# Wavelet Functional ANOVA, Bayesian False Discovery Rate, and Longitudinal Measurements of Oxygen Pressure in Rats

GARY L. ROSNER[1]
AND
BRANI VIDAKOVIC[2]

**Abstract**

Linear models can be functional in terms of independent or response variables or both. In functional ANOVA-type models often used to model longitudinal measurements and general time series, however, all components have a functional form. One of the main problems in inference using such models is the intrinsic dependence in "time" that makes pointwise inference difficult.

We propose performing the inference in the wavelet domain instead of the time domain. Transformations by orthogonal wavelets preserve the structure of the linear model and, at the same time, decorrelate the data.

The proposed methodology is applied to longitudinal measurements from experiments measuring oxygen pressure in tumor-bearing rats.

**KEY WORDS:** Wavelet ANOVA; Shrinkage; Adaptivity; Denoising; Bayesian Hierarchical Models.

## 1 Introduction

In conventional statistical practice, an observation is usually a number or a vector. But in many situations, observed values are curves or vectors of curves. Prototypical examples are growth curves (e.g., measurements of height and weight in children at particular age times), brain potentials, and a variety of responses in biological, chemical, and geophysical measurements. A vibrant research in this area is summarized in the monograph by Ramsay and Silverman (1997).

Two characteristics are common to any functional data analysis (FDA): *a strong link with the multivariate statistical paradigm* and *the need for regularization.* The strong link with multivariate statistics arises from the fact that methods, such as principal component analysis, multivariate linear modeling, canonical correlation analysis, etc. can be applied within the functional data analysis framework. Function values $f(t)$, where $t$ is continuous, are observed at discrete time points. In the process of analysis, the integrals are replaced by discrete quadrature-based approximations, methods and the final conclusions are usually in terms of weighted multivariate analysis measures.

Regularization in FDA consists of assuming a particular class of smooth functions for the estimators. The implementation is carried out in various ways. For example, one can penalize roughness as part of the fitting criteria or use particular representations that have inherent smoothness (e.g., splines, wavelets, neural networks). Often there are links between magnitudes of coefficients in a particular representation and the regularity features of the represented objects (wavelets, pursuit methods).

In making an inference in longitudinal functional data models, two problems are of particular concern for the inference maker. The most important problem is dimensionality. Explanatory variables are sampled curves, and

[1]Gary Rosner is an Associate Professor, Institute of Statistics and Decision Sciences, Duke University, and Biometry Division, Community and Family Medicine, Duke University Medical Center, 207 Hanes House, Box 3958, Durham, NC 27710.

[2]Brani Vidakovic is an Assistant Professor, Institute of Statistics and Decision Sciences, Duke University, Old Chem 223-B, P.O. Box 90251, Durham, NC 27708.

depending on the rate of sampling, the number of variables can exceed the number of responses by factors of tens. Sensible dimension reduction here is of major interest. Connected with the problem of dimensionality is the problem of dependence. The responses from design points close in time necessarily are dependent and should be modeled as such. Even if the errors are assumed independent, the other components of the model are highly correlated in time. A standard answer to these two problems is to make inference in the *domain of principal components*, see Ramsay and Silverman (1997).

The emerging wavelet-based methodologies are capable of successfully dealing with both of the problems simultaneously. In the next chapter we describe the statistical model and discuss its wavelet counterpart.

# 2 The Model, Methodology, and Simulational Example

The functional ANOVA (FANOVA) model has been utilized by many authors. It is a particular case of a general functional linear model. For example, Ramsay *et al.* (1996) use it to model lip motion from the acoustical data.

We give a definition that is in fact "functionalized" version of a standard one way ANOVA formulation, and remark that variety of more complex models can be functionalized and dealt with in a semi-parametric or nonparametric fashion.

Suppose that for any fixed $t \in T \subset \mathbb{R}$, the observations $\mathbf{y}$ are modeled by a fixed effect ANOVA model,

$$y_{il}(t) = \mu(t) + \alpha_i(t) + \epsilon_{il}(t),\ i = 1, \ldots, p,\ l = 1, \ldots, n_i;\ \sum_{i=1}^{p} n_i = n, \tag{1}$$

where $\epsilon_{il}(t)$ are independent $\mathcal{N}(0, \sigma^2)$ errors. On the other hand, when $i$ and $l$ are fixed we assume that functions $\mu(t)$ and $\alpha_i(t)$ are in $\mathbb{L}_2(T)$, and that $\epsilon_{il}(t)$ is an integrated Wiener measure. To ensure identifiability of treatment functions $\alpha_i$, it is standardly imposed:

$$\int |\sum_i n_i \alpha_i(t)| dt = 0. \tag{2}$$

In the rest of the paper we will assume that measurements $\mathbf{y}$ are taken at equidistant times $t_m$,

$$t_m = m2^{-N},\ 1 \le m \le 2^N,$$

and that $N$ is chosen as a power of 2.

The standard least square estimators for $\mu(t)$ and $\alpha_i(t)$

$$\begin{aligned} \hat{\mu}(t) &= \bar{y}_{..}(t) = \frac{1}{n}\sum_{i,l} y_{il}(t), \\ \hat{\alpha_i}(t) &= \bar{y}_{i.}(t) - \bar{y}_{..}(t). \end{aligned}$$

where $\bar{y}_{i.}(t) = \frac{1}{n_i}\sum_l y_{il}(t)$, are obtained by minimizing the discrete version of LMSSE (for example, Ramsay and Silverman, 1997, p. 141),

$$LMSSE = \sum_t \sum_{i,l} [y_{il}(t) - (\mu(t) + \alpha_i(t))]^2, \tag{3}$$

subject to the discrete version of constraint (2), $(\forall t)\ \sum_i n_i \alpha_i(t) = 0$.

The fundamental ANOVA identity becomes a functional identity,

$$\mathrm{SST}(t) = \mathrm{SSTr}(t) + \mathrm{SSE}(t), \tag{4}$$

with $\mathrm{SST}(t) = \sum_{i,l}[y_{il}(t) - \bar{y}_{..}(t)]^2$, $\mathrm{SSTr}(t) = \sum_i n_i [y_{i.}(t) - \bar{y}_{..}(t)]^2$, and $\mathrm{SSE}(t) = \sum_{i,l}[y_{il}(t) - \bar{y}_{i.}(t)]^2$.

