Optimal testing in functional analysis of variance models

Felix Abramovich,

Department of Statistics and Operations Research, Tel Aviv University,

Tel Aviv 69978, Israel.

Anestis Antoniadis,

Laboratoire IMAG-LMC, University Joseph Fourier,

BP 53, 38041 Grenoble Cedex 9, France.

Theofanis Sapatinas,

Department of Mathematics and Statistics, University of Cyprus,

P.O. Box 20537, CY 1678 Nicosia, Cyprus.

and

Brani Vidakovic,

Industrial and Systems Engineering, Georgia Institute of Technology,

Atlanta, Georgia 30332-0205, USA.

Authors' Footnote:

Felix Abramovich is Professor at Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel. email: felix@math.tau.ac.il

Anestis Antoniadis is Professor at Laboratoire IMAG-LMC, University Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France. email: antonia@imag.fr

Theofanis Sapatinas is Assistant Professor in Department of Mathematics and Statistics, University of Cyprus, P.O.B. 20537, CY 1678 Nicosia, Cyprus. email: T.Sapatinas@ucy.ac.cy

Brani Vidakovic is Associate Professor in the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. email: brani@isye.gatech.edu

Abstract

We consider the testing problem in a general functional analysis of variance model. We test the null hypotheses that the main effects and/or the interactions are zeros against the composite nonparametric alternative hypotheses that they are separated away from zero in L^2-norm and also posses some smoothness properties. We adapt the minimax functional hypothesis testing procedures for testing a zero signal in a Gaussian "signal plus noise" model to derive asymptotically (as the noise level goes to zero) minimax nonadaptive and adaptive functional hypothesis testing procedures for the main effects and/or the interactions based on the empirical wavelet coefficients of the data. Wavelet decompositions allow one to characterise different types of smoothness conditions assumed on the response function by means of its wavelet coefficients for a wide range of various function classes.

In order to shade some light on the theoretical results obtained, we carried out a small simulation study to examine the finite sample performance of the proposed functional hypothesis testing procedures. We also apply these tests to two real-life data examples arising from endocrinology and from neuropsychology. Concluding remarks and hints for possible extensions of the proposed methodology are also given.

KEYWORDS: Analysis of Variance; Besov Spaces; Functional Analysis of Variance; Functional Magnetic Resonance Imaging; Metabolite Progesterone Analysis; Nonparametric Hypothesis Testing; Positron Emission Tomography; Wavelets.

1. INTRODUCTION

Analysis of variance (ANOVA) is one of the most widely used tools in applied statistics. While very useful for handling low dimensional data, it has its limitations in analysing *functional* responses. Such responses 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 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. For instance, in the traditional analysis of electro-encephalogram (EEG) data, experts record EEG measurements of healthy men and women of different ages starting from young adults through middle ages to get a refined understanding of the variation in EEG due to age and gender. The general question to be answered is then the influence of age and gender on the shape of the EEG measurements.

In such cases, *functional analysis of variance* (FANOVA) methods provide alternatives to classical ANOVA methods while still allowing a simple interpretation. Due to a large set of applications involving functional data, FANOVA models have recently gained popularity and related literature has been steadily growing. For comprehensive reviews we refer, for example, to Ramsay & Silverman (1997) and Stone *et al.* (1997). Although there is an impressive literature on *fitting* FANOVA models and *estimating* their components (e.g., Wahba *et al.*, 1995; Stone *et al.*, 1997; Huang, 1998; Lin, 2000; Gu, 2002), there is no much work on developing *hypothesis testing* procedures in FANOVA models. This paper considers the testing problem in a general FANOVA model and derives asymptotically optimal tests for its components.

First we describe the following *diffusion* version of the FANOVA model we are going to consider hereafter. Suppose that one observes a series of *sample paths* of a stochastic process driven by

$$dY_i(\mathbf{t}) = m_i(\mathbf{t})\, d\mathbf{t} + \epsilon\, dW_i(\mathbf{t}), \quad i = 1, \ldots, r; \quad \mathbf{t} \in [0,1]^d, \tag{1}$$

where $\epsilon > 0$ is a constant, r and d are finite integers, m_i are (unknown) d-dimensional response functions and W_i are independent d-dimensional standard Wiener processes. In most applications, we are interested in the cases $d = 1$ (a set of *signals*) and $d = 2$ (a set of *images*).

Denote by H_1 the Euclidean space $\mathbb{R}^r$, which is isomorphic to a finite-dimensional Hilbert space of real-valued functions defined on $S = \{1, 2, \ldots, r\}$, endowed with the inner product $\langle f, g \rangle = \sum_{i=1}^r f(i)g(i)$. Let H_2 be the infinite-dimensional separable Hilbert space $L^2([0,1]^d)$ of real-valued functions defined on $[0,1]^d$ with the standard inner product $\langle f, g \rangle = \int_{[0,1]^d} f(\mathbf{t})g(\mathbf{t})d\mathbf{t}$. Let $\otimes$ be the standard algebraic *tensor product*. For the FANOVA model (1), we shall assume that $m \in H_1 \hat{\otimes}_2 H_2$, where $m(i, \mathbf{t}) = m_i(\mathbf{t})$ and the space $H_1 \hat{\otimes}_2 H_2$ is a Hilbert space obtained by completing the prehilbertian space $H_1 \otimes H_2$, endowed with the inner product $\langle h_1 \otimes h_2, k_1 \otimes k_2 \rangle_{H_1 \otimes H_2} = \langle h_1, k_1 \rangle_{H_1} \langle h_2, k_2 \rangle_{H_2}$ for $h_1, k_1 \in H_1$ and $h_2, k_2 \in H_2$. Since in our case H_1 is finite-dimensional, $H_1 \hat{\otimes}_2 H_2$ is isomorphic to $H_1 \otimes H_2$.

In order to complete the FANOVA model (1), we need to define the structure of the response functions m_i. Following Antoniadis (1984) (see also Huang, 1998; Lin, 2000), the functions m_i admit the following unique decompositions :

$$m_i(\mathbf{t}) = m_0 + \mu(\mathbf{t}) + a_i + \gamma_i(\mathbf{t}) \quad i = 1, \ldots, r; \quad \mathbf{t} \in [0,1]^d, \tag{2}$$

where m_0 is a constant function, $\mu(\mathbf{t})$ is either zero or a non-constant function of $\mathbf{t}$ (the *main effect* of $\mathbf{t}$), a_i is either zero or a non-constant function of i (the *main effect* of i) and $\gamma_i(\mathbf{t})$ is either zero or a non-zero function which cannot be decomposed as a sum of a function of i and a function of $\mathbf{t}$ (the *interaction* component). Due to the Hilbert structure of $H_1 \hat{\otimes}_2 H_2$, the following sets of orthogonal (identifiability) conditions hold for the FANOVA model (1) :

$$\int_{[0,1]^d} \mu(\mathbf{t})\, d\mathbf{t} = 0, \quad \sum_{i=1}^r a_i = 0, \quad \forall\, i = 1, \ldots, r; \;\; \mathbf{t} \in [0,1]^d \tag{3}$$

$$\sum_{i=1}^r \gamma_i(\mathbf{t}) = 0, \quad \int_{[0,1]^d} \gamma_i(\mathbf{t})\, d\mathbf{t} = 0, \quad \forall\, i = 1, \ldots, r; \;\; \mathbf{t} \in [0,1]^d. \tag{4}$$

As in the standard ANOVA models, one is naturally interested in testing the significance of the main effects and/or the interactions in the FANOVA model (1)-(4). Since these testing problems involve functional data, we call them *functional hypothesis testing* problems. A first, somewhat naive approach to the functional hypothesis testing is to look at FANOVA as a standard *univariate* ANOVA problem for each specific t (e.g., Ramsay & Silverman, 1997) and perform a series of, say, corresponding pointwise F-tests. A crucial drawback of this approach is that an enormous number of hypotheses (the number of data points per curve can be hundreds or thousands) has to be tested simultaneoulsy that causes a serious multiplicity problem. Ignoring multiplicity leads to the uncontrolled overall Type I error while, for example, Bonferroni type procedures are known to yield an extremely low power of the test. Another approach to FANOVA testing considered in the literature is to treat functional data as multivariate vectors and to apply traditional multivariate ANOVA techniques combined sometimes with various initial dimensionality-reduction procedures (e.g., Barry & Hartigan, 1990; Raz, 1990, Buckley, 1991; Eubank & La Riccia, 1993; Chen, 1994). However, the "curse of dimensionality" makes these attempts also problematic. Faraway (1997) discussed the difficulties of generalizing the ideas of multivariate testing procedures to the functional data analysis context. Recently, Fan & Lin (1998) proposed a powerful overall test for functional hypothesis testing based on the decomposition of the original functional data into orthogonal series (Fourier, wavelets) and applying the adaptive Neyman's testing procedure of Fan (1996) to the resulting empirical coefficients. The underlying idea is based on the *sparsity* of the underlying signal's representation in the Fourier or wavelet domains that allows a significant reduction of dimensionality. Somewhat similar approaches were considered in Eubank (2000) and Dette & Derbort (2001). However, none of the above works investigates the optimality of the proposed tests.

In this paper we derive asymptotically (as the noise level $\epsilon \to 0$) optimal (minimax) *non-adaptive* and *adaptive* functional hypothesis testing procedures for testing the significance of the main effects and/or the interactions in the FANOVA model (1)-(4) against the composite nonparametric alternatives that they are separated away from zero in $L^2([0,1]^d)$-norm and also posses some smoothness properties. To derive the tests, we adapt the minimax functional hypothesis testing procedures for testing a zero signal in a Gaussian "*signal plus* *noise*" model originated by Ingster (1982) and further developed in Ermakov (1990), Ingster (1993a, 1993b, 1993c), Spokoiny (1996), Ingster & Suslina (1998, 2000), Spokoiny (1998), Horowitz & Spokoiny (1999), Lepski & Spokoiny (1999) for various separation distances between the two hypotheses and different function classes under the alternative. The tests are based on the empirical wavelet coefficients of the data. Wavelet decompositions allow one to characterise different types of smoothness conditions assumed on the response function by means of its wavelet coefficients for a wider range of function classes than the ones obtained by, for example, their Fourier counterparts (e.g., Meyer, 1992; DeVore & Lorentz, 1993).

Note that though in practice one always observes discrete data sample of size n, under some general conditions there exists the asymptotic equivalence of the corresponding nonparametric regression model with the variance $\sigma^2 = n\epsilon^2$ and the white noise model (1) (Brown & Low, 1996; Donoho & Johnstone, 1999).

The paper is organised as follows. In Section 2 we formulate the hypotheses to be tested and provide definitions and background on functional hypothesis testing and wavelet analysis necessary for the proposed methodology. In Section 3 we derive asymptotically optimal (minimax) nonadaptive and adaptive functional hypothesis testing procedures for the main effects and/or the interactions. In Section 4 we carried out a small simulation study to examine the *finite sample* performance of the proposed functional hypothesis testing procedures. We also apply these tests to two real-life data examples arising from endocrinology and neuropsychology. Finally, in Section 5, we provide concluding remarks and provide some hints for possible extensions of the proposed methodology.

