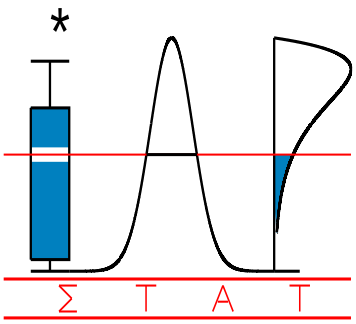
T E C H N I C A L

R E P O R T

0459

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. Wavelet decomposition approaches are used to model both fixed-effects and random-effects in the same functional space. This help us in interpreting the resulting model as a functional data model since it does not contradict the intuition that, if each outcome is a curve, which is the basic unit in functional data analysis, then the population-average curve and the subject-specific curves should have the same smoothness property (i.e., they should lie in the same functional space). 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). We also apply wavelet-based estimation procedures for both fixed-effects and random-effects in the general functional mixed-effects model. We illustrate the usefulness of the proposed estimation and testing procedures by means of two real-life datasets arising from endocrinology and physiology.

KEYWORDS: Functional Data; Functional Mixed-Effects Models; Functional Spaces; Linear Mixed-Effects Models; Linear Wavelet Estimation; Longitudinal Data; Nonlinear Wavelet Estimation; Nonparametric Estimation; Nonparametric Hypothesis Testing; Nonparametric Mixed-Effects Models; Nonparametric Model Selection; Smoothing Spline Estimation; Wavelets.

1 Introduction

There is a growing interest to incorporate *complex designs* and *correlation structures* in *standard* nonparametric and semiparametric regression models. Most of the recent developments in this direction are for *longitudinal data* studies used in many fields of research, including epidemiology, clinical trials, and survey sampling (see, e.g., Diggle *et al.*, 1994; Verbeke & Molenberghs, 2000). Longitudinal data are collected either prospectively, following subjects forward in time, or retrospectively, by extracting multiple measurements on each subject from historical records. They are characterized by the dependence of repeated observations over time within the same subject, and this correlation should be taken into account to draw valid scientific inferences.

Nowadays, another form of data, called *functional data* (see, e.g., Ramsay & Silverman, 1997, 2002), 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.

Although both longitudinal data analysis and functional data analysis are concerned with the analysis of data consisting of repeated measurements of subjects over time, there some important differences, perhaps due to the structural differences of the involved measurements. Longitudinal data usually involve small number of repeated measurements per subject, taken typically at sparsely, and often irregularly, spaced time points for different subjects. Functional data, in contrast, tend to involve a larger 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 longitudinal data analysis and functional data analysis are also often somewhat different, partly because to differences in the nature of scientific questions that are being addressed. The aims of longitudinal data analysis have a stronger inferential component, whereas those of functional data analysis tend to be more exploratory - to represent and display data in order to highlight interesting characteristics, perhaps as input for further analysis (see, e.g., Chapter 1 in Diggle *et al.*, 1994; Chapter 1 in Ramsay & Silverman, 1997). Despite these differences in focus, there are many common 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 longitudinal or 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. For example, in a longitudinal hormone study on progesterone (see Zhang *et al.*, 1998), the progesterone level varies over time in a complicated manner and it is difficult to model its trend using a simple parametric function. It is therefore of potential interest to model the time-effect nonparametrically

while accounting for the correlation of observations within the same subject. Extensions of linear mixed-effects models by including nonparametric fixed-effects and parametric random-effects have been considered by many researchers, including Wypij *et al.* (1993) and Wang & Taylor (1995) (regression splines were used to model fixed-effects); Hart & Wehrly (1986) and Zeger & Diggle (1994) (kernel methods were used to model time-trends); Barry (1996) (an empirical Bayes method was considered to model fixed-effects); and Wang (1998a,b), Zhang *et al.* (1998), Verbyla *et al.* (1999) and Guo (2002a) (cubic smoothing splines were used to model fixed-effects). 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 longitudinal or functional data analysis models have also been proposed. For direct methods we refer to, for example, Rice & Silverman (1991) (the within-subject covariance was modelled as function of the first few eigenfunctions with random amplitudes) while for indirect methods we refer to, for example, Hoover *et al.* (1998) and Wu *et al.* (1998) (the within-subject covariance was taken into account via a bandwidth selection by using a leave-one-subject-out cross-validation approach). The limitation of these approaches is that the population-average curve and the subject-specific curves do not have the same smoothness property because the fixed-effects and the random-effects are not modelled by a unified approach. Obviously, this causes difficulty in interpreting the resulting model as a functional data model since it contradicts the intuition that, if each outcome is a curve (which is the basic unit in functional data analysis), then the population-average curve and the subject-specific curves should have the same smoothness property (i.e., they should lie in the same functional space). The above limitation was, however, addressed by Shi *et al.* (1996) and Rice & Wu (2001) (cubic B-splines were used to model both the population-average curve as well as the subject-specific curves), and by Wu & Zhang (2002) (local-polynomials were used to model both the population-average curve as well as the subject-specific curves).

Although much work has been done on the *estimation* in various functional mixed-effects models, as discussed above, 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). In this model, both fixed-effects and random-effects are modelled in the same functional space, and therefore they both have the same smoothness property. The idea behind this 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) and later used by, amongst others, Brumback & Rice (1998), Wang (1998a,b), Zhang *et al.* (1998), Verbyla *et al.* (1999) and Guo (2002a). Adopting the formulation of Wahba (1978) which was used to show that the estimate of a cubic smoothing spline can be obtained through the posterior estimate of a partially integrated Wiener process, it can be shown that the posterior means of the fixed-effects (based on all the data) and the posterior means of the random effects (based only on

the data from the specific subject) are both cubic smoothing splines which share the same degree of smoothness. Furthermore, conditional on the restricted (or residual) maximum likelihood (RML) (or generalised maximum likelihood (GML) in the spline smoothing literature) estimates and other variance components, the resulting cubic smoothing splines estimates are simply the best linear unbiased predictors (BLUP). Although computationally efficient for small data sets, the resulting estimation approach is not computationally efficiently for large data sets, since it requires inversion of high-dimensional matrices. Alternatively, Guo (2002b) proposed a sequential estimation procedure that exploits the state-space representation of a cubic smoothing spline (see Wecker & Ansley, 1998) and uses the modified Kalman filtering and smoothing algorithm of Koopman & Durbin (2000) to estimate the smoothing parameters and the functional effects. This approach can be efficiently implemented (by avoiding inversion of high-dimensional matrices) and, therefore, can be applied to large data sets. 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). Furthermore, RML estimation was suggested as a unified criterion for estimation, model selection, and inference. 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 ∞ 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). Furthermore, although a conceptual connection between cubic smoothing splines methods and Bayesian formulation of inference about functions is often made in the literature, there always seems to be an implicit or explicit warning not to take this connection too literately. Formal use of such connections is rarely made, a notably exception being the Bayesian connection to construct confidence intervals about spline estimates, as proposed in Wahba (1983), since the frequentist properties of such intervals are well-known to be problematical (see Green, 2000). 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. We point out that a general functional mixed-effects model, similar to the one studied by Guo (2002b), has also been recently studied by Morris & Carroll (2004). Their methodology is based on fully Bayesian wavelet-based approaches, 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, a proper testing methodology for both fixed-effects and random-effects is lacking

from the analysis described in Morris & Carroll (2004).

In this paper, our aim is to study both *estimation* and *inference* in a general functional mixed-effects model through wavelet decomposition approaches. 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, 1982). In the former context, the smoothing parameters are usually based on minimising the mean integrated squared error, thus balancing the trade-off between bias and variance of the corresponding nonparametric estimates, while, in the latter context, the 'optimal' smoothing rules are defined so that contiguous alternative hypotheses with the fastest possible rate of convergence to the null hypothesis can be detected consistently, thus leading to powerful testing procedures.

We now introduce the two real-life datasets that have motivated our methodological thinking, and which we analyse in this paper.

Progesterone Data: Urinary metabolite progesterone curves measured over patients with healthy reproductive function were obtained by the Institute of Toxicology and Environmental Health at the University of California at Davis, USA (see Munro *et al.*, 1991). The data consist of a sample of 69 nonconceptive and 22 conceptive women's menstrual cycles. As is standard practice in endocrinological research, progesterone profiles are aligned by the day of ovulation, here determined by serum luteinizing hormone, then truncated at the end around the day of ovulation to present curves of the same length, so that all individual curves possess 24 scheduled distinct points. There are about 8.3% missing data. The data are highly correlated; the correlation coefficients in the nonconceptive and conceptive groups are bigger than 0.70 and 0.50 respectively. Figure 5.1 shows the urinary metabolite progesterone curves for the nonconceptive and conceptive groups (12 curves have been selected from each group) together with their means. The data have been analysed by, e.g., Brumback & Rice (1998), Fan & Zhang (2000) and Wu & Zhang (2002), and our analysis will be considered in detail in Section 5. One of the aim of the analysis is to characterize differences in conceptive and nonconceptive menstrual cycles prior to implantation, which is typically done a week after ovulation.

Orthosis Data: Human movement data were acquired and computed at the Laboratoire Sport et Performance Motrice, Grenoble University, France (see Cahouët *et al.*, 2002). The purpose of recording such data was the interest to better understand the processes underlying movement generation under various levels of an externally applied moment to the knee. 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 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 5.6 shows the available data set; typical moment plots over gait cycles. The data have been analysed by Abramovich & Angelini (2003) and Abramovich *et al.* (2004), and our analysis will be considered in detail in Section 5. 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.

The paper is organised as follows. In Section 2, we first provide a formulation for a general functional mixed-effects model and discuss some special cases of interest. We then briefly recall some relevant facts about the wavelet series expansion and the discrete wavelet transform that we need further. Wavelet decomposition approaches are then developed to model both fixed-effects and random-effects in the same functional space. This help us in interpreting the resulting model as a functional data model. 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 3, we provide hypothesis testing procedures for both fixed-effects and random-effects in the general functional mixed-effects model. In particular, the hypothesis testing procedures for random-effects (testing whether the random-effects functional components are equal to zero) are based on adapting recent work by Crainiceanu & Ruppert (2004a) who have derived finite sample and asymptotic null distributions for the LR and RLR test statistics for testing the null hypothesis that a variance component is zero, in linear mixed-effects model with one variance component, against that it is positive. On the other hand, the hypothesis testing procedures for fixed-effects (testing whether certain fixed-effects functional components or contrasts are equal to zero) are based on adapting recent work by Baraud *et al.* (2003) who have derived a nonasymptotic procedure, based on model selection methods, for testing the null hypothesis that the expectation of a Gaussian vector with n independent components belongs to a certain linear subspace of $\mathbb{R}^n$ against a nonparametric alternative. By adapting the recent work of Durot & Rozenholc (2004), these hypothesis testing procedures can be also used for a vector that is not assumed Gaussian nor having identically distributed components but consisting of random variables with symmetrical distributions. In Section 4, we apply wavelet-based estimation 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 estimation and testing procedures by applying them on the progesterone and orthosis datasets described above. Some concluding remarks are made in Section 6. Finally, in the Appendix, we briefly recall some relevant facts about the *penalized* and *regularized* linear

wavelet estimators (Appendix – A1) and the (inhomogeneous) Besov spaces on the unit interval (Appendix – A2), necessary to understand some of the technical details used in the development of the proposed methodology, as well as outlines of the proofs of the theoretical results obtained in earlier sections (Appendix – A3).

2 Functional Mixed-Effects Models

In this section, we provide a formulation for a general functional mixed-effects model and discuss some special cases of interest. Wavelet decomposition approaches are then developed to model both fixed-effects and random-effects in the same functional space. A linear mixed-effects representation, that is subsequently used for estimation and inference in the general functional mixed-effects model, is also derived.

2.1 The General Setup And Some Special Cases

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 *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 processes), $\mathbf{X}_{ij} = (X_{ij}[1], \ldots, X_{ij}[p])$ 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 ϵ_{ij} are independent and identically distributed Gaussian random variables (independent of $\boldsymbol{a}(t)$) with zero-mean and variance σ_ϵ^2, denoted by $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. Hereafter, "$^{\mathrm{T}}$" denotes the transpose of a vector or matrix.

Model (1) can be easily extended to accommodate (possibly different number of) repetitions per subject, say l_i $(i = 1, 2, \ldots, n)$, and will be considered in specific cases in subsequent sections. Also, as in Lin & Carroll (2000), where nonparametric regression estimation was considered for clustered data with a single covariate measured without error, we assume that when the design matrices $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ contain covariates they have been measured accurately.

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, some of them are mentioned below:

1. *Linear Mixed-Effects Models:* When the fixed functional components of $\boldsymbol{\beta}(t)$ are considered to be a $p \times 1$ vector $\boldsymbol{\beta}$ of *fixed* unknown population parameters and the random functional

components of $\boldsymbol{a}(t)$ are considered to be a $q \times 1$ vector $\boldsymbol{a}$ of *random* unknown individual effects modelled by zero-mean Gaussian distributions, then model (1) reduces to the classical linear mixed-effects model for longitudinal data studied by Laird & Ware (1982).

2. *Semiparametric and Nonparametric Longitudinal Data Models:* When the population-average curve profile $\mathbf{X}_{ij}\boldsymbol{\beta}(t)$ is taken to be the *single* element $\beta(t)$ that is modelled by either a parametric or a nonparametric approach, the i-th curve-specific deviation $\mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t)$ is taken to be the *single* element $U^{(i)} + W^{(i)}(t)$, where the $U^{(i)}$ are mutually independent and identically distributed zero-mean Gaussian random variables, and the $W^{(i)}(t)$ are mutually independent zero-mean stationary Gaussian processes, then model (1) reduces respectively to the so-called semiparametric or nonparametric longitudinal data model studied by Diggle *et al.* (1994, Section 5). Furthermore, if $\beta(t)$ is modelled by a cubic smoothing spline approach and $W^{(i)}(t)$ are mutually independent zero-mean standard Ornstein-Uhlenbeck or nonhomogeneous Ornstein-Uhlenbeck processes, then model (1) reduces to a special case of the nonparametric longitudinal data model studied by Zhang *et al.* (1998).

3. *Functional Regression Models, Functional Fixed-Effects Analysis of Variance Models, and Functional Analysis of Covariance Models:* When the random functional components of $\boldsymbol{a}(t)$ are set to zero, then model (1) reduces to the functional linear or varying-coefficient model, which includes functional regression models (if the design matrix $\mathbf{X}_{ij}$ contains covariates), functional fixed-effects analysis of variance models (if the design matrix $\mathbf{X}_{ij}$ contains dummy variables), and functional analysis of covariance models (if the design matrix $\mathbf{X}_{ij}$ contains both covariates and dummy variables). Although there is limited work on *fitting* functional regression models (see, e.g., Hastie & Tibshirani, 1993; Ramsay & Silverman, 1997, 2002; Fan & Zhang, 2000), and on *estimating* and *testing* the components in functional fixed-effects analysis of variance models (see, e.g., Fan & Lin, 1998; Dette & Derbort, 2001; Gu, 2002; Abramovich *et al.*, 2004; Cuevas *et al.*, 2004), we are not aware of any related work for functional analysis of covariance models.