For each $t$, the function

$$F(t) = \frac{\mathrm{SSTr}(t)/(p-1)}{\mathrm{SSE}(t)/(n-p)} \tag{5}$$

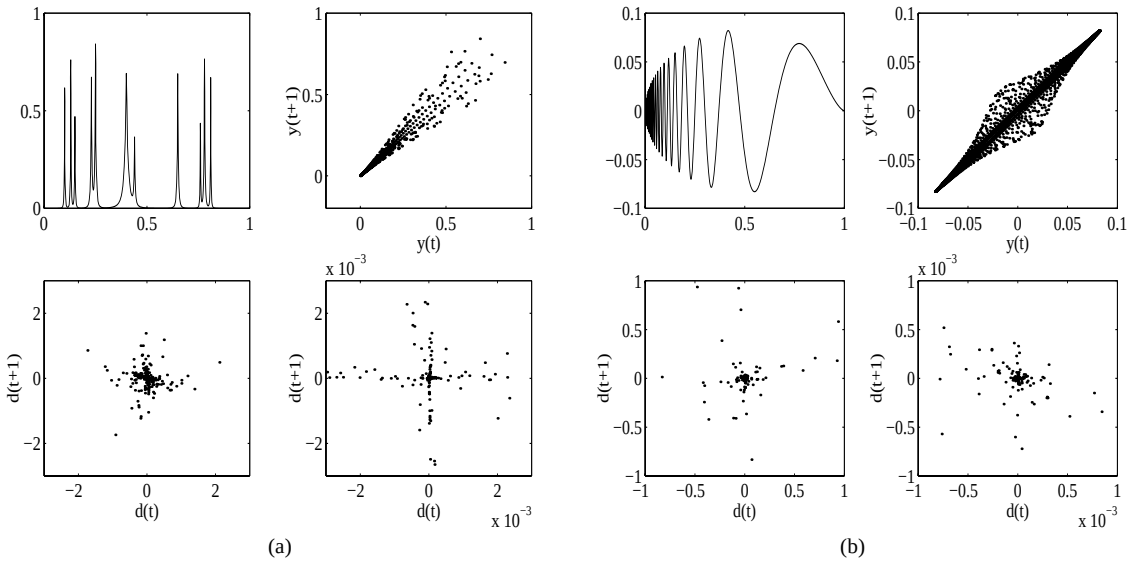


Figure 1: Two standard test-signals of Donoho-Johnstone (bumps and doppler). The plot of one-step correlations are presented in the original and in wavelet domains.

is distributed as non-central $F_{p-1,n-p}\left(\frac{\sum_i n_i \alpha_i^2(t)}{\sigma^2}\right)$.

As hinted in the introduction, an automatic pointwise application of the standard ANOVA method in the above method is hindered by dependence and dimensionality problems. Wavelet transformations are known to be whitening transformation, see Example 1 for an intuitive account. Their decorrelation properties are now well explored and quantified, see, for example, Wornell (1996) and references therein. We propose to transform the discretized FANOVA model to the wavelet domain and proceed with inference there. In addition to preserving the structure of the original linear model, observations in the wavelet domain will be almost uncorrelated and suitably "prepared" for dimension reduction by the wavelet thresholding/shrinkage.

**Example 1.** To illustrate the decorrelation property of the wavelet transformations, we explore two standard test functions `bumps` and *doppler* in the time and wavelet domains. The sample size was 2048.

On the upper left corners of Panels (a) and (b) in Figure 1 `bumps` and *doppler* are given. These test signals are our "data", **y.** On the upper right corners, the values of $y(t+1)$ are plotted against $y(t)$. The coefficients of linear correlations in the time domain are 0.9803 and 0.9858, respectively.

After processing these signals with an orthonormal wavelet transformation (Daubechies' least asymmetric wavelet with 4 vanishing moments) the corresponding plots (lower plots in panels (a) and (b), Fig 1) exhibit reduced correlation. The correlation measurements are -0.0301 and 0.1478, respectively.

Let $\mathbf{d}$ be a wavelet transformation of $\mathbf{y}$, $\mathbf{d} = \mathbb{W}\mathbf{y}$. Due to linearity and orthogonality of $\mathbb{W}$,

$$\begin{aligned} d_{il}(j,k) &= \theta_i(j,k) + \epsilon'_{il}(j,k) \\ &= \theta(j,k) + \tau_i(j,k) + \epsilon'_{il}(j,k), \end{aligned} \tag{6}$$

is the wavelet transform of model (1). The discrete times $t_m$ in the wavelet domain are replaced by a standard multiresolution double indexing $(j,k)$, in which $j$ corresponds to a scale level and $k$ corresponds to a location shift.

Of course, $\sum_i \tau_i(j,k) = 0$ for any $(j,k)$. Moreover, the ANOVA estimators in the time domain, and inverse transformations of the estimators in the wavelet domain coincide. The following result is true:

**Result:** Let $\hat{\mu}$ and $\hat{\alpha_i}$ be least-square estimators of $\mu$ and $\alpha_i$. If $\hat{\theta}$ and $\hat{\tau}_i$ are least-square estimators of $\theta$ and $\tau_i$, and

$$\begin{aligned} \tilde{\mu} &= \mathbb{W}^{-1}\hat{\theta}, \\ \tilde{\alpha}_i &= \mathbb{W}^{-1}\hat{\tau}_i,\ i = 1, \ldots, p, \end{aligned}$$

then $\hat{\mu} \equiv \tilde{\mu}$ and $\hat{\alpha}_i \equiv \tilde{\alpha}_i$. This is a consequence of *energy preservation* property of orthogonal wavelet transformations,

$$\sum_t \text{MSE}(t) = \sum_{j,k} \textit{WMSE}(j,k). \tag{7}$$

where MSE and *WMSE* are the mean squares errors in the time and the wavelet pointwise ANOVAs, respectively.

As discussed next, the benefits of carrying out an analysis in the wavelet domain, rather than in the time domain, comes from decorrelation and regularization considerations.

## 2.1 Inference in the Wavelet Domain

Ramsay and Silverman (1997) discuss several approaches for incorporating regularization in functional linear models. Standard methods involve criteria that penalize roughness. Also choosing relatively few basis functions in some basis representation of a model can lead to regular estimators. This second method, sometimes intertwined with particular roughness penalty methods, is especially convenient when models are wavelet-represented. This comes from the fact that wavelets are unconditional bases for a variety of smoothness spaces, and appropriate regularizations can easily be controlled by the selection and shrinkage of wavelet coefficients.