2. FORMULATIONS AND DEFINITIONS

2.1 Formulation of the hypotheses to be tested

In Section 1 we defined unique orthogonal decompositions (2) of the response functions m_i in the FANOVA model (1). Due to (2), testing the significance of the main effects and/or the interactions is equivalent to testing the following hypotheses

$$H_0 \quad : \quad a_i = 0, \qquad \forall\, i = 1, \ldots, r, \tag{5}$$

$$H_0 \quad : \quad \mu(\mathbf{t}) \equiv 0, \qquad \mathbf{t} \in [0,1]^d, \tag{6}$$

$$H_0 \quad : \quad \gamma_i(\mathbf{t}) \equiv 0, \qquad \forall\, i = 1, \ldots, r, \;\; \mathbf{t} \in [0,1]^d. \tag{7}$$

Assume first that ϵ is known. Integrating (1) with respect to $\mathbf{t}$ and using the identifiability conditions (3)-(4), we have

$$Y_i^* = m_0 + a_i + \epsilon\, \xi_i, \quad i = 1, \ldots, r, \quad \sum_{i=1}^{r} a_i = 0,$$

where $Y_i^* = \int_{[0,1]^d} dY_i(\mathbf{t})$ and ξ_i are independent $N(0,1)$ random variables. This is the classical one-way ANOVA model with known ϵ, so testing (5) can be performed by standard ANOVA procedures.

We focus then on *functional* hypothesis testing for the null hypotheses (6)-(7). We do not specify any parametric forms on μ and γ_i under the alternative hypotheses and wish to test the corresponding null hypotheses against as large class of alternatives as possible. In particular, assume that m_i belong to some Besov ball of radius C on $[0,1]^d$, $B^s_{p,q}(C)$, where $1 \leq p, q \leq \infty$, $s > 0$. Then, obviously, μ and γ_i belong to $B^s_{p,q}(C)$ as well. Besov classes are chosen because of their exceptional expressive power: for the particular choices of parameters s, p and q they include among others the Hölder and Sobolev classes of smooth functions, functions of bounded variation, etc. (e.g., Härdle *et al.*, 1998). In addition, since 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, we also need to assume that the set of alternative hypotheses is separated away from the set of null hypotheses (6)-(7) in the $L^2([0,1]^d)$-distance by ρ. Hence, denoting hereafter the $L^2([0,1]^d)$-norm by $||\cdot||_2$, we consider the alternative hypotheses to be, respectively, of the form

$$H_1 \quad : \quad \mu \in \mathcal{F}(\rho), \tag{8}$$

$$H_1 \quad : \quad \gamma_i \in \mathcal{F}(\rho), \quad \text{at least for one} \quad i = 1, \ldots, r, \tag{9}$$

where $\mathcal{F}(\rho) = \{f \in B^s_{p,q}(C) : ||f||_2 \geq \rho\}$.

The objective of this paper is to derive a consistent test that achieves the optimal minimax rate uniformly over the whole range of Besov balls under the alternatives (8)-(9).

2.2 Basic definitions

We start from basic definitions of the functional hypothesis testing. Suppose we are given data

$$dZ(\mathbf{t}) = f(\mathbf{t})\, d\mathbf{t} + \epsilon\, dW(\mathbf{t}), \quad \mathbf{t} \in [0,1]^d,$$

where W is a d-dimensional standard Wiener process. We wish to test

$$H_0 : f \equiv 0 \quad \text{versus} \quad H_1 : f \in \mathcal{F}(\rho), \tag{10}$$

where $\mathcal{F}(\rho) = \{f \in B^s_{p,q}(C) : ||f||_2 \geq \rho\}$.

A (nonrandomized) test ϕ is a measurable function of the observations with two values $\{0,1\}$, where $\phi = 0$ and $\phi = 1$ correspond to accepting and rejecting the null hypothesis H_0 respectively. As usual, the quality of test ϕ is measured by *Type I* (erroneous rejection of H_0) and *Type II* (erroneous acceptance of H_0) errors. The probability of Type I error is defined as

$$\alpha(\phi) = P_{f\equiv 0}(\phi = 1),$$

while the probability of Type II error for the composite nonparametric alternative hypothesis H_1 is defined as

$$\beta(\phi, \rho) = \sup_{f\in\mathcal{F}(\rho)} P_f(\phi = 0).$$

In this paper we focus on the asymptotic behavior of functional hypothesis testing procedures as the noise level $\epsilon \to 0$. It is clear that the smaller the noise level, the less separated from zero alternatives can be detected without loosing accuracy. Thus, it is natural to consider the optimal (fastest) rate of decay to zero of the "discrimination threshold" ρ (as a function of ϵ) as $\epsilon \to 0$ for prescribed α and β. In other words, we seek such $\rho(\epsilon)$ that

(i) for any $\rho'(\epsilon)$ satisfying

$$\rho'(\epsilon)/\rho(\epsilon) = o_\epsilon(1),$$

one has

$$\inf_{\phi_\epsilon} [\alpha(\phi_\epsilon) + \beta(\phi_\epsilon, \rho'(\epsilon))] = 1 - o_\epsilon(1),$$

where $o_\epsilon(1)$ is a sequence tending to zero as $\epsilon \to 0$.

(ii) for any $\alpha > 0$ and $\beta > 0$ there exists a constant $c > 0$ and a test ϕ^*_ϵ such that

$$\alpha(\phi^*_\epsilon) \leq \alpha + o_\epsilon(1)$$

$$\beta(\phi^*_\epsilon, c\rho(\epsilon)) \leq \beta + o_\epsilon(1).$$

The corresponding test ϕ^*_ϵ is called an asymptotically optimal (minimax) test. Ingster (1993a, 1993b, 1993c) and Lepski & Spokoiny (1999) showed that for $sp > d$ the optimal (minimax) rate for such a testing problem is

$$\rho(\epsilon) = \epsilon^{4s''/(4s''+d)}, \tag{11}$$

where $s'' = \min(s, s - \frac{d}{2p} + \frac{d}{4})$, and derived tests that achieve this optimal rate. However, the proposed tests were *nonadaptive* in the sense that they involved the smoothness parameter s of the corresponding Besov ball which is usually unknown in practice. Spokoiny (1996), Horowitz & Spokoiny (1999) considered the problem of *adaptive* testing where s is unknown *a priori* but assumed to lie within a range $[s_{\min}, s_{max}]$. They showed that no adaptive test can achieve the exact optimal rate (11) uniformly over all s –"lack of adaptivity" property of the functional hypothesis testing problem, and there is always a price to pay for adaptivity. However, it turned out that the price is remarkably low. If one allows to increase $\rho(\epsilon)$ by an additional log-log adaptive factor $t_\epsilon = (\log\log \epsilon^{-2})^{1/4}$, i.e. to consider $\rho(\epsilon t_\epsilon)$, then Horowitz & Spokoiny (1999) showed that the optimal rate of adaptive testing is

$$\rho(\epsilon t_\epsilon) = (\epsilon t_\epsilon)^{4s''/(4s''+d)}, \tag{12}$$

which is only within a log-log factor of (11). Moreover, the resulting adaptive test is uniformly consistent, i.e. $\beta(\phi^*_\epsilon, c\rho(\epsilon t_\epsilon)) = o_\epsilon(1)$ for all $s \in [s_{\min}, s_{max}]$ for some constant c. The adaptive factor t_ϵ is unavoidable and cannot be reduced. For more details we refer to Horowitz & Spokoiny (1999).

2.3 Wavelet background

Since the test statistics we shall develop for the main effects and the interactions of the FANOVA model (1)-(2) will be based on appropriate wavelet decompositions, we recall briefly some relevant facts about wavelets. For detailed expositions of the mathematical aspects of wavelets we refer, for example, to Meyer (1992), Daubechies (1992) and Mallat (1999), while comprehensive expositions and reviews on wavelets applications in statistical settings are given, for example, in Antoniadis (1997), Vidakovic (1999), Abramovich *et al.* (2000) and Antoniadis *et al.* (2001).

To simplify the notation, we consider the case $d = 1$ and work with orthonormal periodic wavelet bases in $L^2([0, 1])$ generated by dilations a compactly supported scaling function ϕ and dilations and translations of the corresponding compactly supported mother wavelet ψ (see, for example, Mallat, 1999, Section 7.5.1) :

$$\phi^{\mathrm{p}}(t) = \sum_{l\in\mathbb{Z}} \phi(t-l) \quad \text{and} \quad \psi^{\mathrm{p}}_{jk}(t) = \sum_{l\in\mathbb{Z}} \psi_{jk}(t-l), \quad j \geq 0, \quad k = 0, ..., 2^j - 1$$

where

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

The collection

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

generates an orthonormal periodic wavelet basis in $L^2([0, 1])$. Despite the poor behavior of periodic wavelets near the boundaries, where they yield large coefficients for nonperiodic functions, they are commonly used because their numerical implementation is particular simple. To unify the notations the scaling function ϕ^p will be denoted by $\psi^p_{-1,0}$ and the superscript "p" will be suppressed for convenience. For any $f \in L^2([0, 1])$ we denote by $\theta_{-10} = \langle f, \psi_{-10}\rangle$ the scaling coefficient and by $\theta_{jk} = \langle f, \psi_{jk}\rangle$, $j \geq 0$, $k = 0, 1, \ldots, 2^j - 1$ the wavelet coefficients of f for the orthonormal periodic wavelet basis defined above.

We end this section by noting that if the scaling function ϕ (and, thus, the mother wavelet ψ) is of regularity $r > 0$, the corresponding wavelet basis is an unconditional basis in the

Besov spaces $B^s_{p,q}([0,1])$ for $0 < s < r$, $1 \le p, q \le \infty$. This allows one to characterize Besov balls in terms of wavelet coefficients (e.g., Meyer, 1992).

3. MAIN RESULTS

In this section we apply the results of Spokoiny (1996) and Horowitz & Spokoiny (1999) to derive asymptotically optimal (minimax) nonadaptive and adaptive functional hypothesis testing procedures for the main effects and the interactions of the FANOVA model (1)-(4). To simplify the exposition and to emphasize the main idea we consider in detail the case $d = 1$. This restriction will be relaxed in Remark 3.5 below wherein we briefly explain straightforward extensions to the case $d > 1$.

3.1 Testing in the functional analysis of variance models

Averaging over the r observed paths of the FANOVA model (1)-(2) yields

$$d\overline{Y}(t) = \frac{1}{r}\sum_{i=1}^{r} dY_i(t) = \left(\frac{1}{r}\sum_{i=1}^{r} m_i(t)\right) dt + \epsilon\, d\overline{W}(t), \quad t \in [0,1],$$

where $\overline{W}$ is the average of r independent standard Wiener processes on $[0,1]$. Given the identifiability conditions (3)-(4) for the components of m_i, the latter equation can be rewritten as

$$d\overline{Y}(t) = (m_0 + \mu(t))\, dt + \epsilon\, d\overline{W}(t), \quad t \in [0,1] \tag{13}$$

which in view of (1)-(2) implies

$$d(Y_i - \overline{Y})(t) = a_i + \gamma_i(t) + \epsilon\, d(W_i - \overline{W})(t), \quad i = 1, \ldots, r; \quad t \in [0,1]. \tag{14}$$

By the basic properties of the increments of a standard Wiener process on $[0,1]$, the stochastic processes $\{W_i - \overline{W};\ i = 1, \ldots, r\}$ are Wiener processes with the same covariance kernel $C(s,t) = \frac{r-1}{r}\min(s,t)$, though they are no longer independent. Hence, the model (13) and each of the i models stated in the equation (14) are still diffusion models which can be written in the following general form

$$dZ(t) = f(t)\, dt + \eta\, dW(t), \quad t \in [0,1], \tag{15}$$

where $Z(t) = \overline{Y}(t),\ f(t) = m_0 + \mu(t),\ \ \eta = \frac{\epsilon}{\sqrt{r}}$ and $Z(t) = (Y_i - \overline{Y})(t),\ f(t) = a_i + \gamma_i(t)$, $\eta = \epsilon\sqrt{(r-1)/r}$ for the models (13) and (14), respectively.