4. *Nonparametric Mixed-Effects Models and Functional Mixed-Effects Analysis of Variance Models:* When the design matrices $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ are of dimension *one* and include only dummy variables, and both the fixed functional component $\beta(t)$ and the random functional component $a(t)$ are modelled in the same functional space (by a nonparametric approach), then model (1) reduces to the nonparametric mixed-effects model studied by Shi *et al.* (1996), Rice & Wu (2001), Wu & Zhang (2002) and Crainiceanu & Ruppert (2004b). In the case where the fixed functional component $\beta(t)$ and the random functional component $a(t)$ are not (necessarily) modelled in the same functional space, then model (1) reduces to the functional mixed-effects analysis of variance model studied by Abramovich & Angelini (2003).

5. *Smoothing Spline Mixed-Effects Analysis of Variance Models:* When the design matrix $\mathbf{X}_{ij}$ includes only dummy variables, $\boldsymbol{\beta}(t)$ is a p-dimensional nonparametric function lying in a reproducing kernel Hilbert space and possesses an ANOVA like decomposition (where some of

its terms could be set to zero to achieve parsimony), and the random functional components of $\boldsymbol{a}(t)$ are considered to be a $q \times 1$ vector $\boldsymbol{a}$ of *random* unknown individual effects modelled by zero-mean Gaussian distributions, then model (1) reduces to the smoothing spline mixed-effects analysis of variance model studied by Wang (1998b) and Guo (2002a).

6. *Nested and Crossed Samples of Curves Models:* When the design matrices $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ are appropriately used to create block-diagonal fixed-effects and block-diagonal random-effects design matrices respectively, model (1) reduces to the nested mixed-effects and crossed mixed-effects models studied by Brumback & Rice (1998). It also includes the nested hierarchical functional model considered by Morris *et al.* (2003).

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). Furthermore, in order to make the population-average curve profile and subject-specific deviation have the same smoothness assumptions, the fixed elements of $\boldsymbol{\beta}(t)$ and the random elements of $\boldsymbol{a}(t)$ should both lie in the same functional space. This can be accomplished, e.g., if the fixed elements of $\boldsymbol{\beta}(t)$ and the random elements of $\boldsymbol{a}(t)$ are both modelled as cubic smoothing splines, as proposed in Guo (2002b). In addition to the drawbacks discussed in Section 1, this formulation also restricts the smoothness of the underlying functions since it is well known that cubic smoothing splines are optimal representations only for functions lying in Hilbert-Sobolev spaces of positive integer regularity index. However, deviations from smooth effects may be present and this behaviour should also be included in the modelling formulation (see, e.g., Morris & Carroll, 2004). If replicates are available at the design points, then the presence of non-smooth effects can be possible accommodated by fitting a cubic smoothing spline to the means at the replicated values, as suggested by Verbyla *et al.* (1999). A more natural framework to include non-smooth effects is, however, 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.

2.2 The wavelet series expansion

Throughout the article 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

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

where

$$\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k), \quad \psi_{jk}(t) = 2^{j/2}\psi(2^j t - k).$$

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 $f(t) \in L^2[0, 1]$, we denote by $u_{j_0k} = \langle f(t), \phi_{j_0k}(t)\rangle$ $(k = 0, 1, \ldots, 2^{j_0} - 1)$ the scaling coefficients and by $w_{jk} = \langle f(t), \psi_{jk}(t)\rangle$ $(j \geq j_0;\ k = 0, 1, \ldots, 2^j - 1)$ the wavelet coefficients of $f(t)$ for the orthonormal periodic wavelet basis defined above; the function $f(t)$ is then expressed in the form

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

For detailed expositions of the mathematical aspects of wavelets we refer to, e.g., Meyer (1992), Daubechies (1992) and Mallat (1999), while comprehensive expositions and reviews on wavelets applications in statistical settings are given in, e.g., Antoniadis (1997), Vidakovic (1999), Abramovich *et al.* (2000) and Antoniadis *et al.* (2001).

2.3 The discrete wavelet transform

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 developed by Mallat (1989) that requires only order n operations. In this case, for a given primary resolution level j_0 and under periodic boundary conditions, the DWT of $\mathbf{f}$ results in an n-dimensional vector $\mathbf{d}$ comprising both discrete scaling coefficients $c_{j_0k},\ k = 0, ..., 2^{j_0} - 1$ and discrete wavelet coefficients $d_{jk},\ j = j_0, ..., J - 1;\ k = 0, ..., 2^j - 1$. We do not provide technical details here of the order n DWT algorithm mentioned above. Essentially

the algorithm is a fast hierarchical scheme for deriving the required inner products which at each step involves the action of low and high pass filters, followed by a decimation (selection of every even member of a sequence). The IDWT may be similarly obtained in terms of related filtering operations. For a nice account of the DWT and IDWT in terms of filter operators we refer to, e.g., Nason & Silverman (1995).

2.4 Wavelet-based Model Specifications for the Fixed and Random Effects: A Karhunen-Loève Representation

Recall that Guo (2002b), adopting the formulation of Wahba (1978) which was used to show that the estimate of a cubic smoothing spline can be obtained through the posterior estimate of a partially integrated Wiener process, has shown that the posterior means of the fixed-effects (based on all the data) and the posterior means of the random effects (based only on the data from the specific subject) are both cubic smoothing splines which share the same degree of smoothness. Furthermore, conditional on the various parameters, the resulting cubic smoothing splines estimates are simply the BLUP.

Similar formulations can also be developed through wavelet decompositions by modelling the fixed-effects as a *single* realisation of a partially diffuse Karhunen-Loève representation, while the random-effects are modelled as *random* realisations from the same partially Karhunen-Loève representation with appropriate normalisation. More specifically, for each $r_1 = 1, 2, \ldots, p$, let the fixed functional components $\beta_{r_1}(t)$ of $\boldsymbol{\beta}(t)$ be modelled as

$$\beta_{r_1}(t) = \sum_{k=0}^{m-1} b_k^{(r_1)} \zeta_k(t) + \delta_\beta^{(r_1)} W^{(r_1)}(t), \quad t \in [0,1] \tag{2}$$

and, for each $r_2 = 1, 2, \ldots, q$, let the random functional components $a_{r_2}(t)$ of $\boldsymbol{a}(t)$ be modelled as

$$a_{r_2}(t) = \sum_{k=0}^{m-1} c_k^{(r_2)} \zeta_k(t) + \delta_\alpha^{(r_2)} W^{(r_2)}(t), \quad t \in [0,1], \tag{3}$$

where $\zeta_k(t)$ are known linearly independent squared-integrable functions, $b_k^{(r_1)}$ are independent and identically distributed $N(0, \tau_1^2)$ random variables with $\tau_1^2 \to 0$ (a vague prior), $\mathbf{c}^{(r_2)} = (c_0^{(r_2)}, \ldots, c_{m-1}^{(r_2)})^{\mathrm{T}} \sim N(0, \tau_{2,r_2}^2 D)$, for some fixed $0 < \tau_{2,r_2}^2 < \infty$ and a nonnegative-definite covariance matrix D (a proper prior), $\delta_\beta^{(r_1)}$ and $\delta_\alpha^{(r_2)}$ are unknown positive constants, and $W^{(r_1)}(t)$, $W^{(r_2)}(t)$ are both copies of a second-order zero-mean stochastic process $Z(t)$ with covariance kernel $\mathbb{E}(Z(s)Z(t)) = \mathcal{W}(s,t)$.

The covariance kernel $\mathcal{W}(s,t)$ is assumed to satisfy the conditions: $\int_0^1 \mathcal{W}(t,t)dt < \infty$ and $\int_0^1 \int_0^1 [\mathcal{W}(s,t)]^2 ds dt < \infty$. The first condition ensures that the sample paths of $Z(t)$ (and hence of $\beta_{r_1}(t)$ and $a_{r_2}(t)$) are in $L^2[0,1]$ almost surely, while the second condition ensures that $\mathcal{W}(s,t)$ has an eigenfunction-eigenvalue decomposition by the Hilbert-Schmidt theorem. That is, there exist functions $\xi_\nu(t)$, $\nu = 1, 2, \ldots$, and a non-increasing sequence of non-negative numbers $\lambda_1 \geq \lambda_2 \ldots \geq 0$ such that $\mathcal{W}(s,t) = \sum_\nu \lambda_\nu \xi_\nu(s) \xi_\nu(t)$. Thus, the process $Z(t)$ has the so-called Karhunen-Loève representation, i.e., $Z(t) \sim \sum_\nu \gamma_\nu \xi_\nu(t)$, where "$\sim$" means "equal in distribution", and $\gamma_\nu(t)$,

$\nu = 1, 2, \ldots$, is a sequence of uncorrelated random variables with zero means and variances λ_ν, $\nu = 1, 2, \ldots$. Observe that each of the representations (2) and (3) is a *nonparametric mixed-effects model* itself; the coefficients $b_k^{(r_1)}$ and $c_k^{(r_2)}$ are the "fixed-effects" and the random-coefficients $\gamma_\nu(t)$ are the "random-effects" (see Huang & Lu, 2000; Angelini *et al.*, 2003).

Fix now a primary resolution level $j_0 \geq 0$, take $m = 2^{j_0}$ and consider the orthonormal periodic wavelet basis $\{\phi_{j_0k}(t),\ k = 0, 1, \ldots, 2^{j_0} - 1;\ \psi_{jk}(t),\ j \geq j_0;\ k = 0, 1, \ldots, 2^j - 1\}$ discussed in Section 2.2 as our choice of basis for $(\zeta_k(t), \xi_\nu(t))$ $(k = 0, 1, \ldots, m-1; \nu = 1, 2, \ldots)$. Based on this choice of basis, representation (2), for the fixed functional components $\beta_{r_1}(t)$ of $\boldsymbol{\beta}(t)$, takes the form

$$\beta_{r_1}(t) = \sum_{k=0}^{2^{j_0}-1} b_k^{(r_1)} \phi_{j_0k}(t) + \delta_\beta^{(r_1)} \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \gamma_{jk} \psi_{jk}(t), \quad t \in [0, 1] \tag{4}$$

while representation (3), for the random functional components $a_{r_2}(t)$ of $\boldsymbol{a}(t)$, takes the form

$$a_{r_2}(t) = \sum_{k=0}^{2^{j_0}-1} c_k^{(r_2)} \phi_{j_0k}(t) + \delta_\alpha^{(r_2)} \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \gamma_{jk} \psi_{jk}(t), \quad t \in [0, 1], \tag{5}$$

where γ_{jk} are uncorrelated random variables with zero means and $\mathbb{V}(\gamma_{jk}) = \lambda_j$.

Theorem 2.1 *Under the above modelling methodology, assume further that γ_{jk} are independent random variables, i.e., they are Gaussian random variables (which is equivalent to assuming that $Z(t)$ is a centered Gaussian process), and that D is a diagonal matrix. Then, for each $r_1 = 1, 2, \ldots, p$,*

$$\hat{\beta}_{r_1}(t) := \lim_{\tau^2 \to \infty} \mathbb{E}(\beta_{r_1}(t) \mid \mathbf{Y}) = \tilde{f}_{\lambda_\beta}(t)$$

and, for each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$,

$$\hat{\alpha}_{r_2}^{(i)}(t) := \mathbb{E}(a_{r_2}(t) \mid \mathbf{Y}_i) = \tilde{f}_{\lambda_\alpha}^{(i)}(t),$$

where $\mathbf{Y}$ is all the data Y_{ij} $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m_i)$, $\mathbf{Y}_i$ is the data from the ith subject $(i = 1, 2, \ldots, n)$, and $\tilde{f}_{\lambda_\beta}(t)$, $\tilde{f}_{\lambda_\alpha}^{(i)}(t)$ are the penalized linear wavelet estimators *for appropriate smoothing parameters $\lambda_\beta,\ \lambda_\alpha > 0$.*

Theorem 2.1 ensures that $\hat{\beta}_{r_1}(t)$ and $\hat{\alpha}_{r_2}^{(i)}(t)$ are both penalized linear wavelet estimators (see Appendix – A1) and, obviously, the subject-specific prediction curve $\mathbf{X}_{ij}\hat{\boldsymbol{\beta}}(t) + \mathbf{Z}_{ij}\hat{\boldsymbol{\alpha}}^{(i)}(t)$ is also a penalized linear wavelet estimator. The differences between the fixed functional components of $\boldsymbol{\beta}(t)$ and the random functional components of $\boldsymbol{a}(t)$ are that (i) each fixed functional component of $\boldsymbol{\beta}(t)$ is modelled as a single realisation of a partially diffuse Karhunen-Loève representation, while the random functional components of $\boldsymbol{a}(t)$ are modelled as *random* realisations from the same partially Karhunen-Loève representation with appropriate normalisation, and (ii) each fixed functional components of $\boldsymbol{\beta}(t)$ is estimated by its posterior mean conditional on all data, while each random functional component of $\boldsymbol{a}(t)$ is estimated as a posterior mean conditional on the i-th subject only. As each realisation of a Karhunen-Loève representation is a curve whose smoothness property is characterised by the covariance kernel of the underlying second-order zero-mean stochastic process,

the random functional components of $\boldsymbol{a}(t)$ for different subjects (different i's) share the same degree of smoothness because they share the same correlation structures. Therefore, the fixed functional components of $\boldsymbol{\beta}(t)$ and the random functional components of $\boldsymbol{a}(t)$ both share the same degree of smoothness, i.e., they lie in the same functional space, the Hilbert-Sobolev space, $H_2^s[0,1]$, with noninteger regularity index $s > 1/2$ (see Appendices – A1 and A2).

Note that the above wavelet-based representation extends the results of Guo (2002b) since the penalized linear wavelet estimator generalizes the corresponding nonparametric regression estimation problem over Hilbert-Sobolev spaces, $H_2^s[0,1]$, with noninteger regularity index $s > 1/2$. Note that this space consists of relatively smooth functions, but not as smooth as the usual Hilbert-Sobolev spaces with integer regularity index $s \geq 1$ used in cubic spline smoothing (see, e.g., Antoniadis, Grégoire & McKeague, 1994). Furthermore, conditional on the parameters and working along the lines given in Theorems 3.1 and 3.2 of Huang & Lu (2000) or Theorem 4.2 of Angelini *et al.* (2003), the resulting penalized linear wavelet estimator in each case is the BLUP. This connection can therefore be used to develop an estimation procedure similar to the one developed by Guo (2002b) for cubic smoothing splines (and it has been applied in the analysis of the examples presented in Section 5). However, as in cubic spline smoothing, we point out that the conceptual connection between linear wavelet methods and Bayesian formulation of inference about functions should not be taken too literately.