Therefore, before separating functions, we threshold their coefficients. Such thresholding can assume independence models, due to the decorrelating property of wavelet transformations. We will apply the thresholding method proposed in Vidakovic (1998) in which the thresholding is performed by a Bayesian testing of hypotheses. Given the multiple testing problem, we propose a Bayesian version of the *false discovery rate* method of Abramovich and Benjamini (1995). We start with a description of thresholding resulting from Bayesian inference in the wavelet domain.

In the Bayesian framework there can not be any "fixed effects" in the literal sense, since all the parameters are regarded as having a probability distribution. The terms "fixed effects" and "random effects" came from sampling theory-based inference and should not concern a Bayesian. The differences between "fixed" and "random" effects come at the stage of choosing prior distributions for the parameters. According to a conventional practice, *appropriate* priors in the case of "fixed effects" are noninformative, as discussed in Box and Tiao (1973).

Due to the fact that wavelet transformation is orthogonal and linear, the distributional structure in the wavelet domain corresponds to that in the time domain. For example, if $\mathbf{d} = \{d_{1,1}, d_{1,2}, \ldots d_{1,n_1}, d_{2,1}, \ldots d_{p,n_p}\}$ is the vector of wavelet coefficients at position $(j,k)$, then the inference on $\tau_1$, for example, will be made with reference to the models using the "wavelet coefficients" $\bar{d}_{1j_1} - \bar{d}_{..},\ j_1 = 1, \ldots, n_1$.

We threshold (i.e., replace with 0) wavelet coefficients (or linear combinations of wavelet coefficients), if a statistical hypothesis that states that their modeled location is 0, is "accepted."

Testing a precise hypothesis in Bayesian fashion requires a prior that has a point mass component. Otherwise, testing is trivial since any continuous prior density will give a precise hypothesis zero prior (and hence the posterior) probability. For a discussion on testing precise hypotheses in a Bayesian fashion see Berger (1985).

Assume that a typical wavelet coefficient in the model (6) can be modeled as normally distributed

$$[d_{il}|\theta_i, \sigma^2] \sim \mathcal{N}(\theta_i, \sigma^2), \tag{8}$$

where $\theta_i = \theta + \tau_i$. To simplify the notation, the time-scale indexing $(j,k)$ is dropped.

Also in the description of thresholding, we generically denote the observed coefficient or the linear combination of coefficients by $d$ and the corresponding location of by $\theta$.

The marginal likelihood after the variance $\sigma^2$ is integrated out is

$$d|\theta \sim f(d|\theta),$$

After observing $d$, we test the hypothesis $H_0 : \theta = 0$, versus $H_1 : \theta \neq 0$. If the hypothesis $H_0$ is rejected, $d$ is retained in the model; otherwise, it is replaced by 0. The prior distributions for the locations of wavelet coefficients are well explored. The standard choice is a mixture that contains a point mass at 0, or a concentrated continuous distribution

that approximates a point mass at 0. This appropriate prior in the context of wavelet shrinkage was first suggested by Peter Müller (1994, personal communication). See also Chipman et al. (1987), Clyde et al. (1998), and Müller and Vidakovic (1999).

Let

$$\theta \sim \pi(\theta) = \pi_0\delta_0 + \pi_1\xi(\theta), \tag{9}$$

where $\xi(\theta)$ is a prior that describes the spread of $\theta$ when $H_0$ is false, $\delta_0$ is a point mass at 0, and $\pi_0$ and $\pi_1$ are prior probabilities of hypotheses $H_0$ and $H_1$, respectively, $(\pi_0 + \pi_1 = 1)$.

The method proposed in Vidakovic (1998) utilizes the posterior probabilities of hypotheses $H_0$ to threshold *a posteriori* insignificant coefficients.

The posterior probability of the null hypothesis is $P(H_0|d) = (1 + \frac{\pi_1}{\pi_0}\,\frac{1}{B})^{-1}$, where $B = \frac{f(d|0)}{\int_{\theta\neq 0} f(d|\theta)\xi(\theta)d\theta}$ is the Bayes factor in favor of $H_0$.

If[3]

$d|\theta \sim \mathcal{DE}(\theta, \frac{1}{\sqrt{2\mu}})$ and

$$\pi(\theta) = \pi_0\delta_0 + \pi_1\xi(\theta),$$

then the posterior probability of $H_0$ is

$$p = \frac{\pi_0 e^{-c|d|}}{\pi_0 e^{-c|d|} + \pi_1(\Pi_1(c) + \Pi_2(c))}, \tag{10}$$

where $\Pi_1(c) = \int_0^\infty \xi(\theta - d)\; e^{-c\theta}d\theta$ and $\Pi_2(c) = \int_0^\infty \xi(\theta + d)\; e^{-c\theta}d\theta$ are one-sided Laplace transformations of $\xi(\theta - d)$ and $\xi(\theta + d)$, and $c = \sqrt{2\mu}$.

As one may expect, the threshold is sensitive with respect to the choice of $\mu$ which is connected with precision $1/\sigma^2$. In the applications, we worked with a normal prior $\xi(\theta)$ with large variance ($\tau^2 = 100$).

The parameters in $\pi_0$ and $\pi_1$ express the prior probabilities of hypotheses $H_0$ and $H_1$. Since parsimony of the thresholded model is expected, the reasonable choice for $\pi_0 = 1 - \pi_1$ should be close to 1. A reasonable default values are 0.90 or 0.95. Additional adaptivity of the model can be obtained by varying $\pi_0$ level-wise, as done in Clyde et al. (1998).

Since the identity (7) holds the wavelet based functional Anova would make sense in two different cases:

(i) estimate ANOVA coefficients in the wavelet domain in a *non-least-square* fashion. Bayesian versions of pointwise ANOVA's are good since the estimators of the model coefficients are shrinkers. Such shrunk versions of the coefficients are then returned to the time domain by the inverse wavelet transformation, thus yielding regularized estimators of the model components (mean and treatment functions).

(ii) Perform shrinkage of the wavelet coefficients for all functional replications, and then apply ANOVA pointwise.

## 2.2 Model Selection by Bayesian False Discovery Rate

As we discussed in the previous section, thresholding of wavelet coefficients can be viewed as a multiple testing problem. For each wavelet coefficient $d_i = \theta_i + \sigma\epsilon_i$, the hypothesis $H_0 : \theta_i = 0$ is tested against the alternative $H_1 : \theta_i \neq 0$. If the hypothesis $H_0$ is rejected, the coefficient $d_i$ is retained in the model. Otherwise, it is discarded.