In both cases, under the null hypotheses (6)-(7), f is a constant function though for the latter case, the composite null hypothesis contains r constraints of this type. Thus, our goal is to derive an optimal test for testing

$$H_0:\ f \equiv \text{constant} \quad \text{versus} \quad H_1:\ f - \int_0^1 f(t)dt \in \mathcal{F}(\rho),$$

in the general diffusion model (15), where $\mathcal{F}(\rho) = \{f \in B^s_{p,q}(C) : ||f||_2 \geq \rho\}$ and, obviously, constant $= \int_0^1 f(t)dt$.

Choose a mother wavelet ψ of regularity $r > s$. Performing the periodic wavelet transform (see Section 2.3) on (15), one has

$$Y_{jk} = \theta_{jk} + \epsilon\, \xi_{jk}, \quad j \geq -1; \quad k = 0, 1, \ldots, 2^j - 1, \tag{16}$$

where $Y_{jk} = \int_0^1 \psi_{jk}(t)dZ(t)$, $\theta_{jk} = \int_0^1 \psi_{jk}(t)f(t)dt$ and ξ_{jk} are independent $N(0,1)$ random variables. Note that under the null hypothesis $H_0:\ f \equiv$ constant, the only possibly nonzero coefficient of the wavelet decomposition of f is the scaling coefficient $\theta_{-10} = \int_0^1 f(t)dt$. Therefore, testing

$$H_0:\ f \equiv \text{constant}$$

is equivalent to testing

$$H_0:\ \theta_{jk} = 0 \quad \forall\, j \geq 0; \quad k = 0, 1, \ldots, 2^j - 1.$$

The above testing problem differs from $H_0:\ f \equiv 0$ in (10) studied by Spokoiny (1996) in a Gaussian "signal plus noise" model only by removing the requirement $\theta_{-10} = 0$ under the null hypothesis. Obviously this only difference does not affect the asymptotic properties of the resulting functional hypothesis testing procedures and we can, therefore, apply the results obtained by Spokoiny (1996) to develop asymptotically optimal tests for our FANOVA setting.

To apply the results of Spokoiny (1996), we assume that the parameters of the corresponding Besov ball $B^s_{p,q}(C)$ satisfy $1 \leq p, q \leq \infty$, $sp > 1$ and $s - \frac{1}{2p} + \frac{1}{4} > 0$. Such assumptions are common in wavelet function estimation (e.g., Donoho *et al.*, 1995; Donoho & Johnstone, 1998). Hence, summarizing, we consider the general diffusion model (15) and want to test

$$H_0: \ f \equiv \text{constant } (= \int_0^1 f(t)dt) \quad \text{versus} \quad H_1: \ (f - \int_0^1 f(t)dt) \in \mathcal{F}(\rho),$$

where $\mathcal{F}(\rho) = \{f \in B^s_{p,q}(C) : ||f||_2 \geq \rho\}$, $sp > 1$ and $s - \frac{1}{2p} + \frac{1}{4} > 0$.

3.2 Minimax tests

Nonadaptive minimax test Consider first the case where all the parameters s, p, q and the radius C of the corresponding Besov ball $B^s_{p,q}(C)$ are known. Let $s'' = \min\{s, s - \frac{1}{2p} + \frac{1}{4}\}$. Note that, when $p \geq 2$, the condition $sp > 1$ leads to $s''p > 1$ while, for $1 \leq p < 2$, the condition $sp > 1$ implies $s''p > \frac{3}{4}$. Under such conditions the minimax rate of testing in (3.1) is given by

$$\rho(\eta) = \eta^{4s''/(4s''+1)}.$$

(see Spokoiny, 1996, Theorem 2.1).

Now we construct the test that achieves this optimal rate. Let j_η be the largest possible integer such that $j_\eta \leq \log(\eta^{-2})$. In fact, asymptotically we can assume that

$$j_\eta = \log(\eta^{-2}).$$

Let also $j(s)$ be the resolution level given by

$$j(s) = \frac{2}{4s'' + 1} \log_2 \left(C\eta^{-2} \right).$$

Here we assume that the right-hand sides of the above expressions are integers; otherwise we take the integer parts of these quantities. Note that, for any admissible value of s, $j(s) < j_\eta$ and that $j(s)$, $j_\eta \to \infty$ as $\eta \to 0$. Let $\mathcal{J} = \mathcal{J}_- \cup \mathcal{J}_+$ where $\mathcal{J}_-$ is the set of resolution levels below $j(s)$ and $\mathcal{J}_+$ is the set of resolution levels between $j(s)$ and j_η, i.e.

$$\mathcal{J}_- = \{0, \ldots, j(s) - 1\}, \quad \mathcal{J}_+ = \{j(s), \ldots, j_\eta\}.$$

For each $j \in \mathcal{J}_-$, define S_j as

$$S_j = \sum_{k=0}^{2^j-1} (Y_{jk}^2 - \eta^2) \tag{17}$$

while, for each $j \in \mathcal{J}_+$ and for given $\lambda > 0$, define $S_j(\lambda)$ as

$$S_j(\lambda) = \sum_{k=0}^{2^j-1} [(Y_{jk}^2 \mathbf{1}(|Y_{jk}| > \eta\lambda) - \eta^2 b(\lambda)], \tag{18}$$

where $\mathbf{1}(A)$ is the indicator function of the set A, $b(\lambda) = \mathbb{E}\,[\xi^2 \mathbf{1}(|\xi| > \lambda)]$ and ξ is a $N(0,1)$ random variable. Note that the terms in $S_j(\lambda)$ are defined by applying hard thresholding on the empirical wavelet coefficients, which is a standard procedure for minimax wavelet estimation in nonparametric regression settings (e.g., Donoho *et al.*, 1995).

With the above notation, introduce the following test statistics

$$T(j(s)) = \sum_{j=0}^{j(s)-1} S_j, \tag{19}$$

and

$$Q(j(s)) = \sum_{j=j(s)}^{j_\eta - 1} S_j(\lambda_j), \tag{20}$$

where $\lambda_j = 4\sqrt{(j - j(s) + 8)\ln 2}$. Let also $v_0^2(j(s))$ and $w_0^2(j(s))$ be the variances of $T(j(s))$ and $Q(j(s))$, respectively, under H_0. It is easy to see that

$$v_0^2(j(s)) = 2\eta^4 2^{j(s)} \quad \text{and} \quad w_0^2(j(s)) = \eta^4 \sum_{j=j(s)}^{j_\eta - 1} 2^j d(\lambda_j),$$

where $d(\lambda_j) = \mathbb{E}\,[\xi^4 \mathbf{1}(|\xi| > \lambda_j)]$. Using the results of Fan (1996), one gets the following asymptotic expressions

$$\mathbb{E}(\xi^{2k} \mathbf{1}(|\xi| > \lambda_j)) = \sqrt{2/\pi}\lambda_j^{2k-1} 2^{-8(j-j(s)+8)} + O\left(\lambda_j^{2k-3} 2^{-8(j-j(s)+8)}\right), \quad k = 1, 2, \ldots . \tag{21}$$

However, such an approximation might be poor in finite sample examples and result in too low values of $d(\lambda_j)$ and $b(\lambda_j)$. A better approximation for $d(\lambda_j)$ is obtained by using the Taylor expansions for $\lambda_j \le 1$ and the Laurent expansions for $\lambda_j > 1$ respectively :

$$d(\lambda_j) = 3 - \sqrt{2/\pi}\Lambda^5/5 + \Lambda^7/(7\sqrt{2\pi}) + o(\Lambda^8), \tag{22}$$

where $\Lambda = \min(\lambda_j, 1/\lambda_j)$. Similar approximations can be derived for $b(\lambda_j) = \mathbb{E}(\xi^2 \mathbf{1}(|\xi| > \lambda_j))$.

Finally, for a given significance level $\alpha \in (0,1)$, let ϕ^* be the test defined by

$$\phi^* = \begin{cases} \mathbf{1}\left\{T(j(s)) > v_0(j(s))Z_{1-\alpha}\right\} & \text{if } p \geq 2 \\ \mathbf{1}\left\{T(j(s)) + Q(j(s)) > \sqrt{v_0^2(j(s)) + w_0^2(j(s))}Z_{1-\alpha}\right\} & \text{if } 1 \leq p < 2, \end{cases} \tag{23}$$

where $Z_{1-\alpha}$ is $(1-\alpha)100\%$-th percentile of the $N(0,1)$ distribution.

The following proposition, whose proof is given in the Appendix, establishes the asymptotic optimality of ϕ^*:

Proposition 3.1 *Let the mother wavelet ψ be of regularity $r > s$ and let the parameters s, p, q and the radius C of the Besov ball $B^s_{p,q}$ be known, where $1 \leq p, q \leq \infty$, $sp > 1$ and $s - \frac{1}{2p} + \frac{1}{4} > 0$. Then, for a fixed significance level $\alpha \in (0,1)$, the test ϕ^*, defined in (23), for testing*

$$H_0 : f \equiv \quad \textit{constant}\ (= \int_0^1 f(t)dt) \quad \textit{versus} \quad H_1 : (f - \int_0^1 f(t)dt) \in \mathcal{F}(\rho),$$

where $\mathcal{F}(\rho) = \{f \in B^s_{p,q}(C) : ||f||_2 \geq \rho\}$, is level-$\alpha$ asymptotically optimal (minimax) test, as $\eta \to 0$. That is, for any $\beta \in (0,1)$, it attains the asymptotically minimax rate of testing

$$\rho(\eta) = \eta^{4s''/(4s''+1)}, \quad \text{as} \quad \eta \to 0,$$

where $s'' = \min\{s, s - \frac{1}{2p} + \frac{1}{4}\}$.

Remark 3.1 For $p \geq 2$, the test defined in (23) differs from that developed by Spokoiny (1996), who proposed to perform *nonlinear* thresholding for all $j \geq j(s)$ regardless of p, while we suggest instead for $p \geq 2$ to simply truncate wavelet series at level $j(s) - 1$. These results are similar to those known for nonparametric estimation of quadratic functionals where *linear* estimators are still optimal for $p \geq 2$ and $sp > 1$ (e.g., Efromovich & Low, 1996).

Adaptive minimax test The structure of the rate-optimal tests developed in the Section 3.2 essentially relies on the smoothness of the underlying function via the parameter s of the corresponding Besov ball while such kind of prior information is typically lacking in practical applications. Our aim now is to develop an *adaptive* functional hypothesis testing procedure that does not require the knowledge of s and achieves an optimal testing rate up to an unavoidable log-log factor (see Section 2.2).

We consider now the case where the parameters s, p, q and the radius C of the corresponding Besov ball $B^s_{p,q}(C)$ are unknown but assume that $0 < s < s_{\max}$, $1 \le p, q \le \infty$, $sp > 1$ and $s - \frac{1}{2p} + \frac{1}{4} > 0$. The corresponding range of these parameters will be denoted by $\mathcal{T}$. Let $t_\eta = (\log\log \eta^{-2})^{1/4}$ and $j_{\min} = \frac{2}{4s''_{\max}+1} \log \eta^{-2}$, where $s''_{\max} = \min\{s_{\max}, s_{\max} - \frac{1}{2p} + \frac{1}{4}\}$. Suppose again that the mother wavelet ψ is of regularity $r > s_{\max}$.