2.5 Wavelet-based Model Specifications for the Fixed and Random Effects: A Sequence-Space Representation

A different approach to modelling the fixed and random effects, that allows a wide range of irregular effects (for both fixed-effects and random-effects) on the one hand and overcomes the drawback mentioned in Section 2.4 on the other hand, 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 ρ_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^{ρ_1}-sense; the additional parameter ρ_2 is secondary in its role, allowing for additional fine tuning of the definition of the space (see Appendix – A2).

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, among others, Di Zio & Frigessi (1999) and Botchkina (2002), 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 $\{\phi_{j_0 k}(t),\ k = 0, 1, \ldots, 2^{j_0} - 1;\ \psi_{jk}(t),\ j \geq j_0;\ k = 0, 1, \ldots, 2^j - 1\}$ discussed in Section 2.2. For each $r_1 = 1, 2, \ldots, p$, assume that $\beta_{r_1}(t) \in B^s_{\rho_1,\rho_2}[0,1]$ 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

$$\alpha_{r_2}^{(i)}(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0 k}^{(r_2,i)} \phi_{j_0 k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \theta_{jk}^{(r_2,i)} \psi_{jk}(t), \quad t \in [0,1],$$

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

$$\theta_{jk}^{(r_2,i)} \sim \pi_{jk}^{(r_2,i)} N(0, v_{jk}^{(r_2,i)}) + (1 - \pi_{jk}^{(r_2,i)})\delta(0),$$

where $\delta(0)$ is a point mass at zero. Furthermore, we assume that, for each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, the quantities $\pi_{jk}^{(r_2,i)}$ and $v_{jk}^{(r_2,i)}$ 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.,

$$\pi_{jk}^{(r_2,i)} = \min(1, C_\theta 2^{-j\beta_i}), \text{ for some } \beta_i \geq 0,$$

and

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

where C_θ and σ_θ^2 are some positive quantities.

A relationship between the Besov space parameters and the hyperparameters of the prior model considered above can be now established. The interesting cases correspond to considering $0 \leq \beta_i \leq 1$ (see Section 4.1 in Abramovich *et al.*, 1998). Then, exploiting the equivalence between the Besov norm of the function $\alpha_{r_2}^{(i)}(t)$ and the corresponding sequence space norm (see Appendix – A2), and using Theorem 1 in Abramovich *et al.* (1998), for any given values of $c_{j_0 k}^{(r_2,i)}$ $(k = 0, 1, \ldots, 2^{j_0} - 1)$, and for each $i = 1, 2, \ldots, n$; $r_2 = 1, 2, \ldots, q$,

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

if and only if

$$s + 1/2 - \beta_i/\rho_1 - \alpha_i/2 < 0$$

or

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

In order to take into account the Besov space parameter ρ_2 as well, we can also allow a more delicate dependence on the variance parameter $v_{jk}^{(r_2,i)}$ by introducing a second parameter $\gamma_i \in \mathbb{R}$, i.e., $v_{jk}^{(r_2,i)} = \sigma_\theta^2 2^{-j\alpha_i} j^{\gamma_i}$. By following Theorem 2 in Abramovich *et al.*, 1998), for any given values of $c_{j_0 k}^{(r_2,i)}$ $(k = 0, 1, \ldots, 2^{j_0} - 1)$, and for each $i = 1, 2, \ldots, n$; $r_2 = 1, 2, \ldots, q$,

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

if and only if

$$s + 1/2 - \beta_i/\rho_1 - \alpha_i/2 < 0$$

or

$$s + 1/2 - \beta_i/\rho_1 - \alpha_i/2 = 0$$

and γ satisfies the appropriate one of the following conditions:

(a) $\gamma < -2/\rho_2$ for (*i*) $\rho_1, \rho_2 < \infty$ and $0 \leq \beta_i < 1$, (*ii*) $\rho_1, \rho_2 < \infty$ and $\beta_i = 1$, and (*iii*) $\rho_1 = \infty$, $\rho_2 < \infty$ and $\beta_i = 1$;

(b) $\gamma < -1 - 2/\rho_2$ for $\rho_1 = \infty$, $\rho_2 < \infty$ and $0 \leq \beta_i < 1$;

(c) $\gamma \leq 0$ for $\rho_1 < \infty$, $\rho_2 = \infty$ and $0 \leq \beta_i < 1$;

(d) $\gamma \leq -1$ for $\rho_1, \rho_2 = \infty$ and $0 \leq \beta_i < 1$;

(e) $\gamma < 0$ for (*i*) $\rho_1 < \infty$, $\rho_2 = \infty$ and $\beta_i = 1$, and (*ii*) $\rho_1, \rho_2 = \infty$ and $\beta_i = 1$.

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 $0 < s < r$, $1 \leq \rho_1, \rho_2 \leq \infty$), by appropriately relating the Besov space parameters s, ρ_1 and ρ_2 to the hyperparameters α_i, β_i and/or γ_i of the prior models discussed above.

Remark 2.1 The hyperparameter $\pi_{jk}^{(r_2,i)}$ is the prior probability that a wavelet coefficient at resolution level j and scale k is important for representing the i-th random curve, and $v_{jk}^{(r_2,i)}$ is the prior variance of the important wavelet coefficient at resolution level j and scale k. This model allows us to adapt between sparse and dense random effects.

Remark 2.2 Note that, for each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, we have used the same mother wavelet $\psi(t)$ in the formulations given above. Different mother wavelets can be used to get the required characterizations as long as their regularities are larger than the regularity of the corresponding Besov space.

2.6 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). 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 of Mallat (1989) mentioned in Section 2.3. Note that this assumption is not especially restrictive, since if the grid is fine enough, interpolation can used to obtain a common grid (of power two) without substantively changing the observed data.

Fix now a primary resolution level $j_0 \geq 0$ and consider the orthonormal periodic wavelet basis $\{\phi_{j_0k}(t),\ k = 0, 1, \ldots, 2^{j_0} - 1;\ \psi_{jk}(t),\ j \geq j_0;\ k = 0, 1, \ldots, 2^j - 1\}$ discussed in Section 2.2. 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 $m^* = 2^{j_1} = [m/\ln(m)]$ 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 Theorems 1 and 2 in Abramovich *et al.* (1998) discussed in Section 2.5. 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 $r_1=1,2,\ldots,p$, $\beta_{r_1}(t)\in B^s_{\rho_1,\rho_2}[0,1]$ for $0<s_1<r$, $1\leq\rho_1,\rho_2\leq\infty$.

Similarly, 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 with probability $\pi_{jk}^{(r_2,i)}$ or a point mass at zero with probability $(1-\pi_{jk}^{(r_2,i)})$. By analogy to Section 2.5, we take $\pi_{jk}^{(r_2,i)} = \min(1, C_\theta 2^{-j\beta_i})$, for some $\beta_i\geq 0$, and $v_{jk}^{(r_2,i)} = \sigma_\theta^2 2^{-j\alpha_i}$ or $v_{jk}^{(r_2,i)} = \sigma_\theta^2 2^{-j\alpha_i} j^{\gamma_i}$, for some $\alpha_i\geq 0$ and $\gamma_i\in\mathbb{R}$, where σ_θ^2 and C_θ are some positive quantities. 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 $B^s_{\rho_1,\rho_2}[0,1]$, with $0<s<r$, $1\leq\rho_1,\rho_2\leq\infty$ (see Section 2.5). For identifiability reasons, that will be 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 $r_2=1,2,\ldots,q$, ϵ_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

$$\mathbf{Y} = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\epsilon}}, \tag{6}$$

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 (6) 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 $\min(1,C_\theta 2^{-j\beta_i})2^{-j\alpha_i}$ or $\min(1,C_\theta 2^{-j\beta_i})2^{-j\alpha_i}j^{\gamma_i}$ 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. Therefore,

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

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 λ can be considered a *signal-to-noise ratio* since σ_θ^2 determines the size of the "signal" given by $\mathbb{E}(\mathbf{Y} \mid \widetilde{\boldsymbol{\theta}}) = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\theta}}$ and σ_ϵ^2 determines the size of the "noise". Note that $\sigma_\theta^2 = 0$ if and only if $\lambda = 0$, and the parameter space for λ is $[0, \infty)$.

Remark 2.3 We point out that 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 which is typical for longitudinal or functional data.

Remark 2.4 An alternative model can be also built by taking different priors with hyperparameters α_i, β_i and γ_i for $i \in \{1, 2, \ldots, n_1\}$, $i \in \{n_1 + 1, n_1 + 2, \ldots, n_2\}$, ..., $i \in \{n_{p-1} + 1, n_{p-1} + 2, \ldots, n_p\}$, where $n_1 + n_2 + \ldots + n_p = n$. In other words, using similar notation, model (6) is now replaced by

$$\mathbf{Y} = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} + \widetilde{\mathbf{Z}}_1\widetilde{\boldsymbol{\theta}}_1 + \ldots + \widetilde{\mathbf{Z}}_q\widetilde{\boldsymbol{\theta}}_q + \tilde{\boldsymbol{\epsilon}}. \tag{7}$$

Model (7) 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 $\alpha_{r_2}^{(i)}(t)$ $(i = 1, 2, \ldots, n;\ r_2 = 1, 2, \ldots, q)$.

3 Inference in Functional Mixed-Effects Models

In this section, we provide hypothesis testing procedures for both fixed-effects and random-effects in the general functional mixed-effects model (1).

3.1 Testing for Random-Effects

According to the modelling formulation in Section 2.6, 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 (6) which, in turn, is equivalent to testing the following hypotheses

$$H_0 :\ \sigma_\theta^2 = 0\ (\lambda = 0) \quad \text{versus} \quad H_A :\ \sigma_\theta^2 > 0\ (\lambda > 0). \tag{8}$$

Testing the above hypotheses is non-standard because the parameter under the null hypothesis is on the boundary of the parameter space. Therefore, using 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 (6) is usually not a vector of independent random variables, the non-standard asymptotic theory of Self & Liang (1987) does not apply. Stram & Lee (1994) showed that the theory of Self & Liang (1987) can be still applied

to testing for a zero variance component in linear mixed-effects model with one variance component if the response vector $\mathbf{Y}$ can be partitioned into independent and identically distributed subvectors and the number of independent subvectors tends to infinity (while the number of observations per subvector can be either fixed or also increasing to infinity). However, they assumed that the random-effects are independent from subject to subject, and they implicitly assumed that the number of subjects increases to infinity. Their results would not hold for a fixed number of subjects, even if the observations per subject increased to infinity.

It is therefore evident that the non-standard asymptotic theory for LR and RLR tests cannot be applied blindly for testing the hypotheses in (8), since the stated assumptions on $\mathbf{Y}$ and on random-effects usually do not hold in the liner mixed-effects model (6). With this in mind, Crainiceanu & Ruppert (2004a) 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 (8).

3.1.1 Profile and Restricted Profile Log-Likelihood Functions

Twice the log-likelihood function for the linear mixed-effects model (6) is (up to a constant that does not depend on the parameters)

$$L(\widetilde{\mathbf{d}}, \lambda, \sigma_\epsilon^2) = -nm \log(\sigma_\epsilon^2) - \log|\mathbf{V}_\lambda| - \frac{(\mathbf{Y} - \widetilde{\mathbf{X}}\widetilde{\mathbf{d}})^{\mathrm{T}} \mathbf{V}_\lambda^{-1} (\mathbf{Y} - \widetilde{\mathbf{X}}\widetilde{\mathbf{d}})}{\sigma_\epsilon^2}. \tag{9}$$

Under the alternative hypothesis H_A in (8), by *fixing* λ and solving the first order minimum conditions for $\widetilde{\mathbf{d}}$ and σ_ϵ^2, one gets the maximum profile likelihood estimates

$$\widehat{\widetilde{\mathbf{d}}}(\lambda) = (\widetilde{\mathbf{X}}^T \mathbf{V}_\lambda^{-1} \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^{\mathrm{T}} \mathbf{V}_\lambda^{-1} \mathbf{Y} \tag{10}$$

and

$$\widehat{\sigma_\epsilon^2}(\lambda) = \frac{1}{nm} (\mathbf{Y} - \widetilde{\mathbf{X}}\widehat{\widetilde{\mathbf{d}}}(\lambda))^{\mathrm{T}} \mathbf{V}_\lambda^{-1} (\mathbf{Y} - \widetilde{\mathbf{X}}\widehat{\widetilde{\mathbf{d}}}(\lambda)). \tag{11}$$

Plugging the expressions (10) and (11) into (13) we obtain (up to a constant that does not depend on the parameters), the profile log-likelihood function

$$L(\lambda) = -\log|\mathbf{V}_\lambda| - nm \log(\mathbf{Y}^{\mathrm{T}} \mathbf{P}_\lambda^{\mathrm{T}} \mathbf{V}_\lambda^{-1} \mathbf{P}_\lambda \mathbf{Y}), \tag{12}$$

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 RML. RML consists of maximizing the likelihood function associated with $(nm - m^*p)$ linearly independent contrasts. It makes no difference which $(nm - m^*p)$ linearly independent contrasts are used because the likelihood function for any such set differs by no more than an additive constant (see Harville, 1977). Twice the restricted log-likelihood function for the linear mixed-effects model (6) is (up to a constant that does not depend

on the parameters)

$$l(\widetilde{\mathbf{d}}, \lambda, \sigma_\epsilon^2) = L(\widetilde{\mathbf{d}}, \lambda, \sigma_\epsilon^2) - m^* p \log(\sigma_\epsilon^2) - \log(|\widetilde{\mathbf{X}}^{\mathrm{T}} \mathbf{V}_\lambda^{-1} \widetilde{\mathbf{X}}|). \tag{13}$$

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 (10) while the maximum restricted profile likelihood estimate of $\sigma_\epsilon^2(\lambda)$ is now given by

$$\widehat{\sigma_\epsilon^2}(\lambda) = \frac{1}{nm - m^* p} (\mathbf{Y} - \widetilde{\mathbf{X}} \widehat{\widetilde{\mathbf{d}}}(\lambda))^{\mathrm{T}} \mathbf{V}_\lambda^{-1} (\mathbf{Y} - \widetilde{\mathbf{X}} \widehat{\widetilde{\mathbf{d}}}(\lambda)). \tag{14}$$

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

$$l(\lambda) = -\log|\mathbf{V}_\lambda| - \log|\widetilde{\mathbf{X}}^T \mathbf{V}_\lambda^{-1} \widetilde{\mathbf{X}}| - (nm - m^* p) \log\{\mathbf{Y}^{\mathrm{T}} \mathbf{P}_\lambda^{\mathrm{T}} \mathbf{V}_\lambda^{-1} \mathbf{P}_\lambda \mathbf{Y}\}. \tag{15}$$

By following Crainiceanu & Ruppert (2004a) and Claeskens (2004), and taking into account that $[n(\ln(m) - q)] > p$, 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 (12) 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 (15), 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 ω_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 $\mathbf{P}_0 = \mathbf{I}_{nm} - \widetilde{\mathbf{X}} (\widetilde{\mathbf{X}}^{\mathrm{T}} \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^{\mathrm{T}}$.