For example, the universal threshold can be viewed as a critical value of a test with the level

$$\alpha = P(|d_i| > \sqrt{2\log n}\;\sigma\;|H_0\;) \approx (n\sqrt{\pi\log n})^{-1}.$$

The power of this test against the alternative $H_1 : \theta_i = \theta\;(\neq 0)$ is $O\left(\frac{1}{n\sqrt{\log n}}\right)$.

Universal thresholding is equivalent to a Bonferroni-type procedure. In testing $n$ statistical hypotheses simultaneously, the Bonferroni procedure guarantees that the overall level of the omnibus test is $\alpha$ by setting the levels for the individual hypotheses as $\frac{\alpha}{n}$. For large $n$, the individual levels $\frac{\alpha}{n}$ become unduly small, which results in loss of "strictness" and dissipation of power.

[3]Double exponential marginal likelihood $f(d|\theta) = \frac{1}{2}\sqrt{2\mu}e^{-\sqrt{2\mu}|d-\theta|}$ is obtained from (8) when the prior on $\sigma^2$ is exponential $\pi(\sigma^2|\mu) = \mu e^{-\mu\sigma^2}$.

A way to control such dissipation of power is proposed by Abramovich and Benjamini (1995, 1996). Their proposal is based on the *false discovery rate* (FDR) method of Benjamini and Hochberg (1995).

Here is a brief description. Let $R$ be the number of wavelet coefficients retained in the model. If $S$ of them are correctly kept, then $V = R - S$ are erroneously kept. The random variable $Q = V/R$ expresses the error in such a procedure. The false discovery rate of coefficients is the expectation of $Q$; that is, the expected proportion of coefficients erroneously kept. Following Benjamini and Hochberg (1995), Abramovich and Benjamini (1995) propose maximizing the number of coefficients kept, subject to condition $\mathbb{E}Q \leq \alpha$, for $\alpha$ small.

We propose a Bayesian version of FDR procedure. Suppose that in testing $n$ hypotheses, we obtain a sequence $p_1, p_2, \ldots, p_n$ of their posterior probabilities. Let $p_{(1)}, p_{(2)}, \ldots, p_{(n)}$ be increasingly ordered posterior probabilities, and $q_{(i)} = 1 - p_{(i)}$.

When deciding about retaining the wavelet coefficients in the model by rejecting corresponding null hypotheses $H_0 : \theta = 0$, we control the number of hypotheses that are erroneously rejected, $V$. If the $R$ hypotheses with smallest posterior probabilities are rejected, we require the expectation (with respect to the posterior measure) of $Q = V/R$ not to exceed $\alpha$. Note that

$$\begin{aligned} \mathbb{E}Q &= \frac{1}{R}\sum_{i=0}^{R} iP(\text{Among } R \text{ rejected hypotheses, the number of erroneously rejected is } i) \\ &= \frac{1}{R}\sum_{i=0}^{R} iP_R(i), \end{aligned} \tag{11}$$

where the probabilities $P_R(i)$ can be calculated efficiently as the coefficients with powers $z^i$ in generating polynomial

$$\varphi_R(z) = \prod_{i=0}^{R}(q_{(i)} + p_{(i)}z) = \sum_{i=0}^{R} P_R(i)z^i. \tag{12}$$

Thus, the Bayesian FDR procedure can be summarized as follows:

- **STEP 1.** Find the posterior probabilities $p_{jk}$ of all hypotheses $H_0 : \theta_{jk} = 0$ and order them according to their size.
- **STEP 2.** Fix $\alpha$ small and set $R = 1$.
- **STEP 3.** Increase $R$ by 1. Find $\varphi_R(z)$ using $p_{(1)}, \ldots, p_{(R)}$, and calculate $\mathbb{E}Q$.
- **STEP 4.** If $\mathbb{E}Q \geq \alpha$ then the maximum posterior probability of rejection is $p_{R-1}$. **STOP.**

Otherwise, if $\mathbb{E}Q < \alpha$, return to **STEP 3.**

## 2.3 A simulation example.

To illustrate the proposed methodology we provide a simulational example with known components of the model. The components were Heaviside function $\mu(t) = 5 * h(t - 0.25)$, $t \in [0, 1]$ as the main effect function and 5 treatment effect functions, $\alpha_i(t)$, $i = 1, \ldots, 5$. The functions $\alpha_1(t), -\alpha_4(t)$ are test functions `blocks, bumps, doppler,` and `heavisine`, respectively, (see Donoho and Johnstone, 1994), while $\alpha_5(t) = -\frac{1}{n_5}(n_1\alpha_1(t) + \cdots + n_4\alpha_4(t))$ because of the constraint (2).

The number of observations was $n = 66$ and the treatment sample sizes were $n_1 = 14, n_2 = 11, n_3 = 6, n_4 = 15$, and $n_5 = 20$.

Each observation from the treatment $i$ is a sequence of $N = 512$ values of $\mu_i(t) + \epsilon(t) = \mu(t) + \alpha_i(t) + \epsilon(t)$ taken at equally spaced points $t \in [0, 1]$. The wavelet selected is Symmlet with 5 vanishing moments. All functional components of the model are rescaled so that the signal-to-noise ratio [SNR] of $\mu + \alpha_i$ (signal) to $\epsilon$ (noise) is 5 and the size of noise is 1.

To get estimators of $\mu(t)$ and $\alpha_i(t)$, we applied the Bayesian FDR procedure with level $\alpha = 0.1$, and posterior probabilities given by (10) with $\mu = (\hat{\sigma^2})^{-1}$ where $\hat{\sigma^2}$ is a robust estimator of the variance (exibited using the level of finest detail in the corresponding wavelet decomposition).

In Table 2.3, the first column of numbers are squared errors of the time-domain Anova estimators while in the second column are the errors of the wavelet processed estimators. For example $\sum_t[\hat{\mu}(t) - \mu(t)]^2 = 7.6983, \sum_t[\tilde{\mu}(t) - \mu(t)]^2 = 0.5327$. The errors are exact since we know the true value of function.