The idea of adaptive test is to consider the whole possible range of $j(s) = j_{\min}, \ldots, j_\eta - 1$ and reject H_0 if it is rejected at least for one level $j(s)$. Since $\mathrm{card}(\{j_{\min}, \ldots, j_\eta - 1\}) = O(\ln \eta^{-2})$, performing a Bonferonni type testing leads to the following asymptotic *adaptive* test :

$$\phi^*_\eta = \mathbf{1}\left[\max_{j_{\min} \le j(s) < j_\eta} \left\{ \frac{T(j(s)) + Q(j(s))}{\sqrt{v_0^2(j(s)) + w_0^2(j(s))}} \right\} > \sqrt{2 \ln\ln \eta^{-2}} \right]. \tag{24}$$

Spokoiny (1996) showed that the test ϕ^*_η defined in (24) is an adaptive optimal test, i.e. such that

$$\alpha(\phi^*_\eta) = o_\eta(1)$$

and

$$\sup_{\mathcal{T}} \beta(\phi^*_\eta, c\rho(\eta t_\eta)) = o_\eta(1),$$

where $\rho(\eta t_\eta) = (\eta t_\eta)^{4s''/(4s''+1)}$, $o_\eta(1)$ is a sequence tending to zero as $\eta \to 0$ and c is a constant. If, in addition, it is known that $p \ge 2$ then, similar to (23), the above adaptive test defined in (24) can be simplified as

$$\phi^*_\eta = \mathbf{1}\left[\max_{j_{\min} \le j(s) < j_\eta} \left\{ \frac{T(j(s))}{\sqrt{v_0^2(j(s))}} \right\} > \sqrt{2 \ln\ln \eta^{-2}} \right]. \tag{25}$$

The proof of this assertion is given in the Appendix.

We finish this section with several remarks :

Remark 3.2 From (25) one can see that unlike adaptive nonparametric function estimation where nonlinear procedures are *always* necessary to achieve the adaptive minimax rates, the situation for testing for the regular case $p \geq 2$, $sp > 1$ is different. This phenomenon is also known in quadratic functionals estimation (e.g., Efromovich & Low, 1996).

Remark 3.3 The test statistic of ϕ^*_η in (24) is essentially a sum of squares of thresholded centered empirical wavelet coefficients of the data curve with the properly chosen *level-dependent* thresholds. In fact, it is similar in spirit to that used in the Neyman's test in Fan (1996), Fan & Lin (2000) though they apply a certain *global* threshold.

Remark 3.4 The results obtained in this section remain true if *different* (sufficiently regular) mother wavelets are used for $\mu(t)$ and different $\gamma_i(t)$.

Remark 3.5 The extension of the above nonadaptive and adaptive tests to the case $d > 1$ is straightforward using the d-dimensional periodic wavelet transform. Note that the optimal (minimax) rate of testing in this case is $\rho(\eta) = \eta^{4s''/(4s''+d)}$ (e.g., Horowitz & Spokoiny, 1999) and the additional factor t_η for the adaptive test remains the same. It is easy to show that in order to achieve this rate, one should perform essentially the same procedures as for $d = 1$, but based on the empirical coefficients of a d-dimensional periodic wavelet transform with similar statistics within the resolution range $j(s) = 2(4s''_{\max}+d)^{-1} \log \eta^{-2}, \ldots, d^{-1} \log \eta^{-2} - 1$, where $s'' = \min\{s, s - \frac{d}{2p} + \frac{d}{4}\}$, $sp > d$ and $s - \frac{d}{2p} + \frac{d}{4} > 0$.

Applications to the functional analysis of variance models Here we apply the derived functional hypothesis testing procedures (23), (24) and (25) to our FANOVA setting.

To test the main effect $H_0 : \mu \equiv 0$, we apply (23), (24) and (25) directly, using the average process $\overline{Y}$ defined in (13) as Z in (15), and setting $\eta = \frac{\epsilon}{\sqrt{r}}$.

To test the interaction components, note first that testing

$$H_0 : \gamma_i \equiv 0, \quad \forall\, i = 1, \ldots, r,$$

is equivalent to testing

$$H_0 : \theta_{jk}^{(i)} = 0, \quad \forall\, i = 1, \ldots, r, \quad j \geq 0; \quad k = 0, \ldots, 2^j - 1,$$

where $\theta_{jk}^{(i)}$ are the wavelet coefficients of γ_i. To apply (23), (24) and (25), define

$$S_j^{(i)} = \sum_{k=0}^{2^j-1} [(Y_{jk}^{(i)})^2 - \eta^2]$$

for each $j \in \mathcal{J}_-$, and, for each $j \in \mathcal{J}_+$,

$$S_j^{(i)}(\lambda_j) = \sum_{k=0}^{2^j-1} [(\{Y_{jk}^{(i)}\}^2 \mathbf{1}(|Y_{j,k}^{(i)}| > \eta\lambda_j) - \eta^2 b(\lambda_j)],$$

where $Y_{jk}^{(i)}$ are the empirical wavelet coefficients of $d(Y_i - \overline{Y})(t)$ in (14), $\eta = \epsilon\sqrt{(r-1)/r}$ and $\lambda_j = 4\sqrt{(j - j(s) + 8)\ln 2}$.

Define now

$$T(j(s)) = \sum_{i=1}^{r} T^{(i)}(j(s)) = \sum_{i=1}^{r} \sum_{j=0}^{j(s)-1} S_j^{(i)}$$

and

$$Q(j(s)) = \sum_{i=1}^{r} Q^{(i)}(j(s)) = \sum_{i=1}^{r} \sum_{j=0}^{j(s)-1} S_j^{(i)}(\lambda_j).$$

Although both $T^{(i)}(j(s))$ and $Q^{(i)}(j(s))$ are correlated for different i (see comments after (14)), we have

$$\mathrm{Var}(T(j(s)) \;\leq\; r^2\, \mathrm{Var}(T^{(i)}(j(s))) \;=\; 2r^2\eta^4 2^{j(s)}$$

and

$$\mathrm{Var}(Q(j(s)) \;\leq\; r^2\, \mathrm{Var}(Q^{(i)}(j(s))) \;=\; r^2\eta^4 \sum_{j=j(s)}^{j_\eta - 1} 2^j d(\lambda_j).$$

We can therefore apply (23), (24) and (25) with

$$v_0^2(j(s)) = 2r^2\eta^4 2^{j(s)} \quad \text{and} \quad w_0^2(j(s)) = r^2\eta^4 \sum_{j=j(s)}^{j_\eta - 1} 2^j d(\lambda_j).$$

The additional r^2 factor does not depend on η and does not affect the asymptotical properties of the proposed tests (see Appendix).

Finally, it is important to point out that in practice one always deals with *discrete* data and, therefore, should apply the *sampled* versions of the tests proposed in Sections 3.2 and 3.2 to the empirical coefficients of the *discrete* periodic wavelet transforms (e.g., Vidakovic, 1999, Chapter 5.6). To justify the optimality of the sampled versions of the tests we can exploit the general asymptotic equivalence results between the Gaussian white noise model and the corresponding Gaussian nonparametric regression setting with equispaced design and variance $\sigma^2 = n\epsilon^2$ (see Brown & Low, 1996), and the more specific analogous results of Donoho & Johnstone (1999) within the wavelet framework. Although the proposed methodology is simple and powerful, the practical use of discrete wavelet transforms requires the equispaced data points. This requirement can be relaxed by preprocessing the data by either binning or interpolation methods (e.g., Antoniadis & Pham, 1998; Cai & Brown, 1998; Kovac & Silverman, 2000; Amato & Antoniadis, 2001; Antoniadis & Fan, 2001). Once the data are preprocessed, the wavelet transform of the preprocessed data yields the appropriate test statistics. Note however, that such preprocessing induces some dependency, and our "white noise model" approach may need some modifications. It remains to be seen how this preprocessing will affect optimality of the tests.

In most applications, σ is unknown and its estimation is crucial for the success of the functional hypothesis testing procedures described above. In wavelet function estimation, the common practice is to robustly estimate σ by the median of absolute deviation of the empirical wavelet coefficients of the data at the highest resolution level divided by 0.6745 (see Donoho & Johnstone, 1994, 1995). In the FANOVA model (1), unless there are replications, we estimate σ by averaging its r robust estimates obtained from each individual data curve. Note that in this case the estimate of σ is *independent* from the test statistics (23)-(25) that do not involve empirical wavelet coefficients from the finest level. In the one dimensional case and for smooth alternatives one could also use the nonparametric estimators described in Hall *et al.* (1990, 1991) and Dette *et al.* (1998), but their methods are not easy to extend to higher dimensional settings. Another reliable method, easily extendable to large dimensions, is the one proposed by Huang & Cressie (2000) which is based on a robust estimator of the semivariogram of the observed process at some specific lags.

4. NUMERICAL EXAMPLES

The purpose of this section is to shade some light on the theoretical results discussed in Section 3. First, we carry out a small simulation study to investigate the *finite sample* performance of the proposed tests. Then, we apply these tests to two real-life data examples arising from endocrinology and from neuropsychology.

The computational algorithms related to wavelet analysis were performed using Version 8 of the WaveLab toolbox for MATLAB that is freely available from `http://www-stat.stanford.edu/softw` The entire study was carried out using the MATLAB programming environment.

4.1 Simulation Study

The simulation study is based on synthetic data formed by using the battery of standard test functions of Donoho & Johnstone (1994, 1995) frequently used in wavelet benchmarking : the `BLOCKS`, `BUMPS`, `DOOPLER` and `HEAVISINE` functions. We add the additional test function `MISHMASH`, defined as

$$\texttt{MISHMASH} = -(\ \texttt{BLOCKS} + \ \texttt{BUMPS} + \texttt{DOOPLER} + \texttt{HEAVISINE}),$$

to satisfy the first part of the identifiability condition (4), i.e., to ensure that the sum of all functions is zero at each point.

The observations are simulated as discretized versions of equations (1)-(2), satisfying discretized identifiability conditions (3)-(4). At n equispaced time points $t_j = j/n,\ j = 1, \ldots, n$, the data are generated as multivariate vectors $y_i(t_j),\ i = 1, \ldots, 5$ defined as the sum of a constant $m_0 = 1$, the mean function $\mu(t) = 2\sin(2\pi t)$ (the main effect of t), a corresponding test function and a Gaussian noise of a given size. The test functions actually represent the *group effects* and can be decomposed as $a_i + \gamma_i(t)$, where the *main group effects* a_i are the integrals of the original test functions and the *interaction components* $\gamma_i(t)$ are their *centered* versions so that $\int_0^1 \gamma_i(t)dt = 0$. The mean function $\mu(t)$ and the functions $\gamma_i(t)$ are depicted in Figure 4.1.

