on the unit interval, $B^s_{\rho_1,\rho_2}[0,1]$, consist of functions that have a specific degree of smoothness in their derivatives. The parameter $\rho_1$ can be viewed as a degree of function's inhomogeneity while $s$ is a measure of its smoothness. Roughly speaking, the (not necessarily integer) parameter $s$ indicates the number of function's (fractional) derivatives, where their existence is required in an $L^{\rho_1}$-sense; the additional parameter $\rho_2$ is secondary in its role, allowing for additional fine tuning of the definition of the space. For a detailed study on (inhomogeneous) Besov spaces we refer to, e.g., Donoho & Johnstone (1998).

By exploiting the relation between the hyperparameters of a prior model and the parameters of those Besov spaces within which realisations from the prior will fall, as originally suggested by Abramovich *et al.* (1998) and further considered by, e.g., Di Zio & Frigessi (1999), the fixed functional components of $\boldsymbol{\beta}(t)$ and the random functional components of $\boldsymbol{a}(t)$ can be both made to share the same degree of smoothness, i.e., they should lie in the same Besov space.

Fix now a primary resolution level $j_0 \geq 0$ and consider the orthonormal periodic wavelet basis $\{\phi_{j_0k},\ k = 0, 1, \ldots, 2^{j_0} - 1;\ \psi_{jk},\ j \geq j_0;\ k = 0, 1, \ldots, 2^j - 1\}$ discussed in Section 3.2. For each $r_1 = 1, 2, \ldots, p$, assume that $\beta_{r_1} \in B^s_{\rho_1,\rho_2}[0,1]$ for $0 < s < r$, $1 \leq \rho_1, \rho_2 \leq \infty$. For each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, consider now the following *random* wavelet series expansion

$$\alpha^{(i)}_{r_2}(t) = \sum_{k=0}^{2^{j_0}-1} c^{(r_2,i)}_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \theta^{(r_2,i)}_{jk}\psi_{jk}(t), \quad t \in [0,1],$$

where, for each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, the wavelet coefficients $\theta^{(r_2,i)}_{jk}$ are assumed to be independent and identically distributed random variables, distributed as

$$\theta^{(r_2,i)}_{jk} \sim N(0, v^{(r_2,i)}_{jk}).$$

Furthermore, we assume that, for each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, the quantities $v^{(r_2,i)}_{jk}$ are functions of the resolution level $j$ only. In particular, we assume that they decrease exponentially as a functions of the resolution level $j$, i.e.,

$$v^{(r_2,i)}_{jk} = \sigma^2_\theta 2^{-j\alpha_i}, \text{ for some } \alpha_i \geq 0,$$

where $\sigma^2_\theta$ is some positive quantity.

A relationship between the Besov space parameters and the hyperparameters of the prior model considered above can be now established. Exploiting the equivalence between the Besov norm of the

function $\alpha_{r_2}^{(i)}(t)$ and the corresponding sequence space norm (see, e.g., Donoho & Johnstone, 1998), and using Theorem 1 in Abramovich *et al.* (1998), for any given values of $c_{j_0k}^{(r_2,i)}$ $(k = 0, 1, \ldots, 2^{j_0}-1)$, and for each $i = 1, 2, \ldots, n$; $r_2 = 1, 2, \ldots, q$,

$$\alpha_{r_2}^{(i)}(t) \in B^s_{\rho_1,\rho_2}[0,1] \quad \text{almost surely}$$

if and only if

$$s + 1/2 - \alpha_i/2 < 0$$

or

$$s + 1/2 - \alpha_i/2 = 0 \ \text{ for } \ 1 \le \rho_1 < \infty \ \text{ and } \ \rho_2 = \infty.$$

The above results show that, in each case, the fixed functional components of $\boldsymbol{\beta}(t)$ and the random functional components of $\boldsymbol{a}(t)$ can be both made to share the same degree of smoothness (i.e., they can both lie in the same functional space, $B^s_{\rho_1,\rho_2}[0,1]$, with $0 < s < r$, $1 \le \rho_1, \rho_2 \le \infty$), by appropriately relating the Besov space parameters $s$, $\rho_1$ and $\rho_2$ to the hyperparameters $\alpha_i$ of the prior model discussed above.

**Remark 3.1** The hyperparameter $v_{jk}^{(r_2,i)}$ is the prior variance of the important wavelet coefficient at resolution level $j$ and scale $k$. The above model corresponds to the prior belief that all wavelet coefficients on all levels have the same probability of being non-zero. This actually characterises self-similar processes, such as Brownian motion, the overall regularity depending on the value of $\alpha_i$. This model allows us to capture key characteristics of variations in multiple curves.

### 3.4 A Linear Mixed-Effects Representation

In this section, we provide a linear mixed-effects representation that is subsequently used for estimation and testing in the general functional mixed-effects model (1). Observations from different subjects are independent, while observations from the same subject are correlated to various degrees. We assume that the within-subject design is equispaced on fine grid, a common model for many instrumental devices usually used to collect functional data. Furthermore we take $m_i = m$ for all $i = 1, 2, \ldots, n$ with $m = 2^J$ for some positive integer $J$. This setting allows one to consider the discrete wavelet transform which can be performed through the computationally fast algorithm mentioned in Section 3.2. Note that this assumption is not especially restrictive, since if the grid is fine enough, interpolation can be used to obtain a common grid (of power two) without substantively changing the observed data.

Fix now a primary resolution level $j_0 \ge 0$ and consider the orthonormal periodic wavelet basis $\{\phi_{j_0k},\ k = 0, 1, \ldots, 2^{j_0} - 1;\ \psi_{jk},\ j \ge j_0;\ k = 0, 1, \ldots, 2^j - 1\}$ discussed in Section 3.2. For each $r_1 = 1, 2, \ldots, p$, we set $\beta_{r_1}(\mathbf{t}) = (\beta_{r_1}(t_1), \ldots, \beta_{r_1}(t_m))^{\mathrm{T}}$, where $\mathbf{t} = (t_1, \ldots, t_m)$ with $t_j = j/m$ for

$j = 1, 2, \ldots, m$. Considerations related to asymptotic minimax optimality theory suggest taking as a maximal resolution level $j_1$ a level such that $m/\ln(m) \leq 2^{j_1} \leq 2m/\ln(m)$ (see Delyon & Juditsky, 1996). We recommend hereafter the choice $m^* = 2^{j_1}$ falling in this interval, because the resulting wavelet estimators perform well for both smooth and piecewise smooth functions with isolated points of singularity. (Obviously, such a choice does not affect the conclusions of Theorem 1 in Abramovich *et al.* (1998) discussed in Section 3.3.) With this notation, we can write $\beta_{r_1}(\mathbf{t}) = W_{m\times m^*}^{\mathrm{T}} \tilde{\mathbf{d}}_{r_1}$, where $W_{m^*\times m}$ is the $m^* \times m$ matrix associated with the orthonormal periodic wavelet basis, and $\tilde{\mathbf{d}}_{r_1}$ is the $m^* \times 1$ vector of the corresponding scaling and wavelet coefficients $\{\tilde{b}_{j_0k}^{(r_1)},\ k = 0, 1, \ldots, 2^{j_0} - 1;\ \tilde{d}_{jk}^{(r_1)},\ j = j_0, \ldots, j_1 - 1;\ k = 0, 1, \ldots, 2^j - 1\}$. We assume that, for each $r_1 = 1, 2, \ldots, p$, $\beta_{r_1}(t) \in B_{\rho_1,\rho_2}^s[0, 1]$ for $0 < s_1 < r$, $1 \leq \rho_1, \rho_2 \leq \infty$.

Regarding the random-effects, the random components of their wavelet coefficients will help us to incorporate the correlation structure and subject-specific features of functional data in the estimation process in the proposed functional mixed-effects model. For each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, we set $\alpha_{r_2}^{(i)}(\mathbf{t}) = (\alpha_{r_2}^{(i)}(t_1), \ldots, \alpha_{r_2}^{(i)}(t_m))^{\mathrm{T}}$ and we can write $\alpha_{r_2}^{(i)}(\mathbf{t}) = W_{m\times m^*}^{\mathrm{T}} \tilde{\boldsymbol{\theta}}_{r_2}^{(i)}$, where again $W_{m^*\times m}$ is the $m^* \times m$ matrix associated with the orthonormal periodic wavelet basis, and $\tilde{\boldsymbol{\theta}}_{r_2}^{(i)}$ is the $m^* \times 1$ vector of the corresponding scaling and wavelet coefficients $\{\tilde{c}_{j_0k}^{(r_2,i)},\ k = 0, 1, \ldots, 2^{j_0} - 1;\ \tilde{\theta}_{jk}^{(r_2,i)},\ j = j_0, \ldots, j_1 - 1;\ k = 0, 1, \ldots, 2^j - 1\}$. We assume that, for each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, the wavelet coefficients $\tilde{\theta}_{jk}^{(r_2,i)}$ are independent and identically distributed $N(0, v_{jk}^{(r_2,i)})$ random variables. This independence assumption implicitly restricts the time-domain covariance matrices to the class of matrices diagonalisable by the DWT. Moreover, since the variances (diagonal elements) are allowed to vary across the wavelet scales, the above model accommodates only stationary covariance structures for the between curves covariance matrix from the same subject. However, even with such restrictions, the model is flexible enough to capturing key characteristics of subject-specific deviations encountered in practice. By analogy to Section 3.3, we take $v_{jk}^{(r_2,i)} = \sigma_\theta^2 2^{-j\alpha_i}$, for some $\alpha_i \geq 0$, where $\sigma_\theta^2$ is some positive quantity. Certain combinations of the Besov parameters and hyperparameters of the prior model can now be exploited in order the fixed functional components of $\boldsymbol{\beta}(t)$ and the random functional components of $\boldsymbol{a}(t)$ lie in the same Besov space $B_{\rho_1,\rho_2}^s[0, 1]$, with $0 < s < r$, $1 \leq \rho_1, \rho_2 \leq \infty$ (see Section 3.3). For identifiability reasons, that will become clear later, we assume that $[n(\ln(m) - q)] > p$, where $[x]$ denotes the integer part of $x$. Finally, we assume that, for each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, $\epsilon_i$ and $\tilde{\theta}_{jk}^{(r_2,i)}$ are independent.