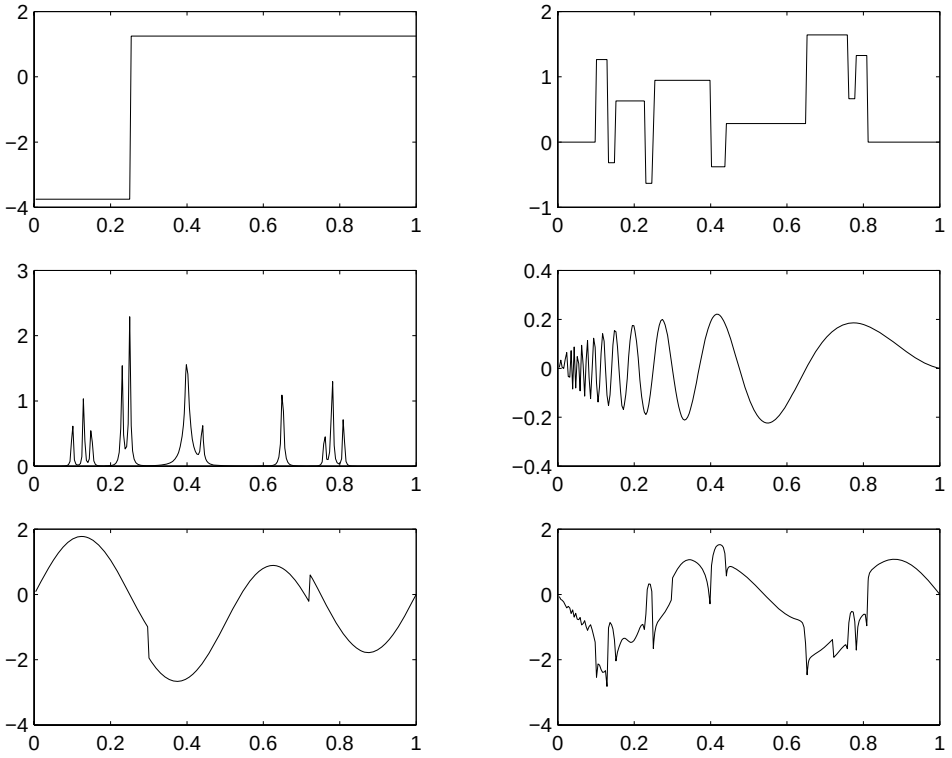

Figure 2: The mean function, $\mu(t)$, five treatment-effect functions $\alpha_1(t) - -\alpha_5(t)$ (four are standard Donoho-Johnstone test functions and the fifth is a "compensator" function necessary to satisfy the condition (2).

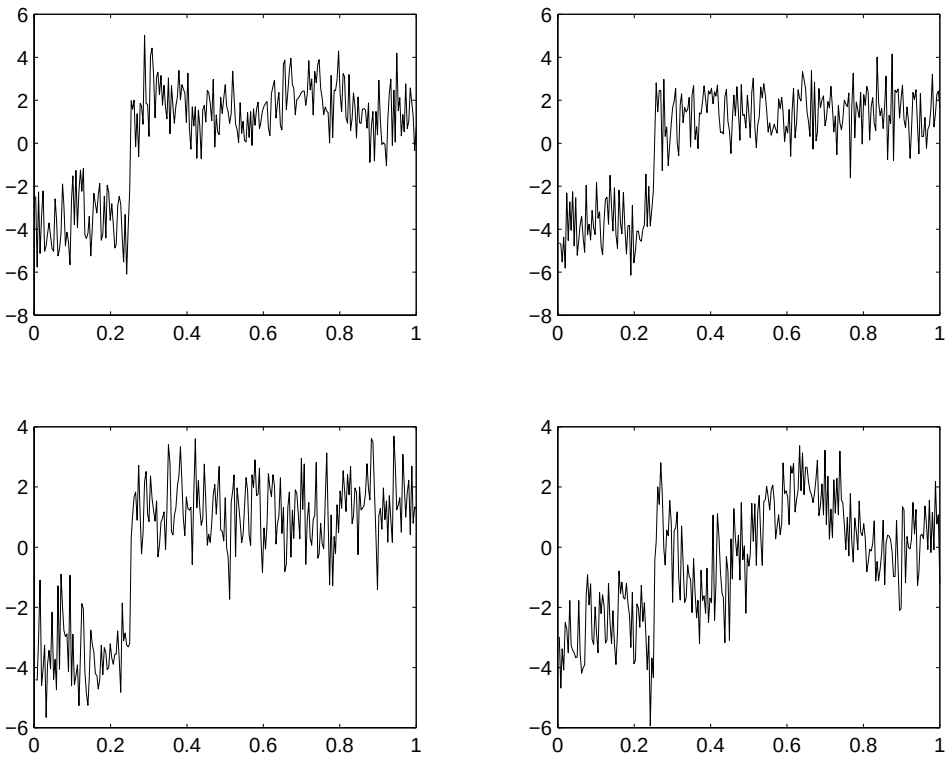


Figure 3: Typical "observations." Each observation is a sum of mean function $\mu(t)$, one of treatment-effects $\alpha_i(t)$, and the noise $\epsilon(t)$.

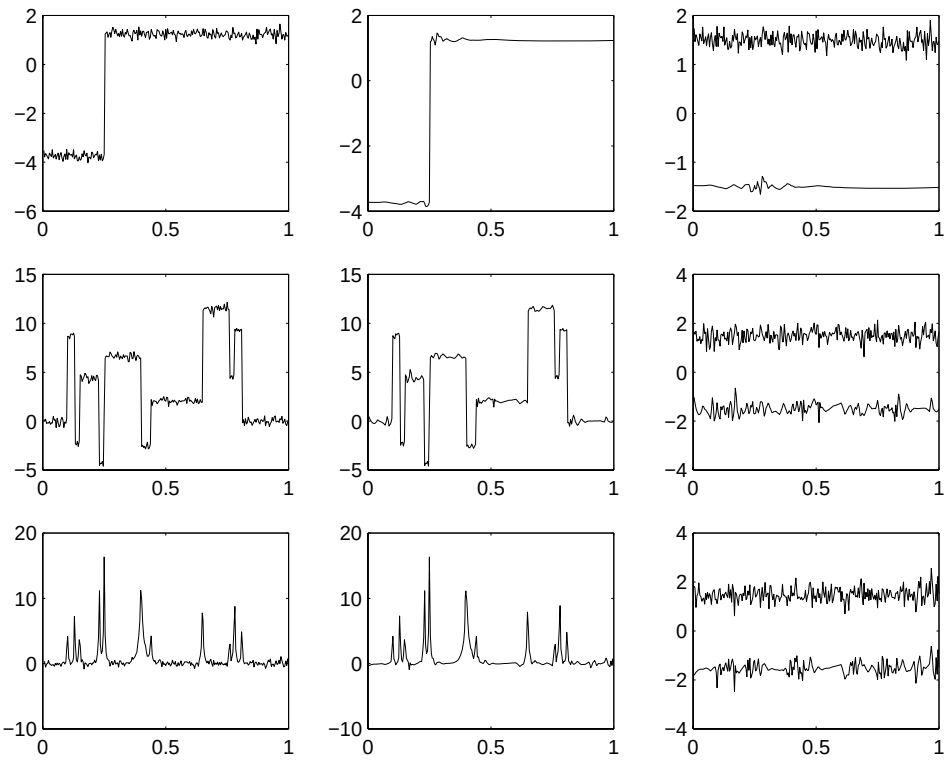

Figure 4: Reconstructions by an Anova procedure in time domain (Column 1), wavelet domain with shrinkage (Column 2), and the errors (Column 3). The errors are translated for +1.5 units for the time domain and -1.5 units for the wavelet domain.