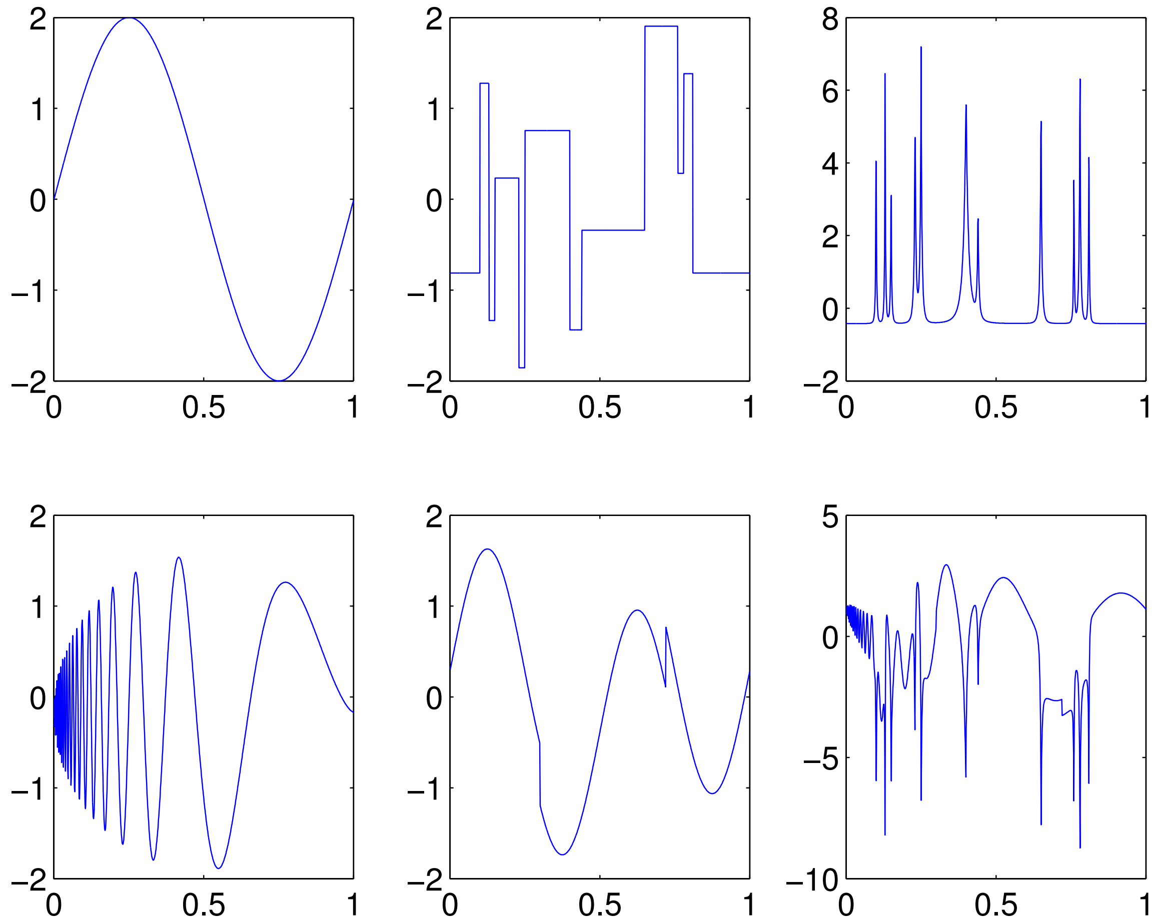


Figure 4.1: (Clockwise) The mean function $\mu(t) = 2\sin(2\pi t)$ and the centered group effects functions $\gamma_i(t)$, $i = 1, \ldots, 5$ (centered `BLOCKS`, `BUMPS`, `MISHMASH`, `HEAVISINE` and `DOPPLER`) sampled at $n = 1024$ data points.

The size of noise is selected in accordance with the energy (or variance) of the test functions, i.e. their squared L^2-norm. The five test functions are in addition rescaled so that for all of them a noise of size 1 achieves the prescribed signal-to-noise ratio (SNR), defined as the ratio of standard deviations of the signal and of the noise. Five simulated observations (one for each test function shown in Figure 4.1) of a specific length ($n = 1024$), with two different SNRs (SNR = 3 and 7), are shown superimposed and separately in Figure 4.2.

We do not assume to know the standard deviation σ of the error terms but estimate it as suggested in Section 3.2. We now apply the nonadaptive FANOVA testing procedure discussed in Section 3.2 for the main effect and interaction terms separately at a significance

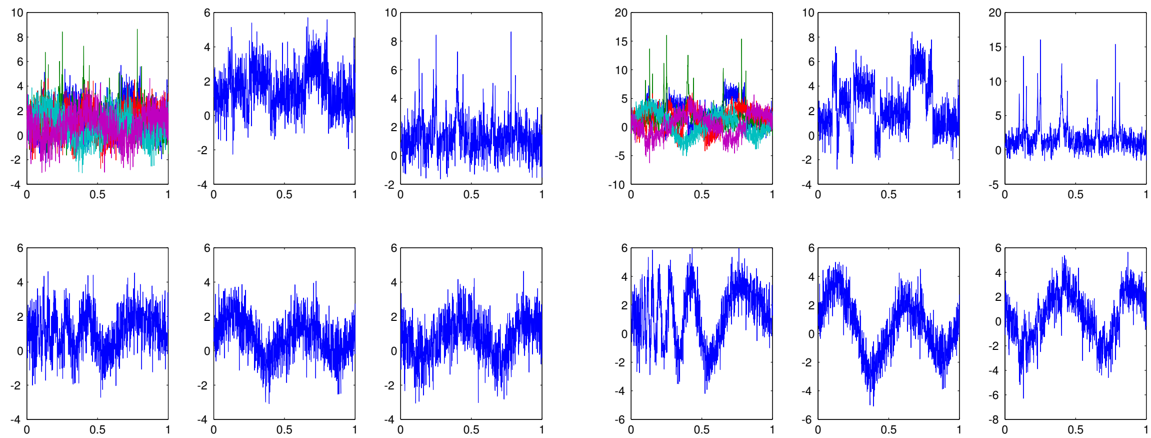


Figure 4.2: Five simulated observations (one for each test function shown in Figure 4.1) sampled at $n = 1024$ data points are shown superimposed (first plot) and separately (remaining five plots) for (a) SNR = 3 and (b) SNR = 7.

level $\alpha = 0.05$ for both SNRs.

As far as the hypothesis $H_0 : a_i = 0, \ \forall \ i = 1, \ldots, 5$ is concerned, there is a standard ANOVA testing procedure. Since $n = 1024$, under H_0 the resulting test statistic is asymptotically distributed as the χ^2_4 distribution. For the SNR=3 case, we get an observed $\chi^2 = 6.59$ with critical value $\chi^2_{4,0.05} = 9.488$ (p-value = 0.159). For the SNR=7 case, we get an observed $\chi^2 = 10.494$ with critical value $\chi^2_{4,0.05} = 9.488$ (p-value= 0.033). Note that after rescaling, $||a||_2 = 0.0934$ for SNR=3 and $||a||_2 = 0.2164$ for SNR=7.

To test the hypothesis $H_0 : \mu(t) = 0$, we used the truncation version ($p \geq 2$) of the nonadaptive FANOVA testing procedure defined in (23). Since μ is smooth, we have used the compactly supported *Symmlet 8-tap filter* mother wavelet and adopted $j(s) = 3$. This choice for $j(s)$ was based on our experience rather than on asymptotic considerations. Indeed, the function μ can be well described by its wavelet coefficients from just the three coarsest levels. For the case SNR=3, the value of the test statistic $T(3)$ was equal to 15.2851 to be compared with the critical value 1.5949 while, for the case SNR=7, the corresponding values of the test statistic and of the critical value were 97.5164 and 1.6316 respectively.

To test the hypothesis $H_0 : \gamma_i(t) = 0, \ \forall \ i = 1, \ldots, 5$, we applied the general (thresholding) version of the nonadaptive FANOVA testing procedure in (23). We used the compactly supported *Daubechies 6-tap filter* mother wavelet and set again $j(s) = 3$. Although more resolution levels are needed for less smooth functions γ_i, even for them, the wavelet coefficients on the seventh and higher levels can be hardly distinguished from noise that lead us to set $j_\eta = 7$. For SNR=3, the value of the test statistic $T(3) + Q(3)$ was equal to 275.3326 to be compared with the critical value 154.6294 while, for the case SNR=7, the corresponding values of the test statistic and of the critical value were 5941.0998 and 156.4943 respectively.

The fact that we estimated the noise level σ in the above tests does not seem to affect the asymptotic critical values in all cases and for all tested hypotheses considered above. Moreover, the corresponding tests performed well even for the low SNR.

The simulations showed that some asymptotical specifications face problems in finite sample situations. For instance, the thresholds $\lambda_j = C\sqrt{(j - j(s) + 8)\ln 2}$ with $C = 4$ used in defining the test statistics in expression (20) were unreasonably high, and finite sample situations called for $C = 1$. Moreover, using the asymptotic expression (21) derived in Fan (1996), one may get, in some finite sample situation, values of the $Q(j(s))$ statistics somewhere between 10^{-20} and 10^{-16}. In such cases we have used the approximation (22) for $d(\lambda_j)$.

We have also performed an extensive power analysis for the above tests against the composite alternatives

$$H_1 : \mu \in \mathcal{F}(\rho), \quad H_1 : \frac{1}{5}\sum_{i=1}^{5} \gamma_i \in \mathcal{F}(\rho) \quad \text{and} \quad H_1 : \frac{1}{5}(\sum_{i=1}^{5} |a_i|^2)^{1/2} = \rho > 0.$$

The significance levels were fixed at $\alpha = 5\%$. Fixing ρ to 1, the magnitude of the signals were changed to achieve a prescribed SNR. The graphs of *empirical* power functions (computed with 500 replications for each test and each SNR) against the SNR are given in Figure 4.3 and demonstrate how fast the power of the proposed procedure increases with increasing of SNR or, equivalently, of the L^2-distance ρ between the null hypothesis and the composite alternative. The sample size was taken to be $n = 512$. When the null hypotheses are true, the power is around the critical significance level of 5%. We see that the tests perform quite well. The difference in power for each type of the tested hypotheses corresponds to the different smoothness of the underlying functions. Thus, the simulation results illustrate that the developed FANOVA asymptotically rate-optimal tests have satisfactory power in the presence of relatively extreme alternatives.

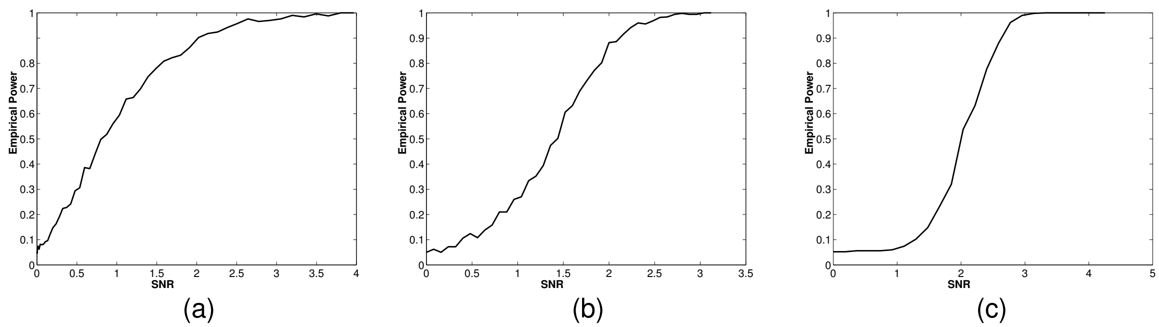


Figure 4.3: Empirical power functions for testing (a) $H_0 : a_i = 0,\ \forall\, i = 1, \ldots, 5$ versus $H_1 : \|a\|_2/5 = \rho$, (b) $H_0 : \mu(t) = 0$ versus $H_1 : ||\mu||_2 = \rho$, and (c) $H_0 : \gamma_i(t) = 0,\ \forall\, i = 1, \ldots, 5$ versus $H_1 : ||\sum_i \gamma_i/5||_2 = \rho$. In all three panels, the sample size was taken to be $n = 512$, the number of trials was set to 500, and the L^2-distance ρ was fixed at 1.