3.1.2 Finite Sample Null Distributions of the LR and RLR Tests

As in Crainiceanu & Ruppert (2004a), the finite sample LR test statistic is defined as

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

while the finite sample RLR test statistic is defined as

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

The following theorem gives the spectral representations of the finite sample null distributions of the LR_{nm} and RLR_{nm} test statistics that can be used for testing the hypotheses in (8).

Theorem 3.1 *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 (8),*

$$\mathrm{LR}_{nm} \stackrel{\mathcal{D}}{=} \sup_{\lambda \geq 0} \left[nm \log\left\{ 1 + \frac{N_{nm}(\lambda)}{D_{nm}(\lambda} \right\} - \sum_{s=1}^{nm^*q} \log(1 + \lambda\xi_{s,nm}) \right] \tag{16}$$

and

$$\mathrm{RLR}_{nm} \stackrel{\mathcal{D}}{=} \sup_{\lambda \geq 0} \left[(nm - m^*p) \log\left\{ 1 + \frac{N_{nm}(\lambda)}{D_{nm}(\lambda} \right\} - \sum_{s=1}^{nm^*q} \log(1 + \lambda\mu_{s,nm}) \right], \tag{17}$$

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

Remark 3.1 Each of the finite sample null distributions of the LR_{nm} and 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. By following Crainiceanu & Ruppert (2004a), the probabilities

$$\pi_0 = P\left(\frac{\sum_{s=1}^{nm^*q} \mu_{s,n}\omega_s^2}{\sum_{s=1}^{nm-m^*p} \omega_s^2} \leq \frac{1}{nm} \sum_{s=1}^{nm^*q} \xi_{s,n} \right) \quad \text{and} \quad \pi_1 = P\left(\frac{\sum_{s=1}^{nm^*q} \mu_{s,n}\omega_s^2}{\sum_{s=1}^{nm-m^*p} \omega_s^2} \leq \frac{1}{nm - m^*p} \sum_{s=1}^{nm^*q} \mu_{s,n} \right) \tag{18}$$

provide good approximations for the probability masses at zero of the finite sample null distributions for the LR_{nm} and RLR_{nm} test statistics respectively, and can be easily obtained by simulations.

The finite sample null distributions of the RL_{nm} and RLR_{nm} test statistics given by (16) and (17) respectively depend only on the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$. Following Crainiceanu & Ruppert (2004a), 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 LR_{nm} and RLR_{nm} test statistics, once the eigenvalues $\mu_{s,nm}$ and $\xi_{s,nm}$ have been calculated.

An Algorithm for the Finite Sample Null Distributions of the LR_{nm} and RLR_{nm} Test Statistics

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

Step 2. Simulate nm^*q independent χ_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 λ_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 (6) (which obviously depends on the number of subjects, n, the number of observations, m (recall that $m^* = [m/\ln(m)]$), per subject, and the number of random-effects, q, in the general functional mixed-effects model (1)).

Alternatively, the asymptotic null distributions of the LR_{nm} and 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 LR_{nm} and 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 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.

3.1.3 Asymptotic Null Distributions of the LR and RLR Tests

The following theorem provides the asymptotic null distributions of the LR_{nm} and RLR_{nm} test statistics.

Theorem 3.2 *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 (8),*

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

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

Remark 3.2 Each of the asymptotic null distributions of the LR_{nm} and 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. By following Crainiceanu & Ruppert (2004a), the probabilities

$$p_0 = P\left(\sum_{s=1}^{nm^*q} \mu_s \omega_s^2 \leq \sum_{s=1}^{nm^*q} \xi_s\right) \quad \text{and} \quad p_1 = P\left(\sum_{s=1}^{nm^*q} \mu_s \omega_s^2 \leq \sum_{s=1}^{nm^*q} \mu_s\right) \tag{19}$$

provide excellent approximations for the probability masses at zero of the asymptotic null distributions for the LR_{nm} and RLR_{nm} test statistics respectively.

Remark 3.3 The results of Theorem 3.2 can be expressed as

$$\mathrm{LR}_{nm} \Rightarrow p_0\delta(0) + (1-p_0)G_0 \quad \text{and} \quad \mathrm{RLR}_{nm} \Rightarrow p_1\delta(0) + (1-p_1)G_1$$

where G_0 and G_1 denote probability distributions that are generally different than χ^2_1, and p_0 and p_1 are the asymptotic probability masses at zero given in (19), which are typically different than 0.5. Although the form of these asymptotic null distributions are similar to the asymptotic null distribution $0.5\delta(0) + 0.5\chi^2_1$ that can be used for testing the variance component in the linear mixed-effects model (6), under the assumption of independent and identically distributed data $\mathbf{Y}$ (for all values of the parameter) or when the response vector $\mathbf{Y}$ can be partitioned into independent subvectors and the number of independent subvectors tends to infinity (see Stram & Lee, 1994), the probability masses at zero p_0 and p_1 and the second distributions G_0 and G_1 in the mixtures differ. The probabilities p_0 and p_1 can be computed either directly using well-known tabulated probability distributions or using an algorithm of Farebrother (1990) or using Monte Carlo (or other more efficient schemes to) simulations. These latter schemes to simulations could also be used to simulate quantiles of the distributions G_0 and G_1 if closed-form expressions are not available.

We have just seen that the asymptotic null distributions discussed above 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 3.1 `(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.,

$$Y_{ijk} = \beta(t_{ij}) + \alpha^{(i)}(t_{ij}) + \epsilon_{ijk}, \quad i = 1, \ldots, n;\ j = 1, \ldots, m;\ k = 1, \ldots, l, \tag{20}$$

where $\beta(t)$ is an unknown functional mean, $\alpha^{(i)}(t)$ are realisations of a zero-mean Gaussian process $a(t)$, and ϵ_{ijk} are independent and identically distributed $N(0, \sigma^2_\epsilon)$ random variables that are also independent of $a(t)$. Using the wavelet transform parameterisation discussed in Section 2.6, the

matrix $\widetilde{\mathbf{X}}$ for fixed-effects is simply an $nml \times m^*$ matrix with nl block columns equal to $W^{\mathrm{T}}_{m\times m^*}$ 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 $W^{\mathrm{T}}_{m\times m^*}$ 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 $W^{\mathrm{T}}_{m\times m^*}$, 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 $W^{\mathrm{T}}_{m\times m^*}$, 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 of Theorem 3.2 reduce then in assuming that there exists a constant $\eta \geq 0$ such that $l^{1-\eta}(nm)^\eta(m^*)^{-\eta} = \mathcal{O}(1)$.

3.1.4 Signal-to-Noise Ratio Estimators

In this section, we discuss three possible ways of obtaining consistent estimators of the signal-to-noise ratio parameter $\lambda = \sigma^2_\theta/\sigma^2_\epsilon$.

Profile and Restricted Profile Maximum Likelihood Estimators

By maximizing (12) or (15) one can obtain the profile and restricted profile maximum likelihood estimators $\hat{\lambda}^{\mathrm{LR}}$ and $\hat{\lambda}^{\mathrm{RLR}}$ respectively of λ. Under some regularity conditions, one can show that $\hat{\lambda}^{\mathrm{LR}}$ and $\hat{\lambda}^{\mathrm{RLR}}$ are consistent estimators of λ (see Claeskens, 2004).

An Estimating Function Based Estimator

Consider first the QR decomposition (see, for example, Thistead, 1986) of the fixed-effects design matrix $\tilde{\mathbf{X}}$

$$\tilde{\mathbf{X}} = [Q_1\ Q_2]\begin{bmatrix} R \\ 0 \end{bmatrix},$$

where R is upper triangular. It follows from the definition of the QR decomposition that the columns of Q_2 define a set of orthonormal vectors that span the orthogonal complement of the column space of $\tilde{\mathbf{X}}$. The image $\mathbf{Y}^* = Q_2^T\mathbf{Y}$ is, therefore, a zero-mean Gaussian vector with covariance matrix $\sigma^2_\epsilon\mathbf{V}^*_\lambda$, where $\mathbf{V}^*_\lambda = \mathbf{I}_{N^*} + \lambda\widetilde{\mathbf{Z}}^*\mathbf{\Sigma}\widetilde{\mathbf{Z}}^{*\mathrm{T}}$ with $\widetilde{\mathbf{Z}}^* = Q_2^T\widetilde{\mathbf{Z}}$ and $N^* = nm - m^*p$.

Assuming for the moment that σ^2_ϵ is known, the vector $\mathbf{Y}^*$ takes its values in the N^*-dimensional vector statistical space $S = \{\mathbb{R}^{N^*}, \mathcal{B}, \mu_\lambda, \lambda \geq 0\}$, with μ_λ denoting the zero-mean Gaussian probability measure on the Borel σ-algebra $\mathcal{B}$ of subsets of $\mathbb{R}^{N^*}$ given by $\mu_\lambda = \mathcal{L}(\mathbf{Y}^*)$, where $\mathcal{L}(\mathbf{X})$ denotes the law (distribution) of the random vector $\mathbf{X}$, with covariance operator given by $\sigma^2_\epsilon\left(\mathbf{I}_{N^*} + \lambda\,\widetilde{\mathbf{Z}}^*\mathbf{\Sigma}\widetilde{\mathbf{Z}}^{*\mathrm{T}}\right)$. Recall that an estimating function (see Godambe, 1960) for the vector statistical space S is any map $g:\ \mathbb{R}^{N^*}\times\mathbb{R} \to \mathbb{R}$ such that, given an observed value $\mathbf{y}^*$ of the random vector $\mathbf{Y}^*$ in S, an estimate of λ is obtained by solving the equation $g(\mathbf{y}^*;\lambda) = 0$ with respect to λ. The estimating function g is said to be unbiased if $\mathbb{E}_\lambda(g) = 0$ for any λ. Following Antoniadis &

Lavergne (1994), the explicit unbiased estimating function that we consider is

$$\begin{aligned} g(\mathbf{y}^*;\lambda) &= \|\mathbf{y}^*\|^2 - \sigma_\epsilon^2 \operatorname{tr}\left(\mathbf{I}_{N^*} + \lambda\, \widetilde{\mathbf{Z}}^* \mathbf{\Sigma} \widetilde{\mathbf{Z}}^{*\mathrm{T}}\right) \\ &= \|\mathbf{y}^*\|^2 - \sigma_\epsilon^2 \left(N^* + \lambda \operatorname{tr}\left(\widetilde{\mathbf{Z}}^* \mathbf{\Sigma} \widetilde{\mathbf{Z}}^{*\mathrm{T}}\right)\right), \quad \mathbf{y}^* \in \mathbb{R}^{N^*},\ \lambda \geq 0, \end{aligned}$$

where $\operatorname{tr}(\mathbf{A})$ denotes the trace of the square matrix $\mathbf{A}$. When σ_ϵ^2 is known, the corresponding estimate $\hat{\lambda}$ that solves $g(\mathbf{Y}^*;\lambda) = 0$ is given by

$$\hat{\lambda} = \frac{\|\mathbf{Y}^*\|^2/\sigma_\epsilon^2 - N^*}{\operatorname{tr}\left(\widetilde{\mathbf{Z}}^* \mathbf{\Sigma} \widetilde{\mathbf{Z}}^{*\mathrm{T}}\right)}.$$

It is clear that the above estimator can assume negative values. Starting form $\hat{\lambda}$ it is then natural to define an admissible estimator $\hat{\lambda}^+$ by the rule $\hat{\lambda}^+ = \max(\hat{\lambda}, 0)$. Suppose now that one has available an independent and identically distributed sample $(\mathbf{y}_1^*, \ldots, \mathbf{y}_r^*)$ from the above considered model (which obviously arises when one has repetitions per subject in the general functional mixed-effects model (1)). An efficient estimator of λ can be obtained as follows. Let $\gamma_k,\ k = 1, \ldots, N^*$ be the eigenvalues of the matrix $\widetilde{\mathbf{Z}}^* \mathbf{\Sigma} \widetilde{\mathbf{Z}}^{*\mathrm{T}}$, and define

$$\tilde{\lambda} = \hat{\lambda}^+ + \frac{1}{\sqrt{r}} \mathcal{I}^{-1}(\hat{\lambda}^+) \Delta(\hat{\lambda}^+), \tag{21}$$

where

$$\mathcal{I}(\lambda) = \frac{1}{2} \sum_{k=1}^{N^*} \frac{\gamma_k^2}{(1+\lambda\gamma_k)^2},$$

and

$$\Delta(\lambda) = \frac{1}{2\sqrt{r}} \sum_{i=1}^{r} \sum_{k=1}^{N^*} (y_{ik}^{*2} - (1+\lambda\gamma_k)) \frac{\gamma_k}{(1+\lambda\gamma_k)^2}.$$

Under fairly general conditions on the mixed-effects design matrix $\tilde{\mathbf{Z}}$, one may then show that $\tilde{\lambda}$ is consistent, asymptotically Gaussian and reaches the highest possible efficiency as $r \to \infty$ (see Antoniadis & Lavergne, 1994).

Obviously, $\tilde{\lambda}$ depends on the unknown σ_ϵ^2. Nevertheless, using a QR decomposition of the matrix $[\tilde{\mathbf{X}} \mid \tilde{\mathbf{Z}}]$, one may derive an unbiased and consistent estimator of σ_ϵ^2 that may be employed in expression (21) to provide a final estimator of λ. However, the resulting Gaussian distribution becomes degenerate as the parameter λ approaches its boundary and we will not pursue this approach furthermore. Let us just say that the use of the estimator $\tilde{\lambda}$ defined in (21) can be very appealing in practical applications when repetitions are available.

A Wavelet Domain Based Estimator

Following the wavelet-based model formulation discussed in Section 2.5, one can take $\alpha_i = \alpha$, $\beta_i = \beta$ and $\gamma_i = \gamma$ for all $i = 1, 2, \ldots, n$. Their values can be chosen, e.g., by combining the prior knowledge of the smoothness of the individual data curves and the visual appearance of simulated functions under this prior formulation, as shown in Abramovich *et al.* (1998).