| function | time domain LS estimate | wavelet domail LS estimate after shrinkage |
|---|---|---|
| $\mu$ | 7.6983 | 0.5327 |
| $\alpha_1$ | 29.6217 | 20.8990 |
| $\alpha_2$ | 40.7046 | 25.2157 |
| $\alpha_3$ | 76.0097 | 12.3358 |
| $\alpha_4$ | 28.2213 | 5.1684 |
| $\alpha_5$ | 17.9339 | 15.5773 |
| total | 200.1895 | 79.7289 |

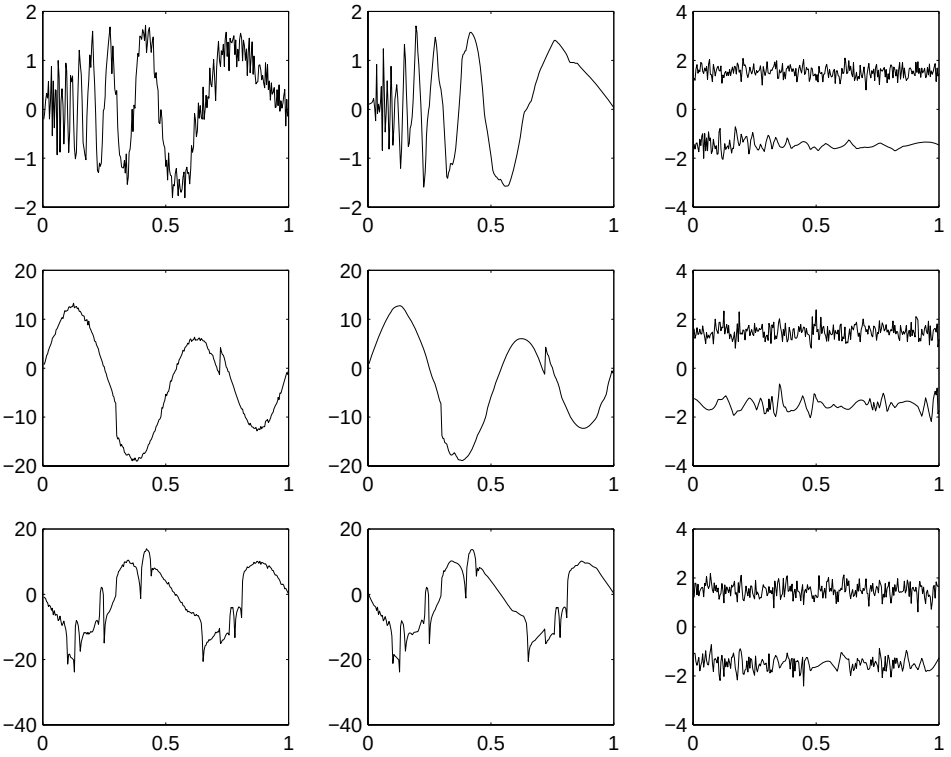

Figure 5: Continuation of Figure 4.

### 2.4 Some Practical Guidelines

## 3 An Application to an Experiment in Tumor Physiology

Experiments carried out *in vitro* with tumor cell lines have demonstrated that tumor cells respond to radiation and anti-cancer drugs differently, depending on the environment. In particular, available oxygen is important. Efforts to increase the level of oxygen within tumor cells have included laboratory rats with implanted tumors breathing pure oxygen. Unfortunately, animals breathing pure oxygen may experience large drops in blood pressure, enough to make this intervention too risky for clinical use (Dewhirst, Lanzen, and Braun, 1998).

Mark Dewhirst, Professor at Department of Radiation Oncology at Duke University, sought to evaluate carbogen (95% pure oxygen and 5% carbon dioxide) as a breathing mixture that might improve tumor oxygenation without causing a drop in blood pressure. The protocol called for making measurements on each animal over 20 minutes of breathing room air, followed by 40 minutes of carbogen breathing. The experimenters took serial measurements of oxygen partial pressure ($PO_2$), tumor blood flow (LDF), mean arterial pressure (MAP), and heart rate. Microelectrodes, inserted into the tumors (one per animal) measured $PO_2$ at a particular location within the tumor throughout the study period. Two laser Doppler probes, inserted into each tumor, provided measurements of blood flow. An arterial