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^{\mathrm{T}}$ and $\widetilde{\mathbf{d}} = (\tilde{\mathbf{d}}_1^{\mathrm{T}}, \ldots, \tilde{\mathbf{d}}_p^{\mathrm{T}})^{\mathrm{T}}$, and let $\widetilde{\mathbf{X}}_i = \mathbf{X}_i\mathbf{W}^{(p)}$ and $\widetilde{\mathbf{Z}}_i = \mathbf{Z}_i\mathbf{W}_n^{(q)}$, where $\mathbf{X}_i = \mathrm{diag}(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im})$ (each element is an appropriately constructed matrix containing dummy variables and/or covariates), $\mathbf{W}^{(p)} = \mathrm{diag}(W_{m\times m^*}^{\mathrm{T}}, \ldots, W_{m\times m^*}^{\mathrm{T}})$ ($p$ blocks), $\mathbf{Z}_i = \mathrm{diag}(\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{im})$ (each element is an appropriately constructed matrix containing dummy variables and/or covariates), $\mathbf{W}_n^{(q)} = \mathrm{diag}(\mathbf{W}^{(q)}, \ldots \mathbf{W}^{(q)})$ ($n$ blocks) and $\mathbf{W}^{(q)} = \mathrm{diag}(W_{m\times m^*}^{\mathrm{T}}, \ldots, W_{m\times m^*}^{\mathrm{T}})$ ($q$

blocks). Let also $\widetilde{\boldsymbol{\theta}}_i = (\tilde{\boldsymbol{\theta}}_{i1}, \ldots, \tilde{\boldsymbol{\theta}}_{iq})^{\mathrm{T}}$ and $\tilde{\boldsymbol{\epsilon}}_i = (\epsilon_{i1}, \ldots, \epsilon_{im})^{\mathrm{T}}$. With this notation, the general functional mixed-effects model (1) can be rewritten as

$$\mathbf{Y} = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\epsilon}}, \tag{2}$$

where $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_n^{\mathrm{T}})^{\mathrm{T}}$, $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1^{\mathrm{T}}, \ldots, \widetilde{\mathbf{X}}_n^{\mathrm{T}})^{\mathrm{T}}$, $\widetilde{\mathbf{Z}} = (\widetilde{\mathbf{Z}}_1^{\mathrm{T}}, \ldots, \widetilde{\mathbf{Z}}_n^{\mathrm{T}})^{\mathrm{T}}$, $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\theta}}_n^{\mathrm{T}})^{\mathrm{T}}$ and $\widetilde{\boldsymbol{\epsilon}} = (\widetilde{\boldsymbol{\epsilon}}_1^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\epsilon}}_n^{\mathrm{T}})^{\mathrm{T}}$. Model (2) is clearly a *linear mixed-effects model* with *one* variance component where the fixed-effects are parameterized by the wavelet coefficients of $\beta_{r_1}(t)$ ($r_1 = 1, 2, \ldots, p$) and the random-effects are parameterized by the wavelet coefficients of $\alpha_{r_2}^{(i)}(t)$ ($i = 1, 2, \ldots, n$; $r_2 = 1, 2, \ldots, q$).

Obviously, $\mathbb{E}(\widetilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\epsilon}})^{\mathrm{T}} = (\mathbf{O}_{nm^*q}, \mathbf{O}_{nm})^{\mathrm{T}}$, where $\mathbf{O}_k$ is a $k \times k$ matrix with zero entries. Moreover, it is not difficult to see that $\mathbb{V}(\widetilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\epsilon}})^{\mathrm{T}} = \mathrm{diag}(\sigma_\theta^2 \boldsymbol{\Sigma}, \sigma_\epsilon^2 \mathbf{I}_{nm})$, where $\boldsymbol{\Sigma} = \mathrm{diag}\left(\boldsymbol{\Sigma}^{(1)}, \ldots, \boldsymbol{\Sigma}^{(n)}\right)$ ($n$-components) with $\boldsymbol{\Sigma}^{(i)}$ being a diagonal matrix with diagonal entries corresponding to the elements $2^{-j\alpha_i}$ for each $i = 1, 2, \ldots, n$ (the same for all $r_2 = 1, 2, \ldots, q$), and $\mathbf{I}_k$ is the $k \times k$ identity matrix. It is easily seen from the above that the corresponding covariance surface for the Gaussian process modelling the random-effects $\alpha_{r_2}^{(i)}(t)$ is given by $\sigma_\theta^2 W_{m \times m^*} \boldsymbol{\Sigma} W_{m \times m^*}^{\mathrm{T}}$. This matrix describes how the functions vary one from another and the parameters $\alpha_i$ and $\sigma_\theta^2$ have a clear impact on any inference that is done but such a specification seems unavoidable since the large dimension of the covariance matrices make it infeasible to estimate them in a completely unstructured fashion. Ideally, estimates of these parameters could be obtained from some prior information or assumptions about, e.g., the regularity of realizations of the random-effects $\alpha_{r_2}^{(i)}(t)$. In practice, however, it is often difficult to elicit such a prior information about the regularity properties. We therefore suggest to estimate these parameters form the data by maximum likelihood. Therefore, the parameters associated with the covariance matrices are substituted by their estimates without taking their precision into account in the inference that is done subsequently.

Summarizing, we can now write

$$\mathbb{E}(\mathbf{Y}) = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} \qquad \text{and} \qquad \mathbb{V}(\mathbf{Y}) = \sigma_\epsilon^2 \mathbf{V}_\lambda,$$

where $\mathbf{V}_\lambda = \mathbf{I}_{nm} + \lambda \widetilde{\mathbf{Z}} \boldsymbol{\Sigma} \widetilde{\mathbf{Z}}^{\mathrm{T}}$ and $\lambda = \sigma_\theta^2 / \sigma_\epsilon^2$. The parameter $\lambda$ can be considered as a ratio of the *curve-to-curve variability* and the *within-curve noise*. Note that $\sigma_\theta^2 = 0$ if and only if $\lambda = 0$, and the parameter space for $\lambda$ is $[0, \infty)$.

**Remark 3.2** The independence assumption of the random effects in the wavelet domain discussed above only implies independence in the data domain when the variance components of the random effects are identical across all resolution levels $j$ and locations $k$, implying that our modelling methodology allows to model correlation between observations over the same individual.

# 4 Inference in Functional Mixed-Effects Models

## 4.1 Testing for Random-Effects

According to the modelling formulation in Section 3.4, testing for random-effects in the general functional mixed-effects model (1) is equivalent to testing for a zero variance component in the linear mixed-effects model (2) which, in turn, is equivalent to testing the following hypotheses

$$H_0 : \sigma_\theta^2 = 0 \ (\lambda = 0) \quad \text{versus} \quad H_A : \sigma_\theta^2 > 0 \ (\lambda > 0). \tag{3}$$

Testing the above hypotheses is non-standard because the parameter under the null hypothesis is on the boundary of the parameter space. Therefore, using the non-standard asymptotic theory developed by Self & Liang (1987) for independent data, one may be tempted to conclude that the finite sample null distributions of the resulting LR and RLR tests could be approximated by a $0.5\delta(0) + 0.5\chi_1^2$ distribution, i.e., a 50:50 mixture of a point mass at zero and a chi-square distribution with one degree of freedom. However, a second problem is lack of independence, at least under the alternative hypothesis. Because the response variable $\mathbf{Y}$ in the linear mixed-effects model (2) is usually not a vector of independent random variables, the non-standard asymptotic theory of Self & Liang (1987) does not apply. With this in mind, Crainiceanu & Ruppert (2004) have recently derived finite sample and asymptotic null distributions for the LR and RLR test statistics in linear mixed-effects model with one variance component. This is the approach that we consider in the following development for testing the hypotheses in (3).

### 4.1.1 Profile and Restricted Profile Log-Likelihood Functions

Twice the log-likelihood function for the linear mixed-effects model (2) is (up to a constant that does not depend on the parameters)

$$L(\widetilde{\mathbf{d}}, \lambda, \sigma_\epsilon^2) = -nm \log(\sigma_\epsilon^2) - \log|\mathbf{V}_\lambda| - \frac{(\mathbf{Y} - \widetilde{\mathbf{X}}\widetilde{\mathbf{d}})^{\mathrm{T}} \mathbf{V}_\lambda^{-1} (\mathbf{Y} - \widetilde{\mathbf{X}}\widetilde{\mathbf{d}})}{\sigma_\epsilon^2}. \tag{4}$$

Under the alternative hypothesis $H_A$ in (3), by *fixing* $\lambda$ and solving the first order minimum conditions for $\widetilde{\mathbf{d}}$ and $\sigma_\epsilon^2$, one gets the maximum profile likelihood estimates

$$\widehat{\widetilde{\mathbf{d}}}(\lambda) = (\widetilde{\mathbf{X}}^T \mathbf{V}_\lambda^{-1} \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^{\mathrm{T}} \mathbf{V}_\lambda^{-1} \mathbf{Y} \tag{5}$$

and

$$\widehat{\sigma_\epsilon^2}(\lambda) = \frac{1}{nm} (\mathbf{Y} - \widetilde{\mathbf{X}} \widehat{\widetilde{\mathbf{d}}}(\lambda))^{\mathrm{T}} \mathbf{V}_\lambda^{-1} (\mathbf{Y} - \widetilde{\mathbf{X}} \widehat{\widetilde{\mathbf{d}}}(\lambda)). \tag{6}$$

Plugging the expressions (5) and (6) into (4) we obtain (up to a constant that does not depend on the parameters), the profile log-likelihood function

$$L(\lambda) = -\log|\mathbf{V}_\lambda| - nm \log(\mathbf{Y}^{\mathrm{T}} \mathbf{P}_\lambda^{\mathrm{T}} \mathbf{V}_\lambda^{-1} \mathbf{P}_\lambda \mathbf{Y}), \tag{7}$$

where $\mathbf{P}_\lambda = \mathbf{I}_{nm} - \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^T\mathbf{V}_\lambda^{-1}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\mathbf{V}_\lambda^{-1}$.

In order to take into account the loss in degrees of freedom due to estimation of the $m^*p$-dimensional $\widetilde{\mathbf{d}}$ parameters, and thereby to obtain unbiased variance components estimators, Patterson & Thompson (1971) introduced the notion of restricted (or residual) maximum likelihood (RML) (or generalised maximum likelihood in the spline smoothing literature). RML consists of maximizing the likelihood function associated with (any) $(nm-m^*p)$ linearly independent contrasts. Twice the restricted log-likelihood function for the linear mixed-effects model (2) is (up to a constant that does not depend on the parameters)

$$l(\widetilde{\mathbf{d}}, \lambda, \sigma_\epsilon^2) = L(\widetilde{\mathbf{d}}, \lambda, \sigma_\epsilon^2) - m^*p\log(\sigma_\epsilon^2) - \log(|\widetilde{\mathbf{X}}^{\mathrm{T}}\mathbf{V}_\lambda^{-1}\widetilde{\mathbf{X}}|). \tag{8}$$

Using arguments similar to the ones used to obtain the maximum profile likelihood estimates, the maximum restricted profile likelihood estimate of $\widetilde{\mathbf{d}}(\lambda)$ is still given by (5) while the maximum restricted profile likelihood estimate of $\sigma_\epsilon^2(\lambda)$ is now given by

$$\widehat{\sigma_\epsilon^2}(\lambda) = \frac{1}{nm - m^*p}(\mathbf{Y} - \widetilde{\mathbf{X}}\widehat{\widetilde{\mathbf{d}}}(\lambda))^{\mathrm{T}}\mathbf{V}_\lambda^{-1}(\mathbf{Y} - \widetilde{\mathbf{X}}\widehat{\widetilde{\mathbf{d}}}(\lambda)). \tag{9}$$

The restricted profile log-likelihood function (up to a constant that does not depend on the parameters) is then given by

$$l(\lambda) = -\log|\mathbf{V}_\lambda| - \log|\widetilde{\mathbf{X}}^T\mathbf{V}_\lambda^{-1}\widetilde{\mathbf{X}}| - (nm - m^*p)\log\{\mathbf{Y}^{\mathrm{T}}\mathbf{P}_\lambda^{\mathrm{T}}\mathbf{V}_\lambda^{-1}\mathbf{P}_\lambda\mathbf{Y}\}. \tag{10}$$

By following Crainiceanu & Ruppert (2004) and Claeskens (2004), and taking into account the restriction $[n(\ln(m)-q)] > p$ (see Section 3.4), one can show that both profile and restricted profile log-likelihood functions can be written as functions of latent eigenvalues. In particular, the profile log-likelihood function (7) can be written, up to a constant that does not depend on the parameters, as

$$L(\lambda) = -\sum_{s=1}^{nm^*q}\log(1+\lambda\xi_{s,nm}) - nm\log\left\{\sigma_\epsilon^2\sum_{s=1}^{nm^*q}\frac{1}{1+\lambda\mu_{s,nm}}\omega_s^2 + \sum_{s=nm^*q+1}^{nm-m^*p}\omega_s^2\right\},$$

while, the restricted profile log-likelihood function (10), can be written as

$$l(\lambda) = -\sum_{s=1}^{nm^*q}\log(1+\lambda\mu_{s,nm}) - (nm-m^*p)\log\left\{\sigma_\epsilon^2\sum_{s=1}^{nm^*q}\frac{1}{1+\lambda\mu_{s,nm}}\omega_s^2 + \sum_{s=nm^*q+1}^{nm-m^*p}\omega_s^2\right\},$$

where $\omega_s$ are independent and identically distributed $N(0,1)$ random variables, and $\mu_{s,nm}$ and $\xi_{s,nm}$ are the eigenvalues of the matrices $\mathbf{K}_\mu = \mathbf{\Sigma}^{1/2}\widetilde{\mathbf{Z}}^{\mathrm{T}}\mathbf{P}_0\widetilde{\mathbf{Z}}\mathbf{\Sigma}^{1/2}$ and $\mathbf{K}_\xi = \mathbf{\Sigma}^{1/2}\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Z}}\mathbf{\Sigma}^{1/2}$ respectively. Here, $\mathbf{\Sigma}^{1/2}$ is the unique symmetric square root of $\mathbf{\Sigma}$ and $\mathbf{P}_0 = \mathbf{I}_{nm} - \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}$.