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 mixtures of

two Gaussian distributions with appropriate means, variances and mixture proportions, depending on $\pi_{jk}^{(r_2,i)}$ and $v_{jk}^{(r_2,i)}$. The maximum likelihood estimators of $\pi_{jk}^{(r_2,i)}$ and $v_{jk}^{(r_2,i)}$ cannot be obtained explicitly, and numerical procedures should be adopted. However, given σ_ϵ^2, it is possible to get their estimators in closed form, say $\hat{\pi}_{jk}^{(r_2,i)}$ and $\hat{v}_{jk}^{(r_2,i)}$ respectively, by the method of moments (see Abramovich & Angelini, 2004). By noting that for $j = 0$ and $\gamma = 0$, one gets $v_{jk}^{(r_2,i)} = \sigma_\theta^2$, an estimator of σ_θ^2 can be obtained for each value of $r_2 = 1, 2, \ldots, q$, and then estimate σ_θ^2 by averaging its q estimates obtained, resulting in the estimator $\hat{\sigma}_\theta^2$. The results of Abramovich & Angelini (2004) justify its consistency property.

However, in most applications, the noise variance σ_ϵ^2 can also be estimated in the wavelet domain. In wavelet function estimation, the common practice is to robustly estimate σ_ϵ 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 σ_ϵ by averaging its n robust estimates obtained from each individual data curve, resulting in the estimator $\hat{\sigma}_\epsilon^2$. A consistent estimator of λ is then simply obtained by $\hat{\lambda}^{wd} = \hat{\sigma}_\theta^2/\hat{\sigma}_\epsilon^2$.

Remark 3.4 Since the linear mixed-effects model (6) has been modelled by the wavelet-based model discussed in Section 2.5, which involves $\pi_{jk}^{(r_2,i)}$ and $v_{jk}^{(r_2,i)}$, we have applied the wavelet domain based estimator $\hat{\lambda}^{wd}$ in the analysis of the examples presented in Section 5. Preliminary simulation results indicate a very good performance, and our findings agree with those of Abramovich & Angelini (2004). On the other hand, the profile and restricted profile maximum likelihood estimators $\hat{\lambda}^{\mathrm{LR}}$ and $\hat{\lambda}^{\mathrm{RLR}}$ are strongly biased and with high variability, whereas the estimating function based estimator $\tilde{\lambda}$ is not expected to perform well due to the low number of repetitions.

Remark 3.5 If one considers the linear mixed-effects model (7) instead, then the above process can be repeated to estimate the various variance components. Then, as in Claeskens (2004), one can get estimates of the various $\hat{\sigma}_{\theta,r_2}^2$ for each $r_2 = 1, 2, \ldots, q$, leading to different estimates $\hat{\lambda}_{r_2}^{wd} = \hat{\sigma}_{\theta,r_2}^2/\hat{\sigma}_\epsilon^2$ for each $r_2 = 1, 2, \ldots, q$.

Let us conclude this section by saying that the testing ideas described above have been applied for testing functional random-effects in the examples section provided below.

3.2 Testing for Fixed-Effects

The wavelet decomposition proposed in Section 2.6 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 (6) and take the case where σ_ϵ^2 is unknown but the signal-to-noise ratio $\lambda = \sigma_\theta^2/\sigma_\epsilon^2$ is known. The general functional mixed-effects model (1) and the specific wavelet-based modelling approach that we have used in Section 2.5 for representing the random-effects functional components show that one has to take $\beta_i = 0$ in order the vector of observations $\mathbf{Y}$ be Gaussian. (However, an analogous test that can also be used when the vector of observations is not assumed Gaussian nor having identically distributed components but only consisting of random variables with symmetrical distributions, that is the case when one considers $\beta_i \neq 0$, will be briefly discussed in Remark 3.7 below.) In this case, the image of the vector $\mathbf{Y}$ by $V_\lambda^{-1/2}$ leads to the linear regression model

$$\begin{aligned} \mathbf{Y}_\lambda &= V_\lambda^{-1/2}\mathbf{Y} \\ &= V_\lambda^{-1/2}\tilde{\mathbf{X}}\tilde{\mathbf{d}} + \sigma_\epsilon \boldsymbol{\eta}, \end{aligned} \tag{22}$$

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 (22) 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

$$H_{0,c} : \ \boldsymbol{\mu} \in \mathcal{V}_c \quad \text{versus} \quad H_{A,\ell} : \boldsymbol{\mu} \in (\mathcal{V}_c + S_\ell)\backslash\mathcal{V}_c, \tag{23}$$

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_ℓ and N_ℓ will respectively denote the dimensions of S_ℓ and $(\mathcal{V}_c + S_\ell)^\perp \cap \mathcal{E}$. Let also k_c be the rank of A_c.

3.2.1 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

$$T_{\bar{\alpha}} = \sup_{\ell \in \mathcal{L}} \{T_{c,\ell} - \bar{F}^{-1}_{D_\ell, N_\ell}(\bar{\alpha}_\ell)\}, \tag{24}$$

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 (23) when $T_{\bar{\alpha}}$ is positive.

3.2.2 Level of the Test

We first study the level of the test statistic defined in (24) and show that it is of level $\bar{\alpha}$. Indeed, the following theorem holds.

Theorem 3.3 *The test statistic $T_{\bar{\alpha}}$ defined in (24), under the null hypothesis (23), satisfies*

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

The proof of Theorem 3.3 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 3.2.4).

3.2.3 Power of the Test

We now study the power of the test statistic defined in (24). Let $0 < \gamma < 1$, and let us first introduce some quantities that depend on $\bar{\alpha}_\ell$, γ, D_ℓ and N_ℓ. For each $u > 0$ and each $\ell \in \mathcal{L}$, we set

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

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

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

With the above notation, the following theorem holds.

Theorem 3.4 *Let $T_{\bar{\alpha}}$ be the test statistic defined by (24), 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 ρ_m^2 be defined as follows*

$$\mathcal{F}_m = \{\boldsymbol{\mu}(\tilde{\mathbf{d}}) \in \mathcal{E};\ d_m^2(\tilde{\mathbf{d}}, \mathcal{V}_c) \geq \rho_m^2\} \tag{25}$$

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 3.4, 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 ρ_m^2). In fact, there is a balance to achieve between $d_m^2(\Pi_{\mathcal{V}_c^\perp\cap\mathcal{E}}\boldsymbol{\mu}(\tilde{\mathbf{d}}), S_\ell)$ and v_ℓ^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_ℓ but v_ℓ^2 increases with D_ℓ. Therefore, the choice of $\mathcal{L}$ has to be done carefully, as we see for a specific case in Section 3.2.4 below.

Remark 3.6 (i) Under the condition

$$[\mathcal{H}_\mathcal{L}]: \quad \text{For all} \quad \ell \in \mathcal{L}, \quad \bar{\alpha}_\ell \geq \exp(-N_\ell/10) \quad \text{and} \quad \gamma \geq 2\exp(-N_\ell/21),$$

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).

(ii) The proposed test statistic (24) cannot directly be computed in practical applications because it depends on the unknown quantity λ. However, this problem can be solved by replacing λ 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 3.1.4.

3.2.4 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 ρ_m. Theorem 3.5 below gives a upper bound for the minimum separation from zero (see, e.g., Ingster 1982; Abramovich *et al.*, 2004), uniformly over $\mathcal{F}(R,s,\rho_m)$, by considering a specific collection of subspaces S_ℓ 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 3.5 *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. 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 σ_ϵ^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).

Remark 3.7 The general functional mixed-effects model (1) and the specific wavelet-based modelling approach that we have used in Section 2.5 for representing the random-effects functional components show that if one takes $\beta_i \neq 0$ then the vector of observations $\mathbf{Y}$ is not Gaussian but consisting of random variables with symmetrical distributions. Then, one can develop similar hypothesis testing procedures for testing functional fixed-effects contrasts by adapting the recent work of Durot & Rozenholc (2004) that parallels the work of Baraud (2002) and Baraud *et al.* (2003) and provides a non-asymptotic procedure, based on model selection methods, for testing the null hypothesis that the expectation of a vector consisting of random variables with symmetrical distributions is zero against a nonparametric composite alternative. Moreover, this test achieves the optimal rate of testing (up to the unavoidable logarithmic factor) over a class of Hölderian functions with smoothness parameter $s \in (1/4, 1]$ in the case when the errors satisfy a Bernstein-type condition, and achieves the optimal rate of testing for $s \geq 1/4 + 1/p$ in the case when the errors possess bounded moments of order $2v$ for $v \geq 2$. However, although more powerful in the case where the errors have strongly non-symmetrical distributions, their test is somewhat more computationally intensive than the one provided by Baraud *et al.* (2003) since the corresponding quantile that is used to reject the null hypothesis is estimated from the observed data. For this reason, and the fact that the two tests do not show any significant differences in the analysis of the examples considered below, we have decided not to develop this methodology any further.

Let us conclude this section by saying that the testing ideas described above have been applied for testing functional fixed-effects in the examples section provided below.

4 Estimation in Functional Mixed-Effects Models

In this section, we briefly discuss how one can apply wavelet-based estimation procedures for both fixed-effects and random-effects in the general functional mixed-effects model (1).

For the estimation of the functional fixed-effects (i.e., the population-average curve profiles), one can apply either the penalised linear wavelet estimator (which is also the BLUP) discussed in Section 2.4 or the classical weighted least-squares methodology based on the linear mixed-effects model (6). Both methods are easily implemented and have been used in the analysis of the examples considered below.

The penalised linear wavelet estimator can obviously be used for the estimation of the functional random-effects (i.e., the curve-specific functions). On other hand, the linear mixed-effects model (6) can be automatically used for the estimation of the functional random-effects via (restricted) maximum likelihood estimation of variance components, as is commonly used in standard liner mixed-models software such as `PROC MIXED` in `SAS` and `lme()` in `S-PLUS` (see, e.g., Ngo & Wand, 2004). Since the entire numerical study considered below was carried out using the MATLAB programming environment, and using the fact that both methods do not show any significant differences in the analysis of the examples considered below, we have decided only to present the application of the penalised linear wavelet estimator for the estimation of the functional random-effects.

5 Examples

The purpose of this section is to illustrate the usefulness of the proposed estimation (see Section 4) and testing (see Section 3) procedures, by applying them on the progesterone and orthosis datasets discussed in Section 1.

The computational algorithms related to wavelet analysis were performed using Version 8 of the WaveLab toolbox for MATLAB (Buckheit, Chen, Donoho, Johnstone & Scargle, 1995) that is freely available from `http://www-stat.stanford.edu/software/software.html`. The entire study was carried out using the MATLAB programming environment.

5.1 Progesterone Data Analysis

From the original dataset of 91 curves that exemplified the methods of Brumback & Rice (1998), Fan & Zhang (2000) and Wu & Zhang (2002), we selected a subset of 24 curves relevant for our mixed-effects FANOVA application. Following the wavelet-based formulation and using the matrix notation of Section 2.6, it is not difficult to see that, in this particular situation, we have $n = 24$, $m = 32$ ($m^\star = 8$), $p = 2$ and $q = 1$, and the general functional mixed-effects model (1) can be now expressed as a linear mixed-effects model with one variance component, written as

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

where

- $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_{24}^{\mathrm{T}})^{\mathrm{T}}$ (a 768×1 vector of data points);
- $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{W}^{(2)}$ (a 768×32 fixed-effects design matrix), $\mathbf{X} = (A_1^{\mathrm{T}}, A_2^{\mathrm{T}})^{\mathrm{T}}$, where A_i ($i = 1, 2$) are 12×2 block-zero matrices apart from their i^{th} column which consists of 12 identity matrices each one of size 32×32, and $\mathbf{W}^{(2)} = \mathrm{diag}(W_{32\times 8}^{\mathrm{T}}, W_{32\times 8}^{\mathrm{T}})$;
- $\widetilde{\mathbf{d}} = (\tilde{\mathbf{d}}_1^{\mathrm{T}}, \tilde{\mathbf{d}}_2^{\mathrm{T}})^{\mathrm{T}}$ (a 16×1 vector of fixed-effects wavelet coefficients);
- $\widetilde{\mathbf{Z}} = \mathbf{I}_{768}\mathbf{W}^{(24)}$ (a 768×192 random-effects design) and $\mathbf{W}_{24}^{(1)} = \mathrm{diag}(W_{32\times 8}^{\mathrm{T}}, \ldots, W_{32\times 8}^{\mathrm{T}})$ (24 blocks);

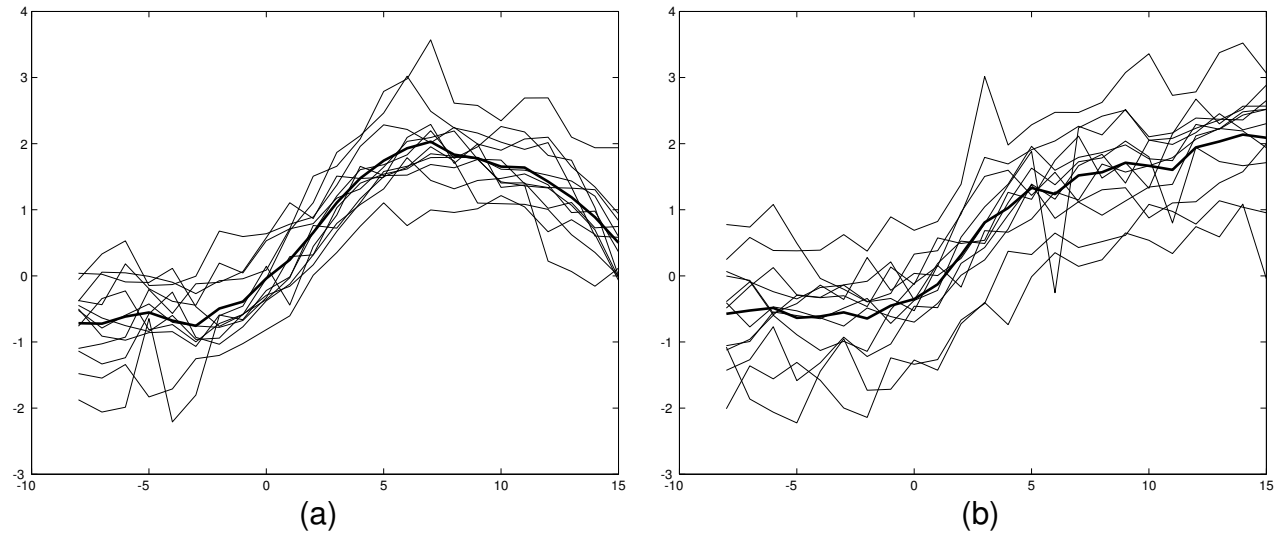


Figure 5.1: *Progesterone Dataset*: Urinary metabolite progesterone curves (thin lines) and the corresponding group means (thick lines): (a) *Nonconceptive* menstrual cycles; (b) *Conceptive* menstrual cycles.

- $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\theta}}_{24}^{\mathrm{T}})^{\mathrm{T}}$ (a 192×1 vector of random-effects wavelet coefficients);
- $\widetilde{\boldsymbol{\epsilon}} = (\widetilde{\boldsymbol{\epsilon}}_1^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\epsilon}}_{24}^{\mathrm{T}})^{\mathrm{T}}$ (a 768×1 vector of standard Gaussian errors).