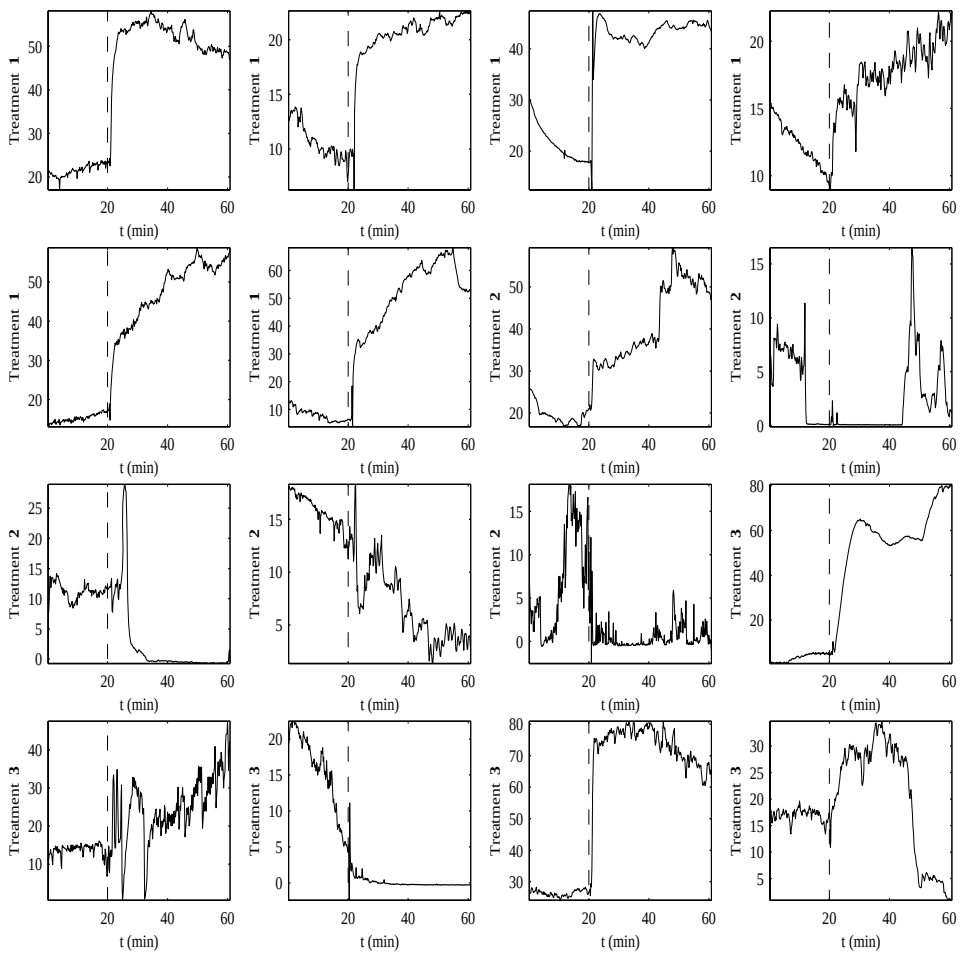


Figure 6: $PO_2$ measurements. Notice that despite a variety of functional responses and a lack of a simple parametric model, at the time $t^* = 20'$ the pattern generally changes.

line into the right femoral artery allowed measurement of MAP. Each animal wore a face mask for administration of breathing gases (room air or carbogen). (See Lanzen, Braun, Ong, and Dewhirst, 1998, for more information about these experiments.)

Nine rats had tumors transplanted within the quadriceps muscle (which we will denote by TM). For comparison, the studies also included eight rats with tumors transplanted subcutaneously (TS) and six rats without tumors (N), in whom measurements were made in the quadriceps muscle.

Figure 6 show some the data ($PO_2$) The plots show several features, including an obvious rise in $PO_2$ at the 20-minute mark among some of the animals. No physiologic model exists that would characterize the shapes of these profiles mathematically. The primary study question concerned evaluating the effect of carbogen breathing on $PO_2$. The analysis is made complicated by the knowledge that there may be acute change in $PO_2$ after carbogen breathing starts. The primary question of interest is whether the tumor tissue behaves differently than normal muscle tissue and/or whether tumor implanted subcutaneously responds to carbogen breathing differently than tumor tissue implanted in muscle tissue in the presence of acute jumps in $PO_2$.

The analyses concern inference on change in some physiologic measurements after an intervention. The problem for the data analysis is how best to define "change" to allow for the inferenc edesirei by the invertigators.

From a statistical modeling point of view, the main issues concern building a flexible model for the multivariate time series $y_{ij}$ of responses and to provide for formal inference on the occurrence of change at time $t^*$. "equality" of the $PO_2$ profiles. From the figures it is clear that the main challenge arises from the highly irregular behavior of responses. Neither physiologically considerations nor any exploratory data analysis motivate any parsimonious parametric form.

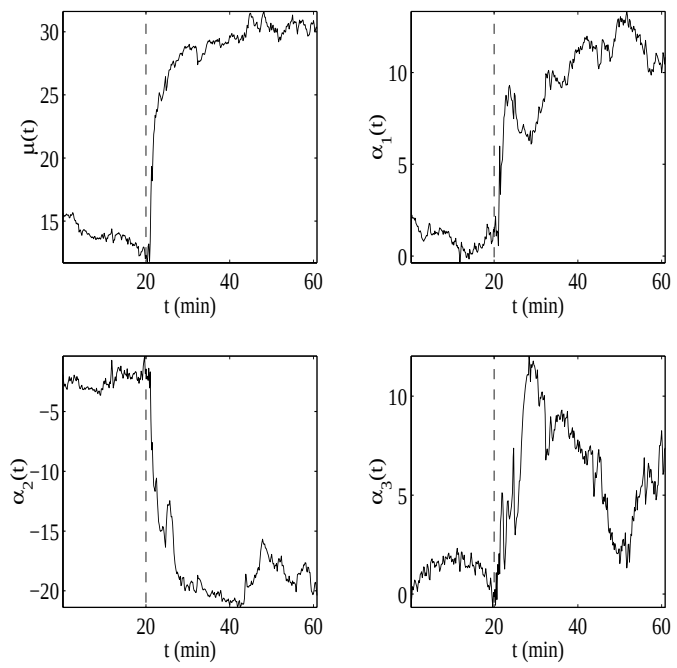


Figure 7: Functional Anova results in the time domain. More to be added...

Different individuals seem to exhibit widely varying response patterns. Still, it is clear from inspection of the data that for some response series a definite change is taking place at time $t^*$.

Figure 7 shows the results of functional ANOVA in the time domain. One can imagine from the figure that adding $\mu(t)$ and $\alpha_2(t)$ will lead to a relatively horizontal expected profile for group $i$, each fitted curve canceling the other to some extent. The wavelet-smoothed expected profiles are shown in Figure 8. The expected profiles for each group are shown in Figure 9, along with the "grand mean" function in the upper left. As expected, $\mu_2(t)$ looks a lot like noise about a horizontal line, whereas the other two groups show a clear jump in $PO_2$ just after the start of carbogen breathing at 20 minutes into the experiment.

Here are some conclusions... • Possible to test: $H_0 : \mu_i(t) = Const.$

• Parseval+Haar $\boxed{\int \mu(t)^2 dt = c_{00}^2 + \sum_{j\geq 0,k} d_{jk}^2}$

if $\mu(t) = c_{00}\phi_{00}(t) + \sum_{j\geq 0;k} d_{jk}\psi_{jk}(t)$.

Brani: We clearly need some statistics relating to the overall ANOVA inference about the equality of the $\mu_i, i = 1, \ldots, p$.

# 4 Discussion

In this article we addressed some benefits of making an inference about functional ANOVA model by considering wavelet domain. Various generalizations are possible: (i) extend inference to various linear functional statistical models, and (ii) consider a variety of shrinkage techniques,

It is also possible to used different wavelet bases for data belonging to different treatments, or select the basis adaptively, from the wavelet packet library.