### 4.1.2 Finite Sample and Asymptotic Null Distributions of the LR and RLR Tests

The finite sample LR test statistic is defined as

$$\mathrm{LR}_{nm} \equiv \sup_{\lambda \geq 0} \mathrm{LR}_{nm}(\lambda) \equiv \sup_{\lambda \geq 0} (L(\lambda) - L(0)),$$

while the finite sample RLR test statistic is defined as

$$\mathrm{RLR}_{nm} \equiv \sup_{\lambda \geq 0} \mathrm{RLR}_{nm}(\lambda) \equiv \sup_{\lambda \geq 0} (l(\lambda) - l(0)).$$

The following theorem gives the spectral representations of the finite sample null distributions of the $\mathrm{LR}_{nm}$ and $\mathrm{RLR}_{nm}$ test statistics that can be used for testing the hypotheses in (3).

**Theorem 4.1** *Let $\mu_{s,nm}$ and $\xi_{s,nm}$ be the eigenvalues of the matrices $\mathbf{K}_\mu = \mathbf{\Sigma}^{1/2}\widetilde{\mathbf{Z}}^{\mathrm{T}}\mathbf{P}_0\widetilde{\mathbf{Z}}\mathbf{\Sigma}^{1/2}$ and $\mathbf{K}_\xi = \mathbf{\Sigma}^{1/2}\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Z}}\mathbf{\Sigma}^{1/2}$ respectively, where $\mathbf{P}_0 = \mathbf{I}_{nm} - \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}$. Then, under the null hypothesis $H_0$ in (3),*

$$\mathrm{LR}_{nm} \stackrel{\mathcal{D}}{=} \sup_{\lambda \geq 0} \left[ nm \log \left\{ 1 + \frac{N_{nm}(\lambda)}{D_{nm}(\lambda} \right\} - \sum_{s=1}^{nm^*q} \log(1 + \lambda \xi_{s,nm}) \right] \tag{11}$$

*and*

$$\mathrm{RLR}_{nm} \stackrel{\mathcal{D}}{=} \sup_{\lambda \geq 0} \left[ (nm - m^*p) \log \left\{ 1 + \frac{N_{nm}(\lambda)}{D_{nm}(\lambda} \right\} - \sum_{s=1}^{nm^*q} \log(1 + \lambda \mu_{s,nm}) \right], \tag{12}$$

*where the notation "$\stackrel{\mathcal{D}}{=}$" denotes equality in distribution,*

$$N_{nm}(\lambda) = \sum_{s=1}^{nm^*q} \frac{\lambda \mu_{s,nm}}{1 + \lambda \mu_{s,nm}} \omega_s^2, \quad D_{nm}(\lambda) = \sum_{s=1}^{nm^*q} \frac{\omega_s^2}{1 + \lambda \mu_{s,nm}} + \sum_{s=nm^*q+1}^{nm-m^*p} \omega_s^2,$$

*and $\omega_s$, $s = 1, 2, \ldots, nm^*q$, are independent and identically distributed $N(0,1)$ random variables.*

Each of the finite sample null distributions of the $\mathrm{LR}_{nm}$ and $\mathrm{RLR}_{nm}$ test statistics has a probability mass at zero, and this mass can be very large indeed. Although there is no simple expression for these probabilities, there is a good approximation (see Crainiceanu & Ruppert, 2004).

The finite sample null distributions of the $\mathrm{RL}_{nm}$ and $\mathrm{RLR}_{nm}$ test statistics depend only on the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$. Following Crainiceanu & Ruppert (2004), the following algorithm, which we have used in the analysis of the examples presented in Section 5, provides a simple way to simulate the finite sample null distributions of the $\mathrm{LR}_{nm}$ and $\mathrm{RLR}_{nm}$ test statistics, once the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$ have been calculated.

**Step 1.** Define a grid $0 = \lambda_1 < \lambda_2 < \ldots < \lambda_K$ of possible values of $\lambda$.

**Step 2.** Simulate $nm^*q$ independent $\chi_1^2$ random variables $\omega_1^2, \ldots, \omega_{nm^*q}^2$.

**Step 3.** Independently of step 2, simulate $X_{nm,m^*p,nm^*q} = \sum_{s=nm^*q+1}^{nm-m^*p} \omega_s^2$ with $\chi^2_{nm-m^*p-nm^*q}$ distribution.

**Step 4.** For every grid point $\lambda_i$, compute

$$N_{nm}(\lambda_i) = \sum_{s=1}^{nm^*q} \frac{\lambda_i \mu_{s,nm}}{1+\lambda_i \mu_{s,nm}} \omega_s^2 \quad \text{and} \quad D_{nm}(\lambda) = \sum_{s=1}^{nm^*q} \frac{\omega_s^2}{1+\lambda_i \mu_{s,nm}} + X_{nm,m^*p,nm^*q}.$$

**Step 5.** Determine $\lambda_{\max}^{\mathrm{LR}}$ which maximizes

$$f_{nm}^{\mathrm{LR}}(\lambda_i) = \left[ nm \log\left\{1 + \frac{N_{nm}(\lambda_i)}{D_{nm}(\lambda_i)}\right\} - \sum_{s=1}^{nm^*q} \log(1+\lambda_i \xi_{s,nm}) \right]$$

and $\lambda_{\max}^{\mathrm{RLR}}$ which maximizes

$$f_{nm}^{\mathrm{RLR}}(\lambda_i) = \left[ (nm - m^*p) \log\left\{1 + \frac{N_{nm}(\lambda_i)}{D_{nm}(\lambda_i)}\right\} - \sum_{s=1}^{nm^*q} \log(1+\lambda_i \mu_{s,nm}) \right]$$

over $\lambda_1, \lambda_2, \ldots, \lambda_K$.

**Step 6.** Compute

$$\mathrm{LR}_{nm} = f_{nm}^{\mathrm{LR}}(\lambda_{\max}^{\mathrm{LR}}) \quad \text{and} \quad \mathrm{RLR}_{nm} = f_{nm}^{\mathrm{RLR}}(\lambda_{\max}^{\mathrm{RLR}}).$$

**Step 7.** Repeat **Steps 2 - 6**.

The above algorithm could, however, be computationally very expensive since its speed depends on the number of random-effects, $nm^*q$, in the linear mixed-effects model (2) (which obviously depends on the number of subjects, $n$, the number of observations, $m$, per subject, and the number of random-effects, $q$, in the general functional mixed-effects model (1)).

Alternatively, the asymptotic null distributions of the $\mathrm{LR}_{nm}$ and $\mathrm{RLR}_{nm}$ test statistics can be easily obtained; they actually depend on the asymptotic behaviour of the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$ used to calculated the finite sample null distributions of the $\mathrm{LR}_{nm}$ and $\mathrm{RLR}_{nm}$ test statistics. Note that all these asymptotic null distributions essentially depend on the asymptotic behaviour of the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$. When these eigenvalues cannot be computed explicitly it may be simple to study the asymptotic behaviour of the corresponding matrices. Once the asymptotic behaviour of the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$ is available, one can either obtain closed-form expressions for or easily simulated from the corresponding asymptotic null distributions.

The following theorem provides the asymptotic null distributions of the $\mathrm{LR}_{nm}$ and $\mathrm{RLR}_{nm}$ test statistics.

**Theorem 4.2** *Let $\mu_{s,nm}$ and $\xi_{s,nm}$ be the eigenvalues of the matrices $\mathbf{K}_\mu = \mathbf{\Sigma}^{1/2}\widetilde{\mathbf{Z}}^{\mathrm{T}}\mathbf{P}_0\widetilde{\mathbf{Z}}\mathbf{\Sigma}^{1/2}$ and $\mathbf{K}_\xi = \mathbf{\Sigma}^{1/2}\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Z}}\mathbf{\Sigma}^{1/2}$ respectively, where $\mathbf{P}_0 = \mathbf{I}_{nm} - \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}$. Suppose that there exists a*

*constant $\eta \geq 0$ so that, for every $s$, the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$ satisfy $\lim_{nm\to\infty}(nm)^{-\eta}\mu_{s,nm} = \mu_s$ and $\lim_{nm\to\infty}(nm)^{-\eta}\xi_{s,nm} = \xi_s$, where $\mu_s \neq 0$ for at least one $s$. Then, under the null hypothesis $H_0$ in (3),*

$$\mathrm{LR}_{nm} \Rightarrow \sup_{d\geq 0}\left\{\sum_{s=1}^{nm^*q}\frac{d\,\mu_{s,nm}}{1+d\,\mu_{s,nm}}\omega_s^2 - \sum_{s=1}^{nm^*q}\log(1+d\,\xi_{s,nm})\right\}$$

*and*

$$\mathrm{RLR}_{nm} \Rightarrow \sup_{d\geq 0}\left\{\sum_{s=1}^{nm^*q}\frac{d\,\mu_{s,nm}}{1+d\,\mu_{s,nm}}\omega_s^2 - \sum_{s=1}^{nm^*q}\log(1+d\,\mu_{s,nm})\right\},$$

*where the notation "$\Rightarrow$" denotes weak convergence and $\omega_s$, $s = 1, 2, \ldots, nm^*q$, are independent and identically distributed $N(0,1)$ random variables.*

Each of the asymptotic null distributions of the $\mathrm{LR}_{nm}$ and $\mathrm{RLR}_{nm}$ test statistics has a probability mass at zero, and this mass can be very large indeed. Although there is no simple expression for these probabilities, there is a good approximation (see Crainiceanu & Ruppert, 2004). We point out that the asymptotic null distributions depend on the asymptotic behaviour of the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$. In the following, we treat a simple example of practical interest showing how these conditions can be reduced to a simple expression.