Figure 5.1 shows the urinary metabolite progesterone curves for the nonconceptive and conceptive groups (12 curves have been selected from each group) together with their means. 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. To our knowledge, the various methods appeared in the literature for analysing the progesterone data are mainly concerned with the estimation of the population curves as well as the estimation of the individual curves. 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 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 3.1 reveals that $\hat{\sigma}_\theta = 0.8241$, $\hat{\sigma}_\epsilon = 0.1854$ resulting in $\hat{\lambda}^{wd} = 19.7578$. The finite sample restricted likelihood ratio test statistic, computing on a grid of 400 points and taking 100000 simulations from the null, takes the value of 22.9039 and the corresponding probability at zero value is 0.5166 which shows that the corresponding testing methodology is feasible. Figure 5.2 shows the histogram with 50 bins of the restricted likelihood ratio test statistic values under the null hypothesis. The corresponding p-value is 0, showing that there is significant evidence of random-effects in this case. By following the discussion in Section 4, Figure 5.3 shows the random-effects estimates for both groups. As observed in the figure, the estimates on both groups reveal a smooth behaviour.

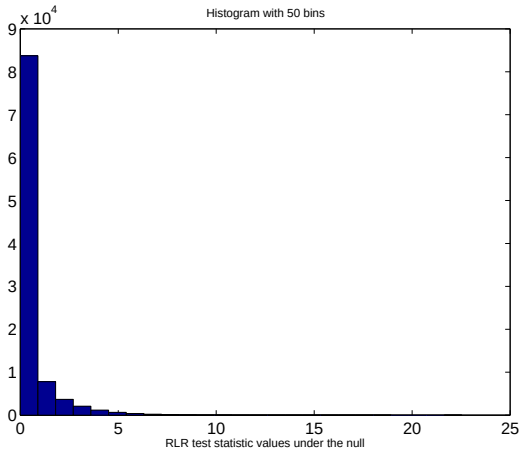
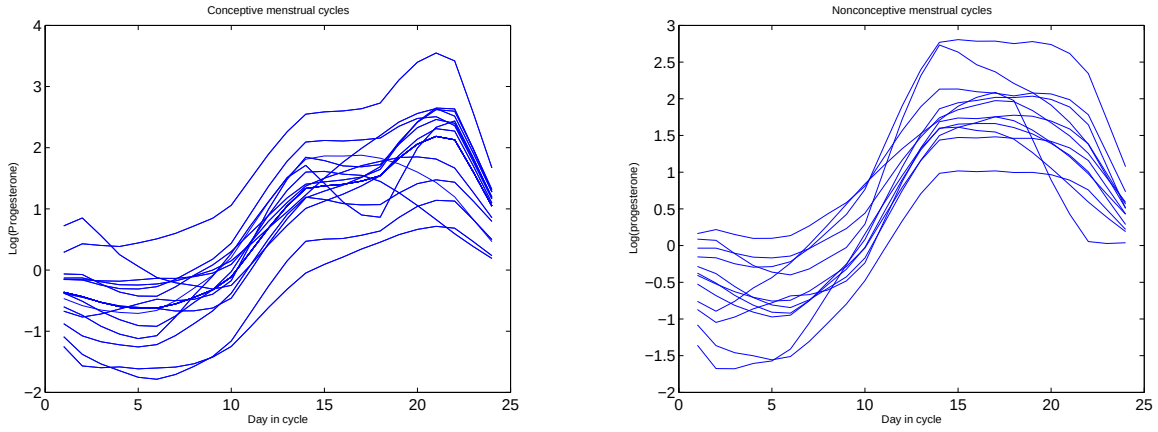


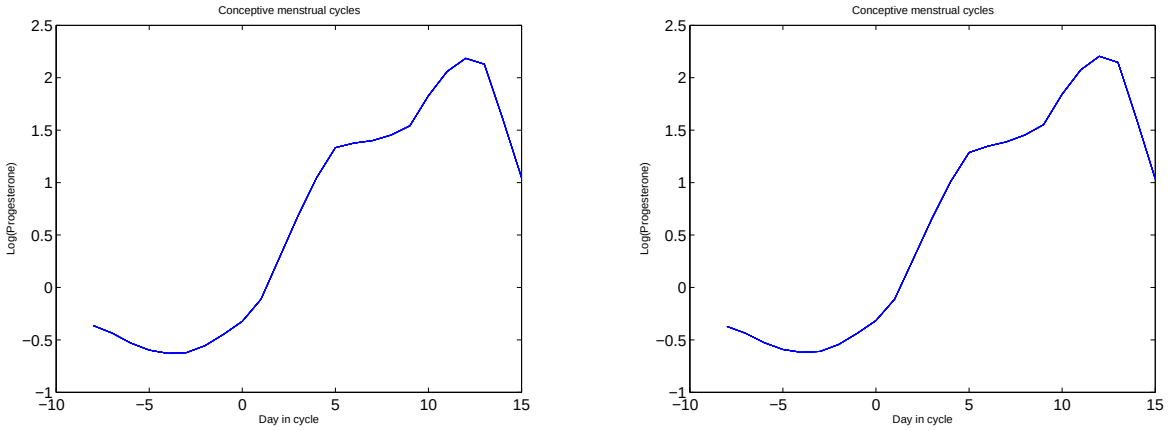
Figure 5.2: Random-Effects Testing for the *Progesterone Dataset*: The histogram with 50 bins of the RLR test statistic values under the null hypothesis.



(a) (b)

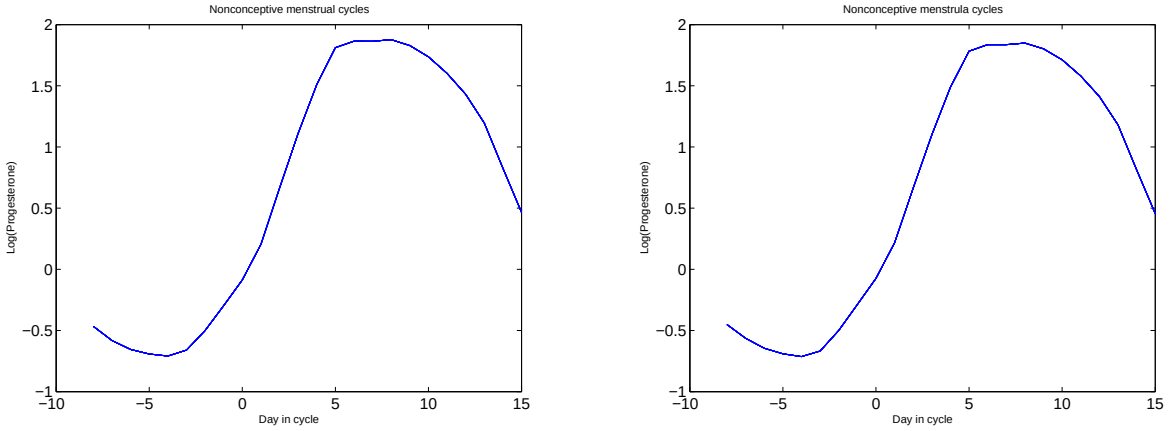
Figure 5.3: Random-Effects Estimates for the *Progesterone Dataset*: The BLUP method for (a) The *Nonconceptive* Group; (b) The *Conceptive* Group.

Regarding the functional fixed-effects, the application of the testing methodology presented in Section 3.2 reveals that a piecewise constant functions collection $\{S_\ell : \ \ell \in \mathcal{L}\}$ of orders 1, 3 and 7 gives a Bonferroni based test statistic value (where each of the corresponding Bonferroni level is taken as 0.0166) of 90.2814. This shows that the fixed-effects hypothesis that the two group means are the same is rejected. By following the discussion in Section 4, Figure 5.4 and Figure 5.5 show the mean estimates for the conceptive and nonconceptive groups based on the BLUP and weighted least-squares methods. As observed in the figures, in each group, both estimation methods are visually identical.



(a) (b)

Figure 5.4: Fixed-Effects Estimates for the *Conceptive* Group: (a) The BLUP method; (b) The Weighted Least-Squares method.



(a) (b)

Figure 5.5: Fixed-Effects Estimates for the *Nonconceptive* Group: (a) The BLUP method; (b) The Weighted-Least Squares method.

5.2 Orthosis Data Analysis

Abramovich *et al.* (2004) analysed this dataset as arising from a fixed-effects FANOVA 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 (2003), 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 FANOVA testing methodology of Abramovich & Angelini (2003).) Averaging over the 10 repetitions for each subject, following the wavelet-based formulation, and using the matrix notation of Section 2.6, it is not difficult to see that, in this particular situation, we have $n = 28$, $m = 64$ ($m^\star = 8$), $p = 4$ 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

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

where

- $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_{28}^{\mathrm{T}})^{\mathrm{T}}$ (a 1792×1 vector of data points);
- $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{W}^{(4)}$ (a 1792×32 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 7×4 block-zero matrices apart from their i^{th} column which consists of 7 identity matrices each one of size 64×64, and $\mathbf{W}^{(4)} = \mathrm{diag}(W_{32\times 8}^{\mathrm{T}}, \ldots, W_{32\times 8}^{\mathrm{T}})$ (4 blocks);
- $\widetilde{\mathbf{d}} = (\tilde{\mathbf{d}}_1^{\mathrm{T}}, \ldots, \tilde{\mathbf{d}}_4^{\mathrm{T}})^{\mathrm{T}}$ (a 32×1 vector of fixed-effects wavelet coefficients);
- $\widetilde{\mathbf{Z}} = \mathbf{I}_{1792}\mathbf{W}^{(28)}$ (a 1792×224 random-effects design) and $\mathbf{W}_{28}^{(1)} = \mathrm{diag}(W_{32\times 8}^{\mathrm{T}}, \ldots, W_{32\times 8}^{\mathrm{T}})$ (28 blocks);
- $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\theta}}_{28}^{\mathrm{T}})^{\mathrm{T}}$ (a 224×1 vector of random-effects wavelet coefficients);
- $\widetilde{\boldsymbol{\epsilon}} = (\widetilde{\boldsymbol{\epsilon}}_1^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\epsilon}}_{28}^{\mathrm{T}})^{\mathrm{T}}$ (a 1792×1 vector of standard Gaussian errors).

Figure 5.6 shows the available data set; typical moment plots over gait cycles. 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 3.1 reveals that $\hat{\sigma}_\theta = 40.3127$, $\hat{\sigma}_\epsilon = 1.0799$ resulting in $\hat{\lambda}^{wd} = 1393.5311$. The finite sample restricted likelihood ratio test statistic, computing on a grid of 400 points and taking 100000 simulations from the null, takes the value of 3.2743 and the corresponding probability at zero value is 0.5255 which shows that the corresponding testing methodology is feasible. Figure 5.7 shows the histogram with 50 bins of the restricted likelihood ratio test statistic values under the null hypothesis. The corresponding p-value is 0.0304, showing that there is significant evidence of random-effects in this case. By following the discussion in Section 4, Figure 5.8 shows the random-effects estimates of the averaged curves in each group. As observed in the figure, the corresponding estimates reveal an inhomogeneous behaviour.

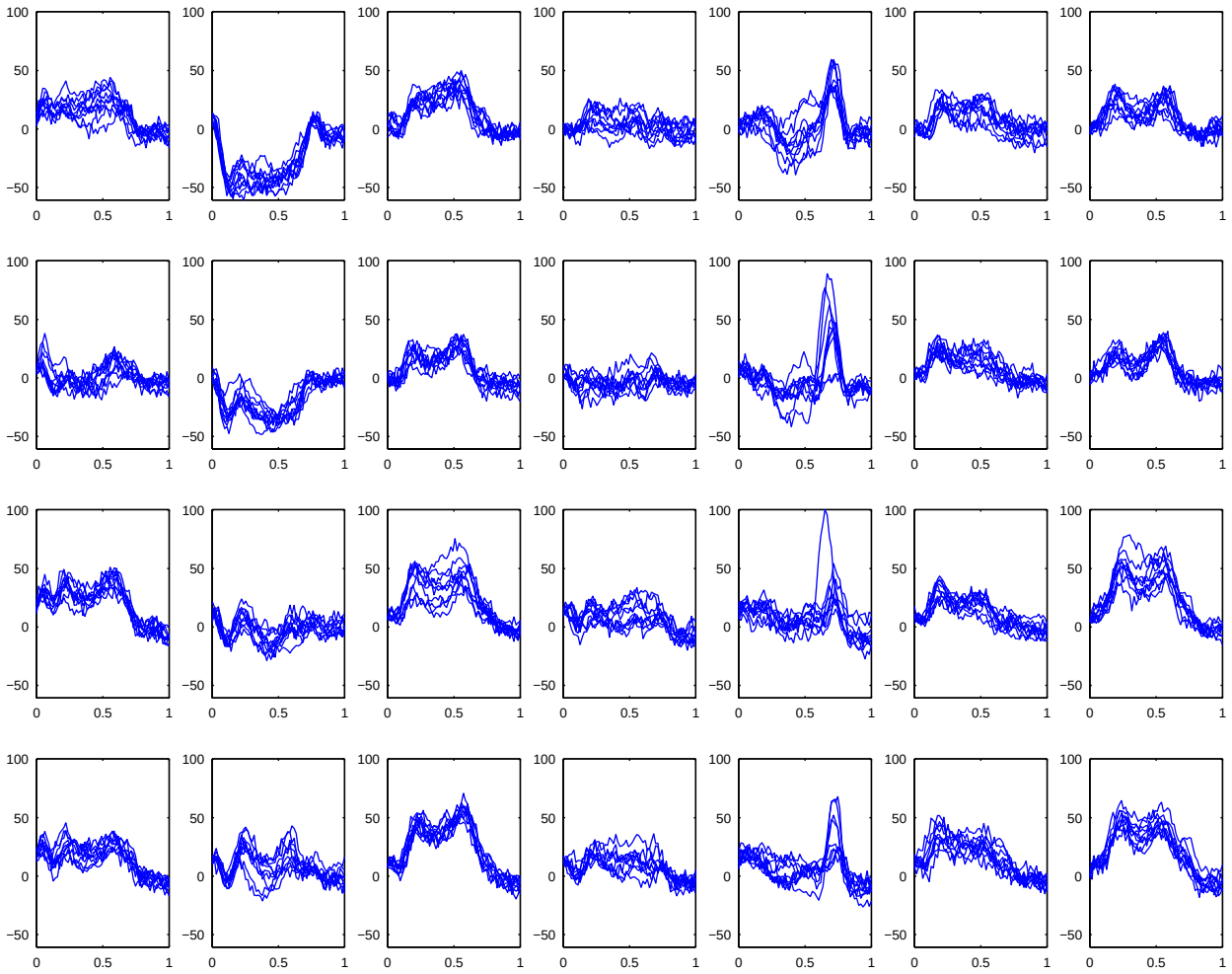


Figure 5.6: *Orthosis Dataset*: The panels in rows correspond to *Treatments* while the panels in columns correspond to *Subjects*; there are ten repeated measurements in each panel.