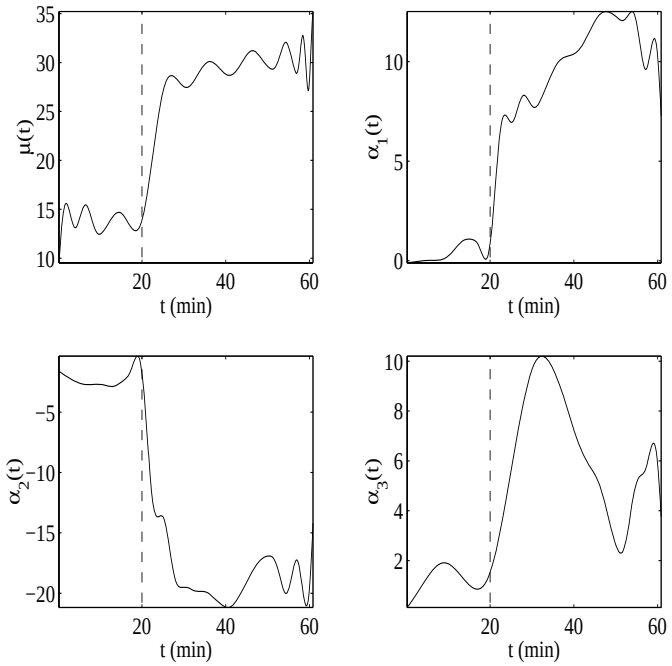


Figure 8: Later.

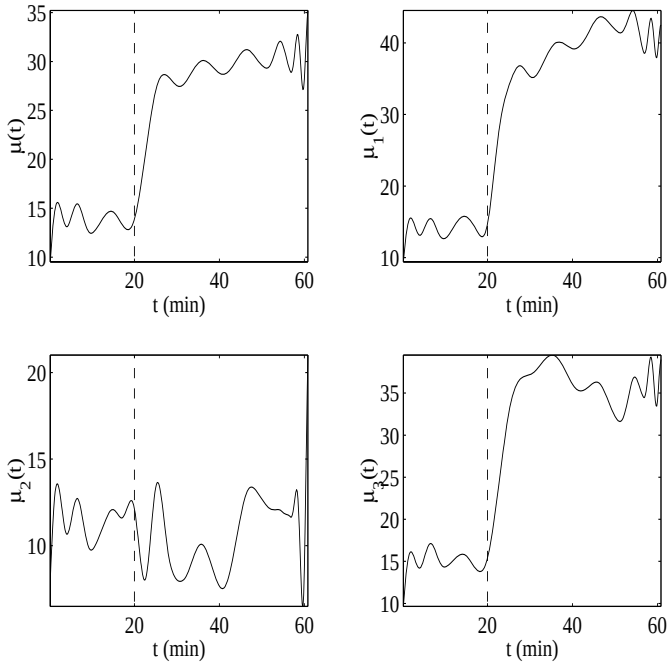


Figure 9: Later.

# References


[1] ABRAMOVICH, F. and BENJAMINI, Y. (1995). Thresholding of wavelet coefficients as a multiple hypotheses testing procedure, In: *Wavelets and Statistics*, Editors A. Antoniadis and G. Oppenheim. Lecture Notes in Statistics, **103**, 5–14. Springer-Verlag, New York.

[2] ABRAMOVICH, F. and BENJAMINI, Y. (1996). Adaptive thresholding of wavelet coefficients. *Comput. Statist. Data Anal.* **22**, 351–361.

[3] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *J. Roy. Statist. Soc. Ser. B.* **57,** 289–300.

[4] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag.

[5] CLYDE, M. A., PARMIGIANI, G., and VIDAKOVIC, B. (1998). Multiple Shrinkage and Subset Selection in Wavelets. *Biometrika*, **85**, 391–402.

[6] DEWHIRST, M. W., BRAUN, R. D., and LANZEN, J. L. (1998). Temporal changes in $pO_2$ of R3230Ac tumors in Fischer-344 rats. *International Journal of Radiation Oncology, Biology, and Physics*, **42,** 723-726.

[7] KNEIP, A. and GASSER, T. (1992). Statistical tools to analyze data representing a sample of curves. *Annals of Statistics*, **20,** 1236–1265.

[8] LANZEN, J. L., BRAUN, R. D., ONG, A. L., and DEWHIRST, M. W. (1998). Variability in blood flow and $pO_2$ in tumors in response to carbogen breathing. *International Journal of Radiation Oncology, Biology, and Physics*, **42,** 855-859.

[9] MÜLLER, P. and VIDAKOVIC, B. (1999). MCMC Methods in Wavelet Shrinkage: Non-Equally Spaced Regression, Density and Spectral Density Estimation, In: *Bayesian Inference In Wavelet Based Models*, Editors P. Müller and B. Vidakovic, Lecture Notes in Statistics,**141**, 187–202, Springer Verlag, New York.

[10] RAMSAY, J. O. and DALZELL, C. J. (1991). Some tools for functional data analysis (with discussion.) *Journal of the Royal Statistical Society, Series B,* **53,** 539–572.

[11] RAMSAY, J.O. and SILVERMAN, B. W. (1997). *Functional Data Analysis*, Springer-Verlag, New York.

[12] RAMSAY, J. O., BOCK, R. D., and GASSER, T. (1995) Comparison of height acceleration curves in the Fels, Zurich and Berkeley growth data. *Annals of Human Biology,* **22,** 413–426.

[13] RAMSAY, J. O. (1995). Principal differential analysis: Data reduction by differential operators. *Journal of the Royal Statistical Society,* **58,** 495–508.

[14] RAMSAY, J. O., MUNHALL, K. G., GRACCO, V. L. and OSTRY, D. J. (1996) Functional data analysis of lip motion. *Journal of the Acoustical Society of America,* **99,** 3718–3727.

[15] RICE, J. A. and 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.

[16] SILVERMAN, B. W. (1995). Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society, Series B.*

[17] WAHBA, G. (1990). *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.

[18] VIDAKOVIC, B. (1998). Nonlinear Wavelet Shrinkage With Bayes Rules and Bayes Factors, *Journal of the American Statistical Association*, **93**, 441, 173–179.

[19] WORNELL, G. W. (1996). *Signal Processing with Fractals: A Wavelet Based Approach*, Prentice Hall, Englewood Cliffs, NJ.



G. L. ROSNER AND B. VIDAKOVIC
INSTITUTE OF STATISTICS
AND DECISION SCIENCES
OLD CHEM BUILDING
DUKE UNIVERSITY
BOX 90251, DURHAM, NC 27708-0251
brani|rosner@stat.duke.edu