4.2 Urinary Metabolite Progesterone Curves

Urinary metabolite progesterone curves measured daily over 12 conceptive and 12 nonconceptive menstrual cycles were obtained by the Institute of Toxicology and Environmental Health at the University of California at Davis, USA. These samples came from patients with healthy reproductive function involved in an artificial insemination clinic where insemination attempts are well-timed for each menstrual cycle. 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. One of the aims of the analysis is to characterize differences in conceptive and nonconceptive menstrual cycles prior to implantation, which is typically done a week after ovulation.

From the original data set of 91 curves that exemplified the methods of Brumback & Rice (1998), we selected a subset of 24 curves (12 curves for each type of menstrual cycles) relevant for the FANOVA application. The selected curves correspond to different subjects and are missing data-free. Assuming the same noise variance σ^2 for both groups and averaging the data curves over subjects within each group, leads to a FANOVA model of the form (1) with $\epsilon = \sigma/\sqrt{12}$. Figure 4.4 shows the superimposed curves for the nonconceptive and conceptive menstrual cycles, together with the corresponding group means $\hat{m}_i(t),\ i = 1, 2$.

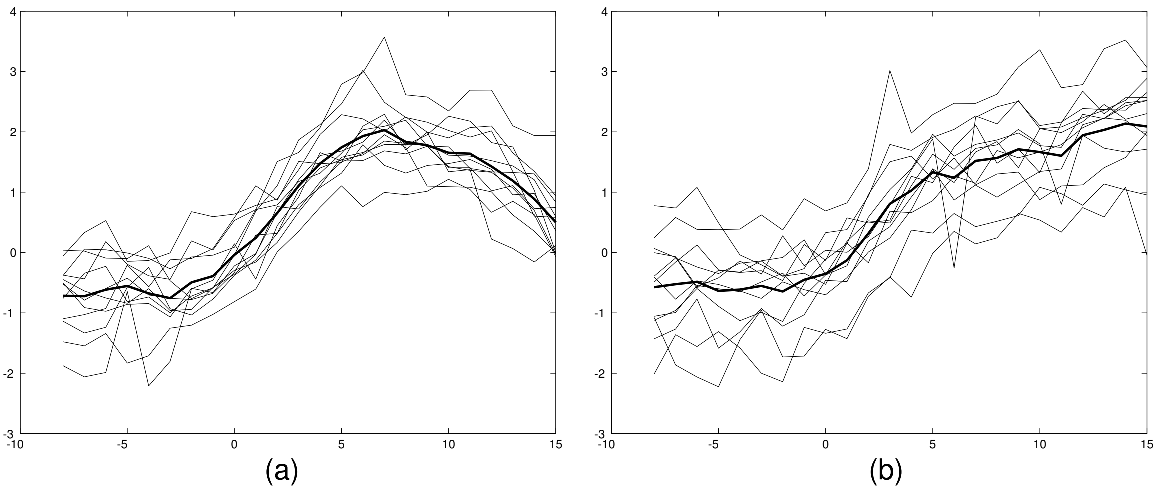


Figure 4.4: Urinary metabolite progesterone curves (thin lines) and the corresponding group means (thick lines): (a) Nonconceptive menstrual cycles; (b) Conceptive menstrual cycles

Figure 4.5 presents the group means $\hat{m}_i(t),\ i = 1, 2$ together with the centered overall mean $\hat{\mu}(t)$ and the centered group effects $\hat{\gamma}_i(t)$ in the FANOVA decompositions (2) of $\hat{m}_i(t)$. Figure 4.5 gives some ideas what one would expect from the tests. One can see that while the two centered group mean curves progress similarly during approximately 8 days before and 8 days after the day of ovulation, they show different trends from the eighth day after ovulation. The overall test $H_0 : a_i = 0,\ i = 1, 2$ for the main group effects does not take into account the longitudinal aspect of the data and yields an observed value of the test statistic $\chi^2 = 3.6367$, where the standard deviation $\hat{\sigma} = 0.0982$ was estimated as suggested in Section 3.2, while the critical value $\chi^2_{1,0.95} = 3.8415$. The corresponding p-value for this test is 0.0564 and the null hypothesis is therefore not rejected. The reason is clear since up to the eighth day after ovulation the two groups behave very similarly and the difference

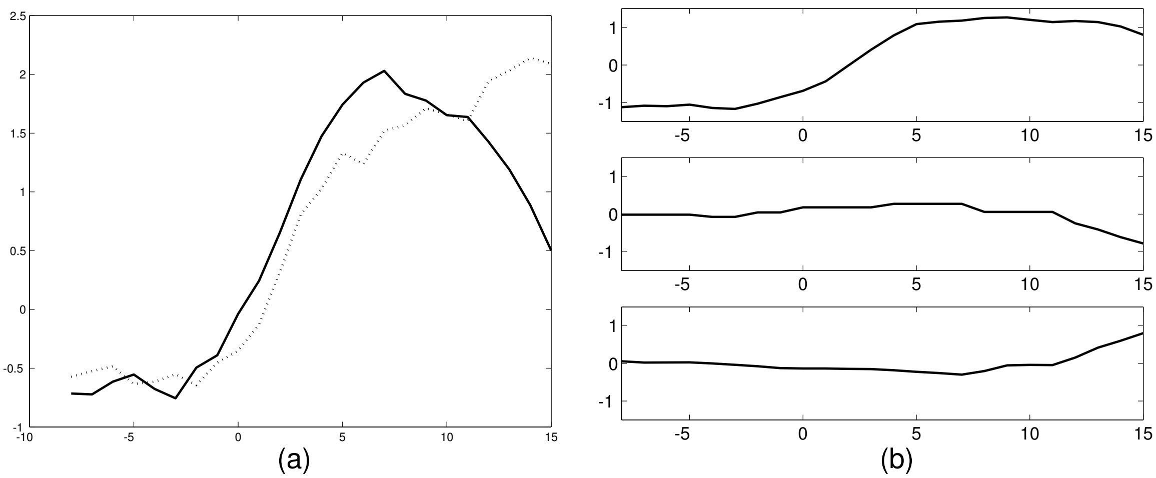


Figure 4.5: Urinary metabolite progesterone curves: (a) the group means for both conceptive and nonconceptive menstrual cycles; (b) the centered mean $\hat{\mu}(t)$ [top panel], the centered group effect $\hat{\gamma}_1(t)$ for nonconceptive menstrual cycles [middle panel], and the centered group effect $\hat{\gamma}_2(t)$ for conceptive menstrual cycles [bottom panel].

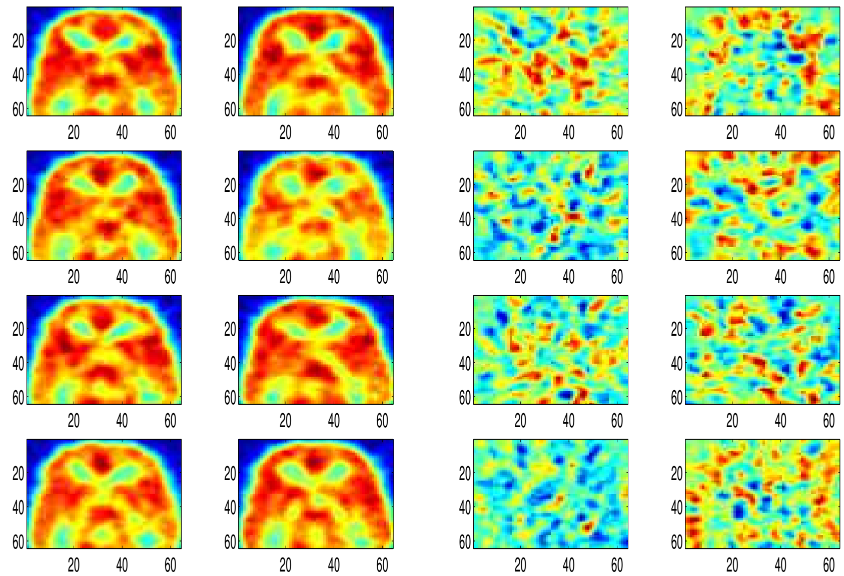
becomes evident only towards the end of a menstrual cycle.

The overall global difference between the two groups tested by our procedure is still dominated by the main part of the longitudinal behaviour of the response and in order to detect the differences accross time one should obviously rely upon some kind of local test. This exactly the purpose of the adaptive FANOVA procedure defined in (24) when it is applied to test the main time effect and the time-group interaction components. Indeed, the analysis with $j_{\min} = 1$ and $j_\eta = 3$ implies that the mean shape of these curves is not constant and they have different tendencies across time, since the null hypotheses $H_0 : \mu(t) = 0$ and $H_0 : \gamma_i(t) = 0,\ i = 1, 2$ are both rejected, where the compactly supported mother wavelets *Symmlet 8-tap filter* and *Daubechies 6-tap filter* were used, respectively, for testing. The corresponding values of the thresholding test statistics in (24) are respectively 457.8175 and 223.9303 to be compared with the threshold 1.9228 (note that for $r = 2$ the thresholds for these two tests are the same - see Section 3.2.3). Hence, the results of the tests coincide with our preliminary conclusions – although the *overall* progesterone profile is similar in both groups, its behaviour during a menstrual cycle is different.

4.3 Positron Emission tomography and Functional Magnetic Resonance Imaging Whole-Brain Data

Neuroimaging is a relatively new tool that provides information about brain functioning not before discovered. In the last decade, a great deal of progress has been made in these techniques and, despite its many limitations, functional neuroimaging seems to be a feasible and valuable tool that enables cognitive psychologists and neuroscientists to study the human brain in action. Two popular techniques are positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). In functional imaging, the information of clinical interest is usually the differences between images of two different activation states of the brain controlled by some experimental paradigm. As an application of testing in a two-dimensional FANOVA model, we consider a small portion of PET-fMRI data from a single subject experiment (see Kinahan & Noll, 1999). The subject used left hand to perform finger opposition task: touch thumb to index finger, to middle finger, to ring finger, and to pinky. Subject performed this task at a rate of 2 Hz, as guided by fixing a visual cue. For baseline, there was no finger movement, but the visual cue was still present. The goal of this experiment was to study the pattern of activity produced by the motor function of finger tapping. The fixation-only (no finger tapping) condition serves as a control: brain activity in response to finger tapping with fixation but not to fixation alone is attributed to the neural processing evoked by tapping.

The principal data from a PET-fMRI experiment is a sequence of three-dimensional images of the subject's brain. Each image consists of measurements of the magnetic resonance signal over a grid of small, regular volume elements, called *voxels*. A magnetic resonance image reveals the anatomic structure of the brain, but in PET-fMRI, one is not interested in the images *per se*, but rather in small systematic changes in the measured magnetic resonance signal. A direct statistical analysis of such data in the spatial domain is problematic because of a poor SNR, the large number of pixels that need to be investigated, and the relative spatial localization of the signals to be detected. This relative spatial localization is the



(a) (b)

Figure 4.6: (a) A selection of z = 8 slices in the 8 PET-fMRI measurements; (b) the centered group effects, $\hat{\gamma}_i(\mathbf{t}),\ i = 1, \ldots, 8$.

motivation for using the proposed wavelet-based FANOVA methods, because the wavelet-based derived tests are efficient in detecting local changes. The data are 8 realigned and normalized Analyze format $65 \times 87 \times 26$ scans where *left* is left and up is the front of the head by neurological convention. We selected one out of 26 possible z-positions ($z = 8$) to construct two-dimensional images for the FANOVA analysis. The x- and y-dimension sizes were curtailed to size of 64 each. The selected images are shown in Figure 4.6(a). Odd scans (1,3,5,7) are activation, even scans (2,4,6,8) are baseline.