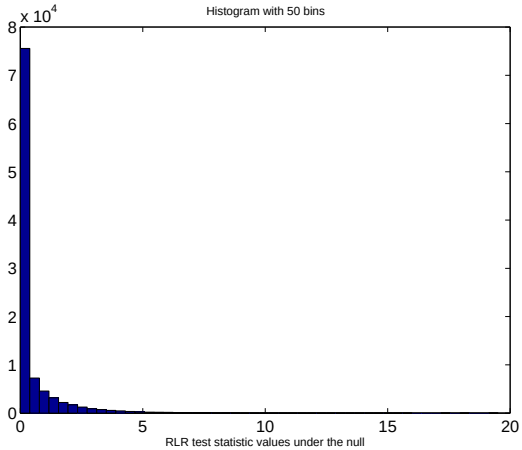


Figure 5.7: Random-Effects Testing for the *Orthosis Dataset*: The histogram with 50 bins of the RLR test statistic values under the null hypothesis.

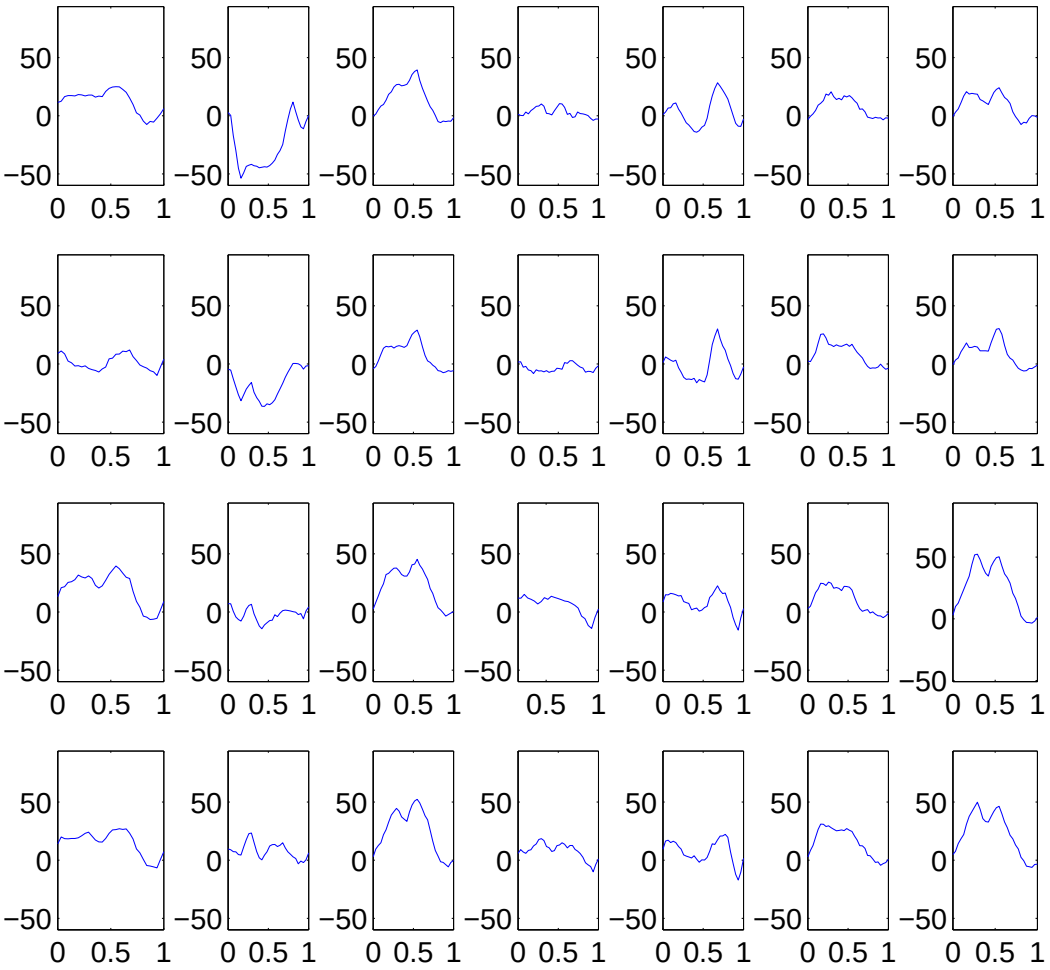
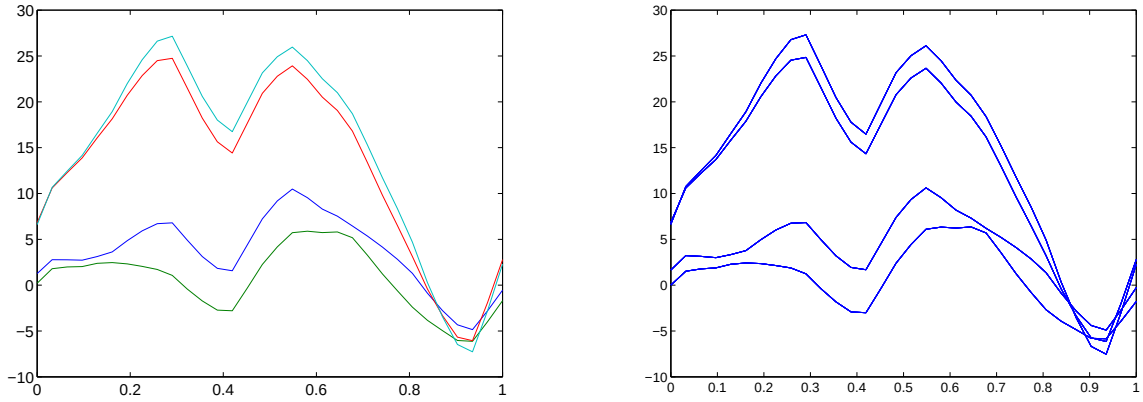


Figure 5.8: Random-Effects Estimates for the *Orthosis Dataset*: The BLUP method.

Regarding the functional fixed-effects, the application of the testing methodology presented in Section 3.2 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): 4.8472 for Spring 1 vs Spring 2 conditions, 8.9922 for Control vs Orthosis conditions, and 48.7512 for Spring 1 + Spring 2 vs Control + Orthosis conditions. These show that the various fixed-effects hypotheses of a similar behaviour under the different conditions are all rejected. 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 & Angelini (2003) and Abramovich *et al.* (2004). By following the discussion in Section 4, Figure 5.9 shows the corresponding group means estimates based on the BLUP and weighted least-squares methods. As observed in the figure, in each group, both estimation methods are visually identical.



(a) (b)

Figure 5.9: Fixed-Effects Estimates for the *Orthosis Dataset*: (a) The BLUP method; (b) The Weighted Least-Squares method.

6 Concluding Remarks

We considered a general functional mixed-effects model that inherits the flexibility of linear mixed-effects models in handling complex designs and correlation structures. Wavelet decomposition approaches were used to model both fixed-effects and random-effects in the same functional space. This help us in interpreting the resulting model as a functional data model since it does not contradict the intuition that, if each outcome is a curve, which is the basic unit in functional data analysis, then the population-average curve and the subject-specific curves should have the same smoothness property (i.e., they should lie in the same functional space). A linear mixed-effects representation was then obtained that was used for estimation and inference in the general functional mixed-effects model. Adapting recent methodologies in linear mixed-effects and nonparametric regression models, 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) were provided. Wavelet-based estimation procedures for both fixed-effects and random-effects in the general functional mixed-effects model were also applied. The usefulness of the proposed estimation and testing procedures was illustrated by means of two real-life datasets arising from endocrinology and physiology.

Although the particular examples were modelled as 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. Furthermore, although 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, the sparsity of the corresponding matrices due to the properties of compactly supported wavelets can be explored to improve upon these limitations. However, this has not been taken care of in our implementation since the particular examples analysed were easily handled.

Acknowledgements

Anestis Antoniadis were 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 'University of Cyprus–2003, 2004 Research Grants', and the 'Cyprus-France CY-FR/0204/04 Zenon program'. Theofanis Sapatinas would like to thank Anestis Antoniadis for excellent hospitality while visiting Grenoble on various occasions to carry out this work. We would like to thank Dr. Amarantini and Dr. Luc for providing us with the Orthosis data and Professor Morris for sending us a copy of his recent work.

Appendix

A1: Penalised and Regularised Linear Wavelet Estimators

We briefly introduce the *penalized* and *regularized* linear wavelet estimators and point out some of their properties. Let $H^s[0,1]$ be the Hilbert-Sobolev spaces of functions $f(t) \in L^2[0,1]$ with noninteger regularity index $s > 1/2$. Endowing $H^s[0,1]$ with the orthonormal periodic wavelet basis defined Section 2.2, it is possible to show (see, e.g., Angelini *et al.*, 2003) that $H^s = H_0 \bigoplus H_1$, where H_0 and H_1 are both reproducing kernel Hilbert spaces with reproducing kernels

$$K^0(s,t) = \sum_{k=0}^{2^{j_0}-1} \phi_{j_0k}(s)\phi_{j_0k}(t) \quad \text{and} \quad K^1(s,t) = \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \frac{\psi_{jk}(s)\psi_{jk}(t)}{2^{2sj}}, \quad (s,t) \in [0,1]^2.$$

Hence, the Hilbert-Sobolev norm $\|f\|_{H^s} \equiv \|f\|_{B^s_{2,2}}$ can be rewritten as $\|f\|_{H^s} = \|P_0 f\|_{H^s} + \|P_1 f\|_{H^s}$, where P_0 and P_1 denote the orthogonal projectors in H_0 and H_1 respectively. In such a setting, the standard nonparametric regression problem can be solved by the classical regularisation approach

$$\min_{f \in H^s} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(t_i))^2 + \lambda \, \|P_1 f\|^2_{H^s}. \tag{26}$$

The term $\|P_1 f\|^2_{H^s}$ penalise the details in the wavelet expansion of $f(t)$, and the regularisation parameter $\lambda > 0$ gives the best compromise between smoothness and goodness-of-fit of the investigated solution $\hat{f}_\lambda(t)$. The *penalized linear wavelet estimator* $\hat{f}_\lambda$ has the same optimal statistical properties as the one obtained with cubic smoothing splines, but it generalizes the corresponding nonparametric regression estimation problem over Hilbert-Sobolev spaces, $H^s[0,1]$, with non-integer regularity index $s > 1/2$.

Since $\hat{f}_\lambda$ is computational very expensive, Antoniadis (1996), Amato & Vuza (1997) and Angelini *et al.* (2003) proposed an alternative estimator, called the *regularised linear wavelet estimator*, $\tilde{f}_\lambda$, that is easier and computationally very fast to compute. This estimator arises as solution to a penalized least squares problem (which is a tight approximation to problem (26)) and it has the form

$$\tilde{f}_\lambda(t) = \sum_{k=0}^{2^{j_0}-1} \tilde{\alpha}_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \tilde{\beta}_{jk}\psi_{jk}(t), \quad t \in [0,1], \tag{27}$$

where $(\tilde{\alpha}_{j_0k})_k$ and $(\tilde{\beta}_{jk})_{j,k}$ are defined by

$$\begin{cases} \tilde{\alpha}_{j_0k} &= c_{j_0k}, & k = 0, 1, \ldots, 2^{j_0} - 1, \\ \tilde{\beta}_{jk} &= \frac{1}{(1 + \lambda 2^{2sj})}\, d_{jk}, & j_0 \leq j \leq J-1;\ k = 0, 1, \ldots, 2^j - 1, \\ \tilde{\beta}_{jk} &= 0, & j \geq J;\ k = 0, 1, \ldots, 2^j - 1. \end{cases} \tag{28}$$

Here $\lambda > 0$ is the regularisation parameter that gives the best compromise between smoothness and goodness-of-fit of the investigated solution, ϕ and ψ are the scaling and wavelet functions respectively of the selected orthonormal periodic wavelet system, and c_{j_0k} $(k = 0, 1, \ldots, 2^{j_0} - 1)$ and d_{jk} $(j = j_0, \ldots, J-1;\ k = 0, 1, \ldots, 2^j - 1)$ are the corresponding scaling and wavelet coefficients of the function f sampled at $n = 2^J$ points for some $J > 0$.

Note that Angelini *et al.* (2003) have obtained the estimator $\tilde{f}_\lambda$ for a general grid design where, earlier, the same (in form) estimator was obtained for an equispaced grid design by Antoniadis (1996) and Amato & Vuza (1997). Furthermore, its optimal global convergence rate has been proved by Antoniadis (1996) and Amato & Vuza (1997), its optimal pointwise convergence rate has been proved by Angelini & De Canditiis (2002), and the appropriateness of the Generalized Cross Validation (GCV) criterion for selecting the smoothing parameter λ from the sample has been studied by Amato & De Canditiis (2001).

A2: Besov Spaces on the Unit Interval

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. More specifically, let the rth difference of a function $f(t)$ be

$$\Delta_h^{(r)} f(t) = \sum_{k=0}^{r} \binom{r}{k} (-1)^k f(t + kh),$$

and let the rth modulus of smoothness of $f(t) \in L^{\rho_1}[0,1]$ be

$$\nu_{r,\rho_1}(f; t) = \sup_{h \leq t} (||\Delta_h^{(r)} f||_{L^{\rho_1}[0, 1-rh]}).$$

Then the Besov seminorm of index (s, ρ_1, ρ_2) is defined for $r > s$, where $1 \leq \rho_1, \rho_2 \leq \infty$, by

$$|f|_{B^s_{\rho_1,\rho_2}} = \left[\int_0^1 \left\{ \frac{\nu_{r,\rho_1}(f; h)}{h^s} \right\}^{\rho_2} \frac{dh}{h} \right]^{1/\rho_2}, \qquad \text{if} \qquad 1 \leq \rho_2 < \infty,$$

and by

$$|f|_{B^s_{\rho_1,\infty}} = \sup_{0<h<1} \left\{ \frac{\nu_{r,\rho_1}(f; h)}{h^s} \right\}.$$

The Besov norm is then defined as

$$||f||_{B^s_{\rho_1,\rho_2}} = ||f||_{L^{\rho_1}} + |f|_{B^s_{\rho_1,\rho_2}}$$

and the Besov space on $[0,1]$, $B^s_{\rho_1,\rho_2}[0,1]$, is the class of functions $f : [0,1] \to \mathbb{R}$ satisfying $f(t) \in L^{\rho_1}[0,1]$ and $|f|_{B^s_{\rho_1,\rho_2}} < \infty$, i.e. satisfying $||f||_{B^s_{\rho_1,\rho_2}} < \infty$. The parameter ρ_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 derivatives, where their existence is required in an L^{ρ_1}-sense; the additional parameter ρ_2 is secondary in its role, allowing for additional fine tuning of the definition of the space.