**Example 4.1** `(Balanced One-Way Functional Mixed-Effects ANOVA Model)`
Consider the following balanced one-way functional mixed-effects ANOVA model with $n$ levels, $m$ discretised values per level, and $l$ repetitions per level, i.e.,

$$Y_{ijk} = \beta(t_{ij}) + \alpha^{(i)}(t_{ij}) + \epsilon_{ijk}, \quad i = 1, \ldots, n;\ j = 1, \ldots, m;\ k = 1, \ldots, l, \tag{13}$$

where $\beta(t)$ is an unknown functional mean, $\alpha^{(i)}(t)$ are realisations of a zero-mean Gaussian process $a(t)$, and $\epsilon_{ijk}$ are independent and identically distributed $N(0, \sigma_\epsilon^2)$ random variables that are also independent of $a(t)$. Using the wavelet transform parameterisation discussed in Section 3.4, the matrix $\widetilde{\mathbf{X}}$ for fixed-effects is simply an $nml \times m^*$ matrix with $nl$ block columns equal to $W_{m\times m^*}^{\mathrm{T}}$ and the matrix $\widetilde{\mathbf{Z}}$ for random-effects is the $nml \times nm^*$ matrix, with $l$ row blocks each made by a block diagonal $nm \times nm^*$ with the matrix $W_{m\times m^*}^{\mathrm{T}}$ on the diagonal. We consider the asymptotic situation, where the number of levels $n$ is fixed, while $m, l \to \infty$. Recall that $\mathbf{P}_0$ denotes the orthogonal projector of $\mathbb{R}^{nml}$ onto the space orthogonal to the column space of $\widetilde{\mathbf{X}}$. By the orthogonality of the columns of $W_{m\times m^*}^{\mathrm{T}}$, it is easy to see that the rank of $\widetilde{\mathbf{X}}$ is $m^*$ and, therefore, $\mathbf{P}_0$ has $nml - m^*$ eigenvalues equal to 1 and $m^*$ eigenvalues equal to 0. Using again the orthogonality of the columns of $W_{m\times m^*}^{\mathrm{T}}$, it is easy to prove that $\widetilde{\mathbf{Z}}^T\mathbf{P}_0\widetilde{\mathbf{Z}}$ has $m^*(n-1)$ eigenvalues equal to $l$ and the remaining eigenvalues equal to 0. Moreover, the eigenvalues of $\mathbf{K}_\mu$ are given by the product of eigenvalues of $\mathbf{\Sigma}$ and $\widetilde{\mathbf{Z}}^T\mathbf{P}_0\widetilde{\mathbf{Z}}$ while the eigenvalues of $\mathbf{K}_\xi$ are given by the product of eigenvalues

of $\mathbf{\Sigma}$ and $\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}}$. Given the behaviour of the diagonal matrix $\mathbf{\Sigma}$, it follows that both $\mu_{s,nm}$ and $\xi_{s,nm}$ are therefore $\mathcal{O}(l(m^*)^{-\eta})$. The conditions on the asymptotic behaviour of the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$ mentioned above reduce now in assuming that there exists a constant $\eta \geq 0$ such that $l^{1-\eta}(nm)^{\eta}(m^*)^{-\eta} = \mathcal{O}(1)$.

### 4.1.3 Estimators for the ratio of the curve-to-curve variability and the within-curve noise.

We discuss below two possible ways of obtaining consistent estimators of the parameter $\lambda = \sigma_\theta^2/\sigma_\epsilon^2$.

**Profile and Restricted Profile Maximum Likelihood Estimators.** By maximizing (7) or (10) one can obtain the profile and restricted profile maximum likelihood estimators $\hat{\lambda}^{\mathrm{LR}}$ and $\hat{\lambda}^{\mathrm{RLR}}$ respectively of $\lambda$. Under some regularity conditions, one can show that $\hat{\lambda}^{\mathrm{LR}}$ and $\hat{\lambda}^{\mathrm{RLR}}$ are consistent estimators of $\lambda$ (see Claeskens, 2004).

**A Wavelet Domain Based Estimator.**

It is not difficult to see that, for each individual data curve, the empirical wavelet coefficients of the data, at each resolution level $j$, are independent random variables, distributed as Gaussian distributions with appropriate means and variances, depending on $v_{jk}^{(r_2,i)}$. The maximum likelihood estimators of $v_{jk}^{(r_2,i)}$ can be obtained explicitly. Given $\sigma_\epsilon^2$, consistent estimators of $\sigma_\theta^2$ and $\alpha_i$ can be now obtained in closed forms. Indeed, by noting that for $j = 0$ one gets $v_{jk}^{(r_2,i)} = \sigma_\theta^2$, an estimator of $\sigma_\theta^2$ can be obtained for each value of $r_2 = 1, 2, \ldots, q$, and then estimate $\sigma_\theta^2$ by averaging its $q$ estimates obtained, resulting in the estimator $\hat{\sigma}_\theta^2$. Noting that $v_{jk}^{(r_2,i)} = \sigma_\theta^2 2^{-j\alpha_i}$, estimators of $\alpha_i$ are now easily derived.

However, in most applications, the noise variance $\sigma_\epsilon^2$ can also be estimated in the wavelet domain. In wavelet function estimation, the common practice is to robustly estimate $\sigma_\epsilon$ by the median of the absolute deviation of the empirical wavelet coefficients of the data at the highest resolution level divided by 0.6745. This can be done for all the individual data curves and then estimate $\sigma_\epsilon$ by averaging its $n$ robust estimates obtained from each individual data curve, resulting in the estimator $\hat{\sigma}_\epsilon^2$. A consistent estimator of $\lambda$ is then simply obtained by $\hat{\lambda}^{wd} = \hat{\sigma}_\theta^2/\hat{\sigma}_\epsilon^2$.

## 4.2 Testing for Fixed-Effects

The wavelet decomposition proposed in Section 3.4 for the general functional mixed-effects model (1) can also provide an efficient way to make meaningful inference on the fixed-effects by testing whether certain fixed-effects or contrasts are equal to zero. The proposed method will be based on an appropriately defined F-test based procedure for testing that the expectation of a Gaussian vector with $nm$ independent components belongs to a linear subspace of $R^{nm}$ against a nonparametric

alternative. The testing procedure is available even when the variance of the observations is unknown and does not depend on any prior information on the alternative. The properties of the test are nonasymptotic and the test will be rate optimal (up to a logarithmic factor) over various classes of alternatives simultaneously.

To begin with, consider the linear mixed-effects model (2) and take the case where $\sigma_\epsilon^2$ is unknown but the parameter $\lambda = \sigma_\theta^2/\sigma_\epsilon^2$ is known. The general functional mixed-effects model (1) and the specific wavelet-based modelling approach that we have used in Section 3.3 for representing the random-effects functional components show that the vector of observations $\mathbf{Y}$ is Gaussian. In this case, the image of the vector $\mathbf{Y}$ by $V_\lambda^{-1/2}$ leads to the linear regression model

$$\mathbf{Y}_\lambda = V_\lambda^{-1/2}\mathbf{Y} = V_\lambda^{-1/2}\tilde{\mathbf{X}}\tilde{\mathbf{d}} + \sigma_\epsilon \boldsymbol{\eta}, \tag{14}$$

where $\boldsymbol{\eta}$ is a random vector with independent and identically distributed standard Gaussian components, i.e., $\eta_i \sim N(0,1)$ for $i = 1, 2, \ldots, nm$. Let $\boldsymbol{\nu}$ denote the expectation of $\mathbf{Y}$ and let $\boldsymbol{\mu}$ be its image by $V_\lambda^{-1/2}$. The space of means $\mathcal{E}$ of model (14) is the $m^*p$-dimensional linear subspace of $\mathbb{R}^{nm}$ spanned by the columns of the matrix $V_\lambda^{-1/2}\tilde{\mathbf{X}}$, i.e.,

$$\mathcal{E} = \{\boldsymbol{\mu} \in \mathbb{R}^{nm} : \ \boldsymbol{\mu} = V_\lambda^{-1/2}\tilde{\mathbf{X}}\tilde{\mathbf{d}} \ \text{ with } \ \tilde{\mathbf{d}} \in \mathbb{R}^{m^*p}\}.$$

Testing for significant fixed-effects functional components or contrasts is formally a test of the null hypothesis $H_c : A_c\tilde{\mathbf{d}} = \mathbf{0}$ for a suitable defined matrix $A_c$, against general alternatives. A powerful approach to such a high-dimensional hypothesis testing is available by adapting the model selection based procedures proposed recently in Baraud *et al.* (2003), which are naturally generalized to our present scenario.

Let $\mathcal{V}_c$ be the linear subspace of $\mathcal{E}$ defined by

$$\mathcal{V}_c = \{V_\lambda^{-1/2}\tilde{\mathbf{X}}\tilde{\mathbf{d}}, \ A_c\tilde{\mathbf{d}} = \mathbf{0}\}$$

for a suitable defined contrast matrix $A_c$. Following the idea of Baraud *et al.* (2003), we propose below a test of $\boldsymbol{\mu} \in \mathcal{V}_c$ against that it does not. The testing procedure relies upon appropriately defined $F$-statistics which have been widely used for hypothesis testing in the linear model framework due to their intuitional appeal and their equivalence to LR for fixed-effects models. It is described as follows.

We consider a finite collection $\{S_\ell : \ \ell \in \mathcal{L}\}$ of linear subspaces included in the orthogonal complement $\mathcal{V}_c^\perp \cap \mathcal{E}$ of $\mathcal{V}_c$ in $\mathcal{E}$, such that for each $\ell \in \mathcal{L}$, $S_\ell \neq \mathcal{V}_c^\perp \cap \mathcal{E}$ and $S_\ell \neq \{0\}$. The index set $\mathcal{L}$ is allowed to depend on the number of observations $nm$. Given a suitable sequence $\{\bar{\alpha}_\ell : \ \ell \in \mathcal{L}\}$ of numbers in $(0,1)$, we consider for each $\ell \in \mathcal{L}$, the Fisher test of level $\bar{\alpha}_\ell$ for testing

$$H_{0,c} : \ \boldsymbol{\mu} \in \mathcal{V}_c \quad \text{versus} \quad H_{A,\ell} : \boldsymbol{\mu} \in (\mathcal{V}_c + S_\ell)\backslash\mathcal{V}_c, \tag{15}$$

and denote by $T_{c,\ell}$ the corresponding test statistic. The resulting test can then be regarded as an adaptive test of linear hypothesis based on a multiple testing procedure rejecting $H_{0,c}$ against $H_{A,\ell}$ as soon as there exists $\ell \in \mathcal{L}$ such that $T_{c,\ell}$ is larger than some threshold.

To pursue, let us first introduce notations that will be repeatedly used throughout this section. The distribution of the vector of observations $\mathbf{Y}_\lambda$ will be denoted by $\mathbb{P}_{\boldsymbol{\mu}}$. For any linear subspace $\mathcal{A}$ of $\mathbb{R}^{nm}$, we denote by $\Pi_{\mathcal{A}}$ the orthogonal projector onto $\mathcal{A}$ (with respect to the Euclidean norm $||\cdot||$). For any $u \in \mathbb{R}$, $\bar{\Phi}(u)$, $\bar{\chi}_D(u)$ and $\bar{F}_{D,N}(u)$ denote respectively the probability for a standard Gaussian variable, a chi-square with $D$ degrees of freedom, and a Fisher with $D$ and $N$ degrees of freedom to be larger than $u$. For any $c$, $d_c$ will denote the dimension of $\mathcal{V}_c$ and, for each $\ell \in \mathcal{L}$, $D_\ell$ and $N_\ell$ will respectively denote the dimensions of $S_\ell$ and $(\mathcal{V}_c + S_\ell)^\perp \cap \mathcal{E}$. Let also $k_c$ be the rank of $A_c$.

### 4.2.1 Description of the Test

Let $\bar{\alpha} \in (0,1)$ be a fixed significance level. Assume that the collection $\{S_\ell : \ \ell \in \mathcal{L}\}$ of linear subspaces of $\mathcal{V}_c^\perp \cap \mathcal{E}$ is such that $1 \leq D_\ell \leq nm - m^\star p + k_c - 1$. We set

$$T_{c,\ell} = \frac{N_\ell \|\Pi_{S_\ell} \mathbf{Y}_\lambda\|^2}{D_\ell \|\Pi_{(\mathcal{V}_c + S_\ell)^\perp \cap \mathcal{E}} \mathbf{Y}_\lambda\|^2},$$

and we define

$$T_{\bar{\alpha}} = \sup_{\ell \in \mathcal{L}} \{T_{c,\ell} - \bar{F}^{-1}_{D_\ell, N_\ell}(\bar{\alpha}_\ell)\}, \tag{16}$$

where $\{\bar{\alpha}_\ell : \ \ell \in \mathcal{L}\}$ is a sequence of numbers in $(0,1)$ such that, for all $\boldsymbol{\mu} \in \mathcal{V}_c$, $\sum_{\ell \in \mathcal{L}} \bar{\alpha}_\ell \leq \alpha$. We then reject the null hypothesis (15) when $T_{\bar{\alpha}}$ is positive.