As in the previous examples, the standard deviation σ of the error terms was estimated as suggested in Section 3.2 and $\hat{\sigma} = 0.0321$. We also used the compactly supported wavelets *Haar* and *Daubechies 6-tap filter* for testing the null hypotheses $H_0 : \mu(\mathbf{t}) = 0$ and $H_0 : \gamma_i(\mathbf{t}) = 0,\ \forall\, i = 1, \ldots, 8$, respectively. The estimates of the centered group effects functions, $\hat{\gamma}_i(\mathbf{t}),\ i = 1, \ldots, 8$, shown in Figure 4.6(b), and the estimate of the centered group mean, $\hat{\mu}(\mathbf{t})$, shown in Figure 4.7(a), are both significantly different from zero at significance level $\alpha = 0.05$, i.e. the hypotheses $H_0 : \gamma_i(\mathbf{t}) = 0,\ i = 1, \ldots, 8$ and $H_0 : \mu(\mathbf{t}) = 0$ are both

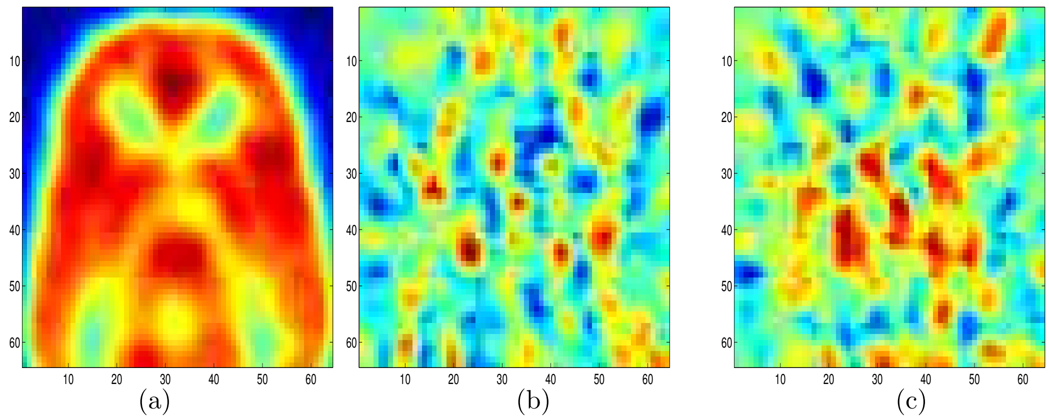


Figure 4.7: (a) The centered group mean, $\hat{\mu}(\mathbf{t})$, for the selection of $z = 8$ slices in the 8 PET-fMRI measurements. (b) Rescaled image of $\hat{\gamma}_2(\mathbf{t}) - \hat{\gamma}_8(\mathbf{t})$; (c) Rescaled image of $\hat{\gamma}_1(\mathbf{t}) - \hat{\gamma}_8(\mathbf{t})$.

rejected at significance level $\alpha = 0.05$ since the nonadaptive version of the FANOVA testing procedure defined in (23) produced p-values close to zero. Indeed, for adopted $j(s) = 3$ and $j_\eta = 7$, the values of $T(3)$ $(p \geq 2)$ for testing $H_0 : \mu(\mathbf{t}) = 0$ was 163.66, and $T(3) + Q(3)$ for testing $H_0 : \gamma_i(\mathbf{t}) = 0,\ i = 1, \ldots, 8$, was 42.22. In both cases the corresponding critical values were smaller than 1. Such outcomes to these two tests are expected.

Although it has not developed in this paper, exactly as in traditional analysis of variance, the proposed functional hypothesis testing procedures can be straightforwardly adapted to testing *contrasts* between the interactions γ_i. As an illustration, in the context of this example, we tested two additional hypotheses on contrasts: (i) $H_0 : \gamma_2(\mathbf{t}) = \gamma_8(\mathbf{t})$, (two baseline effects, say 2 and 8, are the same), and (ii) $H_0 : \gamma_1(\mathbf{t}) = \gamma_8(\mathbf{t})$ (index finger activation effect is equal to a baseline, say 8), using the compactly supported wavelet *Daubechies 6-tap filter*. Adopting again the nonadaptive version of the FANOVA testing procedure defined in (23) with $\eta = \sqrt{2}\epsilon$, both null hypotheses were rejected. However, the test statistics for the test in (i) was 4.740 while for the test in (ii) the corresponding test statistic was 6.054. Although the two baseline effects (2 and 8) are found to be significantly different (even though there was no finger movement), the visual cue, the complexity of the data and other uncontrolled experimental conditions may be responsible for this significant difference. Figures 4.7(a) and 4.7(b) give the rescaled images of $\hat{\gamma}_2(\mathbf{t}) - \hat{\gamma}_8(\mathbf{t})$ and $\hat{\gamma}_1(\mathbf{t}) - \hat{\gamma}_8(\mathbf{t})$, respectively.

5. CONCLUDING REMARKS AND POSSIBLE EXTENSIONS

We considered the testing problem in a general functional analysis of variance model and derived asymptotically optimal (minimax) nonadaptive and adaptive functional hypothesis testing procedures for the main effect and the interactions. The FANOVA decomposition of "ideal" Gaussian white noise processes allowed one to present a wide variety of models in the same format, which facilitates the application of general nonparametric testing procedures to assess the nature of the underlying mean function. An important characteristic of the developed functional hypothesis testing methodology is that it allows one to perform a similar analysis for various types of hypotheses in FANOVA models. Moreover, the resulting procedures are computationally inexpensive and can be easily implemented.

The proposed approach differs from the smoothing spline analysis of variance methodology in FANOVA models (e.g., Wahba *et al.*, 1995; Stone *et al.*, 1997; Lin, 2000; Gu, 2002) or the projection method of Huang (1998). Indeed, the main goal of the above techniques is the *estimation* of the FANOVA model components while the proposed methodology is focused on *functional hypothesis testing*. It also differs from other hypothesis testing procedures in functional data analysis developed recently (e.g., Faraway, 1997; Dette & Derbort, 2001) that treat curved data as a multivariate vector and adopt traditional analysis of variance with various initial dimensionality-reduction techniques. The proposed testing methodology is much closer in spirit to the overall wavelet-based adaptive Neyman's test of Fan (1996) and Fan & Lin (2000). Unlike other approaches to hypothesis testing in FANOVA models, we establish the asymptotic optimality of the proposed functional hypothesis testing procedures.

We would like to finish the paper by pointing at several possible extensions. As we have briefly mentioned in Section 4.3, the proposed functional hypothesis tests can be straightforwardly adapted to testing *contrasts* between the interactions γ_i.

An interesting and practically important extension of model (1) for $d = 1$ is a model of the form

$$dY(s,t) = (m_0 + a(s) + \mu(t) + \gamma(s,t))\; dt\; ds + \epsilon\; dW(s,t), \quad (s,t) \in [0,1]^2,$$

where W is a two-dimensional standard Wiener process, i.e. when both predictors are continuous. We are interested in testing the null-hypotheses of the form: $H_0: \ a \equiv 0$, $H_0: \ \mu \equiv 0$, $H_0: \ \gamma \equiv 0$. Note that the latter essentially corresponds to testing the additivity of the response function $m(s,t)$. Applying the two-dimensional periodic wavelet transform to the data, a specific structure of the matrix V of the resulting wavelet coefficients implies that testing $H_0: \ a \equiv 0$ or $H_0: \ \mu \equiv 0$ will be based only on the first row or the first column of the matrix V respectively, while the remaining (major) part of the matrix of coefficients should be used to test the interaction γ. The extension of the proposed functional hypothesis testing procedures to this case is quite straightforward.

Thresholding in (18) can be possibly performed by other methods than that of Spokoiny (1996), which was adopted in this paper. In particular, in view of recent results in quadratic functional estimation (e.g., Gayraud & Tribouley, 1999; Laurent & Massart, 2000; Johnstone, 2001) we believe it can be performed by grouping empirical wavelet coefficients within each resolution level in a block and using thresholding blockwise rather than individually. It will be also interesting to investigate how *data-driven* thresholding procedures like SURE (see Donoho & Johnstone, 1994, 1995) or FDR (see Abramovich & Benjamini, 1995, 1996), developed in the context of function estimation can be adapted within the functional hypothesis testing framework. This could improve the finite sample properties of the tests.

All the above are avenues for further research that hope will be addressed in the future.

ACKNOWLEDGEMENTS

Felix Abramovich and Anestis Antoniadis were supported by 'Project IDOPT, INRIA-CNRS-IMAG' and 'Project AMOA, IMAG'. Theofanis Sapatinas was supported by a 'University of Cyprus Research Grant'. Brani Vidakovic was supported by the NSF grant DMS 0004131 at Georgia Institute of Technology. Felix Abramovich and Theofanis Sapatinas would like to thank Anestis Antoniadis for excellent hospitality while visiting Grenoble to carry out this work.

The authors would like to thank P. Kinahan and D. Noll for sharing the PET-fMRI data used in Section 4.3. The corresponding compressed data set (1Mb) is available for academic purposes at `ftp://ftp.fil.ion.ucl.ac.uk/spm/data/PET_motor.tar.gz`.

APPENDIX

Here we prove Proposition 3.1 and the assertion (25).

PROOF OF PROPOSITION 3.1 When $1 \leq p < 2$ this is exactly the test proposed by Spokoiny (1996) and the proof follows directly from his results. Thus, we only consider here the case $p \geq 2$.

First, let $p \geq 2$ and $sp > 1$. Note that in this case $s'' = s$ and the condition $s - \frac{1}{2p} + \frac{1}{4} > 0$ holds automatically. Regardless of the true hypothesis, one always has

$$e(j(s)) = \mathbb{E}\left(T(j(s))\right) = \|P_{V_{j(s)}}(f - \int_0^1 f(t)dt)\|_2^2,$$

where $P_{V_{j(s)}}$ denotes the orthogonal projector onto the approximation space $V_{j(s)}$ of the multiresolution analysis (see Mallat, 1989). Using standard results for noncentral chi-squared distributions (e.g., Johnstone, 2001) we have

$$v^2(j(s)) = \mathrm{Var}(T(j(s)) = 2\eta^4 2^{j(s)} + 4\eta^2 e(j(s)). \tag{26}$$

The test statistic $T(j(s))$ is a sum of $j(s)$ independent, squared-integrable random variables and, since $j(s) \to \infty$ as $\eta \to 0$, by the central limit theorem, $T(j(s))$ is asymptotically normal. Moreover, note that when the null hypothesis is true, $e(j(s)) = 0$ and $v^2(j(s)) = v_0^2(j(s))$, and therefore, the test ϕ^* given in (23) is asymptotically of significance level α.