The Besov classes include, in particular, the well-known Hilbert-Sobolev ($H_2^s[0,1]$, $s = 1, 2, \ldots$) and Hölder ($C^s[0,1]$, $s > 0$) spaces of smooth functions ($B_{2,2}^s[0,1]$ and $B_{\infty,\infty}^s[0,1]$ respectively), but in addition less-traditional spaces, like the space of bounded-variation, sandwiched between $B_{1,1}^1[0,1]$ and $B_{1,\infty}^1[0,1]$. The latter functions are of statistical interest because they allow for better models of spatial inhomogeneity (see, e.g., Meyer, 1992; Donoho & Johnstone, 1998).

The Besov norm for the function $f(t)$ is related to a sequence space norm on the wavelet coefficients of the function. As noted in Section 2.2, confining attention to the resolution and spatial indices $j \geq j_0$ and $k = 0, 1, \ldots, 2^j - 1$ respectively, and denoting by $s' = s + 1/2 - 1/\rho_1$, the sequence space norm is given by

$$\begin{aligned} ||w||_{b^s_{\rho_1,\rho_2}} &= ||u_{j_0}||_{\rho_1} + \left\{ \sum_{j=j_0}^{\infty} 2^{js'\rho_2} ||w_j||_{\rho_1}^{\rho_2} \right\}^{1/\rho_2}, \quad \text{if} \quad 1 \leq \rho_2 < \infty, \\ ||w||_{b^s_{\rho_1,\infty}} &= ||u_{j_0}||_{\rho_1} + \sup_{j \geq j_0} \left\{ 2^{js'} ||w_j||_{\rho_1} \right\}, \end{aligned}$$

where

$$||u_{j_0}||_{\rho_1}^{\rho_1} = \sum_{k=0}^{2^{j_0}-1} |u_{j_0 k}|^{\rho_1} \quad \text{and} \quad ||w_j||_{\rho_1}^{\rho_1} = \sum_{k=0}^{2^j-1} |w_{jk}|^{\rho_1}.$$

If the mother wavelet $\psi(t)$ is of regularity $r > 0$, it can be shown that the corresponding orthonormal periodic wavelet basis defined in Section 2.2 is an unconditional basis for the Besov spaces $B^s_{\rho_1,\rho_2}[0,1]$ for $0 < s < r$, $1 \leq \rho_1, \rho_2 \leq \infty$. In other words, we have

$$K_1 ||f||_{B^s_{\rho_1,\rho_2}} \leq ||w||_{b^s_{\rho_1,\rho_2}} \leq K_2 ||f||_{B^s_{\rho_1,\rho_2}},$$

where K_1 and K_2 are constants, not depending on $f(t)$. Therefore the Besov norm of the function $f(t)$ is equivalent to the corresponding sequence space norm defined above; this allows one to characterize Besov spaces in terms of wavelet coefficients (see, e.g., Meyer, 1992; Donoho & Johnstone, 1998). For a more detailed study on (inhomogeneous) Besov spaces we refer to, e.g., DeVore & Popov (1988) and Meyer (1992).

A3: Outlines of the Proofs of the Theoretical Results

For sake of brevity, we omit most of the details and just provide outlines of the proofs of the theoretical results obtained in Sections 2 and 3.

Proof of Theorem 2.1. The proof of the theorem can be obtained by working along the same lines of the proof of Theorem 4 in Angelini *et al.* (2000). □

Proof of Theorem 3.1. The proof of the theorem can be obtained by working along the same lines of the proof of Theorem 1 in Crainiceanu & Ruppert (2004a), taking into account that $[n(\ln(m) - q)] > p$. □

Proof of Theorem 3.2. The proof of the theorem can be obtained by working along the same lines of the proof of Theorem 2 in Crainiceanu & Ruppert (2004a), taking also into account the discussion in their Section 3 and that $[n(\ln(m) - q)] > p$. □

Proof of Theorem 3.3. Under the null hypothesis (23), and for each ℓ, 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 χ^2 variables with D_ℓ and N_ℓ 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_ℓ and N_ℓ 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 3.4. 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 $d^2_m(\Pi_{\mathcal{V}_c^\perp\cap\mathcal{E}}\boldsymbol{\mu}(\tilde{\mathbf{d}}), S_\ell)$ and of v_ℓ. □

Proof of Theorem 3.5. 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,

$$\begin{aligned} d^2_m(\mathbf{f},S_j) &= \inf_{g_j\in S_j}\left\{\frac{1}{m}\sum_{i=1}^m\sum_{k\in K_j}\left(f(x_i)-[\mathbf{g}_m]_i\right)^2 1_{B_{kj}}(x_i)\right\} \\ &\leq \inf_{g_j\in S_j}\left\{\frac{1}{m}\sum_{i=1}^m\sum_{k\in K_j}\sup_{B_{kj}}|f(x_i)-g_j(x_i)|\right\}^2, \end{aligned} \tag{29}$$

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 (29), we easily get

$$d_m^2(\mathbf{f}, S_j) = C^2 R^2 2^{-2js}. \tag{30}$$

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

$$\ln\left(\frac{\ln(m)}{\bar{\alpha}_m}\right) \leq \ln\ln(m). \tag{31}$$

Using inequalities (30) and (31), we obtain

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

The conclusion of the theorem now follows by working along the 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

$$\inf_{j\in\mathcal{M}_m}\left\{R^2 2^{-2js} + \frac{\sigma_\epsilon^2}{m}\sqrt{2^j \ln\ln(m)}\right\} \leq 2\frac{\sigma_\epsilon^2}{m}\sqrt{2^{j'}\ln\ln(m)} \leq 2\sqrt{2}R^{\frac{2}{(1+4s)}}\left(\frac{\sigma_\epsilon^2}{m}\sqrt{\ln\ln(m)}\right)^{\frac{4s}{(1+4s)}}. \tag{32}$$

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

$$\inf_{j\in\mathcal{M}_m}\left\{R^2 2^{-2js} + \frac{\sigma_\epsilon^2}{m}\sqrt{2^j \ln\ln(m)}\right\} \leq CR^2 m^{-2s}. \tag{33}$$

From (31), (32) and (33), the requested inequality is proved. □

References

[1] Abramovich, F. & Angelini, C. (2003). Testing in mixed-effects FANOVA models. *Techical Report*, **RP-SOR-03-03**, Department of Statistics and Operations Research, Tel Aviv University, Israel.

[2] Abramovich, F., Bailey, T.C. & Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, **49**, 1–29.

[3] 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.

[4] 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.

[5] Amato, U. & Vuza, D.T. (1997). Wavelet approximation of a function from samples affected by noise. *Revue Roumaine de Mathématiques Pures et Appliquées*, **42**, 481–493.

[6] Amato, U. & De Canditiis, D. (2001). Convergence in probability of the Mallows and GCV wavelet and Fourier regularization methods. *Statistics and Probability Letters*, **54**, 325–329.

[7] Angelini, C. & De Canditiis, D. (2002). Pointwise convergence of the wavelet regularized estimators. *Communications in Statistics – Theory and Methods*, **31**, 1561–1578. (Erratum: *Communications in Statistics – Theory and Methods*, **32**, 1085–1086, (2003).)

[8] Angelini, C., De Canditiis, D. & Leblanc, F. (2000). Wavelet regression estimation in nonparametric mixed effect models. *Rapporto Tecnico*, **193/00**, Istituto per Applicazioni della Matematica, CNR, Napoli, Italy.

[9] Angelini, C., De Canditiis, D. & Leblanc, F. (2003). Wavelet regression estimation in nonparametric mixed effect models. *Journal of Multivariate Analysis*, **85**, 267–291.

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

[11] Antoniadis, A. (1996). Smoothing noisy data with tapered coiflets series. *Scandinavian Journal of Statistics*, **23**, 313–330.

[12] Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *Journal of the Italian Statistical Society*, **6**, 97–144.

[13] Antoniadis, A. & Lavergne, C. (1994). An estimating function for a scalar parameter in a covariance operator. *Statistics and Decisions*, **12**, 53–65.

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

[15] Antoniadis, A., Grégoire, G. & McKeague, I.W. (1994). Wavelet methods for curve estimation. *Journal of the American Statistical Association*, **89**, 1340–1353.

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

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

[18] Barry, D. (1996). An empirical Bayes approach to the growth curve analysis. *The Statistician*, **45**, 3–19.

[19] Botchkina, N. (2002). Wavelets for non-parametric regression and a test of significance. *Ph.D. Thesis*, School of Mathematics, University of Bristol, United Kingdom.

[20] Brumback, B.A. & Rice, J.A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with comments). *Journal of the American Statistical Association*, **93**, 961–994.

[21] Buckheit, J.B, Chen, S., Donoho, D.L., Johnstone, I.M & & Scargle, J. (1995). About WaveLab. *Technical Report*, Department of Statistics, Stanford University, USA.

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

[23] Claeskens, G. (2004). Restricted likelihood ratio lack of fit tests using mixed spline models. *Technical Report*, Institute of Statistics, Université catholique de Louvain, Belgium.

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

[25] Crainiceanu, C.M. & Ruppert, D. (2004b). Restricted likelihood ratio tests in nonparametric longitudinal models. *Statistica Sinica*, **14**, 713–729.

[26] Cuevas, A., Febrero, M. & Fraiman, R. (2004). An anova test for functional data. *Computational Statistics and Data Analysis*, **47**, 111–122.

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

[28] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

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

[30] Dette, H. & Derbort, S. (2001). Analysis of variance in nonparametric regression models. *Journal of Multivariate Analysis*, **76**, 110–137.

[31] DeVore, R.A. & Popov, V. (1988). Interpolation of Besov Spaces. *Transactions of the American Mathematical Society*, **305**, 397–414.

[32] Diggle, P.J., Liang, K.-Y. & Zeger, S.L. (1994). *Analysis of Longitudinal Data*. New York: Oxford University Press.

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

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

[35] Durot, C. & Rozenholc, Y. (2004). An adaptive test for zero mean. *Technical Report*, **N. 915**, Laboratoire de Probabilites et Modeles Aleatoires, Universite Paris VI & Paris VII, France.

[36] Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.

[37] Fan, J. & Lin, S.-K. (1998). Test of significance when data are curves. *Journal of the American Statistical Association*, **93**, 1007–1021.

[38] Fan, J. & Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society*, Series B, **62**, 303–322.

[39] Farebrother, R.W. (1990). The distribution of a quadratic form in normal variables. *Applied Statistics*, **39**, 294–309.

[40] Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, **31**, 1208–1211.

[41] Green, P.J. (1999). Disussion of "The analysis of designed experiments and longitudinal data by using smoothing splines" by P. Verbyla, B.R. Cullis, M.G. Kenward & S.J. Welham. *Applied Statistics*, **48**, 304–305.

[42] Green, P.J. (2000). Comment on "Bayesian backfitting" by T. Hastie & R. Tibshirani. *Statistical Science*, **15**, 218–221.

[43] Gu, C. (2002). *Smoothing Spline ANOVA Models*. New York: Springer-Verlag.

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

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

[46] Hart, J.D. & Wehrly, T.E. (1986). Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association*, **81**, 1080–1088.

[47] Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320–338.

[48] Hastie, T.J. & Tibshirani, R.J. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society*, Series B, **55**, 757–796.

[49] Hoover, D.R., Rice, J.A., Wu, C.O. & Yang, L.P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809–822.

[50] 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.

[51] Huang, S.-Y. & Lu, H.H.-S. (2000). Bayesian wavelet shrinkage for nonparametric mixed-effects models. *Statistica Sinica*, **10**, 1021–1040.

[52] Ingster, Yu.I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problems of Information Transmission*, **18**, 130–140.

[53] Ingster, Yu.I. & Suslina, I.A. (1998). Minimax detection of a signal for Besov bodies and balls. *Problems of Information Transmission*, **34**, 48–59.

[54] Koopman, S.J. & Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, **21**, 281–296.

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

[56] Lin, X. & Carroll, R.J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, **95**, 520–534.

[57] Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

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

[59] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.

[60] Morris, J.S. & Carroll, R.J. (2004). Wavelet-based functional mixed models. *Technical Report*, MD Anderson Cancer Center, University of Texas, USA.

[61] Morris, J.S., Vannucci, M., Brown, P.J. & Carroll, R.J. (2003). Wavelet-based nonparametric modelling of hierarchical functions in colon carcinogenisis (with discussion). *Journal of the American Statistical Association*, **98**, 573–597.

[62] Munro, C., Stabenfeldt, G., Cragun, J., Addiego, L., Overstreet, J. & Lasley, B. (1991). Relationship of serum estradiol and progesterone concentrations to the excretion profiles of their major urinary metabolites as measured by enzyme immunoassay and radioimmunoassay. *Clinical Chemistry*, **37**, 838–844.

[63] Nason, G.P. & Silverman, B.W. (1995). The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics*, Antoniadis, A. & Oppenheim, G. (Eds.), Lecture Notes in Statistics, **103**, pp. 281–300, New York: Springer-Verlag.

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

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

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

[67] Ramsay, J.O. & Silverman, B.W. (2002). *Applied Functional Data Analysis*. New York: Springer-Verlag.

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

[69] 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.

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

[71] 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.

[72] Shi, M., Weiss, R.E., & Taylor, J.M.G. (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics*, **45**, 151–163.

[73] 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.

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

[75] Stram, D.O. & Lee, J.W. (1994). Variance components testing in longitudinal mixed effects model. *Biometrics*, **50**, 1171–1177.

[76] Thistead, R.A. (1986). *Elements of Statistical Computing*. London: Chapman & Hall.

[77] Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

[78] Verbyla, P., Cullis, B.R., Kenward, M.G. & Welham, S.J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Applied Statistics*, **48**, 269–311.

[79] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York: John Wiley & Sons.

[80] Wahba, G. (1978). Improper priors, splines smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society*, Series B, **40**, 364–372.

[81] Wahba, G. (1983). "Confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society*, Series B, **45**, 133–150.

[82] Wang, Y. (1998a). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, **93**, 341–348.

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

[84] Wang, Y. & Taylor, J.M.G. (1995). Inference for smoothing curves in longitudinal data with application to an AIDS clinical trial. *Statistics in Medicine*, **14**, 1205–1218.

[85] Wecker, W.E. & Ansley, C.F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, **78**, 81–89.

[86] Wu, C.O., Chiang, C.T. & Hoover, D.R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, **93**, 1388–1402.

[87] Wu, H. & Zhang, J.-T. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association*, **97**, 883–897.

[88] Wypij, D., Pugh, M. & Ware, J.H. (1993). Modeling pulmonary function growth with regression splines. *Statistica Sinica*, **3**, 329–350.

[89] Zeger, S.L. & Diggle, P.J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, **50**, 689–699.

[90] Zhang, D., Lin, X.H., Raz, J. & Sowers, M.F. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93**, 710–719.