### 4.2.2 Level of the Test

We first study the level of the test statistic defined in (16) and show that it is of level $\bar{\alpha}$. Indeed, the following theorem holds.

**Theorem 4.3** *The test statistic $T_{\bar{\alpha}}$ defined in (16), under the null hypothesis (15), satisfies*

$$\forall\, \boldsymbol{\mu} \in \mathcal{V}_c, \quad \mathbb{P}_{\boldsymbol{\mu}}\{T_{\bar{\alpha}} > 0\} \leq \bar{\alpha}.$$

The proof of Theorem 4.3 shows clearly that the above procedure is a Bonferroni-like procedure in which the $p$-value $\bar{\alpha}$ is composed by $\#\mathcal{L}$ significance levels, where $\#\mathcal{L}$ is the total number of models that are tested. It is well known that the Bonferroni approach is overly conservative when $\#\mathcal{L}$ is large; the choice of $\mathcal{L}$ is therefore important and connected to optimal model selection procedures (see Section 4.2.4).

### 4.2.3 Power of the Test

We now study the power of the test statistic defined in (16). Let $0 < \gamma < 1$, and let us first introduce some quantities that depend on $\bar{\alpha}_\ell$, $\gamma$, $D_\ell$ and $N_\ell$. For each $u > 0$ and each $\ell \in \mathcal{L}$, we set

$$L_\ell = \log(1/\bar{\alpha}_\ell), \quad L = \log(2/\gamma), \quad r_\ell = 2\exp(4L_\ell/N_\ell),$$

$$K_\ell(u) = 1 + 2\sqrt{\frac{u}{N_\ell}} + 2r_\ell\frac{u}{N_\ell}, \quad \Lambda_1(\ell) = 2.5\left(1 + \max(K_\ell(L_\ell), r_\ell)\right)\frac{D_\ell + L_\ell}{N_\ell},$$

$$\Lambda_2(\ell) = 2.5\sqrt{1 + K_\ell^2(L)}\left(1 + \sqrt{\frac{D_\ell}{N_\ell}}\right), \quad \Lambda_3(\ell) = 2.5\left[\max\left(\frac{r_\ell K_\ell(L)}{2}, 5\right)\right]\left(1 + 2\frac{D_\ell}{N_\ell}\right).$$

Under the condition, $\bar{\alpha}_\ell \geq \exp(-N_\ell/10)$ and $\gamma \geq 2\exp(-N_\ell/21)$, for all $\ell \in \mathcal{L}$, which is usually met for reasonable choices of $\{S_\ell : \ \ell \in \mathcal{L}\}$ and $\{\bar{\alpha}_\ell : \ \ell \in \mathcal{L}\}$, the quantities $\Lambda_1(\ell)$, $\Lambda_2(\ell)$ and $\Lambda_3(\ell)$ behave like constants (see Baraud *et al.*, 2003).

With the above notation, the following theorem holds.

**Theorem 4.4** *Let $T_{\bar{\alpha}}$ be the test statistic defined by (16), and assume that $\tilde{\mathbf{X}}^T V_\lambda^{-1}\tilde{\mathbf{X}}$ converges to a positive definite matrix as $m \to \infty$. Let $\mathcal{F}_m$ and $\rho_m^2$ be defined as follows*

$$\mathcal{F}_m = \{\boldsymbol{\mu}(\tilde{\mathbf{d}}) \in \mathcal{E};\ d_m^2(\tilde{\mathbf{d}}, \mathcal{V}_c) \geq \rho_m^2\} \tag{17}$$

*and*

$$\rho_m^2 = \inf_{\ell \in \mathcal{L}}\left[(1 + \Lambda_1(\ell))\, d_m^2(\Pi_{\mathcal{V}_c^\perp \cap \mathcal{E}}\boldsymbol{\mu}, S_\ell) + v_\ell^2\right],$$

*where*

$$v_\ell^2 = \left[\Lambda_2(\ell)\sqrt{D_\ell \log\left(\frac{2}{\gamma\bar{\alpha}_\ell}\right)} + \Lambda_3(\ell)\log\left(\frac{2}{\gamma\bar{\alpha}_\ell}\right)\right]\frac{\sigma_\epsilon^2}{m^\star p}.$$

*Then,*

$$\lim_{m\to\infty}\sup_{\tilde{\mathbf{d}} \in \mathcal{F}_m} \mathbb{P}_{\boldsymbol{\mu}(\tilde{\mathbf{d}})}\left(T_{\bar{\alpha}} \leq 0\right) = 0.$$

According to Theorem 4.4, one can see that the larger the $\mathcal{F}_m$ is the better the power of the test. The definition of the set $\mathcal{F}_m$ suggests that we would take advantage in considering a collection of linear subspaces $\{S_\ell : \ \ell \in \mathcal{L}\}$ with good approximation properties in order to decrease the bias term, $d_m^2(\Pi_{\mathcal{V}_c^\perp \cap \mathcal{E}}\boldsymbol{\mu}, S_\ell)$ (as well as $\rho_m^2$). In fact, there is a balance to achieve between $d_m^2(\Pi_{\mathcal{V}_c^\perp \cap \mathcal{E}}\boldsymbol{\mu}, S_\ell)$ and $v_\ell^2$. If, for example, the collection $\{S_\ell : \ \ell \in \mathcal{L}\}$ is totally ordered for the inclusion, $d_m^2(\Pi_{\mathcal{V}_c^\perp \cap \mathcal{E}}\boldsymbol{\mu}, S_\ell)$ decreases with $D_\ell$ but $v_\ell^2$ increases with $D_\ell$. Therefore, the choice of $\mathcal{L}$ has to be done carefully, as we see for a specific case in Section 4.2.4 below.

The proposed test statistic (16) cannot directly be computed in practical applications because it depends on the unknown quantity $\lambda$. However, this problem can be solved by replacing $\lambda$

with a consistent estimator, regardless that $H_0$ is true or not (see Horowitz & Spokoiny, 2001, Section 2.5). This is exactly the case for the signal-to-noise estimators $\hat{\lambda}^{\mathrm{LR}}$, $\hat{\lambda}^{\mathrm{RLR}}$ and $\hat{\lambda}^{wd}$ discussed in Section 4.1.3. The method we have developed operates just as if the parameter $\lambda$ substituted is the true one, without taking the uncertainty in estimating $\lambda$ into account in the inference that is done subsequently.

### 4.2.4 Nonasymptotic minimax rates for testing the nullity of functional fixed-effects contrasts

Here, we derive an upper bound for the rate of testing the nullity of a given contrast of the functional fixed-effects in the general functional mixed-effects model (1). We are therefore able to evaluate the general bounds and power of the testing procedure discussed above. The connection with the procedure given above is clear when relating the discrete wavelet coefficients $\tilde{\mathbf{d}}$ with the mean vector $\boldsymbol{\mu}(\tilde{\mathbf{d}})$, which is in this case nothing else than the vector of sampled values of the contrast, at least when the sampling grid is the same for all individuals. Indeed, for two functions $f$ and $g$ sampled on an equidistant grid on $[0,1]$ of size $m$, we set $\|\mathbf{f}-\mathbf{g}\|_m^2 = \sum_{i=1}^m (f(t_i)-g(t_i))^2/m$ and $d_m(\mathbf{f},\mathbf{g}) = \|\mathbf{f}-\mathbf{g}\|_m$. For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^m$ we set $\|\mathbf{u}-\mathbf{v}\|_m^2 = \sum_{i=1}^m (u_i-v_i)^2/m$ and $d_m(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|_m$. Note that if $\mathbf{u}$ and $\mathbf{v}$ (in $\mathbb{R}^m$) are the discrete wavelet coefficients of the functions $f$ and $g$ respectively, sampled on an equidistant grid on $[0,1]$ of size $m$, we then have that $\|\mathbf{u}-\mathbf{v}\|_m^2 = \|\mathbf{f}-\mathbf{g}\|_m^2$ (see Antoniadis, 1994).

Let $s \in (0,1]$ and $R>0$. We assume that the functional fixed-effects contrast that we wish to test its nullity belongs to a class within a Besov space $\mathcal{B}^s_{\infty,\infty}([0,1],R)$ (a Hölder space on $[0,1]$ of regularity $s$),

$$\mathcal{B}^s_{\infty,\infty}([0,1],R) = \{f:\ |f(x)-f(y)| \leq R|x-y|^s\},$$

i.e., the desired class is expressed as

$$\mathcal{F}(R,s,\rho_m) = \{f \in \mathcal{B}^s_{\infty,\infty}([0,1],R):\ d_m(\mathbf{f},\boldsymbol{\mu}(\tilde{\mathbf{d}})) \geq \rho_m\},$$

where $\boldsymbol{\mu}(\tilde{\mathbf{d}})$ is the wavelet reconstruction from the wavelet coefficients of the true mean of the estimated contrast.

Here, our concern will be the rate at which the distance between the null and alternative hypotheses can decrease to zero while still permitting consistent testing, the set of alternatives should be also separated away from the null hypothesis in the $d_m$-distance by $\rho_m$. Theorem 4.5 below gives a upper bound for the minimum separation from zero (see, e.g., Abramovich *et al.*, 2004), uniformly over $\mathcal{F}(R,s,\rho_m)$, by considering a specific collection of subspaces $S_\ell$ and a series of levels $\bar{a}_\ell$.

Denote by $\mathcal{M}_m$ the set of all indices $j$ such that $2^j \leq [m/2]$. For each index $j \in \mathcal{M}_m$, let $K_j = \{k:\ 1 \leq k \leq 2^j\}$ and set $B_{kj} = [(k-1)/2^j, k/2^j)$. Therefore, for each $j \in \mathcal{M}_m$, the intervals

$(B_{kj})_{k\in K_j}$ define a partition of $[0,1) = \cup_{k\in K_j} B_{kj}$. For each $j \in \mathcal{M}_m$, the subspaces $S_j$ that we consider is the linear space spanned by the following set of vectors $\left\{\frac{1}{\#B_{kj}}\sum_{i=1}^{m}\sum_{t_i\in B_{kj}} e_i;\ k \in K_j\right\}$, where $\#B_{kj} = \#\{t_i \in B_{kj},\, i = 1,\ldots,m\}$ and $(e_1,\ldots,e_m)$ be the canonical basis of $\mathbb{R}^m$. Note that $S_j$, as it is defined above, is related to the basis of piecewise constant functions on $[0,1)$. Indeed, for each $j \in \mathcal{M}_m$ and for each $k \in K_j$, denoting by $g_k(x) = 1_{B_{kj}}(x)$, it is easily seen that $S_j$ is the vector space spanned by the vectors $\mathbf{g}_k = (g_k(t_1),\ldots,g_k(t_m))$. (When $\mathcal{V}_c = \{0\}$, i.e., we test the nullity of the corresponding functional contrast $c$, $S_j$ is in $\mathcal{V}_c^{\perp} = \mathbb{R}^m$.)

For a given $f$, let $\rho_m^2(f)$ be the "indifference threshold" for testing $f \equiv 0$ against that $f \in \mathcal{F}(R,s,\rho_m)$. With the above notation, the following theorem holds.