Let $\beta > 0$, denote by $\beta(\phi^*, f) = P_f(\phi^* = 0)$ the Type II error, and let $\beta(\phi^*, \rho) = \sup_{(f - \int_0^1 f(t)dt) \in \mathcal{F}(\rho)} \beta(\phi^*, f)$ be the probability of Type II error for the composite alternative $H_1 : (f - \int_0^1 f(t)dt) \in \mathcal{F}(\rho)$. It is straightforward to see that, for any specific f within the alternative, one has

$$\beta(\phi^*, f) = \Phi\left(\frac{v_0(j(s))}{v(j(s))} z_{1-\alpha} - \frac{e(j(s))}{v(j(s))}\right),$$

where Φ is the cumulative distribution function of a $N(0,1)$ random variable. Set $\kappa(j(s)) = \frac{v_0(j(s))}{v(j(s))}$. Since $v(j(s)) \geq v_0(j(s))$, $\kappa(j(s))$ is bounded above by 1. Hence, as $\eta \to 0$, the asymptotic behavior of $\beta(\phi^*, f)$ depends only on the ratio of squared bias to standard deviation, $\frac{e(j(s))}{v(j(s))}$. Since $\mathcal{F}(\rho)$ belongs to the Besov ball $B^s_{p,q}(C)$ with $p \geq 2$ and $sp > 1$,

$$\sum_{j \geq j(s)} \sum_{k=0}^{2^j} \theta^2_{jk} \leq c_0 2^{-2sj(s)},$$

for some constant c_0 (e.g., Härdle *et al.*, 1998) and, therefore, for any f within the alternative set

$$e(j(s)) \geq (\|f\|_2^2 - c_0 2^{-2sj(s)}) \geq (\rho^2 - c_0 2^{-2sj(s)}).$$

¿From (26) one has

$$v^2(j(s)) \geq 2^{j(s)+1}\eta^4 + 4\eta^2(\rho^2 - c_0 2^{-2sj(s)}).$$

Thus, for $j(s) = \frac{2}{4s+1} \log_2 (C\eta^{-2})$ and the optimal rate of testing $\rho(\eta) = \eta^{4s/(4s+1)}$, one can verify that there exists a constant c_β such that

$$\lim_{\eta \to \infty} \inf_{(f - \int_0^1 f(t)dt) \in \mathcal{F}(c_\beta \rho(\eta))} \frac{e(j(s))}{v(j(s))} > \tilde{c}_\beta,$$

where $\tilde{c}_\beta > 0$ satisfies $\Phi(z_{1-\alpha} - \tilde{c}_\beta) = \beta$ and, hence, $\tilde{c}_\beta = z_{1-\alpha} + z_{1-\beta}$. This shows that the test ϕ^* is indeed asymptotically minimax.

Proof of the assertion (25) Recall again that, since $p \geq 2$, we have $s''_{\max} = s_{\max}$ and, therefore, $j_{\min} = \frac{2}{4s_{\max}+1} \log \eta^{-2}$. To prove that the test (25) is asymptotically adaptive minimax and uniformly consistent, we need to show that

$$\alpha(\phi^*_\eta) = o_\eta(1),$$

and

$$\sup_{\mathcal{T}} \beta(\phi^*_\eta, c\rho(\eta t_\eta)) = o_\eta(1),$$

for some constant $c > 0$. As we have mentioned in the proof of Proposition 3.1, under the null hypothesis, for every $j(s) = j_{\min}, \ldots, j_\eta$, $T(j(s))/v_0(j(s))$ are asymptotically $N(0,1)$ random variables (though dependent) and applying the well known extreme value results for Gaussian random variables (e.g., Leadbetter, *et al.*, 1986) we have

$$\alpha(\phi^*_\eta) = P_{H_0}\left\{ \max_{j_{\min}\le j(s)<j_\eta} \left\{ \frac{T(j(s))}{\sqrt{v_0^2(j(s))}} \right\} > \sqrt{2\ln\ln\eta^{-2}} \right\} \to 0, \quad \text{as} \quad \eta \to 0.$$

Choose now any set of parameters $(s,p,q,C) \in \mathcal{T}$. Note that $\frac{1}{p} < s < s_{\max}$. For the chosen set define $j^*(s)$ by

$$2^{-j^*(s)} = (\eta t_\eta)^{4/(4s+1)}.$$

Then for any f within the alternative set we have

$$\begin{aligned} P_f\left\{ \max_{j_{\min}\le j(s)<j_\eta} \left\{ \frac{T(j(s))}{\sqrt{v_0^2(j(s))}} \right\} \le \sqrt{2\ln\ln\eta^{-2}} \right\} &\le P_f\left\{ \frac{T(j^*(s))}{\sqrt{v_0^2(j^*(s))}} \le \sqrt{2\ln\ln\eta^{-2}} \right\} \\ &\le \Phi\left(\sqrt{2\ln\ln\eta^{-2}} - \frac{e(j^*(s))}{v(j^*(s))} \right) \qquad (27) \end{aligned}$$

Repeating the arguments of the proof of Proposition 3.1 and substituting $c\rho(\eta t_\eta)$ and $j^*(s)$ in (27), the straightforward calculus yields

$$\frac{e(j^*(s))}{v(j^*(s))} = O\left(t_\eta^2\right) = O\left(\sqrt{2\ln\ln\eta^{-2}}\right), \qquad (28)$$

where one can always find a constant c such that the ratio of squared bias to standard deviation in (28) is larger than $\sqrt{2\ln\ln\eta^{-2}}$. Thus, for this c, the probability of Type II error in (27) will tend to zero for any f and any specific set of parameters within $\mathcal{T}$.

Finally, note that the above proofs still hold for $v_0^2(j(s))$ and $v^2(j(s))$ multiplied by r^2 that appears in testing the interaction component (see Section 3.2.3).

REFERENCES

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

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

Abramovich, F. & Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients. *Comput. Statist. Data Anal.*, **22**, 351–361.

Amato, U. & Antoniadis, A. (2001). Adaptive wavelet series estimation in separable nonparametric regression models. *Statist. Comput.*, **11**, 373–394.

Antoniadis, A. (1984). Analysis of variance on function spaces. *Math. Operationsforsch. u. Statist., Ser. Statist.*, **15**, 59–71.

Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *J. Ital. Statist. Soc.*, **6**, 97–144.

Antoniadis, A. & Fan, J. (2001) Regularization of wavelets approximations (with discussion). *J. Am. Statist. Assoc.*, **96**, 939–967.

Antoniadis, A. & Pham, D.T. (1998). Wavelet regression for random or irregular design. *Computat. Statist. Data Anal.*, **28**, 353–369.

Antoniadis, A., Bigot, J. & Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *J. Statist. Soft.*, **6**, Issue 6, 1–83.

Bary, D. & Hartigan, J. (1990). An omnibus test for departures from constant mean. *Ann. Statist*, **18**, 1340–1357.

Brown, L.D. & Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, **24**, 2384–2398.

Brumback, B. & Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Am. Statist. Assoc.*, **93**, 961–983.

Buckley, M. (1991). Detecting a smooth signal: optimality of cusum based procedures.

Biometrika, **78**, 253–262.

Cai, T.T. & Brown, L.D. (1998). Wavelet shrinkage for nonequispaced samples. *Ann. Statist.*, **26**, 1783-1799.

Chen, J.C. (1994). Testing for no effect in nonparametric regression via spline smoothing techniques. *Ann. Inst. Statist. Math.*, **45**, 251–265.

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

Dette, H. & Derbort, S. (2001). Analysis of variance in nonparametric regression models. *J. Multivariate Anal.*, **76**, 110–137.

Dette, H., Munk, A. & Wagner, T. (1998). Estimating the variance in nonparametric regression – what is a reasonable choice? *J. R. Statist. Soc. B*, **60**, 751–764.

DeVore, R.A. & Lorentz, G.C. (1993). *Constructive Approximation.* New York: Springer Verlag.

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

Donoho, D.L. & Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Assoc.*, **90**, 1200–1224.

Donoho, D.L. & Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**, 879–921.

Donoho, D.L. & Johnstone, I.M. (1999). Asymptotic minimaxity of wavelet estimators with sampled data. *Statist. Sinica*, **9**, 1–32.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. R. Statist. Soc. B*, **57**, 301–337.

Efromovich, S. & Low, M. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.*, **24**, 1106–1125.

Ermakov, M.S. (1990). Minimax detection of a signal in a Gaussian white noise. *Theory Probab. Appl.*, **35**, 667–679.

Eubank, R.L. (2000). Testing for no effect by cosine series methods. *Scand. J. Statist.*, **27**, 747–763.

Eubank, R.L. & LaRiccia, V.N. (1993). Testing for no effect in non-parametric regression. *J. Statist. Planning Inference*, **36**, 1–14.

Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Am. Statist. Assoc.*, **91**, 674–688.

Fan, J. & Lin, S.-K. (1998). Test of significance when data are curves. *J. Am. Statist. Assoc.*, **93**, 1007–1021.

Faraway, J.J. (1997). Regression analysis for a functional response. *Technometrics*, **39**, 254–261.

Gayraud, G. & Tribouley, K. (1999). Wavelet methods to estimate an intergrated quadratic functional: adaptivity and asymptotic law. *Statist. Probab. Lett.*, **44**, 109–122.

Gu, C. (2002). *Smoothing spline ANOVA models.* New York: Springer-Verlag.

Hall, P., Kay, J.W. & Titterington, D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, **77**, 521–528.

Hall, P., Kay, J.W. & Titterington, D.M. (1991). On estimation of noise variance in two-dimensional processing. *Adv. Appl. Probab.*, **23**, 476–495.

Härdle, W., Kerkyacharian, G., Pikard, D. & Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications.* Lecture Notes in Statistics **129**, New York: Springer-Verlag.

Horowitz, J.L. & Spokoiny, V.G. (1999). An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica*, **69**, 599–631.

Huang, J.Z. (1998). Projection estimation in multiple regression with application to functional ANOVA models. *Ann. Statist.*, **26**, 242–272.

Huang, H.-C. & Cressie, N. (2000). Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data. *Technometrics*, **42**, 262–276.

Ibragimov, I.A. & Khasminskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* New York: Springer-Verlag.

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

Ingster, Yu.I. (1993a). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.*, **2**, 85–114.

Ingster, Yu.I. (1993b). Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist.*, **2**, 171–189.

Ingster, Yu.I. (1993c). Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.*, **2**, 249–268.

Ingster, Yu.I. & Suslina, I.A. (1998). Minimax signal detection for Besov balls and bodies. *Problems Inform. Transmission*, **34**, 56–68.

Ingster, Yu.I. & Suslina, I.A. (2000). Minimax nonparametric hypothesis testing for ellipsoids and Besov bodies. *ESAIM: Probab. Statist.*, **4**, 53–135.

Kinahan, P.E. & Noll, D.C. (1999). A direct comparison between whole-brain PET and BOLD fMRI measurements of single-subject activation response. *Neuroimage,* **9**, 430–438.

Johnstone, I. (2001). Thresholding for weighted χ^2. *Statist. Sinica*, **11**, 691–704.

Kovac, A. & Silverman, B.W. (2000). Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *J. Am. Statist. Assoc.*, **95**, 172–183.

Laurent, B. & Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, **28**, 1302–1338.

Leadbetter, M.R, Lindgren, G. & Rootzén, H. (1986). *Extremes and Related Properties of Random Sequences and Processes.* New York: Springer-Verlag.

Lepski, O.V. & Spokoiny, V.G. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, **5**, 333–358.

Lin, Y. (2000). Tensor product space ANOVA models. *Ann. Statist.*, **28**, 734–755.

Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pat. Anal. Mach. Intel.*, **11**, 674–693.

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

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

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

Raz, J. (1990). Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach. *J. Am. Statist. Assoc.*, **85**, 132–138.

Spokoiny, V.G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.*, **24**, 2477–2498.

Spokoiny, V.G. (1998). Adaptive and spatially adaptive testing of nonparametric hypothesis hypothesis. *Math. Methods Statist.*, **7**, 245–273.

Stone, C.J., Hansen, M., Kooperberg, C. & Truong, Y. (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion). *Ann. Statist.*, **25**, 1371–1470.

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

Wahba, G., Wang, Y., Gu, C., Klein, R. & Klein, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin epidemiological study of diabetic retinopathy. *Ann. Statist.*, **23**, 1865–1895.