**Theorem 4.5** *Assume that* $R^2 \geq \frac{\sigma_\epsilon^2}{m}\sqrt{\ln\ln(m)}$. *Let* $\bar{\alpha}$ *be an overall significance level. Then, there exists a constant* $C_{\bar{\alpha}}$ *(depending on* $\bar{\alpha}$*), such that for all* $s \in (0,1]$, *one has*

$$\rho_m^{\star 2} := \sup_{f\in\mathcal{F}(R,s,\rho_m)} \rho_m^2(f) \leq C_{\bar{\alpha}}\left[R^{\frac{2}{(1+4s)}}\left(\frac{\sigma_\epsilon^2}{m}\sqrt{\ln\ln(m)}\right)^{\frac{4s}{(1+4s)}} + R^2 m^{-2s} + \frac{\sigma_\epsilon^2}{m}\ln\ln(m)\right].$$

Recall that the optimal rate of testing is the fastest rate at which $\rho_m^\star$ can approach zero while permitting consistent testing uniformly over $\mathcal{F}(R,s,\rho_m^\star)$. Note that when $1/4 \leq s \leq 1$, the rate of testing of our procedure, in the sense of Ingster (1982), is $\left(\frac{1}{m}\sqrt{\ln\ln(m)}\right)^{\frac{2s}{(1+4s)}}$. The minimax rate of testing is, however, $m^{-\frac{2s}{(1+4s)}}$. The loss of efficiency by a $\ln\ln(m)$ factor is unavoidable and is due to the fact that our procedure is adaptive with respect to $s$ and $R$ (see, e.g., Spokoiny, 1996). On the other hand, when $s < 1/4$, the rate of testing is of order $m^{-s}$, but it is not known whether such a rate is optimal or not. When $\sigma_\epsilon^2$ is assumed to be known, the rate of testing for regular functions, as the ones we consider for testing contrasts, is $m^{-1/4}$ (see Baraud, 2002).

## 5 Applications

The purpose of this section is to illustrate the usefulness of the proposed functional hypotheses testing procedures described in Section 4, by applying them on the orthosis and mass spectrometry proteomic datasets discussed in Section 2.

Using classical linear mixed-effects estimation techniques, the linear mixed-effects representation (2) is also used to obtain wavelet-based estimates for both fixed-effects and random-effects. More specifically, for the estimation of the functional fixed-effects (i.e., the population-average curve profiles), we apply the classical weighted least-squares (WLS) methodology, which is easily implemented. On the other hand, for the estimation of the functional random-effects (i.e., the curve-specific functions), we apply RML estimation of variance components, as is commonly used in standard liner mixed-models software such as, e.g., `PROC MIXED` in `SAS` and `lme()` in `S-PLUS`

(see, e.g., Ngo & Wand, 2004). Following the wavelet-based model formulation for random-effects discussed in Section 3.3, and in order to reduce the computational time, we have taken for both examples $\alpha_i = \alpha$ for all $i = 1, 2, \ldots, n$. Its value was estimated by maximum likelihood.

The computational algorithms related to wavelet analysis were performed using Version 8 of the freely available WaveLab toolbox for Matlab.

### 5.1 Orthosis Data

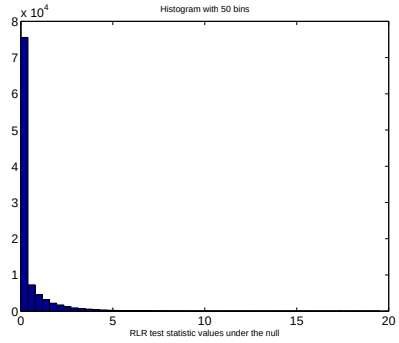
Abramovich *et al.* (2004) analysed this dataset as arising from a fixed-effects functional analysis of variance model with 2 qualitative factors (Subjects and Treatments), 1 quantitative factor (Time) and 10 replications for each level combination. They considered a block design model, treating subjects as blocks, which allowed them to make inference about the treatments of interest; they found significant global differences between treatments although under Spring 1 and Spring 2 conditions the subjects behave similarly, the same being less true under Control and Orthosis conditions. They also found a highly significant global trend over time.

However, as in Abramovich & Angelini (2006), it is more reasonable to treat subjects as random-effects and to apply the proposed estimation and testing procedures. (We point out at this point that testing for functional random-effects is lacking from the mixed-effects functional analysis of variance testing methodology of Abramovich & Angelini (2006).) Averaging over the 10 repetitions for each subject, following the wavelet-based formulation, and using the matrix notation of Section 3.4, it is not difficult to see that, in this particular situation, we have $n = 28$, $m = 64$ ($m^\star = 16$), $p = 4$ and $q = 1$, and the general functional mixed-effects model (1) can be expressed as a linear mixed-effects model with one variance component, written as

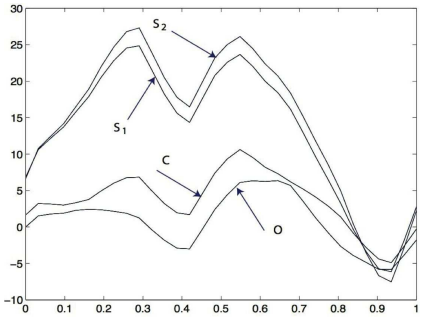
$$\mathbf{Y} = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\epsilon}},$$

using the compactly supported *Symmlet 4-tap filter* mother wavelet, where

- $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_{28}^{\mathrm{T}})^{\mathrm{T}}$ (a $1792 \times 1$ vector of data points);
- $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{W}^{(4)}$ (a $1792 \times 64$ fixed-effects design matrix), $\mathbf{X} = (A_1^{\mathrm{T}}, A_2^{\mathrm{T}}, A_3^{\mathrm{T}}, A_4^{\mathrm{T}})^{\mathrm{T}}$, where $A_i$ ($i = 1, \ldots, 4$) are $7 \times 4$ block-zero matrices apart from their $i^{\mathrm{th}}$ column which consists of 7 identity matrices each one of size $64 \times 64$, and $\mathbf{W}^{(4)} = \mathrm{diag}(W_{64\times 16}^{\mathrm{T}}, \ldots, W_{64\times 16}^{\mathrm{T}})$ (4 blocks);
- $\widetilde{\mathbf{d}} = (\tilde{\mathbf{d}}_1^{\mathrm{T}}, \ldots, \tilde{\mathbf{d}}_4^{\mathrm{T}})^{\mathrm{T}}$ (a $64 \times 1$ vector of fixed-effects wavelet coefficients);
- $\widetilde{\mathbf{Z}} = \mathbf{I}_{1792}\mathbf{W}^{(28)}$ (a $1792 \times 448$ random-effects design) and $\mathbf{W}_{28}^{(1)} = \mathrm{diag}(W_{64\times 16}^{\mathrm{T}}, \ldots, W_{64\times 16}^{\mathrm{T}})$ (28 blocks);
- $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\theta}}_{28}^{\mathrm{T}})^{\mathrm{T}}$ (a $448 \times 1$ vector of random-effects wavelet coefficients);




(a)




(b)

Figure 5.3: (a): Random-effects testing for the Orthosis dataset: The histogram with 50 bins of the RLR test statistic values under the null hypothesis; (b): Fixed-effects estimates for the Orthosis dataset ('S2' (Spring 2), 'S1' (Spring 1), 'C' (Control) and 'O' (Orthosis)).

- $\widetilde{\boldsymbol{\epsilon}} = (\widetilde{\boldsymbol{\epsilon}}_1^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\epsilon}}_{28}^{\mathrm{T}})^{\mathrm{T}}$ (a $1792 \times 1$ vector of standard Gaussian errors).

One of the aim of the analysis is to understand how a subject can cope with the external perturbation, and we need to quantify the ways in which the individual mean cross-sectional functions differ over the various conditions. Below, we apply our general methodology described in previous sections in order to test both functional fixed-effects (i.e., if there is difference between specific functional contrasts of interest) and functional random-effects (i.e., if there is any random-effect present in the dataset) as well as to estimate the various functional components.

Regarding the functional random-effects, the application of the testing methodology presented in Section 4.1 reveals that $\hat{\sigma}_\theta = 40.313$, $\hat{\alpha} = 4.082$ and $\hat{\sigma}_\epsilon = 1.080$ resulting in $\hat{\lambda}^{wd} = 1393.531$. The finite sample RLR test statistic, computing on a grid of 400 points and taking 100000 simulations from the null, takes the value of 3.490 and the corresponding probability at zero value is 0.528 which shows that the corresponding testing methodology is feasible. Figure 5.3(a) shows the histogram with 50 bins of the RLR test statistic values under the null hypothesis. The corresponding $p$-value is 0.028, showing that there is significant evidence of random-effects in this case. Figure 5.4 shows the random-effects estimates of the averaged curves in each group based on RML estimation of variance components.

Regarding the functional fixed-effects, the application of the testing methodology presented in Section 4.2 reveals that a piecewise constant functions collection $\{S_\ell : \ \ell \in \mathcal{L}\}$ of orders 1, 3, 7 and 15 gives the following Bonferroni based test statistic value (where, in each case, each of the corresponding Bonferroni level is taken as 0.0125): 4.847 for Spring 1 vs Spring 2 conditions, 8.992 for Control vs Orthosis conditions, and 48.751 for Spring 1 + Spring 2 vs Control + Orthosis

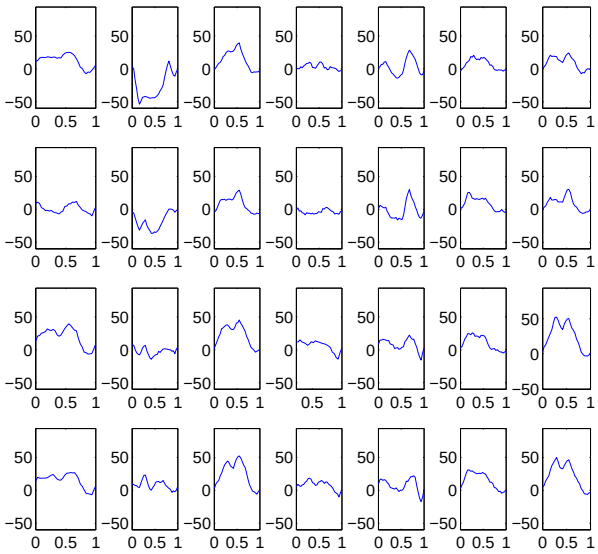


Figure 5.4: Random-effects estimates for the Orthosis dataset.

conditions. These show that the various fixed-effects hypotheses of a similar behaviour under the different conditions are all rejected (the overall p-values were 0.020, 0.001 and 0 respectively). This supports the fact that individuals adjust their posture differently under perturbations of different nature. Note that the different behaviour under Spring 1 and Spring 2 conditions, that is further supported by the empirical evidence of the scientists provided us with the data, it is not captured by the testing methodologies of Abramovich *et al.* (2004) and Abramovich & Angelini (2006). Figure 5.3(b) shows the corresponding group means estimates based on the WLS method.

The entire model fitting for the orthosis data example took 49 minutes in Matlab on a Mac G5 computer 1.8 GHz with 1 GB RAM. The most time consuming step is the estimation of the fixed-effects by the WLS method.

## 5.2 Mass Spectrometry Proteomic Data

Petricoin *et al.* (2002) reported finding patterns in mass spectrometry proteomic data (SELDI-TOF, Ciphergen) that can distinguish between serum samples from normal women and serum samples from women with ovarian cancer, even when the cancers are at early stages (data are available from `http://clinicalproteomics.steem.com`). There are three different datasets and we have used the most recent one dated 08-07-02; see Alexe *et al.* (2004) for a description of all datasets and of various analyses. The dataset comprises 162 cancer samples and 91 control cases.

The expression profile contains measurements of 15,154 intensities on a grid of m/z ratios, but these intensities do not represent independent peptides for this instrument. The peptides yield "peaks" in the spectra, which cover a substantial number of m/z values. A typical SELDI-TOF profile might contain information on several hundred peptides.

Certain preprocessing steps must be performed before analyzing the spectra, including removal of baseline and normalization to calibrate the spectra from different samples. There is no consensus in the literature about how the various pre-processing steps should be done. The hope remains that if a strong signal is truly present in the given data, then it will not be too sensitive to the details of the pre-processing, and information critical for building models with strong prediction capabilities will be retained. We performed baseline correction on all spectra by using a loess procedure and normalization using the appropriate functions in the Bioinformatics toolbox of Matlab (see Figure 2.2 in the introduction for a plot of the preprocessed data).

As now acknowledged by the original authors (see Liotta *et al.*, 2005), the cancers and controls for this experiment were not randomized, but rather run in separate batches, which resulted in a strong systematic bias between the cancer and control spectra. Thus, in this dataset, it is quite easy to separate the cancer and control spectra using any part of the spectrum (see Baggerly *et al.*, 2005), mainly beacuse of this hard-wired bias. However, here we have used this dataset with the only purpose of illustrating the proposed methodology. From the original dataset of curves that exemplified the methods of Petricoin *et al.* (2002), we have randomly selected a subset of 24 curves relevant for our mixed-effects FANOVA application, in an attempt to discriminate between 12 disease spectra (cancer) and 12 normal spectra (control). Figure 5.5 shows some mass spectra for the cancer and control groups together with their means. It has been noted in the literature that the intensities at a range of m/z values around 7000 to 7500 have some interesting features, with a number of peaks in the control samples that are more pronounced than the peaks in the cancer samples. Since the grid is fine enough, and for computational feasibility, the data in this range were interpolated on an equispaced grid of 512 points, without substantively changing the observed data. Following the wavelet-based formulation and using the matrix notation of Section 3.4, in this particular situation we have $n = 24$, $m = 512$ ($m^\star = 128$), $p = 2$ and $q = 1$, and the general functional mixed-effects model (1), as in our previous example, is expressed again as a linear mixed-effects model with one variance component

$$\mathbf{Y} = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\epsilon}},$$

using the compactly supported *Symmlet 4-tap filter* mother wavelet, where

- $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_{24}^{\mathrm{T}})^{\mathrm{T}}$ (a $12288 \times 1$ vector of data points);
- $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{W}^{(2)}$ (a $12288 \times 256$ fixed-effects design matrix), $\mathbf{X} = (A_1^{\mathrm{T}}, A_2^{\mathrm{T}})^{\mathrm{T}}$, where $A_i$ $(i = 1, 2)$


(a) (b)

Figure 5.5: Petricoin 08-07-02 dataset: Mass spectra (thin lines) and the corresponding average spectra (thick lines): (a) Control group; (b) Cancer group.

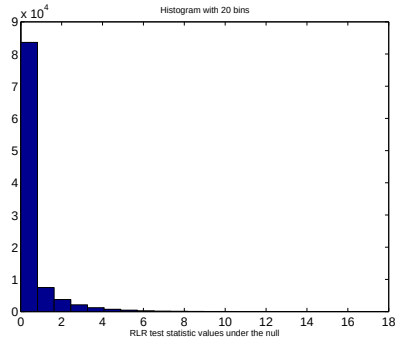
are $12 \times 2$ block-zero matrices apart from their $i^{\text{th}}$ column which consists of 12 identity matrices each one of size $512 \times 512$, and $\mathbf{W}^{(2)} = \text{diag}(W_{512\times 128}^{\text{T}}, W_{512\times 128}^{\text{T}})$;

- $\widetilde{\mathbf{d}} = (\tilde{\mathbf{d}}_1^{\text{T}}, \tilde{\mathbf{d}}_2^{\text{T}})^{\text{T}}$ (a $256 \times 1$ vector of fixed-effects wavelet coefficients);
- $\widetilde{\mathbf{Z}} = \mathbf{I}_{12288}\mathbf{W}^{(24)}$ (a $12288 \times 3072$ random-effects design) and $\mathbf{W}_{24}^{(1)} = \text{diag}(W_{512\times 128}^{\text{T}}, \ldots, W_{512\times 128}^{\text{T}})$ (24 blocks);
- $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^{\text{T}}, \ldots, \widetilde{\boldsymbol{\theta}}_{24}^{\text{T}})^{\text{T}}$ (a $3072 \times 1$ vector of random-effects wavelet coefficients);
- $\widetilde{\boldsymbol{\epsilon}} = (\widetilde{\boldsymbol{\epsilon}}_1^{\text{T}}, \ldots, \widetilde{\boldsymbol{\epsilon}}_{24}^{\text{T}})^{\text{T}}$ (a $12288 \times 1$ vector of standard Gaussian errors).

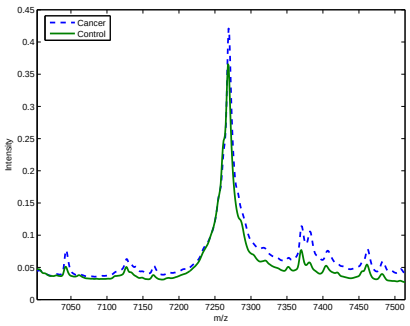
One of the aim of the analysis is to test both functional fixed-effects (i.e., if there is difference between the two group mean curves) and functional random-effects (i.e., if there is any random-effect present in the dataset) as well as to estimate the various functional components.

Regarding the functional random-effects, the application of the testing methodology presented in Section 4.1 reveals that $\hat{\sigma}_\theta = 10.012$, $\hat{\alpha} = 2.542$ and $\hat{\sigma}_\epsilon = 0.629$ resulting in $\hat{\lambda}^{wd} = 253.141$. The finite sample RLR test statistic, computing on a grid of 400 points and taking 100000 simulations from the null, takes the value of 30.541 and the corresponding probability at zero value is 0.531 which shows that the corresponding testing methodology is feasible. Figure 5.6 shows the histogram with 20 bins of the RLR test statistic values under the null hypothesis. The corresponding $p$-value is 0, showing that there is, as suspected from the plots, a significant strong evidence of random-effects in this case.

Regarding the functional fixed-effects, the application of the testing methodology presented in Section 4.2 reveals that a piecewise constant functions collection $\{S_\ell : \ \ell \in \mathcal{L}\}$ of orders 1, 3, 7,




(a)




(b)

Figure 5.6: (a): Random-effects testing for the Petricoin dataset: The histogram with 20 bins of the RLR test statistic values under the null hypothesis; (b): Fixed-effects estimates for the Petricoin 08-07-02 dataset.

15, 31, 63 and 127 gives a Bonferroni based test statistic value (where each of the corresponding Bonferroni level is taken as 0.007) of 104.556. This shows that the fixed-effects hypothesis that the two group means are the same is rejected (the overall $p$-value is 0.009). As expected, the proposed methodology is capable of diagnosis for diseases using SELDI-TOF on tissue samples even in presence of a strong random-effect. Figure 5.6(b) shows the corresponding group means estimates based on the WLS method.

The entire model fitting for the mass spectrometry proteomic data example took 2 hours and 37 minutes in Matlab on a Mac G5 computer 1.8 GHz with 1 GB RAM. The most time consuming step is the estimation of the fixed-effects by the WLS method.

## 6 Concluding Remarks

Although the particular motivated examples were modelled as a functional mixed-effects analysis of variance models, the methodology presented in this paper is very general and can be applied to other functional mixed-effects models, depending on the particular applications at hand. On the other hand, the matrix representation used to construct the general linear mixed-effects model that was subsequently considered for further analysis increases the computational and storage demands, and put some limitation of the proposed methodology in large sample sizes, as it has been observed in our computations. The complexity of the proposed algorithm is $O(nm^{\star}p)$, so when $nm^{\star}p$ is very large the algorithm is not efficient. As mentioned in the application section, the most time consuming step is the estimation of the fixed-effects by the wavelet-based weighted least-squares method. A possible way to improve upon this limitation is to explore the sparsity of the involved

matrices; however, this is not immediately clear to us and it has not been taken care of in our implementation.

We finally point out an interesting extension of the proposed modelling methodology. An alternative model can be built by taking different priors with hyperparameters $\alpha_i$ for $i \in \{1, 2, \ldots, n_1\}$, $i \in \{n_1 + 1, n_1 + 2, \ldots, n_2\}, \ldots$, $i \in \{n_{p-1} + 1, n_{p-1} + 2, \ldots, n_p\}$, where $n_1 + n_2 + \ldots + n_p = n$. In other words, using similar notation, model (2) is now replaced by

$$\mathbf{Y} = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} + \widetilde{\mathbf{Z}}_1\widetilde{\boldsymbol{\theta}}_1 + \ldots + \widetilde{\mathbf{Z}}_q\widetilde{\boldsymbol{\theta}}_q + \tilde{\boldsymbol{\epsilon}}. \tag{18}$$

Model (18) is clearly a *linear mixed-effects model* with $q$ variance components where the fixed-effects are parameterized by the wavelet coefficients of $\beta_{r_1}(t)$ $(r_1 = 1, 2, \ldots, p)$ and the random-effects are parameterized by the wavelet coefficients of $\alpha_{r_2}^{(i)}(t)$ $(i = 1, 2, \ldots, n;\ r_2 = 1, 2, \ldots, q)$. Adapting the recent methodology of Claeskens (2004), ideas similar to the ones presented earlier can now be further developed to provide functional hypothesis testing procedures for both fixed-effects and random-effects as well as to estimate the various components.

A detailed analysis of all these is beyond the scope of this article but present avenues for further research that hope will be addressed in the future.

## Acknowledgements

Anestis Antoniadis was supported by the 'IAP Research Network P5/24', the 'EC-HPRN-CT-2002-00286 Breaking Complexity Network', and the 'Cyprus-France CY-FR/0204/04 Zenon Program'. Theofanis Sapatinas was supported by the 'Cyprus-France CY-FR/0204/04 Zenon Program'. We would like to thank Dr. Amarantini and Dr. Luc for providing us with the Orthosis data. The authors would like to thank the two referees for their useful comments and suggestions.

## Appendix A

For sake of brevity, we provide outlines of the proofs of the theoretical results obtained in Section 4.

**Proof of Theorem 4.1.** The proof of the theorem can be obtained by working along the same lines of the proof of Theorem 1 in Crainiceanu & Ruppert (2004), taking into account that $[n(\ln(m) - q)] > p$. □

**Proof of Theorem 4.2.** The proof of the theorem can be obtained by working along the same lines of the proof of Theorem 2 in Crainiceanu & Ruppert (2004), taking also into account the discussion in their Section 3 and that $[n(\ln(m) - q)] > p$. □

**Proof of Theorem 4.3.** The proof of the theorem can be obtained by arguing along the same lines of the arguments given in Section 2.2 in Baraud *et al.* (2003). More precisely, under the null hypothesis (15), and for each $\ell$, the random variables $\|\Pi_{S_\ell}\mathbf{Y}_\lambda\|^2$ and $\|\Pi_{(\mathcal{V}_c+S_\ell)^\perp\cap\mathcal{E}}\mathbf{Y}_\lambda\|^2$ are independent and distributed as $\chi^2$ variables with $D_\ell$ and $N_\ell$ degrees of freedom respectively. Thus, for each $\ell\in\mathcal{L}$, the test statistics $T_{c,\ell}$ are distributed under the null as Fisher variables with $D_\ell$ and $N_\ell$ degrees of freedom and, therefore, we have

$$\forall\,\boldsymbol{\mu}\in\mathcal{V}_c,\quad \mathbb{P}_{\boldsymbol{\mu}}\{T_{c,\ell}>\bar{F}^{-1}_{D_\ell,N_\ell}(\bar{\alpha}_\ell)\}\leq\bar{\alpha}_\ell.$$

Using now the Bonferroni inequality, it follows that

$$\forall\,\boldsymbol{\mu}\in\mathcal{V}_c,\quad \mathbb{P}_{\boldsymbol{\mu}}\{T_{\bar{\alpha}}>0\}\leq\sum_{\ell\in\mathcal{L}}\mathbb{P}_{\boldsymbol{\mu}}\{T_{c,\ell}>\bar{F}^{-1}_{D_\ell,N_\ell}(\bar{\alpha}_\ell)\}\leq\sum_{\ell\in\mathcal{L}}\bar{\alpha}_\ell\leq\bar{\alpha},$$

which is our claim. □

**Proof of Theorem 4.4.** The proof of the theorem can be obtained by working along the same lines of the proof of Theorem 1 in Baraud *et al.* (2003), by taking into account the way the space $\mathcal{V}_c$ is defined, which follows by noting that the extra assumption on the design matrix $\tilde{\mathbf{X}}$ and the properties of the discrete wavelet transform of a discretised function do not affect the orders of $d^2_m(\Pi_{\mathcal{V}_c^\perp\cap\mathcal{E}}\boldsymbol{\mu},S_\ell)$ and of $v_\ell$. □

**Proof of Theorem 4.5.** Let $D_j$ be the dimension of $S_j$. Note first that for all $j\in\mathcal{M}_m$, $D_j\leq 2^j$. Since, by assumption, for all $j\in\mathcal{M}_m$ we have that $s^j\leq[m/2]$, we get

$$\rho_m(f)\leq\inf_{j\in\mathcal{M}_m}\left\{\left(1+\frac{k_{\bar{\alpha}}}{2}\right)d^2_m(\mathbf{f},S_j)+k_{\bar{\alpha}}\left[\sqrt{2^j\left(\ln\left(\frac{1}{\bar{\alpha}_m}\right)+\ln\ln(m)\right)}+\ln\left(\frac{1}{\bar{\alpha}_m}\right)+\ln\ln(m)\right]\frac{\sigma^2_\epsilon}{m}\right\},$$

where $\bar{\alpha}_m=\frac{\bar{\alpha}}{\#\mathcal{M}_m}$ and $k_{\bar{\alpha}}$ is a positive constant depending only on $\bar{\alpha}$. Now, by definition,

$$d^2_m(\mathbf{f},S_j)=\inf_{g_j\in S_j}\left\{\frac{1}{m}\sum_{i=1}^m\sum_{k\in K_j}(f(x_i)-[\mathbf{g}_m]_i)^2\,1_{B_{kj}}(x_i)\right\}\leq\inf_{g_j\in S_j}\left\{\frac{1}{m}\sum_{i=1}^m\sum_{k\in K_j}\sup_{B_{kj}}|f(x_i)-g_j(x_i)|\right\}^2,\tag{19}$$

where $g_j$ is a piecewise constant on [0,1] which coincides with a constant on each interval $B_{kj}$ and such that $g_j(x_i)=[\mathbf{g}_j]_i$. By Corollary 3.1 in Dahmen *et al.* (1980), we know that each function in $\mathcal{B}^s_{\infty,\infty}([0,1],R)$ is uniformly approximated by a piecewise function on [0,1]. Therefore, there exists a piecewise constant function $\bar{g}$, which is constant on each rectangle $B_{kj}$ for each $j\in K_j$, such that $\sup_{x\in B_{kj}}|f(x)-\bar{g}_j(x)|\leq CR2^{-js}$, where $C>1$ is a constant. Using (19), we easily get

$$d^2_m(\mathbf{f},S_j)=C^2R^2 2^{-2js}.\tag{20}$$

The cardinality of the set $\mathcal{M}_m$ is less than $\log_2([m/2])$. Therefore, for all $j \in \mathcal{M}_m$, we have that $\bar{\alpha}_m \geq C/\log_2([m/2])$, and that

$$\ln\left(\frac{\ln(m)}{\bar{\alpha}_m}\right) \leq \ln\ln(m). \tag{21}$$

Using inequalities (20) and (21), we obtain

$$\rho_m(f) \leq C\left[\inf_{j\in\mathcal{M}_m}\left\{R^2 2^{-2js} + \frac{\sigma_\epsilon^2}{m}\sqrt{2^j \ln\ln(m)}\right\} + \frac{\sigma_\epsilon^2}{m}\ln\ln(m)\right]. \tag{22}$$

The conclusion of the theorem now follows by working along the same lines of the proof of Corollary 2 in Baraud *et al.* (2003). Indeed, note that

$$R^2 2^{-2js} \leq \frac{\sigma_\epsilon^2}{m}\sqrt{2^j \ln\ln(m)}$$

if and only if

$$2^j \geq \rho^\star := \left(\frac{R^2 m}{\sigma_\epsilon^2\sqrt{\ln\ln(m)}}\right)^{\frac{2}{(1+4s)}}.$$

By the assumptions, we therefore have $\rho^\star \geq 1$. If there exists a $j' \in \mathcal{M}_m$ such that $\rho^\star \leq 2^{j'}$, then

$$\inf_{j\in\mathcal{M}_m}\left\{R^2 2^{-2js} + \frac{\sigma_\epsilon^2}{m}\sqrt{2^j \ln\ln(m)}\right\} \leq 2\frac{\sigma_\epsilon^2}{m}\sqrt{2^{j'}\ln\ln(m)} \leq 2\sqrt{2}R^{\frac{2}{(1+4s)}}\left(\frac{\sigma_\epsilon^2}{m}\sqrt{\ln\ln(m)}\right)^{\frac{4s}{(1+4s)}}. \tag{23}$$

Otherwise, take $j' \in \mathcal{M}_m$ such that $m/4 \leq 2^{j'} \leq m/2$. Since $2^{j'} \leq \max(\rho^\star, m/2)$, we obtain the upper bound

$$\inf_{j\in\mathcal{M}_m}\left\{R^2 2^{-2js} + \frac{\sigma_\epsilon^2}{m}\sqrt{2^j \ln\ln(m)}\right\} \leq 2R^2 2^{-2j's} \leq 2R^2\left(\frac{m}{4}\right)^{-2s}.$$

Since $s \leq 1$, we obtain

$$\inf_{j\in\mathcal{M}_m}\left\{R^2 2^{-2js} + \frac{\sigma_\epsilon^2}{m}\sqrt{2^j \ln\ln(m)}\right\} \leq CR^2 m^{-2s}. \tag{24}$$

From (22), (23) and (24), the requested inequality is proved. □

# References


[1] Abramovich, F. & Angelini, C. (2006). Testing in mixed-effects FANOVA models. *Journal of Statistical Planning and Inference* (to appear).

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

[3] Abramovich, F., Antoniadis, A., Sapatinas, T. & Vidakovic, B. (2004). Optimal testing in a fixed-effects functional analysis of variance model. *International Journal of Wavelets, Multiresolution and Information Processing*, **2**, 323–349.

[4] Alexe, G., Alexe, S., Liotta, L.A., Petricoin, E., Reiss, M. & Hammer, P.L. (2004). Ovarian cancer detection by logical analysis of proteomic data. *Proteomics*, **4**, 766–783.

[5] Antoniadis, A. (1994). Smoothing noisy data with coiflets. *Statistica Sinica*, **4**, 651–678.

[6] Baggerly, K.A., Morris, J.S., Edmondson, S.R. & Coombes, K.R. (2005). Signal in noise: evaluating reported reproducibility of serum proteomics tests for ovarian cancer. *Journal of the National Cancer Institute*, **97**, 307–309.

[7] Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, **8**, 577–606.

[8] Baraud, Y., Huet, S. & Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *Annals of Statistics*, **31**, 225–251.

[9] Cahouët, V., Martin, L. & Amarantini, D. (2002). Static optimal estimation of joint accelerations for inverse dynamic problem solution. *Journal of Biomechanics*, **35**, 1507–1513.

[10] Claeskens, G. (2004). Restricted likelihood ratio lack-of-fit tests using mixed spline models. *Journal of the Royal Statistical Society*, Series B, **66**, 909–926.

[11] Crainiceanu, C.M. & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society*, Series B, **66**, 165–185.

[12] Dahmen, W., DeVore, R. & Scherer, K. (1980). Multidimensional spline approximation. *SIAM Journal of Numerical Analysis*, **17**, 380–402.

[13] Delyon, B. & Juditsky, A. (1996). On minimax wavelet estimators. *Applied and Computational Harmonic Analysis*, **3**, 215–228.

[14] Di Zio, M. & Frigessi, A. (1999). Smoothness in Bayesian non-parametric regression with wavelets. *Methodology and Computing in Applied Probability*, **1**, 391–405.

[15] Donoho, D.L. & Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *Annals of Statistics*, **26**, 879–921.

[16] Durban, M., Harezlak, J., Wand, M.P. & Carroll, R.J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, **24**, 1153–1167.

[17] Green, P.J. (1999). Disussion of "The analysis of designed experiments and longitudinal data by using smoothing splines" by Verbyla, *et al. Applied Statistics*, **48**, 304–305.

[18] Green, P.J. & Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models.* London: Chapman & Hall.

[19] Guo, W. (2002a). Inference in smoothing spline analysis of variance. *Journal of the Royal Statistical Society*, Series B, **64**, 887–898.

[20] Guo, W. (2002b). Functional mixed effects models. *Biometrics*, **58**, 121–128.

[21] Horowitz, J.L. & Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**, 599–631.

[22] Ingster, Yu.I. & Suslina, I.A. (2003). *Nonparametric Goodness-of-Fit Testing under Gaussian Models*. *Lecture Notes in Statistics*, **169**, New York: Springer-Verlag.

[23] Laird, N.M. & Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.

[24] Liotta, L.A., Lowenthal, M., Mehta, A., Conrads, T.P., Veenstra, T.D., Fishman, D.A. & Petricoin, E.F. (2005). Importance of communication between producers and consumers of publicly available experimental data. *Journal of the National Cancer Institute*, **97**, 310–314.

[25] Mallat, S.G. (1999). *A Wavelet Tour of Signal Processing*. 2nd ed. San Diego: Academic Press.

[26] Morris, J.S. & Carroll, R.J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society*, Series B, **68**, 179–199.

[27] Ngo, L. & Wand, M.P. (2004). Smoothing with mixed model software. *Journal of Statistical Software*, **9**, Issue 1, 1–54.

[28] Patterson, H.D. & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

[29] Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone C., Fishman, D.A., Kohn, E.C. & Liotta, L.A. (2002). Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, **359**, 572–577.

[30] Ramsay, J.O. & Silverman, B.W. (1997). *Functional Data Analysis*. New York: Springer-Verlag.

[31] Rice, J.A. (2004). Functional and longitudinal data analysis: perspectives on smoothing. *Statistica Sinica*, **14**, 631–647.

[32] Rice, J.A. & Silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society*, Series B, **53**, 233–243.

[33] Rice, J.A. & Wu, C.O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253–259.

[34] Self, S.G. & Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–610.

[35] Speed, T.P. (1991). Discussion of "That BLUP is a good thing: the estimation of random effects", by G.K. Robinson. *Statistical Science*, **6**, 42–44.

[36] Spokoiny, V.G. (1996). Adaptive hypothesis testing using wavelets. *Annals of Statistics*, **24**, 2477–2498.

[37] Wang, Y. (1998). Mixed-effects smoothing spline ANOVA. *Journal of the Royal Statistical Society*, Series B, **60**, 159–174.